

DAG-TransLoc: Domain-Generalized Indoor Localization with Variable Anchor Configurations

Dhriti M Goenka, Abhishek Jindal, Manish Kumar

School of Technology, Dhirubhai Ambani University (Formerly DAICT), Gandhinagar, Gujarat, India

Email: {202201041, abhishek_jindal, manish_kumar}@dau.ac.in

Abstract—Deploying indoor localization systems across real-world environments fails not because models lack accuracy, but because they are brittle — trained for one floor plan, one anchor layout, one signal profile. We present Domain Aware Generalized Transformer based Localization (DAG-TransLoc) that addresses this fundamental deployment gap by jointly handling domain shifts and variable anchor configurations within a single unified architecture. The model uses self-attention to fuse multi-anchor CIR features with a masking mechanism that gracefully handles missing or variable anchors — eliminating the need for architectural changes across deployments. To learn representations that survive environment changes, we introduce feature decoupling with domain-aware attention, separating invariant propagation characteristics from environment-specific ones. A ray-tracing-based pipeline using Sionna and WiSegRT enables systematic evaluation across geometry, anchor density, and receiver sampling variations. Under moderate domain shifts, the framework achieves up to 98.3% accuracy within 1m, with RMSE as low as 0.275m, and under challenging combined BS and geometry shifts, it achieves 95.8% accuracy at 0.493m RMSE. Performance is strongly governed by training-target domain compatibility; complementary domain combinations generalize well, while highly heterogeneous or mismatched combinations can substantially degrade accuracy, underscoring that domain-aware training composition, not diversity alone, is key to robust generalization.

Index Terms—Indoor localization, Positioning, Transformer, Domain generalization, CIR, RSRP.

I. INTRODUCTION

Accurate indoor localization is a key enabler for applications such as smart buildings, robotics, augmented reality, and next-generation wireless systems. However, unlike outdoor environments where GPS is reliable, indoor localization must contend with severe multipath propagation, signal attenuation, and non-line-of-sight (NLoS) conditions [1]. These challenges are particularly pronounced in sub-6 GHz bands (e.g., 2.4 GHz), where richer multipath and stronger penetration produce highly environment-dependent channel characteristics, making learned models difficult to transfer across deployments. In practice, real-world deployments exhibit substantial variability in environmental geometry, anchor density, and receiver sampling, leading to significant domain shifts that severely degrade model performance.¹

¹In this work, domain shift is instantiated as geometry, anchor-density, and receiver-sampling variation within a single physical environment; extending evaluation to distinct buildings or scenes is discussed as future work in Section VII.

While deep learning-based approaches have significantly improved localization accuracy by learning complex mappings from wireless signals such as CSI, CIR, and ToA [2]–[4], they each address only a subset of the challenges in real deployments. Transformer-based models such as Swin-Loc [5] improve multi-anchor feature fusion through attention mechanisms but assume a fixed environment and anchor layout. Domain generalization methods address environment variability through feature disentanglement [6], [7] or adversarial alignment [8], [9], but are evaluated under single-anchor or fixed-anchor settings. Domain-aware attention has been explored for single-anchor UWB localization [10], demonstrating its utility for cross-environment transfer, but does not address variable anchor configurations or multi-anchor fusion. To the best of our knowledge, no existing work jointly handles environment shifts, variable anchor counts, and receiver distribution shifts within a single architecture.

To this end, we propose DAG-TransLoc, a transformer-based indoor localization framework specially designed to eliminate trade-off between generalization and flexibility. The model leverages self-attention to capture inter-anchor relationships while employing a masking mechanism to support a variable number of anchors. In addition, we construct a ray-tracing-based cross-domain evaluation setup using Sionna [11] and WiSegRT [12]. The main contributions of this work are summarized as follows:

- We propose DAG-TransLoc, a transformer-based indoor localization framework for domain generalization that jointly addresses geometry, anchor density, and receiver distribution shifts in sub-6 GHz environments.
- We introduce a masking-based multi-anchor fusion mechanism with domain-aware attention for robust localization under heterogeneous deployment configurations and varying anchor availability.
- We develop a ray-tracing-driven evaluation framework to generate controlled cross-domain scenarios and systematically analyze generalization performance under realistic wireless propagation conditions.

The remainder of this paper is organized as follows. Sections II and III describe the system model and dataset generation process. Section IV presents the proposed DAG-TransLoc framework, while Section V discusses the training strategy. Experimental results are presented in Section VI, followed by the conclusion in Section VII.

II. SYSTEM MODEL

We consider an indoor wireless localization scenario where multiple anchors are deployed within a structured environment, and we estimate the 2D position of a receiver based on observed wireless channel characteristics.

A. Environment and Propagation Model

The indoor environment is modeled using realistic geometric layouts derived from the WiSegRT dataset [12], where each scene is represented using XML-based descriptions containing structural elements and objects. Wireless propagation is simulated using the Sionna RT ray-tracing framework [11], which models line-of-sight (LoS), reflection, diffraction, and scattering to capture complex multipath interactions between anchor–receiver pairs. Sionna RT enables reproducible geometry-consistent domain variations, while WiSegRT provides realistic XML-based scene descriptions.

The system operates at a carrier frequency of 2.4 GHz. Both anchors and receivers are modeled using planar antenna configurations, where anchors employ vertically polarized dipole antennas and receivers use isotropic antennas with cross-polarization. Multipath propagation is computed up to a maximum depth of 4, ensuring accurate modeling of higher-order interactions.

B. Channel Model and Feature Representation

For a given anchor–receiver pair, the wireless channel is modeled as a multipath channel impulse response (CIR) [13]:

$$h(t) = \sum_{k=1}^K a_k \delta(t - \tau_k) \quad (1)$$

where $a_k \in \mathbb{C}$ and τ_k denote the complex gain and propagation delay of the k -th path, respectively, and K is the number of multipath components. Since the number of multipath components varies across links, the CIR is converted into a fixed-dimensional representation. The ray-tracing process $\mathcal{P} = \{(a_k, \tau_k)\}_{k=1}^K$ is discretized as:

$$h[n] = \sum_{k=1}^K a_k \cdot \mathbf{1} \left(\left| \frac{\tau_k}{\Delta\tau} - n \right| < \frac{1}{2} \right) \quad (2)$$

where $\Delta\tau = T_{\max}/L$ is the delay resolution, $L = 256$ is the number of delay bins, and $\mathbf{1}(\cdot)$ is the indicator function. The maximum delay window $T_{\max}$ is estimated using a percentile-based approach to retain the majority of multipath energy while avoiding excessive sparsity. This discretized representation ensures consistent input dimensionality across all samples and enables efficient neural network processing.

Assuming that the system consists B anchors, the CIRs from all anchors corresponding to a receiver location are aggregated as $\mathbf{X}_i \in \mathbb{C}^{B \times L}$ where each row corresponds to one anchor–receiver link. In addition to CIR, received signal strength (RSRP) is computed as:

$$\text{RSRP} = 10 \log_{10} \left(\sum_{n=1}^L |h[n]|^2 \right) \quad (3)$$

which provides a coarse measure of signal power and complements the detailed multipath structure captured by the CIR.

III. DATASET GENERATION

To enable domain generalization evaluation, we utilize Scene 5 from the WiSegRT dataset [12], which represents an indoor environment of approximately $6 \text{ m} \times 4 \text{ m} \times 2.4 \text{ m}$. The scene contains diverse objects² which create rich multipath propagation characteristics.

To study domain generalization, each dataset variant is treated as a separate domain associated with a distinct joint distribution $P_i(\mathbf{X}, \mathbf{y})$, where $\mathbf{X}$ represents the multi-anchor wireless input composed of CIR and auxiliary signal features, and $\mathbf{y}$ denotes the receiver location. Multiple domains are constructed by varying environmental geometry, anchor density, and receiver sampling using the Sionna RT simulation framework [11]. Depending on the experimental setting, different subsets of these domains are designated as source domains $\{\mathcal{D}_s^i\}_{i=1}^{N_s}$ and unseen target domains $\mathcal{D}_t$ for evaluating cross-domain generalization performance.

Environmental variations are created by editing the XML-based scene description to remove selected objects. In some variants, large objects such as cupboards and tables are removed, while in others, smaller objects such as televisions and lamps are eliminated. These modifications alter the multipath environment and simulate realistic changes in indoor layouts.

Anchor density variations reflect practical deployment constraints. While the original scene contains $T = 18$ anchors, reduced configurations with $T = 10$ and $T = 8$ anchors are generated. To construct reduced anchor sets while preserving spatial coverage, structured selection strategies are employed instead of random sub-sampling. First, a greedy max-min distance selection strategy iteratively selects anchors that maximize the minimum distance to already selected anchors, ensuring maximal spatial spread. Additionally, a grid-based strategy partitions the deployment area into spatial cells with at most one anchor per cell, preventing clustering and ensuring uniform coverage. These strategies maintain sufficient spatial diversity essential for reliable localization and enable systematic analysis of anchor density effects on model generalization. *Receiver sampling variations* modify the spatial grid interval from 0.2 m to 0.25 m, creating different spatial distributions of receiver positions.

A. Dataset Domain Analysis/Domain Characterization

To verify that the constructed domains represent meaningful variations, we analyze statistical differences across the generated datasets.

Environmental Geometry: Fig. 1, Fig. 2, and Fig. 3 show the effect of object removal on multipath propagation. The modified environment ($\mathcal{D}_{v3}$) exhibits a stronger dominant LoS component and reduced energy in later delay taps, reflecting

²The scene contains both wall-touching and centrally placed objects, including large and small cabinets, curtains, windows, a central chair-table setup, and a television unit with a lamp, creating both LoS and NLoS propagation conditions.

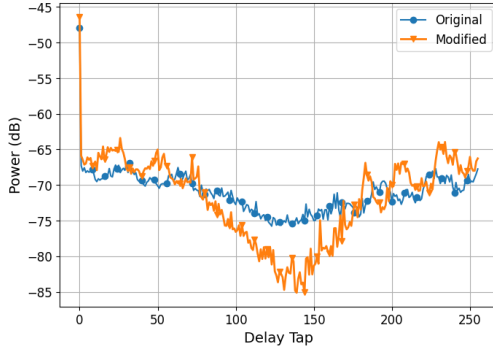


Fig. 1: Average power delay profile (PDP) for the original and modified ($\mathcal{D}_{v3}$) environments.

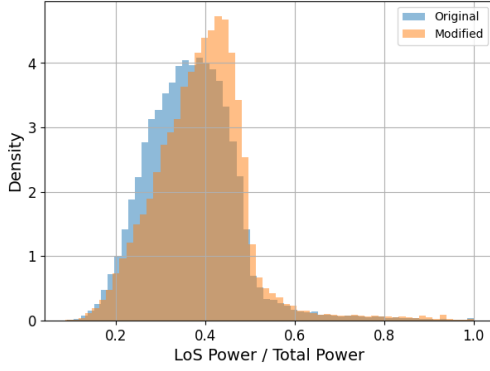


Fig. 2: LoS power ratio distribution for the original and modified environments.

weaker secondary reflections. This is further supported by a higher LoS power ratio (Fig. 2) and a reduced delay spread (peak ~ 22 vs. 28 taps), indicating loss of late-arriving paths. These changes confirm that geometry modifications significantly alter propagation characteristics.

Anchor Density: Figures 4 and 5 analyze the effect of reducing anchor density (18 to 10). The 10 anchor configuration produces a more concentrated RSRP distribution due to increased average link distances, whereas 18 anchor exhibits greater variability. Coverage analysis shows that while both configurations maintain high connectivity at low thresholds ($\sim 93\%$), performance diverges at stricter thresholds, where 10 anchor degrades more rapidly. This indicates that reduced anchor density not only limits anchor availability but also increases sensitivity to signal quality constraints, forming a distinct domain shift.

Receiver Sampling Distribution: Figure 6 shows that channel variability distributions remain nearly identical across grid spacings (0.15 m, 0.20 m, 0.25 m). This confirms that receiver variation represents a pure spatial distribution shift, with no change in underlying propagation statistics.

IV. PROPOSED METHOD

We propose a transformer-based framework for domain-generalized indoor localization that supports variable anchor

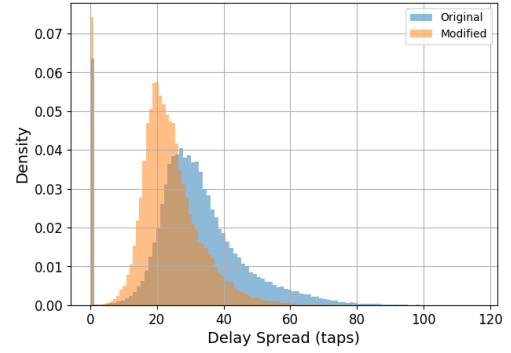


Fig. 3: Delay spread distribution for the original and modified environments.

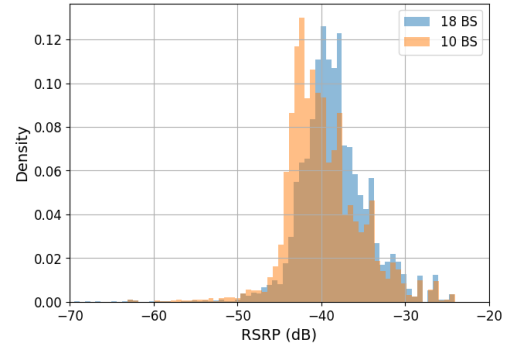


Fig. 4: RSRP distribution for 18 anchor and 10 anchor configurations under identical scene geometry.

configurations. The overall pipeline is illustrated in Fig. 7. Given wireless channel measurements, the objective is to estimate the receiver's coordinates while learning feature representations that remain robust to domain shifts. As shown in Fig. 7, the framework consists of four components: feature representation, transformer-based multi-anchor fusion, feature decoupling, and domain-aware attention, followed by a regression head.

A. Feature Representation

As depicted in the leftmost block of Fig. 7, each anchor provides both CIR and RSRP measurements. For the i -th anchor, the CIR $\mathbf{x}_i \in \mathbb{C}^{1 \times L}$ is processed using a 1D CNN $\mathbf{f}_i^{\text{CIR}} = \text{CNN}_{\text{CIR}}(\mathbf{x}_i) \in \mathbb{R}^{1 \times d_{\text{CIR}}}$. The corresponding RSRP measurement for the i -th anchor is lifted into a higher dimensional embedding via a shallow MLP $\mathbf{f}_i^{\text{aux}} = \text{MLP}(r_i), r_i \in \mathbb{R}, \mathbf{f}_i^{\text{aux}} \in \mathbb{R}^{1 \times d_{\text{aux}}}$. This nonlinear projection allows the model to learn expressive representations of signal strength beyond a simple linear scaling. The extracted features are concatenated and projected into a shared space:

$$\mathbf{z}_i = \mathbf{W}_z[\mathbf{f}_i^{\text{CIR}} \parallel \mathbf{f}_i^{\text{aux}}], \quad \mathbf{z}_i \in \mathbb{R}^{1 \times d_{\text{model}}} \quad (4)$$

These representations form per-anchor tokens, as shown in the *Token Formation* stage of Fig. 7.

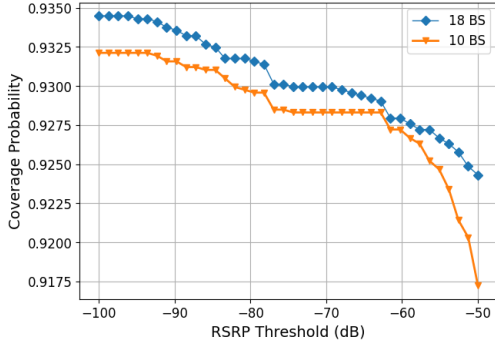


Fig. 5: Coverage probability versus RSRP threshold for 18 BS and 10 BS configurations.

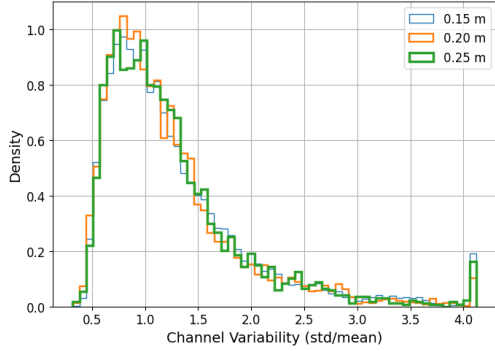


Fig. 6: Channel variability distribution across receiver grid spacings of 0.15 m, 0.20 m, and 0.25 m.

B. Transformer-Based Multi-Anchor Fusion

To incorporate spatial geometry, anchor coordinates $\mathbf{b}_i \in \mathbb{R}^{1 \times 3}$ are encoded and added to the feature tokens:

$$\tilde{\mathbf{z}}_i = \mathbf{z}_i + \mathbf{W}_p \mathbf{b}_i, \quad \tilde{\mathbf{z}}_i \in \mathbb{R}^{1 \times d_{model}} \quad (5)$$

As illustrated in the *Multi-anchor Fusion* block of Fig. 7, the tokens are processed using a transformer encoder [14]. To support varying numbers of anchors, inputs are padded to $B_{\max}$ and a binary mask $\mathbf{m}$ is used to ignore invalid entries. A CLS token aggregates global information:

$$\mathbf{h} = \text{Transformer}([\mathbf{z}_{\text{cls}}, \tilde{\mathbf{z}}_1, \dots], \mathbf{m}) \quad (6)$$

This enables the model to capture inter-anchor relationships while remaining robust to varying deployment configurations. The CLS token is initialized as a learned parameter $\mathbf{z}_{\text{cls}} \in \mathbb{R}^{1 \times d_{model}}$ with zeros and optimized jointly during training.

C. Feature Decoupling

Following the transformer encoder, the global feature representation $\mathbf{h}$ is decomposed into domain-invariant and domain-specific components $\mathbf{z}_{\text{inv}} = f_{\text{inv}}(\mathbf{h})$, $\mathbf{z}_{\text{spec}} = f_{\text{spec}}(\mathbf{h})$ where $f_{\text{inv}}(\cdot)$ and $f_{\text{spec}}(\cdot)$ are implemented as two independent multilayer perceptron (MLP) projection heads, each consisting of stacked linear layers with nonlinear activations. These projection networks transform the shared transformer embedding into separate feature subspaces, enabling complementary

representations while preserving the expressive capacity of the transformer features. Finally, both embeddings are ℓ_2 -normalized to stabilize optimization and encourage structured feature learning in the latent space.

D. Domain-Aware Attention and Regression Head

Although domain-invariant features promote generalization, relying solely on them may discard useful environment-specific information. To address this, we introduce a domain-aware attention mechanism that incorporates domain context in a controlled manner (see Fig. 7). For each domain d , a prototype is computed from domain-specific features, where $|\mathcal{D}_d|$ denotes the number of samples in domain d :

$$\mathbf{p}_d = \frac{1}{|\mathcal{D}_d|} \sum_{i \in \mathcal{D}_d} \mathbf{z}_{\text{spec}}^{(i)} \quad (7)$$

This prototype captures characteristic propagation patterns of the domain. We then compute attention between invariant features and the domain prototype using learnable projections:

$$\mathbf{q}_d = \mathbf{W}_Q \mathbf{p}_d, \quad \mathbf{k}_i = \mathbf{W}_K \mathbf{z}_{\text{inv}}^{(i)}, \quad \mathbf{v}_i = \mathbf{W}_V \mathbf{z}_{\text{inv}}^{(i)} \quad (8)$$

The attention scores are given by

$$s_i = \frac{\mathbf{k}_i^\top \mathbf{q}_d}{\sqrt{D}}, \quad \alpha_i = \frac{\exp(s_i)}{\sum_j \exp(s_j)} \quad (9)$$

Using these weights, a domain context vector is obtained $\mathbf{c}_d = \sum_i \alpha_i \mathbf{v}_i$. Finally, the invariant feature is adaptively fused with the domain context using a gating mechanism:

$$\mathbf{g}_i = \sigma(\mathbf{W}_g[\mathbf{z}_{\text{inv}}^{(i)} \parallel \mathbf{c}_d]), \quad \mathbf{z}_i^{\text{att}} = \mathbf{g}_i \odot \mathbf{z}_{\text{inv}}^{(i)} + (1 - \mathbf{g}_i) \odot \mathbf{c}_d \quad (10)$$

At inference time, domain labels are unavailable; therefore, the batch-level mean of domain-specific features is used as a pseudo-prototype³, allowing the attention mechanism to adapt to the target distribution without explicit supervision.

Finally, as shown in the rightmost block of Fig. 7, the refined representation $\mathbf{z}^{\text{att}}$ is passed through a fully connected regression network to predict the receiver coordinates $\hat{\mathbf{y}} = f_{\text{reg}}(\mathbf{z}^{\text{att}})$ where $\hat{\mathbf{y}} \in \mathbb{R}^{1 \times 2}$ denotes the estimated position.

V. TRAINING STRATEGY

We train the proposed model to achieve both accurate localization and robustness across domains. This is accomplished using a balanced multi-domain training scheme and a combination of localization and domain generalization objectives.

³This construction assumes batch-level domain homogeneity; for a single query or a domain-mixed batch, the pseudo-prototype reduces to a less-informative global average. Adaptive prototype estimation independent of batch composition is left to future work.

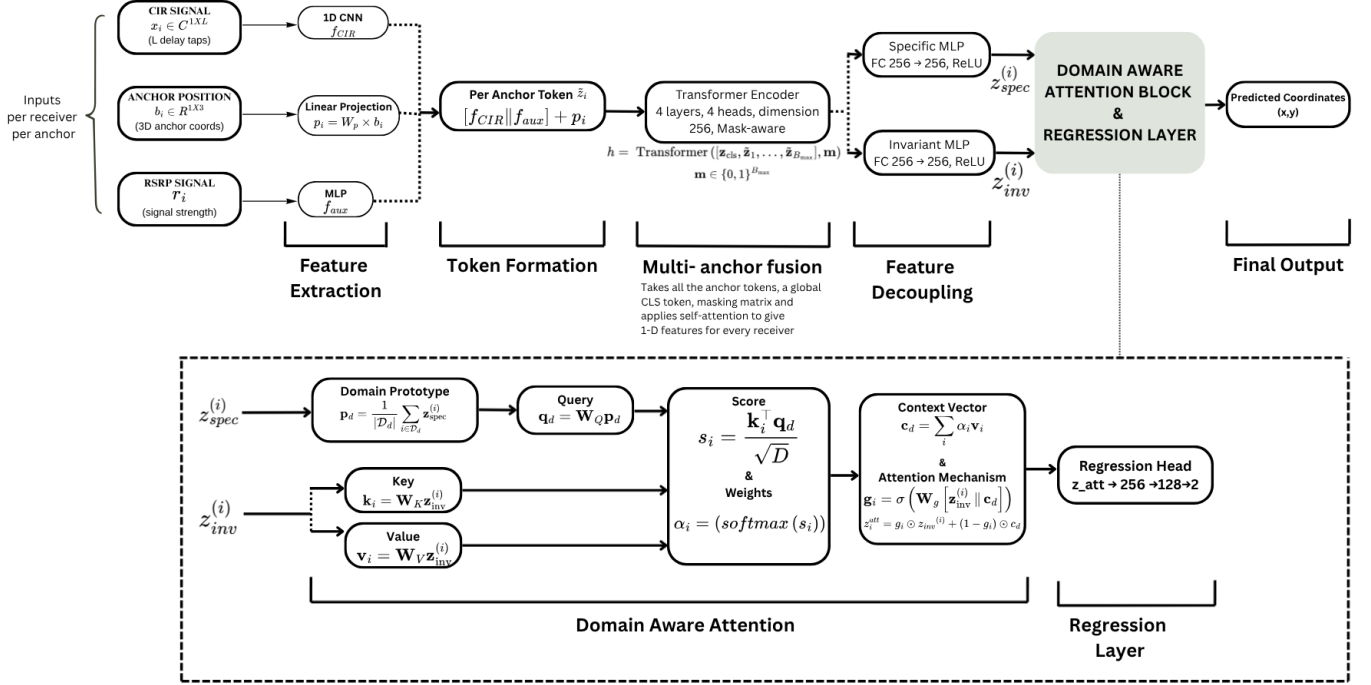


Fig. 7: Overview of the proposed transformer-based domain-generalized indoor localization framework.

A. Balanced Multi-Domain Training

We consider multiple source domains $\{\mathcal{D}_1, \dots, \mathcal{D}_{N_s}\}$, each representing a distinct environment. To prevent bias toward domains with more samples, we adopt a balanced sampling strategy. Given a batch size B , each mini-batch is constructed by sampling an equal number of samples from every domain:

$$B_d = \frac{B}{N_s}, \quad \mathcal{B} = \bigcup_{d=1}^{N_s} \mathcal{B}_d \quad (11)$$

where B_d denotes the number of samples drawn from domain $\mathcal{D}_d$, and $\mathcal{B}_d$ represents the corresponding sample set. This ensures uniform domain contribution and enables stable computation of domain alignment losses.

B. Training Objectives and Optimization

The model is trained using a combination of localization and domain generalization losses. The primary objective is the localization loss $\mathcal{L}_{\text{reg}} = \text{SmoothL1}(\hat{\mathbf{y}}, \mathbf{y})$, operating on normalized coordinates. To additionally penalize large errors in meters, two auxiliary losses are included: a mean distance loss $\mathcal{L}_{\text{mean}}$ over all batch samples, and a tail loss $\mathcal{L}_{\text{tail}}$ restricted to the worst $k = \lfloor 0.2B \rfloor$ predictions, where $\mathcal{T}_k$ denotes their indices:

$$\mathcal{L}_{\text{mean}} = \frac{1}{B} \sum_{i=1}^B \|\hat{\mathbf{y}}_i - \mathbf{y}_i\|_2, \quad \mathcal{L}_{\text{tail}} = \frac{1}{k} \sum_{i \in \mathcal{T}_k} \|\hat{\mathbf{y}}_i - \mathbf{y}_i\|_2 \quad (12)$$

Domain invariance is enforced via an adversarial classification loss $\mathcal{L}_{\text{cls}} = \text{CE}(f_{\text{cls}}(\mathbf{z}_{\text{inv}}), d)$ and an MMD [15] alignment loss over all domain pairs $\mathcal{P} = \{(i, j) : i < j\}$:

$$\mathcal{L}_{\text{MMD}} = \frac{1}{|\mathcal{P}|} \sum_{(i,j) \in \mathcal{P}} \text{MMD}^2(\mathbf{z}_{\text{inv}}^{(i)}, \mathbf{z}_{\text{inv}}^{(j)}) \quad (13)$$

To encourage decoupling, a separation term is maximized by negating it in the total loss.

$$\mathcal{L}_{\text{sep}} = \frac{1}{|\mathcal{D}|} \sum_d \text{MMD}^2(\mathbf{z}_{\text{inv}}^{(d)}, \mathbf{z}_{\text{spec}}^{(d)}) \quad (14)$$

The total loss is:

$$\mathcal{L} = \mathcal{L}_{\text{reg}} + \alpha \mathcal{L}_{\text{mean}} + \beta \mathcal{L}_{\text{tail}} + (\lambda_1 \mathcal{L}_{\text{cls}} + \lambda_2 \mathcal{L}_{\text{MMD}} - \lambda_3 \mathcal{L}_{\text{sep}}) \quad (15)$$

Domain losses are ramped up via $r(t) = \min(1, \frac{t-5}{15})$ after epoch 5, with $\alpha=0.05$, $\beta=0.1$, $\lambda_1=0.05$, $\lambda_2=\lambda_3=0.02$. The model is optimized using AdamW [16] with learning rate 2×10^{-4} , with 5 warm-up epochs of linear scaling followed by cosine annealing decay.

VI. EXPERIMENTAL RESULTS

We evaluate the proposed framework under multiple domain generalization scenarios, analyzing the effects of environmental geometry, anchor density, and receiver distributions.

A. Experimental Setup

We construct multiple domains from WiSegRT Scene 5 [12] using Sionna RT [11]. The following domains are defined:

- $\mathcal{D}_{\text{orig}}$: Original scene with full geometry and 18 BS.

TABLE I: Generalization Performance on $\mathcal{D}_{v4}$

Training Domains	RMSE	Median	P90	Acc < 1m
$\mathcal{D}_{\text{orig}} + \mathcal{D}_{v1} + \mathcal{D}_{v2}$	0.954	0.720	1.280	76.0%
$\mathcal{D}_{\text{orig}} + \mathcal{D}_{v2} + \mathcal{D}_{v3}$	1.465	1.002	2.297	49.7%
$\mathcal{D}_{v1} + \mathcal{D}_{v2} + \mathcal{D}_{v3}$	1.230	0.903	1.775	56.8%

TABLE II: Generalization under Moderate Geometry Shifts

Train $\rightarrow$ Test	RMSE (m)	Median(m)	P90(m)	Acc < 1m
$\mathcal{D}_{\text{orig}} + \mathcal{D}_{v1} + \mathcal{D}_{v2} \rightarrow \mathcal{D}_{v3}$	0.298	0.088	0.450	97.8%
$\mathcal{D}_{v1} + \mathcal{D}_{v2} + \mathcal{D}_{v4} \rightarrow \mathcal{D}_{v3}$	0.275	0.091	0.381	98.28%
$\mathcal{D}_{v1} + \mathcal{D}_{v3} + \mathcal{D}_{v4} \rightarrow \mathcal{D}_{v2}$	0.298	0.112	0.319	98.3%
$\mathcal{D}_{\text{orig}} + \mathcal{D}_{v2} + \mathcal{D}_{v3} \rightarrow \mathcal{D}_{v1}$	0.501	0.081	0.254	97.5%
$\mathcal{D}_{v1} + \mathcal{D}_{v3} + \mathcal{D}_{v4} \rightarrow \mathcal{D}_{\text{orig}}$	0.289	0.115	0.346	98.46%

- $\mathcal{D}_{v1}$: Modified geometry with selected large objects removed (18 BS).
- $\mathcal{D}_{v2}$: Small object removal configuration (18 BS).
- $\mathcal{D}_{v3}$: Multiple objects removed (18 BS).
- $\mathcal{D}_{v4}$: Reduced BS configuration (8 BS) with significant object removal (combined BS and geometry shift).
- $\mathcal{D}_{v5}$: Reduced BS configuration (10 BS) with original geometry preserved (isolated BS sparsity).
- $\mathcal{D}_{r15}$: Receiver grid with finer spacing (0.15 m)
- $\mathcal{D}_{r25}$: Receiver grid with coarser spacing (0.25 m)

All domains contain 5,510 receiver locations, except $\mathcal{D}_{r15}$ (14,625) and $\mathcal{D}_{r25}$ (2,760). Models are trained using the multi-domain strategy described in Section V and evaluated using Euclidean localization error. We report RMSE, median, P90, and accuracy within 1 m to capture both average and worst-case behavior. Although RMSE and sub-meter accuracy remain strong across most experiments, the P90 metric is higher in cross-domain settings due to a small number of challenging samples under severe propagation mismatch. In these cases, most predictions remain accurate, while a limited set of difficult receiver locations with altered multipath or reduced anchor visibility dominate the tail-error statistics [17].

B. Generalization Across Environmental Variations

We first evaluate generalization across geometry variations while keeping anchor configuration fixed (18 anchors). Results on $\mathcal{D}_{v4}$ (Table I) show a clear degradation due to reduced multipath richness caused by object removal. Despite this, the model maintains around 75% accuracy within 1 m, indicating robustness under severe domain shifts. For moderate geometry variations (Table II), the model achieves consistently strong performance, with sub-meter RMSE and accuracy above 97%. Notably, high accuracy is also maintained when evaluating on $\mathcal{D}_{\text{orig}}$ after training on modified domains, indicating that the model learns transferable representations without overfitting to simplified environments.

C. Impact of Training Domain Diversity

We analyze how the number and composition of training domains affect generalization to $\mathcal{D}_{v4}$. Tables III and IV summarize the results. When using only two training domains (Table III), certain combinations outperform more diverse

TABLE III: Two-Domain Generalization to $\mathcal{D}_{v4}$

Training Domains	RMSE (m)	Median (m)	P90 (m)	Acc < 1m (%)
$\mathcal{D}_{\text{orig}} + \mathcal{D}_{v2}$	0.802	0.652	1.058	86.2%
$\mathcal{D}_{v1} + \mathcal{D}_{v2}$	0.933	0.737	1.190	77.7%

TABLE IV: Effect of Increasing Training Domains on $\mathcal{D}_{v4}$

Training Domains	RMSE (m)	Median(m)	P90(m)	Acc
$\mathcal{D}_{\text{orig}} + \mathcal{D}_{v1}$	0.909	0.695	1.332	75.5%
$\mathcal{D}_{\text{orig}} + \mathcal{D}_{v1} + \mathcal{D}_{v2}$	0.954	0.720	1.280	76.0%
$\mathcal{D}_{\text{orig}} + \mathcal{D}_{v1} + \mathcal{D}_{v2} + \mathcal{D}_{v3}$	1.430	1.294	2.006	30.8%

settings. For example, $\mathcal{D}_{\text{orig}} + \mathcal{D}_{v2}$ achieves better accuracy than other pairs. This indicates that domain compatibility is more important than simply increasing diversity. When domains are too dissimilar, the model struggles to align conflicting propagation patterns, leading to higher tail errors. As the number of training domains increases (Table IV), performance initially remains stable, but begins to degrade as more heterogeneous domains are introduced. While accuracy is comparable between two and three domains, a significant drop is observed when all domains are included, along with a sharp increase in P90 error. These results highlight a key trade-off: while diversity aids generalization, excessive heterogeneity destabilizes alignment, particularly in tail errors. For moderate shifts, all configurations exceed 96% accuracy, confirming that degradation is driven by mismatch severity rather than domain count.

D. Base Station and Geometry Generalization

We analyze the combined impact of BS density and geometry. Results are summarized in Table V. Two key observations emerge. First, generalization from high to low BS density is feasible but depends on the training distribution. Training on $\mathcal{D}_{\text{orig}} + \mathcal{D}_{v2}$ leads to performance degradation on $\mathcal{D}_{v4}$ due to reduced anchors and simplified geometry. However, adding $\mathcal{D}_{v5}$ significantly improves performance, indicating that intermediate BS configurations help bridge dense and sparse deployments and improve robustness. Second, generalization is asymmetric. Models trained only on reduced-BS domains ($\mathcal{D}_{v4} + \mathcal{D}_{v5}$) perform poorly on $\mathcal{D}_{\text{orig}}$, failing to capture the richer multipath structure of dense environments. This suggests that exposure to complex propagation patterns is essential for learning transferable representations. Overall, both BS configuration and geometry jointly affect generalization. Strong performance is achieved when training domains provide complementary variations, whereas limited structural diversity or insufficient complexity leads to degraded accuracy, particularly in tail errors.

These results also validate the masking-based architecture. The model generalizes across configurations ranging from 18 anchors ($\mathcal{D}_{\text{orig}}$) to 8 anchors ($\mathcal{D}_{v4}$) without architectural changes or retraining. The strong performance retained across these settings, including 95.8% accuracy on $\mathcal{D}_{v4}$, demonstrates that the masking mechanism effectively handles anchor sparsity while preserving learned inter-anchor relationships.

TABLE V: Results for BS and Geometry Generalization

Train $\rightarrow$ Test	RMSE (m)	Median (m)	P90 (m)	Acc < 1m
$\mathcal{D}_{\text{orig}} + \mathcal{D}_{v2} \rightarrow \mathcal{D}_{v4}$	0.802	0.652	1.058	86.2%
$\mathcal{D}_{\text{orig}} + \mathcal{D}_{v5} \rightarrow \mathcal{D}_{v4}$	0.493	0.286	0.720	95.8%
$\mathcal{D}_{v4} + \mathcal{D}_{v5} \rightarrow \mathcal{D}_{\text{orig}}$	1.982	1.837	2.799	16.2%
$\mathcal{D}_{\text{orig}} + \mathcal{D}_{v5} \rightarrow \mathcal{D}_{v3}$	0.304	0.172	0.451	98.3%

TABLE VI: Results for Receiver Distribution Generalization

Train $\rightarrow$ Test	RMSE (m)	Median (m)	P90 (m)	Acc < 1m
$\mathcal{D}_{\text{orig}} + \mathcal{D}_{r25} \rightarrow \mathcal{D}_{r15}$	0.452	0.095	0.261	97.6%
$\mathcal{D}_{\text{orig}} + \mathcal{D}_{r15} \rightarrow \mathcal{D}_{r25}$	0.397	0.056	0.152	98.5%
$\mathcal{D}_{r25} + \mathcal{D}_{v2} \rightarrow \mathcal{D}_{v3}$	0.307	0.106	0.443	97.5%
$\mathcal{D}_{r15} + \mathcal{D}_{v3} \rightarrow \mathcal{D}_{v2}$	0.498	0.076	0.218	97.5%

TABLE VII: Performance under Mixed Domain Shifts (Target: $\mathcal{D}_{v4}$)

Training Domains	Shift Type	RMSE (m)	Acc < 1m
$\mathcal{D}_{\text{orig}} + \mathcal{D}_{v2}$	Geometry	0.802	86.2%
$\mathcal{D}_{\text{orig}} + \mathcal{D}_{v3}$	Severe Geometry	1.264	54.9%
$\mathcal{D}_{v2} + \mathcal{D}_{v3}$	Geometry	0.871	78.2%
$\mathcal{D}_{\text{orig}} + \mathcal{D}_{v5}$	Anchor sparsity	0.493	95.8%
$\mathcal{D}_{v5} + \mathcal{D}_{v2}$	Anchor sparsity + Geometry	0.463	96.6%
$\mathcal{D}_{r25} + \mathcal{D}_{\text{orig}}$	Receiver + Geometry	0.800	80.1%
$\mathcal{D}_{r25} + \mathcal{D}_{v2}$	Receiver + Geometry	1.062	76.5%

E. Receiver Distribution Generalization

We next analyze receiver distribution shifts (Table VI). When only sampling density varies, the model generalizes well across grids, with slightly better performance when trained on finer resolutions due to improved spatial coverage. When combined with environmental variations, performance remains strong for moderate shifts, indicating that the model can separate spatial sampling effects from propagation characteristics. However, under severe domain mismatch (e.g., $\mathcal{D}_{v4}$), performance degrades, showing that geometry remains the dominant factor.

F. Mixed Domain Shift: Unified Evaluation

Finally, we evaluate combined domain shifts (Table VII). The results show that performance is largely governed by the most severe mismatch between training and target domains. Training configurations that include intermediate domains, particularly $\mathcal{D}_{v5}$, consistently achieve the best performance, as they help bridge structural differences between domains. Receiver distribution shifts alone have a relatively smaller impact, but when combined with environmental changes, they contribute to additional degradation. Overall, the results demonstrate that the proposed framework achieves strong robustness under mixed domain conditions, with best performance obtained when training domains are complementary and aligned with key characteristics of the target environment.

G. Ablation Study

To evaluate the contribution of the proposed domain-generalization components, we selectively remove domain-aware attention and the individual domain-generalization loss terms, namely $\mathcal{L}_{\text{cls}}$, $\mathcal{L}_{\text{MMD}}$, and $\mathcal{L}_{\text{sep}}$, while keeping all other

TABLE VIII: Ablation Study under Moderate Shift ($\rightarrow \mathcal{D}_{v3}$)

Configuration	RMSE (m)	P90 (m)	Acc < 1m (%)
Full Model	0.295(13)	0.429(17)	97.8(0.5)
w/o Dom. Attn.	0.285(18)	0.432(40)	98.1(0.4)
w/o $\mathcal{L}_{\text{cls}}$	0.290(9)	0.427(3)	97.9(0.4)
w/o $\mathcal{L}_{\text{MMD}}$	0.298(7)	0.433(21)	97.7(0.2)
w/o $\mathcal{L}_{\text{sep}}$	0.301(2)	0.444(14)	97.9(0.1)
w/o All Losses	0.299(15)	0.448(24)	97.7(0.3)

TABLE IX: Ablation Study under Severe Shift ($\rightarrow \mathcal{D}_{v4}$)

Configuration	RMSE (m)	P90 (m)	Acc < 1m (%)
Full Model	0.490(21)	0.724(6)	96.0(0.2)
w/o Dom. Attn.	0.472(32)	0.726(29)	96.2(0.5)
w/o $\mathcal{L}_{\text{cls}}$	0.487(19)	0.720(13)	96.0(0.3)
w/o $\mathcal{L}_{\text{MMD}}$	0.505(22)	0.742(51)	95.5(1.0)
w/o $\mathcal{L}_{\text{sep}}$	0.477(15)	0.732(36)	96.0(0.6)
w/o All Losses	0.477(22)	0.719(46)	96.1(0.5)

hyperparameters, training domains, and evaluation settings unchanged. We additionally evaluate the effect of removing all domain-generalization losses simultaneously to examine their combined contribution. The masking-based multi-anchor fusion mechanism is not separately ablated, as its contribution to handling varying anchor configurations is evaluated independently in Section VI-D.

Table VIII presents the ablation results under a representative moderate domain shift, $\mathcal{D}_{\text{orig}} + \mathcal{D}_{v1} + \mathcal{D}_{v2} \rightarrow \mathcal{D}_{v3}$, while Table IX evaluates a more challenging combined domain shift, $\mathcal{D}_{\text{orig}} + \mathcal{D}_{v5} \rightarrow \mathcal{D}_{v4}$. All results are reported as mean(std) over three random seeds (42, 123, and 7).

Under the moderate domain shift, all configurations achieve strong localization performance, indicating that the task remains relatively robust when the target domain is close to the training domains. However, removing $\mathcal{L}_{\text{MMD}}$ or $\mathcal{L}_{\text{sep}}$ increases both RMSE and P90 relative to the full model. In addition, removing all domain-generalization losses simultaneously produces the largest P90 error, suggesting that these objectives provide complementary benefits, particularly in reducing larger localization errors. The contribution of the loss terms becomes more pronounced under the severe domain shift. Removing the MMD-based invariance-alignment objective $\mathcal{L}_{\text{MMD}}$ produces the clearest degradation, increasing RMSE from 0.490m to 0.505m and P90 from 0.724m to 0.742m, while reducing the accuracy within 1m from 96.0% to 95.5%. Notably, this is the only evaluated ablation that consistently degrades all three metrics relative to the full model under the severe shift. This result highlights the importance of explicitly aligning invariant feature distributions when the target domain differs substantially from the source domains. Overall, the ablation results indicate that the proposed domain-generalization objectives provide complementary benefits, with $\mathcal{L}_{\text{MMD}}$ making the most consistent contribution to robust generalization under challenging domain shifts.

VII. CONCLUSION

We proposed DAG-TransLoc, a transformer-based indoor localization framework jointly addressing domain generalization and variable anchor configurations. The masking-

based architecture with feature decoupling and domain-aware attention achieves up to 98.3% accuracy within 1m under moderate shifts and 95.8% under challenging combined BS and geometry shifts, though performance depends strongly on training-target domain compatibility, as detailed in Section VI. Results show that domain compatibility matters more than diversity, transfer is asymmetric, and high-complexity training domains are essential for robust generalization. Future work includes real-world hardware validation to characterize the sim-to-real gap (e.g., phase noise and clock synchronization errors not captured by ray-tracing simulation), dynamic user tracking, and lightweight adaptation for highly heterogeneous environments.

ACKNOWLEDGEMENT

The authors thank the Telecom Centres of Excellence (TCoE), Department of Telecommunications, Government of India, for supporting this research work through the Telecom Technology Development Fund (TTDF) grant TTDF/6G/532.

REFERENCES

- [1] F. Zafari, A. Gkelias, and K. K. Leung, "A survey of indoor localization systems and technologies," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 3, pp. 2568–2599, 2019.
- [2] X. Wang, L. Gao, S. Mao, and S. Pandey, "Deepfi: Deep learning for indoor fingerprinting using channel state information," in *2015 IEEE Wireless Communications and Networking Conference (WCNC)*, 2015, pp. 1666–1671.
- [3] A. Kirmaz, T. Şahin, D. S. Michalopoulos, and W. Gerstacker, "Toa and tdoa estimation using artificial neural networks for high-accuracy ranging," *IEEE J. Sel. Areas Commun.*, vol. 41, no. 12, pp. 3816–3830, 2023.
- [4] W. Long, X. Wen, H. Liu, M. Li, S. Li, Y. Wu, F. Chen, L. Liu, and R. Zhu, "Reliable indoor localization in multibuilding environments: Leveraging environment-invariant and position-related features," *IEEE Internet of Things Journal*, vol. 12, no. 14, pp. 26401–26414, 2025.
- [5] X. Xu, F. Zhu, S. Han, Z. Yu, H. Zhao, B. Wang, and P. Zhang, "Swin-loc: Transformer-based csi fingerprinting indoor localization with mimo isac system," *IEEE Trans. Veh. Technol.*, vol. 73, no. 8, pp. 11664–11679, 2024.
- [6] Z. Xu, M. Xue, X. Xiong, H. Hu, Y. Liu, J. Zhang, H. Wang, and Y. Shen, "Enhancing cross-scenario generalization in indoor localization via feature disentanglement," *IEEE Trans. Mobile Comput.*, pp. 1–17, 2026.
- [7] M. Xue, Z. Xu, J. Zhang, H. Wang, and Y. Shen, "A generalization method for indoor localization via domain-invariant feature learning," *IEEE Commun. Lett.*, vol. 29, no. 4, pp. 704–708, 2025.
- [8] Y. Ganin, E. Ustinova, H. Ajakan, P. Germain, H. Larochelle, F. Laviolette, M. Marchand, and V. Lempitsky, "Domain-adversarial training of neural networks," *Journal of Machine Learning Research*, vol. 17, no. 59, pp. 1–35, 2016.
- [9] G. Prasad, A. Pandey, and S. Kumar, "Domain adaptation for localization using combined autoencoder and gradient reversal layer in dynamic iot environment," *IEEE Trans. Netw. Sci. Eng.*, vol. 11, no. 1, pp. 685–695, 2024.
- [10] H. Hu, Z. Xu, and Y. Shen, "Dan: A domain-level attention network for cross-environment uwb indoor localization," in *2025 IEEE/CIC International Conference on Communications in China (ICCC)*, 2025, pp. 1–5.
- [11] J. Hoydis, F. A. Aoudia, S. Cammerer, M. Nimier-David, N. Binder, G. Marcus, and A. Keller, "Sionna RT: Differentiable ray tracing for radio propagation modeling," in *IEEE Globecom Workshops (GC Wkshps)*, 2023, pp. 317–321.
- [12] L. Zhang, H. Sun, J. Sun, and R. Q. Hu, "WiSegRT: Dataset for site-specific indoor radio propagation modeling with 3D segmentation and differentiable ray-tracing," in *IEEE WCNC Workshops*, 2024, arXiv:2312.11245.
- [13] A. Goldsmith, *Wireless communications*. United Kingdom: Cambridge University Press, Jan. 2005, vol. 9780521837163, publisher Copyright: © Cambridge University Press 2005.
- [14] A. Vaswani *et al.*, "Attention is all you need," in *Proc. NeurIPS*, 2017.
- [15] A. Gretton, K. M. Borgwardt, M. J. Rasch, B. Schölkopf, and A. Smola, "A kernel two-sample test," *Journal of Machine Learning Research*, vol. 13, no. 25, pp. 723–773, 2012, <http://jmlr.org/papers/v13/gretton12a.html>.
- [16] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," in *International Conference on Learning Representations*, 2017. [Online]. Available: <https://api.semanticscholar.org/CorpusID:53592270>
- [17] Y. Wang, J. Fu, and Y. Cao, "A new indoor positioning method reducing the errors due to nlos environments under sparse sensor deployment conditions," *IEEE Trans. Ind. Informat.*, vol. 21, no. 12, pp. 9767–9777, 2025.