BiLi: Bridging the Last Mile in LiDAR Localization

Minghang Zhu[1,2,*] **Junchuan Lan**[1,2,*] **Zhijing Wang**[1,2] **Haoze Chang**[1,2]
Wen Li[3] **Sheng Ao**[1,2,†] **Cheng Wang**[1,2]

[1]Fujian Key Laboratory of Urban Intelligent Sensing and Computing,
Xiamen University, Xiamen, China
[2]Key Laboratory of Multimedia Trusted Perception and Efficient Computing,
Ministry of Education of China, Xiamen University, Xiamen, China
[3]School of Engineering Mathematics and Technology, University of Bristol,
Bristol, United Kingdom

mihoo@stu.xmu.edu.cn 23020251154360@stu.xmu.edu.cn

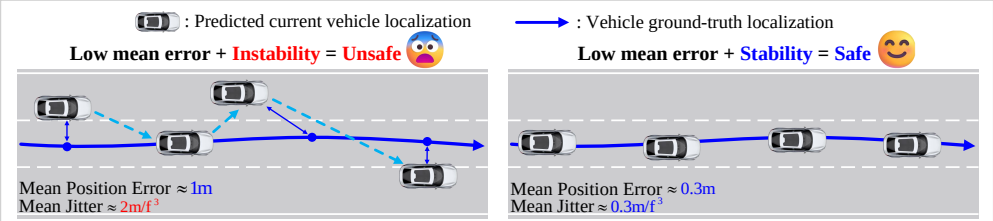

Figure 1: **The "last-mile" problem in LiDAR localization.** (A) Existing methods may achieve low mean errors overall, yet they suffer from severe localization instability. These sudden, unphysical location jumps pose critical safety risks for lane-level autonomous driving. (B) Our proposed BiLi framework fundamentally resolves this challenge. It delivers high-precision, jump-free, and exceptionally stable localization, meeting the strict safety prerequisites for real-world deployment.

Abstract

In large-scale outdoor LiDAR localization, Scene Coordinate Regression (SCR) achieves sub-meter accuracy, but its deployment in high-precision autonomous driving is hindered by a "last-mile" problem: the point-wise prediction paradigm causes unstable localization jitter, rendering localization kinematically discontinuous despite low mean errors. To bridge the gap between high-precision localization and temporal kinematic consistency, we propose BiLi, a spatiotemporal synergistic relocalization framework. Spatially, BiLi utilizes Implicit Manifold Regularization (IMR) driven by an offline neural implicit field. Formulated as a novel point-to-manifold projection constraint, IMR strictly anchors predicted coordinates to continuous physical zero-level sets. Temporally, to bypass computationally heavy implicit optimization, we distill the teacher's kinematics into a lightweight network, compelling it to learn robust, equivariant physical priors. Finally, a Differentiable Robust Manifold Gating mechanism dynamically fuses the global spatial predictions with the temporal kinematic states via adaptive state updating, facilitating efficient real-time deployment. Extensive experiments demonstrate that BiLi fundamentally resolves the "last-mile" challenge. Our method outperforms state-of-the-art approaches by 25% and 65% on the Oxford and NCLT benchmarks, respectively, delivering highly accurate, real-time, and stable localization. https://reckozmh.github.io/BiLi

[*]Equal contribution.
[†]Corresponding author.

40th Conference on Neural Information Processing Systems (NeurIPS 2026).

1 Introduction

Accurate and robust LiDAR localization [1, 2, 3, 4, 5] is a fundamental task in autonomous driving and robotics [6, 7]. While traditional map-matching methods [8, 9, 10, 11] are effective, they require massive 3D map storage and high communication bandwidth, which limits their use in large-scale outdoor environments. Therefore, end-to-end regression networks have become a popular alternative [12, 13, 14, 15]. By memorizing the scene directly within the network weights, they remove the need to store explicit maps during inference.

These regression-based methods fall into two main categories: Absolute Pose Regression (APR) and Scene Coordinate Regression (SCR). APR directly predicts the 6-DoF sensor pose from the input point cloud [16, 17]. However, this direct regression design struggles to explicitly capture complex scene geometry, limiting its localization accuracy. In contrast, SCR predicts the global coordinates for each local point and then uses RANSAC [18] to calculate the pose. By explicitly leveraging geometric consistency, SCR consistently outperforms APR and has become the mainstream approach. Recently, by incorporating semantic or textual information, advanced SCR methods have successfully reduced localization errors to the sub-meter level.

However, a critical issue remains when deploying these highly accurate SCR models in real-world autonomous driving: low mean errors do not guarantee safe navigation. As shown in Figure 1, current SCR approaches [19, 20, 21, 22, 23] might achieve sub-meter average accuracy, but it still occasionally produces sudden, lane-level location jumps. Fundamentally, this issue stems from representation confounding, a problem inherent in predicting each point independently. Temporally, these spatial errors translate into high-frequency, unphysical fluctuations in localization. For downstream planning and control modules, such unstable localization is disastrous. Even infrequent location jumps can trigger dangerous steering shakes or emergency braking, posing severe safety risks. Therefore, delivering both high precision and absolute localization stability is the essential prerequisite for real-world application [24, 25].

We refer to the simultaneous achievement of high precision and localization stability as the **"last-mile"** challenge of outdoor localization. To address this, we identify two fundamental limitations in current SCR methods: 1) Spatially, Current models [26, 23] primarily use point-to-point L1 distance as the training loss. These microscopic losses blindly minimize numerical errors while ignoring the underlying continuous topological manifold of the scene, allowing predicted coordinates to unnaturally "float" in free space and causing spatial ambiguity. 2) Temporally, Predicting each frame independently ignores the historical motion context. Without explicit kinematic constraints, the network easily produces physically impossible location jumps that violate the continuous nature of real vehicle motion.

To resolve the "last-mile" challenge, we propose BiLi, a spatiotemporal synergistic architecture for high-precision and exceptionally stable LiDAR localization. Guided by an offline neural implicit Teacher, BiLi ensures rigorous spatial consistency and temporal stability during real-time inference. Spatially, we introduce Implicit Manifold Regularization (IMR). IMR adopts a point-to-manifold projection to address the representation confounding problem in SCR. Temporally, we propose Mamba-Driven Kinematic Distillation (MDKD) to efficiently extract continuous motion priors with linear complexity. To fuse the motion priors with SCR predictions, we design Dynamic Manifold Gating (DMG), which dynamically adjusts its fusion tolerance based on sequential hidden states to achieve high-precision and stable localization. Our main contributions are summarized as follows:

- We define the "last-mile" challenge in deploying LiDAR relocalization and propose BiLi, a spatiotemporal synergistic architecture, to resolve it.
- We introduce Implicit Manifold Regularization (IMR) to perform point-to-manifold projection, fundamentally resolving spatial representation confounding in SCR.
- We design MDKD to distill continuous motion priors and construct DMG for real-time pose fusion, achieving high-precision and stable localization.
- Extensive experiments demonstrate that BiLi achieves state-of-the-art performance, notably improving localization accuracy by 25% and 65% on the Oxford Radar RobotCar and NCLT datasets, respectively.

2 Related Work

LiDAR Localization. Traditional LiDAR localization frameworks primarily rely on matching incoming point clouds against pre-constructed 3D global maps [8, 9, 11, 27, 28]. While these map-based strategies can deliver reliable pose estimation [29, 30], their practical deployment is severely bottlenecked by exorbitant memory footprints and substantial communication bandwidth requirements, especially in large-scale outdoor environments. To circumvent the dependency on explicit map storage, regression-based (map-free) localization has emerged as a promising alternative. By directly embedding the environmental geometry into neural network weights [16, 19, 31], these approaches significantly reduce online memory consumption and are generally categorized into Absolute Pose Regression (APR) and Scene Coordinate Regression (SCR).

Absolute Pose Regression (APR). APR pipelines are designed to directly output the 6-DoF sensor pose from the input scan in an end-to-end manner [16, 32, 33, 17]. Early pioneering explorations demonstrated the viability of this paradigm by pairing foundational point cloud feature extractors with pose regression heads [16]. Subsequent research has continuously pushed the boundaries of APR by incorporating multi-modal fusion mechanisms [34, 35, 36, 37], spatio-temporal consistency constraints [31], and even iterative diffusion optimization processes [17]. However, because APR typically compresses the entire scene observation into a holistic global descriptor, it inherently loses fine-grained geometric details. This black-box mapping struggles to fully exploit local structural constraints, which fundamentally limits its absolute localization precision.

Scene Coordinate Regression (SCR). To overcome APR's precision bottlenecks, SCR predicts dense correspondences between local LiDAR points and global coordinates, subsequently solving for poses using RANSAC [18]. SGLoc [19] pioneered this direction by decoupling correspondence regression from pose estimation to better capture geometric structures. Recognizing vulnerabilities to visual ambiguity and structural degradation, recent studies have explored various feature enhancements: LiSA [20] incorporates semantic distillation to better understand scene layouts, RALoc [21] introduces rotation-aware feature canonicalization against viewpoint variations, LightLoc [26] utilizes pre-trained encoders to drastically reduce training costs, and GTRLoc [23] introduces geospatial textual regularizations to resolve ambiguities in similar regions. Despite these advancements, current SCR methods still predominantly optimize point-wise Euclidean losses. This frame-by-frame paradigm ignores continuous physical manifolds and kinematic constraints, inevitably causing the unstable localization jumps that our BiLi framework aims to resolve.

Neural Implicit Representations. Neural implicit representations model continuous 3D geometry within coordinate-based networks. Techniques like DeepSDF [38] represent spatial topology through a continuous Signed Distance Function (SDF), where the physical surface is defined as the zero-level set. To handle large-scale and unbounded environments, PIN [39] anchors latent features to sparse physical points for scalable and detailed scene reconstruction. Inspired by these advancements, our work introduces neural implicit representations into LiDAR localization. Rather than using them for novel view synthesis, we leverage a pre-built offline neural SDF field as a rigorous microscopic geometric constraint, guiding our network to project coordinate predictions onto actual physical manifolds and eliminating spatial ambiguity.

3 Methodology

To overcome the "last-mile" challenge, as shown in Figure 2, we propose **BiLi**, a spatiotemporal synergistic architecture. Guided by an offline neural implicit Teacher, BiLi introduces Implicit Manifold Regularization (Sec. 3.1) for strict spatial consistency, Mamba-Driven Kinematic Distillation (Sec. 3.2) for continuous temporal priors, and Dynamic Manifold Gating (Sec. 3.3) for adaptive pose fusion, ultimately achieving high-precision, stable localization.

3.1 Implicit Manifold Regularization (IMR) for Spatial Consistency

Offline Neural Implicit Mapping and Pose Estimation. Drawing inspiration from DeepSDF [38] and PIN [39], we represent the scene geometry as the zero-level set of a continuous Signed Distance Function (SDF) field $\mathcal{S} : \mathbb{R}^3 \to \mathbb{R}$. Physically, the SDF encodes the shortest orthogonal distance from any spatial coordinate to the nearest physical surface, serving as a robust geometric prior. The scene is encoded into an elastic, point-based neural implicit map $\mathcal{M} = \{(\mathbf{p}_j, \mathbf{q}_j, \mathbf{f}_j, \mathbf{c}_j)\}_{j=1}^{M}$. Here,

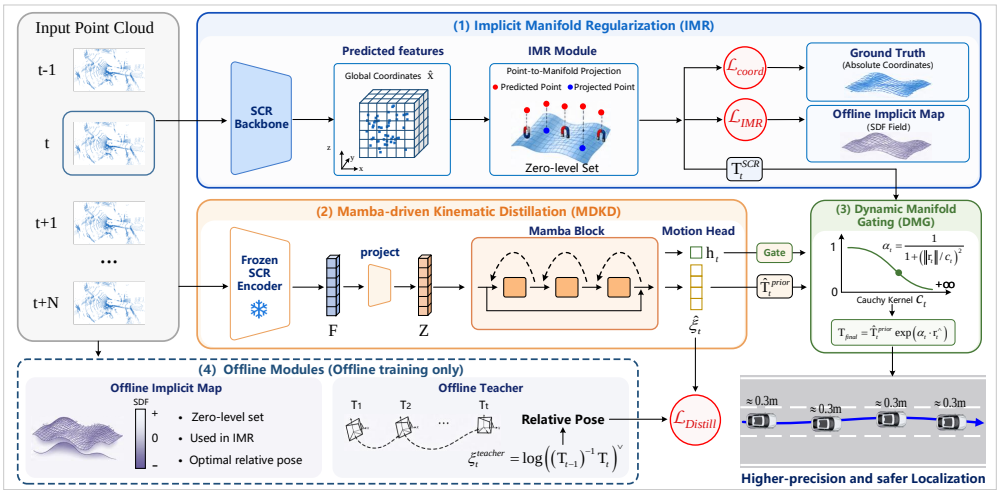

Figure 2: **Overview of the proposed BiLi framework.** (1) In the spatial domain, IMR constrains representation learning by anchoring absolute predictions to the physical zero-level set via manifold projection. (2) In the temporal domain, MDKD distills offline kinematics into a Mamba block to extract a stable kinematic prior $\hat{\mathbf{T}}_t^{prior}$ and hidden states $\mathbf{h}_t$. (3) DMG dynamically fuses the spatial observation $\mathbf{T}_t^{SCR}$ and the kinematic prior via an $\mathbf{h}_t$-guided Cauchy kernel. (4) The Offline Modules act as a high-fidelity Teacher, providing spatiotemporal supervision exclusively during training.

each neural point is defined in the world frame by a spatial position $\mathbf{p}_j \in \mathbb{R}^3$ and a quaternion $\mathbf{q}_j \in \mathbb{R}^4$ representing its local frame orientation. $\mathbf{f}_j \in \mathbb{R}^8$ is the optimizable latent geometric feature, and $\mathbf{c}_j \in [0,1]^8$ is the local semantic descriptor derived from the encoder [26] via MLP projection. To serve as an authoritative Teacher, the system derives the optimal relative pose $\mathbf{T}^* \in SE(3)$ by minimizing a joint geometric-semantic residual for each LiDAR point p_i:

$$\mathbf{T}^* = \arg\min_{\mathbf{T}} \sum_i \left(\mathcal{L}_{sdf}(\mathbf{T} \circ p_i) + \mathcal{L}_{feat}(\mathbf{T} \circ p_i) \right), \tag{1}$$

where $\circ$ denotes the standard Lie group action mapping the local incoming point p_i into the global implicit map. $\mathcal{L}_{sdf}$ enforces alignment with the implicit surface, and $\mathcal{L}_{feat}$ leverages the injected semantic descriptors to ensure tracking robustness in geometrically degenerate scenarios. The resulting high-fidelity trajectory provides the necessary spatiotemporal prior for subsequent SCR regularization, the optimization details in Appendix Sec. A.

Implicit Manifold Regularization (IMR). Conventional SCR frameworks rely exclusively on point-to-point Euclidean distance losses. However, such naive Euclidean penalties completely ignore the underlying geometric topology of the scene, allowing predictions to "float" arbitrarily as long as the numerical error is minimized. To eliminate the spatial noise caused by representation confounding, we shift the paradigm from point-to-point mapping to point-to-manifold projection via our Implicit Manifold Regularization (IMR). For each predicted global coordinate $\hat{\mathbf{x}}$, we query its K-nearest neural points $\{(\mathbf{p}_j, \mathbf{q}_j, \mathbf{f}_j)\}_{j \in \mathcal{K}(\hat{\mathbf{x}})}$ from the pre-built implicit map. To guarantee strict $SE(3)$-invariance, the query $\hat{\mathbf{x}}$ is transformed into the local tangent space of each retrieved neural point:

$$\mathbf{d}_j = \mathbf{R}(\mathbf{q}_j)^T (\hat{\mathbf{x}} - \mathbf{p}_j), \tag{2}$$

where $\mathbf{R}(\mathbf{q}_j) \in SO(3)$ is the rotation matrix uniquely derived from the quaternion $\mathbf{q}_j$. The local Signed Distance Function value is then evaluated using a globally shared, shallow geometric decoder $\Psi_{\theta_{sdf}}$, which maps the concatenated latent feature $\mathbf{f}_j$ and relative coordinate $\mathbf{d}_j$ to a continuous scalar distance. Rather than a simple summation, we formulate the continuous regularization as a Nadaraya-Watson kernel-weighted manifold projection:

$$\mathcal{L}_{IMR} = \sum_{j \in \mathcal{K}(\hat{\mathbf{x}})} w_j \cdot |\Psi_{\theta_{sdf}}(\mathbf{f}_j, \mathbf{d}_j)|, \tag{3}$$

where the distance-based weight w_j is dynamically determined by spatial proximity:

$$w_j = \frac{(\|\hat{\mathbf{x}} - \mathbf{p}_j\|^2 + \epsilon)^{-1}}{\sum_{k \in \mathcal{K}(\hat{\mathbf{x}})} (\|\hat{\mathbf{x}} - \mathbf{p}_k\|^2 + \epsilon)^{-1}}, \tag{4}$$

where $\epsilon = 10^{-6}$ ensures numerical stability. This differentiable formulation elegantly evaluates how far the predicted point deviates from the nearest physical zero-level set.

SCR Total Loss. To establish the fundamental spatial reference, we first define the base coordinate regression loss $\mathcal{L}_{coord}$. This loss serves as a macroscopic structural anchor by explicitly penalizing the absolute L1 distance between the predicted global coordinates $\hat{\mathbf{x}}_i$ and their corresponding ground-truth $\mathbf{x}_{gt,i}$:

$$\mathcal{L}_{coord} = \frac{1}{N} \sum_{i=1}^{N} \|\hat{\mathbf{x}}_i - \mathbf{x}_{gt,i}\|_1. \tag{5}$$

Building upon this macroscopic foundation, the final optimization objective for the SCR network is formulated as a dual-constraint loss that seamlessly integrates our continuous geometric prior:

$$\mathcal{L}_{scr} = \mathcal{L}_{coord} + \mathcal{L}_{IMR}. \tag{6}$$

Here, $\mathcal{L}_{coord}$ acts as a macroscopic anchor for global convergence. Concurrently, $\mathcal{L}_{IMR}$ serves as a microscopic attractor, penalizing non-zero SDF to force "floating" predictions to strictly adhere to the continuous physical zero-level set, thereby resolving spatial representation confounding.

3.2 Mamba-Driven Kinematic Distillation (MDKD) for Temporal Consistency.

Since the exact offline Teacher is too computationally heavy for real-time deployment, we distill its continuous kinematics into a lightweight stateful prior to mitigate SCR's unphysical localization jumps. Rather than memorizing explicit maps, we leverage scene-agnostic features to predict plausible relative motion. Given a T-frame window, a frozen backbone [26] extracts relative topological features $\mathbf{F}_{1:T} \in \mathbb{R}^{T \times d}$, which are linearly condensed into a compact latent sequence $\mathbf{Z}_{1:T} \in \mathbb{R}^{T \times m}$.

To efficiently infer continuous motion from sequential features, we introduce a Selective Structured State Space Model Mamba [40, 41]. Bypassing the complexity bottlenecks of Transformers, Mamba propagates hidden states with linear $O(T)$ complexity. To guarantee mathematical rigor, rather than regressing raw transformation matrices, the network maps the sequence directly to the tangent space, predicting the instantaneous inter-frame velocity $\hat{\xi}_t \in \mathbb{R}^6$:

$$[\hat{\xi}_t, \mathbf{h}_t] = \Phi_{mamba}(\mathbf{Z}_{1:T}), \tag{7}$$

where the hidden state $\mathbf{h}_t$ actively encapsulates the ongoing kinematic context and predictive uncertainty for subsequent adaptive fusion. Concurrently, the exact high-fidelity motion from the Teacher is extracted via the logarithmic map:

$$\xi_t^{teacher} = \log\left((\mathbf{T}_{t-1}^{teacher})^{-1} \mathbf{T}_t^{teacher}\right)^{\vee} \in \mathbb{R}^6, \tag{8}$$

where $(\cdot)^{\vee}$ maps the skew-symmetric matrix to a 6D vector representing linear and angular velocities. The distillation process is elegantly supervised using a standard Mean Squared Error (MSE) loss in the Lie algebra space:

$$\mathcal{L}_{Distill} = \|\hat{\xi}_t - \xi_t^{teacher}\|_2^2, \tag{9}$$

By sequentially processing scene-agnostic deep features through a state-space model, the network successfully learns generalized, geometry-dependent physical priors (e.g., decelerating at corners or maintaining inertia along straight corridors). This Mamba-driven distillation effectively provides a stable, robust kinematic constraint for the subsequent real-time gating mechanism.

3.3 Dynamic Manifold Gating (DMG) for Real-Time Pose Fusion

The ultimate objective is to resolve the "last-mile" spatial jittering strictly in real-time. During online inference, the computationally heavy Teacher is entirely bypassed. We obtain a global, drift-free but potentially jittery absolute pose $\mathbf{T}_t^{SCR}$ from the SCR backbone, alongside a locally stable temporal prediction propagated via the kinematic prior:

$$\hat{\mathbf{T}}_t^{prior} = \mathbf{T}_{t-1} \exp(\hat{\xi}_t^{\wedge}), \tag{10}$$

where $(\cdot)^{\wedge}$ is the exact inverse operator of $(\cdot)^{\vee}$, mapping the 6D kinematic vector back to its corresponding 4×4 skew-symmetric matrix in the Lie algebra to enable the matrix exponential.

Traditional linear interpolation cannot be applied to fuse $SE(3)$ matrices as it strictly violates the manifold geometry. Instead, we formulate the fusion as an adaptive geodesic interpolation process on the Lie group. We first compute the geometric pose innovation (i.e., the manifold residual) mapping the temporal prior to the SCR observation:

$$\mathbf{r}_t = \log\left((\hat{\mathbf{T}}_t^{prior})^{-1}\mathbf{T}_t^{SCR}\right)^{\vee} \in \mathbb{R}^6. \tag{11}$$

A sudden, unphysical jump in SCR will manifest as an abnormally large norm $\|\mathbf{r}_t\|$. To act as a dynamic "kinematic shock absorber", we design a differentiable robust gating mechanism. Rather than relying on a static, hand-crafted threshold, we formulate an adaptive manifold fusion weight $\alpha_t \in [0, 1]$ using a scaled Cauchy [42] robust kernel:

$$\alpha_t = \frac{1}{1 + (\|\mathbf{r}_t\|/c_t)^2}. \tag{12}$$

Crucially, the tolerance margin c_t is not a fixed scalar, but a kinematic-aware dynamic parameter predicted by a lightweight gating network:

$$c_t = \text{Softplus}(\text{MLP}_{gate}(\mathbf{h}_t)), \tag{13}$$

where the Mamba hidden state $\mathbf{h}_t$ encodes the ongoing kinematic context. Softplus ensures a positive margin $c_t > 0$, enabling the network to act as an adaptive shock absorber: it tightens ($c_t \downarrow$) to reject SCR anomalies during stable motion, and relaxes ($c_t \uparrow$) to accommodate valid complex maneuvers. The final pose is then updated via $SE(3)$ manifold:

$$\mathbf{T}_{final} = \hat{\mathbf{T}}_t^{prior} \exp(\alpha_t \cdot \mathbf{r}_t^{\wedge}). \tag{14}$$

This end-to-end trainable architecture guarantees jump-free, kinematically coherent localization at the native frequency of the LiDAR, successfully solving the "last-mile" problem.

4 Experiments

4.1 Experimental Settings

Datasets. We evaluate our method on two large-scale outdoor LiDAR localization benchmarks, the Oxford Radar RobotCar dataset [43] and the NCLT dataset [44]:

- **Oxford** (Quality-enhanced Oxford RobotCar) is an urban driving dataset comprising a 10-kilometer route with a coverage area of 200 hm^2 under various weather conditions. The point clouds were collected using a Velodyne HDL-32E LiDAR. The ground truth poses were obtained through INS interpolations and further quality-enhanced by the SGLoc [19]. In all experiments, we use the quality-enhanced ground-truth version of the Oxford dataset. Following previous works [19, 20, 21], we adopt the identical training and testing splits.
- **NCLT** is a campus dataset collected at the University of Michigan's North Campus using a Segway robot platform. It features a 5.5-kilometer route spanning a coverage area of 45 hm^2 and encompasses long-term seasonal variations. The point clouds were captured by a Velodyne HDL-32E LiDAR, and the ground truth poses were refined via SLAM optimization. Following [19, 20, 21], the standard training and testing splits are utilized.

Evaluation Metrics. Following existing outdoor LiDAR localization methods [19, 20, 21, 26], we adopt the Mean position error (m) $\downarrow$ and Mean orientation error ($^\circ$) $\downarrow$ as our primary evaluation metrics. Furthermore, to comprehensively assess the stability of the predicted localization, we align with standard robotics evaluation protocols and introduce the Relative Pose Error (RPE) (m) $\downarrow$ and the Mean Jitter Rate (Jitter) ($\mathrm{m/f^3}$) $\downarrow$ as auxiliary metrics [45]. Specifically, the Jitter metric reflects the degree of high-frequency, small-scale spatial jumps in the trajectory, which is a critical factor determining the safety and reliability of lane-level high-precision localization in autonomous driving.

Implementation. Our proposed BiLi is implemented in PyTorch, utilizing MinkowskiEngine[46] for sparse convolutional operations. All experiments are conducted on a single NVIDIA RTX 4090 GPU. We optimize the network using the AdamW optimizer [47] coupled with a one-cycle learning rate schedule, where the learning rate ranges from 5×10^{-4} to 5×10^{-3}. For the Scene Coordinate

Table 1: Quantitative results on the Oxford dataset. We report the mean position and mean orientation error↓(m, °). Best results in **bold**, second best results underlined.

Mechanism	Method	Venue & Year	15-13-06-37	17-13-26-39	17-14-03-00	18-14-14-42	Mean Error (m, °)↓
APR	PointLoc [16]	2022'Sens.J	10.75, 2.36	11.07, 2.21	11.53, 1.92	9.82, 2.07	10.79, 2.14
	PosePN++ [48]	2022'PR	4.54, 1.83	6.44, 1.78	4.89, 1.55	4.64, 1.61	5.13, 1.69
	STCLoc [31]	2022'TITS	5.14, 1.27	6.12, 1.21	5.32, 1.08	4.76, 1.19	5.34, 1.19
	NidaLoc [33]	2023'TITS	3.71, 1.50	5.40, 1.40	3.94, 1.30	4.08, 1.30	4.28, 1.38
	HypLiLoc [34]	2023'CVPR	5.03, 1.46	4.31, 1.43	3.61, 1.11	2.61, 1.09	3.89, 1.27
SCR	SGLoc [19]	2023'CVPR	1.79, 1.67	1.81, 1.76	1.33, 1.59	1.19, 1.39	1.53, 1.60
	LiSA [20]	2024'CVPR	0.94, 1.10	1.17, 1.21	0.84, 1.15	0.85, 1.11	0.95, 1.14
	RALoc [21]	2025'ICCV	1.52, 1.28	1.71, 1.27	1.37, 1.20	1.41, 1.23	1.51, 1.24
	LightLoc [26]	2025'CVPR	0.82, 1.12	0.85, 1.07	0.81, 1.11	0.82, 1.16	0.83, 1.12
	GTRLoc [23]	2025'NeurIPS	0.77, 1.02	0.77, 1.01	0.67, 1.01	0.80, 1.07	0.75, 1.03
	BiLi	Ours	**0.63**, **0.91**	**0.55**, **0.88**	**0.51**, **0.90**	**0.57**, **0.96**	**0.56**, **0.91**

Table 2: Quantitative results on the NCLT dataset. We report the mean position and mean orientation error↓(m, °). Best results in **bold**, second best results underlined.

Mechanism	Method	Venue & Year	2012-02-12	2012-02-19	2012-03-31	2012-05-26	Mean Error (m,°)↓
APR	PointLoc [16]	2022'Sens.J	7.23, 4.88	6.31, 3.89	6.71, 4.32	9.55, 5.21	7.45, 4.58
	PosePN++ [48]	2022'PR	4.97, 3.75	3.68, 2.65	4.35, 3.38	8.42, 4.30	5.36, 3.52
	STCLoc [31]	2022'TITS	4.91, 4.34	3.25, 3.10	3.75, 4.04	7.53, 4.95	4.86, 4.11
	NidaLoc [33]	2023'TITS	4.48, 3.59	3.14, 2.52	3.67, 3.46	6.32, 4.67	4.40, 3.56
	HypLiLoc [34]	2023'CVPR	1.71, 3.56	1.68, 2.69	1.52, 2.90	2.29, 3.34	1.80, 3.12
SCR	SGLoc [19]	2023'CVPR	1.20, 3.08	1.20, 3.05	1.12, 3.28	3.48, 4.43	1.75, 3.46
	LiSA [20]	2024'CVPR	0.97, **2.23**	0.91, **2.09**	0.87, **2.21**	3.11, 2.72	1.47, **2.31**
	RALoc [21]	2025'ICCV	1.61, 4.71	1.61, 4.88	1.58, 4.71	3.51, 5.40	2.07, 4.92
	LightLoc [26]	2025'CVPR	0.98, 2.76	0.89, 2.51	0.86, 2.67	3.10, 3.26	1.46, 2.80
	GTRLoc [23]	2025'NeurIPS	0.95, 2.53	0.82, 2.45	0.82, 2.52	3.01, 2.99	1.40, 2.62
	BiLi	Ours	**0.53**, 2.51	**0.46**, 2.26	**0.46**, 2.32	**0.50**, **2.41**	**0.49**, 2.38

Regression (SCR) module, the input point clouds are voxelized with grid sizes of 0.25 m and 0.3 m for the Oxford and NCLT datasets, respectively. Conversely, for the implicit neural point pose estimation, the voxel size is universally set to 0.4 m, with the number of optimization iterations fixed at 10. In Implicit Manifold Regularization(IMR), the number of queried K-nearest neural points is set to 6. The entire training process is conducted with a batch size of 512 for 180 epochs. The Mamba network adopts a T-frame of 5. The architectures of Mamba (MDKD) and DMG gating are detailed in Appendix Sec B.

4.2 Comparison with State-of-the-art Methods

Baselines. We extensively compare our BiLi against various state-of-the-art learning-based localization approaches, which are broadly categorized into two mechanisms: (1) Absolute Pose Regression (APR) methods, including PointLoc [16], PosePN++ [48], STCLoc [31], NIDALoc [33], and HypLiLoc [34]; and (2) Scene Coordinate Regression (SCR) methods, comprising SGLoc [19], LiSA [20], LightLoc [26], RALoc [21], and GTRLoc [23]. Notably, LightLoc, RALoc, and GTRLoc represent the most recent and highly competitive state-of-the-art methods demonstrating superior localization performance in current literature.

Results on Oxford dataset. As shown in Tab. 1, BiLi establishes a new benchmark with a mean error of 0.56 m and 0.91°, outperforming the strongest SCR baseline (GTRLoc) by over 25% in translation. More crucially, BiLi fundamentally resolves the "last-mile" challenge by delivering unprecedented localization stability (Tab. 3). It achieves a 90.9% Recall@1<1m, while slashing the Mean Relative Pose Error (RPE) to 0.16 m and trajectory jitter to 0.22 m/f³, massive reductions of 62% and 67% compared to GTRLoc. These metrics quantitatively validate the effectiveness of our spatiotemporal synergistic architecture in delivering highly accurate and exceptionally stable localization.

Figure 3 visualizes trajectories on sequence 17-14-03-00. Despite acceptable average errors, recent baselines exhibit dangerous, high-frequency spatial outliers (blue segments, error > 3 m) due to point-wise independent predictions. Conversely, BiLi generates an exceptionally stable, jump-free trajectory tightly aligned with the ground truth. This stark contrast visually confirms that our spatiotemporal synergistic architecture successfully eliminates unphysical spatial jumps, ensuring the strict structural stability required for safe autonomous navigation.

Table 3: Localization trajectory evaluation on the Oxford and NCLT dataset. We report localization stability metrics, including Recall@1<1m(%)↑, mean RPE (m)↓, and Mean Jitter (m/f³)↓. Best results in **bold**, second best results underlined.

Dataset	Quality-Enhanced Oxford				NCLT			
Venue & Year	25'ICCV	25'CVPR	25'NeurIPS	Ours	25'ICCV	25'CVPR	25'NeurIPS	Ours
Method	RALoc	LightLoc	GTRLoc	BiLi	RALoc	LightLoc	GTRLoc	BiLi
Recall@1<1m (%)↑	26.9	80.9	83.6	**90.9**	37.4	77.0	84.3	**93.1**
Mean Position Error (m)↓	1.51	0.83	0.75	**0.56**	2.07	1.46	1.40	**0.49**
Mean RPE (m)↓	0.56	0.45	0.42	**0.16**	2.44	1.21	0.80	**0.26**
Mean Jitter (m/f³)↓	0.86	0.68	0.66	**0.22**	8.17	3.93	2.56	**0.76**

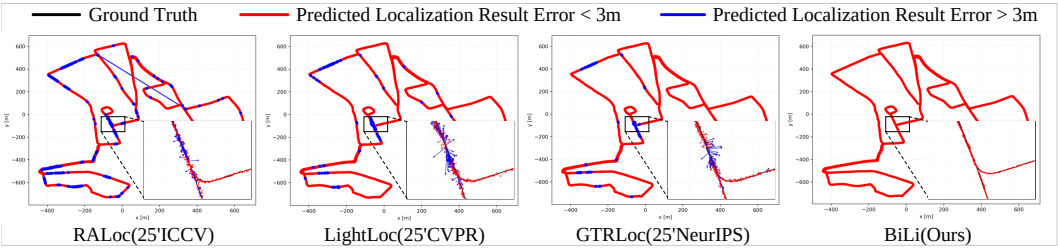

Figure 3: Visualization of localization results on Oxford.

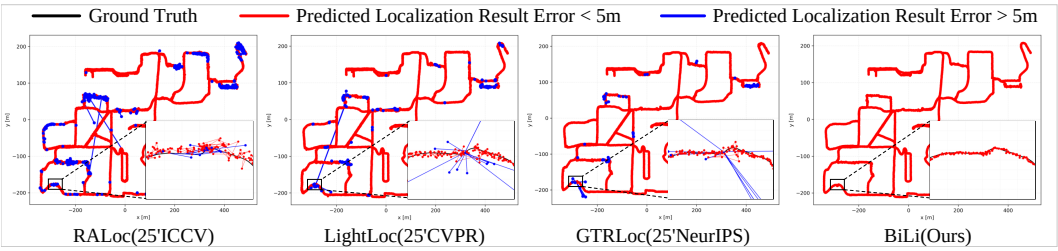

Figure 4: Visualization of localization results on NCLT.

Results on NCLT Dataset. Table 2 demonstrates BiLi's exceptional robustness against the severe long-term seasonal variations of the NCLT benchmark. It comprehensively outperforms all APR and SCR baselines, achieving a remarkable 65% reduction in mean position error compared to the SOTA GTRLoc (from 1.40 m to 0.49 m), making it the first method to break the sub-0.5-meter barrier on this dataset. Furthermore, Table 3 highlights BiLi's extraordinary localization stability under these seasonal perturbations. It achieves a 93.1% Recall@1<1m, while drastically slashing the Mean Relative Pose Error (RPE) to 0.26 m and Mean Trajectory Jitter to 0.76 m/f³, representing massive drops of 67% and 70% against GTRLoc, respectively.

Figure 4 visualizes the predicted trajectories on sequence 2012-03-31. While recent SOTA baselines frequently exhibit massive localization errors (blue segments, error > 5 m), BiLi maintains a highly continuous, jump-free trajectory tightly aligned with the ground truth. Both quantitative and qualitative results consistently prove that our spatiotemporal synergistic architecture successfully resolves the "last-mile" challenge for safe autonomous navigation in large-scale outdoor LiDAR localization.

Real-Time Deployment Feasibility. Table 4 evaluates the inference efficiency and model size. Given the native LiDAR frequencies, the strict latency upper bounds are 50 ms for Oxford and 100 ms for NCLT. BiLi comfortably satisfies these constraints, achieving ultra-fast inference speeds of 30 ms and 48 ms, respectively, with a compact footprint of only 76 MB. Notably, despite introducing the robust pose fusion mechanism to eliminate trajectory jitter, BiLi adds negligible computational overhead, matching the speed of the most lightweight baseline (LightLoc).

Table 4: Real-time and Model-size on Oxford and NCLT.

Method	Real-time Inference Speed		Model↓
	Oxford(<50)↓	NCLT(<100)↓	
SGLoc [19]	38ms	75ms	116MB
LiSA [20]	38ms	75ms	116MB
RALoc [21]	26ms	55ms	123MB
LightLoc [26]	29ms	48ms	74MB
GTRLoc [23]	35ms	56ms	80MB
BiLi (Ours)	30ms	48ms	76MB

Table 6: Ablation on Key Components of Our Proposed BiLi.

IMR	MDKD	DMG	Mean Error(m,°)↓	
			Oxford	NCLT
			0.78, 1.11	1.53, 2.84
✓			0.64, 0.98	0.73, 2.45
	✓	✓	0.70, 1.04	1.17, 2.77
✓	✓	✓	**0.56**, **0.91**	**0.49**, **2.38**

Table 7: Ablation on Mamba-Driven Kinematic Distillation (MDKD).

Method	BiLi(w/o MDKD)	BiLi
Recall@1<1m(%) ↑	**91.4**	90.9
Mean Position(m)↓	0.57	**0.56**
Mean RPE(m) ↓	**0.15**	0.16
Mean Jitter(m/f³) ↓	**0.19**	0.22
Real-Time(< 50ms) ↓	124 ✗	**30** ✓

This confirms that BiLi perfectly balances sub-meter accuracy, stable localization performance, and real-time execution, practically resolving the "last-mile" challenge for autonomous driving.

4.3 Ablation studies

Ablation on BiLi Key Components. Table 6 presents the ablation study of our proposed BiLi framework on the Oxford and NCLT datasets to validate the contribution of each core component: Implicit Manifold Regularization (IMR), Mamba-Driven Kinematic Distillation (MDKD), and Dynamic Manifold Gating (DMG). We provide additional experiments in Appendix sec. C.

- **Effectiveness of IMR.** Integrating IMR alone substantially reduces the mean position error from 0.78 m to 0.64 m on Oxford and from 1.53 m to 0.73 m on NCLT. This confirms that point-to-manifold projection effectively anchors predictions to continuous zero-level sets, mitigating spatial representation confounding even under severe scene variations.
- **Effectiveness of MDKD and DMG.** Employing temporal modules (MDKD and DMG) alone also outperforms the baseline, dropping errors to 0.70 m on Oxford and 1.17 m on NCLT. This verifies that distilling kinematic priors into a lightweight Mamba model and adaptively fusing them enforces continuous temporal kinematic consistency.

Ablation on Mamba-Driven Kinematic Distillation (MDKD). Table 7 assesses the necessity of distillation on Oxford by comparing BiLi with a variant (w/o MDKD) that runs the teacher's iterative SDF optimization online. Both achieve comparable accuracy (0.56 m vs 0.57 m) and stability (RPE: 0.16 m vs 0.15 m, Jitter: 0.22 m/f^3 vs 0.19 m/f^3), but critically diverge in efficiency. To meet the native 20 Hz LiDAR rate (< 50 ms), direct online implicit tracking is computationally prohibitive (124 ms/frame). By distilling this heavy optimization into a feed-forward Mamba network, MDKD slashes latency to 30 ms without sacrificing accuracy, successfully breaking the computational bottleneck for strict real-time deployment.

Ablation on Dynamic Manifold Gating (DMG). Table 5 evaluates the impact of the tolerance margin c_t during pose fusion on the Oxford dataset. Relying on fixed empirical constants exposes a fragile trade-off: a permissive margin ($c = 2$) blindly trusts SCR anomalies, causing catastrophic localization failure (mean error: 7.91 m), while an overly strict margin ($c \leq 0.3$) minimizes jitter (0.20 m/f³) but induces severe dead-reckoning drift (0.64 m). By dynamically predicting c_t via the kinematic hidden states, DMG successfully escapes this dilemma. Acting as an intelligent "kinematic shock absorber", it optimally balances drift correction with jump suppression, achieving the best overall accuracy (0.56 m) and stable localization. (RPE: 0.16 m, Jitter: 0.22 m/f³).

Table 5: Ablation on Dynamic Manifold Gating (DMG).

Evaluation Metrics	Different Parameters of c					BiLi
	0.1	0.3	0.5	1	2	
R@1<1m (%)↑	87.1	88.3	88.2	90.1	6.9	**90.9**
Mean Position (m) ↓	0.65	0.64	0.65	0.68	7.91	**0.56**
Mean RPE(m)↓	0.25	0.29	0.36	0.36	6.65	**0.16**
Mean Jitter(m/f3)↓	0.26	**0.20**	0.36	1.00	23.23	0.22

5 Conclusion

This paper presents BiLi to resolve the "last-mile" challenge in outdoor LiDAR localization. By integrating Implicit Manifold Regularization (IMR) to eliminate spatial representation confounding, along with Mamba-distilled motion priors and Dynamic Manifold Gating (DMG), BiLi fundamentally eradicates unphysical location jumps. Outperforming state-of-the-art methods on standard benchmarks, BiLi sets a new standard for highly accurate, real-time, and stable autonomous deployment.

6 Acknowledgements

This work was partially supported by the Fundamental Research Funds for the Central Universities (20720260165) and the National Natural Science Foundation of China (No. 62501502). The authors declare no competing interests.

References

[1] H. Yin, X. Xu, S. Lu, X. Chen, R. Xiong, S. Shen, C. Stachniss, and Y. Wang, "A survey on global lidar localization: Challenges, advances and open problems," *International Journal of Computer Vision*, vol. 132, no. 8, pp. 3139–3171, 2024.

[2] R. W. Wolcott and R. M. Eustice, "Fast lidar localization using multiresolution gaussian mixture maps," in *2015 IEEE international conference on robotics and automation (ICRA)*. IEEE, 2015, pp. 2814–2821.

[3] J. Nubert, E. Walther, S. Khattak, and M. Hutter, "Learning-based localizability estimation for robust lidar localization," in *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2022, pp. 17–24.

[4] Z. Shi, X. Zhang, W. Xu, Y. Xia, Y. Zang, S. Shen, and C. Wang, "L2rsi: Cross-view lidar-based place recognition for large-scale urban scenes via remote sensing imagery," *Advances in Neural Information Processing Systems*, vol. 38, pp. 160 681–160 704, 2026.

[5] D. Liu, S. Huang, W. Li, S. Shen, and C. Wang, "Text to point cloud localization with multi-level negative contrastive learning," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 39, no. 5, 2025, pp. 5397–5405.

[6] N. Slonim, Y. Bilu, C. Alzate, R. Bar-Haim, B. Bogin, F. Bonin, L. Choshen, E. Cohen-Karlik, L. Dankin, L. Edelstein *et al.*, "An autonomous debating system," *Nature*, vol. 591, no. 7850, pp. 379–384, 2021.

[7] F. Thomas and L. Ros, "Revisiting trilateration for robot localization," *IEEE Transactions on robotics*, vol. 21, no. 1, pp. 93–101, 2005.

[8] Z. Liu, S. Zhou, C. Suo, P. Yin, W. Chen, H. Wang, H. Li, and Y.-H. Liu, "Lpd-net: 3d point cloud learning for large-scale place recognition and environment analysis," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2019, pp. 2831–2840.

[9] W. Zhang and C. Xiao, "Pcan: 3d attention map learning using contextual information for point cloud based retrieval," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 12 436–12 445.

[10] L. Luo, S.-Y. Cao, X. Li, J. Xu, R. Ai, Z. Yu, and X. Chen, "Bevplace++: Fast, robust, and lightweight lidar global localization for unmanned ground vehicles," *IEEE Transactions on Robotics*, 2025.

[11] Z. Fan, Z. Song, H. Liu, Z. Lu, J. He, and X. Du, "Svt-net: Super light-weight sparse voxel transformer for large scale place recognition," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 36, no. 1, 2022, pp. 551–560.

[12] S. Ao, Q. Hu, B. Yang, A. Markham, and Y. Guo, "Spinnet: Learning a general surface descriptor for 3d point cloud registration," in *CVPR*, 2021, pp. 11 753–11 762.

[13] S. Ao, Q. Hu, H. Wang, K. Xu, and Y. Guo, "Buffer: Balancing accuracy, efficiency, and generalizability in point cloud registration," in *CVPR*, 2023, pp. 1255–1264.

[14] S. Ao, Y. Guo, Q. Hu, B. Yang, A. Markham, and Z. Chen, "You only train once: Learning general and distinctive 3d local descriptors," *TPAMI*, vol. 45, no. 3, pp. 3949–3967, 2022.

[15] J. Wu, M. Zhu, D. Liu, W. Li, S. Ao, S. Shen, C. Wen, and C. Wang, "Leader: Learning reliable local-to-global correspondences for lidar relocalization," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, June 2026, pp. 9932–9942.

[16] W. Wang, B. Wang, P. Zhao, C. Chen, R. Clark, B. Yang, A. Markham, and N. Trigoni, "Pointloc: Deep pose regressor for lidar point cloud localization," *IEEE Sensors Journal*, vol. 22, no. 1, pp. 959–968, 2021.

[17] W. Li, Y. Yang, S. Yu, G. Hu, C. Wen, M. Cheng, and C. Wang, "Diffloc: Diffusion model for outdoor lidar localization," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 15 045–15 054.

[18] O. Chum and J. Matas, "Optimal randomized ransac," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 8, pp. 1472–1482, 2008.

[19] W. Li, S. Yu, C. Wang, G. Hu, S. Shen, and C. Wen, "Sgloc: Scene geometry encoding for outdoor lidar localization," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 9286–9295.

[20] B. Yang, Z. Li, W. Li, Z. Cai, C. Wen, Y. Zang, M. Muller, and C. Wang, "Lisa: Lidar localization with semantic awareness," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 15 271–15 280.

[21] Y. Yang, W. Li, S. Ao, Q. Xu, S. Yu, Y. Guo, Y. Zhou, S. Shen, and C. Wang, "Raloc: Enhancing outdoor lidar localization via rotation awareness," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2025, pp. 3304–3313.

[22] Y. Chen, Q. Li, Y. Yang, W. Li, S. Ao, and C. Wang, "Unleashing the power of data generation in one-pass outdoor lidar localization," in *Proceedings of the 33rd ACM International Conference on Multimedia*, 2025, pp. 199–208.

[23] S. Yu, W. Li, X. Sun, Z. Yuan, X. Wang, S. Wang, R. She, and C. Wang, "Gtr-loc: Geospatial text regularization assisted outdoor lidar localization," *Advances in Neural Information Processing Systems*, vol. 38, pp. 136 468–136 490, 2026.

[24] G. D. Arana, O. A. Hafez, M. Joerger, and M. Spenko, "Localization safety validation for autonomous robots," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2020, pp. 6276–6281.

[25] T. G. Reid, S. E. Houts, R. Cammarata, G. Mills, S. Agarwal, A. Vora, and G. Pandey, "Localization requirements for autonomous vehicles," *SAE International Journal of Connected and Automated Vehicles*, vol. 2, no. 3, pp. 173–190, 2019.

[26] W. Li, C. Liu, S. Yu, D. Liu, Y. Zhou, S. Shen, C. Wen, and C. Wang, "Lightloc: Learning outdoor lidar localization at light speed," in *Proceedings of the Computer Vision and Pattern Recognition Conference*, 2025, pp. 6680–6689.

[27] M. Bosse and R. Zlot, "Place recognition using keypoint voting in large 3d lidar datasets," in *2013 IEEE international conference on robotics and automation*. IEEE, 2013, pp. 2677–2684.

[28] Y. Huang, C. Liu, M. Zhu, S. Ao, C. Wen, and C. Wang, "Difflo: Semantic-aware lidar odometry with diffusion-based refinement," in *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2025, pp. 17 050–17 059.

[29] K. Vidanapathirana, M. Ramezani, P. Moghadam, S. Sridharan, and C. Fookes, "Logg3d-net: Locally guided global descriptor learning for 3d place recognition," in *2022 International Conference on Robotics and Automation (ICRA)*. IEEE, 2022, pp. 2215–2221.

[30] Y. Cui, X. Chen, Y. Zhang, J. Dong, Q. Wu, and F. Zhu, "Bow3d: Bag of words for real-time loop closing in 3d lidar slam," *IEEE Robotics and Automation Letters*, vol. 8, no. 5, pp. 2828–2835, 2022.

[31] S. Yu, C. Wang, Y. Lin, C. Wen, M. Cheng, and G. Hu, "Stcloc: Deep lidar localization with spatio-temporal constraints," *IEEE Transactions on Intelligent Transportation Systems*, vol. 24, no. 1, pp. 489–500, 2022.

[32] C. R. Qi, L. Yi, H. Su, and L. J. Guibas, "Pointnet++: Deep hierarchical feature learning on point sets in a metric space," *Advances in neural information processing systems*, vol. 30, 2017.

[33] S. Yu, X. Sun, W. Li, C. Wen, Y. Yang, B. Si, G. Hu, and C. Wang, "Nidaloc: Neurobiologically inspired deep lidar localization," *IEEE Transactions on Intelligent Transportation Systems*, vol. 25, no. 5, pp. 4278–4289, 2023.

[34] S. Wang, Q. Kang, R. She, W. Wang, K. Zhao, Y. Song, and W. P. Tay, "Hypliloc: Towards effective lidar pose regression with hyperbolic fusion," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 5176–5185.

[35] X. Liu, Z. Hu, H. Ling, and Y.-M. Cheung, "Mtfh: A matrix tri-factorization hashing framework for efficient cross-modal retrieval," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 3, pp. 964–981, 2021.

[36] X. Liu, X. Wang, and Y.-M. Cheung, "Fddh: Fast discriminative discrete hashing for large-scale cross-modal retrieval," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 33, no. 11, pp. 6306–6320, 2022.

[37] H. Xi, Z. Shao, D. A. Hensher, J. D. Nelson, H. Chen, and K. Wijayaratna, "A multi-task transformer with mixture-of-experts for personalized periodic predictions of individual travel behavior in multimodal public transport," *Transportation Research Part C: Emerging Technologies*, vol. 179, p. 105287, 2025.

[38] J. J. Park, P. Florence, J. Straub, R. Newcombe, and S. Lovegrove, "Deepsdf: Learning continuous signed distance functions for shape representation," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 165–174.

[39] Y. Pan, X. Zhong, L. Wiesmann, T. Posewsky, J. Behley, and C. Stachniss, "Pin-slam: Lidar slam using a point-based implicit neural representation for achieving global map consistency," *IEEE Transactions on Robotics*, vol. 40, pp. 4045–4064, 2024.

[40] A. Gu and T. Dao, "Mamba: Linear-time sequence modeling with selective state spaces," *arXiv preprint arXiv:2312.00752*, 2023.

[41] Z. Shao, H. Xi, D. A. Hensher, Z. Wang, X. Gong, and J. Gao, "A spatial–temporal dynamic attention-based mamba model for multi-type passenger demand prediction in multimodal public transit systems," *Transportation Research Part E: Logistics and Transportation Review*, vol. 202, p. 104282, 2025.

[42] M. J. Black and P. Anandan, "The robust estimation of multiple motions: Parametric and piecewise-smooth flow fields," *Computer vision and image understanding*, vol. 63, no. 1, pp. 75–104, 1996.

[43] D. Barnes, M. Gadd, P. Murcutt, P. Newman, and I. Posner, "The oxford radar robotcar dataset: A radar extension to the oxford robotcar dataset," in *2020 IEEE international conference on robotics and automation (ICRA)*. IEEE, 2020, pp. 6433–6438.

[44] N. Carlevaris-Bianco, A. K. Ushani, and R. M. Eustice, "University of michigan north campus long-term vision and lidar dataset," *The International Journal of Robotics Research*, vol. 35, no. 9, pp. 1023–1035, 2016.

[45] A. Zeng, L. Yang, X. Ju, J. Li, J. Wang, and Q. Xu, "Smoothnet: A plug-and-play network for refining human poses in videos," in *European Conference on Computer Vision*. Springer, 2022, pp. 625–642.

[46] C. Choy, J. Gwak, and S. Savarese, "4d spatio-temporal convnets: Minkowski convolutional neural networks," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 3075–3084.

[47] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," *arXiv preprint arXiv:1711.05101*, 2017.

[48] S. Yu, C. Wang, C. Wen, M. Cheng, M. Liu, Z. Zhang, and X. Li, "Lidar-based localization using universal encoding and memory-aware regression," *Pattern Recognition*, vol. 128, p. 108685, 2022.

A Optimization Details of Offline Neural Implicit Mapping

In Section 3.1 of the main text, we introduced the offline neural implicit Teacher, which optimizes the spatial parameters of the implicit map $\mathcal{M}$ and estimates the highly accurate relative pose $\mathbf{T}^* \in SE(3)$. Here, we provide the detailed mathematical formulation of the residuals and the Lie algebraic optimization process.

A.1 Detailed Formulation of Geometric-Semantic Residuals

Given an incoming LiDAR scan, we aim to register each point p_i to the global implicit map $\mathcal{M} = \{(\mathbf{p}_j, \mathbf{q}_j, \mathbf{f}_j, \mathbf{c}_j)\}_{j=1}^{M}$. Let $\mathbf{p}'_i = \mathbf{T} \circ p_i$ be the globally transformed coordinate of p_i under the current pose estimate $\mathbf{T}$. To compute the continuous fields $\mathcal{S}(\mathbf{p}'_i)$ and $\mathcal{C}(\mathbf{p}'_i)$, we query the K-nearest neural points and transform $\mathbf{p}'_i$ into their local frames:

$$\Delta\mathbf{p}_{ij} = \mathbf{R}(\mathbf{q}_j)^{-1}(\mathbf{p}'_i - \mathbf{p}_j). \tag{15}$$

Using a shared geometry MLP (Ψ_{sdf}) and a semantic MLP (Ψ_{feat}), we predict local distances and features, which are then fused via Inverse Distance Weighting (IDW) to form the continuous global query values:

$$\mathcal{S}(\mathbf{p}'_i) = \sum_{j=1}^{K} w_{ij}\Psi_{sdf}(\mathbf{f}_j, \Delta\mathbf{p}_{ij}), \quad \mathcal{C}(\mathbf{p}'_i) = \sum_{j=1}^{K} w_{ij}\Psi_{feat}(\mathbf{c}_j, \Delta\mathbf{p}_{ij}), \tag{16}$$

where $w_{ij} \propto \exp(-\|\mathbf{p}'_i - \mathbf{p}_j\|^2/\sigma^2)$ is the spatial weight with bandwidth σ.

The optimization objective minimizes the joint geometric-semantic residual. The Geometric SDF Loss forces the transformed points to lie exactly on the zero-level set (physical surface):

$$\mathcal{L}_{sdf}(\mathbf{p}'_i) = H\big(\mathcal{S}(\mathbf{p}'_i)\big), \tag{17}$$

where $H(\cdot)$ is the robust Huber loss to suppress dynamic outliers. The Semantic Feature Loss ensures tracking stability in degenerate geometries (e.g., featureless corridors) by anchoring points to semantic consistency:

$$\mathcal{L}_{feat}(\mathbf{p}'_i) = \lambda_{feat}\left\|\mathcal{C}(\mathbf{p}'_i) - \mathbf{c}_i^{in}\right\|_2^2, \tag{18}$$

where $\mathbf{c}_i^{in}$ is the original semantic descriptor extracted from the incoming scan by the encoder, and λ_{feat} is a balancing factor.

A.2 Pose Optimization on the Lie Group $\mathfrak{se}(3)$

Finding the optimal $\mathbf{T}^*$ that minimizes $\sum_i(\mathcal{L}_{sdf} + \mathcal{L}_{feat})$ is a non-linear least-squares problem. To strictly respect the $SE(3)$ manifold geometry and avoid singularity, we optimize the pose by applying iterative left-perturbations on the Lie algebra $\mathfrak{se}(3)$. Let the perturbation be $\xi \in \mathbb{R}^6$, the pose is updated as:

$$\mathbf{T}_{k+1} = \exp(\xi^\wedge)\mathbf{T}_k. \tag{19}$$

To employ the Levenberg-Marquardt (LM) algorithm, we derive the Jacobians using the chain rule. For the geometric residual $r_{sdf} = \mathcal{S}(\mathbf{p}'_i)$, the Jacobian with respect to the perturbation ξ is:

$$\mathbf{J}_{sdf} = \frac{\partial r_{sdf}}{\partial \xi} = \frac{\partial \mathcal{S}}{\partial \mathbf{p}'_i}\frac{\partial \mathbf{p}'_i}{\partial \xi}, \tag{20}$$

where $\frac{\partial \mathcal{S}}{\partial \mathbf{p}'_i} \in \mathbb{R}^{1\times 3}$ is the spatial gradient analytically provided by the backpropagation of Ψ_{sdf} and the IDW process. The second term is the standard point perturbation Jacobian on $SE(3)$:

$$\frac{\partial \mathbf{p}'_i}{\partial \xi} = \begin{bmatrix}\mathbf{I}_{3\times 3} & -(\mathbf{p}'_i)^\times\end{bmatrix} \in \mathbb{R}^{3\times 6}, \tag{21}$$

with $(\cdot)^\times$ denoting the skew-symmetric matrix. The Jacobian for the semantic residual $\mathbf{J}_{feat}$ is derived symmetrically via Ψ_{feat}.

At each iteration, the optimal increment $\Delta\xi^*$ is solved via the normal equation:

$$(\mathbf{J}^T\mathbf{W}\mathbf{J} + \mu\mathbf{I})\Delta\xi^* = -\mathbf{J}^T\mathbf{W}\mathbf{r}, \tag{22}$$

where $\mathbf{J}$ and $\mathbf{r}$ are the stacked Jacobians and residuals for all points, $\mathbf{W}$ is the diagonal weight matrix inherently provided by the derivative of the Huber loss, and μ is the LM damping parameter. Once converged, this offline procedure accurately disentangles spatial noise, providing the precise spatiotemporal ground truth ($\mathbf{T}^*$ and aligned manifolds) required for the real-time Implicit Manifold Regularization (Sec. 3.1) and Mamba-Driven Kinematic Distillation (Sec. 3.2).

B Network Architectures for MDKD and DMG

In this section, we detail the specific network architectures for the Mamba-Driven Kinematic Distillation (MDKD) module and the Dynamic Manifold Gating (DMG) mechanism introduced in Sections 3.2 and 3.3 of the main text.

B.1 Architecture of the Mamba (MDKD) Network

The objective of the MDKD network is to efficiently process the sequential topological features and output both the instantaneous 6D inter-frame velocity $\hat{\xi}_t$ and the kinematic context state $\mathbf{h}_t$. As illustrated in Figure 5, the network is designed with a lightweight, linear-complexity state-space architecture.

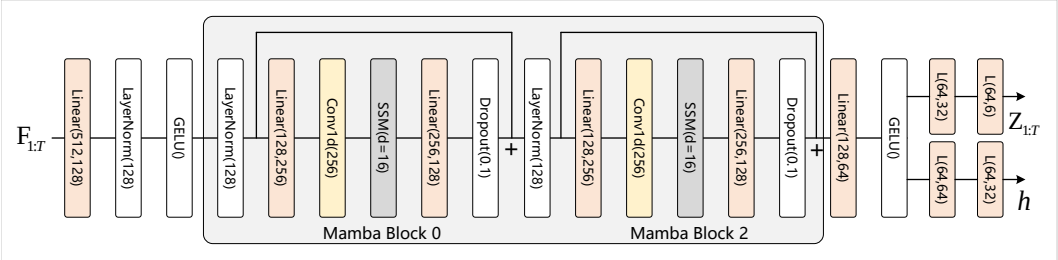

Figure 5: Visualization of Mamba network in MDKD.

1. Input Embedding: The input to the network is a sequence of frame-to-frame topological features $\mathbf{F}_{1:T} \in \mathbb{R}^{T\times 512}$ extracted by the frozen backbone. To reduce computational overhead, this input is first projected into a compact latent dimension via an embedding block consisting of a linear layer (`Linear(512, 128)`), followed by `LayerNorm` and a `GELU` activation function.

2. Core Mamba Blocks: The core sequence modeling relies on two identical, stacked Mamba Blocks (denoted as Block 0 and Block 1). Each block employs a residual connection. Within the block, the 128-dimensional latent vector is first normalized (`LayerNorm`) and then expanded to a higher-dimensional space via `Linear(128, 256)`. To capture local temporal dependencies, a 1-dimensional convolution (`Conv1d`) is applied, followed by the Structured State Space Model (`SSM`) with a state dimension of $d = 16$. The features are then projected back to the original dimension via `Linear(256, 128)`, followed by a `Dropout` layer with a rate of 0.1 to prevent overfitting before being added back to the residual stream.

3. Bifurcated Output Heads: After traversing the Mamba blocks, the sequence passes through a shared projection layer (`Linear(128, 64)` and `GELU`) and bifurcates into two specific task heads:

- **Kinematic Head:** This branch predicts the continuous motion prior. It consists of two linear layers (`Linear(64, 32)` $\rightarrow$ `Linear(32, 6)`), ultimately outputting the 6D Lie algebra vector $\hat{\xi}_t \in \mathbb{R}^6$ that dictates the relative $SE(3)$ transformation.
- **Context Head:** This branch extracts the hidden kinematic context and uncertainty. It employs two linear layers (`Linear(64, 64)` $\rightarrow$ `Linear(64, 32)`), yielding the compact hidden state $\mathbf{h}_t \in \mathbb{R}^{32}$ that serves as the condition for the subsequent DMG module.

B.2 Architecture of the Dynamic Manifold Gating (DMG)

To perform the real-time adaptive pose fusion detailed in Section 3.3, the DMG mechanism relies on a highly efficient Multilayer Perceptron (MLP_{gate}) to dynamically predict the tolerance margin c_t. As

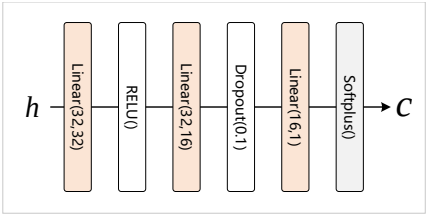

Figure 6: Visualization of DMG gating network.

illustrated in Figure 6, the network is designed to be lightweight to meet strict real-time requirements while maintaining sufficient capacity for non-linear mapping.

The input to MLP_{gate} is the 32-dimensional hidden state $\mathbf{h}_t \in \mathbb{R}^{32}$, which is extracted by the Context Head of the MDKD module and encapsulates the instantaneous kinematic context and predictive uncertainty. The network sequentially processes this state through the following layers:

- **Feature Transformation:** The input state first passes through a `Linear(32, 32)` layer followed by a `ReLU` activation. This preserves the dimensional capacity while extracting higher-level non-linear contextual features.
- **Dimensionality Reduction & Regularization:** The features are then compressed via a `Linear(32, 16)` layer. To prevent the gating mechanism from overfitting to specific trajectory patterns and becoming over-confident, a `Dropout` layer with a probability of 0.1 is applied.
- **Margin Prediction:** A final projection layer (`Linear(16, 1)`) maps the 16-dimensional vector to a single scalar value.

Mathematically, the forward pass is formulated as:

$$c_t = \text{Softplus}\Big(\texttt{Linear}_{16\to 1}\big(\text{Dropout}\left(\texttt{Linear}_{32\to 16}\left(\text{ReLU}(\texttt{Linear}_{32\to 32}(\mathbf{h}_t))\right)\right)\big)\Big). \tag{23}$$

The final `Softplus` function explicitly enforces a strictly positive output ($c_t > 0$). By mapping the sequence-derived hidden state to a dynamic margin, this compact network successfully acts as an intelligent "shock absorber", gracefully rejecting abnormal SCR jumps while accommodating valid complex maneuverings on the $SE(3)$ manifold.

C Superiority of Deep Spatiotemporal Synergy over Post-hoc Pose Fusion

In the main text, we argue that mitigating the "last-mile" spatial jittering requires an end-to-end spatiotemporal synergistic architecture rather than naive post-processing. To empirically validate this claim, we conduct a comprehensive comparison on the NCLT dataset between our proposed Dynamic Manifold Gating (DMG) mechanism and several conventional pose-level SCR-Odometry fusion strategies.

C.1 SCR-Odometry Fusion Baselines

To examine whether the last-mile trajectory instability can be solved by conventional post-processing, we construct three pose-level fusion baselines on top of the output pose sequences of RALoc, LightLoc, and GTRLoc. These baselines introduce odometry relative-motion constraints only after SCR pose estimation has been completed. Therefore, unlike BiLi, they do not access point-cloud features, scene-coordinate predictions, implicit manifold fields, or Mamba hidden states. This setting provides a strong diagnostic comparison: if pose-level smoothing alone were sufficient, these enhanced SCR baselines should largely eliminate localization jitter.

In data processing, *Origin* denotes the original output of each SCR method evaluated on frames synchronized with the corresponding odometry trajectory. For B1/B2/B3, we first re-anchor the odometry trajectory offline to reduce the influence of long-term drift, so that the evaluation focuses on local relative-motion stabilization rather than global odometry drift. This re-anchoring is applied only

Table 8: Evaluation of pose-level SCR-Odometry fusion on the NCLT dataset. Metrics are averaged over four NCLT sequences, including Recall@1<1m(%)↑, Mean Position Error (m)↓, Mean RPE (m)↓, and Mean Jitter (m/f^3)↓.

Method	RALoc(ICCV'25)				LightLoc(CVPR'25)				GTRLoc(NeurIPS'25)				BiLi(Ours)
Baseline	Origin	B1	B2	B3	Origin	B1	B2	B3	Origin	B1	B2	B3	
Recall@1 <1 m (%) ↑	37.37	37.68	38.09	38.38	77.02	77.35	77.56	78.27	84.29	84.55	84.63	85.16	**93.10**
Mean Position Error (m) ↓	2.07	2.82	2.82	2.78	1.46	1.70	1.70	1.68	1.40	1.25	1.26	1.24	**0.49**
Mean RPE (m) ↓	2.44	2.42	2.43	2.26	1.21	1.19	1.20	1.11	0.80	0.78	0.80	0.73	**0.26**
Mean Jitter (m/f^3) ↓	8.14	8.07	8.13	7.44	3.94	3.87	3.92	3.56	2.57	2.50	2.55	2.29	**0.76**

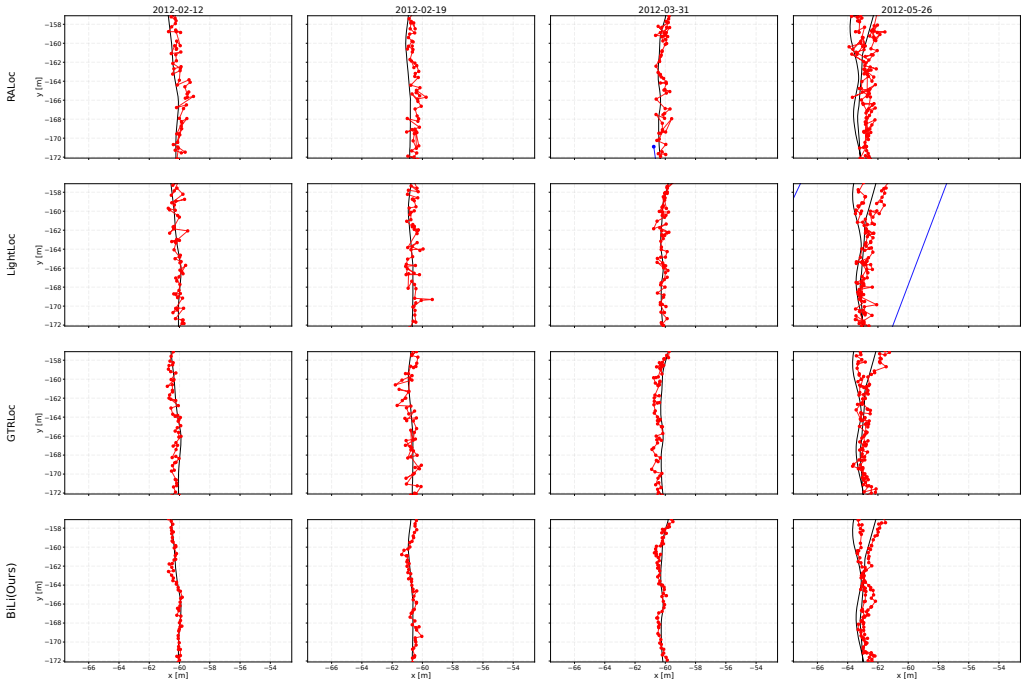

Figure 7: Visualization of localization trajectories on NCLT. We compare the original SCR baselines RALoc, LightLoc, and GTRLoc with BiLi(Ours) across four test sequences.

to the odometry trajectory and never modifies the original SCR outputs. The final reported metrics are computed from the complete sequence-level fused trajectories.

B1 is a rule-based robust fusion baseline that treats SCR poses as absolute anchors and odometry as relative-motion priors. It computes an $SE(3)$ residual between the SCR observation and the motion-propagated pose, and clips or rejects corrections when the residual becomes abnormally large.

B2 replaces this fixed rule with a learned MLP-based fusion gain under the same pose-level input setting. It is trained with a leave-one-sequence-out protocol to avoid test-sequence leakage.

B3 further strengthens the post-hoc baseline by optimizing a short $SE(3)$ pose chain with SCR absolute anchors and odometry relative-motion factors under a robust loss.

Overall, these baselines intentionally represent strong post-hoc alternatives to BiLi. They can test how much trajectory stability can be recovered by pose-level fusion after SCR prediction, while also exposing the limitation of such post-processing: because they operate only on already-estimated poses, they cannot correct the point-level spatial representation confounding or learn scene-aware kinematic priors as BiLi does.

Table 9: Evaluation of pose-level SCR-Odometry fusion on the Oxford dataset. Metrics are averaged over four Oxford sequences, including Recall@1<1m(%)↑, Mean Position Error (m)↓, Mean RPE (m)↓, and Mean Jitter (m/f^3)↓.

Method	RALoc(ICCV'25)				LightLoc(CVPR'25)				GTRLoc(NeurIPS'25)				BiLi(Ours)
Baseline	Origin	B1	B2	B3	Origin	B1	B2	B3	Origin	B1	B2	B3	
Recall@1 <1 m (%) ↑	26.91	26.95	27.31	27.07	80.85	81.18	81.04	81.87	83.55	83.80	83.66	84.34	**90.90**
Mean Position Error (m) ↓	1.50	1.50	1.49	1.49	0.79	0.78	0.78	0.77	0.75	0.75	0.75	0.74	**0.56**
Mean RPE (m) ↓	0.56	0.70	0.71	0.65	0.45	0.60	0.62	0.56	0.42	0.56	0.57	0.52	**0.16**
Mean Jitter (m/f^3) ↓	0.86	2.16	2.22	1.98	0.68	1.79	1.85	1.64	0.66	1.65	1.71	1.52	**0.22**

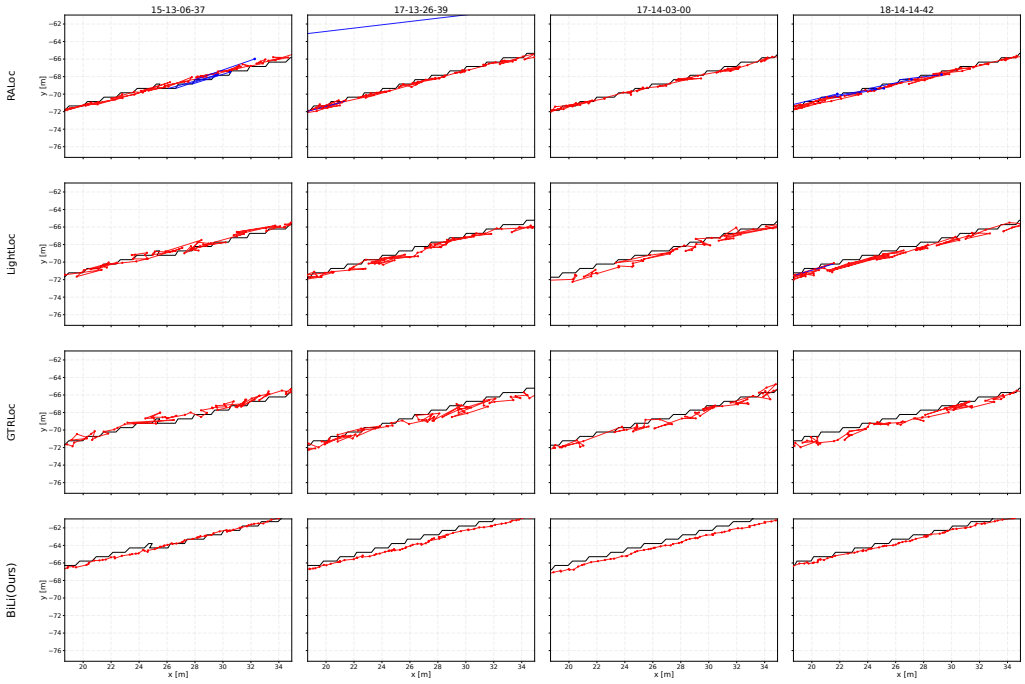

Figure 8: Visualization of localization trajectories on Oxford. We compare the original SCR baselines RALoc, LightLoc, and GTRLoc with BiLi(Ours) across four test sequences.

C.2 Results on NCLT under enhanced pose-level fusion baselines

As shown in Tab. 8, conventional pose-level SCR-Odometry fusion does improve the original SCR trajectories, but the improvement remains limited even under the strongest enhanced setting. Among these enhanced baselines, B3 generally achieves the strongest post-processing results, suggesting that window-based robust optimization defines a strong empirical ceiling for post-hoc pose-level stabilization. For the strongest SCR baseline GTRLoc, B3 improves Recall@1<1m from 84.29% to 85.16%, reduces the mean position error from 1.26 m to 1.24 m, and decreases RPE/Jitter from 0.80 m/2.57 m/f^3 to 0.73 m/2.29 m/f^3. These results show that pose-level fusion can smooth the already-predicted pose sequence, but it still cannot fundamentally remove the instability caused by frame-wise SCR predictions.

More importantly, BiLi remains substantially superior even when compared with these enhanced SCR baselines. Compared with the strongest enhanced baseline GTRLoc+B3, BiLi achieves 93.1% Recall@1<1m, 0.49 m mean position error, 0.26 m RPE, and 0.76 m/f^3 trajectory jitter. This corresponds to a 7.94 percentage-point improvement in Recall@1<1m, together with 60%, 64%, and 67% reductions in mean position error, RPE, and jitter, respectively. This large margin demonstrates that the last-mile instability cannot be solved by post-processing pose sequences alone. Instead, BiLi resolves the problem from inside the localization pipeline by combining implicit manifold regularization with learned temporal kinematic priors.

Figure 7 provides a direct qualitative comparison against the original SCR outputs. Across the four NCLT sequences, RALoc, LightLoc, and GTRLoc still exhibit visible high-frequency oscillations, fragmented trajectory traces, and occasional large spatial deviations under severe seasonal variations. In contrast, BiLi generates a much more continuous and jump-free trajectory that remains tightly aligned with the ground truth. This visual evidence confirms that BiLi does not merely smooth unstable outputs, but produces intrinsically stable localization suitable for safety-critical autonomous navigation.

C.3 Results on Oxford under enhanced pose-level fusion baselines

Table 9 shows a consistent trend on Oxford. Pose-level odometry fusion brings moderate improvements over the original SCR outputs, especially on smoothness-related metrics. For the strongest baseline GTRLoc, B3 improves Recall@1<1m from 83.55% to 84.34%, reduces mean position error from 0.75 m to 0.74 m, and lowers RPE/Jitter from 0.57 m/1.72 m/f^3 to 0.52 m/1.52 m/f^3. This verifies that enhanced pose-level fusion can alleviate part of the local temporal fluctuation. However, its gains are still bounded by the quality of the underlying SCR observations, because it operates only after pose estimation and has no access to point-level geometry or learned kinematic states.

By contrast, BiLi achieves a much stronger accuracy-stability trade-off even against the enhanced baselines. Compared with GTRLoc+B3, BiLi improves Recall@1<1m from 84.34% to 90.9%, while reducing mean position error from 0.74 m to 0.56 m. More crucially, BiLi slashes RPE from 0.52 m to 0.16 m and trajectory jitter from 1.52 m/f^3 to 0.22 m/f^3, corresponding to reductions of 69% and 86%, respectively. These substantial improvements demonstrate that BiLi goes far beyond conventional pose-level smoothing: IMR suppresses spatially inconsistent scene-coordinate predictions at the point/manifold level, while MDKD and DMG enforce temporally plausible pose updates through learned motion priors.

Figure 8 further visualizes the difference against the original SCR methods. Although RALoc, LightLoc, and GTRLoc can roughly follow the global route, their trajectories still contain jagged local deviations and sporadic high-error segments, showing that low average error does not necessarily imply deployment-level stability. Conversely, BiLi produces a smooth, structurally coherent, and tightly aligned trajectory across all Oxford sequences. This qualitative result is consistent with the quantitative reductions in RPE and jitter, confirming that BiLi effectively bridges the last-mile gap between accurate localization and stable autonomous navigation.

D Limitations and Future Work

While BiLi effectively resolves the spatial jittering challenge and achieves state-of-the-art real-time performance, it has a few limitations that point to interesting future research directions:

1. Static Map Assumption. BiLi extracts spatial constraints from an offline neural implicit map via the IMR module. If the physical environment undergoes massive, long-term structural changes after the mapping phase (e.g., construction zones or heavy snow), the pre-built Signed Distance Function (SDF) zero-level set becomes outdated, potentially misguiding the spatial regularization.

2. Extreme Joint Degeneracy. The DMG module acts as a robust "shock absorber" by dynamically fusing SCR absolute poses and MDKD relative priors. However, in prolonged featureless environments (e.g., an extremely long, blank tunnel), both the SCR backbone and the vision/LiDAR-driven MDKD prior may degrade simultaneously. Without a reliable signal from either modality, the system will eventually accumulate drift.

3. Offline Computational Cost. To guarantee strict real-time online inference, BiLi pushes the heavy burden of $SE(3)$ Lie algebraic optimization and manifold generation to the offline Teacher. This makes the initial mapping phase computationally intensive, limiting its immediate scalability for massive, city-scale datasets without large compute clusters.

Future Work. To address these boundary conditions, our future work will focus on three areas: (1) developing "lifelong learning" mechanisms to dynamically update the implicit map on the fly; (2) tightly coupling the Mamba state-space model with inertial sensors (IMU) to provide absolute kinematic anchors in featureless tunnels; and (3) adopting highly efficient rendering paradigms, such as 3D Gaussian Splatting, to significantly reduce the offline training overhead.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The abstract and introduction clearly reflect the contributions and scope of this paper.

   Guidelines:

   - The answer [NA] means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A [No] or [NA] answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: In Appendix D, we discuss the limitations of our work.

   Guidelines:

   - The answer [NA] means that the paper has no limitation while the answer [No] means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [NA]

Justification: Our work does not include theoretical results.

Guidelines:

- The answer [NA] means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: We have disclosed the experimental protocols, dataset splits, implementation details, and hyperparameters needed to reproduce the main results.

   Guidelines:

   - The answer [NA] means that the paper does not include experiments.
   - If the paper includes experiments, a [No] answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
   - While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
     (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
     (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
     (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
     (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: [Yes]

Justification: The code will be open-sourced upon acceptance of the paper.

Guidelines:

- The answer [NA] means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so [No] is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer) necessary to understand the results?

   Answer: [Yes]

   Justification: The main details are shown in Section 4 and Appendix B.

   Guidelines:

   - The answer [NA] means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [No]

   Justification: Following common practice among outdoor LiDAR relocalization benchmarks, we do not report error bars; per-sequence results are provided for the main comparisons.

   Guidelines:

   - The answer [NA] means that the paper does not include experiments.
   - The authors should answer [Yes] if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
   - The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
   - The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
   - The assumptions made should be given (e.g., Normally distributed errors).
   - It should be clear whether the error bar is the standard deviation or the standard error of the mean.

- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g., negative error rates).
- If error bars are reported in tables or plots, the authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: Details of the computational resources and runtime-memory statistics are provided in Sections 4, with complementary storage details in Appendix 4.

   Guidelines:

   - The answer [NA] means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: Our work conforms to the NeurIPS Code of Ethics.

   Guidelines:

   - The answer [NA] means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer [No] , they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [NA]

    Justification: There is no societal impact of this work.

    Guidelines:

    - The answer [NA] means that there is no societal impact of the work performed.
    - If the authors answer [NA] or [No] , they should explain why their work has no societal impact or why the paper does not address societal impact.
    - Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.

- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate Deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pre-trained language models, image generators, or scraped datasets)?

    Answer: [NA]

    Justification: The paper poses no such risks.

    Guidelines:

    - The answer [NA] means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [Yes]

    Justification: We cite the original papers that produce the codes and datasets.

    Guidelines:

    - The answer [NA] means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.
    - For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
    - If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
    - For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.

- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [Yes]

    Justification: We will be open-sourced upon acceptance of the paper.

    Guidelines:

    - The answer [NA] means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [NA]

    Justification: This paper does not involve crowdsourcing nor research with human subjects.

    Guidelines:

    - The answer [NA] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [NA]

    Justification: This paper does not involve crowdsourcing nor research with human subjects.

    Guidelines:

    - The answer [NA] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
    - We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
    - For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does *not* impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

Answer: [NA]

Justification: The core method development in this research does not involve LLMs as any important, original, or non-standard components.

Guidelines:

- The answer [NA] means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
- Please refer to our LLM policy in the NeurIPS handbook for what should or should not be described.