DIAL: DRIFT-AWARE CONTINUAL ALIGNMENT FOR IMPLICIT LIDAR LOCALIZATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Implicit LiDAR localizers compress a scene into a metric field stored in network parameters, but this field becomes stale as environments evolve. Repeated pose annotation and full retraining are costly across recurring revisits. Continual updates face two coupled problems: supervision is missing where the historical model fails, and unconstrained adaptation can overwrite a coordinate readout that remains valid elsewhere. We present the first study of continual learning for implicit LiDAR localization. We find that the historical localizer supports both needs: its predictions reveal where the implicit map has changed, while its metric readout preserves valid coordinate associations when old scans are unavailable. DIAL's Temporal Consistency-guided Pseudo-label Pipeline (TCPP) cross-validates retrieval and scene-coordinate evidence over time to recover pseudo poses for changed or unseen frames. Head-Decoupled Encoder Regularization (HDER) freezes the readout and adapts selected encoder blocks when old scans are inaccessible. With historical memory, Memory-Conditioned Rehearsal (MCR) reinstates old-scene constraints; Redundancy-Aware Exemplar Replay (RAER) selects the replay needed as route coverage expands. On NCLT, pseudo labels for about 20% of revisit frames suffice for maintenance. A buffer-based DIAL route updates the localizer for the first NCLT environmental change in about 12 minutes of continual optimization; repeated maintenance then sustains localization across roughly one year of NCLT revisits, while segmented QEOxford approaches full-retraining accuracy. Code will be released.

1 INTRODUCTION

LiDAR localization supports autonomous navigation and robotics in GNSS-denied environments (Yin et al., 2024). Implicit point-cloud localizers are attractive at scale: scene coordinate regression and absolute pose regression encode scene geometry in network parameters (Brachmann et al., 2014; 2016; Wang et al., 2021c; Li et al., 2023), avoiding a large map search at inference and enabling efficient relocalization after training.

This compact representation also makes the network a scene-specific map. As urban geometry changes, its learned scan-to-coordinate mapping becomes stale, and small representation shifts can produce large pose errors. Faster scene fitting (Li et al., 2025a), cheaper trajectory generation (Chen et al., 2025), and retrieval-assisted feed-forward localization (Barroso-Laguna et al., 2026) reduce the cost of building a localizer. Maintaining an existing metric field raises a further question: how can later observations correct changed geometry while preserving coordinates that remain valid? Repeated annotation and full retraining make this question consequential for long-term deployment. We formulate continual implicit LiDAR localization for the first time, treating the learned metric field itself as the object that must evolve while retaining previously valid coordinate associations.

Figure 1 quantifies this maintenance frontier. Without maintenance, PELoc and SGLoc-v2 (Chen et al., 2025; Li et al., 2023) reach aggregate errors of 4.75 m and 3.38 m, respectively. DIAL lowers these errors to approximately 1.9 m in 12 minutes and 1.3 m in 27 minutes of continual optimization. Across later revisits, repeated maintenance keeps the implicit localizer usable as the environment continues to evolve.

DIAL considers two forms of continual implicit localization (Figure 3): changed revisits, where an established route evolves over time, and segmented stages, where a long route is acquired incremen-

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.

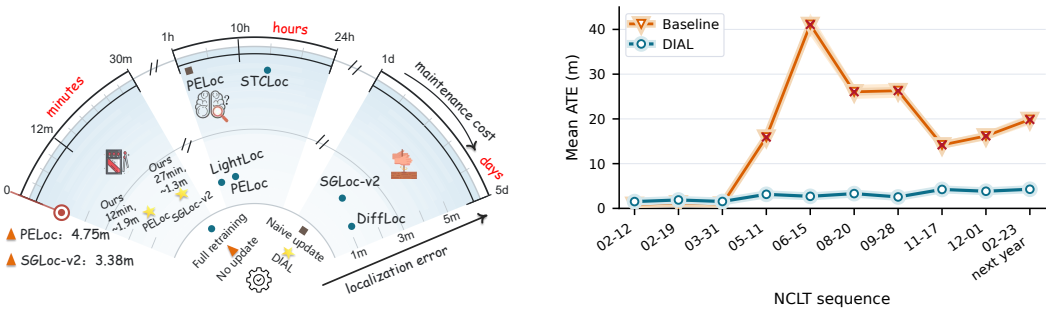


Figure 1: **Maintenance frontier and long-term localization.** Left: across implicit localization algorithms and update strategies, DIAL reduces PELoc error from 4.75 m to approximately 1.9 m in 12 minutes and SGLoc-v2 error from 3.38 m to approximately 1.3 m in 27 minutes of continual optimization, avoiding full retraining. Right: after repeated continual maintenance, DIAL preserves localization across roughly one year of NCLT revisits under scene evolution.

tally with spatial overlap. Figure 2 illustrates the corresponding update dilemma. Without an update, the coordinate field becomes stale around changed or unseen observations. A naive update instead lets gradients from the new revisit rewrite precise historical mappings. Continual maintenance must decide both which observations can supply supervision and which model components can change.

The historical localizer informs both decisions. Its failures are diagnostic: scene coordinate regression (SCR) remains locally precise where the implicit map is valid, whereas retrieval preserves global association under change but has long-tailed metric error. Their temporal behavior helps identify scans that need new supervision and poses that can supply it. The historical readout also retains metric memory. Without old scans, freezing this readout prevents new supervision from directly overwriting historical coordinate associations. Replay supplies explicit constraints on those associations when readout adaptation is needed, particularly as segmented learning extends route coverage.

DIAL implements this policy through four components. The Temporal Consistency-guided Pseudo-label Pipeline (TCPP) combines temporally validated retrieval and SCR evidence with ICP refinement to recover pseudo poses for changed or unseen frames. Head-Decoupled Encoder Regularization (HDER) fixes the metric readout when historical data are absent, and Memory-Conditioned Rehearsal (MCR) restores historical constraints when replay is available. During segmented learning, Redundancy-Aware Exemplar Replay (RAER) removes old scans whose locations are already covered by the current stage. This allocation of supervision and plasticity avoids rebuilding the complete localizer after each change.

Our contributions are summarized as follows:

1. We formulate continual implicit LiDAR localization across changed revisits and segmented route stages. Prediction discrepancies from the historical model expose stale geometry, and its coordinate readout stabilizes the metric field when old data are unavailable.

2. We realize this principle through TCPP, which recovers supervision from complementary retrieval and regression evidence, and through HDER, MCR, and RAER, which preserve historical geometry under different memory and route-overlap conditions.

3. On long-term NCLT and segmented QEOxford, DIAL retains old-scene accuracy, offers 12-minute buffer-based optimization for the first NCLT environmental change, maintains localization for a year, and approaches the full-retraining upper bound on QEOxford.

2 Related Work

LiDAR localization. Outdoor LiDAR localization spans odometry, SLAM, map matching, retrieval, and learning-based regression (Xu & Zhang, 2021; Zheng et al., 2024; Yin et al., 2024). We focus on compact global localization. Regression methods such as LightLoc, PELoc, and DiffLoc (Li et al., 2025a; Chen et al., 2025; Li et al., 2024), together with RALoc, LISA, LEADER, and

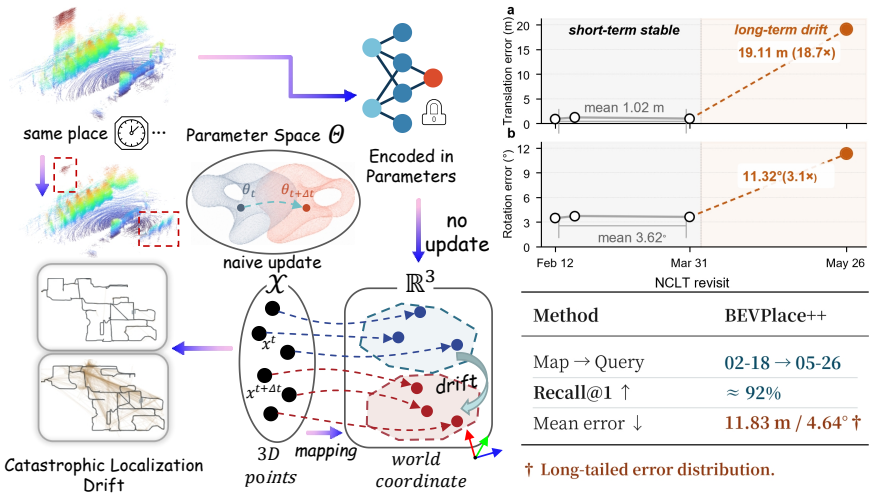


Figure 2: **Two failures under scene evolution.** No update makes the SCR field stale and causes catastrophic localization drift; naive update can overwrite historical mappings and cause forgetting. Retrieval remains globally robust but exhibits long-tailed metric error, motivating TCPP.

TACO (Yang et al., 2025; 2024; Wu et al., 2026; Xing et al., 2026), learn metric poses from scene-dependent representations but generally assume a static scene. Retrieval and map-matching methods, including BEVPlace++, EgoNN, and C-LaV (Luo et al., 2025; Komorowski et al., 2021; Cao et al., 2026), tolerate revisits but retain explicit maps or features. DIAL maintains an implicit metric field and invokes retrieval as global evidence for pseudo-label construction.

Visual localization and scene coordinate regression. Scene coordinate regression in visual localization offers a useful reference for implicit metric localization. ACE (Brachmann et al., 2023) trains a scene coordinate regressor in minutes, while GLACE and R-SCoRe extend compact regression to large-scale scenes (Wang et al., 2024; Jiang et al., 2025). ACE-G (Bruns et al., 2025) and Reloc3r (Dong et al., 2025) improve cross-view or cross-scene generalization, and FastForward (Barroso-Laguna et al., 2026) couples retrieval with feed-forward coordinate prediction. These methods address fast fitting or generalization, typically from a fixed posed collection. DIAL instead maintains the inherited metric field as its environment evolves. Unlike fitting a new regressor, continual maintenance must preserve established coordinate associations while restricting adaptation to observations whose geometry is no longer represented.

Continual visual localization. Image-based continual localization (Wang et al., 2021b) adds scenes with coverage-based sampling. ConDo (Li et al., 2025b) combines teacher-generated pseudo labels with uniform sampling of historical and newly collected data, but does not diagnose new observations from prediction discrepancies. In contrast, DIAL uses the historical localizer to identify changed or unseen regions and prioritizes these observations, reducing the amount of training required. It also examines continual adaptation without replayable historical data and the role of model plasticity across long-term scene evolution and segmented route expansion. This diagnostic role turns historical predictions into evidence for both supervision routing and the choice of model plasticity, beyond their conventional use as replacement labels.

Continual learning beyond localization. General continual learning mitigates forgetting through distillation, rehearsal, or parameter isolation (Li & Hoiem, 2018; Buzzega et al., 2020; Wang et al., 2022b; Smith et al., 2023; McDonnell et al., 2023; He et al., 2025). Source-free adaptation instead updates a deployed model from unlabeled streams, progressing from entropy minimization to safeguards against temporal correlation, dynamic shifts, and error accumulation (Wang et al., 2021a; 2022a; Niu et al., 2022; Gong et al., 2022; Niu et al., 2023; Yuan et al., 2023; Liu et al., 2024). These methods are largely organized around discrete prediction. Implicit localization instead requires retaining continuous coordinate associations whose errors affect downstream metric pose recovery. DIAL maintains this scene-coordinate readout, freezing it in HDER and updating it only when historical replay makes that safe.

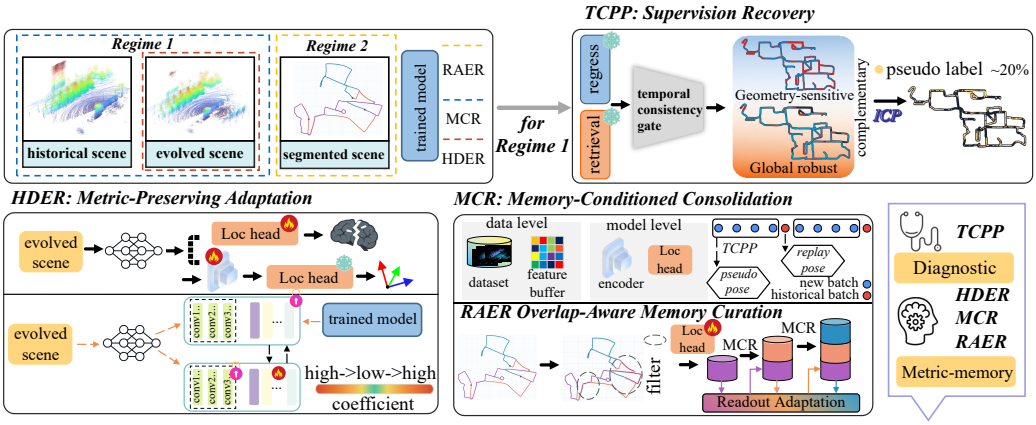


Figure 3: **DIAL organizes continual implicit localization around two coupled decisions: where supervision is trustworthy and where model plasticity is safe.** TCPP diagnoses changed or unseen observations from complementary retrieval and SCR evidence over time and recovers pseudo poses with global anchors and local ICP propagation. HDER, MCR, and RAER then preserve metric memory under different historical-support conditions, treating the historical localizer jointly as a diagnostic model and a stability anchor.

3 METHOD

3.1 PROBLEM FORMULATION AND OVERVIEW

DIAL defines a policy for supervision recovery and plasticity allocation, selecting an update route from the available historical support (Figure 3). TCPP cross-validates global retrieval and local SCR evidence over time, with ICP supplying pseudo poses for changed or unseen frames. Without replayable history, DIAL pairs TCPP with HDER and freezes the readout. When replay and existing coordinate support are available, TCPP operates with MCR. Overlapping segmented expansion uses MCR and RAER with an adaptive readout.

At stage t, let $f_{\theta^{t-1}} = h_{\psi^{t-1}} \circ g_{\phi^{t-1}}$ denote the historical localizer, with parameters $\theta^{t-1} = (\phi^{t-1}, \psi^{t-1})$. The encoder g_ϕ maps a scan to features, and the readout h_ψ predicts scene coordinates from which a metric pose is recovered. DIAL updates the model on $\mathcal{D}_t = \{(\mathbf{P}_i, \mathbf{T}_i)\}_{i=1}^{N_t}$, where N_t is the number of scans, $\mathbf{P}_i$ is scan i, and $\mathbf{T}_i \in SE(3)$ is its ground-truth or pseudo 6-DoF pose. The objective is to adapt to $\mathcal{D}_t$ while retaining accuracy on all earlier-stage data $\mathcal{D}_{<t}$; available historical samples determine the maintenance route.

3.2 SUPERVISION RECOVERY AND MAINTENANCE ROUTES

Temporal Consistency-guided Pseudo-label Pipeline (TCPP). TCPP uses predictions from the historical localizer to identify observations whose scene representation has become stale. Evolved geometry disrupts the temporal smoothness of SCR predictions, allowing a trajectory residual to determine where supervision is needed. The residual measures local trajectory inconsistency rather than absolute pose error, so supervision can be routed without ground-truth poses. For an ordered revisit $\mathcal{X}_t = \{\mathbf{P}_i\}_{i=1}^{N}$, let $\mathbf{p}_i^m \in \mathbb{R}^{d_m}$ denote the translation supplied by source m. For temporal gating, SCR uses its planar XY translation ($m = \mathrm{S}$, $d_m = 2$), whereas retrieval uses the full 3D translation ($m = \mathrm{R}$, $d_m = 3$). For each interior frame, TCPP computes

$$e_i^m = \|\mathbf{p}_i^m - \mathbf{p}_{i-1}^m\|_2 + \|\mathbf{p}_{i+1}^m - \mathbf{p}_i^m\|_2 - \|\mathbf{p}_{i+1}^m - \mathbf{p}_{i-1}^m\|_2. \tag{1}$$

Here e_i^m is the triangle-inequality residual and $\|\cdot\|_2$ is Euclidean distance. Source m accepts a frame when $q_i^m = \mathbb{I}[e_i^m \le \tau^m]$, with $\tau^m = \mathrm{median}(\mathbf{e}^m) + \kappa_m \,\mathrm{MAD}(\mathbf{e}^m)$ and $\mathbf{e}^m = \{e_i^m\}_{i=2}^{N-1}$. The indicator $\mathbb{I}[\cdot]$ converts this robust test into a binary decision, κ_m controls its tolerance, and $\mathrm{MAD}(\mathbf{e}^m) = \mathrm{median}_i\,|e_i^m - \mathrm{median}(\mathbf{e}^m)|$. Persistent SCR failures are consolidated by a five-frame drift test before entering the update set $\mathcal{U}_t$; Appendix D.2 provides the detailed temporal-drift criterion.

Pseudo-pose assignment exploits the complementary reliability of global retrieval and local regression. For each $i \in \mathcal{U}_t$, the routing rule is

$$\widetilde{\mathbf{T}}_i = \begin{cases} \mathbf{T}_i^{\mathrm{R}}, & q_i^{\mathrm{R}} = 1, \\ \mathcal{G}_{\mathrm{ICP}}(i \mid a, \mathbf{T}_a), & q_i^{\mathrm{R}} = 0, \end{cases} \tag{2}$$

where a is a trusted interval-boundary anchor with global pose $\mathbf{T}_a$, and $\mathcal{G}_{\mathrm{ICP}}$ composes adjacent-scan relative transforms from that anchor to frame i. This produces $\widetilde{\mathcal{D}}_t = \{(\mathbf{P}_i, \widetilde{\mathbf{T}}_i) \mid i \in \mathcal{U}_t\}$, concentrating supervision on changed or unseen regions. Appendix D gives the propagation rule, decision states, and gating details; Appendix A visualizes the recovered pseudo trajectories and diagnosed update regions.

Head-Decoupled Encoder Regularization (HDER). Without replay, the parameter update itself must retain old localization knowledge. The pose head is a continuous coordinate readout; overwriting it can blur previously precise metric associations. HDER keeps $h_{\psi^{t-1}}$ fixed in the forward path. It also freezes shallow convolutions that capture reusable LiDAR structure and confines adaptation to higher encoder blocks $\mathcal{A}$. With historical encoder parameters ϕ^{t-1} and voxel targets $\mathbf{c}_v$ from ground truth or TCPP, HDER optimizes

$$\mathcal{L}_{\mathrm{HDER}} = \frac{1}{|\mathcal{V}_t|} \sum_{v \in \mathcal{V}_t} \|h_{\psi^{t-1}}(g_\phi(\mathbf{P})_v) - \mathbf{c}_v\|_1 + \lambda(u) \sum_{\ell \in \mathcal{A}} \omega_\ell \|\phi_\ell - \phi_\ell^{t-1}\|_2^2, \qquad \psi = \psi^{t-1}. \tag{3}$$

Here $\mathcal{V}_t$ is the set of supervised voxels in current scan $\mathbf{P}$, $|\mathcal{V}_t|$ its cardinality, and v the voxel index. The first term is the mean ℓ_1 scene-coordinate loss to target $\mathbf{c}_v$; the second is a squared ℓ_2 drift penalty over adaptable layers $\ell \in \mathcal{A}$. Parameters ϕ_ℓ and ϕ_ℓ^{t-1} are the current and historical weights of layer ℓ, ω_ℓ is its anchoring weight, and $\lambda(u)$ controls regularization at normalized progress $u \in [0, 1]$. The constraint $\psi = \psi^{t-1}$ freezes the historical readout. Under the cosine sandwich schedule in Equation 6, strong initial anchoring controls early pseudo-label gradients, the relaxed middle phase admits new geometry, and the final rise consolidates the adapted encoder. Appendix B details the schedule phases (Table A1) and the resulting gradient path.

Memory-Conditioned Rehearsal (MCR). When historical samples are available, MCR periodically restores old-scene constraints. Dataset replay retains raw geometric supervision; bounded raw or feature buffers trade fidelity for speed. Replay may enter at the encoder or readout: its entry site determines the historical gradient path, while readout plasticity follows the maintenance regime. Let $\mathcal{M}_s = \{m_i \in \mathcal{D}_{<t} : i \equiv 0 \pmod{s}\}$ contain memory items sampled every s historical frames, and let $\mathcal{B}_t$ and $\mathcal{B}_{\mathcal{M}}$ denote current and replay batches. At zero-indexed optimization step k, MCR minimizes

$$\mathcal{L}^{(k)} = \mathcal{L}_{\mathrm{loc}}(\mathcal{B}_t) + \beta \rho_k \mathcal{L}_{\mathrm{loc}}(\mathcal{B}_{\mathcal{M}}) + \lambda(u) \mathcal{L}_{\mathrm{reg}}(\phi, \phi^{t-1}), \qquad \rho_k = \begin{cases} 1, & k \in \{0, K, 2K, \ldots\}, \\ 0, & \text{otherwise.} \end{cases} \tag{4}$$

Here $\mathcal{L}^{(k)}$ is the total step-k objective, $\mathcal{L}_{\mathrm{loc}}$ the localization loss, $\mathcal{L}_{\mathrm{reg}}(\phi, \phi^{t-1})$ the historical encoder prior, and β the replay weight. The gate ρ_k inserts one historical batch every K current updates; $\lambda(u)$ weights encoder regularization. We use $s = 4$, with $K = 4$ on NCLT and $K = 2$ on QEOxford. Setting $\lambda = 0$ yields pure MCR, while $\beta = 0$ yields memory-free HDER. DIAL-M denotes DIAL with MCR. DIAL-$\mathrm{M}_{\mathrm{enc}}$ uses encoder-side dataset replay, whereas DIAL-$\mathrm{M}_{\mathrm{head}}$ uses buffer-based head-side replay. Appendix F separates memory realization, gradient entry, and cadence.

Redundancy-Aware Exemplar Replay (RAER). Segmented route learning creates spatial overlap between temporally distinct stages. Replaying old scans already covered by the current stage spends memory without extending coordinate support. RAER retains historical samples according to their geometric complementarity with the current stage. For old exemplar m_i at XY position $\mathbf{z}_i$ and current-stage positions $\mathcal{Z}_t$,

$$d_i = \min_{\mathbf{z} \in \mathcal{Z}_t} \|\mathbf{z}_i - \mathbf{z}\|_2, \qquad \mathcal{M}_{t-1}^{\mathrm{RAER}} = \{m_i \in \mathcal{M}_{t-1} : d_i \geq r\}. \tag{5}$$

Here d_i is the nearest XY distance from old exemplar m_i to the current-stage position set $\mathcal{Z}_t$, $\mathcal{M}_{t-1}$ is the incoming historical memory, and r is the exclusion radius. The output $\mathcal{M}_{t-1}^{\mathrm{RAER}}$ retains exactly those exemplars with $d_i \geq r$; closer frames are removed as their locations are already covered by the current stage. Substantial spatial distribution shifts across segments call for readout adaptation. RAER uses MCR with a trainable readout to extend coordinate support while replaying complementary geometry; Appendix E specifies the selection order and radius behavior.

Table 1: **Maintenance frontier on NCLT.** Aggregate localization error versus continual-optimization time, historical replay, and pose-head plasticity.

Strategy	Optim. time	Historical replay	Pose head	Aggregate TE/RE ↓
No update (PELoc)	0	None	–	4.75/5.39
Naive adaptation	1.53h	None	Trainable	5.96/9.03
Full rebuild (upper bound)	6.03h	Full GT	Trainable	1.50/4.07
DIAL-H	∼19 min	None	Frozen	2.59/6.03
DIAL-M$_{\text{head}}$	∼12 min	Buffer	Trainable	1.90/5.06
DIAL-M$_{\text{enc}}$	∼25 min	Dataset	Frozen	**1.73/4.92**

4 EXPERIMENTS

4.1 EXPERIMENTAL SETUP

Datasets and protocols. We evaluate DIAL on NCLT (Carlevaris-Bianco et al., 2016) and segmented QEOxford (Barnes et al., 2020). NCLT provides long-term revisits; QEOxford partitions a route into overlapping stages. We apply DIAL to PELoc and SGLoc-v2 (Chen et al., 2025; Li et al., 2023), trained on one and three historical traversals, respectively. For primary NCLT maintenance, TCPP constructs pseudo supervision from `2012-05-26`, which contains evolved and previously unseen regions. We test changed/unseen localization on `2012-05-11` and retention on `02-12`/`02-19`/`03-31`; aggregate error averages these four routes. Figure 4 extends the protocol across roughly one year. Segmented QEOxford reports mean TE/RE on four hold-out sequences. Exact initialization and evaluation sequence IDs are listed in Appendix G.

Baselines. Maintenance baselines cover PELoc, LightLoc, DiffLoc, STCLoc, SGLoc-v2, LiSA, and HypLiLoc (Chen et al., 2025; Li et al., 2025a; 2024; Yu et al., 2022; Li et al., 2023; Yang et al., 2024; Wang et al., 2023), together with no update, naive adaptation, full retraining, and DIAL. Continual-learning controls include CoTTA (Wang et al., 2022a), SDM (Bricken et al., 2023), Fisher-EWC (Niu et al., 2022), ConDo (Li et al., 2025b), and CL-HSCNet (Wang et al., 2021b).

Metrics. We report mean translation error in meters and rotation error in degrees, separating performance on old and newly introduced scenes. Maintenance cost is measured by adaptation time, pose-supervision demand, and retained historical data or memory. For segmented learning, we report final average performance after the three-stage update sequence. Appendix G records the matched protocol controls, Appendix K defines the coordinate and error conventions, and Table A3 reports HDER/RAER sensitivity.

4.2 MAINTENANCE FRONTIER

Table 1 summarizes the deployment trade-off on PELoc. No update leaves the metric field stale, naive adaptation damages historical mappings, and full rebuild requires complete historical supervision. With TCPP pseudo labels (∼20% of frames), DIAL reaches low aggregate error with 12–25 minutes of continual optimization using no historical replay, a buffer, or dataset replay. Figure 1 extends this comparison to a second implicit localizer; Appendix H separates maintenance computation from the unchanged deployment graph.

4.3 LOCALIZATION MAINTENANCE UNDER SCENE EVOLUTION

Table 2 evaluates long-term scene evolution under static deployment, naive adaptation, full retraining, and the DIAL maintenance routes.

Without adaptation, PELoc yields 4.75/5.39 m/°. Naive adaptation takes 1.53 hours and increases the error to 5.96/9.03 m/°; full rebuild reaches 1.50/4.07 m/° after 6.03 hours with complete historical ground truth. Using TCPP pseudo labels for approximately 20% of the revisit frames, DIAL-H uses no historical replay data, DIAL-M$_{\text{head}}$ completes optimization in 12 minutes, and DIAL-M$_{\text{enc}}$ reaches **1.73/4.92** m/° in 25 minutes. Here "None" denotes no raw historical scans or replay buffer during localizer optimization; TCPP may query a pre-built retrieval index for pseudo-pose recovery. On SGLoc-v2, DIAL reduces aggregate error to 1.31/3.61 m/°; Appendix I specifies the shared interface and backbone-specific choices.

Table 2: Localization maintenance on NCLT. We report mean TE/RE (m/°), historical replay, supervision, and continual-optimization time. Naive adaptation uses full-revisit pseudo poses; DIAL uses the TCPP-selected subset (∼20% of frames). The first three DIAL rows use PELoc; the last transfers DIAL-M_{enc} to SGLoc-v2. In each error column, **bold**/underlined values are best/second best.

Maintenance strategy	DIAL variant	Base localizer	Pose head	Maintenance cost: Historical replay	Maintenance cost: Pose supervision	Maintenance cost: Optim. time	Localization error ↓ (m/°): Historical	Localization error ↓ (m/°): Evolved / unseen	Localization error ↓ (m/°): Aggregate
No Update	–	PELoc	–	None	None	0	1.02/3.61	15.92/10.74	4.75/5.39
	–	LightLoc	–	None	None	0	1.86/4.54	24.34/11.77	7.48/6.35
	–	SGLoc-v2	–	None	None	0	**0.69/2.65**	11.45/7.81	3.38/3.94
Naive adaptation	–	PELoc	Trainable	None	Full pseudo revisit	1.53h	4.98/9.38	8.91/7.97	5.96/9.03
Full rebuild	–	PELoc	Trainable	Full	Full GT	6.03h	–	–	∼1.50/4.07
	–	LightLoc	Trainable	Full	Full GT	1h	–	–	∼1.46/2.80
	–	LiSA	Trainable	Full	Full GT	44h	–	–	∼1.47/**2.30**
	–	HypLiLoc	Trainable	Full	Full GT	12h	–	–	∼1.80/3.12
Continual (Ours)	**DIAL-H**	PELoc	Frozen	None	TCPP	**∼19 min**	2.20/5.62	3.77/7.24	2.59/6.03
	DIAL-M_{head}	PELoc	Trainable	Buffer	TCPP	**∼12 min**	1.46/4.60	3.24/6.45	1.90/5.06
	DIAL-M_{enc}	PELoc	Frozen	Dataset	TCPP	**∼25 min**	1.32/4.53	2.91/6.07	1.73/4.92
	DIAL-M_{enc}	SGLoc-v2	Frozen	Dataset	TCPP	**∼27 min**	0.96/3.32	**2.36/4.50**	**1.31**/3.61

NCLT revisit	No update TE/RE	DIAL-M_{enc} TE/RE
2012-02-12	**0.89/3.49**	1.52/5.33
2012-02-19	**1.22/3.73**	1.88/5.64
2012-03-31	**0.96/3.63**	1.54/5.20
2012-05-11	15.92/10.74	**3.15/7.45**
2012-06-15	41.09/20.30	**2.70/6.05**
2012-08-20	26.05/15.70	**3.30/6.94**
2012-09-28	26.30/16.53	**2.56/6.61**
2012-11-17	14.18/11.15	**4.24/8.64**
2012-12-01	16.19/12.84	**3.83/8.91**
2013-02-23	19.86/12.42	**4.28/9.08**

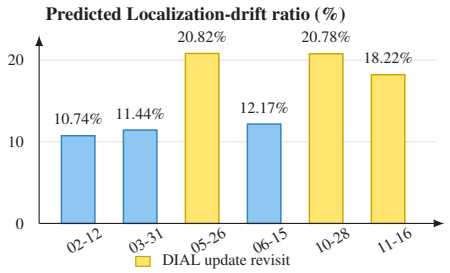


Figure 4: **Long-term NCLT experiment.** Left: sequence-wise localization error of a no-update PELoc baseline and DIAL-M_{enc} after three continual updates. Right: revisits above the 15% predicted localization-drift threshold trigger maintenance. TE/RE denotes mean translation error (m) / rotation error (°).

Figure 4 reports the long-term NCLT experiment. Each triggered update uses DIAL-M_{enc}, with encoder-side dataset replay and a frozen readout. The protocol activates maintenance when the predicted localization-drift ratio exceeds 15%, selecting `2012-05-26`, `2012-10-28`, and `2012-11-16`. DIAL bounds mean translation error within 1.52–4.28 m as the no-update baseline collapses.

4.4 Comparison with Continual-Learning Methods

Table 3 compares continual-learning mechanisms on the same PELoc localizer and NCLT revisit. CoTTA, SDM, and Fisher-EWC update on the full revisit with their respective objectives; DIAL uses the TCPP-selected subset. ConDo and DIAL share the pseudo-label pool and reported historical-memory budget but use uniform sampling and TCPP routing, respectively.

In our CoTTA adaptation (Wang et al., 2022a), the metric readout is updated, yet evolved/unseen error remains 14.05/10.45 m/°. SDM incurs large errors on both historical and evolved routes: its sparse, discrete memory is poorly matched to metric localization, which depends on precise coordinate associations. Fisher-EWC retains historical accuracy at 1.24/4.01 m/° but reaches 20.11/15.64 m/° on the evolved route, indicating limited adaptation. DIAL-H freezes the readout and reaches 3.77/7.24 m/° without replay. From the same pseudo-label pool, ConDo applies uniform sampling (Li et al., 2025b), whereas TCPP directs DIAL-M_{enc} supervision toward changed or unseen frames, reducing evolved/unseen error from 4.87/6.64 to 2.91/6.07 m/°.

Table 3: Continual localization on NCLT with the same PELoc localizer and revisit. CoTTA, SDM, and Fisher-EWC use full-revisit updates. ConDo and DIAL share the pseudo-label pool and matched replay budget, using uniform sampling and the TCPP-selected subset, respectively. Entries report mean translation/rotation error (m/°).

		Retention	Adaptation	Overall
CL regime	**Method**	**Historical**	**Evolved / unseen**	**Aggregate**
Reference	Full rebuild (upper bound)	–	–	1.50/4.07
Regularization (no replay)	CoTTA	1.41/**3.94**	14.05/10.45	4.57/5.57
	SDM	30.73/22.10	39.24/36.83	32.85/25.78
	Fisher-EWC	**1.24**/4.01	20.11/15.64	5.95/6.91
	DIAL-H	2.20/5.62	**3.77/7.24**	**2.59**/6.03
Rehearsal (matched budget)	ConDo	**1.19/4.18**	4.87/6.64	2.11/**4.79**
	DIAL-M$_{\text{enc}}$	1.32/4.53	**2.91/6.07**	**1.73**/4.92

Table 4: **HDER stability ablation on NCLT.** HDER variants use the same TCPP supervision and matched update steps without replay; full-revisit naive adaptation is included as a reference.

	Stability–plasticity design				Localization error ↓ (m/°)		
Variant	**Pose head**	**Reg. site**	**Conv1–3**	**Prior**	**Historical**	**Evolved / unseen**	**Aggregate**
Naive adaptation	Trainable	–	Trainable	–	4.98/9.38	8.91/7.97	5.96/9.03
Head-side regularization	Trainable	Head	Trainable	Fixed	6.53/9.80	9.02/11.82	7.15/10.31
HDER w/o layer selection	Frozen	Encoder	Trainable	Cosine	2.59/6.49	**3.55**/7.57	2.83/6.76
HDER w/o cosine schedule	Frozen	Encoder	Frozen	Fixed	2.27/5.96	3.79/**7.23**	2.65/6.27
HDER	Frozen	Encoder	Frozen	Cosine	**2.20/5.62**	3.77/7.24	**2.59/6.03**

4.5 Ablation Studies

HDER stability. Table 4 examines where to impose stability without replayable historical scans. Among the TCPP-matched variants, head-side regularization is insufficient, whereas HDER reaches 2.59/6.03 m/° by freezing the readout and regularizing selected encoder blocks. Freezing Conv1–3 retains shallow features while higher layers adapt, and the cosine schedule improves on a fixed prior. Full-revisit naive adaptation is an external reference and remains destructive despite using more new-scene supervision.

Segmented continual localization. Table 5 tests readout plasticity and replay selection on segmented QEOxford. With all 18.8k historical frames, a trainable readout reaches 1.08 m / 1.67°; freezing it degrades to 8.81 m / 17.8°, confirming that route expansion requires readout plasticity. RAER removes 23% of replay frames (18.8k to 14.4k) while retaining 1.10 m / 1.64°. At 8.0k frames, it improves over coverage sampling from 1.83/2.31 to 1.67/2.15 m/°.

Table 5: Three-stage continual localization on QEOxford. The 8.0k coverage policy and RAER share the implicit-localization backbone, training schedule, replay count, and update steps. For final TE/RE (m/°), **bold**/underlined values are best/second best.

		Maintenance cost		Pose	Localization
Method	**Exemplar policy**	**Replay frames**	**Optim. time**	**head**	**Final TE/RE (m/°) ↓**
Joint training (upper bound)	All data	Full	∼6.03 h	Trainable	**1.06/1.63**
CL-HSCNet-style coverage	Coverage sampling	8.0k	∼4.90 h	Trainable	1.83/2.31
DIAL-M	Random	18.8k	∼5.55 h	Frozen	8.81/17.8
DIAL-M	Random	18.8k	∼5.55 h	Trainable	1.08/1.67
DIAL-M + RAER	Redundancy-aware	8.0k	∼4.90 h	Trainable	1.67/2.15
DIAL-M + RAER	Redundancy-aware	14.4k	∼5.20 h	Trainable	1.10/1.64

TCPP and MCR. Table 6 isolates pseudo-pose construction and routing. Single-source labels either inherit stale SCR bias or lack temporal cross-validation. At the same 20.8% budget, out-of-pool sampling raises evolved-route error to 9.25/7.66 m/°, whereas TCPP reaches 2.91/6.07 m/°. Supplementary Figure A2 analyzes full-trajectory pseudo-labels; Appendix D details interval recovery and gating.

Table 6: TCPP ablation on NCLT with 5,504 pseudo labels (20.8%). TCPP outperforms single-source and out-of-pool controls.

	Pseudo-label evidence				Budget	Localization error ↓ (m/°)		
Variant	**SCR**	**Retrieval**	**Temporal**	**ICP**	**Frames**	**Historical**	**Evolved / unseen**	**Aggregate**
SCR only	▬				20.8%	2.14/5.09	15.84/10.14	5.56/6.35
Retrieval + ICP		▬		▬	20.8%	1.42/4.72	5.14/6.75	2.35/5.23
Out-of-pool random	▬	▬	▬	▬	20.8%	1.45/**4.34**	9.25/7.66	3.40/5.17
TCPP	▬	▬	▬	▬	20.8%	**1.32**/4.53	**2.91/6.07**	**1.73/4.92**

Table 7: MCR memory ablation on NCLT. Encoder replay favors accuracy, whereas head replay reduces optimization time.

	Memory realization			Localization error ↓ (m/°)			
Variant	**Memory form**	**Replay site**	**Access**	**Historical**	**Evolved / unseen**	**Aggregate**	**Time**
No replay	None	–	None	4.98/9.38	8.91/7.97	5.96/9.03	1.53h
Dataset replay	Raw scans	Encoder	Dataset	**1.32/4.53**	2.91/**6.07**	**1.73/4.92**	∼25 min
Raw buffer replay	Raw scans	Encoder	Buffer	1.43/4.66	**2.87**/6.28	1.79/5.07	∼26 min
Raw buffer replay	Raw scans	Head	Buffer	1.46/4.60	3.24/6.45	1.90/5.06	∼**12 min**
Feature buffer replay	Features	Head	Buffer	1.50/4.61	3.26/6.46	1.94/5.08	∼13 min
MCR (w/o HDER)	Raw scans	Encoder	Dataset	1.32/4.53	**2.91/6.07**	**1.73**/4.92	∼25 min
MCR + HDER	Raw scans	Encoder	Dataset	**1.29/4.39**	3.26/6.17	1.78/**4.84**	∼25 min

Table 7 compares MCR realizations. Encoder-side dataset and buffer replay reach 1.73/4.92 and 1.79/5.07 m/° aggregate error; head-side replay reduces optimization to 12–13 minutes at a modest accuracy cost. Adding HDER improves historical error from 1.32/4.53 to 1.29/4.39 but raises evolved-route error from 2.91/6.07 to 3.26/6.17, favoring retention over new-scene plasticity. Table A10 reports replay cadence.

Discussion on streaming updates. DIAL updates after processing a revisit. TCPP needs neighboring predictions and a drift window, while an unseen interval may require a later trusted anchor for ICP. Scan-wise updates would precede this evidence and alter the historical model during diagnosis. Revisit-level processing preserves the diagnostic model and temporal context; the buffer-based route optimizes selected frames in about 12 minutes. Appendix J discusses streaming constraints.

Interpreting the stability asymmetry. The frozen- and trainable-head results reflect distinct regimes. Changed NCLT revisits retain the coordinate domain but lack old scans; pseudo-label updates can displace the calibrated readout, so HDER fixes it and adapts higher encoder blocks. Segmented QEOxford expands route coverage under replay, allowing a trainable head to extend the coordinate mapping. Tables 4 and 5 connect plasticity to historical support and coordinate-space expansion.

5 Conclusion

DIAL coordinates supervision recovery and model plasticity to maintain implicit LiDAR localizers across changed revisits and segmented route expansion. The historical model supplies diagnostic evidence and a coordinate readout worth preserving. TCPP routes changed or unseen scans; HDER retains the readout without historical data, while MCR and RAER provide replay across memory and overlap conditions. These mechanisms sustain year-long NCLT localization with minute-scale updates and approach joint-training accuracy on segmented QEOxford, showing how stored geometry can guide continual localization.

Limitation and outlook. DIAL relies on revisit observations and reliable retrieval or local anchors, both of which may fail after severe topology changes. Future work should recognize unsupported regions and extend temporal routing to causal windows for memory-efficient online maintenance.

More details and experiments in the appendix. Appendices A–C cover long-term maintenance, HDER stability, and historical-support regimes; D–F detail TCPP, RAER, and MCR; G–K document protocols, computation, backbone transfer, streaming constraints, and coordinate conventions.

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

Dan Barnes, Matthew Gadd, Paul Murcutt, Paul Newman, and Ingmar Posner. The oxford radar robotcar dataset: A radar extension to the oxford robotcar dataset. In *2020 IEEE international conference on robotics and automation (ICRA)*, pp. 6433–6438. IEEE, 2020.

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.

Eric Brachmann, Alexander Krull, Frank Michel, Stefan Gumhold, Jamie Shotton, and Carsten Rother. Learning 6d object pose estimation using 3d object coordinates. In *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part II 13*, pp. 536–551. Springer, 2014.

Eric Brachmann, Frank Michel, Alexander Krull, Michael Ying Yang, Stefan Gumhold, et al. Uncertainty-driven 6d pose estimation of objects and scenes from a single rgb image. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3364–3372, 2016.

Eric Brachmann, Tommaso Cavallari, and Victor Adrian Prisacariu. Accelerated coordinate encoding: Learning to relocalize in minutes using rgb and poses. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5044–5053, 2023.

Trenton Bricken, Xander Davies, Deepak Singh, Dmitry Krotov, and Gabriel Kreiman. Sparse distributed memory is a continual learner. *arXiv preprint arXiv:2303.11934*, 2023.

Leonard Bruns, Axel Barroso-Laguna, Tommaso Cavallari, Aron Monszpart, Sowmya Munukutla, Victor Adrian Prisacariu, and Eric Brachmann. Ace-g: Improving generalization of scene coordinate regression through query pre-training. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 26751–26761, 2025.

Pietro Buzzega, Matteo Boschini, Angelo Porrello, Davide Abati, and Simone Calderara. Dark experience for general continual learning: A strong, simple baseline. In *Advances in Neural Information Processing Systems*, volume 33, pp. 15920–15930, 2020.

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.

Nicholas Carlevaris-Bianco, Arash K Ushani, and Ryan M Eustice. University of michigan north campus long-term vision and lidar dataset. *The International Journal of Robotics Research*, 35 (9):1023–1035, 2016.

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.

Siyan Dong, Shuzhe Wang, Shaohui Liu, Lulu Cai, Qingnan Fan, Juho Kannala, and Yanchao Yang. Reloc3r: Large-scale training of relative camera pose regression for generalizable, fast, and accurate visual localization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 16739–16752, 2025.

Taesik Gong, Jongheon Jeong, Taewon Kim, Yewon Kim, Jinwoo Shin, and Sung-Ju Lee. NOTE: Robust continual test-time adaptation against temporal correlation. In *Advances in Neural Information Processing Systems*, volume 35, pp. 27253–27266, 2022.

Jiangpeng He, Zhihao Duan, and Fengqing Zhu. Cl-lora: Continual low-rank adaptation for rehearsal-free class-incremental learning. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pp. 30534–30544, 2025.

Xudong Jiang, Fangjinhua Wang, Silvano Galliani, Christoph Vogel, and Marc Pollefeys. R-SCoRe: Revisiting scene coordinate regression for robust large-scale visual localization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 11536–11546, 2025.

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, 2021.

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, 2025a.

Zhizhong Li and Derek Hoiem. Learning without forgetting. *IEEE transactions on pattern analysis and machine intelligence*, 40(12):2935–2947, 2018. doi: 10.1109/TPAMI.2017.2773081.

Zijun Li, Zhipeng Cai, Bochun Yang, Xuelun Shen, Siqi Shen, Xiaoliang Fan, Michael Paulitsch, and Cheng Wang. Condo: Continual domain expansion for absolute pose regression. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 14628–14636, 2025b.

Jiaming Liu, Senqiao Yang, Peidong Jia, Renrui Zhang, Ming Lu, Yandong Guo, Wei Xue, and Shanghang Zhang. ViDA: Homeostatic visual domain adapter for continual test time adaptation. In *International Conference on Learning Representations*, 2024.

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

Mark D. McDonnell, Dong Gong, Amin Parvaneh, Ehsan Abbasnejad, and Anton van den Hengel. RanPAC: Random projections and pre-trained models for continual learning. In *Advances in Neural Information Processing Systems*, volume 36, pp. 12022–12053, 2023.

Shuaicheng Niu, Jiaxiang Wu, Yifan Zhang, Yaofo Chen, Shijian Zheng, Peilin Zhao, and Mingkui Tan. Efficient test-time model adaptation without forgetting. In *International conference on machine learning*, pp. 16888–16905. PMLR, 2022.

Shuaicheng Niu, Jiaxiang Wu, Yifan Zhang, Zhiquan Wen, Yaofo Chen, Peilin Zhao, and Mingkui Tan. Towards stable test-time adaptation in dynamic wild world. In *International Conference on Learning Representations*, 2023.

James Seale Smith, Leonid Karlinsky, Vyshnavi Gutta, Paola Cascante-Bonilla, Donghyun Kim, Assaf Arbelle, Rameswar Panda, Rogerio Feris, and Zsolt Kira. CODA-Prompt: Continual decomposed attention-based prompting for rehearsal-free continual learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 11909–11919, 2023.

Dequan Wang, Evan Shelhamer, Shaoteng Liu, Bruno Olshausen, and Trevor Darrell. Tent: Fully test-time adaptation by entropy minimization. In *International Conference on Learning Representations*, 2021a.

Fangjinhua Wang, Xudong Jiang, Silvano Galliani, Christoph Vogel, and Marc Pollefeys. GLACE: Global local accelerated coordinate encoding. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 21562–21571, 2024.

Qin Wang, Olga Fink, Luc Van Gool, and Dengxin Dai. Continual test-time domain adaptation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 7201–7211, 2022a.

Shuzhe Wang, Zakaria Laskar, Iaroslav Melekhov, Xiaotian Li, and Juho Kannala. Continual learning for image-based camera localization. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 3252–3262, 2021b.

Sijie Wang, Qiyu Kang, Rui She, Wei Wang, Kai Zhao, Yang Song, and Wee Peng Tay. Hypliloc: Towards effective lidar pose regression with hyperbolic fusion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5176–5185, 2023.

Wei Wang, Bing Wang, Peijun Zhao, Changhao Chen, Ronald Clark, Bo Yang, Andrew Markham, and Niki Trigoni. Pointloc: Deep pose regressor for lidar point cloud localization. *IEEE Sensors Journal*, 22(1):959–968, 2021c.

Zifeng Wang, Zizhao Zhang, Chen-Yu Lee, Han Zhang, Ruoxi Sun, Xiaoqi Ren, Guolong Su, Vincent Perot, Jennifer Dy, and Tomas Pfister. Learning to prompt for continual learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 139–149, 2022b.

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.

Wei Xu and Fu Zhang. Fast-lio: A fast, robust lidar-inertial odometry package by tightly-coupled iterated kalman filter. *IEEE Robotics and Automation Letters*, 6(2):3317–3324, 2021.

Bochun Yang, Zijun Li, Wen Li, Zhipeng Cai, Chenglu Wen, Yu Zang, Matthias Muller, and Cheng Wang. Lisa: Lidar localization with semantic awareness. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 15271–15280, 2024.

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.

Huan Yin, Xuecheng Xu, Sha Lu, Xieyuanli Chen, Rong Xiong, Shaojie Shen, Cyrill Stachniss, and Yue Wang. A survey on global lidar localization: Challenges, advances and open problems. *International Journal of Computer Vision*, 132(8):3139–3171, 2024.

Shangshu Yu, Cheng Wang, Yitai Lin, Chenglu Wen, Ming Cheng, and Guosheng Hu. Stcloc: Deep lidar localization with spatio-temporal constraints. *IEEE Transactions on Intelligent Transportation Systems*, 24(1):489–500, 2022.

Longhui Yuan, Binhui Xie, and Shuang Li. Robust test-time adaptation in dynamic scenarios. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 15922–15932, 2023.

Chunran Zheng, Wei Xu, Zuhao Zou, Tong Hua, Chongjian Yuan, Dongjiao He, Bingyang Zhou, Zheng Liu, Jiarong Lin, Fangcheng Zhu, et al. Fast-livo2: Fast, direct lidar-inertial-visual odometry. *IEEE Transactions on Robotics*, 2024.

APPENDIX CONTENTS

This appendix connects DIAL's maintenance decisions to trajectory-level evidence, optimization behavior, controlled protocols, and implementation details. Section titles and page numbers below link directly to the corresponding material.

A Long-Term NCLT Maintenance 14
A.1 From Recovered Supervision to Maintained Trajectories 15
B HDER and Frozen-Readout Stability 16
B.1 Regularization schedule 16
B.2 Gradient path under missing history 16
B.3 Sensitivity of continual controls 17
C Historical Support and Plasticity Regimes 18
C.1 The historical model across stages 18
C.2 Transitions between maintenance routes 19
D Temporal Diagnosis and Pseudo-Label Recovery 19
D.1 Recovering an unreliable interval 19
D.2 Temporal residual and decision states 20
E RAER under Segmented Continual Localization 21
E.1 Spatial exclusion rule 21
E.2 Selection under a fixed budget 21
E.3 Interpreting the exclusion radius 22
F MCR Memory Realizations and Replay Cadence 22
F.1 Replay as a gradient constraint 22
F.2 Fixed-capacity replay cadence 23
F.3 Cadence and memory capacity 23
F.4 Gradient visibility at the replay site 24
G Experimental Protocols and Controls 25
G.1 Sequence conditions and shared controls 26
H Computational Structure of DIAL 26
H.1 Inference and checkpoint state 27
I Interface to Implicit Localizers 27
I.1 Transferred and backbone-specific components 27
J Discussion on Streaming Updates 28
J.1 Evidence boundaries and observable safeguards 28
K Coordinate and Transform Conventions 29

A LONG-TERM NCLT MAINTENANCE

Figure A1 compares representative NCLT revisits under the original PELoc checkpoint and the DIAL-M_{enc} checkpoint after the continual update on `2012-11-16`. The paired trajectories expose both retention on historical routes and recovery on evolved revisits, complementing the aggregate errors in Figure 4.

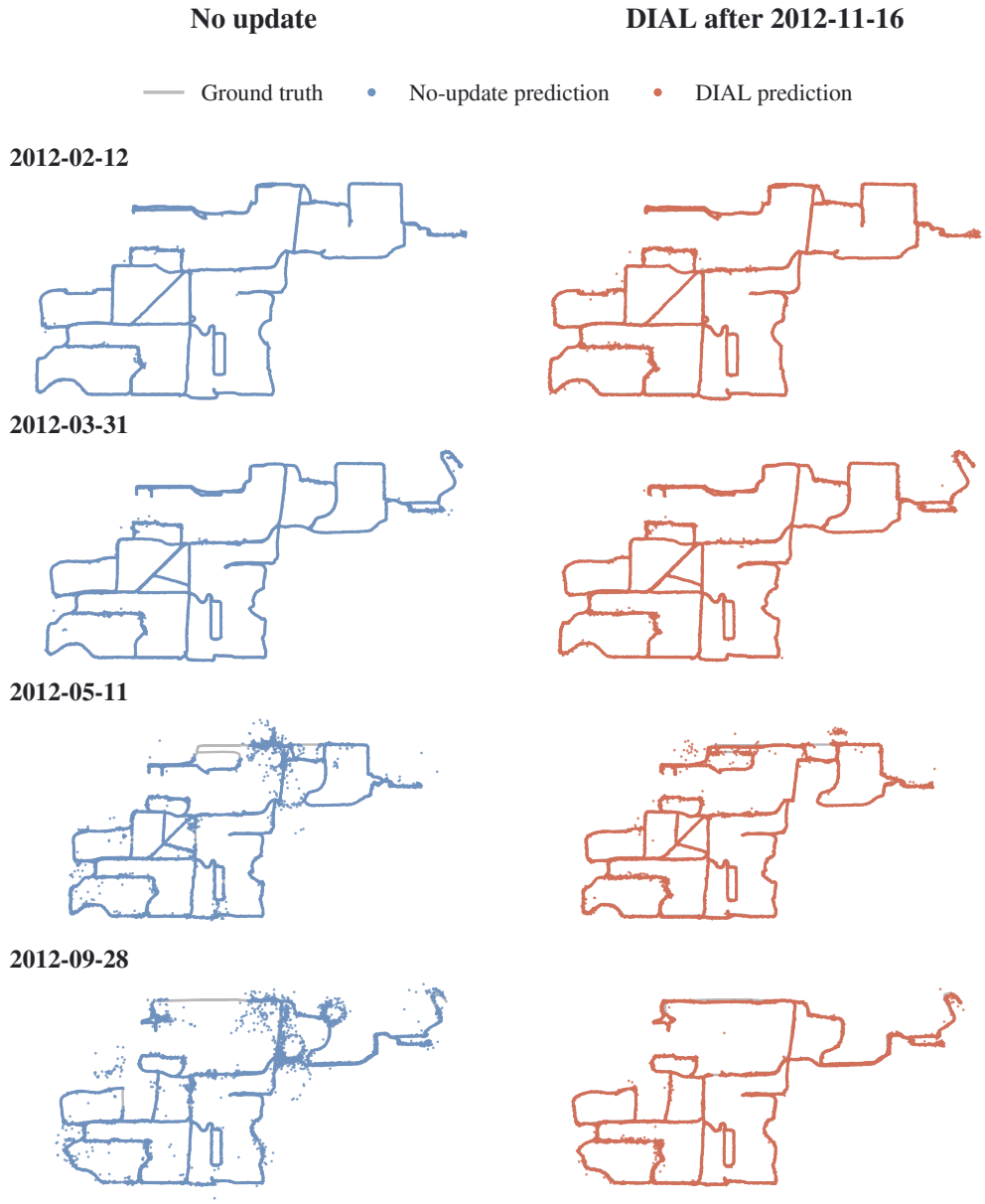


Figure A1: **Trajectory-level comparison before and after continual maintenance.** Each row evaluates the same NCLT revisit using the original PELoc checkpoint (left) and DIAL after the `2012-11-16` update (right). DIAL retains the historical route structure while correcting the severe drift on the evolved `2012-05-11` and `2012-09-28` revisits.

Pseudo-label quality and diagnostic use. Figure A2 compares the full TCPP pseudo trajectories with GT on two NCLT update revisits. Against GT used only for analysis, their median/P90 translation errors are 0.15/1.67 m on `2012-05-26` and 0.15/3.49 m on `2012-10-28`. The pseudo labels preserve the global route structure, with larger residuals concentrated around geometrically complex junctions and loops. They provide supervision when GNSS is unavailable and new ground-truth

poses are not collected during a revisit. When GT is available, the historical localizer still serves as a diagnostic model: temporal inconsistencies identify changed or previously unseen regions, focusing supervised continual learning on affected observations and reducing update cost.

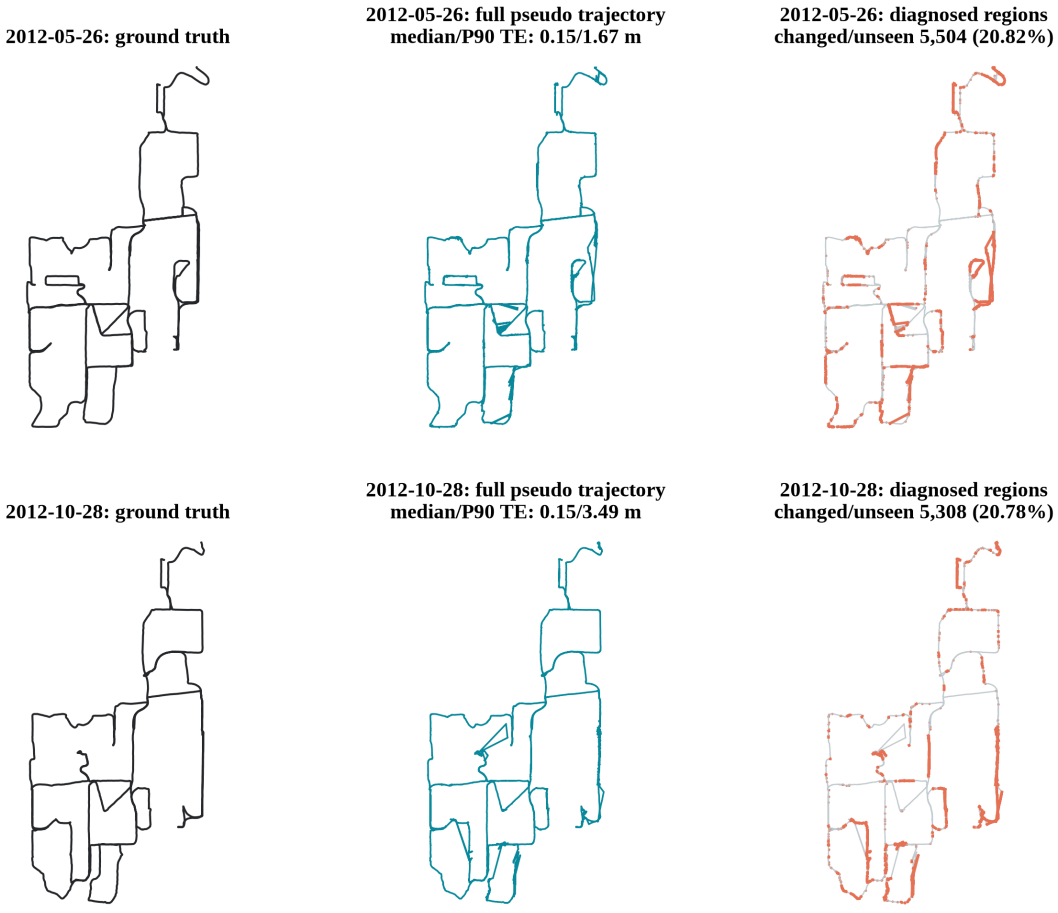


Figure A2: **Full pseudo-label quality and diagnosed regions.** Rows show the `2012-05-26` and `2012-10-28` update revisits; columns show GT, the full TCPP pseudo trajectory, and the regions diagnosed as changed or unseen. Gray traces the full pseudo trajectory, and coral highlights scans routed to the update set. Reported translation errors use GT only for analysis, not pseudo-label generation.

A.1 From recovered supervision to maintained trajectories

The two figures expose complementary aspects of maintenance. Figure A1 tests the final checkpoint on routes that differ in season and elapsed time, whereas Figure A2 examines the supervision used to construct intermediate checkpoints. Successful maintenance requires both: recovered supervision must reach the changed regions, and the resulting update must preserve previously valid historical geometry.

The diagnosed regions form contiguous route segments rather than isolated outliers. This structure follows from the temporal gate: a local SCR failure is accepted as evidence only when it persists within the drift window, after which retrieval or anchor-based ICP supplies a pose in the global frame. The resulting labels serve a narrower role than survey-grade ground truth: they restore coordinate supervision where the historical field has ceased to provide a stable readout. Junctions and loops remain the difficult cases, since similar geometry can support several locally plausible alignments.

B HDER AND FROZEN-READOUT STABILITY

B.1 REGULARIZATION SCHEDULE

With normalized training progress $u \in [0, 1]$, HDER uses

$$\lambda(u) = \begin{cases} \lambda_{\max}, & 0 \le u \le \frac{1}{3}, \\ \lambda_{\min} + \frac{\lambda_{\max} - \lambda_{\min}}{2} \left[1 + \cos\big(3\pi(u - \frac{1}{3})\big)\right], & \frac{1}{3} < u \le \frac{2}{3}, \\ \lambda_{\min} + \frac{\lambda_{\max} - \lambda_{\min}}{2} \left[1 - \cos\big(3\pi(u - \frac{2}{3})\big)\right], & \frac{2}{3} < u \le 1. \end{cases} \tag{6}$$

Here u is normalized training progress, $\lambda(u)$ is the active encoder-prior weight, and $\lambda_{\max}$ and $\lambda_{\min}$ are its upper and lower bounds. We use $\lambda_{\max} = 0.03$ and $\lambda_{\min} = 0.0125$ in the reported experiments. The three intervals allocate equal portions of training to initial anchoring, controlled plasticity, and final consolidation; π is the circular constant in the cosine interpolation.

The schedule is non-monotonic by design. A strong prior first prevents the pseudo-labeled gradient from immediately displacing the historical solution. The middle interval relaxes that constraint while the changed observations are fitted. The final interval strengthens the prior again, favoring adapted solutions that remain close to the inherited encoder. A fixed coefficient does not explicitly encode these distinct optimization phases.

Table A1: Interpretation of the HDER schedule. The intervals have equal duration; only the strength of the historical encoder prior changes.

Progress	**Prior**	**Optimization role**
$[0, 1/3]$	$\lambda_{\max}$	Stabilize the inherited representation as pseudo-supervised fitting begins.
$(1/3, 2/3]$	Cosine decrease	Admit encoder plasticity after the update gradient has become structured.
$(2/3, 1]$	Cosine increase	Consolidate the adapted encoder near the historical parameter basin.

B.2 GRADIENT PATH UNDER MISSING HISTORY

HDER preserves the readout as the metric interface learned by the historical localizer. Its encoder update is confined to higher layers and measured from the historical parameters, allowing changed appearance and geometry to alter the representation without directly rewriting the scene-coordinate map. The update changes the optimization path while supervision remains tied to routed current observations. Algorithm 1 gives the resulting procedure.

Input: historical parameters (ϕ^{t-1}, ψ^{t-1}), pseudo-labeled set $\widetilde{\mathcal{D}}_t$, bounds $(\lambda_{\min}, \lambda_{\max})$
Output: maintained parameters (ϕ^t, ψ^t)

1 Initialize $\phi \leftarrow \phi^{t-1}$ and $\psi \leftarrow \psi^{t-1}$.
2 Freeze the readout h_ψ and shallow encoder blocks Conv1–3.
3 For normalized progress u, evaluate $\lambda(u)$ using Equation 6.
4 Compute the scene-coordinate regression loss on $\widetilde{\mathcal{D}}_t$ and the parameter prior on the trainable encoder blocks.
5 Update only the unfrozen components of ϕ; keep $\psi = \psi^{t-1}$.
6 Return $\phi^t \leftarrow \phi$ and $\psi^t \leftarrow \psi^{t-1}$.

Algorithm 1: HDER update without replayable history

Table A2: Gradient access in HDER. The historical checkpoint supplies both the initialization and the encoder prior; no historical scan enters the update.

Component	**Gradient**	**Role during maintenance**
Pose readout h_ψ	Blocked	Retains the calibrated scene-coordinate mapping.
Encoder Conv1–3	Blocked	Preserves low-level geometric responses shared across revisits.
Higher encoder blocks	Enabled	Absorbs changed-revisit evidence under the historical parameter prior.
Historical scans	Absent	HDER does not assume replayable old observations.

B.3 Sensitivity of continual controls

Two controls govern historical preservation at different stages of DIAL. The HDER prior regulates parameter drift on NCLT, while the RAER radius regulates spatial redundancy under the three-stage QEOxford protocol (Table A3).

Table A3: Sensitivity to HDER and RAER controls. HDER results are reported on NCLT; RAER results use segmented QEOxford. TE/RE denotes mean translation error in meters and rotation error in degrees.

Module	**Control**	**Setting**	**Historical TE/RE↓**	**Evolved / new TE/RE↓**	**Aggregate / final TE/RE↓**
HDER	Encoder prior λ	0 (naive update)	4.98/9.38	8.91/7.97	5.96/9.03
		0.03	2.27/5.96	3.79/**7.23**	2.65/6.27
		Cosine schedule	**2.20/5.62**	**3.77**/7.24	**2.59/6.03**
		0.05	2.25/5.85	3.78/7.29	2.64/6.19
RAER	XY exclusion radius r	0 m	–	–	**1.09**/1.67
		1 m	–	–	1.10/**1.64**
		2 m	–	–	1.10/**1.64**
		3 m	–	–	1.13/1.73

Naive update yields 5.96/9.03 m/° aggregate error. Fixed nonzero priors recover most of this loss, and the cosine schedule further reduces aggregate error by varying the stability constraint during optimization. RAER changes little between $r = 1$ m and $r = 2$ m at the reported precision; $r = 3$ m begins to remove useful historical coverage.

Table A4: Historical support determines the locus of plasticity across DIAL's maintenance routes.

Setting	Historical scans	Coordinate support	Readout policy	Encoder policy	DIAL route and principal risk
Changed revisit	Unavailable	Existing domain; locally stale observations	Frozen	HDER-constrained	TCPP + HDER. Unconstrained readout updates may overwrite calibrated historical coordinate associations.
Changed revisit with replay	Available	Existing domain with explicit historical constraints	Replay-supported	MCR	TCPP + MCR. Replay cadence and memory realization determine the retention–cost trade-off.
Segmented route expansion	Available	New route coverage extends the coordinate domain	Trainable	MCR with RAER	MCR + RAER. Freezing the readout prevents coordinate-domain extension; redundant replay wastes capacity.

C Historical Support and Plasticity Regimes

The readout policy follows the historical support available to the update. Changed revisits retain an established coordinate domain but may lack old scans; segmented learning extends coordinate support while replay supplies direct constraints from earlier stages. Table A4 places these conditions on the same stability–plasticity axis.

Let H_t denote the historical support exposed to stage t: parameter state only, replayable samples, or replayable samples together with an expanding coordinate domain. DIAL selects an update configuration $\rho_t = (\mathcal{U}_t, \mathcal{G}_t, \mathcal{R}_t)$, where $\mathcal{U}_t$ is the supervision route, $\mathcal{G}_t$ is the set of trainable parameter groups, and $\mathcal{R}_t$ specifies replay. The maintained model can be written compactly as

$$\theta^t = \mathrm{Adapt}\Big(\theta^{t-1}, \widetilde{\mathcal{D}}_t, \mathcal{M}_{t-1}; \rho_t(H_t)\Big). \tag{7}$$

Here θ^{t-1} is the historical parameter state, $\widetilde{\mathcal{D}}_t$ is the current pseudo-labeled update set, and $\mathcal{M}_{t-1}$ is empty when no historical samples are available for replay. In Equation 7, Adapt denotes optimization under $\rho_t(H_t)$. The policy is selected before optimization: the same current supervision admits different gradient paths under different H_t.

C.1 The historical model across stages

The checkpoint inherited at stage t connects diagnosis to optimization. Before the update, it produces the SCR trajectory used to diagnose local failure. During HDER, its encoder parameters define the stability prior and its readout defines the fixed metric interface. After maintenance, the new checkpoint assumes these roles for the next stage. Replay extends this parameter-level memory with direct observations; the checkpoint continues to provide the diagnostic trajectory.

Table A5: Roles of the historical localizer across one maintenance stage.

Role	Observed object	Decision informed
Diagnostic model	Temporally ordered SCR poses	Which scans require renewed supervision.
Stability anchor	Encoder parameters and coordinate readout	Which components can change without replay.
Stage initialization	Complete maintained checkpoint	Where optimization for stage t begins.
Inherited state	Checkpoint after stage t	What becomes historical at stage $t+1$.

C.2 Transitions between maintenance routes

Historical support can change between stages even when the backbone remains fixed. A first changed revisit may expose only a checkpoint, placing the update in the HDER route. If representative scans are retained after that stage, a later revisit can use MCR. Segmented expansion adds a second change: the coordinate readout must represent positions outside the earlier segments, which makes replay-supported readout adaptation appropriate.

These transitions are determined by the historical support available to optimization, rather than by a fixed module choice. A checkpoint alone constrains parameters but cannot reproduce the historical input distribution. Replayable scans reinstate direct coordinate losses on earlier observations. Spatial overlap between stages then determines whether the retained set mostly repeats current coverage or contributes geometry that the current stage does not contain. RAER addresses this final distinction.

D Temporal Diagnosis and Pseudo-Label Recovery

TCPP separates diagnosis from pose assignment. The SCR trajectory indicates where the implicit field remains valid, while retrieval supplies globally stable anchor candidates. Temporal inconsistency routes observations to maintenance; retrieval or anchor-based ICP then assigns their pseudo poses (Algorithm 2).

Input: ordered revisit $\mathcal{X}_t$, historical SCR localizer $f_{\theta^{t-1}}$, retrieval localizer r
Output: update set $\mathcal{U}_t$ and pseudo-labeled set $\widetilde{\mathcal{D}}_t$

1 Predict $\mathcal{T}^{\mathrm{S}} \leftarrow f_{\theta^{t-1}}(\mathcal{X}_t)$ and $\mathcal{T}^{\mathrm{R}} \leftarrow r(\mathcal{X}_t)$.
2 Compute $\mathbf{q}^{\mathrm{S}} \leftarrow \mathrm{TemporalGate}(\mathcal{T}^{\mathrm{S}}, \kappa_{\mathrm{S}})$ and apply the five-frame drift test in Equation 9.
3 Compute $\mathbf{q}^{\mathrm{R}} \leftarrow \mathrm{TemporalGate}(\mathcal{T}^{\mathrm{R}}, \kappa_{\mathrm{R}})$.
4 Set $\mathcal{U}_t$ to the union of SCR residual failures and drift-marked windows; initialize anchors with reliable SCR poses.
5 For each $i \in \mathcal{U}_t$ with $q_i^{\mathrm{R}} = 1$, set $\widetilde{\mathbf{T}}_i \leftarrow \mathbf{T}_i^{\mathrm{R}}$ and add i as an anchor.
6 For each unanchored interval, estimate adjacent relative transforms by ICP.
7 Propagate global poses from the trusted interval boundary using Equation 8.
8 Return $\mathcal{U}_t$ and $\widetilde{\mathcal{D}}_t = \{(\mathbf{P}_i, \widetilde{\mathbf{T}}_i) : i \in \mathcal{U}_t\}$.

Algorithm 2: Temporal Consistency-guided Pseudo-label Pipeline

D.1 Recovering an unreliable interval

The update set commonly contains consecutive indices. TCPP partitions these indices into maximal intervals and resolves their coordinate frame before training. A reliable retrieval inside an interval becomes a new global anchor. Where retrieval remains rejected, adjacent-scan ICP supplies relative motion from the trusted interval boundary. For pseudo pose $\widetilde{\mathbf{T}}_i$, trusted anchor a with global pose $\mathbf{T}_a$, and adjacent relative transform $\Delta\mathbf{T}_{j\leftarrow j-1} \in SE(3)$, propagation follows

$$\widetilde{\mathbf{T}}_i = \begin{cases} \mathbf{T}_a \displaystyle\prod_{j=a+1}^{i} \Delta\mathbf{T}_{j\leftarrow j-1}^{-1}, & a < i, \\ \mathbf{T}_a \Delta\mathbf{T}_{a\leftarrow a-1} \cdots \Delta\mathbf{T}_{i+1\leftarrow i}, & a > i. \end{cases} \tag{8}$$

The ordered product is evaluated in acquisition order, preserving the transform convention used by the localizer.

For a bounded interval, propagation proceeds backward from its later trusted boundary. A terminal suffix without a later boundary is propagated forward from the preceding anchor. These two cases correspond to the two ICP branches in Equation 8.

Table A6: Pose recovery for a maximal SCR-rejected interval.

Interval evidence	**Recovery path**
Reliable retrieval at frame i	Assign $\mathbf{T}_i^{\mathrm{R}}$ directly and register frame i as a trusted global anchor.
Rejected retrieval with an earlier anchor	Compose adjacent ICP transforms forward from the anchor to frame i.
Rejected retrieval with a later anchor	Apply the same ordered composition in the reverse temporal direction.
No trustworthy anchor in the bounded context	Defer the interval until additional context is available; do not train from an unreferenced local chain.

This separation keeps two errors conceptually distinct. Temporal diagnosis can correctly identify a stale SCR region even when its first recovered global pose is uncertain. Pose recovery can also produce a smooth trajectory outside the region that actually requires adaptation. TCPP accepts a training pair only after the diagnostic route and the pose source have both been resolved.

TCPP hyperparameters. For all NCLT scene-evolution experiments, SCR temporal gating uses the MAD coefficient $\kappa_{\mathrm{S}} = 4.0$. Continuous SCR drift is detected with a window of 5 frames and a drift threshold of 4.0 m. Retrieval gating uses $\kappa_{\mathrm{R}} = 18.0$ for the BEVPlace++ trajectory. The main experiments use this default configuration. For sensitivity analysis, we vary one control at a time when constructing continual supervision from `2012-05-26`: a 3.0 m drift threshold, $\kappa_{\mathrm{R}} = 12.0$, $\kappa_{\mathrm{S}} = 3.0$, or a 7-frame window. Each resulting model is evaluated on the changed/unseen route `2012-05-11`; this route does not provide pseudo labels or training data. The default gives the lowest mean translation error, while the alternatives differ by 0.02–0.24 m and retain similar median translation and rotation accuracy.

D.2 Temporal residual and decision states

The implementation augments the residual gate with a centered five-frame drift test. Let $W_i = \{j : \max(1, i-2) \le j \le \min(N, i+2)\}$ and $\bar{\mathbf{p}}^{\mathrm{S}}_{W_i} = |W_i|^{-1} \sum_{j \in W_i} \mathbf{p}^{\mathrm{S}}_j$. With $\delta = 4.0$ m, the routed set is

$$\mathcal{U}_t = \{i : q_i^{\mathrm{S}} = 0\} \cup \bigcup_{i : \|\mathbf{p}_i^{\mathrm{S}} - \bar{\mathbf{p}}^{\mathrm{S}}_{W_i}\|_2 > \delta} W_i. \tag{9}$$

Thus an excessive local displacement contributes its complete centered window, matching the temporal grouping used to construct the training files.

For three consecutive positions, the residual in Equation 1 is the excess length of the two-step path over its endpoint displacement. The triangle inequality gives $e_i^m \ge 0$. It approaches zero under locally straight, monotone motion and increases when the middle estimate introduces a turn, reversal, or jump unsupported by its neighbors. The statistic is therefore insensitive to the absolute map origin and reacts to local trajectory structure. Median–MAD normalization further limits the influence of a small number of large failures when the gate threshold is formed.

The SCR and retrieval gates answer different questions. SCR determines whether the historical metric field is locally credible; retrieval determines whether a global pose candidate can anchor recovery. Table A7 enumerates the resulting states. Only SCR-rejected observations enter the update set, preventing already stable portions of the revisit from dominating training.

Table A7: TCPP decision states. A reliable retrieval is used as a global anchor; otherwise pose recovery proceeds from the trusted interval boundary through ordered ICP transforms.

q_i^{S}	q_i^{R}	**Diagnosis**	**Action**
1	1	Historical field remains locally stable	Keep the observation outside the maintenance set.
1	0	SCR is stable despite uncertain retrieval	Retain the historical prediction; no update is requested.
0	1	Historical field is unreliable; retrieval supplies a stable global candidate	Route the scan to maintenance and use $\mathbf{T}_i^{\mathrm{R}}$ as its pseudo pose.
0	0	Both direct estimates are temporally unreliable	Propagate from a trusted anchor by ICP; defer if no trusted anchor is available.

The fixed-budget ablation separates these two decisions. Replacing TCPP routing with samples outside its diagnosed pool preserves the label count and pose-recovery machinery but directs optimization away from regions where the historical field fails. Removing temporal cross-validation instead weakens the quality of the global anchors. The controls isolate targeted supervision from pseudo-pose construction.

E RAER UNDER SEGMENTED CONTINUAL LOCALIZATION

The three-stage protocol extends the represented route while retaining spatial overlap with earlier stages. At each stage, RAER compares replay candidates from all earlier stages with the current trajectory and removes those within the 2 m nearest-distance threshold used in training.

E.1 SPATIAL EXCLUSION RULE

Input: old memory $\mathcal{M}_{t-1}$, current stage $\mathcal{D}_t$, overlap radius r
Output: filtered replay memory $\mathcal{M}_{t-1}^{\mathrm{RAER}}$

1 Extract old and current XY positions $\mathcal{Z}_{\mathcal{M}}$ and $\mathcal{Z}_t$.
2 Build a nearest-neighbor index over $\mathcal{Z}_t$.
3 Initialize $\mathcal{M}_{t-1}^{\mathrm{RAER}} \leftarrow \emptyset$.
4 For each old exemplar m_i, query $d_i = \min_{\mathbf{z}\in\mathcal{Z}_t} \|\mathbf{z}_i - \mathbf{z}\|_2$.
5 If $d_i \geq r$, insert m_i into $\mathcal{M}_{t-1}^{\mathrm{RAER}}$; otherwise discard it as spatially redundant.
6 Return $\mathcal{M}_{t-1}^{\mathrm{RAER}}$.

Algorithm 3: Redundancy-Aware Exemplar Replay

E.2 SELECTION ORDER UNDER A FIXED BUDGET

RAER precedes the final coverage selector. Historical candidates are first subsampled along each earlier segment, then tested against the current-stage XY trajectory. Candidates inside the exclusion radius are removed before farthest-point coverage selection imposes the memory cap. Reversing these two operations would allow current-stage redundancy to consume the budget before complementary historical regions are considered.

Table A8: RAER memory construction used in the matched 8.0k-frame comparison, from historical candidate extraction to bounded raw-scan replay.

Step	Operation	Effect on historical support
1	Stride-4 candidate extraction	Reduces temporal duplication within each earlier trajectory.
2	2 m XY exclusion	Removes old frames whose positions are already represented by the current stage.
3	Pose-space farthest-point selection	Distributes the retained candidates over complementary route coverage.
4	8.0k memory cap	Matches the replay count used by the coverage-sampling baseline.
5	Raw-scan buffer replay every two batches	Supplies encoder and trainable-readout constraints during stages 2 and 3.

The nearest-distance test is implemented with a spatial index over current-stage positions. For N_t current positions and M historical candidates, index construction is $O(N_t \log N_t)$ and the batch of nearest-neighbor queries is $O(M \log N_t)$ under the standard balanced-tree assumption. This filtering is performed once when the replay set is assembled, rather than inside every optimization step. The matched comparison tests whether this remaining support is more useful than memory selected without the exclusion prior.

E.3 Interpreting the exclusion radius

The radius controls a geometric equivalence test, not the final memory size. At $r = 0$, RAER retains every historical candidate and reduces to the unfiltered replay pool. Increasing r removes old observations whose positions already fall within current-stage coverage; the downstream selector can then spend its fixed capacity on more distant historical support. Table A3 shows that $r = 1$ m and $r = 2$ m are indistinguishable at the reported precision, whereas $r = 3$ m begins to discard useful support.

RAER uses position-space overlap. Its memory therefore remains independent of the current encoder as representation learning proceeds. Spatially close scans from substantially different viewpoints can still contain distinct evidence; the segmented protocol keeps the XY criterion fixed so that this geometric prior is evaluated separately from learned similarity.

F MCR Memory Realizations and Replay Cadence

F.1 Replay as a gradient constraint

MCR inserts historical geometry into selected optimization steps. At an active replay step, the current pseudo-labeled batch and a historical batch contribute separate coordinate-regression losses; their weighted sum updates the permitted model components. Dataset replay and a raw-scan buffer reconstruct the historical forward path through the encoder. Feature replay begins after that path and can therefore constrain only the readout associated with the stored representation. Algorithm 4 gives the interleaved update.

Input: historical checkpoint $f_{\theta^{t-1}}$, update set $\widetilde{\mathcal{D}}_t$, memory $\mathcal{M}_{t-1}$, replay interval K, weight β
Output: maintained parameters θ^t

Initialize $\theta \leftarrow \theta^{t-1}$ and construct current-stage mini-batches.
For optimization step b, compute $\mathcal{L}_{\text{new}}$ on the current pseudo-labeled batch.
If $b \bmod K = 0$, sample a historical batch from $\mathcal{M}_{t-1}$ and compute $\mathcal{L}_{\text{rep}}$ at the configured replay site.
Use $\mathcal{L} = \mathcal{L}_{\text{new}} + \beta\mathcal{L}_{\text{rep}}$ at replay steps; otherwise use $\mathcal{L} = \mathcal{L}_{\text{new}}$.
Update the parameter groups enabled by the selected replay and readout policy.
Return $\theta^t \leftarrow \theta$.

Algorithm 4: Memory-Constrained Replay update

Table A9: MCR memory realizations and their optimization paths.

Memory	Replay entry	Historical constraint
Dataset scans	Encoder input	Reconstructs the complete historical forward path from stored scans and poses.
Raw-scan buffer	Encoder input	Uses bounded in-memory scans and poses while preserving encoder-side gradients.
Feature buffer	Readout input	Reuses stored encoder features; avoids historical feature extraction but fixes the representation at storage time.
Raw scan with head-side replay	Readout training path	Computes historical features while restricting the replay update to the pose readout.

The entries indicate which gradient paths the replay site makes available; the selected maintenance regime still determines whether each parameter group is trainable.

F.2 Fixed-capacity replay cadence

MCR separates how densely history is stored from how often it is replayed. Table A10 holds the historical stride at $s = 4$ (6,697 raw scans) and varies only the replay interval K. HDER regularization is disabled in every variant, isolating replay cadence.

Table A10: Sensitivity to the replay interval in MCR. Historical sampling density and memory capacity are fixed; replay frequency is $1/K$. TE/RE denotes mean translation error in meters and rotation error in degrees.

Historical stride s	Replay interval K	Replay frequency	Historical TE/RE↓	Evolved / unseen TE/RE↓	Aggregate TE/RE↓
4	4 **(default)**	0.250	**1.43/4.66**	**2.87/6.28**	**1.79/5.07**
4	6	0.167	1.67/5.47	3.06/7.16	2.02/5.89
4	8	0.125	1.68/5.38	3.11/6.87	2.04/5.75

Historical error averages the three old NCLT sequences, while evolved/unseen error is measured on `2012-05-11`; the aggregate averages all four and excludes the revisit used for continual supervision. Reducing replay frequency degrades both retention and adaptation: moving from $K = 4$ to $K = 6$ or 8 increases error on both historical routes and the evolved route. Historical replay acts as more than a constraint on forgetting; at a sufficient cadence, it anchors the geometric optimization trajectory and prevents new-scene updates from drifting toward a poorer coordinate solution.

F.3 Separating cadence from memory capacity

With deterministic replay every K steps, a run of B optimization steps contains $\lceil B/K \rceil$ historical updates. Changing K therefore alters the temporal spacing of historical gradients while leaving the stored examples unchanged. This differs from changing stride s, which alters which parts of the old trajectory can be sampled even if replay is invoked equally often.

The distinction is important for interpreting Table A10. Its rows hold $s = 4$, the 6,697-scan memory, the replay loss weight, and the current-stage batches fixed. The observed change is attributable to how long optimization proceeds without an explicit historical constraint. A large interval permits several consecutive pseudo-supervised steps to move the representation before the next replay correction; the correction may then arrive after the optimizer has entered a less favorable region.

For segmented learning, $K = 2$ increases historical exposure relative to the NCLT scene-evolution setting. Each segment extends coordinate support, and the trainable readout consequently receives more frequent old-stage constraints. The segmented protocol therefore uses $K = 2$. RAER determines which examples are retained; K determines when they influence training.

F.4 GRADIENT VISIBILITY AT THE REPLAY SITE

The MCR objective at step b can be written as

$$\mathcal{L}_b = \mathcal{L}_{\mathrm{new},b} + \mathbb{I}[b \bmod K = 0]\, \beta \mathcal{L}_{\mathrm{rep},b} + \lambda_b \mathcal{L}_{\mathrm{prior}}, \tag{10}$$

where $\mathcal{L}_{\mathrm{new},b}$ and $\mathcal{L}_{\mathrm{rep},b}$ are the current and historical coordinate losses, $\mathbb{I}[\cdot]$ activates replay every K steps, β is the replay weight, and $\lambda_b \mathcal{L}_{\mathrm{prior}}$ is present only when MCR is combined with HDER. The replay site determines which parameter groups receive the second term; it does not change the current-stage loss.

Table A11: Gradient visibility for the historical loss in Equation 10.

Replay realization	**Encoder**	**Readout**	**Consequence**
Encoder-side raw replay	Historical gradient	Historical gradient	Historical loss constrains both the encoder representation and coordinate readout.
Head-side raw replay	No historical gradient	Historical gradient	Historical scans are encoded online, while replay gradients update only the readout.
Head-side feature replay	No historical gradient	Historical gradient	Stored historical features bypass encoder computation and update only the readout.

Combining MCR with HDER activates two distinct forms of retention in Equation 10: replay constrains predictions on historical samples, while the parameter prior limits movement of the trainable encoder. Their joint use favors old-route retention in the main ablation, with a corresponding loss of new-route plasticity. This combination suits settings that place greater weight on retention; MCR alone leaves more capacity for the new route.

G Experimental Protocols and Controls

The experiments instantiate two forms of maintenance. NCLT scene evolution starts from one historical PELoc localizer; its long-term extension triggers updates when the predicted localization-drift ratio exceeds 15%. Segmented QEOxford instead expands the coordinate domain over three stages and evaluates the localizer after the complete sequence.

Initialization and evaluation splits. On NCLT, PELoc is initialized from `2012-02-18`; SGLoc-v2 uses `2012-01-22`, `2012-02-02`, and `2012-02-18`. The QEOxford stages partition `2019-01-11-14-02-26`. Evaluation uses the four held-out sequences `2019-01-15-13-06-37`, `2019-01-17-13-26-39`, `2019-01-17-14-03-00`, and `2019-01-18-14-14-42`.

Table A12 summarizes the acquisition conditions of the NCLT traversals used to construct the long-term protocol. The sequence set spans seasonal appearance, foliage, illumination, and snow conditions rather than a single type of environmental change.

Table A12: NCLT sequence metadata used to contextualize long-term localization and maintenance experiments. Route lengths and acquisition conditions follow the NCLT sequence records (Carlevaris-Bianco et al., 2016).

Date	Length	Time	Sky	Foliage	Snow
2012-01-08	6.4 km	Midday	Partly cloudy	No	No
2012-01-15	7.5 km	Afternoon	Sunny	No	Yes
2012-01-22	6.1 km	Afternoon	Cloudy	No	Yes
2012-02-02	6.2 km	Afternoon	Sunny	No	No
2012-02-04	5.5 km	Afternoon	Sunny	No	No
2012-02-05	6.5 km	Morning	Sunny	No	No
2012-02-12	5.8 km	Midday	Sunny	No	Yes
2012-02-18	6.2 km	Evening	Sunny	No	No
2012-02-19	6.2 km	Midday	Partly cloudy	No	No
2012-03-17	5.8 km	Morning	Sunny	No	No
2012-03-25	5.8 km	Midday	Sunny	No	No
2012-03-31	6.0 km	Midday	Cloudy	No	No
2012-04-29	3.1 km	Morning	Sunny	Yes	No
2012-05-11	6.0 km	Midday	Sunny	Yes	No
2012-05-26	6.3 km	Evening	Sunny	Yes	No
2012-06-15	4.1 km	Morning	Sunny	Yes	No
2012-08-04	5.5 km	Morning	Sunny	Yes	No
2012-08-20	6.0 km	Evening	Sunny	Yes	No
2012-09-28	5.6 km	Evening	Sunny	Yes	No
2012-10-28	5.6 km	Midday	Cloudy	No	No
2012-11-04	4.8 km	Morning	Cloudy	No	No
2012-11-16	4.8 km	Evening	Sunny	No	No
2012-11-17	5.7 km	Midday	Sunny	No	No
2012-12-01	5.0 km	Evening	Sunny	No	No
2013-01-10	1.1 km	Afternoon	Cloudy	No	Yes
2013-02-23	5.2 km	Afternoon	Sunny	No	Yes
2013-04-05	4.5 km	Afternoon	Sunny	No	Yes

Table A13: Supervision, historical replay, and controlled resources in the three continual-localization protocols.

Experiment	Supervision and historical support	Matched controls and evaluation
NCLT scene evolution	TCPP pseudo poses from `2012-05-26`; no replayable historical scans for HDER, or buffer/dataset replay for MCR	Same PELoc localizer and evaluation routes. CoTTA, SDM, and Fisher-EWC use full-revisit updates; ConDo uses the shared pseudo-label pool with uniform sampling; DIAL uses the TCPP-selected subset. Retention: `02-12`, `02-19`, `03-31`; changed/unseen: `2012-05-11`.
NCLT long-term maintenance	DIAL-M_{enc} with TCPP pseudo poses, encoder-side dataset replay, and a frozen readout	An update is triggered above a 15% predicted localization-drift ratio; replay accumulates earlier update routes. Sequence-wise TE/RE is reported across 2012–2013 revisits.
Segmented QEOxford	Current-stage supervision with encoder replay from earlier stages	Same training steps; the 8.0k comparison matches CL-HSCNet replay count. Final mean TE/RE is measured after three stages.

G.1 Shared controls

For NCLT, TCPP uses $\kappa_S = 4.0$, a five-frame drift window, and a 4.0 m drift threshold; retrieval gating uses $\kappa_R = 18.0$. MCR uses historical stride $s = 4$ and replay interval $K = 4$ on NCLT, whereas segmented QEOxford uses $K = 2$. RAER uses a 2 m XY exclusion radius. HDER applies its encoder prior through the cosine schedule in Equation 6. Optimization time is wall-clock continual- training time; it excludes data collection and any full-rebuild preparation.

All NCLT localization errors are mean translation / rotation error in meters / degrees. For the primary NCLT scene-evolution experiments, historical error averages the three retention traversals, evolved/unseen error is evaluated on `2012-05-11`, and aggregate error averages these four routes. The long-term protocol instead reports sequence-wise errors across the 2012–2013 revisits. QEOxford reports final mean TE/RE after the three-stage sequence. Replay comparisons control memory through the retained-frame count, which directly fixes the number of historical observations available to the update.

H Computational Structure of DIAL

DIAL adds computation during diagnosis and optimization. TCPP processes the ordered revisit before training; MCR and RAER alter the historical observations used by the update. HDER evaluates a parameter prior over the trainable encoder blocks. None of these operations changes the inference graph of the maintained localizer.

Let N be the number of revisit scans, U the number routed to maintenance, M the number of historical candidates, B the number of optimization steps, and C_S, C_R, and C_I the costs of one SCR, retrieval, and adjacent-scan ICP evaluation. Table A14 expresses the additional computation relative to ordinary fine-tuning. These terms isolate algorithmic cost from hardware-dependent wall time.

Table A14: Additional maintenance cost. d_ϕ is the number of trainable encoder parameters, K is the replay interval, and C_{rep} is one historical replay forward–backward pass.

Module	One-time computation	Effect on optimization	Persistent state
TCPP	$O(N(C_{\mathrm{S}} + C_{\mathrm{R}}) + UC_{\mathrm{I}})$	Optimizes on the routed subset U instead of all N scans	Pseudo poses and binary routing decisions
HDER	Parameter masking at initialization	$O(Bd_\phi)$ for the encoder prior	Historical parameter snapshot
MCR	Memory construction or feature extraction	$O(\lceil B/K \rceil C_{\mathrm{rep}})$	Raw scans, features, or dataset access
RAER	$O(N \log N + M \log N)$ nearest-position filtering	No per-step filtering cost; replay follows MCR	Filtered historical exemplar set

H.1 Inference and checkpoint state

The historical checkpoint, TCPP diagnostics, and any replay memory are used only during maintenance and are not part of the deployed forward pass. The updated implicit localizer retains the same scan encoder, coordinate readout, and metric pose solver as its backbone, leaving deployment-time model size unchanged. For a future update, the maintained checkpoint becomes the next historical model, and any retained MCR or RAER memory becomes the incoming historical support.

I Interface to Implicit Localizers

DIAL assumes an implicit localizer with three accessible operations: scan encoding, scene-coordinate prediction, and metric pose recovery. TCPP observes the pose sequence produced by the historical model and does not depend on its internal feature dimensionality. HDER and MCR require a boundary between the encoder and coordinate readout, while RAER operates on trajectory positions before a replay batch is formed. Both evaluated backbones, PELoc and SGLoc-v2, expose this interface.

I.1 What transfers and what remains backbone-specific

The maintenance policy transfers without redefining the backbone's localization objective. Coordinate-loss details, voxelization, receptive fields, and the robust pose solver remain properties of the underlying localizer. The layer boundary used by HDER and the feature tensor stored by head-side MCR must be chosen for each architecture. In contrast, TCPP's route diagnosis and RAER's spatial exclusion act on metric outputs and are unchanged once coordinate-frame conventions are aligned.

Transfer is most direct when the backbone exposes an encoder–readout boundary and globally referenced metric poses. PELoc and SGLoc-v2 satisfy these conditions, although their update dynamics need not coincide. Architectures that entangle pose recovery with the encoder require another stable interface; localizers without global pose outputs require a different temporal gate.

Table A15: DIAL's interface to an implicit localization backbone. The rightmost column identifies the assumption required for transferring each module.

Module	Backbone access	Backbone-independent operation	Required assumption
TCPP	Predicted metric poses	Temporal gating, retrieval validation, and ICP propagation	Poses share a common global frame and scans are temporally ordered.
HDER	Encoder/readout parameter boundary	Frozen-readout policy and scheduled historical prior	The readout contains the scene-coordinate mapping whose overwrite causes forgetting.
MCR	Encoder input or readout input	Replay cadence and loss composition	Historical samples or compatible stored features are available.
RAER	Historical and current trajectory positions	XY exclusion and budgeted exemplar selection	Earlier and current stages use the same metric frame.

J Discussion on Streaming Updates

DIAL updates after a revisit interval. TCPP estimates its robust threshold over the trajectory. The drift decision requires a temporal window, and unreliable retrieval intervals may need a later trusted anchor before ICP propagation can assign global poses. Immediate parameter updates would act before these observations can be distinguished from transient localization noise.

The interval need not span an entire survey. A practical implementation can maintain a bounded queue, finalize an interval once its temporal context and an anchor are available, and launch the same short optimization used here. Such a system is delayed and causal at the interval level, but not scan-wise online. This preserves TCPP's diagnostic role while reducing the time between scene change and model maintenance.

Table A16: Evidence availability under scan-wise and interval-based maintenance.

Evidence	At arrival	After context	Role
Historical SCR pose	Yes	Yes	Initial local prediction and diagnostic signal.
Centered temporal residual	No	Yes	Distinguishes persistent drift from an isolated estimate.
Robust route threshold	Partial	Yes	Calibrates the gate to the revisit trajectory.
Future trusted anchor	No	When observed	Initializes ICP propagation through an unreliable interval.
Replay constraint	Yes, if stored	Yes	Stabilizes optimization after the update set is finalized.

J.1 Evidence boundaries and observable safeguards

DIAL depends on informative diagnostic signals and adequate historical support. Repetitive geometry, missing anchors, short temporal disruptions, and unsupported coordinate expansion each weaken a specific source of evidence. Table A17 summarizes plausible operating boundaries implied by these evidence requirements and the safeguards available in the current design.

Table A17: Potential operating boundaries implied by DIAL's evidence requirements and the mechanisms that limit their influence.

Condition	**Observable signature**	**Current safeguard**	**Unresolved boundary**
Perceptual aliasing	Retrieval is smooth yet inconsistent with local geometry	Cross-validation with SCR and interval-level ICP	Both cues may agree on the same wrong place.
Long anchor-free interval	Consecutive SCR and retrieval rejections	Ordered propagation from a trusted interval boundary	Relative-pose error can accumulate with interval length.
Abrupt motion or missing scans	Large residual confined to a short temporal neighborhood	Persistent-window drift test	Genuine change may be delayed until sufficient context arrives.
Severe coordinate-domain expansion	Many scans lie outside historical support	Trainable readout with MCR and RAER when historical replay is available	Expansion without replayable history remains outside the frozen-readout regime.

K Coordinate and Transform Conventions

Implicit localization connects three spaces that should not be conflated. A voxel position $\mathbf{x}_v \in \mathbb{R}^3$ is expressed in the local LiDAR frame. The readout predicts its scene coordinate $\widehat{\mathbf{c}}_v \in \mathbb{R}^3$ in the global map frame. A metric pose $\widehat{\mathbf{T}}_i \in SE(3)$ is then recovered from correspondences between local voxel positions and predicted scene coordinates. The subscript v indexes a voxel, while i indexes the scan containing it.

Table A18: Notation shared by the four maintenance modules. Superscript $t-1$ denotes state inherited from the preceding stage.

Symbol	**Meaning**
$f_{\theta^{t-1}} = h_{\psi^{t-1}} \circ g_{\phi^{t-1}}$	Historical implicit localizer, decomposed into encoder and coordinate readout.
$\mathbf{P}_i \subset \mathbb{R}^3$; $\mathbf{T}_i \in SE(3)$	LiDAR point set and its ground-truth or pseudo 6-DoF pose.
$\mathbf{p}_i^{\mathrm{S}} \in \mathbb{R}^2$; $\mathbf{p}_i^{\mathrm{R}} \in \mathbb{R}^3$	Planar SCR and full 3D retrieval translations used for temporal gating.
$q_i^{\mathrm{S}}, q_i^{\mathrm{R}}$	Binary temporal reliability decisions for SCR and retrieval.
$\mathcal{U}_t, \widetilde{\mathcal{D}}_t$	Routed update indices and their recovered pseudo-labeled scans.
$\mathcal{M}_{t-1}$	Replayable historical samples or representations available at stage t.
s, K	Historical sampling stride and replay interval.
r	RAER exclusion radius in the shared XY frame.
$\lambda(u)$	HDER encoder-prior weight at normalized training progress u.

For scan i, $\mathbf{T}_i \in SE(3)$ contains a rotation $\mathbf{R}_i \in SO(3)$ and a global translation $\mathbf{t}_i \in \mathbb{R}^3$. Its six degrees of freedom comprise three translation and three rotation components. Ground-truth and TCPP pseudo poses share this representation, so either source can supervise the same coordinate-regression loss.

Localization errors. For an estimated pose $(\widehat{\mathbf{R}}_i, \widehat{\mathbf{t}}_i)$ and its reference $(\mathbf{R}_i, \mathbf{t}_i)$, translation error is the Euclidean distance $\|\widehat{\mathbf{t}}_i - \mathbf{t}_i\|_2$. Rotation error is the geodesic angle between $\widehat{\mathbf{R}}_i$ and $\mathbf{R}_i$, reported in degrees. We report sequence means unless a median is explicitly named; historical and evolved aggregates use the designated evaluation sequences and exclude update revisits.