# DLE-DG: Decoupling Location and Environment Representations for Domain Generalization in RFID Fingerprint Indoor Localization

Shuai Wang[1,2], Weiqing Huang[1,2], Siye Wang[1,2], Yanfang Zhang[1,2]*, Jing Li[1,2], Shanjing Yang[1], and Zhaoxu Wang[1,3]

[1] Institute of Information Engineering, Chinese Academy of Sciences, Beijing, China
[2] School of Cyber Security, University of Chinese Academy of Sciences, Beijing, China
[3] School of Industry-education Integration, University of Chinese Academy of Sciences, Beijing, China
{wangshuai,zhangyanfang}@iie.ac.cn

**Abstract.** RFID fingerprint indoor localization has attracted widespread attention due to its device-free nature for the target and no power supply for reference tags. However, dynamic environmental changes alter signal propagation paths and induce substantial fingerprint distribution shift, degrading accuracy and stability. Existing domain generalization (DG) methods treat environments as domains and learn domain-invariant representations across environments. Yet they often leave residual environmental interference entangled with location information in the learned representation, resulting in poor generalization performance in unseen environments. To address this limitation, we propose DLE-DG, a decoupling RFID fingerprint localization framework designed for dynamic-environment domain generalization. DLE-DG explicitly separates location and environment factors to learn location representations with minimal residual environmental interference, thereby enabling robust generalization to unseen environments. It comprises two synergistic components: (i) a decoupling-oriented fingerprint construction strategy that denoises and normalizes multi-channel observations to produce stable inputs; (ii) a domain-adversarially enhanced decoupling network with a shared representation extractor and a two-branch architecture: the location branch for localization and the environment domain-adversarial branch that suppresses residual environmental interference in the shared representation via min–max adversarial learning, thereby minimizing environmental contamination of the location representation. Experiments demonstrate that DLE-DG achieves state-of-the-art generalization in unseen environments, with an average localization error of 0.08 m, 63% lower than the best advanced localization method.

**Keywords:** Wireless sensing · Indoor fingerprint localization · Decoupled representation · Domain generalization · RFID.

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29921-5_33

## 1 Introduction

With the rapid advancement of smart cities, applications such as indoor navigation [4] and asset tracking [1] demand increasingly robust indoor intelligent sensing capabilities. As a foundational enabler, indoor localization directly determines the functionality and user experience of such applications. RFID fingerprint indoor localization collects radio frequency signals, including received signal strength indicator (RSSI) and phase, via reader antennas and passive reference tags, constructs a location fingerprint mapping database, and performs localization [5]. This approach imposes no hardware burden on users and requires no power supply for reference tags, resulting in low operational overhead and strong practical viability for large-scale indoor deployment [9].

Traditional RFID fingerprint localization methods assume distributional stationarity between training and testing data, i.e., the indoor environment remains static from training to deployment. In practice, however, indoor spaces frequently undergo dynamic changes in furniture quantity and spatial layout. These variations alter multipath propagation paths, inducing substantial distribution shifts in RFID fingerprints [13]. Consequently, localization models trained in controlled environments often suffer severe degradation in accuracy and stability when deployed in unseen dynamic environments (hereafter referred to as unseen environments), constituting a critical barrier to real-world adoption.

To alleviate the above issues, DG has been adapted to RFID localization: distinct indoor environments are treated as separate domains, with training environments designated as source domains and unseen environments as target domains. DG aims to learn a model that generalizes stably to target domains using only source domain data [15]. Existing approaches fall into two primary categories. First, data augmentation methods [14,19] increase training sample diversity via perturbation, mixing, and recombination. However, such techniques primarily expand coverage within known source distributions and poorly model the signal shifts arising from altered propagation paths, limiting their transferability to truly unseen environments. Second, adversarial learning methods [18,11] jointly train across multiple source domains and employ domain-adversarial objectives to extract representations invariant to environmental variation. While these improve robustness, their performance remains constrained by source domain diversity: when target environments diverge significantly from all source domains, learned invariant representations often retain residual environmental interference. This manifests as entanglement of location and environment factors in the representation space, undermining generalization to unseen environments. Thus, effective decoupling of location and environment factors is essential for reliable generalization in unseen environments.

In this work, we address the poor generalization of RFID fingerprint indoor localization in unseen environments by explicitly decoupling location and environment factors to enable localization using a location representation that minimizes residual environmental interference. While decoupling has been widely explored for high-dimensional data (e.g., images) [16,6], it remains underexplored and fundamentally more challenging for RFID signals: RSSI and phase are

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29921-5_33

low-dimensional, one-dimensional time-series observations formed by multipath superposition, where location, environment, and hardware factors are intrinsically entangled [10]. This entanglement produces sparse, information-poor signals with highly nonlinear structure, posing severe challenges for effective factor decoupling. Moreover, conventional decoupling methods suffer from equivalent decomposition and information leakage: representations may appear structurally distinct yet retain latent correlations between location and environment factors, causing residual environmental interference to persist in the location representation and degrade cross-environment performance.

To overcome these limitations, we propose DLE-DG, a decoupling RFID fingerprint localization framework designed for dynamic-environment domain generalization. DLE-DG comprises two core components. First, a decoupling-oriented fingerprint construction strategy (DFCS) leverages multi-channel antenna–tag observations acquired via antenna and tag arrays. Beyond enriching fingerprint information, it applies systematic denoising and normalization to suppress interference from non-location-related factors, including hardware imperfections and acquisition noise, yielding unified, stable fingerprint inputs that facilitate decoupling learning. Second, a domain-adversarially enhanced decoupling network (DADN) employs a shared representation extractor and a two-branch architecture: the location branch serves as the primary localization pathway, learning a location representation optimized for localization; the environment domain-adversarial branch functions as an auxiliary branch, imposing min–max adversarial constraints to suppress residual environmental factors in the shared representation. Crucially, this suppression directly minimizes environmental contamination of the location representation. Furthermore, we enforce orthogonality between the location and environment representations in the joint objective to explicitly minimize mutual information leakage and strengthen factor separation. Collectively, these mechanisms enable DLE-DG to learn a location representation that is intrinsically robust to environmental changes, yielding significantly improved generalization to unseen environments.

The main contributions of this paper are summarized as follows:

1. **A decoupling RFID fingerprint localization framework for dynamic-environment domain generalization.** We propose DLE-DG, which explicitly separates location and environment factors to enable localization in previously unseen environments using a location representation that minimizes residual environmental interference.
2. **A domain-adversarially enhanced decoupling network.** We design DADN with a shared representation extractor and a two-branch architecture: the location branch performs localization, while the environment domain-adversarial branch suppresses residual environmental interference in the shared representation via min–max adversarial learning, directly reducing environmental contamination of the location representation.
3. **A decoupling-oriented fingerprint construction strategy.** We design DFCS to construct stable fingerprint inputs that enrich fingerprint informa-

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29921-5_33

tion and mitigate non-location-related interference to enable robust decoupling learning.
4. **Comprehensive evaluation and public dataset.** We conduct comprehensive evaluation demonstrating that DLE-DG achieves state-of-the-art localization performance in both static-environment and unseen-environment settings. The dataset used in this study is publicly available at DLE-DG dataset.

## 2 Related Work

### 2.1 RFID Fingerprint Indoor Localization

RFID fingerprint indoor localization is a passive, contactless method that requires no device on the target. It builds a location fingerprint database by exploiting spatial variations in radio frequency signals, primarily RSSI and phase, and estimates location via similarity matching against the database [7]. Deep learning has significantly advanced this paradigm by capturing complex, nonlinear signal-location relationships: CNNs model local spatial structure in fingerprint inputs [21]; DNNs and LSTMs learn high-capacity nonlinear transformations and temporal dynamics [2,8]; Transformers use self-attention to encode global signal dependencies, yielding state-of-the-art performance [12]. However, most existing methods implicitly assume distributional stationarity between training and deployment: i.e., the environment remains unchanged after training. In practice, changes in furniture quantity and layout continuously alter wireless propagation paths, inducing substantial distribution shifts in RFID fingerprints [13]. This mismatch causes severe degradation in accuracy and stability during unseen environments, constituting a fundamental barrier to adoption.

### 2.2 Decoupled Representation Learning

In recent years, decoupled representation learning has attracted attention in vision tasks. It aims to learn multiple statistically independent latent representations, each aligned with a distinct generative factor of the data, so that variation in one factor induces minimal change in others [17]. Gan et al. [6] decouple tumor appearance from organ-specific anatomical context in medical images using attention-guided representation separation, improving tumor localization accuracy. Xiao et al. [16] decouple eyeball radius, pupil, and iris in 3D eyeball modeling to enhance gaze estimation robustness. However, unlike high-dimensional visual data, RFID signals are low-dimensional, one-dimensional time-series observations formed by multipath superposition. Their sparse, nonlinear information content exhibits intrinsic entanglement among location, environment, and hardware factors, making effective decoupling fundamentally more challenging. Moreover, conventional decoupling methods suffer from equivalent decomposition (i.e., structurally distinct but functionally coupled representations) and information leakage (i.e., residual environmental interference contaminating the location representation). These limitations motivate the need for more targeted constraints and decoupling-oriented designs for RFID fingerprint localization.

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29921-5_33

## 3 Method

We propose DLE-DG, a decoupling RFID fingerprint localization framework designed for dynamic-environment domain generalization. As illustrated in Fig. 1, DLE-DG comprises two tightly integrated components: a decoupling-oriented fingerprint construction strategy and a domain-adversarially enhanced decoupling network.

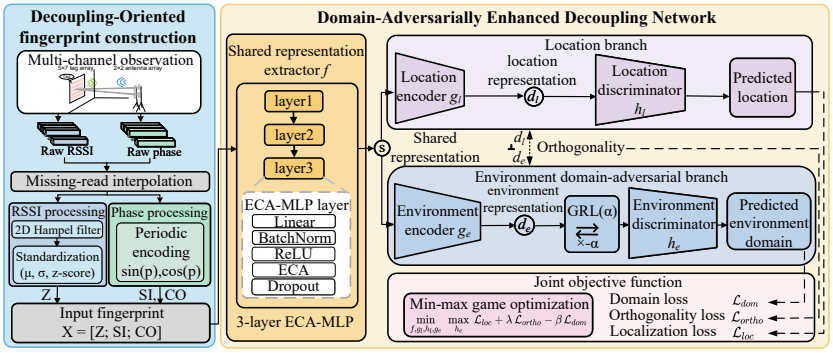



Fig. 1: Architecture of DLE-DG.

### 3.1 Decoupling-Oriented Fingerprint Construction Strategy

The decoupling-oriented fingerprint construction strategy follows a three-stage pipeline of data collection, signal preprocessing, and fingerprint construction, which comprises five parts: multi-channel observation, missing-read interpolation, RSSI processing, phase processing, and input fingerprint construction.

**Multi-Channel Observation.** We deploy a $2 \times 2$ antenna array and a $5 \times 7$ tag array to acquire rich, spatially diverse antenna–tag observations to enrich fingerprint information for decoupling (configuration details are provided in Section 4.1). Let $C$ denote the number of antennas ($C = 4$), and let $H$ and $W$ denote the height and width of the tag grid ($H = 5$, $W = 7$). A single sampling thus yields $C \times H \times W = 140$ distinct antenna–tag channels, each capturing synchronized RSSI and phase observations. Let $r_{c,i,j}$ and $p_{c,i,j}$ denote the RSSI and phase values observed by antenna $c$ at tag-array coordinate $(i, j)$, where $c \in [1, C]$, $i \in [1, H]$, and $j \in [1, W]$. Signal preprocessing is performed independently per antenna; for concision, we omit the antenna index $c$ in subsequent notation and use $r_{i,j}$ and $p_{i,j}$ to represent RSSI and phase at an arbitrary antenna.

**Missing-Read Interpolation.** During RFID data collection, certain antenna–tag pairs may yield missing reads. Naive imputation with a constant distorts the signal distribution and harms training stability. To preserve spatial coherence while respecting the physical continuity of multipath propagation

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29921-5_33

within local neighborhoods [13], we adopt 2D neighborhood-mean interpolation, grounded in the topological layout of the $5 \times 7$ tag array.

Let binary mask matrix $M$ indicate observation availability, where $mask_{i,j} = 1$ if the $(i,j)$-th tag yields a valid observation and 0 otherwise. For each coordinate $(i,j)$, define its 4-connected neighborhood index set:

$$N(i,j) = \{(i-1,j), (i+1,j), (i,j-1), (i,j+1)\} \tag{1}$$

For a missing coordinate $(i,j)$, define the set of valid neighboring observation indices $L_{i,j}$ as those within array bounds and not missing:

$$L_{i,j} = \{(u,v) \in N(i,j) \mid 1 \le u \le H, 1 \le v \le W, mask_{u,v} = 1\} \tag{2}$$

where $(u,v)$ indexes a neighbor of tag $(i,j)$.

Let $K_{i,j} = |L_{i,j}|$ denote the count of valid neighbors. Given RSSI and phase matrices $R$ and $P$ in the tag array, interpolation proceeds as follows:

$$\hat{r}_{i,j} = \begin{cases} r_{i,j}, & mask_{i,j} = 1, \\ \frac{1}{K_{i,j}} \sum_{(u,v) \in L_{i,j}} r_{u,v}, & mask_{i,j} = 0,\ K_{i,j} > 0, \\ \text{mean}(R), & mask_{i,j} = 0,\ K_{i,j} = 0. \end{cases}$$
$$\hat{p}_{i,j} = \begin{cases} p_{i,j}, & mask_{i,j} = 1, \\ \frac{1}{K_{i,j}} \sum_{(u,v) \in L_{i,j}} p_{u,v}, & mask_{i,j} = 0,\ K_{i,j} > 0, \\ \text{mean}(P), & mask_{i,j} = 0,\ K_{i,j} = 0. \end{cases} \tag{3}$$

where $\hat{r}_{i,j}$ and $\hat{p}_{i,j}$ are the interpolated RSSI and phase values at $(i,j)$; $r_{u,v}$ and $p_{u,v}$ are valid neighboring observations; and mean$(\cdot)$ computes the arithmetic mean over the respective matrix ($R$ or $P$). When no valid neighbors exist ($K_{i,j} = 0$), matrix-mean imputation ensures numerical stability.

**RSSI Processing.** Even after missing-read interpolation, RSSI values may retain impulse-type outliers. Such outliers severely degrade fingerprint stability and localization accuracy. To suppress only anomalous values while preserving the intrinsic 2D spatial structure of the tag array, we extend Hampel filtering to two dimensions. Our 2D Hampel filter replaces only statistically identified outliers, thereby avoiding over-smoothing of observations and maintaining fidelity to the original spatial signal distribution. The procedure is as follows:

1. For each interpolated RSSI value $\hat{r}_{i,j}$, define its neighborhood window index set $W_{i,j}$ (a $3 \times 3$ sliding window centered at $(i,j)$; at boundaries, $W_{i,j}$ contains only existing indices). Compute the local window median: $m_{i,j} = \text{median}(\hat{r}_{u,v} : (u,v) \in W_{i,j})$.
2. Compute the local median absolute deviation: $MAD_{i,j} = \text{median}(|\hat{r}_{u,v} - m_{i,j}| : (u,v) \in W_{i,j})$
3. Using an anomaly-threshold scaling constant $t = 3$ (empirically determined), declare $\hat{r}_{i,j}$ an outlier if $|\hat{r}_{i,j} - m_{i,j}| > t \cdot MAD_{i,j}$. Outliers are replaced by the local median; all other values remain unchanged:

$$\tilde{r}_{i,j} = \begin{cases} m_{i,j}, & |\hat{r}_{i,j} - m_{i,j}| > t \times MAD_{i,j}, \\ \hat{r}_{i,j}, & \text{otherwise.} \end{cases} \tag{4}$$

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29921-5_33

where $\tilde{r}_{i,j}$ denotes the denoised RSSI value after 2D Hampel filtering.

To mitigate RSSI scale variations across environments, we apply global standardization over the entire training set. Specifically, all interpolated and filtered RSSI matrices are flattened into 1D vectors $V_r^{(n)}$ for sample $n$. Let $\tilde{r}_i^{(n)}$ denote the $i$-th element of $V_r^{(n)}$, where $i \in [1, CHW]$ and $CHW = 140$ is the total number of antenna–tag channels per sample. The mean $\mu$ and standard deviation $\sigma$ are computed over the full training set of $N$ samples:

$$\begin{aligned} \mu &= \frac{\sum_{n=1}^{N} \sum_{i=1}^{CHW} \tilde{r}_i^{(n)}}{N \cdot CHW} \\ \sigma &= \sqrt{\frac{\sum_{n=1}^{N} \sum_{i=1}^{CHW} (\tilde{r}_i^{(n)} - \mu)^2}{N \cdot CHW}} \\ z_i^{(n)} &= \frac{\tilde{r}_i^{(n)} - \mu}{\sigma} \end{aligned} \tag{5}$$

where $z_i^{(n)}$ is the standardized value of $\tilde{r}_{i,j}^{(n)}$, the resulting standardized vector $Z^{(n)} = [z_1^{(n)}, \ldots, z_{CHW}^{(n)}]$ serves as the final RSSI input for the decoupling network.

**Phase Processing.** Phase is a periodic quantity subject to angular wrapping, which introduces discontinuities (e.g., jumps from $2\pi$ to 0) that violate spatial continuity and impair learning. To preserve the inherent periodicity of phase while eliminating discontinuity, we apply periodic-consistency encoding: each interpolated phase value is mapped onto the unit circle via sine and cosine transformations. Specifically, interpolated phase matrices are flattened into 1D vectors $V_p^{(n)}$ for sample $n$. Let $\hat{p}_i^{(n)}$ denote the $i$-th element of $V_p^{(n)}$, where $i \in [1, CHW]$ and $CHW = 140$. We compute:

$$\begin{aligned} si_i^{(n)} &= \sin(\hat{p}_i^{(n)}) \\ co_i^{(n)} &= \cos(\hat{p}_i^{(n)}) \end{aligned} \tag{6}$$

where $si_i^{(n)}$ and $co_i^{(n)}$ are the sine and cosine periodic encodings of $\hat{p}_i^{(n)}$. Correspondingly, the resulting vectors $SI^{(n)} = [si_1^{(n)}, \ldots, si_{CHW}^{(n)}]$ and $CO^{(n)} = [co_1^{(n)}, \ldots, co_{CHW}^{(n)}]$ constitute the final phase encodings. Since both $\sin(\cdot)$ and $\cos(\cdot)$ naturally constrain outputs to $[-1, 1]$, no further standardization is required.

**Input Fingerprint Construction.** After preprocessing RSSI and phase, we construct the final multi-channel fingerprint input $x^{(n)}$ by horizontally concatenating the standardized RSSI vector $Z^{(n)}$, the sin-encoded phase vector $SI^{(n)}$, and the cos-encoded phase vector $CO^{(n)}$:

$$x^{(n)} = [Z^{(n)}; SI^{(n)}; CO^{(n)}] \tag{7}$$

This yields a $3 \times CHW = 420$-dimensional input per sample, preserving signal richness and suppressing non-location-related interference to enable robust decoupling learning.

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29921-5_33

### 3.2 Domain-Adversarially Enhanced Decoupling Network

The domain-adversarially enhanced decoupling network (as shown in Fig. 1) comprises four tightly coupled components: a shared representation extractor, a location branch, an environment domain-adversarial branch, and a joint objective function.

**Shared Representation Extractor.** Because antenna–tag channels exhibit heterogeneous contributions to localization, we design an Efficient Channel Attention–augmented multilayer perceptron (ECA-MLP) as the shared representation extractor. ECA is embedded in each MLP layer to perform progressive, cross-layer channel re-weighting (illustrated in Fig. 1, ECA-MLP layer). This enables early suppression of environment-sensitive channel activations, preventing their propagation into deeper representations, while amplifying stable, location-discriminative channels. The shared representation $s$ is computed as $s = f(x)$, where $x$ denotes the preprocessed fingerprint input and $f(\cdot)$ implements a three-layer ECA-MLP with hidden dimensions $420 \to 512 \to 384 \to 256$.

**Location Branch.** The location branch processes $s$ to learn a location representation $d_l$ for localization. It comprises a location encoder $g_l(\cdot)$ and a location discriminator $h_l(\cdot)$, which jointly map $s$ to 2D coordinates $\hat{y}_l$.

$$\begin{aligned} d_l &= g_l(s) \\ \hat{y}_l &= h_l(d_l) \end{aligned} \tag{8}$$

Here, $g_l(\cdot)$ implements a two-layer MLP with hidden dimensions $256 \to 192 \to 128$, and $h_l(\cdot)$ implements a two-layer MLP with hidden dimensions $128 \to 64 \to 2$.

**Environment Domain-Adversarial Branch.** To explicitly suppress environmental interference in location representation $d_l$, we design an environment domain-adversarial branch as an auxiliary branch. It engages the shared representation extractor and an environment encoder $g_e(\cdot)$ in a min–max game with an environment discriminator $h_e(\cdot)$, trained to make the environment representation $d_e = g_e(s)$ indistinguishable across environments. This forces $s$ to discard the environment factor, thereby reducing residual environmental interference in $d_l$.

Specifically, $g_e(\cdot)$ (dimensions $256 \to 192 \to 128$) produces $d_e$, while $h_e(\cdot)$ (dimensions $128 \to 64 \to 4$) predicts domain labels $y_e$ after gradient reversal layer (GRL) processing: $\hat{y}_e = h_e\big(\mathrm{GRL}(d_e; \alpha)\big)$.

The GRL acts as the identity mapping during forward pass ($\mathrm{GRL}(d_e; \alpha) = d_e$), but during backpropagation multiplies gradients from $h_e$ by $-\alpha$, effectively reversing the update direction for $g_e$ and $f$. Thus, while $h_e$ minimizes domain classification loss $\mathcal{L}_{dom}$, $g_e$ and $f$ maximize it to realize adversarial training via standard gradient descent. To stabilize optimization, we schedule $\alpha$ progressively:

$$\alpha(k) = \frac{2}{1 + \exp(-10k)} - 1, \qquad k = \frac{t}{T}. \tag{9}$$

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29921-5_33

where $t$ is the current epoch and $T$ is the total number of epochs. This sigmoidal schedule smoothly increases $\alpha$ from 0 to 1, prioritizing localization supervision early and gradually strengthening domain confusion later.

**Joint Objective Function.** We jointly optimize localization loss, domain-adversarial loss, and orthogonal regularization loss in the form of a min–max game:

$$\min_{f,g_l,h_l,g_e} \max_{h_e} \mathcal{L}_{loc} + \lambda \mathcal{L}_{ortho} - \beta \mathcal{L}_{dom} \tag{10}$$

where $\lambda = 0.1$ and $\beta = 1$. Specifically:

1. Localization loss: $\mathcal{L}_{loc} = \mathbb{E}_{(x,y_l)}[\|\hat{y}_l - y_l\|_2^2]$ (mean squared error);
2. Domain-adversarial loss: $\mathcal{L}_{dom} = \mathbb{E}_{(x,y_e)}[\mathrm{CE}(\hat{y}_e, y_e)]$ (cross-entropy);
3. Orthogonal regularization loss: $\mathcal{L}_{ortho} = \mathbb{E}_{(x)}[|\cos(d_l, d_e)|]$, which penalizes the cosine similarity between $d_l$ and $d_e$, enforcing orthogonality between $d_l$ and $d_e$ to structurally decouple location and environment factors.

Eq. 10 implements a min–max game optimization: the environment discriminator is trained to minimize $\mathcal{L}_{dom}$ ($\min_{h_e} \mathcal{L}_{dom}$), while all other modules are jointly optimized to minimize $\mathcal{L}_{loc}$ and $\mathcal{L}_{ortho}$ and simultaneously maximize $\mathcal{L}_{dom}$ ($\min_{f,g_l,h_l,g_e} \mathcal{L}_{loc} + \lambda \mathcal{L}_{ortho} - \beta \mathcal{L}_{dom}$). By suppressing environment information in the shared representation $s$, this process reduces residual environmental interference in $d_l$, thereby improving generalization across unseen environments.

## 4 Evaluation

### 4.1 Experimental Setup

**Implementation Details.** Data collection and processing are performed on a laptop equipped with an Intel(R) Core i7-6500U CPU and an NVIDIA GeForce 940MX GPU, interfaced with an Impinj R420 UHF RFID reader. The signal transmission and reception system comprises four Larid S9028 square antennas arranged in a $2 \times 2$ array and 35 Impinj H47 UHF omnidirectional tags deployed in a $5 \times 7$ array. The inter-antenna distance is set to 16 cm to balance spatial correlation maximization and grating-lobe-induced interference suppression [3]; the inter-tag distance is fixed at 30 cm to mitigate mutual coupling. Data collection software is implemented using Impinj ItemTest v2.6.0. DLE-DG is implemented in PyTorch 2.4.1 and Python 3.8.13.

**Environment Configuration and Data Collection.** We establish an indoor testbed comprising 16 fingerprint points spaced 0.5 m apart and four furniture placement points (Fig. 2). To rigorously evaluate unseen dynamic-environment generalization, we design four distinct environmental configurations, each defined by systematic variations in furniture quantity and spatial layout. Real scenes of all four environments are provided in Fig. 3.

At each fingerprint point, we collect 600 samples (RSSI and phase) per environment, yielding a total dataset of $4 \times 16 \times 600 = 38{,}400$ samples. For unseen environment generalization evaluation, we use a 4-fold leave-one-environment-out

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29921-5_33

Table 1: 4-fold LOEO cross-validation grouping.

| Fold | Training set | Test set |
|---|---|---|
| Fold1 | Env2,Env3,Env4 | Env1 |
| Fold2 | Env1,Env3,Env4 | Env2 |
| Fold3 | Env1,Env2,Env4 | Env3 |
| Fold4 | Env1,Env2,Env3 | Env4 |

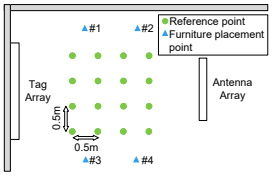
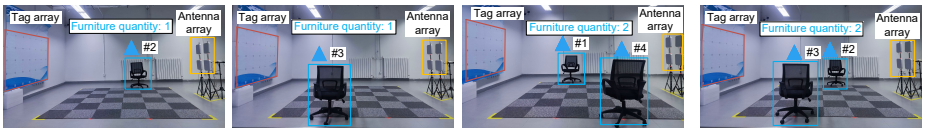



Fig. 2: Layout of fingerprint points and furniture placement points.




(a) Env1 (b) Env2 (c) Env3 (d) Env4

Fig. 3: Real scenes of the four dynamic environments.

(LOEO) cross-validation: in each fold, three environments are used for training, and the remaining environment acts as the unseen environment for testing (Table 1).

**Evaluation Metrics.** We report two complementary metrics: average localization error (ALE) and median error (Med).

ALE quantifies overall accuracy:

$$ALE = \frac{1}{N}\sum_{i=1}^{N} \|\hat{y}_{l_i} - y_{l_i}\|_2^2 \tag{11}$$

where $\hat{y}_{l_i}$ and $y_{l_i}$ are the predicted coordinates and the true coordinates, respectively, and $N$ is the total number of test samples.

Med reflects robustness to outliers and typical-case performance:

$$\text{Med} = \begin{cases} e_{(N+1)/2}, & N \text{ is odd}, \\ \frac{1}{2}\left(e_{N/2} + e_{N/2+1}\right), & N \text{ is even}. \end{cases} \tag{12}$$

where the errors $\{e_i\}$ are sorted in ascending order.

### 4.2 Localization Performance in Static Environments

We first evaluate DLE-DG against state-of-the-art localization baselines under the conventional static environments, such as DSCNN [21], DNN [2], AAResCNN [20], LSTM [8], and AaTs [12]. For each environment (env1–env4), data are split into 80% training and 20% test sets. As shown in Table 2, DLE-DG achieves the lowest average localization error (ALE = 0.023 m) and median error (Med = 0.021 m) among all methods. Its ALE is 36% lower than DSCNN (second-best, 0.036 m) and 83% lower than LSTM (worst-performing, 0.132 m). The

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29921-5_33

Table 2: Localization performance comparison under the static environments (ALE and Med in meters).

| Method | ALE | Med |
|---|---|---|
| DSCNN [21] | 0.036 | 0.025 |
| DNN [2] | 0.088 | 0.061 |
| AAResCNN [20] | 0.067 | 0.038 |
| LSTM [8] | 0.132 | 0.087 |
| AaTs [12] | 0.097 | 0.034 |
| DLE-DG(Ours) | **0.023** | **0.021** |

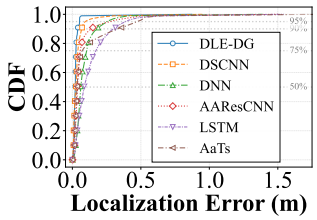



Fig. 4: CDF of localization errors of DLE-DG baselines under the static environments.

cumulative distribution function (CDF) of localization errors (Fig. 4) further confirms superiority: DLE-DG attains the lowest error at all critical percentiles (50th, 75th, 90th, and 95th), demonstrating consistent accuracy and robustness across the error spectrum. These results validate that DLE-DG delivers state-of-the-art performance in static indoor environments.

### 4.3 Generalized Localization Performance in Unseen Environments

To assess unseen dynamic-environments DG, we extend comparisons to DG baselines: DANN [18] and VI-GAN [11]. All models are evaluated under identical 4-fold LOEO (Table 1), where each fold trains on three environments and tests on the held-out unseen environment.

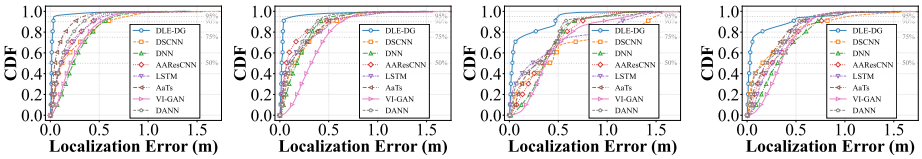
As summarized in Table 3, DLE-DG achieves the lowest ALE in every unseen environment, with an overall average ALE of 0.081 m, reducing error by 63% relative to AaTs (best baseline, 0.221 m) and by 64% relative to DANN (second-best, 0.227 m). DLE-DG also achieves the lowest median errors across all folds: 0.014 m (Fold1), 0.019 m (Fold2), 0.031 m (Fold3), and 0.029 m (Fold4), underscoring high accuracy and outlier resilience in unseen environments. In contrast, traditional deep models (DSCNN, DNN, AAResCNN, LSTM) show poor cross-environment performance: their average ALEs exceed 0.26 m, and degrade sharply versus their performance in static environments (Section 4.2). Critically, in Fold3 and Fold4, baseline methods exhibit severe ALE fluctuations (e.g., AaTs: 0.335 m vs. 0.257 m), whereas DLE-DG maintains remarkable ALE stability (0.123 m vs. 0.128 m). This minimal fluctuation confirms that DLE-DG's performance gains stem not from environment-specific adaptation, but from fundamental decoupling of location and environment factors.

The CDFs in Fig. 5 provide granular insight: DLE-DG consistently achieves steeper curves, indicating higher proportions of low-error predictions. In Fold1, it reaches 95% coverage within 0.05 m, whereas AaTs requires  0.3 m and other baselines fail to reach 90% coverage even at 0.3 m. In Fold2, DLE-DG maintains 90% coverage within 0.05 m; competitors require  0.4 m. In the most challenging Fold3, DLE-DG attains 75% coverage within 0.1 m, while baselines need  0.5 m. Fold4 mirrors this trend. Collectively, these results demonstrate that DLE-DG

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29921-5_33

Table 3: Localization performance under 4-fold LOEO (ALE and Med in meters).

| Method | Fold1 | | Fold2 | | Fold3 | | Fold4 | | Avg. ALE |
|---|---|---|---|---|---|---|---|---|---|
| | ALE | Med | ALE | Med | ALE | Med | ALE | Med | |
| DSCNN [21] | 0.231 | 0.120 | 0.232 | 0.154 | 0.535 | 0.405 | 0.292 | 0.154 | 0.323 |
| DNN [2] | 0.282 | 0.236 | 0.212 | 0.174 | 0.394 | 0.354 | 0.361 | 0.309 | 0.312 |
| AAResCNN [20] | 0.175 | 0.110 | 0.174 | 0.068 | 0.371 | 0.324 | 0.323 | 0.249 | 0.261 |
| LSTM [8] | 0.186 | 0.112 | 0.209 | 0.085 | 0.416 | 0.234 | 0.300 | 0.261 | 0.278 |
| AaTs [12] | 0.094 | 0.047 | 0.199 | 0.110 | 0.335 | 0.281 | 0.257 | 0.197 | 0.221 |
| VI-GAN [11] | 0.256 | 0.209 | 0.352 | 0.322 | 0.462 | 0.370 | 0.367 | 0.344 | 0.359 |
| DANN [18] | 0.146 | 0.102 | 0.178 | 0.123 | 0.332 | 0.299 | 0.252 | 0.181 | 0.227 |
| DLE-DG(Ours) | **0.028** | **0.014** | **0.046** | **0.019** | **0.123** | **0.031** | **0.128** | **0.029** | **0.081** |




(a) Fold1. (b) Fold2. (c) Fold3. (d) Fold4.

Fig. 5: CDF of localization errors for methods under 4-fold LOEO.

delivers robustness and generalization fidelity for RFID fingerprint localization under unseen environments.

### 4.4 Ablation Analysis

We conduct systematic ablation studies on both core components of DLE-DG, the decoupling-oriented fingerprint construction strategy and the domain-adversarially enhanced decoupling network, to rigorously assess the contribution of each module to unseen dynamic-environments DG. All ablations follow the same 4-fold LOEO protocol (Table 1).

**Impact of DFCS.** Here, we mainly evaluate the signal preprocessing in DFCS: missing-read interpolation (MI), RSSI processing (RP), and phase processing (PP). As shown in Table 4, w/o DFCS degrades average ALE from 0.081 m to 0.445 m, confirming its necessity for stable DG performance. w/o RP increases average ALE by 0.196 m, while w/o PP increases it by 0.238 m. This indicates that RSSI and phase carry complementary information, and their joint use can enrich fingerprint robustness across environments. w/o MI increases average ALE by 0.120 m, verifying that unhandled missing data distort signal distribution and undermine generalization.

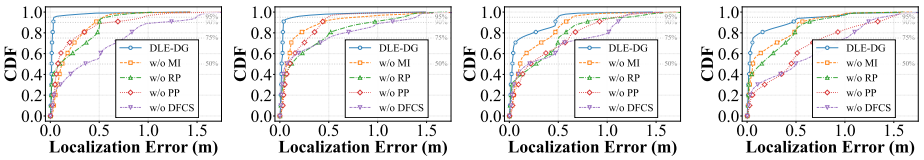
Figure 6 visualizes the CDF degradation: all ablations flatten the curve, but full removal of DFCS collapses performance entirely, confirming that high-quality, stable input is a prerequisite for effective decoupling.

**Impact of DADN.** We evaluate four key design choices in the DADN:

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29921-5_33

Table 4: Ablation of the DFCS under 4-fold LOEO (ALE and Med in meters).

| Method | Fold1 | | Fold2 | | Fold3 | | Fold4 | | Avg. ALE |
|---|---|---|---|---|---|---|---|---|---|
| | ALE | Med | ALE | Med | ALE | Med | ALE | Med | |
| w/o MI | 0.194 | 0.120 | 0.169 | 0.071 | 0.227 | 0.111 | 0.213 | 0.059 | 0.201 |
| w/o RP | 0.211 | 0.094 | 0.300 | 0.119 | 0.342 | 0.267 | 0.253 | 0.145 | 0.277 |
| w/o PP | 0.209 | 0.078 | 0.183 | 0.103 | 0.365 | 0.214 | 0.520 | 0.448 | 0.319 |
| w/o DFCS | 0.468 | 0.351 | 0.326 | 0.069 | 0.412 | 0.199 | 0.572 | 0.498 | 0.445 |
| DLE-DG(Ours) | **0.028** | **0.014** | **0.046** | **0.019** | **0.123** | **0.031** | **0.128** | **0.029** | **0.081** |




(a) Fold1 (b) Fold2 (c) Fold3 (d) Fold4

Fig. 6: CDF of localization errors for DFCS ablations under 4-fold LOEO.

1. w/o ECA: removal of Efficient Channel Attention from the shared representation extractor;
2. w/o DA: removal of the min–max adversarial mechanism from the environment domain-adversarial branch;
3. w/o OR: removal of orthogonal decoupling regularization from the decoupling joint objective function;
4. w/o DR: full removal of the decoupling structure (i.e., no environment domain-adversarial branch and no orthogonality loss).

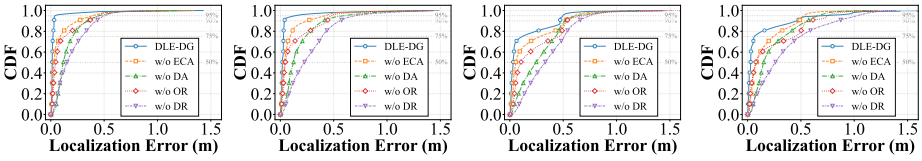
As Table 5 shows, w/o DR causes the largest degradation (+0.214 m average ALE), proving that decoupling is necessary, not optional: without explicit separation of location and environment factors, generalization collapses. w/o DA (+0.135 m) confirms that conventional decoupling is insufficient and exists residual environmental interference in $d_l$. w/o OR (+0.097 m) demonstrates that orthogonality is essential to suppress information leakage. Finally, w/o ECA (+0.051 m) validates that ECA improves representation stability by suppressing environment-sensitive antenna–tag channels early in the representation hierarchy.

Figure 7 reinforces these findings: w/o DR exhibits the heaviest tail and rightmost shift, confirming structural necessity. w/o DA and w/o OR occupy intermediate positions, validating their complementary roles: adversarial learning suppresses environmental interference, while orthogonality enforces representation independence. Together, they form a synergistic decoupling mechanism that cannot be replaced by either component alone.

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29921-5_33

Table 5: Ablation of the DADN under 4-fold LOEO (ALE and Med in meters).

| Method | Fold1 | | Fold2 | | Fold3 | | Fold4 | | Avg. ALE |
|---|---|---|---|---|---|---|---|---|---|
| | ALE | Med | ALE | Med | ALE | Med | ALE | Med | |
| w/o ECA | 0.087 | 0.298 | 0.094 | 0.039 | 0.177 | 0.054 | 0.168 | 0.086 | 0.132 |
| w/o DA | 0.164 | 0.117 | 0.188 | 0.123 | 0.278 | 0.245 | 0.235 | 0.151 | 0.216 |
| w/o OR | 0.113 | 0.047 | 0.142 | 0.047 | 0.235 | 0.103 | 0.222 | 0.077 | 0.178 |
| w/o DR | 0.197 | 0.151 | 0.279 | 0.224 | 0.353 | 0.328 | 0.352 | 0.258 | 0.295 |
| DLE-DG(Ours) | **0.028** | **0.014** | **0.046** | **0.019** | **0.123** | **0.031** | **0.128** | **0.029** | **0.081** |




(a) Fold1. (b) Fold2. (c) Fold3. (d) Fold4.

Fig. 7: CDF of localization errors for DADN ablations under 4-fold LOEO.

## 5 Conclusion

In this paper, we propose DLE-DG, a decoupling RFID fingerprint localization framework for dynamic-environment domain generalization. By explicitly separating RFID fingerprint signals into location and environment factors, DLE-DG produces a location representation with minimal residual environmental interference, directly addressing the core problem of existing methods: severe accuracy degradation in unseen environments due to distribution shifts induced by furniture reconfiguration and layout changes. Our work can be further improved. Real-world deployments involve additional complexities, including human motion, temperature and humidity fluctuations, and reader or antenna reconfiguration. To enhance practical robustness, we plan to extend the benchmark to include cross-room shifts and other environmental dynamics in future work.

**Acknowledgments.** This study was supported by National Key Research and Development Program of China (Grant No.2023YFC3321401).

## References

1. Avellaneda, D., Mendez, D., Fortino, G.: A tinyml deep learning approach for indoor tracking of assets. Sensors **23**(3), 1542 (2023)
2. Azghadi, S.A.R., Mih, A.N., Kawnine, A., Wachowicz, M., Palma, F., Cao, H.: An adaptive indoor localization approach using wifi rssi fingerprinting with slam-enabled robotic platform and deep neural networks. In: 2024 34th International Conference on Collaborative Advances in Software and COmputiNg (CASCON). pp. 1–10. IEEE (2024)
3. Balanis, C.A.: Antenna theory: analysis and design. John wiley & sons (2016)

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29921-5_33

4. Chiasson, D., Lin, Y., Kok, M., Shull, P.B.: Asynchronous hyperbolic uwb source-localization and self-localization for indoor tracking and navigation. IEEE Internet of Things Journal **10**(13), 11655–11668 (2023)
5. Dao, T.H., Nguyen, Q.C., Le, M.T., Hoang, C.A., et al.: Indoor localization system based on passive rfid tags. In: 2014 5th International Conference on Intelligent Systems, Modelling and Simulation. pp. 579–584. IEEE (2014)
6. Gan, X., Zhu, S., Zhang, Y., Wang, Z., Ye, X., Hu, K., Gao, X.: Dual-perspective decoupling network for kidney tumor segmentation on ct images. Neural Networks p. 108042 (2025)
7. Huang, Y., Lui, Z., Ling, G., Lv, S.: An improved bayesian-based rfid indoor location algorithm. Chinese Journal of Electronics **18**(3), 509–512 (2009)
8. Jia, B., Qiao, W., Zong, Z., Liu, S., Hijji, M., Del Ser, J., Muhammad, K.: A fingerprint-based localization algorithm based on lstm and data expansion method for sparse samples. Future Generation Computer Systems **137**, 380–393 (2022)
9. Li, C., Mo, L., Zhang, D.: Review on uhf rfid localization methods. IEEE Journal of Radio Frequency Identification **3**(4), 205–215 (2019)
10. Liu, H., Meng, Z., Li, Z., Zhang, C., Ni, Y., Gao, N., Zhang, Z.: Single-antenna sar rfid system for simultaneous orientation and position sensing in iiot applications. IEEE Internet of Things Journal **12**(23), 51730–51745 (2025)
11. Mallik, M., Chakraborty, S., Sasidhar, K., Chowdhury, C.: Vl-gan: A generative classification approach for fingerprint-based indoor localization. Expert Systems with Applications p. 128400 (2025)
12. Nguyen, S.M., Le, D.V., Havinga, P.J.: Seeing the world from its words: All-embracing transformers for fingerprint-based indoor localization. Pervasive and Mobile Computing **100**, 101912 (2024)
13. Rao, X., Zhou, Q., Yi, Y., Lei, G., Wu, Y., Cao, Y.: Adcloc: A robust and adaptive csi-based device-free passive indoor localization approach for dynamic environments. Digital Signal Processing p. 105568 (2025)
14. Shankar, S., Piratla, V., Chakrabarti, S., Chaudhuri, S., Jyothi, P., et al.: Generalizing across domains via cross-gradient training. ICLR (2018)
15. Wang, J., Lan, C., Liu, C., Ouyang, Y., Qin, T., Lu, W., Chen, Y., Zeng, W., Yu, P.S.: Generalizing to unseen domains: A survey on domain generalization. IEEE transactions on knowledge and data engineering **35**(8), 8052–8072 (2022)
16. Xiao, Y., Bai, X., Chen, B., Su, H., He, H., Xie, L., Yin, E.: Deˆ 2gaze: Deformable and decoupled representation learning for 3d gaze estimation. In: Proceedings of the Computer Vision and Pattern Recognition Conference. pp. 3091–3100 (2025)
17. Xiao, Y., Yang, J., Zhang, S.: Siamese feature decoupling and adaptive prototype alignment for clothes changed person re-identification. IEEE Signal Processing Letters **33**, 181–185 (2025)
18. Xue, M., Xu, Z., Zhang, J., Wang, H., Shen, Y.: A generalization method for indoor localization via domain-invariant feature learning. IEEE Communications Letters (2025)
19. Yue, X., Zhang, Y., Zhao, S., Sangiovanni-Vincentelli, A., Keutzer, K., Gong, B.: Domain randomization and pyramid consistency: Simulation-to-real generalization without accessing target domain data. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 2100–2110 (2019)
20. Zhang, B., Sifaou, H., Li, G.Y.: Csi-fingerprinting indoor localization via attention-augmented residual convolutional neural network. IEEE Transactions on Wireless Communications **22**(8), 5583–5597 (2023)
21. Zhu, H., Cheng, L., Li, X., Yuan, H.: Neural-network-based localization method for wi-fi fingerprint indoor localization. Sensors **23**(15), 6992 (2023)

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29921-5_33