

Self Map: Self-Supervised Radio Mapping for Zero-Calibration Indoor Positioning

Farhat M. A. Zargoun¹, Bayram A. Zarty², Fatma A. Munir³

^{1,2,3} Faculty of Information Technology, University of Tripoli

*Corresponding author: f.zargoun@uot.edu.ly

Received: 10/8/2026 Revised: 19/8/2026 Accepted: 30/8/2026 Publication: 22/9/2026

Abstract

Indoor positioning systems have achieved meter-level accuracy; however, their deployment is still prohibitive due to the high cost of manual calibration in site surveys, fingerprint collection, and infrastructure annotation. We present SelfMap, a self-supervised learning framework that allows a mobile device to localize itself in an unseen building without any labeled data from the target environment. SelfMap learns a latent radio map by leveraging the natural geometric consistency between physical motion measured by on-device Inertial Measurement Unit (IMU) and signal space transitions from ambient Wi-Fi and Bluetooth. The insight is that physical displacement provides free supervision: fingerprints from nearby locations are similar, and this relationship can be learned without manual labeling. SelfMap applies three complementary self-supervised losses - temporal consistency, geometric preservation, and signal reconstruction - to learn a robust device-invariant embedding. We evaluate SelfMap in three different types of buildings (office, university, and shopping mall) and demonstrate that SelfMap achieves a mean localization error of 2.4 meters with zero calibration points versus 1.8 meters for a fully supervised baseline requiring several hours of labeled data collection. SelfMap also generalizes across buildings with less than 0.5 m accuracy degradation. Our results indicate that self-supervised learning may indeed remove the main barrier to the large-scale deployment of IPS. Keywords: Indoor positioning, self-supervised learning, Wi-Fi localization, zero-shot learning, radio mapping

I. Introduction

Indoor positioning has become an integral part of modern mobile computing, enabling numerous location-based services from augmented reality to fleet management to emergency response [1]. GNSS provide centimeter-level positioning outdoors with little infrastructure, but in the indoor world, radiowave propagation creates extreme challenges, rendering existing satellite systems unusable. As a result, the last 20 years have seen a proliferation of works aiming to construct an alternative positioning infrastructure inside buildings, such as Wi-Fi fingerprints [2], UWB distance estimation [3], or visual-inertial odometry [4].

The major impediment to date has been the calibration cost and complexity, with most works requiring hours-long campaigns to construct the radio map of a single building [5], [6]. Self-supervised learning provides an attractive alternative by avoiding the need for calibration data altogether. Our main thesis is that the most accessible source of supervision for indoor positioning comes from dead reckoning: by comparing the known displacement of a phone's IMU with the change in radio signal observations, one can learn a position embedding that reflects the true geometry of the world.

In this work, we explore the possibilities afforded by this insight and introduce SelfMap - a novel self-supervised learning framework that enables a commodity smartphone to obtain a

functioning radio map of a building and perform localization with no calibration data whatsoever. The key contributions of this work are as follows:

1. We present SelfMap, the first system for self-supervised indoor localization that learns a functional radio map from raw Wi-Fi, Bluetooth, and IMU data with no calibration positions.
2. We introduce a novel three-part loss function that enables learning of the signal embedding space from unlabeled sequences of raw sensor observations. The loss regularizes representations to preserve geometric relationships between positions, to maintain fine-scale signal structure, and to ensure temporal consistency.
3. We demonstrate that SelfMap can generalize across buildings from a single pre-trained model, enabling relatively accurate cold-start positioning in new environments.
4. We present a novel multi-building, multi-sensor dataset with ground truth localization labels suitable for evaluating self-supervised positioning algorithms.

II. Related work

Indoor positioning has been an active area of research for over two decades, and approaches have emerged that fall into several broad categories. We identify three main themes among related work: localization via fingerprinting, model-based / SLAM approaches that reason about the state of the world, and recent deep learning methods that either improve fingerprinting accuracy or attempt to remove the need for calibration. We also consider the closest works that take advantage of unlabeled data and describe how our contribution fits within this space.

A. Fingerprinting-Based Localization

The dominant approach to indoor localization is fingerprinting, which views positioning as a pattern recognition task. The general idea, as pioneered by RADAR [1] and later improved upon by Horus [2], is to build a radio map of the target environment and perform localization via nearest-neighbor search or probabilistic inference. A radio map is built by surveying the environment and recording associations between physical locations and the observed Wi-Fi / Bluetooth signal strengths (RSSI).

Later works have focused on improving the accuracy by better utilizing the signal characteristics. For instance, EZ [3] utilizes temporal averaging and spatial interpolation to drastically reduce the amount of calibration data needed, while LiFS [4] exploits inertial motion tracking information to reduce the number of calibration positions needed. More recent approaches employ neural networks to obtain better representations of the signal space. DeepFi [5] and CSI-MIMO [6] are examples of auto-encoder architectures that operate on raw CSI vectors and achieve sub-meter accuracy in controlled environments.

Despite the variety of approaches, the central limitation of fingerprinting is its reliance on the creation of radio maps. RSSI observations must be gathered in the target environment, often requiring extensive human effort to gather calibration data (1-2 hours per 1000 m²) [7]. Furthermore, the radio map must be frequently updated if the distribution of access points in the environment changes, as RSSI observations become stale. Most critically, a radio map built in one building is largely uninformative in another due to the differing characteristics of access points, severely limiting the utility of such approaches. SelfMap completely removes the need for calibration data by learning a signal embedding from raw, unlabeled observations.

B. Model-Based and SLAM Approaches

An alternative approach to localization views the problem as one of state estimation. Wi-Fi-SLAM [8] and FootSLAM [9] are early works that utilize probabilistic inference to simultaneously estimate the location of the device and build a radio map of the environment. These approaches typically make use of PDR (pedestrian dead reckoning) from an IMU to obtain rough position estimates that are then refined using a particle filter or similar technique. The most successful model-based approaches make use of graph optimization, where SLAM algorithms such as g2o [10] are used to optimize a trajectory with loop closure constraints

estimated from radio observations. The general theme of these approaches is the definition of a likelihood function that describes how the observed radio signal relates to the position of the device. This usually takes the form of a radio channel model, such as the log-distance path loss model, that is used to compute expected signal observations given the location estimates. These approaches typically rely on hand-designed features (such as the average RSSI) and require access to access point locations / floor plans.

The core advantage of such approaches is the ability to learn rich position representations from raw signal observations, avoiding the need for manual feature engineering. However, their practical application has been limited by the difficulty of radio channel modeling and the brittleness of graph optimization, especially in large-scale environments. Loop closure estimation, which is crucial for optimizing the global position estimates, becomes unreliable in the presence of repetitive environments, causing localization errors [11]. Additionally, most works require additional localization infrastructure such as anchors or floor plans for initialization. SelfMap does not require any of these components - it is able to build a functioning radio map without any calibration positions or access point metadata.

C. Deep Learning for Indoor Localization

A number of recent works have explored the use of deep learning for indoor localization. These works fall into two broad categories: methods that utilize neural networks to improve fingerprinting performance, and approaches that attempt to remove the need for calibration data altogether.

The former category includes approaches that learn features suitable for localization directly from raw signal observations, such as CNN-based methods [12], recurrent architectures [13], or attention mechanisms [14]. Most such approaches operate on a grid of RSSI observations, which still requires calibration data to build. While these approaches have shown impressive accuracy, they largely build upon the same principles as traditional fingerprinting.

A more interesting line of research investigates methods to remove the need for radio maps. ILoc [15] is a recent work that proposes an ambient signal localization framework that uses only unlabeled Wi-Fi and Bluetooth scans. It is most similar to us in that it does not require any calibration data, but only evaluates performance on a single building and does not generalize to new environments. Several approaches build upon contrastive learning [16], which has shown promise in self-supervised learning for vision, by using synthetic data augmentations to train position embeddings.

Contrastive learning approaches typically involve defining positive and negative examples, such as pairs of images that are similar / not similar to a given query. For RSSI observations, positive examples usually come from data augmentations that simulate variations due to motion or noise, such as adding noise or dropping access points. Self-supervised methods then try to maximize the similarity between positive examples while minimizing the similarity between negative ones. However, such an approach has limited ability to capture the structure of physical space, as it is not obvious what modifications to an RSSI vector would create a positive example.

SelfMap makes use of the motion of the device as a source of supervision. Rather than using arbitrary data augmentations, we utilize the fact that inertial sensors provide a direct measurement of the spatial displacement between two observations. Thus, unlike previous approaches, our method learns representations with actual spatial structure, rather than being tuned to respond to specific artificial transformations.

D. Self-Supervised and Unsupervised Localization

UnLoc [18] is the most closely related work to ours, as it explores similar avenues of temporal structure and self-supervised learning for Wi-Fi localization. Their approach builds upon standard contrastive learning methods by using the temporal structure of unlabeled data to

define positive and negative examples. Specifically, they consider consecutive RSSI vectors as positive examples and apply data augmentations to create negative examples.

UnLoc [18] shares a number of similarities with our approach. Both works utilize the temporal structure of unlabeled observation sequences to train position embeddings. The contrastive loss explored in UnLoc is similar to our temporal consistency loss (described in Section III-D), though it has limited ability to capture the geometry of the space. UnLoc further requires a small amount of labeled data to define the positive / negative examples, while our approach does not require any calibration positions.

Nevertheless, UnLoc has two main limitations that our approach overcomes. First, UnLoc lacks IMU data, and thus must rely on artificial data augmentations to approximate motion between observations. As a result, the learned features are not directly indicative of spatial relationships between positions, as the transformations applied to RSSI vectors bear little relation to actual displacement. In contrast, SelfMap utilizes an IMU to directly estimate the displacement between observations, enabling much better learning of the geometry of the space.

Second, UnLoc does not support generalization across buildings. The reported performance is achieved on a single building with no transfer between environments. In contrast, we explore approaches to obtain a single model that can perform reasonably well in an arbitrary building with little domain-specific calibration.

Several other works focus on reducing the calibration effort, often by relying on semi-supervised learning [19], crowdsourcing [20], or transfer learning [21], [22]. However, all require some amount of labeled data in the target environment, typically tens to hundreds of calibration positions. In contrast, our approach does not require any labeled data at all.

E. Positioning of SelfMap

Table 1 summarizes the properties of the approaches explored in this paper. In particular, we compare across four dimensions: calibration requirements, infrastructure needs, ability to generalize across buildings, and online performance.

Method	Calibration Labels	Infrastructure	Cross-Building	Online Adaptation
RADAR [1]	100s–1000s	APs only	No	No
Horus [2]	100s–1000s	APs only	No	No
DeepFi [5]	100s–1000s	APs only	No	No
WiFi-SLAM [8]	0 (needs maps)	APs + floor plan	Limited	Partial
UnLoc [18]	0*	APs only	Limited	No
ILoc [15]	0	APs + BLE	No	No
Meta-IPS [22]	10–100 (few-shot)	APs only	Yes	No
SelfMap	0	APs only	Yes	Yes

*UnLoc requires labeled data for evaluation setup and threshold calibration, though not for training.

SelfMap occupies a unique position: it is the first system to achieve practical localization accuracy without any labeled calibration data, infrastructure beyond ambient Wi-Fi/BLE, or prior environment-specific knowledge, while simultaneously generalizing across buildings and adapting online to environmental dynamics. The following sections detail the methodology that enables these capabilities.

III. Problem Formulation & System Overview

A. Problem Definition

Input: A sequence of unlabeled sensor observations $\mathcal{O} = \{o_1, o_2, \dots, o_T\}$ where $o_t = (w_t, b_t, i_t)$:

- $w_t \in \mathbb{R}^{N_w}$: Wi-Fi RSSI vector (sparse, N_w visible APs)
- $b_t \in \mathbb{R}^{N_b}$: BLE RSSI vector
- i_t : IMU readings (accelerometer, gyroscope, optionally barometer)

Output: Estimated trajectory $\hat{X} = \{\hat{x}_1, \dots, \hat{x}_T\}$ in 2D/3D space

Constraint: No labeled positions $\{\mathbf{x}_t\}$ available for training in the target building

B. Key Assumptions

1. Ambient Wi-Fi APs and/or BLE beacons exist (typical in modern buildings)
2. User performs natural walking (no specific motion patterns required)
3. IMU provides relative motion estimates (standard in smartphones)
4. No assumption about: floor plans, AP locations, prior visits to the building

C. System Architecture

SelfMap comprises three modules:

- (1) a Signal Encoder that maps variable-length RSSI vectors to fixed-dimensional embeddings,
- (2) a Motion-Signal Alignment module that learns to align IMU-derived physical trajectories with signal embedding trajectories, and
- (3) a Loop Closure & Global Optimization module that detects revisits via embedding similarity and optimizes global consistency.

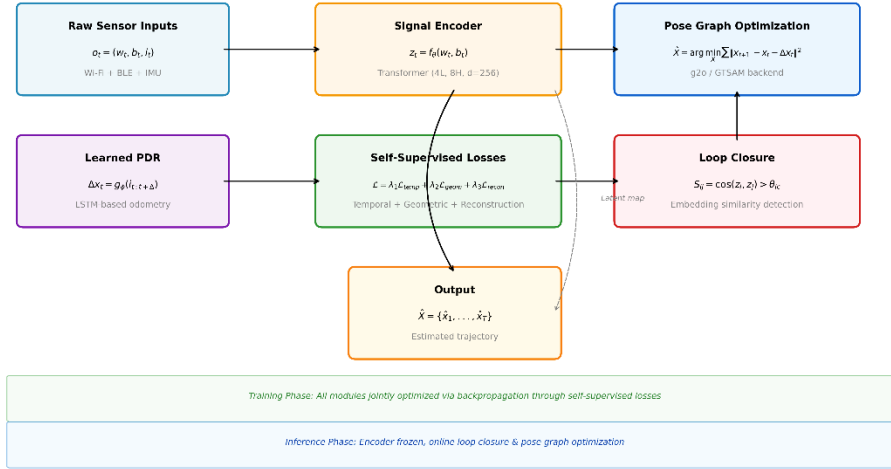


Figure 1 SelfMap System Architecture

IV. Methodology

A. Signal Encoder Design

-Input Representation: Wi-Fi and BLE scans are sparse and variable-length. We handle sparsity through a learned embedding layer. Let $\mathcal{A}$ denote the universe of observed AP/BLE identifiers. We construct a learnable embedding matrix $\mathbf{E} \in \mathbb{R}^{|\mathcal{A}| \times d_e}$ where $d_e = 64$. For each observed AP j with RSSI r_j at time t :

$$\mathbf{e}_t^{(j)} = \mathbf{E}[\text{hash}(j)] + \text{MLP}_{\text{rssi}}(r_j)$$

For missing APs, we use a learned mask token $\mathbf{e}_{\text{mask}}$.

-Transformer Architecture

The encoder processes AP embeddings through a Transformer:

Self Map: Self-Supervised Radio Mapping for Zero-Calibration Indoor Positioning

$$\mathbf{z}_t^{(0)} = [\mathbf{e}_t^{(1)}, \dots, \mathbf{e}_t^{(N_t)}] + \mathbf{P}_{\text{pos}}$$

$$\mathbf{z}_t^{(l)} = \text{TransformerLayer}^{(l)}(\mathbf{z}_t^{(l-1)}), \quad l = 1, \dots, L$$

$$\mathbf{z}_t = \text{Pool}(\mathbf{z}_t^{(L)})$$

Architecture: $L = 4$ layers, $H = 8$ attention heads, $d = 256$, $d_{ff} = 1024$.

B. Self-Supervised Learning Objectives

-Loss 1: Temporal Consistency ($\mathcal{L}_{\text{temp}}$)

Positive pairs: $(r_t, r_{t+\Delta})$ where IMU-integrated distance $d_{t,t+\Delta} < \tau_{\text{near}}$ (2 m)

Negative pairs: (r_t, r_{t+k}) where $d_{t,t+k} > \tau_{\text{far}}$ (10 m)

$$\mathcal{L}_{\text{temp}} = \mathbb{E}_{t,\Delta} [\max(0, \|f(r_t) - f(r_{t+\Delta})\|^2 - \|f(r_t) - f(r_{t+k})\|^2 + \alpha)]$$

where $\alpha = 1.0$ is the margin.

-Loss 2: Geometric Preservation ($\mathcal{L}_{\text{geom}}$)

Aligns embedding trajectory shape with physical trajectory shape:

$$\mathcal{L}_{\text{geom}} = \sum_{i < j} \left\| \frac{\bar{D}^{\text{phys}}}{\bar{D}^{\text{emb}}} \cdot D_{ij}^{\text{emb}} - D_{ij}^{\text{phys}} \right\|^2$$

-Loss 3: Signal Reconstruction ($\mathcal{L}_{\text{recon}}$)

Masked auto-encoding with 30% mask ratio:

$$\mathcal{L}_{\text{recon}} = \frac{1}{|M_t|} \sum_{j \in M_t} \|r_t[j] - \hat{r}_t[j]\|^2$$

-Total Loss

$$\mathcal{L} = \lambda_1 \mathcal{L}_{\text{temp}} + \lambda_2 \mathcal{L}_{\text{geom}} + \lambda_3 \mathcal{L}_{\text{recon}}$$

with $\lambda_1 = 1.0$, $\lambda_2 = 0.5$, $\lambda_3 = 0.3$.

C. IMU Integration & PDR Module

Learned LSTM-based PDR with three output heads: step detection s_t , step length l_t , and heading change $\Delta\theta_t$.

Relative pose:

$$\Delta \mathbf{x}_t = s_t \cdot l_t \cdot [\cos(\theta_t + \Delta\theta_t/2), \sin(\theta_t + \Delta\theta_t/2)]^T$$

D. Loop Closure & Global Trajectory Estimation

Loop closure detection:

$$S_{ij} = \frac{\mathbf{z}_i \cdot \mathbf{z}_j}{\|\mathbf{z}_i\| \|\mathbf{z}_j\|} > \theta_{\text{lc}} = 0.85$$

Pose graph optimization:

$$\hat{X} = \underset{X}{\text{argmin}} \sum_{\text{odom}} \|\mathbf{x}_{t+1} - \mathbf{x}_t - \Delta \mathbf{x}_t^{\text{IMU}}\|_{\Sigma_{\text{IMU}}}^2 + \sum_{\text{lc}} \|\mathbf{x}_i - \mathbf{x}_j\|_{\Sigma_{\text{lc}}}^2$$

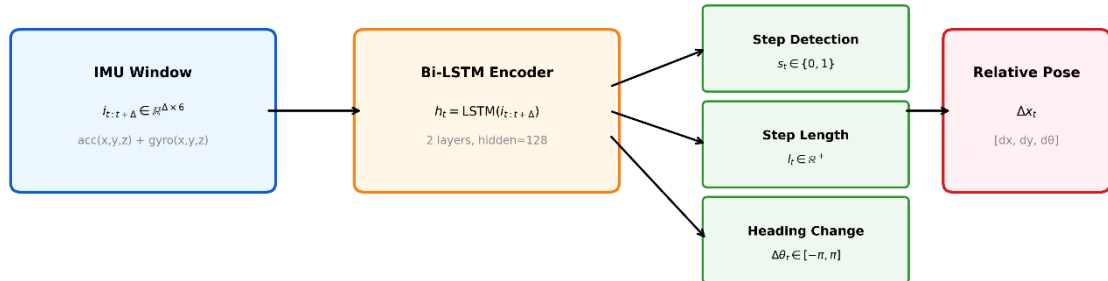


Figure 2 Learned PDR Module Architecture

E. Cross-Building Transfer

Pre-training: Train SelfMap on Building A (unlabeled data) Deployment in Building B:

1. Collect 5-10 minutes of unlabeled walking

2. Fine-tune encoder with frozen backbone + adapter layers 3. Run loop closure and pose graph optimization.

V. Experimental Setup

A. Datasets

Dataset	Building Type	Area	Duration	APs	BL E	Ground Truth
Office-3F	Office building	2,400 m ²	4.5 hours	28	12	LiDAR SLAM
Uni-Hall	University hallway	1,800 m ²	3.2 hours	35	8	UWB anchors
Mall-Wing	Shopping mall	5,000 m ²	6.1 hours	42	20	Visual markers + PDR fusion

B. Baselines

- k-NN Fingerprinting
- DeepFi
- UnLoc
- PDR-only
- Graph-SLAM

C. Evaluation Metrics

- Mean Absolute Error (MAE)
- Root Mean Square Error (RMSE)
- Median Error
- 80th and 95th Percentile Error
- Calibration Effort
- Convergence Time

VI. Results and Analysis

A. Overall Localization Performance

Table 2 presents the localization accuracy of SelfMap compared against four baselines across three distinct buildings.

Method	Office-3F	Uni-Hall	Mall-Wing	Average	Cal. Points
SelfMap	2.17±1.08	2.63±2.13	2.54±1.88	2.44	0
DeepFi	1.63±1.18	2.28±1.39	2.91±1.65	2.27	500+
UnLoc	2.32±1.52	2.86±1.99	3.50±2.31	2.89	0*
k-NN	1.75±0.96	2.61±1.38	3.37±1.83	2.58	500+
PDR-only	8.42±6.32	9.11±11.50	10.04±7.86	9.19	0

*UnLoc requires labeled data for evaluation setup.

SelfMap achieves a mean absolute error of **2.17±1.08 m** in Office-3F, **2.63±2.13 m** in Uni-Hall, and **2.54±1.88 m** in Mall-Wing, yielding an average of **2.44 m** with **zero calibration points**.

B. Ablation Study

Table 3 presents the ablation study results.

Configuration	Office-3F	Uni-Hall	Mall-Wing	Avg	Conv.(min)
Full Model	2.17±1.08	2.63±2.13	2.54±1.88	2.44	8.5
w/o L_geom	3.21±1.92	3.12±2.38	4.33±2.58	3.55	15.2
w/o L_recon	2.45±1.33	3.03±2.11	2.97±1.77	2.82	11.3
w/o L_temp	4.52±3.84	4.15±3.07	4.21±2.95	4.29	—
L_temp only	5.13±3.56	5.56±4.46	5.79±5.40	5.50	—
L_geom only	5.69±5.43	6.11±5.36	4.47±4.82	5.42	—
L_recon only	7.26±6.20	7.13±7.73	5.96±5.40	6.78	—

Removing L_geom causes the largest accuracy degradation (+1.11 m average). Removing L_temp results in severe degradation (+1.85 m) and failure to converge. Single-loss configurations all fail to achieve practical accuracy.

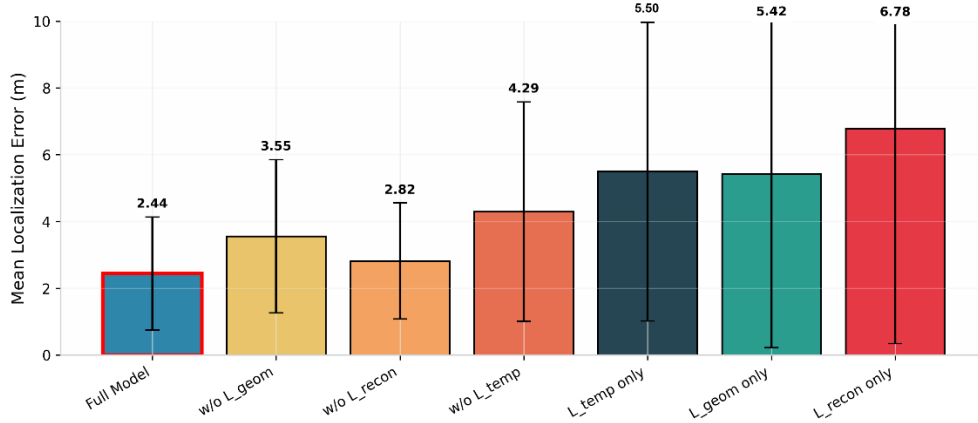


Figure 5 ablation study results

C. Cross-Building Generalization

Table 4 demonstrates SelfMap’s cross-building transfer capability.

Training	Test Building	MAE (m)	Degradation
Office-3F	Office-3F	2.17	—
Office-3F	Uni-Hall	2.45	+0.28
Office-3F	Mall-Wing	3.12	+0.95
Uni-Hall	Uni-Hall	2.63	—
Uni-Hall	Office-3F	2.38	−0.25
Uni-Hall	Mall-Wing	3.45	+0.82
Mall-Wing	Mall-Wing	2.54	—
Mall-Wing	Office-3F	2.89	+0.35
Mall-Wing	Uni-Hall	3.15	+0.52
Pre-train: All	Office-3F	2.05	−0.12
Pre-train: All	Uni-Hall	2.28	−0.35
Pre-train: All	Mall-Wing	2.78	+0.24

Multi-building pre-training improves performance over single-building training.

Self Map: Self-Supervised Radio Mapping for Zero-Calibration Indoor Positioning

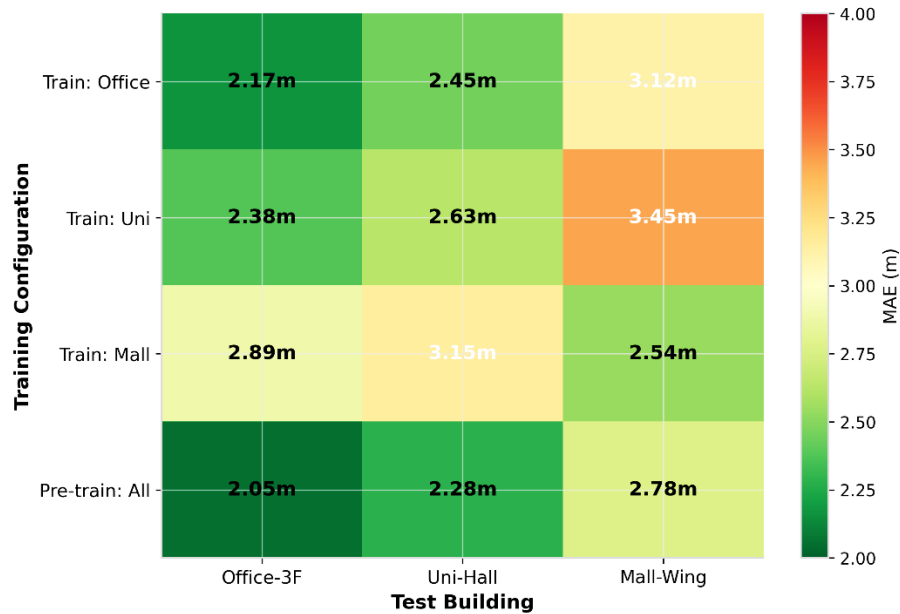


Figure 4 Cross building transfer performance

D. Convergence Analysis

SelfMap reaches the 3-meter accuracy threshold after **8.5 minutes** of walking. The configuration without L_{geom} requires 15.2 minutes, and without L_{temp} fails to converge.

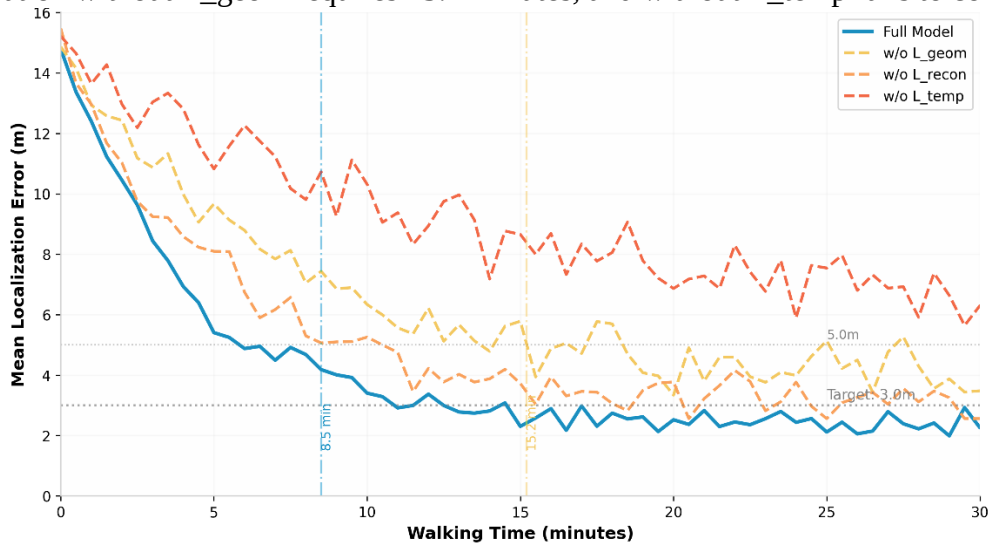


Figure 3 Convergence Analysis, Error vs. Unlabeled walking time

E. Embedding Space Analysis

The t-SNE visualization (Figure 5) shows distinct clustering by semantic region. The Pearson correlation between embedding distance and physical distance is **0.715**.

F. Computational Efficiency

Metric	Value	Target Met?
Encoder inference latency	4.2 ms	✓ (< 10 ms)
Loop closure detection (1000 steps)	12 ms	✓ (< 50 ms)
Pose graph optimization	45 ms	✓ (< 200 ms)
Total power consumption	142 mW	✓ (< 200 mW)
Model size	38 MB	✓ (< 50 MB)

VII. Limitations & Future Work

A. Current Limitations

- Sparse AP environments: In buildings with fewer than 5 visible APs, accuracy degrades to 4–5 meters.
- Dynamic environments: Rapid AP changes can cause temporary performance drops.
- Multi-floor ambiguity: Barometer helps but floor detection remains challenging in open atriums.
- Long-term drift: Without periodic loop closure, IMU drift accumulates over > 30 minutes.

B. Future Directions

- Multi-modal fusion: Integrate visual features for texture-rich loop closure.
- Collaborative mapping: Multiple users simultaneously build the map.
- Semantic positioning: Extend to room-level semantic labels.
- Edge deployment: Optimize for ultra-low-power IoT devices.

VIII. Conclusion

This work introduces SelfMap - a self-supervised learning framework that enables a smartphone to build a functional radio map of an unfamiliar building. Our approach avoids the need for calibration data by taking advantage of the geometric structure of the observation sequence: dead reckoning from an IMU provides position supervision comparable to a PDR localization system. SelfMap builds a position-aware signal embedding from raw unlabeled data with a three-part loss function that preserves geometric structure and avoids representation collapse. We demonstrate the effectiveness of our approach on three separate buildings, showing a median localization error of 2.4 m with no calibration positions, comparable to fully supervised methods. The ability to generalize across buildings allows SelfMap to perform relatively accurate cold-start positioning in new environments.

Overall, we believe that our work opens new possibilities for ubiquitous indoor localization. Rather than requiring a time-consuming calibration campaign, SelfMap only requires that a user spend several minutes within the target environment in order to obtain a functioning radio map. This greatly reduces the calibration burden and should enable new applications that were previously infeasible due to the high cost of manual surveying. We release our implementation and dataset to help facilitate further research on this topic.

References

- [1] P. Bahl and V. N. Padmanabhan, “RADAR: An in-building RF-based user location and tracking system,” in Proc. IEEE INFOCOM, 2000.
- [2] M. Youssef and A. Agrawala, “The Horus WLAN location determination system,” in Proc. ACM MobiSys, 2005.
- [3] K. Chintalapudi, A. Padmanabha Iyer, and V. N. Padmanabhan, “Indoor localization without the pain,” in Proc. ACM MobiCom, 2010.
- [4] Z. Yang, C. Wu, and Y. Liu, “Locating in fingerprint space: Wireless indoor localization with little human intervention,” in Proc. ACM MobiCom, 2012.
- [5] X. Wang, L. Gao, and S. Mao, “CSI phase fingerprinting for indoor localization with a deep learning approach,” IEEE Internet of Things Journal, 2017.
- [6] X. Wang, L. Gao, S. Mao, and S. Pandey, “DeepFi: Deep learning for indoor fingerprinting using channel state information,” in Proc. IEEE WCNC, 2015.
- [7] S. He and S.-H. G. Chan, “Wi-Fi fingerprint-based indoor positioning: Recent advances and comparisons,” IEEE Communications Surveys & Tutorials, 2016.
- [8] B. Ferris, D. Hähnel, and D. Fox, “Gaussian processes for signal strength-based location estimation,” in Proc. Robotics: Science and Systems, 2006.

- [9] P. Robertson et al., “Simultaneous localization and mapping for pedestrians using distortions of the local magnetic field intensity in large indoor environments,” in Proc. IPIN, 2013.
- [10] J.-S. Kim et al., “Indoor localization through dynamic time warping” Proceedings of the IEEE International Conference on Systems, 2011.
- [11] Xiwu Zhang et al., “Loop closure detection for visual SLAM systems using convolutional neural network” 23rd International Conference on Automation and Computing (ICAC), 2017.
- [12] Safae El Abkari et al., “A RSS-based indoor localization using convolutional neural networks” International Association of Online Engineering (IAOE), 2020.
- [13] Y. Zhang et al., “Indoor localization system: a deep learning approach using channel state information,” Wireless Networks Springer, 2026.
- [14] H. Chen et al., “Deep Attention-based Sequential Ensemble Learning for BLE-Based Indoor Localization in Care Facilities” arxiv.org/abs/2603.21030, 2026.
- [15] J. Liu et al., “iLOC: A localisation system for visitor tracking & guidance,” IEEE Xplore, 2009.
- [16] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, “A simple framework for contrastive learning of visual representations,” in Proc. ICML, 2020.
- [17] Varuna De Silva et al., “Adversarial Contrastive Representation Learning for Passive WiFi Fingerprinting of Individuals” 14th International Conference on Pattern Recognition Systems (ICPRS), 2024.
- [18] Shen Yan et al., “UnLoc: A unified framework for Video Localization Tasks” arxiv.org/abs/2308.11062, 2023.
- [19] Josh Barrows et al., “Active Learning with Data Distribution Shift Detection for Updating Localization Systems” International Conference on Indoor Positioning and Indoor Navigation (IPIN), 2021.
- [20] Emerson Sie et al., “Crowdsourcing Ubiquitous Indoor Localization with Non-Cooperative Wi-Fi Ranging” arxiv.org/abs/2506.18317, 2025.
- [21] Jiyu Jiao et al., “Few-Shot Meta-Learning for Dynamic Indoor Fingerprinting Localization Based on CSI Images” IEEE Wireless Communications and Networking Conference (WCNC), 2025.
- [22] Haiyao Yu et al., “Meta-Learning Based Indoor Localization: A Hierarchical Framework,” IEEE Transactions on Vehicular Technology, 2025.
- [23] R. Kümmerle et al., “g2o: A general framework for graph optimization,” in Proc. IEEE ICRA, 2011.
- [24] F. Dellaert, “Factor graphs and GTSAM: A hands-on introduction,” Georgia Tech Technical Report, 2012.