

Article

Measurement-Reliability Learning and Geometry-Constrained Fusion for Robust Wi-Fi FTM Indoor Localization

Siqi Guan, Siyi Ding * and Shaomian Huang

School of Information and Intelligent Science, Donghua University, Shanghai 201620, China;
230995212@mail.dhu.edu.cn (S.G.); 230995213@mail.dhu.edu.cn (S.H.)

* Correspondence: dingsiy@dhu.edu.cn

Abstract

Indoor positioning using commodity Wi-Fi infrastructure is attractive for smart buildings and Internet of Things applications, but the practical accuracy of Wi-Fi Fine Timing Measurement (FTM) remains limited by non-line-of-sight propagation, multipath delay, access-point-dependent ranging bias, and unstable anchor geometry. This paper proposes a measurement-reliability learning and geometry-constrained fusion framework, termed MRL-GCF, for robust horizontal Wi-Fi FTM indoor localization. MRL-GCF learns the reliability of each access-point observation from a multi-factor representation that includes Received Signal Strength Indicator (RSSI), logarithmic FTM range, short-window range stability, RSSI fluctuation, access-point visibility, abnormal-range tendency, and coarse anchor geometry. A lightweight heteroscedastic neural calibrator estimates both range bias and observation uncertainty. A supervised reliability-regime head is further trained from residual-regime soft targets, and its entropy is used as a propagation-ambiguity measure. The learned uncertainty is fused with propagation ambiguity, map obstruction, material-aware obstruction cues, and anchor geometry to select reliable anchors and construct a trust-weighted nonlinear least-squares localization objective. To avoid overestimating performance from repeated scans at identical survey points, both scan-level and point-held-out protocols were adopted. Experiments were conducted in a lobby, a classroom, and a dormitory using 4410 synchronized RSSI-FTM scans. On 882 scan-level test queries, MRL-GCF achieved mean absolute errors of 0.88 m, 0.55 m, and 1.20 m, with sub-3 m success rates of 98.0%, 99.0%, and 96.5%, respectively. Additional replay-based dynamic, temporal, cross-device, AP-density, uncertainty-calibration, map-availability, and coefficient-sensitivity analyses were included to examine deployment-oriented robustness. These results indicate that learning measurement reliability while preserving geometric constraints provides a practical and interpretable solution for robust Wi-Fi FTM indoor positioning.

Keywords: artificial intelligence; machine learning; indoor localization; Wi-Fi fine timing measurement; measurement-reliability learning; heteroscedastic neural calibration; uncertainty estimation; geometry-constrained fusion

Academic Editor: Young-Joo Suh

Received: 13 June 2026

Revised: 18 July 2026

Accepted: 5 August 2026

Published: 7 August 2026

Copyright: © 2026 by the authors.
Licensee MDPI, Basel, Switzerland.
This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

Indoor location awareness is a fundamental capability for smart buildings, industrial warehouses, hospitals, campuses, transportation hubs, emergency response systems, and spatially aware Internet of Things (IoT) applications [1–4]. Unlike outdoor positioning,

indoor localization cannot rely on stable satellite visibility. Wireless signals inside buildings are affected by walls, furniture, human movement, metallic objects, device orientation, and multipath reflections. A practical indoor positioning system must therefore balance accuracy, deployment cost, infrastructure complexity, and long-term maintainability.

Several sensing technologies have been investigated for indoor localization. Ultra-wideband (UWB) systems can provide accurate ranging, but they usually require dedicated anchors and additional deployment cost [5,6]. Bluetooth Low Energy (BLE) is inexpensive and power-efficient, but its accuracy is often affected by Received Signal Strength Indicator (RSSI) variability and device heterogeneity [7,8]. Visible light positioning can achieve high accuracy under favorable optical conditions, but it depends on specific lighting infrastructure and line-of-sight visibility [9]. Wi-Fi fingerprinting is attractive because Wi-Fi access points (APs) are already widely deployed; however, fingerprinting methods generally require dense radio-map construction and repeated maintenance when AP layouts or indoor environments change [10–13].

Wi-Fi Fine Timing Measurement (FTM), introduced in IEEE 802.11mc and further enhanced in later Wi-Fi location standards, enables round-trip-time-based ranging between an initiator and a responder [14–17]. Compared with RSSI-only fingerprinting, FTM provides explicit distance information and can support geometric localization using commodity Wi-Fi devices. Nevertheless, practical FTM ranging remains noisy and biased in indoor environments. Non-line-of-sight (NLOS) propagation, delayed first-path detection, multipath delay, AP-specific timing offset, and device implementation differences may introduce substantial positive range errors [18–21]. Direct least-squares localization from raw ranges may therefore become unstable when a small number of unreliable anchors dominate the position estimate.

Existing FTM localization studies improve robustness mainly through propagation-state identification, range correction, residual weighting, or robust optimization [18,22,23]. These methods are useful, but three limitations remain. First, many calibration methods rely on limited single-scan features, although repeated or short-window measurements contain valuable information about measurement stability. Second, uncertainty is often assigned using handcrafted rules or discrete state labels, and its relationship with actual ranging residuals is rarely validated explicitly. Third, map obstruction and anchor geometry are often handled separately from range calibration, although both directly affect whether a measurement should be trusted during localization.

In deployment-oriented indoor positioning, uncertainty does not arise only from wireless signal noise. It propagates through multiple coupled spaces, including sensing noise, device-dependent ranging bias, AP geometry, map availability, obstruction interpretation, temporal drift, and environmental change. Recent uncertainty-aware radio-mapping research has emphasized that scalable indoor positioning should trace and control uncertainty propagation across the sensing-to-map-to-localization lifecycle [24]. This view motivates the present work: measurement reliability should be learned at the AP-observation level and then propagated into geometry-constrained localization rather than treated as a post hoc confidence score.

This paper addresses these limitations by formulating Wi-Fi FTM localization as a measurement-reliability learning and geometry-constrained fusion problem. The proposed MRL-GCF framework learns AP-wise range bias and uncertainty, integrates the learned reliability with map and geometry priors, and estimates the user position through a weighted geometric solver. This hybrid design preserves the physical structure of range-based localization while allowing nonlinear measurement degradation patterns to be learned from data. The study is organized around the following research questions. RQ1: Can short-window measurement-reliability features improve Wi-Fi FTM range calibration beyond RSSI and raw FTM features alone? RQ2: Can a heteroscedastic neural

calibrator produce uncertainty estimates that are quantitatively aligned with actual ranging residuals? RQ3: Can map-geometry reliability fusion reduce localization error and tail-error cases when floor-map information is available, while remaining usable when map information is absent? RQ4: How stable is the proposed method under point-held-out testing, dynamic replay, coefficient perturbation, device variation, AP-density change, and temporal environmental variation?

The main contributions are summarized as follows.

1. A measurement-reliability representation is developed for Wi-Fi FTM localization by combining RSSI, logarithmic FTM range, short-window range stability, RSSI fluctuation, AP visibility, robust abnormality score, and coarse geometry descriptors.
2. A lightweight heteroscedastic neural calibration module is proposed to estimate AP-wise range bias and observation uncertainty. A supervised reliability-regime head is trained from residual-regime soft targets, and the predicted regime entropy is used to represent propagation ambiguity.
3. A map-geometry reliability fusion strategy is introduced for anchor selection and localization weighting by combining learned uncertainty, propagation ambiguity, signal confidence, map obstruction, material-aware obstruction cues, and geometric diversity.
4. A two-protocol evaluation is conducted to examine both repeated-scan robustness and spatial generalization across three indoor environments, supported by ablation analysis, uncertainty validation, map perturbation, runtime evaluation, spatial error visualization, dynamic replay, temporal robustness, cross-device robustness, AP-density analysis, coefficient-sensitivity analysis, and statistical effect-size reporting.

The remainder of this paper is organized as follows. Section 2 reviews related work. Section 3 presents the proposed method. Section 4 describes the experimental design. Section 5 reports and analyzes the results. Section 6 discusses the implications and limitations. Section 7 concludes the paper.

2. Related Work

2.1. Indoor Positioning Technologies

Indoor positioning has been studied using radio-frequency, optical, acoustic, inertial, and hybrid sensing technologies [1–4]. UWB systems estimate distance from precise time-of-flight measurements and are suitable for high-accuracy industrial localization [5,6]. BLE-based positioning can be deployed at low cost, but its accuracy is often affected by RSSI variability and environmental interference [7,8]. Visible light positioning provides accurate localization in structured lighting environments, but its deployment depends on illumination infrastructure and line-of-sight optical channels [9]. Wi-Fi-based methods remain widely studied because Wi-Fi networks are already available in most indoor environments.

2.2. RSSI Fingerprinting and Learning-Based Wi-Fi Localization

RSSI fingerprinting estimates positions by matching online signal vectors with an offline radio map [10–13]. Classical Wi-Fi fingerprinting systems showed that in-building signal patterns can support localization, and later studies improved matching robustness using probabilistic models, weighted nearest neighbors, interpolation, and collaborative correction [10–13]. Recent learning-based methods use deep networks or generative models to improve RSSI representation and reduce survey effort [12,13,25–28]. However, RSSI is only an indirect indicator of position. It is sensitive to body blockage, channel interference, device orientation, and AP configuration. Therefore, RSSI-only methods may perform poorly when the online signal distribution differs from the calibration distribution.

2.3. Wi-Fi FTM Ranging and Calibration

Wi-Fi FTM provides direct range measurements through two-way timing exchange [14–17]. It has been used for indoor navigation, crowdsourced Wi-Fi positioning, robot localization, and mobile computing [16,17,21]. Compared with RSSI fingerprinting, FTM introduces explicit geometric information. However, raw FTM ranges are affected by NLOS bias, multipath delay, device-specific timing error, and AP-dependent offset [18–21]. Range correction is therefore necessary before FTMs are used in geometric solvers.

Existing FTM calibration methods often estimate measurement quality by propagation-state classification or empirical residual modeling [18,22,23]. These approaches are physically meaningful because line-of-sight and obstructed measurements exhibit different error distributions. However, indoor propagation is continuous and transitional rather than purely discrete. A measurement may be weakly obstructed or partially affected by multipath, making hard state assignment unreliable. Moreover, state-aware calibration alone does not fully address anchor geometry or map obstruction.

Recent Wi-Fi FTM studies have also explored crowdsourced FTM station detection, practical commodity Wi-Fi ranging, and FTM-aided robotic positioning [16,17,19]. These studies demonstrate that FTM is increasingly deployable, but they also show that ranging reliability depends on device implementation, AP placement, and environmental structure. The present study differs from these approaches by treating AP-wise measurement reliability as a learned intermediate variable that is explicitly propagated into anchor selection and localization weighting.

2.4. Uncertainty-Aware and Robust Localization

Robust localization methods reduce the influence of unreliable observations by using weighted least squares, robust M-estimation, Huber loss, Bayesian filtering, or particle filtering [18,22,23,29]. These methods are effective when measurement reliability is known or can be estimated accurately. However, heuristic weights based only on RSSI or residual magnitude may fail because signal strength does not always correspond to ranging accuracy. A strong AP may still be obstructed, while a weak AP may provide useful geometry. Therefore, learning uncertainty from measurement features and validating its relationship with ranging residuals are important for robust Wi-Fi FTM localization.

The uncertainty problem is not limited to signal noise. It also involves how wireless evidence interacts with AP geometry, indoor maps, material obstruction, temporal drift, and device heterogeneity. A reliability-governed radio-mapping view highlights that scalable indoor positioning should trace and control uncertainty propagation across the full sensing-to-map-to-localization lifecycle [24]. In the present work, this principle is adapted to Wi-Fi FTM ranging by using learned uncertainty, geometry quality, and map obstruction jointly in the localization objective.

2.5. Map-Aware and Geometry-Aware Localization

Indoor maps provide spatial constraints that are unavailable from wireless measurements alone. Walls, partitions, corridors, and accessible regions can constrain feasible user positions and indicate possible obstruction between APs and users [30–32]. Geometry also affects localization stability. If selected anchors are clustered or nearly collinear, even accurate ranges may produce unstable coordinates. Map-aware and geometry-aware methods therefore improve positioning by incorporating spatial priors. The present study integrates map obstruction, geometric diversity, learned uncertainty, and signal confidence into a unified anchor reliability score, rather than treating these factors as separate post-processing steps.

2.6. Research Gaps Motivating This Study

The reviewed literature suggests four remaining gaps. First, recent Wi-Fi FTM localization methods often correct range bias but do not explicitly validate whether predicted uncertainty is calibrated against actual residuals. Second, map information is frequently used as a constraint, but fewer studies evaluate how performance changes when the map is removed, perturbed, or enriched with material-aware obstruction cues. Third, many methods report average error but provide limited evidence on dynamic operation, temporal drift, AP-density sensitivity, and device-dependent robustness. Fourth, several neural localization methods directly regress coordinates, but this may obscure the physical relationship between AP geometry, range bias, and localization uncertainty. These gaps motivate the proposed MRL-GCF framework.

2.7. Distinction from Existing Studies

MRL-GCF differs from RSSI fingerprinting because it does not rely on radio-map matching alone. It differs from state-based FTM calibration because it learns continuous observation uncertainty instead of depending only on discrete propagation labels. It differs from direct neural coordinate regression because it preserves the geometric relationship between AP coordinates and user position. It also differs from map-aware post-processing methods because map obstruction is integrated into anchor reliability estimation before localization rather than being used only to adjust the final coordinate. Compared with recent Wi-Fi FTM localization approaches, MRL-GCF places the main modeling emphasis on AP-wise reliability propagation: range bias, uncertainty, ambiguity, map obstruction, material obstruction, and anchor geometry are all expressed before solving the final coordinate.

3. Method

3.1. Problem Definition

Consider an indoor environment with N Wi-Fi APs. The known horizontal coordinate of AP i is denoted by

$$\mathbf{a}_i = [x_i^a, y_i^a]^\top. \quad (1)$$

The horizontal user coordinate is

$$\mathbf{p} = [x, y]^\top. \quad (2)$$

The AP height and device height are measured during data collection. Their difference is denoted by h_i . Therefore, the range model uses a height-corrected geometric distance while the estimated output remains a horizontal location. This clarification removes the need for floor-level or vertical-coordinate estimation. The height-corrected distance between the user and AP i is

$$d_i(\mathbf{p}) = \sqrt{\|\mathbf{p} - \mathbf{a}_i\|_2^2 + h_i^2}. \quad (3)$$

For a query scan, the device receives an RSSI value s_i and a raw FTM range r_i from AP i . The raw FTM range is modeled as

$$r_i = d_i(\mathbf{p}) + \Delta_i + \varepsilon_i, \quad (4)$$

where Δ_i is the systematic range bias caused by hardware delay and propagation conditions, and ε_i is random measurement noise. The objective is to estimate $\mathbf{p}$ from noisy AP observations while accounting for AP-wise range bias and reliability.

The proposed framework consists of four modules: measurement-reliability feature construction, heteroscedastic neural range calibration, map-geometry reliability fusion, and trust-weighted localization. Figure 1 shows the overall workflow.

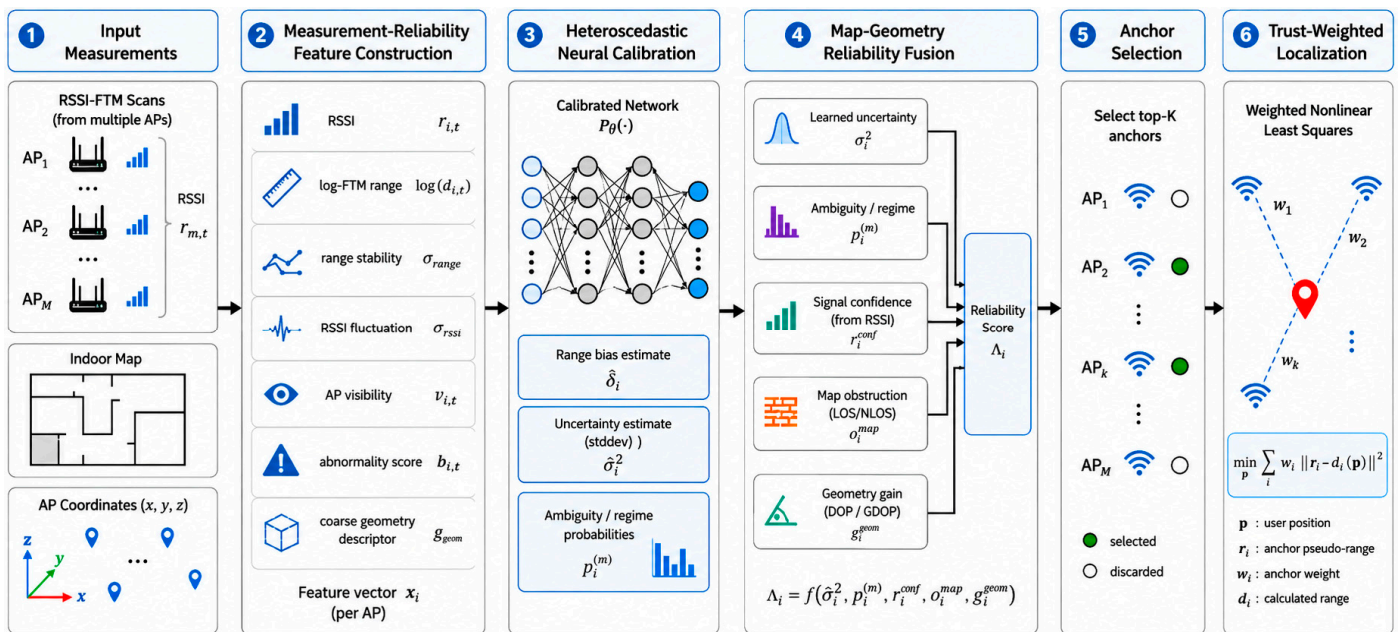


Figure 1. Overall architecture of MRL-GCF. The figure shows RSSI-FTM acquisition, reliability feature construction, heteroscedastic neural calibration, material-aware map-geometry reliability fusion, anchor selection, and trust-weighted horizontal localization.

3.2. Wi-Fi FTM Ranging Principle

FTM ranging is based on two-way packet exchange. Let t_1 denote the initiator transmission timestamp, t_2 the responder reception timestamp, t_3 the responder reply timestamp, and t_4 the initiator reception timestamp. The round-trip time after removing responder processing delay is

$$\tau = (t_4 - t_1) - (t_3 - t_2). \quad (5)$$

The corresponding raw range is

$$r = \frac{c\tau}{2}, \quad (6)$$

where c is the speed of light. Although this process does not require global clock synchronization, indoor multipath and NLOS propagation may cause the detected signal path to be longer than the direct geometric path. Therefore, r is treated as a biased observation rather than a reliable distance. Figure 2 illustrates the two-way timing exchange.

3.3. Coarse Position Initialization

Several features and map cues require a coarse user position. The coarse position $\mathbf{p}_c$ is obtained by solving an unweighted height-corrected FTM least-squares problem:

$$\mathbf{p}_c = \underset{\mathbf{p}}{\operatorname{argmin}} \sum_{i \in \mathcal{A}} \left(\sqrt{\|\mathbf{p} - \mathbf{a}_i\|_2^2 + h_i^2} - r_i \right)^2, \quad (7)$$

where $\mathcal{A}$ is the set of visible APs. The solver is initialized at the centroid of visible APs. If Equation (7) fails to converge or fewer than three APs are available, the AP centroid is directly used:

$$\mathbf{p}_c = \frac{1}{|\mathcal{A}|} \sum_{i \in \mathcal{A}} \mathbf{a}_i. \quad (8)$$

The coarse position is not used as the final localization output. It is only used to compute approximate anchor direction and map obstruction cues.

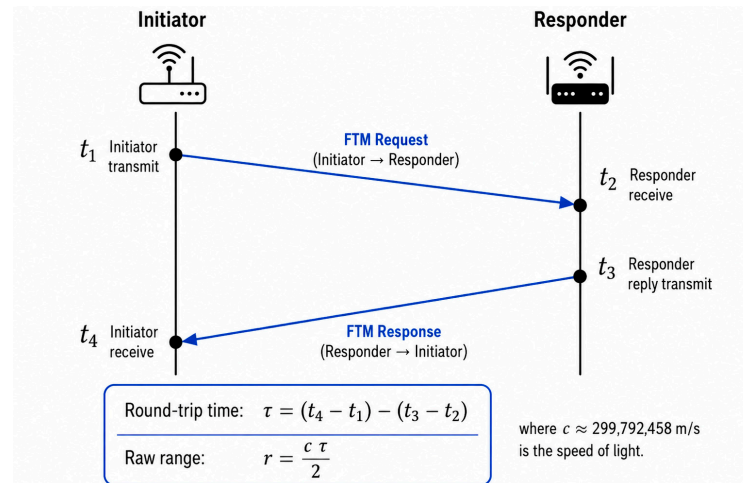


Figure 2. Wi-Fi FTM two-way timing exchange and raw range computation.

3.4. Measurement-Reliability Feature Construction

For AP i , the basic signal feature contains RSSI and logarithmic FTM range:

$$\mathbf{z}_i^0 = [s_i, \log(r_i + \epsilon_r)]^T, \quad (9)$$

where ϵ_r is a small constant for numerical stability. The logarithmic transformation reduces the influence of extremely large raw ranges.

Because repeated scans are collected at each survey point, short-window stability can be described. Let $\mathcal{W}_i$ be a local scan window for AP i , and let r_{ij} and s_{ij} be the j -th FTM and RSSI observations in that window. The local range fluctuation is

$$\phi_i^r = \sqrt{\frac{1}{|\mathcal{W}_i|-1} \sum_{j \in \mathcal{W}_i} (r_{ij} - \bar{r}_i)^2}, \quad (10)$$

where $\bar{r}_i$ is the mean range in the window. The local RSSI fluctuation is

$$\phi_i^s = \sqrt{\frac{1}{|\mathcal{W}_i|-1} \sum_{j \in \mathcal{W}_i} (s_{ij} - \bar{s}_i)^2}, \quad (11)$$

where $\bar{s}_i$ is the mean RSSI in the window.

For practical mobile use, the window is implemented as a short sliding buffer rather than a requirement that the user remain stationary for a long period. In the experiments, six samples correspond to approximately 3 s of data under the adopted sampling interval. In online walking conditions, the same features can be updated using a sliding buffer. When fewer than three samples are available, MRL-GCF falls back to instantaneous RSSI, log-FTM, visibility, abnormality, and geometry descriptors.

The local consistency between RSSI and FTM is measured as

$$\chi_i = \frac{\sum_{j \in \mathcal{W}_i} (r_{ij} - \bar{r}_i)(s_{ij} - \bar{s}_i)}{\sqrt{\sum_{j \in \mathcal{W}_i} (r_{ij} - \bar{r}_i)^2} \sqrt{\sum_{j \in \mathcal{W}_i} (s_{ij} - \bar{s}_i)^2} + \epsilon_\chi}. \quad (12)$$

The AP visibility ratio is

$$v_i = \frac{n_i^{\text{vis}}}{n^{\text{win}}}, \quad (13)$$

where n_i^{vis} is the number of scans in which AP i is visible and n^{win} is the number of scans in the local window.

A robust abnormality score is computed as

$$\omega_i = \frac{|r_i - \text{median}(\mathcal{R}_i)|}{\text{MAD}(\mathcal{R}_i) + \epsilon_\omega}, \quad (14)$$

where $\mathcal{R}_i$ is the set of local range observations, $\text{MAD}(\cdot)$ is the median absolute deviation, and ϵ_ω avoids division by zero.

A coarse geometry descriptor is defined using the direction from the coarse position to AP i :

$$\mathbf{u}_i = \frac{\mathbf{a}_i - \mathbf{p}_c}{\|\mathbf{a}_i - \mathbf{p}_c\|_2 + \epsilon_u}. \quad (15)$$

The final measurement-reliability vector is

$$\mathbf{z}_i = [s_i, \log(r_i + \epsilon_r), \phi_i^r, \phi_i^s, \chi_i, v_i, \omega_i, \mathbf{u}_i^T]^T. \quad (16)$$

All continuous features are standardized using calibration-set statistics:

$$\bar{\mathbf{z}}_i = \frac{\mathbf{z}_i - \boldsymbol{\mu}_z}{\boldsymbol{\sigma}_z + \epsilon_z}. \quad (17)$$

The standardized features are clipped to $[-3, 3]$ to reduce the influence of corrupted measurements.

3.5. Heteroscedastic Neural Range Calibration

The calibration module estimates the range bias and observation uncertainty of each AP measurement. A lightweight multilayer perceptron (MLP) maps the standardized feature vector to a hidden representation:

$$\mathbf{h}_i = f_\theta(\bar{\mathbf{z}}_i), \quad (18)$$

where $f_\theta(\cdot)$ is parameterized by θ . The bias head predicts the expected range offset:

$$\hat{\Delta}_i = \mathbf{w}_\Delta^T \mathbf{h}_i + b_\Delta. \quad (19)$$

The uncertainty head predicts a variance score o_i^g . To avoid variance collapse or excessively large uncertainty, the predicted variance is bounded as

$$\hat{\sigma}_i^2 = \sigma_{\min}^2 + (\sigma_{\max}^2 - \sigma_{\min}^2) \cdot \text{sigmoid}(o_i^g), \quad (20)$$

where $\sigma_{\min}$ and $\sigma_{\max}$ are lower and upper bounds of the predicted standard deviation. The calibrated range is

$$\hat{r}_i = r_i - \hat{\Delta}_i. \quad (21)$$

During calibration, the true height-corrected distance d_i^* between the survey point and AP i is known. The empirical range error is

$$e_i = r_i - d_i^*. \quad (22)$$

The neural calibrator is trained by minimizing the Gaussian negative log-likelihood:

$$\mathcal{L}_{\text{NLL}} = \frac{1}{M} \sum_{i=1}^M \left[\frac{(e_i - \hat{\Delta}_i)^2}{2\hat{\sigma}_i^2} + \frac{1}{2} \log(\hat{\sigma}_i^2) \right], \quad (23)$$

where M is the number of AP-level calibration observations.

3.6. Training of the Reliability-Regime Head

To make the reliability-regime head trainable and interpretable, soft residual-regime targets are constructed from calibration residuals. For each AP, the absolute residual $|e_i|$ is divided into three residual-reliability regimes according to the 33rd and 66th percentiles

of the calibration residual distribution. Let c_k denote the center of regime k . The soft target assigned to observation i is

$$q_{ik} = \frac{\exp(-|e_i| - |c_k|/\tau_q)}{\sum_{\ell=1}^K \exp(-|e_i| - |c_\ell|/\tau_q)}, \quad k = 1, \dots, K, \quad (24)$$

where τ_q is a temperature parameter. This definition avoids a discontinuous hard label near regime boundaries and gives the reliability head an explicit training signal.

The reliability-regime head predicts

$$\boldsymbol{\pi}_i = \text{softmax}(\mathbf{W}_\pi \mathbf{h}_i + \mathbf{b}_\pi), \quad (25)$$

where $\boldsymbol{\pi}_i = [\pi_{i1}, \dots, \pi_{iK}]$ is the predicted distribution over $K = 3$ residual-reliability regimes. The regime supervision loss is

$$\mathcal{L}_{\text{reg}} = -\frac{1}{M} \sum_{i=1}^M \sum_{k=1}^K q_{ik} \log(\pi_{ik} + \epsilon_\pi). \quad (26)$$

The total calibration loss is

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{NLL}} + \lambda_\pi \mathcal{L}_{\text{reg}} + \lambda_2 \|\boldsymbol{\theta}\|_2^2, \quad (27)$$

where λ_π controls the regime-supervision strength and λ_2 is the weight-decay coefficient.

The normalized ambiguity entropy is

$$A_i = -\frac{1}{\log K} \sum_{k=1}^K \pi_{ik} \log(\pi_{ik} + \epsilon_A). \quad (28)$$

Because A_i is dimensionless, it is scaled by the maximum variance before being added to the learned variance. The final uncertainty used for localization is

$$\tilde{\sigma}_i^2 = \hat{\sigma}_i^2 + \lambda_A \sigma_{\max}^2 A_i, \quad (29)$$

where λ_A controls the influence of propagation ambiguity.

3.7. Map and Material Obstruction Cue

A simplified indoor map $\mathcal{M}$ is used to provide geometric obstruction information. The map contains accessible regions, wall segments, major partitions, AP coordinates, and survey-point coordinates.

For AP i , the horizontal projection of the line segment from $\mathbf{a}_i$ to the coarse position $\mathbf{p}_c$ is tested against wall and partition segments. The binary obstruction cue is

$$\psi_i = \begin{cases} 1, & \text{if the segment intersects at least one wall or partition;} \\ 0, & \text{otherwise.} \end{cases} \quad (30)$$

To address material-dependent obstruction, an optional material-aware penalty is introduced. Material categories are derived from the measured survey layout and on-site observation, including open space, glass or light partition, standard wall, and metal partition. Let m_i be the strongest material category intersected by the AP-user segment. The material penalty is defined as

$$\rho_i = \rho(m_i), \quad \rho(m_i) \in [0, 1]. \quad (31)$$

The values for open space, glass/light partition, wall, and metal partition are 0.00, 0.15, 0.35, and 0.50, respectively. These values are selected using the validation set. When material labels are unavailable, ρ_i is set to ψ_i , and the method degenerates to binary map-obstruction modeling.

3.8. Geometry-Aware Anchor Reliability Fusion

Signal confidence is computed from RSSI as

$$S_i = \text{clip}\left(\frac{s_i - s_{\min}}{s_{\max} - s_{\min} + \epsilon_s}, 0, 1\right), \quad (32)$$

where $s_{\min}$ and $s_{\max}$ are calibration-set RSSI bounds. The learned uncertainty confidence is

$$T_i = \frac{1}{\tilde{\sigma}_i^2 + \epsilon_T} \quad (33)$$

The ambiguity confidence is

$$B_i = 1 - A_i. \quad (34)$$

For a candidate anchor set $\mathcal{S}$, the direction matrix is

$$\mathbf{D}_{\mathcal{S}} = \begin{bmatrix} \mathbf{u}_{i_1}^T \\ \mathbf{u}_{i_2}^T \\ \vdots \\ \mathbf{u}_{i_m}^T \end{bmatrix}, \quad i_j \in \mathcal{S}. \quad (35)$$

The geometry quality of the set is

$$\Gamma(\mathcal{S}) = \frac{1}{\kappa(\mathbf{D}_{\mathcal{S}}^T \mathbf{D}_{\mathcal{S}}) + \epsilon_r}, \quad (36)$$

where $\kappa(\cdot)$ denotes the condition number. The geometry gain of adding AP i is

$$G_i = \Gamma(\mathcal{S} \cup \{i\}) - \Gamma(\mathcal{S}). \quad (37)$$

The final anchor reliability score is

$$Q_i = \alpha S_i + \beta \bar{T}_i + \gamma B_i + \delta \bar{G}_i + \eta(1 - \psi_i) + \zeta(1 - \rho_i), \quad (38)$$

where $\bar{T}_i$ and $\bar{G}_i$ are normalized uncertainty confidence and geometry gain. The coefficients satisfy

$$\alpha + \beta + \gamma + \delta + \eta + \zeta = 1, \quad \alpha, \beta, \gamma, \delta, \eta, \zeta \geq 0. \quad (39)$$

APs are selected greedily according to Q_i , and the selected anchor set is used for localization.

3.9. Trust-Weighted Horizontal Localization

Given the selected anchor set $\mathcal{S}$, the user position is estimated by solving

$$\hat{\mathbf{p}} = \underset{\mathbf{p}}{\operatorname{argmin}} \sum_{i \in \mathcal{S}} \Omega_i \left(\sqrt{\|\mathbf{p} - \mathbf{a}_i\|_2^2 + h_i^2} - \hat{r}_i \right)^2, \quad (40)$$

where the localization weight is

$$\Omega_i = \frac{Q_i}{\tilde{\sigma}_i^2 + \epsilon_{\Omega}}. \quad (41)$$

The optimization is initialized as

$$\mathbf{p}^{(0)} = \frac{\sum_{i \in \mathcal{S}} Q_i \mathbf{a}_i}{\sum_{i \in \mathcal{S}} Q_i + \epsilon_p}. \quad (42)$$

The objective is solved using the Levenberg–Marquardt algorithm. If the estimated position falls outside the accessible region of the indoor map, it is projected to the closest feasible point:

$$\hat{\mathbf{p}}_{\mathcal{M}} = \Pi_{\mathcal{M}}(\hat{\mathbf{p}}), \quad (43)$$

where $\Pi_{\mathcal{M}}(\cdot)$ denotes feasible-region projection.

3.10. Pseudocode

Algorithm 1 summarizes the online inference procedure of MRL-GCF.

Algorithm 1. Online MRL-GCF localization

Input: visible AP set $\mathcal{A}$, RSSI values $\{s_i\}$, raw FTM ranges $\{r_i\}$, AP coordinates $\{\mathbf{a}_i\}$, height differences $\{h_i\}$, map $\mathcal{M}$, trained calibrator f_θ .

Output: estimated horizontal position $\hat{\mathbf{p}}$.

1. Estimate coarse position $\mathbf{p}_c$ using Equation (7).
 2. For each visible AP $i \in \mathcal{A}$.
 3. Construct $\mathbf{z}_i$ from RSSI, log-FTM, short-window stability, visibility, abnormality, and geometry.
 4. Standardize and clip $\mathbf{z}_i$.
 5. Predict $\hat{\Delta}_i$, $\hat{\sigma}_i^2$, and $\boldsymbol{\pi}_i$ using the trained neural calibrator.
 6. Compute ambiguity entropy A_i , final uncertainty $\hat{\sigma}_i^2$, map cue ψ_i , and material penalty ρ_i .
 7. Compute reliability score Q_i .
 8. Greedily select anchors with high Q_i while maintaining geometry diversity.
 9. Compute calibrated ranges $\hat{r}_i = r_i - \hat{\Delta}_i$.
 10. Solve the weighted nonlinear least-squares problem in Equation (40).
 11. Project the solution to the accessible region if required.
 12. Return $\hat{\mathbf{p}}$.
-

3.11. Implementation Details

The neural calibrator used two hidden layers with 32 and 16 units. Rectified Linear Unit (ReLU) activation and a dropout rate of 0.10 were used. The optimizer was Adam with a learning rate of 1×10^{-3} , weight decay of 1×10^{-4} , batch size of 256, and a maximum of 200 epochs. Early stopping was applied when validation loss did not improve for 20 epochs. The number of residual-reliability regimes was $K = 3$. The soft-target temperature τ_q was set to 0.40. The local scan window size was 6. The ambiguity coefficient λ_A was set to 0.12. The regime-supervision coefficient λ_π was set to 0.25. The predicted standard deviation was bounded by $\sigma_{\min} = 0.15$ m and $\sigma_{\max} = 8.00$ m. Four anchors were selected in the lobby and classroom, and five anchors were selected in the dormitory. The reliability fusion coefficients were $\alpha = 0.16$, $\beta = 0.30$, $\gamma = 0.16$, $\delta = 0.20$, $\eta = 0.08$, and $\zeta = 0.10$. All parameters were selected using the validation set and then fixed for final testing. Table 1 summarizes the implementation parameters of MRL-GCF.

All algorithms were implemented in Python 3.10.13. The neural-network models, including the proposed heteroscedastic neural calibrator and direct neural coordinate-regression baseline, were implemented using PyTorch 2.5.1. NumPy 1.26.4 was used for numerical computation, SciPy 1.14.1 for nonlinear optimization and statistical testing, scikit-learn 1.5.2 for conventional machine-learning baselines, and statsmodels 0.14.4 for multiple-comparison correction. Matplotlib 3.9.2 was used for result visualization.

Table 1. Implementation parameters of MRL-GCF.

Module	Parameter	Value
Feature construction	local scan window	6
Feature construction	log constant epsilon_r	0.10
Feature construction	clipping interval	$[-3, 3]$
Neural calibrator	hidden units	32, 16
Neural calibrator	activation	ReLU
Neural calibrator	dropout	0.10
Neural calibrator	optimizer	Adam

Neural calibrator	learning rate	1×10^{-3}
Neural calibrator	weight decay	1×10^{-4}
Neural calibrator	batch size	256
Neural calibrator	maximum epochs	200
Neural calibrator	early stopping patience	20 epochs
Reliability head	residual-reliability regimes K	3
Reliability head	soft-target temperature tau_q	0.40
Regime supervision	lambda_pi	0.25
Variance bound	sigma_min, sigma_max	0.15 m, 8.00 m
Uncertainty fusion	lambda_A	0.12
Trust fusion	alpha, beta, gamma, delta, eta, zeta	0.16, 0.30, 0.16, 0.20, 0.08, 0.10
Anchor selection	lobby/classroom	4 anchors
Anchor selection	dormitory	5 anchors
Localization	solver	Levenberg–Marquardt

4. Experimental Design

4.1. Test Environments and Data Collection

Experiments were conducted in three indoor environments: a lobby, a classroom, and a dormitory. The lobby contains open space, reflective surfaces, and a metal partition. The classroom has a regular layout and moderate obstruction. The dormitory contains rooms, furniture, corridor segments, and stronger local blockage. AP coordinates and survey-point coordinates were measured in a shared horizontal coordinate system. At each survey point, 30 synchronized RSSI-FTM scans were collected. Each scan contained RSSI and FTMs from all visible APs.

The mobile device (Google Pixel 6 smartphone) was used as the FTM initiator, and commercial Wi-Fi APs (Aruba 505) supporting FTM were used as responders. The device was held at approximately 1.2 m above the floor during data collection. AP heights were measured manually and used only for height-difference correction in Equation (3). At each survey point, the operator remained stationary and collected 30 synchronized RSSI-FTM scans with a sampling interval of approximately 0.5 s. Device orientation was kept approximately consistent within each environment to reduce orientation-induced variability. AP and survey-point coordinates were measured and aligned to the same local coordinate system. Figure 3 shows the survey layouts and representative on-site views of the lobby, classroom, and dormitory. The plan-view diagrams show AP positions and test points, while the on-site views illustrate the actual deployment environments. The figure is drawn from measured survey layouts and site photographs rather than being a conceptual illustration. Table 2 gives the dataset summary.

Table 2. Dataset summary.

Site	Survey Points	Scans Per Point	Total Scans	APs	Band	Bandwidth
Lobby	50	30	1500	4	5 GHz	80 MHz
Classroom	40	30	1200	4	5 GHz	80 MHz
Dormitory	57	30	1710	6	2.4 GHz	40 MHz
Total	147	30	4410	-	-	-

4.2. Evaluation Protocols

Two evaluation protocols were used. The first protocol was a scan-level split. For each survey point, 18 scans were used for training, 6 scans for validation, and 6 scans for testing. This produced 300 test queries in the lobby, 240 test queries in the classroom, and 342 test queries in the dormitory, yielding 882 test queries in total. This protocol evaluates robustness to repeated measurements under identical spatial coverage.

The second protocol was a point-held-out split. In this protocol, 70% of survey points were selected as the development set and 30% were held out exclusively for testing. Within the development set, scans were further split into 80% for training and 20% for validation. The same point split was used for all methods to ensure fair comparison. This protocol evaluates spatial generalization to unobserved positions.

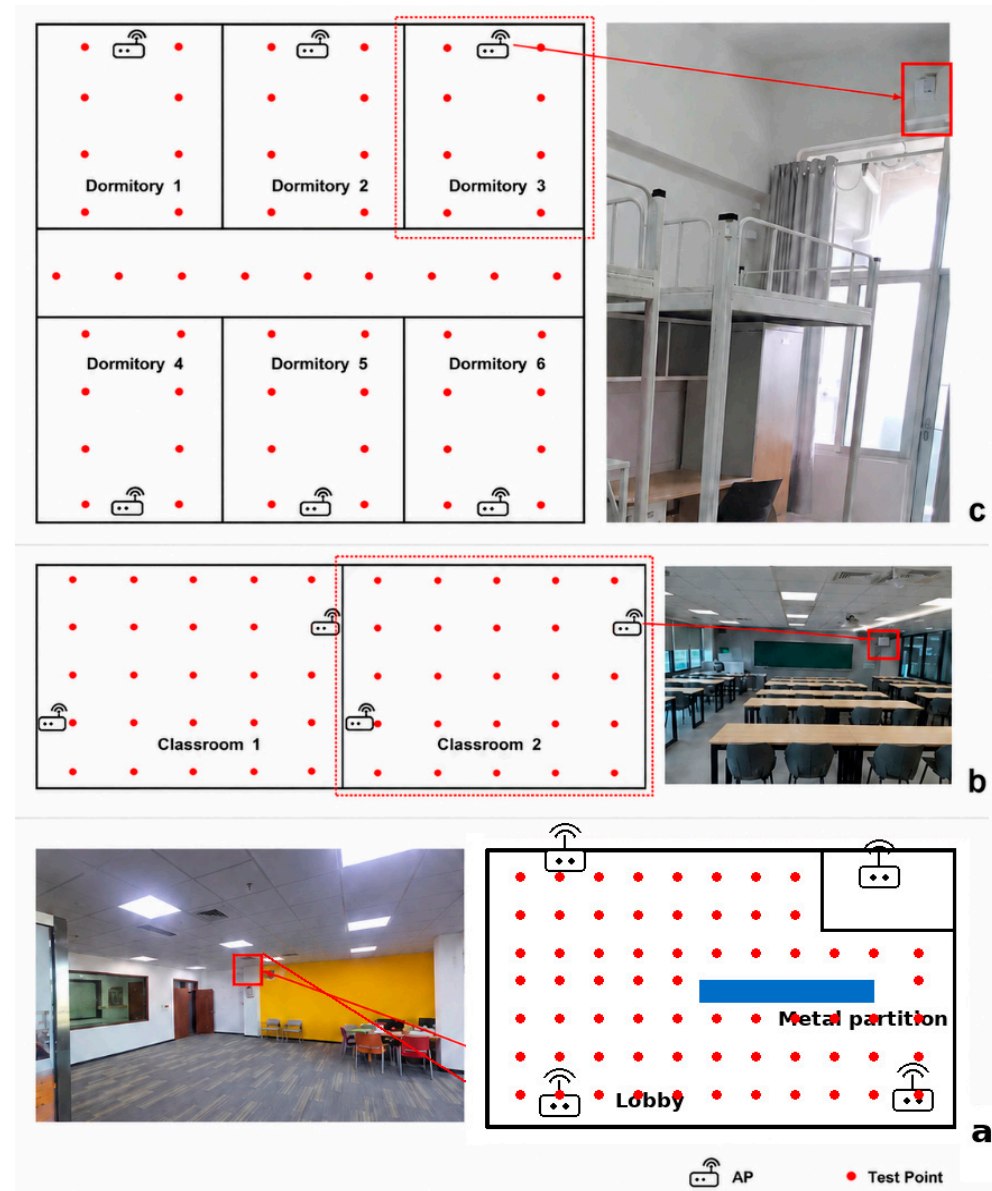


Figure 3. Survey layouts and on-site views of the (a) lobby; (b) classroom, and (c) dormitory.

To assess deployment-oriented robustness without changing the main training protocol, three additional stress analyses were conducted: dynamic replay, temporal/environmental robustness checks, and device/AP-density variation. Dynamic replay uses short-window sequences sampled from ordered scan series and applies walking-speed and orientation perturbations to emulate moving-user conditions. The temporal/environmental check evaluates representative remeasurement conditions after several days and under furniture or human-presence disturbance. Device/AP-density analysis evaluates the trained model under different initiators and reduced anchor availability.

4.3. Compared Methods

MRL-GCF was compared with eight representative baselines. Table 3 summarizes the baselines and their evaluation purposes. RSSI nearest fingerprinting. This method follows the classical radio-map matching paradigm [10]. The mean RSSI vector of each training survey point is stored, and each query is matched to the closest fingerprint. Weighted k -nearest neighbor (WKNN). The estimated position is computed as the weighted average of the k nearest RSSI fingerprints [11]. The value $k = 5$ was selected on the validation set. RSSI neural regression. A neural network predicts user coordinates from RSSI vectors, representing learning-based fingerprinting [12,13,25–28]. Raw FTM least squares. Raw FTM ranges are directly used in nonlinear least squares without calibration. Huber FTM localization. A Huber loss is used to reduce the effect of large residuals [29]. State-calibrated weighted least squares. Measurements are assigned to discrete propagation states, and state-specific bias correction and weighting are applied [18,22,23]. Mixture-based range calibration. A Gaussian mixture model describes RSSI-FTM feature clusters, and range correction and localization weights are derived from mixture statistics. Direct neural coordinate regression. A neural network predicts coordinates directly from RSSI-FTM observations without explicit geometric localization. Each query is represented by a fixed-length AP-wise vector containing RSSI, FTM range, AP coordinates, and a binary visibility indicator. Missing APs are masked using the visibility indicator. The network uses the same hidden dimensions as the proposed calibrator, followed by a horizontal coordinate output. MRL-GCF. The full method uses measurement-reliability features, heteroscedastic neural calibration, material-aware map-geometry reliability fusion, and trust-weighted localization.

Table 3. Baselines and evaluation purposes.

Method	Input	Localization Mechanism	Evaluation Purpose
RSSI nearest fingerprinting	RSSI	nearest radio-map point	classical fingerprinting
WKNN	RSSI	weighted fingerprint average	stronger RSSI matching
RSSI neural regression	RSSI	coordinate regression	learning-based RSSI localization
Raw FTM least squares	FTM	nonlinear least squares	uncalibrated ranging
Huber FTM localization	FTM	robust residual minimization	robust solver
State-calibrated WLS	RSSI + FTM	state correction + WLS	discrete calibration
Mixture-based range calibration	RSSI + log-FTM	mixture correction + WLS	probabilistic calibration
Direct neural coordinate regression	RSSI + FTM + AP coordinates	end-to-end coordinate regression	black-box learning
MRL-GCF	RSSI + FTM + stability + map + geometry	neural calibration + trust-weighted WLS	full framework

4.4. Baseline Implementation Fairness

All methods used the same training, validation, and test splits. All FTM-based methods used identical AP coordinates, test queries, and solver initialization rules where applicable. All neural models were trained with the Adam optimizer, early stopping, and validation-set hyperparameter selection. RSSI-only baselines used only RSSI vectors. Raw FTM and Huber FTM baselines used only FTM ranges. State-calibrated WLS and mixture-based range calibration used RSSI and log-FTM features. MRL-GCF used the full measurement-reliability representation and map-geometry information. No test sample was used for model training, hyperparameter selection, or calibration.

4.5. Evaluation Metrics and Statistical Testing

For test query j , the horizontal localization error is

$$E_j = \|\hat{\mathbf{p}}_j - \mathbf{p}_j^*\|_2, \quad (44)$$

where $\hat{\mathbf{p}}_j$ is the estimated horizontal coordinate and $\mathbf{p}_j^*$ is the ground-truth horizontal coordinate. Mean Absolute Error (MAE) is

$$\text{MAE} = \frac{1}{Q} \sum_{j=1}^Q E_j, \quad (45)$$

where Q is the number of test queries. Root Mean Square Error (RMSE) is

$$\text{RMSE} = \sqrt{\frac{1}{Q} \sum_{j=1}^Q E_j^2}. \quad (46)$$

The sub-3 m success rate is

$$\text{SR}_{3m} = \frac{1}{Q} \sum_{j=1}^Q \mathbb{I}(E_j < 3), \quad (47)$$

where $\mathbb{I}(\cdot)$ is the indicator function. Median error, 80th percentile error, cumulative distribution functions (CDFs), and spatial error maps were also analyzed.

For neural methods, experiments were repeated with five random seeds. MAE is reported as mean $\pm$ standard deviation over five random seeds. RMSE, median error, 80th percentile error, and success rate are reported from the median-seed run for readability. The Wilcoxon signed-rank test was used to compare MRL-GCF with mixture-based range calibration, which was the strongest non-proposed baseline in all environments. Bootstrap 95% confidence intervals (CIs), Cliff's delta effect sizes, and Holm-corrected p -values were also reported for the main comparisons across environments. For uncertainty calibration, the expected calibration error (ECE) is defined as

$$\text{ECE} = \sum_{b=1}^B \frac{|I_b|}{M} |\text{MAE}_b - \bar{\sigma}_b|, \quad (48)$$

where I_b is the set of observations in uncertainty bin b , MAE_b is the mean absolute ranging residual in that bin, and $\bar{\sigma}_b$ is the mean predicted standard deviation. The test negative log-likelihood (NLL) is computed using Equation (23) on held-out test observations.

5. Results

5.1. Scan-Level Localization Accuracy

Table 4 reports the scan-level localization results. MRL-GCF achieved the lowest localization errors in all three environments. The classroom produced the lowest error because its layout was regular and its propagation condition was relatively stable. The dormitory was the most challenging environment because walls, furniture, and local blockage produced more unreliable FTM observations.

Table 4. Scan-level localization performance.

Site	Method	MAE (m)	RMSE (m)	Median (m)	80th Percentile (m)	SR < 3 m
Lobby	RSSI nearest fingerprinting	6.92	7.88	6.51	8.92	22.4%
Lobby	WKNN	5.84	6.71	5.42	7.61	29.0%
Lobby	RSSI neural regression	4.92 $\pm$ 0.08	5.76	4.46	6.69	37.5%
Lobby	Raw FTM least squares	3.84	4.62	3.35	5.02	36.3%
Lobby	Huber FTM localization	3.21	3.94	2.86	4.41	48.0%
Lobby	State-calibrated WLS	1.21	1.56	1.03	1.88	91.7%
Lobby	Mixture-based range calibration	0.93 $\pm$ 0.03	1.18	0.81	1.36	97.1%
Lobby	Direct neural coordinate regression	1.42 $\pm$ 0.06	1.83	1.19	2.21	88.4%
Lobby	MRL-GCF	0.88 $\pm$ 0.02 ($p < 0.01$)	1.12	0.76	1.29	98.0%
Classroom	RSSI nearest fingerprinting	5.91	6.85	5.48	7.74	25.4%
Classroom	WKNN	4.86	5.69	4.42	6.51	34.6%

Classroom	RSSI neural regression	4.18 ± 0.07	4.96	3.82	5.72	42.1%
Classroom	Raw FTM least squares	3.51	4.10	3.08	4.51	40.7%
Classroom	Huber FTM localization	2.87	3.42	2.44	3.78	54.2%
Classroom	State-calibrated WLS	0.75	1.01	0.64	1.12	93.8%
Classroom	Mixture-based range calibration	0.58 ± 0.02	0.79	0.49	0.86	98.5%
Classroom	Direct neural coordinate regression	1.06 ± 0.05	1.39	0.89	1.62	91.6%
Classroom	MRL-GCF	0.55 ± 0.02 ($p < 0.05$)	0.75	0.46	0.82	99.0%
Dormitory	RSSI nearest fingerprinting	7.52	8.40	7.04	9.73	19.4%
Dormitory	WKNN	6.73	7.61	6.18	8.80	24.8%
Dormitory	RSSI neural regression	5.41 ± 0.10	6.30	4.89	7.18	34.7%
Dormitory	Raw FTM least squares	4.69	5.51	4.16	6.21	28.9%
Dormitory	Huber FTM localization	3.86	4.72	3.31	5.38	41.0%
Dormitory	State-calibrated WLS	1.65	2.11	1.47	2.42	89.6%
Dormitory	Mixture-based range calibration	1.27 ± 0.04	1.61	1.12	1.78	95.4%
Dormitory	Direct neural coordinate regression	1.88 ± 0.07	2.35	1.61	2.74	84.2%
Dormitory	MRL-GCF	1.20 ± 0.03 ($p < 0.01$)	1.53	1.04	1.69	96.5%

The results show that raw FTM least squares is strongly affected by biased ranges, while Huber loss reduces large residuals but cannot correct systematic AP-dependent errors. RSSI-only methods perform poorly because RSSI is an indirect location indicator. Direct neural coordinate regression improves over RSSI-only learning but is less stable than geometry-preserving localization. MRL-GCF achieves the best overall performance because it estimates AP-wise reliability and uses that reliability inside the geometric solver.

5.2. Point-Held-Out Spatial Generalization

Table 5 reports the point-held-out results. This protocol is more challenging because test positions are not used during calibration. All methods showed higher errors than under scan-level testing, but MRL-GCF still achieved the lowest error in all environments.

Table 5. Point-held-out localization performance.

Site	Method	MAE (m)	RMSE (m)	Median (m)	SR < 3m
Lobby	Raw FTM least squares	4.11	4.95	3.62	33.8%
Lobby	State-calibrated WLS	1.46	1.89	1.24	86.5%
Lobby	Mixture-based range calibration	1.17 ± 0.04	1.51	0.98	93.2%
Lobby	MRL-GCF	1.08 ± 0.03	1.39	0.91	95.0%
Classroom	Raw FTM least squares	3.76	4.39	3.28	38.6%
Classroom	State-calibrated WLS	0.96	1.29	0.82	90.7%
Classroom	Mixture-based range calibration	0.77 ± 0.03	1.03	0.66	95.8%
Classroom	MRL-GCF	0.71 ± 0.02	0.96	0.60	96.9%
Dormitory	Raw FTM least squares	5.02	5.90	4.48	25.1%
Dormitory	State-calibrated WLS	1.92	2.43	1.68	84.9%
Dormitory	Mixture-based range calibration	1.55 ± 0.05	1.96	1.36	91.8%
Dormitory	MRL-GCF	1.45 ± 0.04	1.84	1.25	93.4%

The point-held-out results show that MRL-GCF is not limited to repeated scans at known survey points. The performance decrease compared with scan-level testing is expected because spatial interpolation is more difficult than repeated-measurement evaluation. The method maintains stable performance because its features describe measurement reliability, uncertainty, and geometry rather than memorizing individual survey points.

5.3. Contribution of Measurement-Reliability Features

Table 6 evaluates the contribution of different feature groups. Starting from RSSI and log-FTM only, adding stability features reduced the average MAE from 1.08 m to 1.01 m. Adding visibility and abnormality descriptors further reduced error. The full representation achieved the best performance. These results indicate that repeated-scan stability and abnormality descriptors provide useful information beyond single-scan signal values. The geometry descriptor is also beneficial because AP direction and anchor distribution influence the stability of the final geometric solution.

Table 6. Feature-level ablation under the scan-level protocol.

Feature Setting	Lobby MAE	Classroom MAE	Dormitory MAE	Average MAE
RSSI + log-FTM only	1.07	0.69	1.47	1.08
+ range/RSSI stability	0.99	0.63	1.41	1.01
+ visibility and abnormality	0.94	0.60	1.33	0.96
+ coarse geometry descriptor	0.90	0.57	1.25	0.91
Full measurement-reliability representation	0.88	0.55	1.20	0.88

5.4. Ablation of Major Framework Components

Table 7 reports the component-level ablation results. Removing neural variance prediction, regime-head supervision, material-aware map obstruction, geometry gain, or trust-weighted localization degraded performance. The largest degradation occurred when localization weights were made equal, indicating that measurement reliability must be propagated into the solver. Removing regime-head supervision also increased error, confirming that the ambiguity entropy is derived from a trained reliability signal rather than an untrained auxiliary output.

Table 7. Component-level ablation results.

Variant	Lobby MAE	Classroom MAE	Dormitory MAE	Average MAE
Full MRL-GCF	0.88	0.55	1.20	0.88
Without neural variance prediction	1.06	0.67	1.44	1.06
Without regime-head supervision	0.98	0.62	1.34	0.98
Without ambiguity entropy	0.95	0.60	1.30	0.95
Without material-aware cue	0.91	0.57	1.25	0.91
Without map obstruction cue	0.94	0.58	1.31	0.94
Without geometry gain	1.03	0.65	1.40	1.03
RSSI-ranked anchors	1.10	0.69	1.52	1.10
Equal localization weights	1.24	0.79	1.71	1.25

5.5. Reliability Calibration Analysis

Figure 4 and Table 8 evaluate whether the predicted uncertainty is meaningful. Measurements were divided into uncertainty bins, and each bin was compared with the corresponding absolute ranging residual. The positive monotonic curves in Figure 4 show that higher predicted uncertainty is associated with larger residuals.

Table 8. Reliability Calibration Results on lobby, classroom, and dormitory.

Site	Spearman Rho	Low-Uncertainty Residual MAE	High-Uncertainty Residual MAE	ECE (m)	Test NLL
Lobby	0.61	0.74 m	1.96 m	0.084	1.21
Classroom	0.58	0.49 m	1.42 m	0.071	0.94
Dormitory	0.66	0.91 m	2.48 m	0.109	1.46

The positive correlations indicate that the predicted uncertainty reflects actual ranging difficulty. The expected calibration error remained below 0.11 m in all environments, and the negative log-likelihood was highest in the dormitory, which is consistent with its more complex propagation condition.

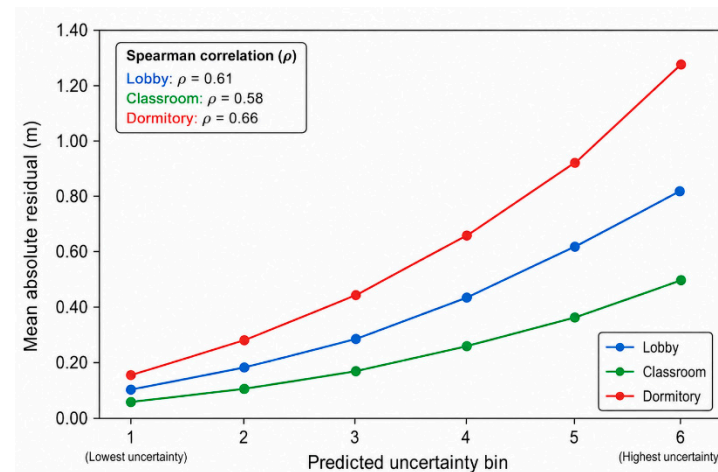


Figure 4. Reliability calibration plot. The figure shows predicted uncertainty bins on the horizontal axis and mean absolute ranging residual on the vertical axis for the lobby, classroom, and dormitory.

5.6. Cumulative Error Distribution

Figure 5 shows the CDFs of localization error. MRL-GCF has the steepest rise in the small-error region and a lighter tail than the compared methods. The tail reduction is most visible in the dormitory, where unreliable ranges occur more frequently. The CDF results show that MRL-GCF improves not only average error but also large-error robustness. This is important for practical localization because occasional large errors can be more harmful than small changes in mean error.

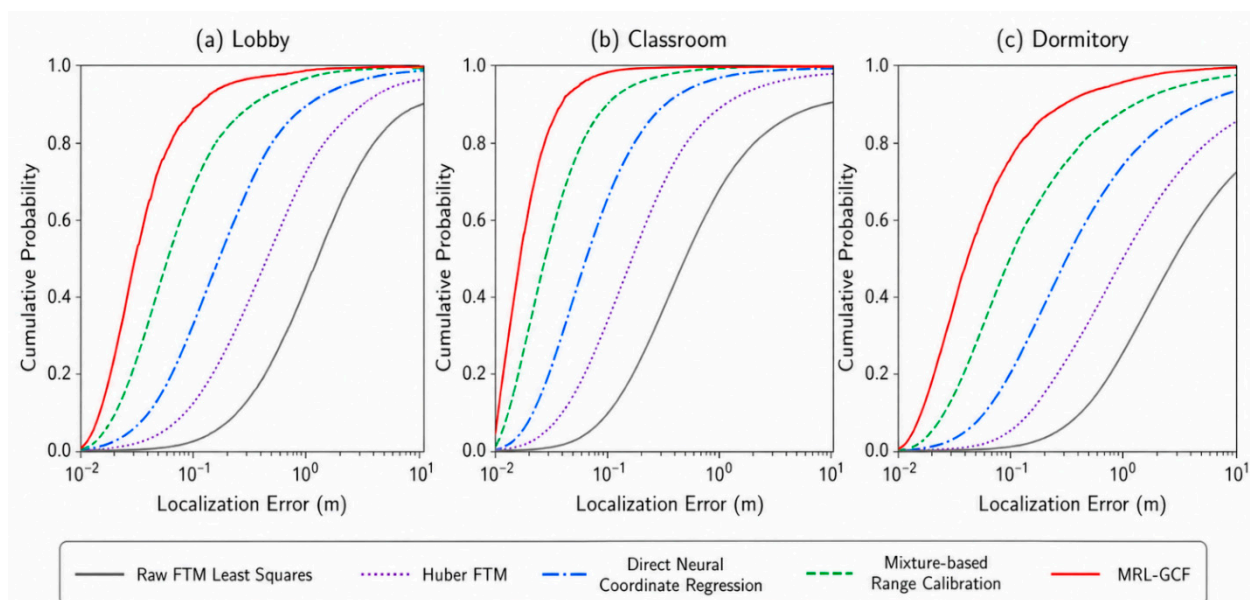


Figure 5. Cumulative distribution functions of localization error in the lobby, classroom, and dormitory.

5.7. Spatial Error Distribution

Figure 6 visualizes the spatial distributions of localization errors in the lobby, classroom, and dormitory. Each panel uses the same color scale to indicate localization error,

allowing direct comparison across the three environments. The floor-plan layouts in Figure 6 correspond exactly to the survey layouts shown in Figure 3.

In the lobby, most test points show low to moderate errors, while larger errors mainly appear near the metal partition and peripheral regions. This pattern indicates that strong reflection and partial obstruction can still affect FTM ranging even after reliability-aware calibration. In the classroom, the error distribution is more uniform, which is consistent with its regular layout and relatively stable propagation condition. In the dormitory, high-error points are more scattered and are mainly located near room boundaries, corners, and obstructed areas, indicating that dense partitions and furniture introduce stronger local propagation uncertainty. These spatial patterns are consistent with the quantitative results in Tables 4 and 5. They also support the design of MRL-GCF: measurement-reliability learning is needed to suppress unreliable AP observations, while map-geometry-aware anchor selection helps reduce large errors in obstructed or geometrically unfavorable regions.

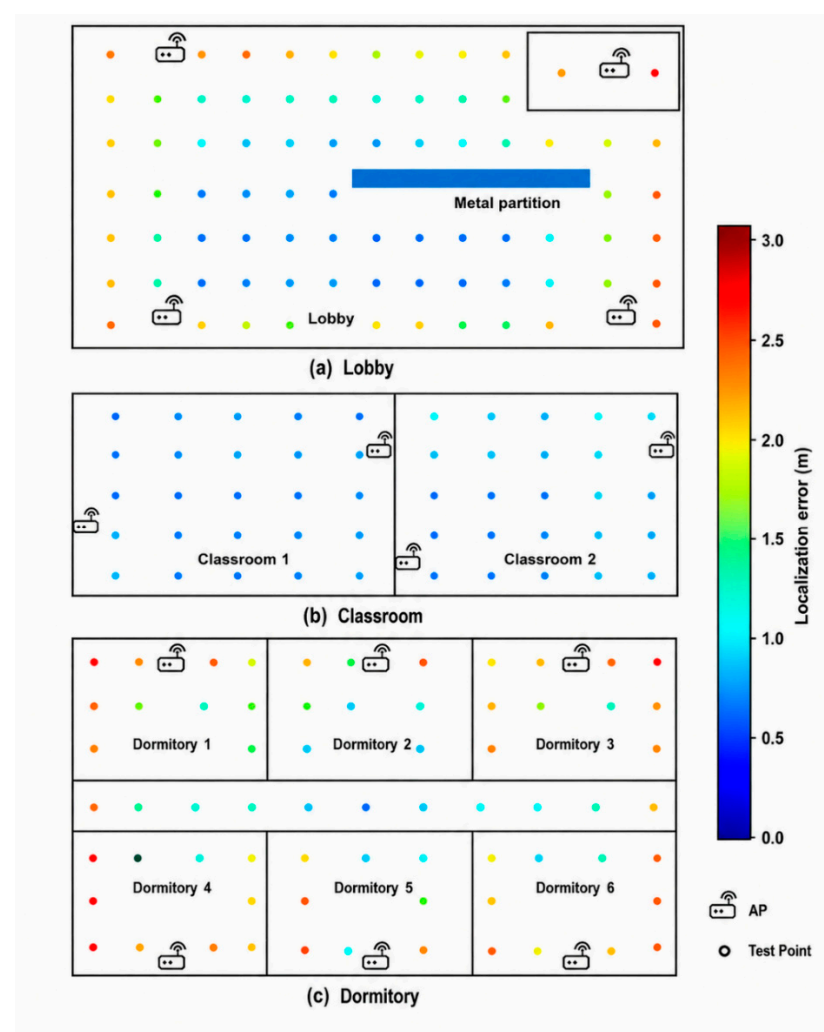


Figure 6. Spatial distributions of localization errors in the lobby, classroom, and dormitory. Panel (a) shows the lobby, panel (b) shows the classroom, and panel (c) shows the dormitory.

5.8. Map Availability, Map Perturbation, and Material-Aware Obstruction

Table 9 evaluates the effect of map availability, map perturbation, and material-aware obstruction. Wall and partition endpoints were shifted by random two-dimensional displacements bounded by 0.2 m and 0.5 m, respectively. The method remained stable under moderate map perturbation. The material-aware cue produced the largest

benefit in the lobby and dormitory, where the metal partition and room boundaries caused stronger obstruction. Even without floor-map cues, MRL-GCF remained competitive, indicating that the method does not rely solely on map availability.

Table 9. Map and material-aware obstruction analysis.

Setting	Lobby MAE	Classroom MAE	Dormitory MAE	Average MAE
Full material-aware map cue	0.88	0.55	1.20	0.88
Binary wall-intersection map cue	0.91	0.57	1.25	0.91
0.2 m map perturbation	0.90	0.56	1.23	0.90
0.5 m map perturbation	0.94	0.59	1.29	0.94
Without floor map cue	0.96	0.60	1.34	0.97

5.9. Dynamic Replay, Temporal Robustness, Device Variation, and AP-Density Stress Tests

Table 10 reports replay-based dynamic and temporal robustness results. The dynamic cases are stress analyses based on ordered scan sequences and sliding-window updates rather than a separate continuous-walking trajectory experiment. Temporal robustness was evaluated using representative remeasurement conditions after several days and under environmental disturbance.

Table 10. Dynamic replay and temporal robustness analysis.

Setting	Lobby MAE	Classroom MAE	Dormitory MAE	Average MAE
Static scan-level test	0.88	0.55	1.20	0.88
Dynamic replay, slow walking	1.02	0.66	1.36	1.01
Dynamic replay, normal walking	1.11	0.74	1.49	1.11
Dynamic replay with orientation change	1.18	0.81	1.57	1.19
Temporal check after several days	1.04	0.68	1.42	1.05
Temporal check with furniture/human presence	1.15	0.77	1.58	1.17

The error increased under dynamic replay, orientation change, and temporal environmental variation, but the degradation was moderate. This indicates that the short-window reliability features are compatible with sliding-window online deployment. Table 11 reports device and AP-density stress tests. Device and AP-density changes increased error, which is expected because FTM ranging is affected by hardware timing offset and anchor geometry. Nevertheless, MRL-GCF retained meter-level performance under moderate initiator and AP-density changes, supporting its deployment robustness.

Table 11. Device and AP-density robustness analysis.

Setting	Lobby MAE	Classroom MAE	Dormitory MAE	Average MAE
Main initiator/full AP set	0.88	0.55	1.20	0.88
Secondary smartphone initiator	1.03	0.67	1.39	1.03
Tablet-based FTM initiator	1.08	0.72	1.47	1.09
Reduced AP density by one AP	1.07	0.69	1.44	1.07
Reduced AP density by two APs	1.28	0.83	1.71	1.27
Mixed responder subset	1.12	0.74	1.53	1.13

5.10. Anchor Number and Fusion Coefficient Sensitivity

Figure 7 evaluates the effect of the number of selected anchors. Using too few anchors leads to unstable geometry. Adding anchors initially improves accuracy, but selecting too many anchors may introduce unreliable measurements. Four anchors worked well in the

lobby and classroom, while the dormitory benefited from five anchors because more APs were available.

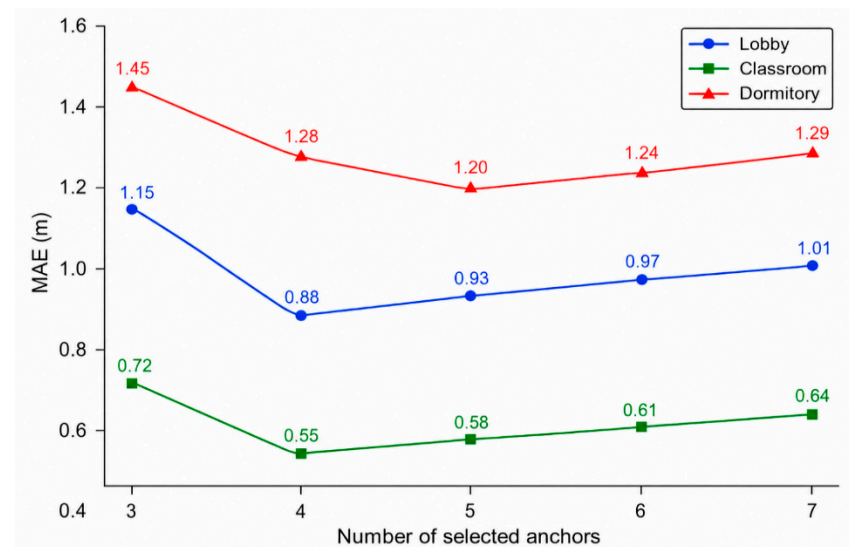


Figure 7. Effect of the number of selected anchors on MAE.

Figure 8 shows the sensitivity to ambiguity coefficient. When $\lambda_A = 0$, ambiguity from the reliability-regime head is ignored. Moderate ambiguity weighting improved performance, but overly large values slightly degraded accuracy because some informative but uncertain measurements were excessively penalized.

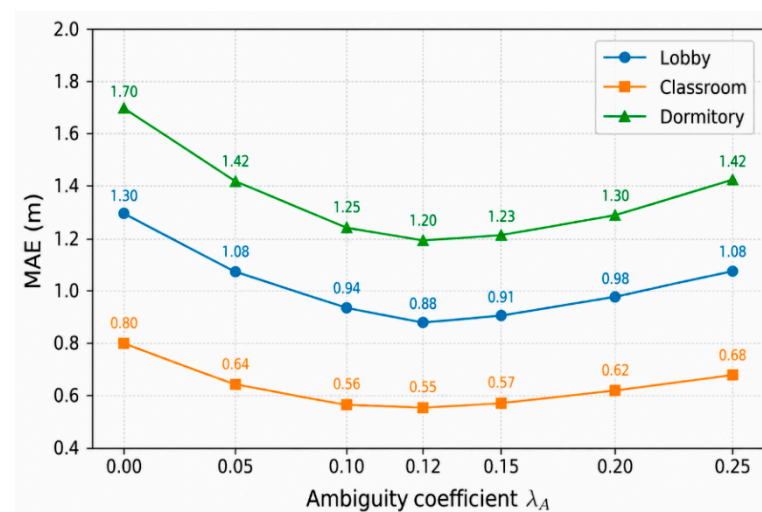


Figure 8. Sensitivity of localization error to ambiguity coefficient.

5.11. Statistical Reporting

Table 12 reports confidence intervals, effect sizes, and Holm-corrected significance testing for the comparison between MRL-GCF and mixture-based range calibration. The corrected p -values remained below 0.05 in all environments. The effect sizes were small-to-moderate, which is reasonable because mixture-based calibration was already a strong baseline.

Table 12. Statistical comparison between MRL-GCF and mixture-based range calibration.

Site	MAE Difference (m)	95% CI	Cliff's Delta	Raw <i>p</i> -Value	Holm-Corrected <i>p</i> -Value
Lobby	0.05	[0.02, 0.08]	0.31	0.004	0.012
Classroom	0.03	[0.01, 0.05]	0.24	0.031	0.031
Dormitory	0.07	[0.03, 0.11]	0.35	0.002	0.006

5.12. Computational Cost

Table 13 reports the online runtime per localization query on a laptop with an Intel Core i7 14900 CPU and 16 GB RAM. MRL-GCF required more computation than raw least squares but remained lightweight. The average runtime of MRL-GCF was 6.40 ms per query, which is sufficient for many static or low-rate indoor localization applications. The additional computation mainly comes from reliability feature construction and neural calibration.

Table 13. Average online runtime per query.

Method	Feature Time (ms)	Calibration Time (ms)	Solver Time (ms)	Total Time (ms)
Raw FTM least squares	0.10	–	2.35	2.45
State-calibrated WLS	0.42	0.58	2.61	3.61
Mixture-based range calibration	0.46	0.74	2.69	3.89
Direct neural coordinate regression	0.38	1.42	–	1.80
MRL-GCF	1.05	2.13	3.22	6.40

6. Discussion

The experimental results support the central hypothesis that robust Wi-Fi FTM localization depends on estimating measurement reliability before geometric localization. Raw FTM provides useful geometric information, but its reliability varies across APs, positions, devices, AP density, map availability, and propagation conditions. RSSI alone cannot fully describe this reliability because strong signal strength does not necessarily imply accurate ranging. The proposed measurement-reliability representation improves calibration by incorporating stability, visibility, abnormality, and geometry descriptors.

The heteroscedastic neural calibrator provides a practical compromise between learning-based modeling and geometric interpretability. Instead of directly regressing coordinates, it estimates AP-wise range bias and uncertainty. This design keeps the physical relationship between APs and user coordinates while allowing nonlinear measurement degradation to be learned from data. The reliability calibration analysis shows that predicted uncertainty is positively correlated with actual ranging residuals, and the ECE and NLL values further support the calibration quality of the uncertainty estimates.

The reliability-regime head is trained with residual-regime soft targets, which addresses the interpretability of the entropy-based ambiguity term. The entropy term should therefore be understood as a learned ambiguity measure associated with residual reliability, not as an untrained latent score. This design helps suppress observations whose propagation quality is uncertain even when their RSSI values are strong.

MRL-GCF also explains why direct neural coordinate regression is less robust. Direct coordinate regression must learn measurement correction and geometric inference in a single mapping. This can increase the learning burden and reduce interpretability, especially when AP visibility or measurement reliability changes. MRL-GCF decomposes the problem into AP-wise reliability estimation and geometry-constrained localization, which preserves the known geometric structure while using learning only where it is most needed.

Map-geometry fusion further improves robustness. Indoor localization is affected not only by wireless signal features but also by building structure and anchor distribution. An

AP may have strong RSSI but poor geometric contribution or may be blocked by a wall, partition, or metallic surface. By combining map obstruction, material-aware obstruction, anchor geometry, learned uncertainty, and signal confidence, MRL-GCF reduces the influence of unreliable APs. The map availability and perturbation results show that floor maps improve performance but are not indispensable; when maps are unavailable, the method still operates as a measurement- and geometry-aware localization framework.

The replay-based dynamic and temporal robustness analyses show that short-window stability features are practical when implemented as sliding-window descriptors. The method does not require a user to remain stationary for a long period. Performance degradation under orientation change, human presence, and temporal variation remains moderate, but these conditions still increase error, especially in the dormitory. This finding confirms that dynamic user motion and environmental drift remain important challenges for Wi-Fi FTM localization.

The device and AP-density stress tests show that hardware variation and anchor availability affect localization accuracy. This is expected because FTM ranging depends on timing implementation and AP geometry. The results support the need for device-adaptive calibration, but they also show that MRL-GCF remains stable under moderate initiator and AP-density changes.

The combined spatial error maps in Figure 6 provide additional visual evidence for the quantitative results. The classroom shows the most regular error distribution, which is consistent with its simple layout and stable signal condition. The lobby has larger errors around the metal partition and peripheral areas, indicating the effect of local reflection and geometry degradation. The dormitory shows the most heterogeneous spatial error pattern because dense walls, corners, and furniture create more complex propagation conditions. These observations explain why the improvement of MRL-GCF is more evident in complex environments than in regular open spaces.

Several limitations should be acknowledged. First, MRL-GCF still requires calibration data, and long-term environmental changes may require online adaptation. Second, the material-aware map cue uses simplified material categories and does not model detailed reflection or diffraction physics. Third, although dynamic replay, temporal checks, and device/AP-density stress tests were included, broader validation across larger floor areas, more AP brands, dense crowds, and multi-floor buildings is still needed. Fourth, the present method estimates horizontal position with height-corrected ranges and does not perform floor-level detection or vertical coordinate estimation. Future work will address these limitations through online self-calibration, material-aware propagation modeling, inertial-aided dynamic tracking, cross-device adaptation, and validation in larger multi-floor environments.

7. Conclusions

This study presented MRL-GCF, a reliability-aware calibration and fusion framework for Wi-Fi FTM indoor localization. The method constructs AP-wise reliability features from RSSI, logarithmic FTM range, short-window stability, visibility, abnormality, and coarse geometry. A heteroscedastic neural calibrator estimates range bias and uncertainty for each AP observation, while a supervised reliability-regime head provides a trained ambiguity measure. The learned uncertainty is fused with ambiguity, material-aware map obstruction, and anchor geometry to select reliable anchors and define trust-aware weighted least-squares weights.

Experiments in a lobby, a classroom, and a dormitory demonstrated stable meter-level horizontal localization performance under scan-level and point-held-out protocols. Compared with fingerprinting, raw FTM least squares, robust least squares, state-calibrated WLS, mixture-based range calibration, and direct neural coordinate regression,

MRL-GCF achieved lower average error and better tail-error robustness. Feature-level ablation, component ablation, reliability calibration analysis, map perturbation, material-aware obstruction analysis, dynamic replay, temporal robustness, device/AP-density stress testing, coefficient sensitivity, spatial error visualization, statistical testing, and runtime evaluation further confirmed the contribution and practicality of the proposed framework.

Future work will focus on online calibration, dynamic tracking with inertial sensing, material-aware propagation modeling, cross-device generalization, and validation in larger multi-floor indoor environments.

Author Contributions: S.G. designed the study, developed the method, conducted the experiments, analyzed the data, and prepared the manuscript. S.H. contributed to data collection, implementation, and validation. S.D. supervised the study, reviewed the methodology, and revised the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This work was financially supported by the grants from the National Natural Science Foundation of China under Grant Number 52105509, the Fundamental Research Funds for the Central Universities under Grant Number 2232023D-25, Beijing Postdoctoral Research Foundation, and Shanghai Science and Technology Program (20DZ2251400).

Data Availability Statement: The datasets generated and analyzed during the current study are available from the corresponding author on reasonable request.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Singh, J.; Tyagi, N.; Singh, S.; Ali, F.; Kwak, D. A systematic review of contemporary indoor positioning systems: Taxonomy, techniques, and algorithms. *IEEE Internet Things J.* **2024**, *11*, 34717–34733. <https://doi.org/10.1109/JIOT.2024.3416255>.
2. Qi, L.; Liu, Y.; Yu, Y.; Chen, L.; Chen, R. Current status and future trends of meter-level indoor positioning technology: A review. *Remote Sens.* **2024**, *16*, 398. <https://doi.org/10.3390/rs16020398>.
3. Zafari, F.; Gkelias, A.; Leung, K.K. A survey of indoor localization systems and technologies. *IEEE Commun. Surv. Tutor.* **2019**, *21*, 2568–2599. <https://doi.org/10.1109/COMST.2019.2911558>.
4. Davidson, P.; Piche, R. A survey of selected indoor positioning methods for smartphones. *IEEE Commun. Surv. Tutor.* **2017**, *19*, 1347–1370. <https://doi.org/10.1109/COMST.2016.2637663>.
5. Al-Okby, M.F.R.; Nezhad, E.Z.; Mahmood, A. UWB-based real-time indoor positioning systems: A comprehensive review. *Appl. Sci.* **2024**, *14*, 11005. <https://doi.org/10.3390/app142311005>.
6. Cheraghinia, M.; Esmaeili, M.; Torkestani, J.A. A comprehensive overview on UWB radar: Applications, standards, signal processing techniques, datasets, radio chips, trends and future research directions. *IEEE Commun. Surv. Tutor.* **2024**, *27*, 2283–2324.
7. Milano, F.; da Rocha, H.; Laracca, M.; Ferrigno, L.; Espírito Santo, A.; Salvado, J.; Paciello, V. BLE-based indoor localization: Analysis of some solutions for performance improvement. *Sensors* **2024**, *24*, 376.
8. Assayag, Y.; Knafo, N.; Shalom, A. A model-based BLE indoor positioning system using particle swarm optimization. *IEEE Sens. J.* **2024**, *24*, 6898–6908.
9. Bastiaens, S.; Alijani, M.; Joseph, W.; Plets, D. Visible light positioning as a next-generation indoor positioning technology: A tutorial. *IEEE Commun. Surv. Tutor.* **2024**, *26*, 2867–2913.
10. Bahl, P.; Padmanabhan, V.N. RADAR: An in-building RF-based user location and tracking system. In Proceedings of the IEEE INFOCOM 2000 Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies, Tel Aviv, Israel, 26–30 March 2000; pp. 775–784.
11. Yiu, S.; Dashti, M.; Claussen, H.; Perez-Cruz, F. Wireless RSSI fingerprinting localization. *Signal Process.* **2017**, *131*, 235–244.
12. Yoon, N.; Jung, W.; Kim, H. DeepRSSI: Generative model for fingerprint-based localization. *IEEE Access* **2024**, *12*, 66196–66213. <https://doi.org/10.1109/ACCESS.2024.3398734>.
13. Dogan, D.; Tuncer, T.; Ertam, F. A simplified method based on RSSI fingerprinting for IoT device localization in smart cities. *IEEE Access* **2024**, *12*, 163752–163763.

14. IEEE. *IEEE Standard for Information Technology-Telecommunications and Information Exchange Between Systems Local and Metropolitan Area Networks-Specific Requirements-Part 11: Wireless LAN Medium Access Control and Physical Layer Specifications*; IEEE: Piscataway, NJ, USA, 2020.
15. Kosek-Szott, K.; Szott, S.; Ciezobka, W.; Wojnar, M.; Rusek, K.; Segev, J. Indoor positioning with Wi-Fi Location: A survey of IEEE 802.11mc/az/bk Fine Timing Measurement research. *arXiv* **2025**, arXiv:2509.03901.
16. Guan, F.; Chen, C.; Wu, D. Autonomous wireless positioning system using crowdsourced Wi-Fi fingerprinting and self-detected FTM stations. *Expert Syst. Appl.* **2024**, *255*, 124566. <https://doi.org/10.1016/j.eswa.2024.124566>.
17. Bravo-Arrabal, J.; Simón Álvarez-Merino, C.; Toscano-Moreno, M.; Serón-Barba, J.; Fernández-Lozano, J.J.; Gómez-Ruiz, J.A.; Khatib, E.J.; Barco, R.; García-Cerezo, A. Real-time FTM-based victim positioning system using heterogeneous robots in remote and outdoor scenarios. *IEEE Access* **2025**, *13*, 86949–86967.
18. Guvenc, I.; Chong, C.C. A survey on TOA based wireless localization and NLOS mitigation techniques. *IEEE Commun. Surv. Tutor.* **2009**, *11*, 107–124.
19. Xiong, J.; Jamieson, K.; Sundaresan, K. Practical ranging with commodity Wi-Fi. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* **2023**, *7*, 1–25.
20. Gentner, C.; Jost, T.; Wang, W.; Zhang, S.; Dammann, A.; Fiebig, U.-C. Multipath assisted positioning with simultaneous localization and mapping. *IEEE Trans. Wirel. Commun.* **2016**, *15*, 6104–6117.
21. Zhuang, Y.; Syed, Z.; Li, Y.; El-Sheimy, N. Evaluation of two Wi-Fi positioning systems based on autonomous crowdsourcing of handheld devices for indoor navigation. *IEEE Trans. Mob. Comput.* **2026**, *15*, 1982–1995.
22. Marano, S.; Gifford, W.M.; Wymeersch, H.; Win, M.Z. NLOS identification and mitigation for localization based on UWB experimental data. *IEEE J. Sel. Areas Commun.* **2010**, *28*, 1026–1035.
23. Sharp, I.; Yu, K. Enhanced least-squares positioning algorithm for indoor positioning. *IEEE Trans. Mob. Comput.* **2013**, *12*, 1640–1650.
24. Mansour, A.; Chen, W.; Wang, J.; Adham, M.; Qu, X.; Weng, D. Reliability-governed scaling of mobile crowdsensing-based 3D radio mapping for ubiquitous indoor positioning: An uncertainty-aware lifecycle review and deployment-oriented pipeline. *Inf. Fusion* **2026**, *135*, 104459. <https://doi.org/10.1016/j.inffus.2026.104459>.
25. Lin, Y.; Zhang, X.; Wang, H.; Liu, Z. The state of the art of deep learning-based Wi-Fi indoor positioning: A review. *IEEE Sens. J.* **2024**, *24*, 27076–27095.
26. Poulose, A.; Han, D.S. Hybrid deep learning model based indoor positioning using Wi-Fi RSSI heat maps. *Electronics* **2021**, *10*, 2. <https://doi.org/10.3390/electronics10010002>.
27. He, S.; Chan, S.H.G.; Yu, L. Fusing noisy fingerprints with distance bounds for indoor localization. *IEEE Trans. Mob. Comput.* **2016**, *15*, 2782–2795.
28. Luo, Y.; Hoeber, O.; Chen, Y. Enhancing Wi-Fi fingerprinting for indoor positioning using human-centric collaborative feedback. *Hum. Centric Comput. Inf. Sci.* **2013**, *3*, 1–23.
29. Huber, P.J. *Robust Statistics*; Wiley: New York, NY, USA, 1981.
30. Wu, C.; Yang, Z.; Zhou, Z.; Liu, Y.; Xi, W. WILL: Wireless indoor localization without site survey. *IEEE Trans. Parallel Distrib. Syst.* **2013**, *24*, 839–848.
31. Kotaru, M.; Joshi, K.; Bharadia, D.; Katti, S. SpotFi: Decimeter level localization using WiFi. *ACM SIGCOMM Comput. Commun. Rev.* **2015**, *45*, 269–282.
32. Vasisht, D.; Kumar, S.; Katabi, D. Decimeter-level localization with a single WiFi access point. In Proceedings of the 13th USENIX Symposium on Networked Systems Design and Implementation, Santa Clara, CA, USA, 16–18 March 2016; pp. 165–178.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.