The following publication Z. Lyu, T. T. L. Chan, G. C. M. Leung, Y. L. Chan, D. P. K. Lun and M. G. Pecht, "High-Dimensional Radio Frequency Fingerprint Synthesis for Indoor Positioning," in IEEE Transactions on Instrumentation and Measurement, vol. 74, pp. 1-16, 2025, Art no. 2517416 is available at https://doi.org/10.1109/TIM.2025.3551824.

High-dimensional radio frequency fingerprint synthesis for indoor positioning

Zhongyuan Lyu, Tom T.L. Chan, Gary C.M. Leung, Yui-Lam Chan, *Member, IEEE*, Daniel P.K. Lun, *Senior Member, IEEE*, and Michael G. Pecht, *Fellow, IEEE*

***Abstract*— Fingerprint-based indoor positioning systems are being explored to aid in location-based services due to their robustness in non-line-of-sight conditions. Current systems utilize high-dimensional radio frequency (HDRF) fingerprints, such as Wi-Fi channel state information, to achieve higher positioning precision. Since data acquisition is labor-intensive, researchers proposed to enrich the dataset with generative models. It however faced challenges arising from capturing the intricate HDRF distribution using simplistic models and the lack of a framework that simultaneously addresses the generative model training, sample evaluation and selection. To synthesize high-quality HDRF fingerprints, this paper proposes an HDRF fingerprint generation framework using a conditional diffusion model that learns the packet-level feature distribution by decomposing HDRF fingerprints using grid points, anchors, and frequency channel information, while preserving the feature spatial correlation within a fingerprint. A sample selection process using the Mahalanobis distance, and the Principal Component Analysis Q-statistic is used to ensure the sample fidelity. An adaptive learning strategy is further developed to integrate the generated synthetic HDRF fingerprints into downstream positioning tasks. Experimental results on two HDRF datasets quantitatively and qualitatively showcase the diversity and fidelity of the synthetic samples. Furthermore, compared to solely utilizing the original dataset, integrating the synthetic HDRF fingerprints from the developed framework to train downstream positioning models can decrease the positioning error by up to 16.9%.**

***Index Terms*— Fingerprint-based indoor positioning, Conditional diffusion model, Adaptive learning, Channel state information, BLE 5.1**

I. INTRODUCTION

Recent years, location-based services have fueled various indoor positioning system (IPS) developments [4]. IPSs leveraging ranging measurement methods, such as trilateration and triangulation, have achieved meter and submeter-level positioning accuracy in ideal indoor conditions but face challenges in complex environments with obstructions [5]. Recent advancements in communication technologies have facilitated the acquisition of radio frequency signals, such as Wi-Fi Channel State Information (CSI) [6] and the in-phase and quadrature-phase (I/Q) sample data from Bluetooth 5.1 [7]. These signals offer rich features with hundreds of dimensions for establishing location-specific fingerprints and enable fingerprint-based IPSs to robustly localize targets under non-line-of-sight (NLOS) conditions [8, 9]. This paper refers to these features as High-Dimensional Radio Frequency (HDRF) features and focuses on HDRF fingerprint-based IPS.

HDRF fingerprints are typically constructed by sequentially combining packet-level features, which are obtained from multiple packets that originate from different antennas and communication channels. Different packet-level features display distinct characteristics, primarily due to factors such as antenna direction, multipath effects, interference from other radio frequency signals, and obstructions [10]. Therefore, HDRF fingerprints collected from proximate locations can exhibit a pronounced nonlinear discrepancy. Such discrepancies allow the system to distinguish between proximate locations and achieve high positioning precision. Correspondingly, the system needs to gather data from more locations with increased spatial density. However, collecting sufficient HDRF fingerprints is resource-intensive and time-consuming, especially for large-scale environments. Researchers have explored data augmentation techniques, such as adding Gaussian noise [11], dropout [12], and crossover among fingerprints [13]. While these techniques enhance the model training performance, they do not capture the distribution of HDRF fingerprints, limiting the model generalization ability.

Previous research also explored using semi-supervised learning, which combines a small subset of labeled fingerprints with a larger set of unlabeled fingerprints [14]. Although collecting unlabeled data can be less challenging, it still demands substantial effort. Researchers further investigated using generative models, such as Variational Autoencoders (VAEs) [15] and Generative Adversarial Networks (GANs) [16], to directly generate HDRF fingerprints and reduce the data acquisition cost [1, 2]. Despite the efforts, previous research still struggles to synthesize location-discriminating HDRF fingerprints and improve the positioning performance due to the following challenges:

1) Previous research trains separate generative models for each location (class) to generate distinct fingerprints [1-3]. Besides, due to the limited data collected at each location, the

Zhongyuan Lyu is with the Centre for Advances in Reliability and Safety, Hong Kong, China (e-mail: zhongyuan.lyu@cairs.hk).

Tom T.L. Chan is with the Centre for Advances in Reliability and Safety, Hong Kong, China (e-mail: tom.chan@cairs.hk).

Gary C.M. Leung is with Blue Pin (HK) Limited, Hong Kong, China (e-mail: gary.leung@bluepin.hk).

Daniel P.K. Lun is with the Department of Electrical and Electronic Engineering, The Hong Kong Polytechnic University, and with the Centre for Advances in Reliability and Safety, Hong Kong, China (e-mail: enpklun@polyu.edu.hk).

Michael G. Pecht is with the Center for Advanced Life Cycle Engineering, University of Maryland, College Park, MD 20742 USA, and with the Centre for Advances in Reliability and Safety, Hong Kong, China (e-mail: pecht@umd.edu).

developed generative models opted for a shallow architecture with convolutional layers [2]. However, under such a strategy, the correlations among the HDRF fingerprints at different locations have not been learned. Moreover, shallow networks with convolutional layers can only learn representations within a narrow input range [17]. It is sufficient for general images since they often exhibit local features (such as edges, lines, textures, etc.). In contrast, more distinctive global features can be found in HDRF fingerprints. Generative models for HDRF fingerprints would benefit from constructing a unified, deeper model that incorporates data from all locations.

2) When assessing the performance of HDRF fingerprint generative models, prior research predominantly focused on visual assessments [2, 18], lacking a quantitative method to measure diversity and fidelity. Moreover, previous studies do not include the sample selection process, risking the inclusion of low-quality data into downstream positioning tasks.

3) Limited datasets are common in HDRF fingerprint-based IPS, which can lead to imperfect fingerprint distribution capture with generative models. In prior research, direct training of positioning models with synthetic fingerprints tends to cause initiation from a suboptimal point in the parameter space, leading to inferior performance and overfitting. More suitable training strategies need to be developed when using synthesized HDRF fingerprints.

This research addresses the above challenges and develops a framework named HDRF Conditional Diffusion (HDRF-CD) for HDRF fingerprint generation. Under HDRF-CD, a unified generative model is developed for all locations and learns the packet-level feature distribution of HDRF fingerprints using the conditional diffusion model (CDM), which allows for fine-grained control over the generation process and has demonstrated remarkable generation capabilities to learn the high-dimensional complex data distribution. To mitigate the discrepancy between the packet-level distribution and the noise prior distribution employed in the diffusion model training process, a scale regularization loss using the predicted clean data is introduced to ensure sample fidelity. During the fingerprint sampling process, the spatial correlation of packet-level features within an HDRF fingerprint is maintained through a strategic sampling approach that employs a diffused version of the original HDRF fingerprints. Sampled packet-level features are then reassembled to form HDRF fingerprints. A sample selection process utilizing the Mahalanobis distance and Principal Component Analysis (PCA) Q-statistic is proposed to ensure the acquisition of high-quality samples. Furthermore, this paper develops an adaptive learning strategy to incorporate synthetic HDRF fingerprints into the positioning model training.

To evaluate the quality of synthetic HDRF fingerprints, this paper introduces a metric called Diversity Fidelity Integrative Score (DFIS), leveraging the fixed meanings of each dimension in HDRF fingerprints. Then, we qualitatively and quantitatively showcase the generated samples by comparing them with those given by the previous generative methods with different metrics. Finally, the usage of synthetic HDRF fingerprints in downstream positioning tasks is investigated.

The contributions of this paper can be summarized as follows:

1) This paper introduces the first HDRF fingerprint generation framework HDRF-CD that utilizes the conditional diffusion model for HDRF fingerprint synthesis to enhance positioning performance. HDRF-CD trains a unified generation model that learns packet-level features of HDRF fingerprints with data from all locations. A scale regularization loss is incorporated on the predicted clean data to mitigate the disparity between the packet-level distribution and the noise prior distribution. HDRF-CD also systematically addresses the sample generation and selection process. Packet-level features of a single synthesis fingerprint are sampled from the diffused fingerprints to ensure sample diversity while preserving spatial correlation. A sample selection process employing the Mahalanobis distance and PCA Q-statistic is further implemented to obtain high-quality samples.

2) This paper conducts experiments on an open-source Wi-Fi CSI IPS dataset and a Bluetooth Low Energy (BLE) Angle of Arrival (AoA) dataset collected in this research. The BLE AoA dataset will be made publicly available. The comprehensive experimental results demonstrate that HDRF-CD can generate plausible HDRF samples and outperform other methods evaluated using the proposed metric Diversity Fidelity Integrative Score (DFIS). Samples from HDRF-CD yield the best performance on both the BLE 5.1 dataset (average 3.3 DFIS vs. the previous best of 1.1 DFIS) and the Wi-Fi CSI dataset (3.6 DFIS vs. the previous best of 2.6 DFIS), indicating superior data diversity among the generated samples while a higher resemblance to the real data.

3) The developed model adaption strategy integrates the synthetic HDRF fingerprints from HDRF-CD into downstream positioning tasks and improves the positioning performance. Specifically, incorporating an additional set of synthetic HDRF fingerprints, which is half the size of the training dataset, leads to a reduction of 16.9% in Root Mean Squared Error (RMSE) on the BLE dataset and 12.4% on the Wi-Fi CSI dataset, respectively.

The remainder of this paper is organized as follows: Section II introduces related research about the generative models for fingerprint-based IPS and generative diffusion model development. Section III provides the illustration of fingerprint-based indoor positioning and HDRF fingerprint generation. Section IV introduces the unified HDRF data synthesis framework and the proposed adaptive learning process for using HDRF fingerprints in positioning model training. Section V presents the numerical experiments and the comparison results. Section VI gives insights into future research directions and concludes this paper.

II. Literature Review

A. Generative Models for Radio Frequency Fingerprint-Based IPS

Different from fingerprint augmentation methods such as value exchange [13] and masking [11], generative models learn the

data distribution and produce data that closely resembles realistic data. Previous research has investigated the application of generative models in Received Signal Strength (RSS) fingerprint-based IPS. Alhomayani and Mahoor [3] studied the imbalanced data issue in BLE RSS-based IPS and utilized VAE as an oversampling tool. To construct a fine-grained indoor radio map with higher spatial granularity, Lan et al. [19] proposed a fingerprint augmentation framework, which comprises three modules: fingerprint-to-image conversion, super-resolution, and image-to-fingerprint conversion. Zou et al. [20] utilized Gaussian process regression conditioned least-squares generative adversarial networks (GPR-GANs), which are trained on collected data in free space, to generate realistic RSS data in constrained spaces.

Previous research also applied generative models to generate HDRF fingerprints. Compared to the single RSS value, HDRF fingerprints encompass high-dimension correlated features, introducing complex spatial-temporal dependencies. Chen and Chang [14] explored VAE and GAN-based semi-supervised techniques for Wi-Fi CSI fingerprints. Their results suggest that generative models have the potential to enhance the efficiency and accuracy of localization systems. Chen et al. [1] proposed a Wi-Fi-based localization system that addressed the inconsistency in Wi-Fi CSI fingerprints caused by changes in body shapes and environments. They developed a data augmenter with individual VAEs for each location to generate synthetic fingerprints. Li et al. [2] developed an Amplitude-Feature Deep Convolutional Generative Adversarial Network (AF-DCGAN) model to enhance the diversity of the CSI amplitude feature maps and reduce human effort. They conducted experiments to visualize the generated diverse samples and improved accuracy compared to other similar techniques.

The generative models from previous research for HDRF fingerprints were mainly developed to achieve context-specific data generation for individual locations. However, training with limited data for each location can cause the generative model to overlook the data diversity of different locations. Consequently, all individual models lack robust generalization capabilities across entire environments. This highlights the need for a comprehensive and unified generative development framework for HDRF fingerprints that can leverage data from all locations. Furthermore, previous research lacked comprehensive evaluations of the generated data to control the quality of samples for downstream positioning tasks. This research proposes a novel framework known as HDRF-CD, which allows the development of a unified generative diffusion model to generate HDRF fingerprints for all locations. HDRF-CD also comprises a sample selection strategy utilizing the Mahalanobis distance and a PCA-based Q-statistic measure to select data that closely resemble the original datasets.

B. Diffusion Models and Conditional Guidance

Previous works have explored the diffusion model, a class of likelihood-based models, for image synthesis tasks and showed their capabilities to produce high-quality images [21-23]. A diffusion model can be considered a Markovian Hierarchical Variational Autoencoder, and the latent dimension is exactly equal to the data dimension. The forward diffusion process gradually corrupts data into standard Gaussian noise over a series of timesteps. The backward sampling procedure, also known as denoising, gradually removes the noise from pure Gaussian random variables to obtain synthesis data. Ho et al. [21] introduced the denoising diffusion probabilistic models for high-quality image synthesis. They achieved a state-of-the-art Fréchet Inception Distance (FID) score and the results demonstrate the potential of diffusion models in generating high-quality images. Subsequent research further improved the diffusion model and developed several variants. Song et al. [24] introduced denoising diffusion implicit models (DDIMs), which employ non-Markovian diffusion processes to achieve the same training objective as DDPM with fewer denoising steps. Nichol and Dhariwal [25] improved the log-likelihood performance of DDPM by proposing a cosine noise schedule that improves Negative Log-Likelihood (NLL) and FID compared to the linear schedule. Rombach et al. [26] developed Latent Diffusion Models (LDMs) to reduce the computational complexity of training diffusion models by training them in the latent space of a pretrained autoencoder. Kingma et al. [27] derived a simple expression for the variational lower bound in terms of the signal-to-noise ratio (SNR) of the diffusion process. This enables the noise schedule to be optimized efficiently together with the diffusion model. Based on stochastic differential equations, Song et al. [28] introduced a unified framework for score-based generative models and DDPM. The framework employs a continuous evolution of distributions over time.

Previous studies also investigated techniques for conditioning/guiding diffusion models on class/context information. Dhariwal and Nichol [29] introduced a classifier guidance method to allow for a trade-off between diversity and fidelity. Specifically, the diffusion score is incorporated with the gradient of the log-likelihood of an auxiliary classifier model and used in the sampling process. Ho and Salimans [30] further introduced the “classifier-free guidance” for balancing mode coverage and sample fidelity in diffusion models. Instead of training an extra classifier, this approach involves using the score estimates of both a conditional and an unconditional diffusion model to achieve a similar trade-off between sample quality and diversity. Bansal et al. [31] presented a universal guidance algorithm that utilizes the predicted clean image obtained from the predicted noise as input to the guidance function without retraining. Hong et al. [32] proposed a general formulation of diffusion guidance that leverages information within intermediate samples, allowing for a more general approach that does not require external conditions or additional training. Goel et al. [33] proposed a Structure-and-Appearance Paired Diffusion model (PAIR-Diffusion) that provides fine-grained control over individual objects in an image.

This research focuses on integrating the diffusion model into the generation of complex radio frequency fingerprints. HDRF-CD specifically emphasizes the incorporation of conditional context information derived from signal packets to enhance the

diffusion model's capability to aggregate fingerprints from various locations, thereby capturing the intricate dependencies within the data.

TABLE I
ATTRIBUTES OF THE RECEIVED PACKET USED FOR FINGERPRINT-BASED IPS

Attribute	Description
Receiving anchor	Unique identifier of the node that successfully received the data packet
Receiving antenna	Specific antenna within the receiving node that captured the signal.
Frequency Channel	The designated frequency channel utilized for the transmission of the data packet.

III. FINGERPRINT-BASED INDOOR POSITIONING AND HDRF FINGERPRINT GENERATION

This section first provides an overview of the fingerprint-based IPS with relevant concepts. Then, the problem of HDRF fingerprint generation addressed in this paper is illustrated.

A. Fingerprint-based IPS

The fingerprint-based IPS aims to localize targets based on the constructed fingerprints in indoor environments such as offices, malls, and warehouses. It utilizes multiple receivers, which are also called anchors, deployed at different locations to localize targets based on the constructed fingerprints. Each considered location is defined as a grid point for data collection and fingerprint construction. Targets move around the environment, carrying the Internet of Things (IoT) tag as the transmitter. The tag transmits signal packets periodically. The anchors receive packets from different directions, and the collected data is forwarded to a gateway, which then uploads the data to a central server. The server further processes the data into HDRF fingerprints for analyzing the targets' locations.

The system development involves offline and online stages. In the offline phase, dataset construction and positioning model training are performed. For dataset construction, collectors equipped with tags traverse all points and pause at each grid point for a predefined period. The system logs the packets received from all anchor nodes and the corresponding grid points. Table 1 summarizes the commonly used metadata associated with the received packets to describe the attributes of the packets. It includes the anchor node that receives the packet, the received antenna of the anchor (each anchor can have multiple antennae to receive packets), and the communication channel (wireless advertising channels that broadcast signals at different frequencies). Apart from the grid point (label) information, the packet-level features also exhibit distinct variations based on these attributes. All attributes can be easily obtained directly from the received packets and work as conditions to construct fingerprints.

A predefined number of packets collected over a continuous period is considered a packet set. All collected packets of each grid point are segmented into packet sets. One packet set is denoted as $F = \{x^1, \dots, x^N\}$, where N represents the number of packets and $x^i \in \mathbb{R}^M$ represent the extracted M dimensions feature from the i^{th} packet. Then the packet set is processed into the HDRF fingerprint as a multidimensional vector with each dimension representing fixed means, such as anchor and antenna/channel. There can be redundant packets in one packet set and these features can be processed by taking the mean values or randomly selecting one of them. Then, all HDRF fingerprints along with their grid point labels are used to train the positioning model.

During the online positioning phase, the system continuously monitors the packets transmitted by targets and collects the extracted features from these packets over a predetermined time frame. These features are preprocessed to form an HDRF fingerprint, which is subsequently input into a trained positioning model. The positioning model then estimates the tag's location by generating an output vector, which represents the probabilities associated with different grid points. The dimension of the output corresponds to the number of grid points.

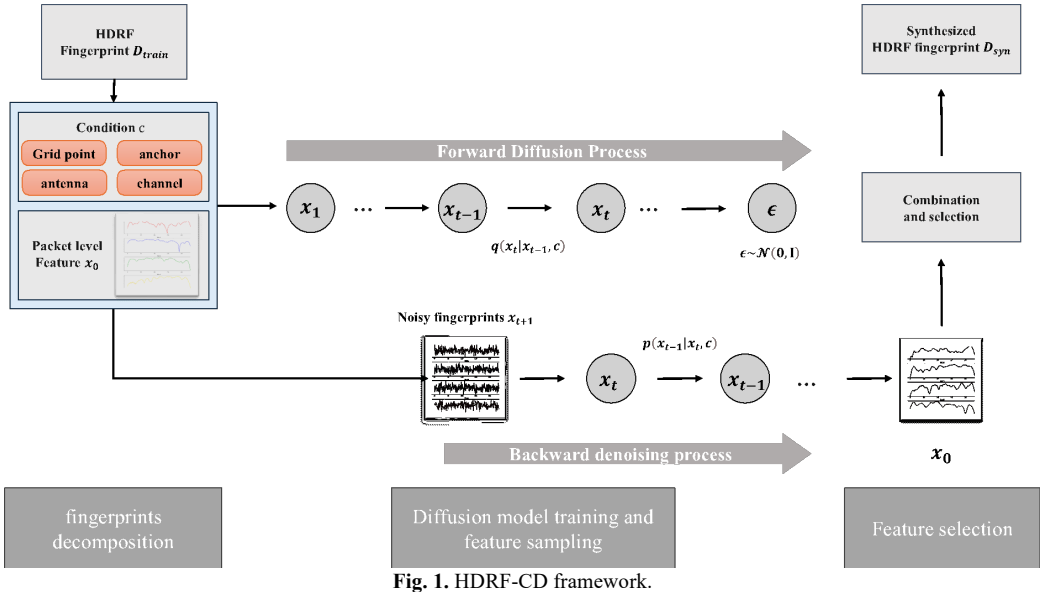


Fig. 1. HDRF-CD framework.

B. HDRF fingerprint generation

This research works on the HDRF fingerprint generation problems for HDRF fingerprint-based IPS to alleviate the labor intensity of data collection and enhance the generalization capability of positioning models. During data collection, users equipped with tags move randomly at each designed grid point. Consequently, only a portion of the space near grid points is traversed by the targets and the collected datasets lack the fingerprints for the unvisited locations in the vicinity of each grid point. This research proposes to synthesize these fingerprints using generative models. Let $D_{train} = \{(F_j, y_j)|j \in [1, N_{train}]\}$ represent the collected training dataset and $D_{syn} = \{(\hat{F}_j, y_j)|j \in [1, N_{syn}]\}$ represent the synthetic HDRF dataset. N_{train} and N_{syn} represent the amount of data in each dataset. This research aims to utilize the training dataset D_{train} to learn the data distribution of all grid points through generative models and to generate a synthetic HDRF dataset D_{syn} that can be generalized to the unvisited spaces around each grid point.

IV. HDRF Conditional Diffusion Framework

This section presents the proposed framework HDRF-CD for HDRF fingerprint synthesis. Subsection A elucidates the overall framework. Subsection B first introduces the CDM training and data generation process. Then, a scale-driven conditional training process is developed to improve the fidelity of HDRF samples. In subsection C, the sample selection strategy based on the Mahalanobis distance and PCA-based Q-statistic measure is proposed for identifying high-quality packet-level features. Subsection D introduces the adaptive learning process that leverages the generated HDRF fingerprints with the original training data to improve the downstream positioning model's performance.

A. HDRF Fingerprints Decomposition and Combination

The HDRF-CD framework, as shown in Fig. 1, comprises three steps: fingerprint decomposition, diffusion model training and feature sampling, and feature selection. In the upper left corner is the input original HDRF fingerprint dataset. The output is the synthesized HDRF fingerprints, shown in the upper right corner.

The fingerprint decomposition is illustrated on the left-hand side of Fig. 1. An HDRF fingerprint is constructed with a collection of packet-level features (instances) where a correlation exists among their values but no necessary dependencies (in fact, with the high-dimensional nature, one packet-level feature is sufficient to characterize a location). Fingerprints can also be easily disaggregated back into their packet-level features. It should be noted that this process is different from directly working on the individual packet-level feature. The packet-level features still need to be combined to construct HDRF fingerprints and employed for downstream positioning tasks. Due to the structural intricacy and large parameter space, training the diffusion generative model necessitates a substantial quantity of HDRF fingerprints, which are limited in practical scenarios. Therefore, HDRF-CD learns the packet-level feature distribution with explicit conditions and views an HDRF fingerprint as a multiple-instance set [34]. For each packet-level feature x^i, let $c_i = \{y_i, ac_i, at_i, f_i\}$ represent the explicit conditions for control purposes, where y_i denotes the grid point index, ac_i denotes the anchor, at_i denotes the antenna, and f_i denotes the frequency channel. Different kinds of HDRF fingerprints can include partial conditions. For example, in a BLE 5.1 packet, all antennas are sequentially utilized to obtain I/Q samples, and a processed BLE 5.1 I/Q fingerprint does not need to include the antenna condition information.

Then, the packet-level features in a fingerprint are learned and generated by the proposed diffusion generative model. Under HDRF-CD, a unified diffusion generative model is developed to learn the packet-level features of all grid points. In the forward diffusion process, the packet-level feature is gradually transformed into pure noise through a series of timesteps. In the backward denoising process, the learned diffusion model iteratively recovers the original data from the noisy data and synthesizes the packet-level features. Then, the obtained clean packet-level features from each anchor/antenna/channel conditional sampling process are reassembled to reconstruct the fingerprints for downstream localization model training.

To preserve the spatial relationships among the features within a fingerprint, HDRF-CD employs a straightforward strategy that samples from the intermediate timesteps of the diffusion model. Compared to the final step of diffusion, where the features become pure noise, the intermediate time steps still retain the correlation among the packet-level features derived from the same HDRF fingerprint. As illustrated in the middle of Fig. 1, for an HDRF fingerprint, HDRF-CD utilizes the noisy version of its packet-level features at a predefined timestep t. Then, the backward denoising process is performed on these noisy packet-level features to obtain clean data, which are further combined into an HDRF fingerprint. Finally, as shown in the right part of Figure 1, HDRF-CD addresses the sample selection process for HDRF fingerprints, which is performed using the Mahalanobis distance and a PCA-based Q-statistic measure.

B. Conditional Diffusion Model and Scale-Driven Conditional Training

This subsection first introduces the CDM training and sampling process. Then, the scale-driven training loss for alleviating the scale explosion of sample values utilizing the predicted clean data is illustrated.

As depicted in the top-middle part of Fig. 1, the forward process of the diffusion generative model is defined as a Markov chain where Gaussian noise is gradually added to the original data based on a discrete variance schedule.

$$q(x_t|x_{t-1}) = \mathcal{N}(x_t;\ (1-\beta_t)x_{t-1}, \beta_t\mathbf{I}), \qquad (1)$$

where t denotes the time step for adding noise, ranging from 1 to T. The variable x_t represents the noisy data at each time step, and β_t represents the corresponding variance schedule. This research keeps β_t as a constant hyperparameter with a linear schedule [21]. Other schedules, for example, the cosine schedule [25] or joint learned through SNR, can also be employed. Then, the transition probability from the clean data $x_0 \sim p_{data}$ to x_t is given by

$$q(x_t|x_0) = \mathcal{N}\left(x_t;\ \sqrt{\overline{\alpha}_t}x_0, \left(1-\sqrt{\overline{\alpha}_t}\right)\mathbf{I}\right),$$
$$\text{where } \overline{\alpha}_t = \prod_{s=1}^{t}(1-\beta_s). \quad (2)$$

In the final time step, x_0 is perturbed to a standard normal distribution (prior distribution). Through reparameterization, instead of predicting the original data inputs from the noisy data, the model θ tries to predict the noise $\epsilon \sim \mathcal{N}(0, I)$ and optimize the model parameter using the following objective [21, 24]:

$$\min_{\theta} \mathbb{E}_{x_o,\epsilon,t}\left[\omega(t)\left|\left|\epsilon_\theta(x_t,t)-\epsilon\right|\right|_2^2\right], \quad (3)$$

where $\omega(t)$ is a time-based weight parameter. For the reverse process, the model gradually recovers the original data from the noisy data.

$$p(x_{t-1}|x_t) = \mathcal{N}(x_{t-1};\ \mu_\theta(x_t,t)\ x_t, \sigma_t^2\mathbf{I})), \quad (4)$$

where $\mu_\theta(\cdot)$ is parameterized using the model θ, and σ_t^2 is set to untrained time-dependent constants (related to β_t). Then, the Markov chain sampling scheme of the diffusion model is:

$$p(x_{0:T}) = p(x_T)\prod_{t=1}^{T} p(x_{t-1}|x_t), \quad (5)$$

where $p(x_{t-1}|x_t)$ is defined in (4). HDRF-CD aims to learn the conditional distribution $p(x_{t-1}|x_t, c)$, which allows explicit control of the generated packet level feature by conditioning it on specific information. Then, at each transition step, conditioning information c is incorporated,

$$p(x_{0:T}|c) = p(x_T)\prod_{t=1}^{T} p(x_{t-1}|x_t, c). \quad (6)$$

And the model can be trained with the loss:

$$\min_{\theta} \mathbb{E}_{x_o,\epsilon,t}\left[\omega(t)\left|\left|\epsilon_\theta(x_t,t,c)-\epsilon\right|\right|_2^2\right] \quad (7)$$

In the sampling process, as the time step approaches 0, the noise in the samples gradually diminishes and the clean data are generated. Without additional domain knowledge, the scaled data (range of [-1, 1]) is commonly assumed to have a zero mean when generating images. However, the distribution of the HDRF features p_{data} often exhibits a non-zero mean. Fig. 2 illustrates the histogram of mean values across all dimensions of Wi-Fi CSI packet features, which have been scaled to the range of [-1, 1]. Notably, the mean values have a concentration of around 0.6.

Therefore, HDRF-CD enhances the distribution learning by introducing constraints on the predicted clean data. With the fact that diffusion model $\epsilon_\theta(x_t,t,y)$ tries to predict the noise $\epsilon = \frac{x_t-\sqrt{\overline{\alpha}_t}x_0}{\sqrt{1-\overline{\alpha}_t}}$ [21], the predicted clean data can be derived from $\hat{x}_0 = \frac{x_t-\sqrt{1-\overline{\alpha}_t}\epsilon_\theta(x_t,t,y)}{\sqrt{\overline{\alpha}_t}}$. To measure the similarity between the clean predicted data $\hat{x}_0$ and original data x_0, one intuitive approach is to use the reconstruction loss. However, directly applying the reconstruction loss to each dimension of the data is inappropriate for HDRF data since it is inherently noisy. Perfectly reconstructing the original data can inadvertently emphasize and amplify the noise, hindering the model's ability to learn meaningful features and patterns. Based on these considerations, HDRF-CD minimizes the L1 norm between $\frac{1}{M}\sum_{d=1}^{M}\hat{x}_{0,d}$ and $\frac{1}{M}\sum_{d=1}^{M}x_{0,\mathrm{d}}$, where M is the packet-level feature dimension. Finally, the training objective is given by

$$\min_{\theta}\left\{\begin{array}{c}\mathbb{E}_{x_0,\epsilon,t}\left[\omega(t)\left|\left|\epsilon_\theta(x_t,t,y)-\epsilon\right|\right|_2^2\right] \\ +\left|\frac{1}{M}\sum_{d=1}^{M}\hat{x}_{0,d}-\frac{1}{M}\sum_{d=1}^{M}x_{0,\mathrm{d}}\right|\end{array}\right\}. \quad (8)$$

C. Sample Selection Strategy

In this section, a feature selection strategy for HDRF is proposed, which has not been extensively explored in previous fingerprint-based IPS research. While research in other areas has attempted to create additional discriminators [35-37] for sample selection, this often complicates the training and sampling processes for HDRF features. This paper proposes a strategy specifically designed for HDRF fingerprints, utilizing the Mahalanobis distance and PCA-based Q-statistic measure, based on the fixed meaning of each dimension of HDRF fingerprints.

Since each dimension of the packet-level feature has a fixed meaning, distance-based metrics can be directly used to measure similarity. Furthermore, this research posits that

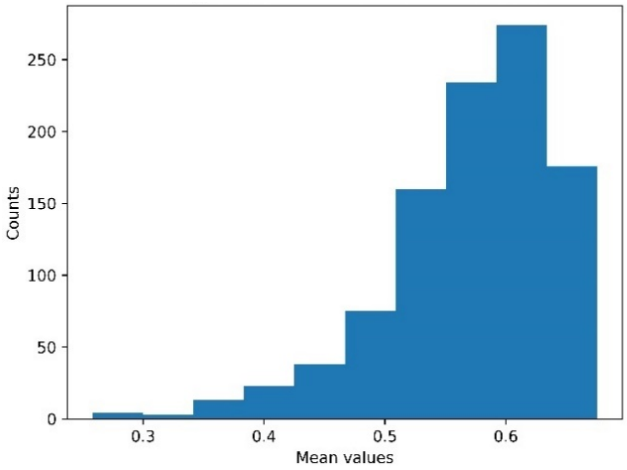


Fig. 2. Histogram of the mean values of Wi-Fi CSI data with a batch size of 512. The X-axis represents the mean values of packet-level features scaled to a range of [-1, 1] across all dimensions. The Y-axis represents the count of data points.

samples that deviate far from the training set are less likely to represent true data. This point has also been corroborated by the previous research [35-37]. Therefore, this research employs PCA Q-statistic, also known as the Squared Prediction Error (SPE), on the residual subspace to filter out a subset of samples based on their SPE values. A high SPE value for a new sample indicates a substantial deviation from the normal patterns captured in the training data. Besides, the computational complexity of the selection process is influenced by the size of the training dataset, generated sample candidates, and feature dimension. PCA enhances the selection effectiveness by reducing the feature dimension, alleviating the computational burden while still capturing the essential characteristics.

The proposed selection strategy is implemented on the packet-level features. Each HDRF fingerprint consists of

packets sampled under all conditions. One HDRF fingerprint is included in the synthesized dataset only when all packet-level features are selected. The overall process is summarized in Algorithm 1. Firstly, a PCA model is fitted with a predefined number of components using the training features and is applied to both the training features and the generated packet-level sample candidates. Then, Algorithm 1 calculates the SPE on the data residual subspace. Next, the covariance matrix is calculated using the projected training features. Finally, the minimum Mahalanobis distance between each test sample and the training features is computed. By utilizing the proposed sample selection strategy, HDRF-CD eliminates the need to develop additional discriminators, thereby streamlining the overall processes. Furthermore, the entire selection process can also be implemented on GPUs with negligible additional computational time.

Algorithm 1 Sample selection strategy using Mahalanobis distance and PCA Q-statistics

Objective: Select samples from the generated sample set that are sufficiently similar to the training dataset with the predefined threshold.

Input: PCA component $\boldsymbol{K}$, training features $\boldsymbol{p} = \{\boldsymbol{p_1}, \dots, \boldsymbol{p_N}\}$, sample features $\boldsymbol{s} = \{\boldsymbol{s_1}, \dots, \boldsymbol{s_M}\}$, selection threshold $\boldsymbol{R}$, SPE ratio $\boldsymbol{\sigma}$%

#1 Calculate the covariance matrix $\boldsymbol{C}$ of $\boldsymbol{p}$.

#2 Principal Components Selection: chosen the $\boldsymbol{K}$ principal components (eigenvectors) $\boldsymbol{V} = (\boldsymbol{v_1}, \dots, \boldsymbol{v_K})$ obtained by singular value decomposition on the covariance matrix $\boldsymbol{C}$.

#3 Obtain the projected train data $\hat{\boldsymbol{p}}$ and samples $\hat{\boldsymbol{s}}$.

#4 Calculate the SPE $\hat{\boldsymbol{s}}^{\boldsymbol{T}}(\boldsymbol{I} - \boldsymbol{V}\boldsymbol{V}^{\boldsymbol{T}})\hat{\boldsymbol{s}}$ and the samples with the top $\boldsymbol{\sigma}$% smallest SPEs are retained.

#5 Calculate the covariance matrix $\boldsymbol{M}$ of $\hat{\boldsymbol{p}}$.

#6 Create Selected_Sample_List = [].

#7 **For** $\hat{\boldsymbol{s}}_{\boldsymbol{i}}$ in $\hat{\boldsymbol{s}}$ **do**.

    Distance_list=[]

    For $\hat{\boldsymbol{p}}_{\boldsymbol{j}}$ in $\hat{\boldsymbol{p}}$ **do**

        Add $\boldsymbol{d_i} = \sqrt{(\hat{\boldsymbol{p}}_{\boldsymbol{j}} - \boldsymbol{s_j})^{\boldsymbol{T}} \boldsymbol{M} (\hat{\boldsymbol{p}}_{\boldsymbol{j}} - \boldsymbol{s_j})}$ to Distance_list

    If Min(Distance_list)< $\boldsymbol{R}$ **do**

        Add $\boldsymbol{s_j}$ to Selected_Sample_List

Output: Selected_Sample_List

D. Adaptive Learning Process for Positioning Model

This paper proposes employing the model adaptation strategy [38, 39], which leverages the model obtained from training with moderate condition domains in hard condition domains. For fingerprint-based indoor positioning problems, the concept of domain adaptation can be elucidated as follows: Considering the original training dataset, in each grid point, the HDRF fingerprints of the dataset are scattered around the center of the grid point, covering only a partial vicinity due to limited data availability. Therefore, this paper defines the source domain as the partial location vicinity. Then, the target domain, which corresponds to the full vicinity regions of all grid points, is also defined. Locations in the target domain, while not in the source domain, can correspond to the unseen HDRF fingerprints in the test dataset. The aim is to develop a positioning model capable of recognizing the fingerprints from the source domain while also being generalizable to the target domain.

The source domain is denoted as S (partial location vicinity) and the target domain (full location vicinity) is denoted as T. This research focuses on the single-stage adaptation method, and the positioning model is adapted from the source domain to the target domain by leveraging the original training dataset and the synthesized dataset.

After training on the dataset D_{train}, the learned positioning model is further adapted to the target domain by learning from both the synthetic HDRF dataset D_{syn} and the original dataset D_{train}. To ensure a balanced influence of both datasets in the training process, a hyperparameter λ is introduced to the loss function to control the weighting assigned to each data point.

$$Loss = \sum_{j=1}^{N_{train}} L(y_j, y_j') + \lambda \sum_{j=1}^{N_{syn}} L(y_j, y_j'). \quad (9)$$

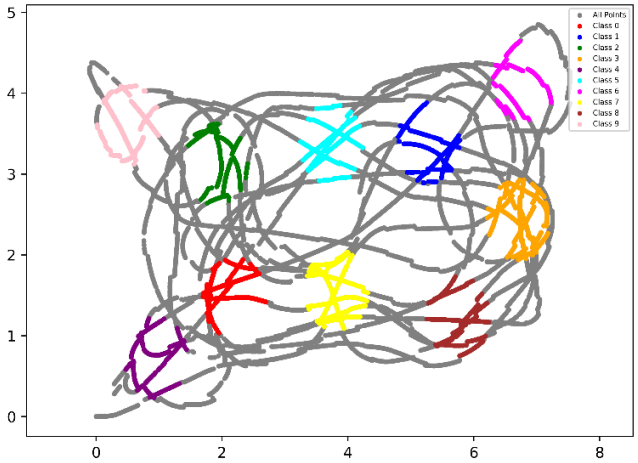
The cross-entropy loss $L(.)$ is applied to evaluate the difference between the predicted outputs and ground truth labels.

V. Numerical Experiments

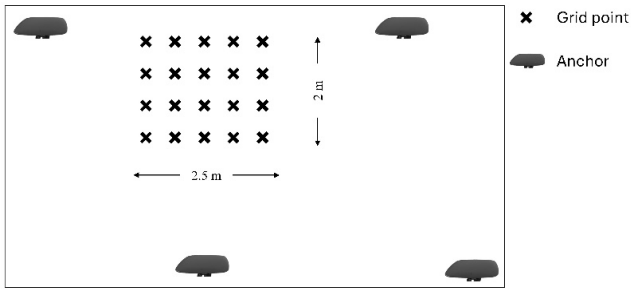
This section presents experimental evaluations. First, subsection A provides details on benchmark indoor positioning fingerprint datasets, baseline methods, model implementations, and evaluation metrics. Second, in subsection B, a qualitative and quantitative assessment of the generated data is conducted, demonstrating the ability of HDRF-CD to synthesize high-quality data. Finally, subsection C investigates enhancing the positioning model's performance by using both the generated data and the original training dataset. All experiments were performed on DGX A100 GPUs with 40 GB memory.

A. Experiment Setup

1) Datasets: This paper conducts performance evaluations using two HDRF datasets: (1) channels_July16 Wi-Fi CSI dataset; and (2) BLE 5.1 AoA location vicinity dataset. The first dataset is from [40] and consists of an 8 m x 5 m environment setup, with three anchors equipped with four antennas each. Following the processing approach of [40], the CSI obtained from each antenna was represented by a vector with a size of 234, which is the number of sub-frequencies [2]. As a result, the shape of each fingerprint is 3 x 4 x 234. Then, the data was scaled to fit within the range of [-1, 1]. In the collected dataset, the label information



(a)



(b)

Fig. 3. The 2D top view of data collection environments. (a) Wi-Fi CSI dataset. The black traces are formed by data points. The clusters of different colors represent different grid points. The unit of the axis is meters. (b) BLE 5.1 dataset. Each black point representing a grid point with 0.5 grid spacing.

associated with each fingerprint was represented by continuous 2D coordinates. Ten grid points were randomly selected, each with a radius of 0.5 m, and a separate class was assigned to each grid point. The grid points are shown in Fig. 3 (a). In total, the dataset comprises 9,116 fingerprints and 78,744 packets for all grid points.

The second dataset was constructed by the authors' team for this study. The data were collected in a 2.5 m x 1.5 m rectangular laboratory area in Building 19W at the Hong Kong Science Park, depicted in Fig. 3 (b). A grid was laid down consisting of 20 (5*4) points, with a grid spacing of 50 cm. The coordinates of each grid point were measured in centimeters, following a coordinate system with the bottom left as the origin. For data collection, the Minew BLE 5.1 AoA G2 system was used, employing four anchors as receivers, and the E5 Beacons tag as the transmitter. The anchors were positioned at the corners, facing the ground at a height of 2.8 m. During the data collection process, the BLE tag was placed on a robot that moved around, slightly deviating from the exact grid points. The tag transmitted BLE packets on three broadcasting channels. Each fingerprint involves one BLE packet from each anchor. Applying the I/Q sample processing method from prior studies [7, 41], a 32-length feature vector ranging within [-1, 1] is derived within every BLE packet based on I/Q samples among all antennas. Consequently, each fingerprint takes the form of 4 x 32. In total, the dataset comprises 8,509 fingerprints and 32,376 packets for all grid points. The BLE dataset will be made publicly available. Example samples of two datasets are shown in the supplementary file.

Although the HDRF datasets are collected continuously over time, each packet collected is independent of the others and can be considered temporally independent, not relying on previous or subsequent packets. Previous research has demonstrated the temporal stability of HDRF features [42, 43]. Therefore, this work adopted 70% training, 10% validation, and 20% testing random split for both datasets, which is commonly adopted in previous research [44]. To ensure the reliability of the results while considering computational resources, three random splits of the datasets were conducted to create 3-fold cross-validation datasets.

2) Baseline Methods and Model Implementation: The proposed HDRF fingerprint generation framework is compared with the following methods: VAE-based generative model from Chen et al. [1]; Deep generative models from Alhomayani et al. [3]; Amplitude-Feature Deep Convolutional Generative Adversarial Network (AF-DCGAN) model from Li et al. [2]. In [2], the authors modified the last sigmoid layer of the discriminator in the DCGAN model, replacing it with a normalization layer. They also used the RMSProp optimization algorithm instead of Adam during the model training process.

The timestep number T was set to 1000 to implement the developed CDM. Such a number was widely adopted in previous works [21, 45]. It is large enough to address the truncation error, which is introduced by the discretization, and scales super linearly to the step size $\frac{1}{T}$ [46]. This work employed a linear discrete variance schedule during the forward process, and the variance gradually increased from β=8 x 10^{-4} to 0.012. Similar to [21], the U-Net backbone was adopted for the model architecture [47] and three feature map resolutions for both datasets were used (from 8 x 8 to 2 x 2 for BLE datasets and from 16 x 16 to 4 x 4 for CSI datasets, respectively). The final network has 15 convolutional residual blocks, 10 self-attention blocks, and four convolutional layers. Parameters are shared across time steps. With batch size 1000, the developed CDM was trained on the Wi-Fi CSI dataset for 16K steps and the BLE 5.1 dataset for 54K steps. For both datasets, the generation was initiated from the time step of 600 using the noisy HDRF fingerprints to preserve spatial correlations. Then, the sample selection process was applied to each packet-level feature of a fingerprint separately. Ultimately, those fingerprints with every packet-level feature satisfying the selection threshold

were chosen. The SPE ratio σ was set to 95%. The selection threshold R was determined as the 70th percentile after sorting all samples in ascending order. The learning rate was set to 0.0003 for all generative models.

The positioning models developed for both datasets utilize a sequential structure comprising two main components: a feature extraction module and a classification module. The detailed structures are shown in Fig. 4. The feature extraction module consists of two three-layer groups. Each group starts with a linear layer, followed by a sigmoid activation function, batch normalization, and a dropout layer with a dropout rate of 0.3. The feature dimensionality gradually decreases in each group.

Similarly, each classification module also consists of two three-layer groups. Finally, a linear layer produces an output vector for classification. In the training phase, if there are no improvements when verifying the model with the validation set after 200 epochs, the model's training will be terminated. The balancing parameter λ in the training of the positioning models was set to 0.5 for both datasets. The learning rate in the initial training stage was set to 0.0003 for both positioning models. In the adaption stage, the learning rate was set to 0.00005 for the BLE dataset and 0.0003 for the CSI dataset.

3) Performance Metric: To quantitively evaluate the quality of synthetic HDRF datasets, this paper introduces the Diversity Fidelity Integrative Score (DFIS) as follows:

$$DFIS = \frac{MED + k \times VED}{MMED}, \tag{10}$$

where

$$MED = \frac{1}{G * A}\sum_{g=1}^{G}\sum_{a=1}^{A} E[D^{g,a}];$$

$$VED = \frac{1}{G*A}\sum_{g=1}^{G}\sum_{a=1}^{A} E_{i,j\in N_{sample}^{g,a}}\left[\left|d_{i,j} - E[D^{g,a}]\right|^2\right];$$

$$MMED = \frac{1}{G * A}\sum_{g=1}^{G}\sum_{a=1}^{A} E_{i\in N_{test}^{g,a}}\left[\min_{, \forall j \in N_{sample}^{g,a}} \left(\left|x_j - x_i\right|_2\right)\right].$$

k is a weight parameter and can be simply set as 1. The numerator consists of two terms. The first term MED involves calculating the Euclidean distances $d_{i,j}$ between every two generated packet-level samples in the sample set $N_{sample}^{g,a}$ belonging to point g and anchor a, resulting in a distance matrix $D^{g,a} = \{d_{i,j}\ \forall\ i,j \in N_{sample}^{g,a}\}$. In contrast to image data evaluation, which necessitates the development of additional feature extraction models [48], HDRF features have fixed meanings for each dimension, and the distances among samples can be easily calculated. $E[\cdot]$ means the average value calculation and $E[D^{g,a}]$ denotes the mean of the distance matrix $D^{g,a}$. Then, MED calculates the average value among all grid points and anchors. G and A denote the numbers of grid points and anchors respectively. The second term VED calculates the sample distance variance in the same manner. The denominator term $MMED$ gives the

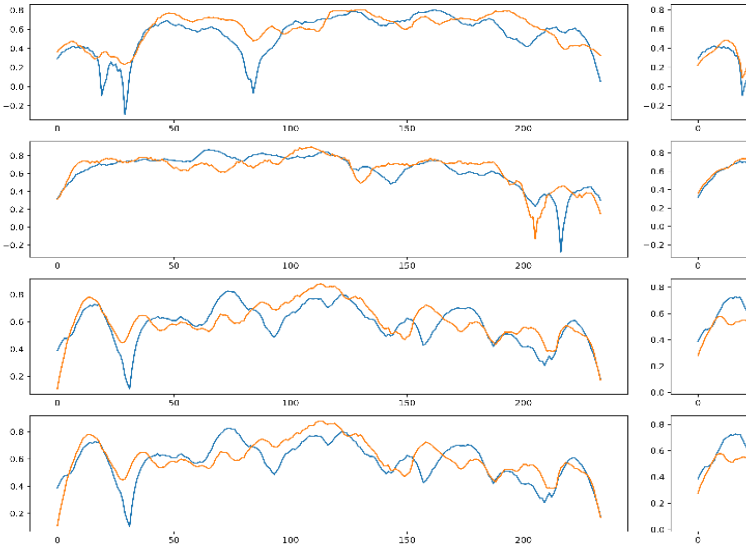


Fig. 5. CSI amplitude feature generated by

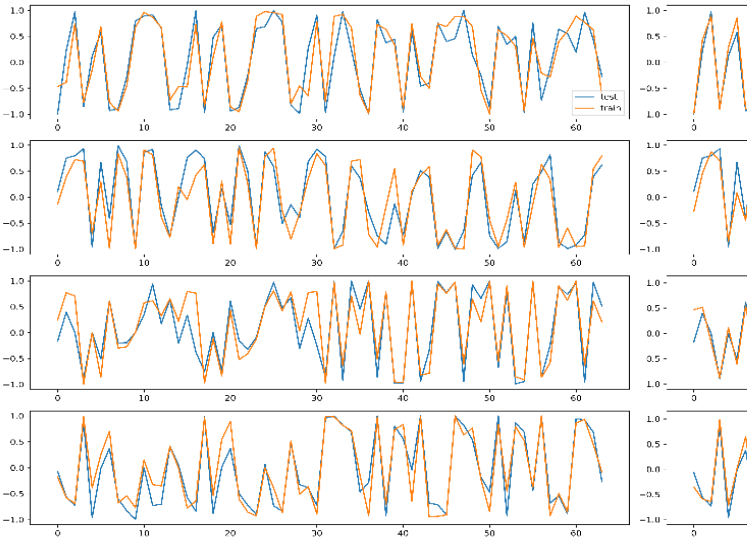


Fig. 6. BLE AoA feature generated by th

similarity between synthetic samples and the data in the test set. In the equation of $MMED$, $N_{test}^{g,a}$ represents the data set of grid point g and anchor a in the test set. $MMED$ is obtained by first computing the minimum Euclidean distance between each test set instance and its nearest neighbor in the sample set. Then, the average distance across all instances for grid point g and anchor a is evaluated. Finally, the mean values across all anchors and grid points are calculated to obtain $MMED$. As shown in (10), the numerator of DFIS measures the diversity of synthetic samples, while the denominator represents the sample fidelity. A higher DFIS value (larger MED and VED, smaller MMED) indicates better quality of the generated data.

Remark 2: Due to the less sensitive spatial variations of CSI amplitudes, the features extracted from different antennas of the same anchor are similar. This work compared the CSI data by pooling all the antennas' samples together for

evaluation. Therefore, in the following experiments, both samples of the CSI dataset and BLE dataset are considered with different grid point and anchor conditions.

Finally, to evaluate the positioning performance, this paper utilizes the widely adopted metric Root Mean Squared Error (RMSE),

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2}, \qquad (11)$$

where $(\hat{x}_i, \hat{y}_i)$ and (x_i, y_i) are the estimated and ground truth coordinates, respectively; N is the size of the test dataset. The grid point with the highest probability is selected from the probability representation as the final location result.

B. Assessment of Synthetic Data

This subsection first presents a qualitative demonstration of the generated packet-level features, showcasing how HDRF-CD effectively captures the patterns of radio frequency data. Then, the quantitative assessment is conducted using the abovementioned metrics and comparing the results with those given by other generative methods.

1) Quality Performance: The generated data for the two datasets obtained from HDRF-CD are showcased in Fig. 5 and Fig. 6 respectively. On the left-hand side of the figures, each subplot shows a selected instance from the test set (blue line) and its most similar sample from the training set (orange line). On the right-hand side, each subplot displays the same instance from the test set (blue line) and its associated most similar sample from the generated data (orange line). It can be observed that the developed CDM has effectively captured the trends of data variation and generated high-quality data.

2) Quantitative Sample Evaluation with Performance Metric: This subsection quantitively evaluates the generated samples from all methods using the DFIS metric. For each method under each cross-validation dataset, we independently sample five times to obtain the statistical DFIS performance. Each time, the number of samples equivalent to the size of the training data sets is generated. The mean DFIS results are presented in Table II. Method "HDRF-CD_NC" represents the DFIS performance of the synthetic samples generated from CDM without using the selection strategy. Each column 'Cross' means the results obtained under different cross-validation datasets.

It can be observed from the table that HDRF-CD demonstrates superior performance on both datasets. In contrast, the best previous method AF-DCGAN achieved approximately 2.6251 DFIS on the Wi-Fi CSI dataset but only achieved 1.0473 DFIS when applied to the BLE AoA dataset. Besides, the samples of HDRF-CD consistently outperform other methods with the highest DFIS, indicating higher sample diversity while maintaining similarity to real HDRF fingerprints. For instance, for the BLE AoA dataset, HDRF-CD achieves DFIS of 3.2623, 3.3139, and 3.2773 for Cross=0, Cross=1, and Cross=2, respectively. In contrast, the AF-DCGAN method obtains 1.0473, 1.1243, and 1.0734. The superior sample diversity exhibited by CDM can be attributed to its ability to capture intricate details of packet-level features and condition the sampling process with packet information to generate a diverse set of outputs.

TABLE III
MEAN RMSE PERFORMANCE ON THE BLE DATASET WITH DIFFERENT AMOUNT OF TRAINING AND SYNTHETIC DATA

Method and data amount	RMSE performance (m)		
	Cross 0	Cross 1	Cross 2
Baseline 0.5	0.1317	0.1237	0.1514
Baseline 1	0.0921	0.0909	0.1139
AF-DCGAN [2] 1 (0.5 gen)	0.1277	0.1229	0.1485
Chen et al. [1] 1 (0.5 gen)	0.1271	0.125	0.1428
Alhomayani et al. [3] 1 (0.5 gen)	0.1215	0.1154	0.1373
Noise 1(0.5 gen)	0.1125	0.1108	0.134
HDRF-CD 1(0.5 gen)	**0.1011**	**0.0977**	**0.1214**
AF-DCGAN [2] 1.5 (0.5 gen)	0.0913	0.0875	0.1142
Chen et al. [1] 1.5 (0.5 gen)	0.0952	0.0958	0.1153
Alhomayani et al. [3] 1.5 (0.5 gen)	0.0872	0.0912	0.1118
Noise 1.5 (0.5 gen)	0.0836	0.0828	0.1017
HDRF-CD 1.5 (0.5 gen)	**0.0742**	**0.0753**	**0.0979**

TABLE IV
MEAN RMSE PERFORMANCE ON THE CSI DATASET WITH DIFFERENT AMOUNT OF TRAINING AND SYNTHETIC DATA

Method and data amount	RMSE performance (m)		
	Cross 0	Cross 1	Cross 2
Baseline 0.5	1.2531	1.2093	1.3586
Baseline 1	0.874	0.7913	0.8514
AF-DCGAN [2] 1 (0.5 gen)	1.1618	1.1546	1.2127
Chen et al. [1] 1 (0.5 gen)	1.2415	1.242	1.3381
Alhomayani et al. [3] 1 (0.5 gen)	1.2162	1.2255	1.3515
Noise 1(0.5 gen)	1.2118	1.1976	1.293
HDRF-CD 1(0.5 gen)	**1.0834**	**1.0673**	**1.1423**
AF-DCGAN [2] 1.5 (0.5 gen)	0.8252	0.7431	0.8131
Chen et al. [1] 1.5 (0.5 gen)	0.8943	0.8156	0.8779
Alhomayani et al. [3] 1.5 (0.5 gen)	0.8656	0.7876	0.8375
Noise 1.5 (0.5 gen)	0.8503	0.7645	0.8642
HDRF-CD 1.5 (0.5 gen)	**0.7381**	**0.7192**	**0.7457**

TABLE II
MEAN DFIS PERFORMANCE COMPARISONS ON TWO HDRF DATASETS

Dataset	Method	DFIS		
		Cross=0	Cross=1	Cross=2
BLE AoA	Chen et al. [1]	0.4301	0.4341	0.4242
	Alhomayani et al. [3]	0.3441	0.3286	0.3251
	AF-DCGAN [2]	1.0473	1.1243	1.0734
	HDRF-CD_NC	2.9799	2.9527	2.9159
	HDRF-CD	**3.2623**	**3.3139**	**3.2773**
Wi-Fi CSI	Chen et al. [1]	0.7234	0.7231	0.7016
	Alhomayani et al. [3]	0.4701	0.4093	0.4041
	AF-DCGAN [2]	2.6251	2.6555	2.5499
	HDRF-CD_NC	3.1447	3.2309	3.2126
	HDRF-CD	**3.4636**	**3.6626**	**3.6444**

Besides, by comparing the performance of "HDRF-CD_NC" and "HDRF-CD", the proposed sample selection strategy can achieve higher DFIS. For example, for the Wi-Fi CSI dataset, the DFIS of HDRF-CD is higher by around 12% compared to HDRF-CD_NC, demonstrating the effectiveness of the developed selection strategy in improving the sample quality.

C. Positioning Performance Improvements

To further evaluate the synthetic HDRF fingerprints on downstream positioning tasks, the positioning models introduced in subsection V.A.2 were trained with the synthetic fingerprints and the adaption strategy proposed in subsection IV.D. We also compared the performance of a commonly used fingerprint augmentation technique, which directly adds Gaussian noise to the datasets. Given that the fingerprint values in both datasets are scaled to the same range of [-1, 1], in the following experiments, noise of zero mean and 0.01 standard deviation is applied. The positioning model was independently trained 5 times for each method under each cross-validation dataset to obtain the statistical performance. The mean RMSE results are summarized in Table III and Table IV. In the column "Method and data amount", each row contains numbers representing the amount of training data, which are multiples of the original training set size. "Baseline 0.5" and "Baseline 1" show the positioning performance using half of the original training dataset and the entire dataset without synthesized fingerprints. Other rows show the performance when using the synthesized fingerprints. The number before the parentheses indicates the total data used, while the number in the parentheses indicates the synthesized fingerprints in the total dataset. For example, "1 (0.5 gen)" means that the training data amount is the same as the original training dataset, and half of them is the synthetic data. "1.5 (0.5 gen)" represents the use of the full training dataset alongside synthetic HDRF fingerprints, which is half the size of the original training dataset. In these cases, the positioning model is first trained using only the original training dataset, followed by training with both the original data and the synthesized data. The results show that:

1) The synthetic HDRF fingerprints from HDRF-CD could consistently improve the RMSE performance on both datasets. In the BLE cross 0 dataset, the "HDRF-CD 1 (0.5 gen)" achieved a 23.2% reduction in positioning error (from 0.1317 to 0.1011) compared to the "Baseline 0.5". Similarly, in the CSI cross 2 dataset, HDRF-CD 1(0.5 gen) showed a 15.9% reduction in positioning error compared to the "Baseline 0.5". In contrast, the generative model AFDCGAN improved performance on the CSI dataset (for example, AF-DCGAN 1 (0.5 gen) achieved around 7.5% reduction compared to the "Baseline 0.5" among three cross datasets) but showed limited improvements on the BLE dataset (around 1.9% reduction in the same case). Additionally, the "Noise" data augmentation approach achieved performance improvement on the BLE dataset but had less impact on the CSI dataset. The results demonstrate the ability of HDRF-CD to generate different types of HDRF fingerprints.

2) HDRF-CD learned the HDRF fingerprint patterns in the vicinity of locations and the synthesized fingerprints achieved the highest performance improvement in generalizing the model to unseen data. By comparing the case of "Baseline 1" with HDRF-CD 1.5 (0.5 gen), the utilization of additional synthetic HDRF fingerprints resulted in further improvements of around 16.9% and 12.4% on the three BLE cross datasets and CSI cross datasets, respectively.

3) The improvement from using the synthetic HDRF fingerprints of HDRF-CD (for example, for the BLE cross 0 dataset, compared to the case of Baseline 0.5, "HDRF-CD 1 (0.5 gen)" yields an improvement of 23.2%) was less than the improvement achieved using the same amount of actual data ("Baseline 1" yields an improvement of 30%). The difference can be interpreted that while sampling from a noisy fingerprint of intermediate timesteps maintains spatial correlation at the fingerprint level, it concurrently constrains feature diversity. Additionally, at the packet feature level, despite employing sample selection methods based on similarity to the training set and using distance-based criteria, the verification of data authenticity remained uncertain. Future research can focus on addressing the presence of adversarial samples within the synthetic HDRF fingerprint datasets.

VI. CONCLUSIONS

This research addresses the fingerprint generation challenges in indoor positioning systems to enhance positioning performance. The key contributions involve generative model training, fingerprint sampling, and fingerprint selection. By decomposing fingerprints into packet-level features, the developed diffusion generative model learns the conditional packet-level data distribution leveraging the readily accessible High-Dimensional Radio Frequency (HDRF) packet information, such as grid points, antennas, anchors, and

frequency channels. Samples from the noisy fingerprints are then taken at intermediate time steps to preserve the spatial correlation of features within each HDRF fingerprint. To address the disparity between the packet-level distribution and the noise prior distribution employed in training, a scale regularization on the predicted clean data is introduced. To incorporate the synthetic HDRF fingerprints in downstream positioning tasks, an adaptive learning process in which the learned positioning model using the original training dataset is further retrained by learning from both the synthetic HDRF dataset and the training dataset.

The approach was assessed against the well-studied numerical experiments of NAME THEM HERE. Using a Diversity Fidelity Integrative Score (DFIS) that quantitatively assesses the quality of the synthesized HDRF samples, the results show that the samples from HDRF-CD exhibit the highest DFIS, indicating superior data diversity among the generated samples while a higher resemblance to the real data. In particular, with the BLE 5.1 and Wi-Fi CSI datasets, HDRF-CD achieves an average DFIS of 3.3 and 3.6 respectively, surpassing the best previous generation model (DFIS of 1.1 and 2.6 respectively) for HDRF fingerprints.

Applying the synthetic HDRF fingerprints to downstream positioning tasks further validates the effectiveness of HDRF-CD. The experimental results show that compared to using only the original training dataset, incorporating the synthetic HDRF fingerprints from HDRF-CD into the training of downstream positioning models can further reduce the positioning error by around 17% and 12% on the BLE and CSI datasets, respectively.

Future research can extend the HDRF-CD framework to generate data for unseen grid points. The spatial coordinate relationships of grid points can be encoded as generation conditions to enhance the generalization capability to unseen areas.

ACKNOWLEDGMENT

The work presented in this article is supported by the Centre for Advances in Reliability and Safety (CAiRS) admitted under AIR@InnoHK Research Cluster.

REFERENCES

[1] X. Chen, H. Li, C. Zhou, X. Liu, D. Wu, and G. Dudek, "Fidora: Robust WiFi-based indoor localization via unsupervised domain adaptation," *IEEE Internet of Things Journal,* vol. 9, no. 12, pp. 9872-9888, 2022.

[2] Q. Li *et al.*, "AF-DCGAN: Amplitude feature deep convolutional GAN for fingerprint construction in indoor localization systems," *IEEE Transactions on Emerging Topics in Computational Intelligence,* vol. 5, no. 3, pp. 468-480, 2019.

[3] F. Alhomayani and M. H. Mahoor, "Oversampling highly imbalanced indoor positioning data using deep generative models," in *2021 IEEE Sensors*, 2021: IEEE, pp. 1-4.

[4] R. Faragher and R. Harle, "Location fingerprinting with bluetooth low energy beacons," *IEEE journal on Selected Areas in Communications,* vol. 33, no. 11, pp. 2418-2428, 2015.

[5] P. S. Farahsari, A. Farahzadi, J. Rezazadeh, and A. Bagheri, "A survey on indoor positioning systems for IoT-based applications," *IEEE Internet of Things Journal,* vol. 9, no. 10, pp. 7680-7699, 2022.

[6] X. Wang, L. Gao, and S. Mao, "CSI phase fingerprinting for indoor localization with a deep learning approach," *IEEE Internet of Things Journal,* vol. 3, no. 6, pp. 1113-1123, 2016.

[7] M. Cominelli, P. Patras, and F. Gringoli, "Dead on arrival: An empirical study of the Bluetooth 5.1 positioning system," in *Proceedings of the 13th international workshop on wireless network testbeds, experimental evaluation & characterization*, 2019, pp. 13-20.

[8] P.-H. Tseng, Y.-C. Chan, Y.-J. Lin, D.-B. Lin, N. Wu, and T.-M. Wang, "Ray-tracing-assisted fingerprinting based on channel impulse response measurement for indoor positioning," *IEEE Transactions on Instrumentation and Measurement,* vol. 66, no. 5, pp. 1032-1045, 2017.

[9] Y. Ruan, L. Chen, X. Zhou, G. Guo, and R. Chen, "Hi-Loc: Hybrid indoor localization via enhanced 5G NR CSI," *IEEE Transactions on Instrumentation and Measurement,* vol. 71, pp. 1-15, 2022.

[10] N. Paulino, L. M. Pessoa, A. Branquinho, and E. Gonçalves, "Design and experimental evaluation of a Bluetooth 5.1 antenna array for angle-of-arrival estimation," in *2022 13th International Symposium on Communication Systems, Networks and Digital Signal Processing (CSNDSP)*, 2022: IEEE, pp. 625-630.

[11] M. Abbas, M. Elhamshary, H. Rizk, M. Torki, and M. Youssef, "WiDeep: WiFi-based accurate and robust indoor localization system using deep learning," in *2019 IEEE International Conference on Pervasive Computing and Communications (PerCom*, 2019: IEEE, pp. 1-10.

[12] A. Hilal, I. Arai, and S. El-Tawab, "DataLoc+: A data augmentation technique for machine learning in room-level indoor localization," in *2021 IEEE Wireless Communications and Networking Conference (WCNC)*, 2021: IEEE, pp. 1-7.

[13] G. G. Anagnostopoulos and A. Kalousis, "ProxyFAUG: Proximity-based fingerprint augmentation," in *2021 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, 2021: IEEE, pp. 1-7.

[14] K. M. Chen and R. Y. Chang, "A comparative study of deep-learning-based semi-supervised device-free indoor localization," in *2021 IEEE Global Communications Conference (GLOBECOM)*, 2021: IEEE, pp. 1-6.

[15] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," *arXiv preprint arXiv:1312.6114,* 2013.

[16] I. Goodfellow *et al.*, "Generative adversarial nets," *Advances in neural information processing systems,* vol. 27, 2014.
[17] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Semantic image segmentation with deep convolutional nets and fully connected crfs," *arXiv preprint arXiv:1412.7062,* 2014.
[18] R. Y. Chang, S.-J. Liu, and Y.-K. Cheng, "Device-free indoor localization using Wi-Fi channel state information for Internet of Things," in *2018 IEEE Global Communications Conference (GLOBECOM)*, 2018: IEEE, pp. 1-7.
[19] T. Lan, X. Wang, Z. Chen, J. Zhu, and S. Zhang, "Fingerprint augment based on super-resolution for WiFi fingerprint based indoor localization," *IEEE Sensors Journal,* vol. 22, no. 12, pp. 12152-12162, 2022.
[20] H. Zou *et al.*, "Adversarial learning-enabled automatic WiFi indoor radio map construction and adaptation with mobile robot," *IEEE Internet of Things Journal,* vol. 7, no. 8, pp. 6946-6954, 2020.
[21] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," *Advances in neural information processing systems,* vol. 33, pp. 6840-6851, 2020.
[22] C. Luo, "Understanding diffusion models: A unified perspective," *arXiv preprint arXiv:2208.11970,* 2022.
[23] J. Sohl-Dickstein, E. Weiss, N. Maheswaranathan, and S. Ganguli, "Deep unsupervised learning using nonequilibrium thermodynamics," in *International conference on machine learning*, 2015: PMLR, pp. 2256-2265.
[24] J. Song, C. Meng, and S. Ermon, "Denoising diffusion implicit models," *arXiv preprint arXiv:2010.02502,* 2020.
[25] A. Q. Nichol and P. Dhariwal, "Improved denoising diffusion probabilistic models," in *International Conference on Machine Learning*, 2021: PMLR, pp. 8162-8171.
[26] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer, "High-resolution image synthesis with latent diffusion models," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 10684-10695.
[27] D. Kingma, T. Salimans, B. Poole, and J. Ho, "Variational diffusion models," *Advances in neural information processing systems,* vol. 34, pp. 21696-21707, 2021.
[28] Y. Song, J. Sohl-Dickstein, D. P. Kingma, A. Kumar, S. Ermon, and B. Poole, "Score-based generative modeling through stochastic differential equations," *arXiv preprint arXiv:2011.13456,* 2020.
[29] P. Dhariwal and A. Nichol, "Diffusion models beat gans on image synthesis," *Advances in neural information processing systems,* vol. 34, pp. 8780-8794, 2021.
[30] J. Ho and T. Salimans, "Classifier-free diffusion guidance," *arXiv preprint arXiv:2207.12598,* 2022.
[31] A. Bansal *et al.*, "Universal guidance for diffusion models," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 843-852.
[32] S. Hong, G. Lee, W. Jang, and S. Kim, "Improving sample quality of diffusion models using self-attention guidance," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 7462-7471.
[33] V. Goel *et al.*, "Pair-diffusion: Object-level image editing with structure-and-appearance paired diffusion models," *arXiv preprint arXiv:2303.17546,* 2023.
[34] O. Maron and T. Lozano-Pérez, "A framework for multiple-instance learning," *Advances in neural information processing systems,* vol. 10, 1997.
[35] Y. Chen, Z. Yan, Y. Zhu, Z. Ren, J. Shen, and Y. Huang, "Data Augmentation for Environmental Sound Classification Using Diffusion Probabilistic Model with Top-K Selection Discriminator," in *International Conference on Intelligent Computing*, 2023: Springer, pp. 283-295.
[36] D. J. Im, H. Ma, G. Taylor, and K. Branson, "Quantitatively evaluating GANs with divergences proposed for training," *arXiv preprint arXiv:1803.01045,* 2018.
[37] D. Lopez-Paz and M. Oquab, "Revisiting classifier two-sample tests," *arXiv preprint arXiv:1610.06545,* 2016.
[38] C. Sakaridis, D. Dai, S. Hecker, and L. Van Gool, "Model adaptation with synthetic and real data for semantic dense foggy scene understanding," in *Proceedings of the european conference on computer vision (ECCV)*, 2018, pp. 687-704.
[39] M. Wulfmeier, A. Bewley, and I. Posner, "Incremental adversarial domain adaptation for continually changing environments," in *2018 IEEE International conference on robotics and automation (ICRA)*, 2018: IEEE, pp. 4489-4495.
[40] R. Ayyalasomayajula *et al.*, "Deep learning based wireless localization for indoor navigation," in *Proceedings of the 26th Annual International Conference on Mobile Computing and Networking*, 2020, pp. 1-14.
[41] Z. Lyu, T. T. Chan, G. C. Leung, D. P. Lun, and M. G. Pecht, "Counteracting Packet Loss in Fingerprint-based Indoor Positioning via Spatially Regularized Entropy and Ground-truth Prior Variational Inference," *IEEE Sensors Journal,* 2024.
[42] H. Abdel-Nasser, R. Samir, I. Sabek, and M. Youssef, "MonoPHY: Mono-stream-based device-free WLAN localization via physical layer information," in *2013 IEEE wireless communications and networking conference (WCNC)*, 2013: IEEE, pp. 4546-4551.
[43] K. Wu, J. Xiao, Y. Yi, D. Chen, X. Luo, and L. M. Ni, "CSI-based indoor localization," *IEEE Transactions on Parallel and Distributed Systems,* vol. 24, no. 7, pp. 1300-1309, 2012.
[44] J. Ding, Y. Wang, H. Si, S. Gao, and J. Xing, "Three-dimensional indoor localization and tracking for mobile target based on WiFi sensing," *IEEE Internet*

of Things Journal, vol. 9, no. 21, pp. 21687-21701, 2022.

[45] Y. Song and S. Ermon, "Generative modeling by estimating gradients of the data distribution," *Advances in neural information processing systems,* vol. 32, 2019.

[46] T. Karras, M. Aittala, T. Aila, and S. Laine, "Elucidating the design space of diffusion-based generative models," *Advances in Neural Information Processing Systems,* vol. 35, pp. 26565-26577, 2022.

[47] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*, 2015: Springer, pp. 234-241.

[48] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter, "Gans trained by a two time-scale update rule converge to a local nash equilibrium," *Advances in neural information processing systems,* vol. 30, 2017.