FM-Based Positioning via Deep Learning

Shilian Zheng, Jiacheng Hu, Luxin Zhang, Kunfeng Qiu, Jie Chen, Peihan Qi, *Member, IEEE*, Zhijin Zhao, and Xiaoniu Yang

***Abstract*— Frequency Modulation (FM) broadcast signals, regarded as opportunistic signals, hold significant potential for indoor and outdoor positioning applications. The existing FM-based positioning methods primarily rely on Received Signal Strength (RSS) for positioning, the accuracy of which needs improvement. In this paper, we introduce FM-Pnet, an end-to-end FM-based positioning method that leverages deep learning. This method utilizes the time-frequency representation of FM signals as network input, enabling automatically learning of deep features for positioning. We also propose two strategies, noise injection and enriching training samples, to enhance the model's generalization performance over long time spans. We construct datasets for both indoor and outdoor scenarios and conduct extensive experiments to validate the performance of our proposed method. Experimental results demonstrate that FM-Pnet significantly outperforms traditional RSS-based positioning methods in terms of both positioning accuracy and stability.**

***Index Terms*— FM signal, positioning, deep learning, convolutional neural network.**

I. Introduction

POSITIONING stands as a cornerstone technology extensively employed in modern society, with positioning navigation firmly ingrained in the fabric of everyday human life [1]. In a series of emerging applications such as Augmented Reality (AR) [2], Virtual Reality (VR) [3], and the Internet of Things (IoT) [4], positioning also plays a crucial role. Satellite navigation systems, such as the Global Positioning System (GPS) [5] and the Beidou Navigation Satellite System (BNSS) [6], serve as predominant positioning technologies widely employed in navigation systems. Notably, in outdoor environments with optimal signal reception, these systems exhibit high positioning accuracy, typically within a margin of no more than 10 meters (2D) for commercial GPS. However, GPS and BNSS rely on space-based signals, incurring substantial construction costs. Moreover, they demand precise timing accuracy for receiving devices. In addition, in densely populated or obstructed areas, particularly indoor environments, signal reception for satellite signals diminishes significantly, leading to a marked reduction in positioning accuracy. Thus, the imperative to develop technologies independent of existing satellite positioning arises, aiming to mitigate positioning costs, broaden the scope of positioning applicability, and enhance positioning robustness.

Navigation via Signals of Opportunity (NAVSOP) has been widely researched as a positioning methodology independent of satellite reliance [7]. Signals of opportunity refer to pre-existing non-navigation Radio Frequency (RF) signals permeating the electromagnetic environment. These signals include magnetic signals [8], analog modulated broadcast signals [9], [10], Digital Terrestrial Multimedia Broadcast (DTMB) signals [11], [12], Bluetooth signals [13], Ultra-WideBand (UWB) signals [14], [15], WiFi signals [16], [17], mobile communication signals [18], [19], visible light signals [20], all potentially exploitable for NAVSOP. Among these signal sources, WiFi signals have gained widespread attention in indoor positioning applications, primarily due to advantages such as the accessibility of router location information and extensive indoor coverage [21]. In cases where the emission location of the WiFi signal is known, the position of the observation point is usually inferred by analyzing the energy of the WiFi signal through the Received Signal Strength (RSS) method [22]. However, owing to the intricate propagation environment, RSS values received at a given location for continuously transmitted WiFi packets tend to be unstable, offering only approximate positional information. Consequently, researchers have proposed to use deep learning methods to learn from the Channel State Information (CSI) features of WiFi signals for precise positioning, thereby significantly improving the accuracy of WiFi-based indoor positioning [23], [24], [25]. Nonetheless, WiFi signals are characterized by limited coverage area and weak penetration capabilities, rendering them predominantly effective indoors. In contrast, Frequency Modulation (FM) broadcast signals (87.5MHz-108MHz) exhibit higher transmission power, longer wavelengths (around 3 m), enhanced penetration, wider coverage range, and robust reception across various environments, including open outdoor spaces, indoor settings with obstructive building structures, and fully enclosed underground garages. Therefore, FM-based positioning holds promise as a versatile positioning methodology applicable both indoors and outdoors [10].

Manuscript received 13 November 2023; revised 19 March 2024; accepted 9 May 2024. Date of publication 13 June 2024; date of current version 21 August 2024. This work was supported in part by the National Natural Science Foundation of China under Grant U19B2016 and in part by the National Basic Scientific Research of China under Grant JCKY2023110C099. *(Corresponding authors: Shilian Zheng; Xiaoniu Yang.)*

Shilian Zheng, Luxin Zhang, Kunfeng Qiu, and Xiaoniu Yang are with the Innovation Studio of Academician Yang, National Key Laboratory of Electromagnetic Space Security, Jiaxing 314033, China (e-mail: lianshizheng@126.com; lxzhangMr@126.com; yexijoe@163.com; yxn2117@126.com).

Jiacheng Hu, Jie Chen, and Zhijin Zhao are with the College of Communication Engineering, Hangzhou Dianzi University, Hangzhou 310018, China (e-mail: 222080242@hdu.edu.cn; 211080032@hdu.edu.cn; zhaozj03@hdu.edu.cn).

Peihan Qi is with the State Key Laboratory of Integrated Service Networks, Xidian University, Xi'an 710071, China (e-mail: phqi@xidian.edu.cn).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/JSAC.2024.3413961.

Digital Object Identifier 10.1109/JSAC.2024.3413961

Positioning technologies leveraging FM sources typically rely on the RSS of FM signals at the test point as a primary feature for positioning, distinguishing varying RSS values across different positions using probabilistic or deterministic methods [26], [27], [28]. These existing FM-based positioning methods have certain limitations. Firstly, radio towers are often far from the positioning points, and the inherent unpredictability of signal fading due to multipath interference along the signal propagation path leads to inevitable errors in methods reliant on FM signal propagation models. Secondly, FM signals lack timing information, thereby necessitating the utilization of the energy value of the signal as a feature for FM-based positioning in current methodologies. However, the granularity of RSS features is often insufficient, providing limited information. In open areas, the RSS values of closely located positions can vary subtly, and due to the uncertainty of noise, the energy values at the same location can also differ significantly over different instances. Whether using probabilistic or deterministic methods, FM signal fingerprint positioning based on RSS typically suffers from inadequate positioning accuracy.

In addition to energy differences, FM signals received at different positions encapsulate a wealth of additional information. When wideband FM signals are received at a fixed location with a high sampling frequency, the amount of information contained is enormous. Leveraging this wealth of information could significantly aid in discerning between closely situated points. Deep learning methods have demonstrated remarkable feature learning capabilities and have found extensive applications across diverse domains such as image processing, natural language processing, and wireless signal processing [29], [30], [31]. Motivated by this, in this paper, we propose an FM-based positioning method using deep learning, denoted as FM-Pnet. Diverging from traditional FM-based positioning methods primarily reliant on RSS, our method harnesses the high-precision time-frequency representation of FM signals acquired with a certain bandwidth, the maximum of which can cover the entire FM broadcast frequency band. Through the application of Short-Time Fourier Transform (STFT) to obtain the time-frequency representation of the signal, we utilize it as the input for a deep neural network to estimate the position. Aligned with the tenets of deep learning, FM-Pnet consists of two stages: offline training and online inference. In the offline training stage, FM signals from known positions are gathered to construct a dataset for training the model parameters. In the subsequent online inference stage, FM signals are received at the target location, subjected to time-frequency transformation, and input to the well-trained model for position estimation based on output confidence scores. In order to improve the model's generalization ability over a large time span, we also propose two strategies to enhance its robustness. Our contributions include the following aspects:

- We introduce a new FM-based positioning framework, namely FM-Pnet, which applies deep learning to automatically learn features from time-frequency representations of FM signals for positioning. It achieves satisfactory positioning performance in both indoor and outdoor scenarios.
- We further employ two enhancement strategies, namely noise injection and enriching the training samples, to improve the FM-Pnet model's generalization over long time spans.
- We provide an attention mechanism that can be embedded at the beginning of the positioning model to process the input time-frequency representation, which can be used to visually analyze which parts of the time-frequency representation play a crucial role in positioning.
- We collect FM signals from different locations in both indoor and outdoor scenarios and establish multiple datasets with both raw FM signals and energy features which can be used to evaluate the performance of FM-based positioning methods.
- We conduct extensive experiments to evaluate the performance of our proposed FM-Pnet in various scenarios. The experimental results show that the proposed FM-Pnet outperforms three existing RSS-based methods. The performance of FM-Pnet can be further enhanced in the frequency domain or in the time domain by increasing the bandwidth of the collected signals or extending the length of each sample. Additionally, through the two proposed enhancement methods, the generalization of FM-Pnet over a large time span can be improved.

The rest of this paper are organized as follows. In Sec. II, we discuss the related work of FM-based positioning. In Sec. III, we formulate the problem. In Sec. IV, we introduce our proposed FM-Pnet in detail. In Sec. V, we compare the positioning performance of FM-Pnet with existing methods. Finally, we provide the concluding remarks in Sec. VI.

II. Related Work

Currently, FM-based positioning has been extensively studied in both outdoor and indoor environments. The positioning technology using FM sources usually utilizes the RSS of the FM signal at the test point as a feature for location and distinguishes different RSS at different positions using probabilistic or deterministic methods.

In outdoor scenarios, probabilistic methods are commonly employed for positioning. These methods consider that each position corresponds to a random variable RSS vector. The location with the highest probability can be determined by searching for the actual RSS vector values. For example, as stated by Fang et al. [26], the observed RSS values at each position are modeled as a Gaussian distribution, and the location corresponding to each RSS value is estimated using the probability distribution. Feasibility experiments were conducted in two different large-scale outdoor environments, urban campuses, and rural mountainous areas. Besides, many studies have combined probabilistic methods with large-scale FM signal propagation models to adapt to the vast outdoor environments [32], [33], [34]. Youssef et al. [32] marked the positions of radio towers on a map and constructed an energy fingerprint map using the FM signal propagation model. In actual measurements, they collected signals from different positions within a large city area and estimated the positions using energy histograms. However, due to complex

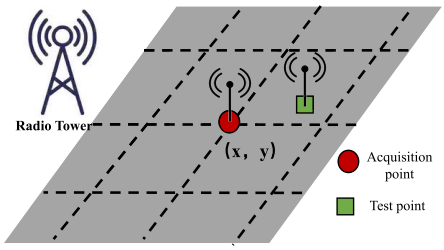


Fig. 1. Problem modeling for FM-based Positioning.

channel propagation, there was a notable disparity between the energy fingerprint map and reality, leading to an experimental distance error of 8 miles. Kumar et al. [34] proposed an algorithm based on a large-scale simulated map construction model, which achieved an average positioning error of 5 miles through experiments using over 900 power spectrums distributed over a range of over 200 miles. Mukherjee et al. [33] used the FM spectrum estimation model and probabilistic methods (Bayesian decision) to achieve a positioning error of 0.12 miles.

In indoor scenarios, where spatial dimensions are typically smaller, establishing a signal database is comparatively straightforward, facilitating the selection of a broader array of positioning methods, both probabilistic and deterministic. Yoon et al. [35] extended the probabilistic approach to indoor settings and established an indoor propagation model for FM signal RSS fingerprints, achieving average errors of approximately 6 meters (m) and 10 m in two indoor scenarios – a campus and a downtown area. In contrast to probabilistic methods, deterministic methods estimate location by considering actual measured values. According to [27], after normalizing all collected RSS values, three methods – K-Nearest Neighbors (KNN), Support Vector Machine (SVM), and Gaussian Process (GP) regression – were applied for positioning tests. Accounting for the temporal variability of RSS energy fingerprints, Popleteev [36], [37] collected long-term FM data at fixed locations to evaluate positioning performance. Chen et al. [28] conducted experiments in three different indoor scenarios to compare positioning using FM signal and WiFi signal RSS fingerprints, achieving similar accuracy. Attempts were made to combine WiFi and FM RSS fingerprints, resulting in improved positioning performance. Additionlly, Moghtadaiee and Dempster [9] independently utilized probabilistic methods (Histogram) and deterministic methods (KWNN), subsequently amalgamating them (Combined) to achieve an average distance error of approximately 2.5 m.

III. Problem Formulation

The positioning problem based on FM signals can be described as inferring the current location based on the features extracted from the received FM signals. As shown in Fig. 1, let's consider a flat area to be positioned, which can be divided into several grid blocks. The target point to be positioned lies within one of these grids. For a target point (x, y), it receives FM signals. Adopting a discrete time formulation, the received FM signal can be represented as

$$r(n) = \sum_{i=1}^{M} h_i(n) * s_i(n) + w(n),\ n = 0, 1, \ldots, L-1, \quad (1)$$

where $s_i(n)$ represents the narrowband FM signal in discrete time format, $h_i(n)$ denotes the channel response, $*$ denotes convolution, M represents the number of narrowband FM signals, L is the signal length, and $w(n)$ is the noise such as Additive White Gaussian Noise (AWGN). From Eq. (1), it's evident that transmission paths differ among various points relative to the radio tower, resulting in diverse channel responses and received signals. Consequently, distinct locations can be identified through the received FM signals. However, due to the variability of obstacles in the transmission path and the uncertainty of noise, there exists temporal variability in signals received at the same points, posing a considerable challenge for positioning. As shown in Fig. 1, if FM signals can be acquired from each known point through prior preparation, the problem that needs to be solved for FM-based positioning is to establish the mapping between the received FM signals and their respective receiving location coordinates:

$$\mathcal{M}_{\mathrm{FM}} : r_\alpha(n) \rightarrow \mathrm{Loc}_\alpha(x, y), \quad (2)$$

where $r_\alpha(n)$ represents the received FM signal in location α and $\mathrm{Loc}_\alpha(x, y)$ represents the coordinate of location α. Additionally, it is desirable to ensure that the distance between the inferred location obtained through this mapping during testing and the actual location is minimized as much as possible:

$$\min ||\mathcal{M}_{\mathrm{FM}}(r_\beta(n)) - \mathrm{Loc}_\beta(x, y)||. \quad (3)$$

In this paper, we propose to establish the mapping relationship $\mathcal{M}_{\mathrm{FM}}$ using deep learning methods.

IV. Methodology

A. Overall Framework

The method of using FM signals for positioning based on deep learning primarily hinges on the powerful feature extraction ability of neural network models. By leveraging a large amount of known position information associated with FM signal data, the network model continuously learns to map the input FM signal data to corresponding location coordinates through self-adaptation.

The proposed FM-Pnet in this paper enables end-to-end positioning by using FM signals, as illustrated in Fig. 2. The framework selects the corresponding time-frequency representation computed from the raw FM signal as input to the deep learning model. This ensures that the original signal information is preserved while facilitating the extraction of hidden information by the subsequent neural network. For the network part, we employ the ResNeXt model, renowned for its exceptional performance in image classification tasks, to process the input data of the dual-channel time-frequency representation. Unlike the original ResNeXt, we adapt its stack depth and feature dimension to better suit the characteristics of the input data.

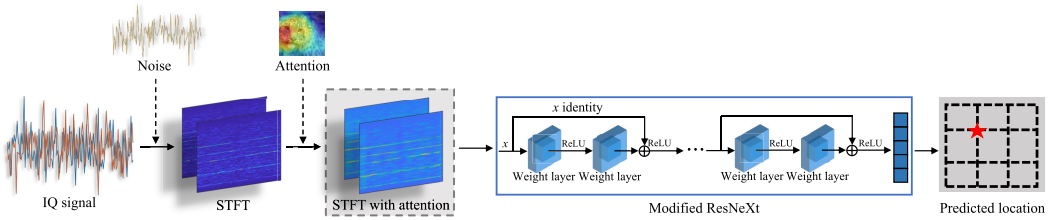


Fig. 2. The overall framework of the proposed positioning method FM-Pnet.

Considering the impact of time span on FM-based positioning, FM-Pnet can choose to introduce noise to the original signal prior to mapping the FM signal into a time-frequency representation. This augmentation enhances the robustness, adaptability, and generalization capability of the trained deep learning model, enabling it to better adapt to FM-based positioning over long time spans. Furthermore, recognizing the sparsity of broadcast signals within the FM frequency band, FM-Pnet can incorporate an attention mechanism after obtaining the time-frequency representation data. This attention mechanism can be trained together with the subsequent neural network model through self-learning, enabling it to focus on the more important parts of the time-frequency representation.

FM-Pnet consists of two main processes: offline training and online positioning. During the model training, we formulate the positioning problem as a classification problem, leveraging the training paradigm of deep learning to obtain a powerful classification model. Specifically, we treat each localization coordinate as a separate class, and FM-Pnet predicts the corresponding predicted position as the class with the highest probability. During the online positioning stage, where the trained classification model is utilized, we employ a weighted localization method similar as [23] to avoid being restricted to selecting a position only from the given coordinates.

In summary, the deep learning-based positioning method proposed in this paper, FM-Pnet, can leverage the powerful feature extraction ability of neural network models to uncover deeper and more effective information embedded within the original FM signals, thereby enhancing positioning accuracy.

B. Time-Frequency Representation

FM signals employ a frequency modulation technique to modulate the desired broadcast information onto the carrier signals. This means that the variations in the frequency of an FM signal can faithfully mirror the dynamic characteristics of sound or musical content. Frequency modulation introduces non-stationarity to the broadcast signal since the frequency varies over time. As a conventional approach for analyzing non-stationary signals, STFT is proficient at capturing the attributes of FM signals in both the time and frequency domains. In our approach, We apply STFT to perform time-frequency transformation on the original FM signal and utilize the transformed time-frequency representation as the input for the positioning model. The STFT involves the partitioning of the signal into numerous short, overlapping segments, each subjected to windowing to mitigate spectral leakage, followed by individual application of a Fourier transform:

$$\mathrm{STFT}_r(m,k)=\sum_{n=0}^{L-1}\left[r(n)g(n-m)\right]e^{-j(2\pi(n-m)k/p)}, \quad (4)$$

where $k=0,1,\cdots,p-1$, p denotes the length of the window function, which is conventionally chosen as a power of 2 to enhance the efficient utilization of the Fast Fourier Transform (FFT) and is also known as FFT size, $r(n)$ is the original FM signal, $g(\cdot)$ represents the window function employed in this paper. This transformation, transitioning from the time domain to the frequency domain provides insights into the signal's frequency distribution across different time intervals. Fig. 3 presents time-frequency representations of FM signals received at three distinct time instances. As shown in Fig. 3, the horizontal axis corresponds to the time dimension, while the vertical axis represents frequency information. Various colors indicate the different values of sample sequence of the FM signal, with brighter yellow shades corresponding to larger value.

After performing STFT, we can obtain several signal segments whose quantity is relevant to the FFT size and the overlap ratio. The overlap ratio refers to the proportion of overlapping samples between adjacent time windows relative to the window size. These two factors also affect the resolution of the obtained time-frequency representation. Specifically, The FFT size affects the frequency resolution of the time-frequency representation, with larger value yielding finer frequency resolution. A higher overlap ratio indicates a larger portion of overlap between adjacent windows, which can yield higher resolution in time. Given that the number of signal segments after STFT is q, the obtained time-frequency representation can be represented as a complex matrix with dimensions of $p\times q$. To preserve all pertinent information, we separately extract the real and imaginary components of this complex matrix and create a real matrix with dimensions of $p\times q\times 2$ as the input of the proposed FM positioning model:

$$\mathcal{S}(i,j,0)=(\mathrm{Real}(\mathrm{STFT}_r(\cdot)))_{i,j}, \quad (5)$$

$$\mathcal{S}(i,j,1)=(\mathrm{Imag}(\mathrm{STFT}_r(\cdot)))_{i,j}, \quad (6)$$

where $\mathcal{S}$ is the final time-frequency representation which is used as the model input, $\mathrm{Real}(\cdot)$ and $\mathrm{Imag}(\cdot)$ are the operations of extracting the real and imaginary parts, $(\cdot)_{i,j}$ represents the element at the i-th row, j-th column and $0\le i\le p-1$, $0\le j\le q-1$.

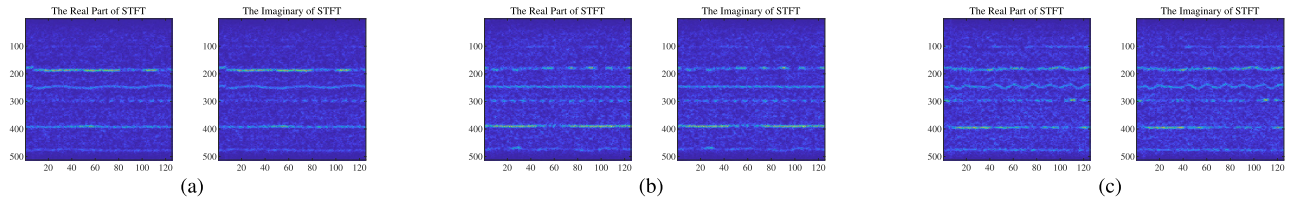


Fig. 3. The time-frequency representation under different reception moments. The sampling bandwidth is 4MHz. (a) Day1, (b) Day2, and (c) Day3.

C. Adopted Network Structure

Convolutional Neural Network (CNN) relies on its hierarchical structural design to excel in feature extraction for images and spatial structured data [38]. However, in CNNs, as the number of convolutional layers increases, issues such as degradation and gradient explosion phenomena may occur. Residual Network (ResNet) effectively solves this problem by incorporating skip connections, enabling the creation of deeper networks while maintaining stronger performance [39]. With the development of ResNet, ResNeXt [40] applies the concept of grouped convolution, which significantly enhances performance without increasing computational cost, demonstrating outstanding robustness in computer vision tasks [41]. The ResNeXt network consists of a series of residual blocks, which have the same topology and follow two fundamental principles: if the spatial mapping of the module is of the same size, the topology in the module shares the same hyperparameters (such as width and filter size); every time the spatial mapping is downsampled by a factor of 2, the module's width is doubled. In ResNeXt, the form of splitting-transforming-merging can be represented as:

$$R(x) = \sum_{i=1}^{C} T_i(x), \tag{7}$$

where T_i represents the same topological structure, C denotes the number of identical branches within a module, typically referred to as the cardinality, which serves as an additional measure of model complexity. Although the size of the width is related to the number of simple transformations (inner products), the size of the cardinality governs the quantity of complex transformations. Empirical evidence demonstrates that cardinality is a fundamental dimension, and increasing cardinality is more effective than enlarging the width and depth of CNNs, resulting in enhanced model expressiveness.

Since the time-frequency representation $\mathcal{S}$ of an FM signal can also be considered as an image, in terms of the network structure, we adopt a network similar to ResNeXt to learn and extract features for positioning. For our FM-Pnet network, C is set to 32. To better learn from the time-frequency representation of input FM signals, we modify the network structure by increasing the number of stacking layers. Table I presents the detailed structure of the network.

TABLE I
BASIC STRUCTURE OF THE NETWORK IN FM-PNET

Layer	Detail
conv1	$3 \times 3, 64$, stride
conv2	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64\ (C = 32) \\ 1 \times 1, 128 \end{bmatrix} \times 3$
conv3	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128\ (C = 32) \\ 1 \times 1, 512 \end{bmatrix} \times 4$
conv4	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256\ (C = 32) \\ 1 \times 1, 512 \end{bmatrix} \times 5$
conv5	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512\ (C = 32) \\ 1 \times 1, 1024 \end{bmatrix} \times 6$
conv6	$\begin{bmatrix} 1 \times 1, 1024 \\ 3 \times 3, 1024\ (C = 32) \\ 1 \times 1, 2048 \end{bmatrix} \times 3$
pooling	Adaptive Average Pool
FC	G-d fc, softmax

D. Training and Inference

FM-Pnet consists of two main processes: offline training and online positioning. The offline training process involves the training of a deep learning positioning model using collected FM signal data. Conversely, online positioning entails employing the trained model to estimate the location based on the received FM signal in real time.

1) Offline Training: As shown in Fig. 1, the positioning problem can be transformed into a positional classification problem with G categories by dividing the region of interest into G grids. For example, if the positioning area is divided into a 3×3 positioning grid with a total of 9 positioning points, the positioning problem becomes a 9-classification task. FM-Pnet classifies FM signals received at location i into the i-th category. Through this problem transformation, the deep learning paradigm is applied to positioning based on FM signals to obtain a powerful positioning model. We can train the model using cross-entropy loss to measure the error between the predicted location and the true label, which can be represented as:

$$\ell = -\sum_{i=1}^{G} \eta_i \log(\rho_i(\mathcal{S}|f_\theta)), \tag{8}$$

where η_i represents the i-th element of the label of the time-frequency representation $\mathcal{S}$, $\rho_i(\mathcal{S}|f_\theta)$ represents the probability of the FM-Pnet network $f_\theta(\cdot)$ prediction $\mathcal{S}$ belongs to class i. For batch data, the batch loss is calculated separately for individual samples and averaged. By optimizing the network parameters using this loss function, we can obtain a well-trained network model.

2) Online Positioning: After the training stage of FM-Pnet, we test the well-trained positioning model with the untrained FM signals that share the same time-frequency format with the training data. To enhance the accuracy of location estimation,

we refine the results through Bayesian reevaluation instead of relying solely on the labeled locations obtained from the trained model. In particular, we calculate the weighted sum of the confidence scores for each predicted class and their corresponding labeled positions to obtain the final position estimation results [23]:

$$x_\rho = \sum_{i=1}^{G} \rho(i) * x(i), \tag{9}$$

$$y_\rho = \sum_{i=1}^{G} \rho(i) * y(i), \tag{10}$$

where x_ρ and y_ρ are the horizontal and vertical coordinates of ultimate position outcome, G is the number of labeled positions during the training stage, $\rho(i)$ denotes the confidence score of i-th prediction position and $x(i)$, $y(i)$ are the horizontal and vertical coordinates of i-th labeled position.

E. Generalization Enhancement Strategies

The electromagnetic environment at the target location are subject to continual variation over time, influenced by factors such as weather conditions, which may affect the propagation characteristics of FM signals. Therefore, FM signals received at the same location but at different times will inevitably exhibit variations. When constructing a training set for FM signals, using data collected over a short time span may limit the diversity of samples, resulting in models that are ill-equipped to adapt to long-term changes in the channel environment. Fig. 3 presents the time-frequency representations of signal samples collected at the same indoor location on different days, clearly showing significant differences in their characteristics. To address this issue, this paper proposes two strategies to enhance the model's generalization performance over long time spans: noise injection augmentation and enriching the training samples.

1) Noise Injection Enhancement: Given the inherent fluctuations in noise power during FM signal reception, the time-frequency representation tends to manifest differences at the same location. In real-world electromagnetic environments, such fluctuations are commonplace. When training models based on FM signal datasets with limited samples, the models cannot adapt to various scenarios under different noise power levels. In other words, the non-stationary characteristics of the noise are likely to elevate the likelihood of errors at the inference of the FM positioning model. To enhance the FM positioning model's robustness against variations in noise power, we introduce random noise at various power levels into the original IQ signals during the training stage. This enhancement strategy may enable the effective performance of deep learning models in positioning tasks across prolonged time intervals. The operation of adding noise is expressed as:

$$r'(n) = r(n) + \mathcal{N}(n), \tag{11}$$

where $\mathcal{N}(n)$ denotes the additive noise and $r'(n)$ is the FM signal after adding noise.

2) Enriching Training Samples: The noise injection method enhances the training dataset by artificially introducing variations in noise, with the expectation that the model can adapt to real changes. Another strategy we consider is to maximize the diversity of training samples with actual data. Diversity does not simply refer to a higher quantity but rather to a distribution that encompasses as much variability as possible, including samples with longer time spans. One implementation approach is to collect multiple rounds of FM signals at the same location during the signal acquisition stage (e.g., batch data collection every day for several days) and then incorporate the data from multiple rounds into the training set. It improves the robustness of the deep learning model trained to adapt to changes in the electromagnetic environment, enabling it to achieve accurate positioning even over longer time spans. Compared to the noise injection method, enriching the training samples requires more human effort during the data collection stage.

F. Spatial Attention Mechanism

The attention mechanism aims to make neural network models learn to ignore irrelevant information as much as possible while focusing more on important information. Considering that the received FM signals within the sampling frequency band only contain a small number of narrowband FM signals, while most of the signals are noise, we introduce the spatial attention mechanism to observe the model's ability to focus on signals within the band.

Due to the local receptive field of convolutional neural networks, each channel unit cannot utilize information outside the receptive field region. To address this, the spatial attention mechanism first takes the maximum and average values of the dual-channel time-frequency representation input matrix in the channel dimension to achieve information aggregation across all channels. Then, the matrix of the dual channels is concatenated in the channel dimension as

$$Q_{\max} = \max_c(M_s), \tag{12}$$

$$Q_{mean} = \mathrm{mean}_c(M_s), \tag{13}$$

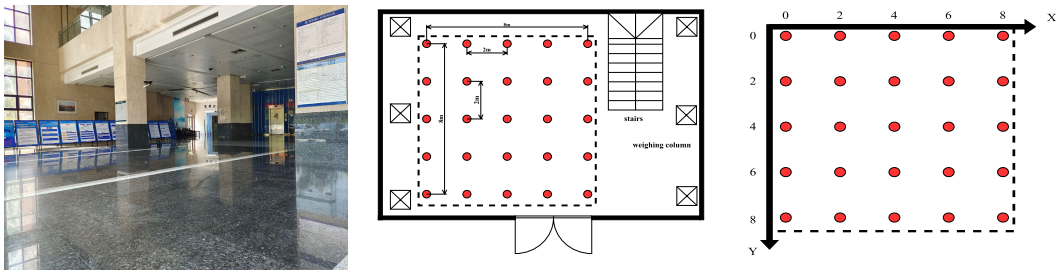
$$Q_{mix} = \mathrm{Cat}(Q_{\max}, Q_{mean}), \tag{14}$$

where $M_s \in \Re^{2\times p\times q}$ represents the input dual-channel time-frequency representation, $\max_c(\cdot)$ and $\mathrm{mean}_c(\cdot)$ respectively denote taking the maximum and average values in the channel dimension, and $\mathrm{Cat}(\cdot)$ represents concatenation in the channel dimension. Then, two layers of 2D convolutional layers and one layer of non-linear activation function are used to aggregate information from different channels, resulting in a single-channel feature matrix. Finally, the attention weight matrix is obtained by applying the Sigmoid function, which restricts the element values between 0 and 1:

$$Q_1 = \mathrm{ReLU}(Conv2D_1(Q_{mix})), \tag{15}$$

$$Q_2 = \mathrm{Sigmoid}(Conv2D_2(Q_1)), \tag{16}$$

where $Q_2 \in \Re^{p\times q}$ is the attention weight map, $Conv2D_1(\cdot)$ is a 2-input and 32-output channel, 2D convolutional layer with a kernel size of 5, a stride of 1, and a padding of 2. $Conv2D_2(\cdot)$ is a 32-input and 1-output channel, 2D convolutional layer with a kernel size of 7, a stride of 1, and a



(a) (b) (c)

Fig. 4. Indoor scenario of a lobby. (a) Real scene image of the lobby, (b) floor plan of the lobby, and (c) (X, Y) coordinates of the collection points.

padding of 3. By element-wise multiplication of the original spectro-temporal input M_s and the obtained attention weight matrix Q_2, a feature matrix with attention information can be obtained. The calculation steps are represented as

$$Q_a = M_s \odot Q_2, \tag{17}$$

where $\odot$ represents the Hadamard product.

V. Experiment

A. Experimental Setup

1) Experimental Scenarios: To verify the performance of the proposed method, we conduct experiments in both indoor and outdoor scenarios.

For the indoor scenario, we select the first-floor lobby of the Science and Technology Museum at Hangzhou Dianzi University, Hangzhou, China, as shown in Fig. 4(a). The floor plan of the lobby is illustrated in Fig. 4(b). Within the lobby, we select a square area and place 25 collection points which are marked as dot circles. These 25 collection points form an 8×8 m^2 square area, with a standardized lateral (vertical) distance of 2 m between adjacent collection positions. As for outdoor scenario, we choose the outdoor plaza outside Jiaxing Railway Station, Jiaxing, China, as shown in Fig. 5(a). There are two rows of trees on both sides of the plaza, and the pedestrian flow on the plaza is relatively small. The plan view of the plaza is shown in Fig. 5(b). Within the unobstructed area without trees, we select a square area measuring 15×12 m^2 which is divided into 30 FM signal collection points. The horizontal (vertical) spacing between each collection point is set to 3 m. The third scenario is a large scale scenario chosen within the Smart Industry Park in Jiaxing, China, which features a more extensive and complex layout, including buildings, circular flower beds, green belts, and various trees, as illustrated in Fig. 6(a). The plan view is depicted in Fig. 6(b), wherein 357 signal collection points are strategically placed in the open space between buildings and flower beds, spanning an area of 24×30 m^2. In contrast to the outdoor scenario at the train station plaza, both the length and width have doubled in this large scale environment. The horizontal and vertical spacing between each collection point in this environment is set to 1.5 m. As shown in Fig. 4(c), Fig. 5(c) and Fig. 6(c), we abstract the horizontal and vertical directions of three areas into the X-axis and Y-axis. Similarly, we set the first collection point in the top left corner as the coordinate origin and established a Cartesian coordinate system on the plane. For each collection point, we assign a two-dimensional coordinate based on its actual physical position.

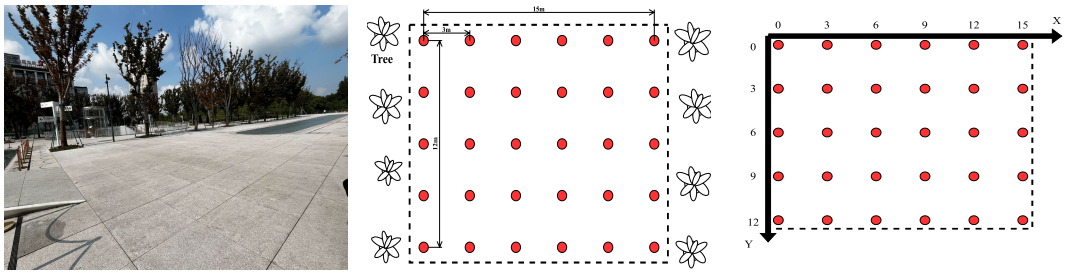
2) Datasets: We choose the center frequency for signal collection to be the center frequency of the FM signal band (87.5MHz-108MHz), specifically 97.5MHz. Data collection is conducted both indoors and outdoors over multiple days. We briefly mark each day as Day1, Day2, and so on. At each collection point, we gather signals continuously for 30 seconds each day. The FM signals collected at each collection point are then compiled to form the positioning datasets.[1]

To examine the impact of signal bandwidth on positioning, we collect FM signals with bandwidths of 320kHz (sampled at 400ksps) and 4MHz (sampled at 5Msps). Additionally, in the outdoor scenario, we acquire data samples with an expanded bandwidth of 20MHz (sampled at 25Msps) to assess the positioning performance with FM signals covering the entire frequency range. To examine the impact of FM signal duration on positioning accuracy, we extract signal samples with varying lengths: 1,024, 4,096, and 16,384. Furthermore, to investigate the effect of the time span on the positioning method, we collect FM signals in the same location over multiple consecutive days. Distinct datasets are created for each experiment. In subsequent experiments, unless explicitly stated otherwise, each dataset used comprises 8,000 training samples and 2,000 testing samples. The corresponding datasets will be made available upon the paper's publication.

3) Operating Environment: The experiments are conducted on a computer featuring an Intel Core i9-9900k CPU clocked at 3.6GHz and an NVIDIA GeForce RTX 2080 GPU. Model training is executed using the PyTorch framework. The Adam optimization algorithm is utilized for parameter updates during training, with a batch size of 32 and an initial learning rate of 0.001. The learning rate was halved every two rounds during the training process with a total of 10 epochs.

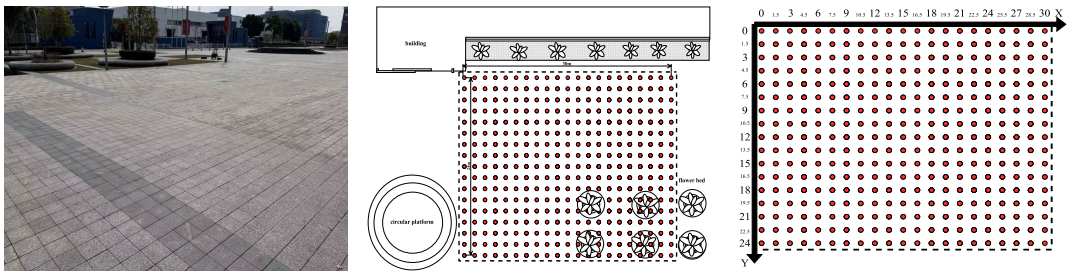
4) Baselines: We conduct a comparative analysis of our proposed method against existing techniques, namely deterministic KWNN [27], the probability-based histogram method (Histogram) [32], and the Combined method [9], which combines the KWNN and Histogram approaches. The Combined method entails computing the arithmetic average of the positioning prediction outcomes from both the KWNN and Histogram methods to derive its ultimate positioning result. For the KWNN method, we iterate over all possible values of K and select the one that yields optimal positioning

[1]The datasets are available at https://github.com/lianshizheng/FM-Pnet.



(a) (b) (c)

Fig. 5. Outdoor scenario of a square. (a) Real scene image of the square, (b) plan view of the square, and (c) (X, Y) coordinates of the collection points.



(a) (b) (c)

Fig. 6. Large-scale scenario. (a) Real scene image of the large-scale, (b) plan view of the large-scale, and (c) (X, Y) coordinates of the collection points.

performance. In the Histogram method, we determine the number of bins for calculating the histogram as the square root of the number of training samples. These three existing methods necessitate the use of RSS from FM signals as both training and testing samples for positioning. The number of channels for RSS calculation varies with the collected bandwidth [9]: 1 channel for the 320kHz bandwidth, 5 channels for the 4MHz bandwidth, and 22 channels for the 20MHz bandwidth. Following the calculation, we obtain RSS vectors with dimensions of 1, 5, and 22, respectively. In the experiments, random guessing is deemed as unsuccessful positioning, and the error associated with random guessing varies across different-sized areas.

5) Performance Metrics: We consider three performance metrics, all of which are derived from distance error calculations.

a) Mean distance error (MDE): MDE is a metric used to measure positioning error. It is calculated by taking the average of the Euclidean distance between the predicted and actual coordinates. The expression for MDE is as follows:

$$\text{MDE} = \frac{1}{N}\sum_{i=1}^{N}\sqrt{(x_i-\hat{x}_i)^2+(y_i-\hat{y}_i)^2}, \tag{18}$$

where (x_i, y_i) denotes the true coordinates of the test point, $(\hat{x}_i, \hat{y}_i)$ represents the predicted coordinates, and N is the sample size. The smaller the MDE, the closer the test result is to the true position, indicating better positioning performance. For random guessing, we randomly generate predicted coordinates and calculate the distance error with the corresponding true coordinates. The MDE values of random guessing for the indoor and outdoor scenarios are 4.6795 and 7.8234, respectively. Results exceeding these values are considered positioning failures. The unit of MDE is m.

b) Standard deviation (STD): STD is a widely used statistical measure employed to characterize the extent of data dispersion. It plays a crucial role in assessing the performance of positioning algorithms. It can be used to measure the dispersion of positioning errors by calculating the standard deviation of the test distance errors for all samples in the test:

$$\text{STD} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(D_i-\text{MDE})^2}, \tag{19}$$

where D_i represents the distance error of sample $\mathcal{S}_i$. A higher value of STD suggests a broader distribution of errors, whereas a lower value indicates a more concentrated distribution of errors.

c) Cumulative distribution function (CDF): CDF represents the distribution function of distance errors. As the distance error increases, the probability accumulates more rapidly towards 1, indicating better positioning performance. For simplicity, in the rest of the paper when referring to MDE, STD, and distance errors in CDF, we will exclude their units and provide only the numerical values.

B. Impact of STFT on FM-Pnet

1) Impact of FFT Size on FM-Pnet: We first examine the influence of varying FFT sizes on the subsequent positioning outcomes. The training and test sets are generated using data acquired with a 4MHz bandwidth on Day1. The signal length is fixed at 4,096. We explore FFT sizes of 32, 128, and 512. The CDF curve of the experimental results is shown in Fig. 7. As shown in Fig. 7(a), in the indoor scenario, compared to FFT sizes of 32 and 128, setting the FFT size to 512 results in more test samples with distance errors less than 0.5. Moreover, the CDF curve converges faster with an FFT size of 512. For

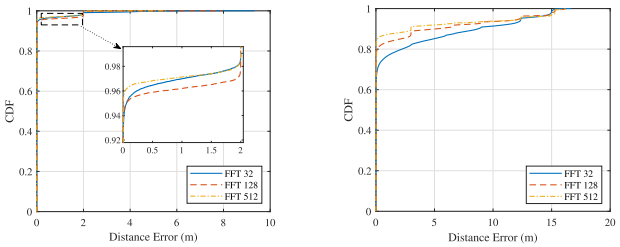


Fig. 7. The influence of different FFT sizes on the positioning performance of FM-Pnet. (a) Indoor scenario, (b) outdoor scenario.

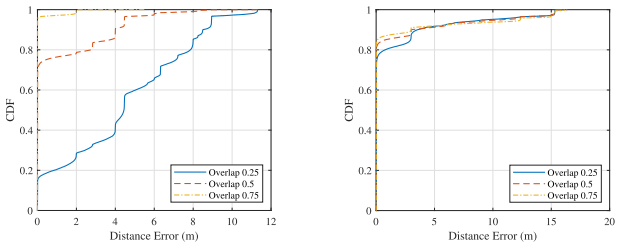


Fig. 8. The influence of different overlap ratios on the positioning performance of FM-Pnet. (a) Indoor scenario, (b) outdoor scenario.

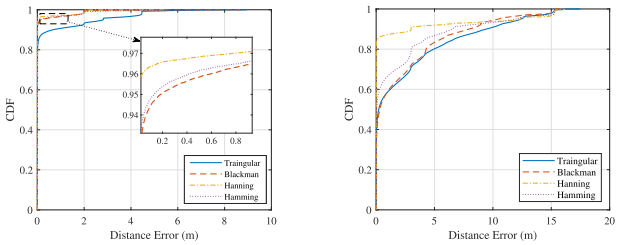


Fig. 9. The influence of different window functions on the positioning performance of FM-Pnet. (a) Indoor scenario, (b) outdoor scenario.

outdoor scenario, as shown in Fig. 7(b), when using 512 FFT points, more than 85% of the samples have distance errors less than 1. This is superior to the results of the other two FFT sizes. These results for indoor and outdoor scenarios show that the performance of FM-Pnet improves with an increase in FFT size. It indicates that increasing the frequency resolution helps with positioning. We will employ an FFT size of 512 for subsequent experiments.

2) Impact of Overlap Ratio on FM-Pnet: Next, we consider the impact of different overlap ratios on positioning. We use the same datasets as the previous experiment and consider three overlap ratios: 0.25, 0.50, and 0.75. The results are shown in Fig. 8. As depicted in Fig. 8(a), in indoor environments, an overlap ratio of 0.75 results in significantly more test samples with distance errors less than 2 compared to overlap ratios of 0.25 and 0.50. The CDF curve converges much faster in this case. For outdoor environments, as shown in Fig. 8(b), an increased overlap ratio also proves beneficial for positioning. These results indicate that the performance of FM-Pnet improves with an increase in overlap ratio, with the improvement being more pronounced in indoor scenarios. In subsequent experiments, we will employ an overlap ratio of 0.75.

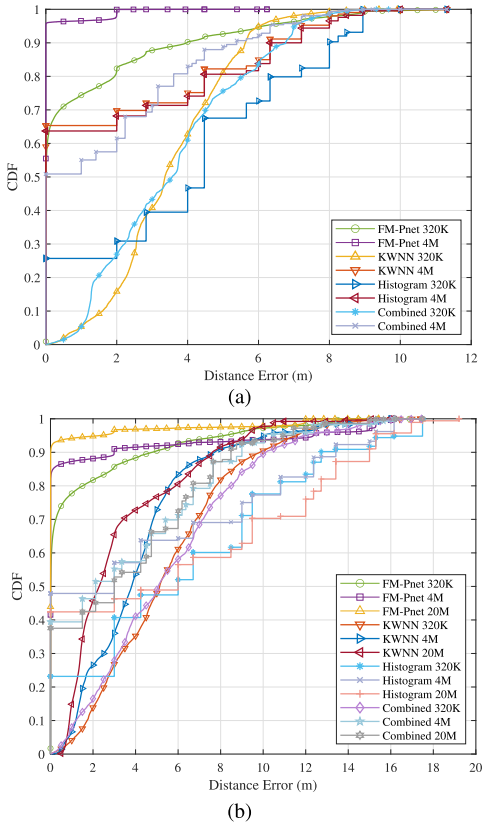


Fig. 10. The influence of different bandwidths on the positioning performance. (a) Indoor scenario, (b) outdoor scenario.

3) Impact of Window Function on FM-Pnet: Finally, we investigate the impact of different window function choices on the positioning performance, including Hanning, Hamming, Blackman, and Triangular. As depicted in Fig. 9(a), using the Hanning window for indoor positioning yields the best performance. With the Hanning window, over 96% of the sample testing errors are within 0.05, showcasing a significant advantage compared to the other three windows. Additionally, the performance of Blackman and Hamming windows for positioning is similar, while the Triangular window exhibits the poorest performance. For outdoor scenario, as shown in Fig. 9(b), using the Hanning window results in over 90% of the samples having distance errors less than 5, surpassing the results obtained with the other three window functions. The results from both indoor and outdoor scenarios indicate that the performance of FM-Pnet is optimal when the Hanning window is chosen. We will use Hanning window in the following experiments.

C. Performance Under Different Time and Frequency Scales

1) Effect of Sample Bandwidth: We use the data collected on Day1 in both indoor and outdoor scenarios. The sample length is fixed at 4,096. Datasets with different bandwidths are used for experiment, each trained separately. It's worth noting that in the case of a 320kHz bandwidth, the total number of samples is less than 4,000. Consequently, we use 2,000 samples for training and 1,000 samples for testing. The experimental results are presented in Table II and Fig. 10.

TABLE II
TEST RESULTS OF DIFFERENT BANDWIDTHS

Metric	Method	Indoor		Outdoor		
		320kHz	4MHz	320kHz	4MHz	20MHz
MDE	KWNN	3.5158	1.8437	5.4423	2.9303	1.7010
	Histogram	3.9809	1.9522	6.3325	4.6804	6.0622
	Combined	3.5971	1.7985	5.5354	3.5987	3.6974
	FM-Pnet	1.0375	0.0689	1.2376	1.1522	0.4003
STD	KWNN	1.5860	2.7882	3.1332	4.3577	4.1142
	Histogram	2.9612	2.8788	5.1928	5.5042	6.0868
	Combined	2.0135	2.2681	3.5267	3.8928	3.7479
	FM-Pnet	2.0654	0.3560	2.7698	3.5516	1.8963

TABLE III
TEST RESULTS OF DIFFERENT SAMPLE LENGTHS

Length	Method	Indoor		Outdoor	
		MDE	STD	MDE	STD
1024	KWNN	2.3403	2.9493	2.9544	4.5737
	Histogram	2.1080	2.9017	4.4428	5.4733
	Combined	2.0781	2.3218	3.5294	4.2336
	FM-Pnet	0.0968	0.4584	1.4118	3.7112
4096	KWNN	1.8370	2.7882	2.9304	4.3577
	Histogram	1.9522	2.8788	4.6804	5.5042
	Combined	1.7985	2.2681	3.5987	3.8928
	FM-Pnet	0.0689	0.3595	1.1522	3.5516
16384	KWNN	0.8254	2.1197	2.5623	4.0603
	Histogram	0.8031	1.9866	4.8419	5.1765
	Combined	0.7917	1.6430	3.4460	3.6519
	FM-Pnet	0.0028	0.1063	0.3669	2.0021
	Segmented	0.0784	0.3868	0.9588	3.1796

As shown in Table II, in both indoor and outdoor scenarios, FM-Pnet achieves lower MDE and STD compared to the baseline methods across all bandwidths. Specifically, for a 4MHz bandwidth, the MDE of FM-Pnet in indoor and outdoor scenarios are 0.0689 and 1.1522, respectively, while the best-performing Combined method among baseline approaches have MDE values of 1.7985 and 3.5987 in indoor and outdoor scenarios. Notably, in the indoor scenario, FM-Pnet has a positioning STD of 0.3560, which is an order of magnitude better than the Combined method, demonstrating FM-Pnet's superior positioning stability. When employing baseline methods for outdoor positioning, there is no reduction in MDE, and no improvement in positioning performance is observed when transitioning from a 4MHz to a 20MHz bandwidth. However, for our proposed FM-Pnet, a comparison of results between the 20MHz and 4MHz bandwidths reveals a significant decrease in MDE from 1.1522 to 0.4003 and a reduction in STD from 3.5516 to 1.8963. This indicates a noteworthy enhancement in both positioning accuracy and stability.

The CDF in the indoor scenario is shown in Fig. 10(a). FM-Pnet outperforms the baseline methods at different bandwidths. In the FM-Pnet method, with a 4MHz bandwidth, nearly 100% of the samples have a distance error of less than 2, and with a 320kHz bandwidth, over 80% of the samples have a distance error of less than 2. In contrast, when using the KWNN method, with a 4MHz bandwidth, only 70% of the samples are less than 2 in distance error. When using a 4MHz bandwidth, FM-Pnet achieves faster convergence of the CDF curve compared to all other traditional methods. With a 320kHz bandwidth, only the Combined method can achieve similar convergence to our method.

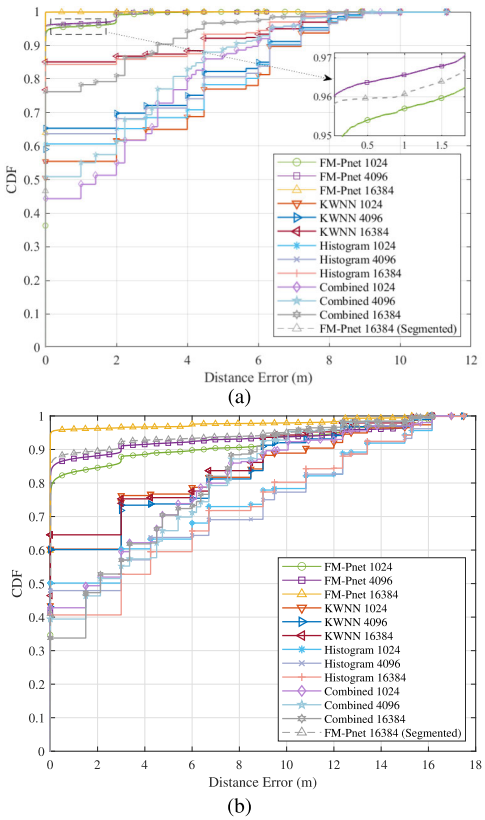
The CDF in the outdoor scenario is shown in Fig. 10(b). Compared to the baseline methods, FM-Pnet demonstrates superior performance using three different bandwidths. The CDF curves have higher initial values and faster convergence speeds. With a 20MHz bandwidth, FM-Pnet achieves a distance error of less than 2 for 95% of the samples, a 10% improvement compared to the results obtained with a 4MHz bandwidth. In contrast, even the best-performing KWNN method among the existing baseline methods still has 15% of samples with distance error less than 2 when using a 20MHz bandwidth.

Overall, FM-Pnet consistently exhibits superior positioning performance compared to the three RSS-based baseline methods across varying bandwidths. Comparing the indoor and outdoor scenarios using the same positioning method, it can be observed that all MDE values are lower in the indoor scenario. This is because in the indoor environment, multipath fading is more pronounced and signals obtained from neighboring positions exhibit higher distinctiveness. Moreover, with the increase in bandwidth–despite maintaining the same signal length–the positioning accuracy of our proposed FM-Pnet continues to improve. This improvement is primarily attributed to the wider bandwidth encompassing more narrowband FM signals, thus providing richer information that FM-Pnet can effectively learn for enhanced positioning accuracy.

2) Effect of Sample Length: Next, we analyze the impact of different sample lengths on positioning performance. We consider three sample lengths: 1,024, 4,096, and 16,384. Data collected with a 4MHz bandwidth on Day1 is used. It's important to note that when considering a sample length of 16,384, the total number of samples does not reach 4,000. Consequently, we opt for 2,000 as the training set and 1,000 as the test set. The experimental results are shown in Table III and Fig. 11.

As shown in Table III, in the indoor scenario, FM-Pnet achieves MDE values of 0.0968, 0.0689, and 0.0028 for sample lengths of 1,024, 4,096, and 16,384, respectively. In contrast, the most effective Combined method among existing baseline approaches has MDE values of 2.0781, 1.7985, and 0.7917 for the corresponding sample lengths. Moreover, in the indoor scenario, FM-Pnet exhibits a positioning STD of less than 0.5 for all sample lengths, while the three baseline methods have STD greater than 1.64. This confirms that FM-Pnet demonstrates stronger positioning stability. In the outdoor scenario, FM-Pnet achieves smaller MDE and STD values for all sample lengths compared to the three baseline methods. Particularly, when testing a sample length of 16,384 in the outdoor scenario, FM-Pnet achieves an MDE of 0.3669 and an STD of 2.0021, whereas the KWNN method only achieves MDE and STD values of 2.5632 and 4.0603, respectively. This indicates a significant performance improvement of FM-Pnet when using a sample length of 16,384.

The CDF curve of the indoor scenario is shown in Fig. 11(a). Compared to the other three methods, each CDF curve of FM-Pnet has a higher initial value and faster



(a)

(b)

Fig. 11. The influence of different sample lengths on the positioning performance. (a) Indoor scenario, (b) outdoor scenario.

convergence speed. Regardless of the sample length tested by FM-Pnet, almost 100% of the samples have testing result below 2. In contrast, even when the other three methods adopt samples with a length of 16,384, only 85% of the samples have an error below 2. The CDF curve of the outdoor scenario is shown in Fig. 11(b). FM-Pnet achieves over 90% of the test samples with distance error lower than the random guessing value of 7.8234, while KWNN method only has 80% and the Histogram method has 70%. The Combined method has to some extent improved the convergence of the CDF, but there are still 40% of the samples with distance error exceeding 3, while this percentage is only 10% for FM-Pnet.

The above results reveal a significant improvement in positioning performance with increased signal length. However, this enhancement comes at the cost of retraining. To overcome this issue, we can adopt the strategy of splitting a long signal into sub-samples of short length and feed these segmented samples into a trained model. The final positioning result is the average of the output positions. To better explore this effect in the experiment, we choose to train the model using samples of length 1,024, and then test it with samples of length 16,384. The experimental results are shown in Fig. 11(a) and 11(b). Whether in indoor scenario or outdoor scenario, when using a model with 1,024-length samples, samples with a length of 16,384 can achieve better testing results. Therefore, we believe that the segmented testing strategy is a feasible approach, especially when considering training complexity.

TABLE IV
TEST RESULTS AT DIFFERENT ANTENNA HEIGHTS

Antenna Height	Method	Primary Model		Mixed Model	
		MDE	STD	MDE	STD
Higher	KWNN	4.0763	3.0712	3.2145	3.0244
	Histogram	4.3466	2.9286	2.1858	2.8216
	Combined	3.9161	2.6125	2.5574	2.3831
	FM-Pnet	3.4665	3.1075	0.0574	0.6140
Lower	KWNN	4.1331	2.9865	2.9827	3.1279
	Histogram	4.2078	2.7717	2.4476	2.9604
	Combined	3.9276	2.4995	2.6080	2.5356
	FM-Pnet	2.1296	2.4938	0.0102	0.1667
Original	KWNN	1.2387	2.4384	2.3589	3.0010
	Histogram	1.1413	2.2912	2.0849	2.8481
	Combined	1.1532	2.0082	2.1127	2.4725
	FM-Pnet	0.0021	0.0859	0.0105	0.1841

D. Performance Under Different Antenna Heights and Grid Resolutions

1) Impact of Antenna Height: To investigate the impact of different antenna heights on positioning accuracy, we conduct data collection in the indoor scenario by setting three distinct antenna heights: 1.5 m (lower), 1.8 m (original), and 2.1 m (higher). Throughout the data collection process, we maintain a fixed signal bandwidth of 4MHz and a fixed length of 4,096, respectively. We train two models for comparison: the primary model, trained solely on data from the original antenna height, and the mixed model, trained using data from all three antenna heights. We then individually assess the positioning performance of these models at the three antenna heights. The results are presented in in Table IV and Fig. 12.

Firstly, when using the primary model, FM-Pnet demonstrates outstanding performance at the original antenna height, while the positioning accuracy significantly decreases for other heights. As shown in Table IV, the MDE value for testing the original antenna height data using the primary model is 0.0021, whereas it is 3.4665 and 2.1296 for testing higher and lower antenna height data, respectively. Simultaneously, the baseline methods trained on the original height dataset also perform poorly at the other two antenna heights. For instance, the combined method, trained using the original antenna height data, yields MDE values of 3.9161, 3.9276, and 1.1532 for the three heights, confirming the impact of antenna height on positioning. The CDF plots in Fig. 12(a) further illustrate the impact of antenna height on positioning performance.

Secondly, when using the mixed model trained on data from the three antenna heights, as shown in Table IV, FM-Pnet achieves impressive MDE values of 0.0574, 0.0102, and 0.0105 for higher, lower, and original antenna heights, respectively. Under the same experimental conditions, the best-performing baseline method yields MDE values of 2.1858, 2.4476, and 2.1127 for higher, lower, and original antenna heights. This emphasizes FM-Pnet's robust adaptability to all three antenna heights when using a mixed height dataset. In Fig. 12(b), after training with mixed height data, FM-Pnet consistently achieves MDE values below 0.02 for over 97% of samples at each testing height. In contrast, all baseline methods struggle to have fewer than 60% of samples with MDE below 0.02. This highlights FM-Pnet's stronger adaptability to different antenna heights after training with mixed data.

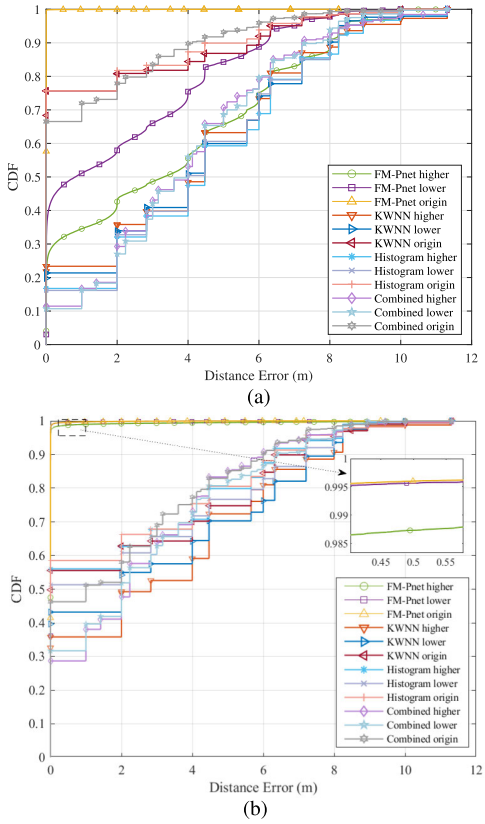


Fig. 12. The impact of antenna height on positioning performance. (a) Primary model, (b) mixed model.

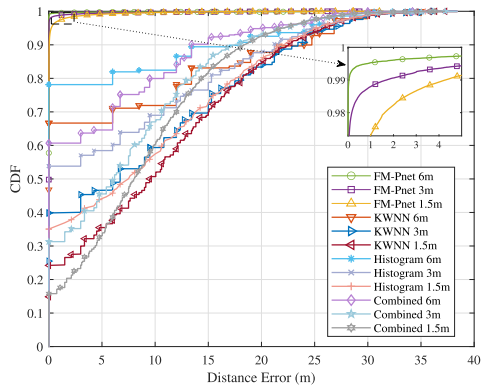


Fig. 13. Performance under different grid resolutions in a larger-scale scenario.

2) Impact of Grid Resolution: To analyze the performance under different grid resolutions, we now conduct experiments in the large-scale scenario. As previously mentioned, a grid is set every 1.5 m during data collection. The signal bandwidth is 4MHz, with a length of 4,096. Based on this setup, we additionally construct two other grid resolutions by spanning across the intermediate grids: 3 m and 6 m. At each grid position, there are a total of 8,000 samples for training and 2,000 samples for testing. The experimental results are shown in Table V and Fig. 13.

As shown in Table V, our proposed FM-Pnet outperforms the baseline methods by a significant margin under all three grid resolutions. Specifically, under the most challenging 1.5 m resolution, our method achieves an MDE of 0.1486, whereas the Combined method has an MDE of 8.8153, nearly an order of magnitude higher. Comparing the results at different resolutions, it is evident that as the resolution increases (from 6 m to 1.5 m), the positioning performance of the methods declines. Specifically, for FM-Pnet, the MDE values at resolutions of 6 m, 3 m, and 1.5 m are 0.0511, 0.0881, and 0.1486, respectively, while for the Histogram method, the corresponding MDE values are 3.6638, 6.8504, and 9.0127. The errors of both methods increase with resolution. However, FM-Pnet exhibits very minimal increases in localization error, indicating its strong adaptability to various resolutions. Furthermore, the CDFs depicted in Fig. 13 also corroborate the experimental findings mentioned above.

TABLE V
TEST RESULTS WITH DIFFERENT GRID RESOLUTIONS

Grid Resolution	Method	MDE	STD
1.5 m	KWNN	10.2117	8.6741
	Histogram	9.0127	8.9442
	Combined	8.8153	6.9677
	FM-Pnet	0.1486	1.0829
3 m	KWNN	8.8773	9.4044
	Histogram	6.8054	8.9910
	Combined	7.3788	7.3625
	FM-Pnet	0.0881	1.0143
6 m	KWNN	5.8013	9.3069
	Histogram	3.6638	7.8174
	Combined	4.5769	7.0622
	FM-Pnet	0.0511	0.9710

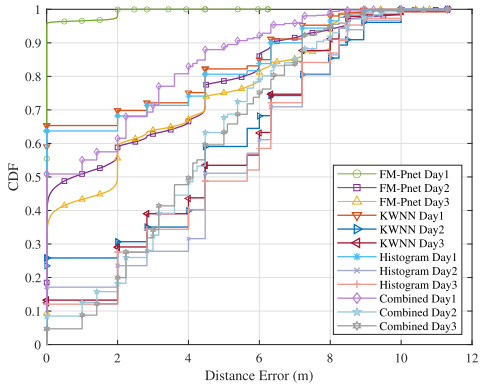
TABLE VI
THE RESULTS OF CROSS-DATE TESTING

Test Date	Method	Indoor		Outdoor	
		MDE	STD	MDE	STD
Day1	KWNN	1.8370	2.7882	2.9304	4.3577
	Histogram	1.9522	2.8788	4.6804	5.5042
	Combined	1.7985	2.2681	3.5987	3.8928
	FM-Pnet	0.0689	0.3595	1.1522	3.5516
Day2	KWNN	4.5137	2.7612	6.3925	4.8716
	Histogram	4.8040	2.8288	7.9498	5.3306
	Combined	4.2250	2.2895	6.6308	4.2188
	FM-Pnet	2.4397	2.9016	5.6300	4.9822
Day3	KWNN	4.3361	3.1726	5.3261	4.7933
	Histogram	4.8421	2.9125	6.6603	5.1077
	Combined	4.1738	2.4121	5.4003	4.0354
	FM-Pnet	2.6973	2.9491	4.0460	4.7674

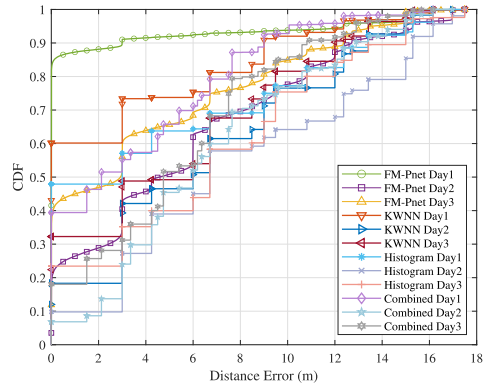
E. Generalization Performance in Cross-Date Testing

1) Cross-Date Testing: In the aforementioned experiments, both the training set and the test set are derived from a single day, namely Day1. To evaluate the generalization performance of FM-Pnet, we extend our testing experiments to incorporate FM signals received over two additional days. These experiments encompass FM signal data from both indoor and outdoor scenarios, with a sample bandwidth of 4MHz and a data length of 4,096.

As shown in Table VI, all positioning methods experience a decline in performance during cross-date testing. But the proposed FM-Pnet demonstrates superior positioning accuracy compared to the other three traditional methods. Specifically, in the indoor scenario, the MDE obtained by testing



(a)



(b)

Fig. 14. Generalization performance in cross-date testing. (a) Indoor scenario, (b) outdoor scenario.

TABLE VII
THE RESULTS OF NOISE INJECTION IN INDOOR SCENARIO

Test Date	Method	MDE	STD
Day1	AWGN	0.0259	0.2167
	AGGN	0.0298	0.2460
	PN	0.0627	0.3461
	Mixed Noise	0.0083	0.1349
	Without Noise	0.0689	0.3595
Day2	AWGN	1.5886	2.4840
	AGGN	1.3557	2.2166
	PN	1.9884	2.6945
	Mixed Noise	1.4134	2.4714
	Without Noise	2.4397	2.9016
Day3	AWGN	1.7279	2.4748
	AGGN	1.6529	2.2885
	PN	2.4797	2.8307
	Mixed Noise	1.6348	2.3815
	Without Noise	2.6973	2.9491

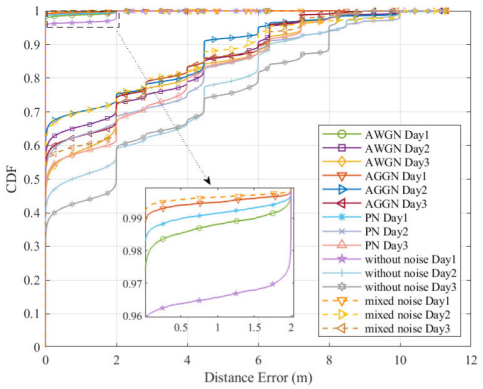


Fig. 15. Performance of noise injection enhancement in indoor scenario.

the Day2 and Day3 samples using trained FM-Pnet from Day1 is 2.6973 and 2.4397, respectively. The best-performing Combined method among baseline methods has MDE values of 4.2250 and 4.1738, respectively, which are close to the random guessing results. In the outdoor scenario, among the three baseline positioning methods, KWNN exhibits the best performance, with MDE values reaching 6.3925 and 5.3261 on the test sets of Day2 and Day3, respectively. In the same situation, our proposed FM-Pnet can achieve MDE of 5.6300 and 4.0460, respectively. Thus, the FM-Pnet also demonstrates the better generalization performance in the outdoor scenario. The CDF results for indoor and outdoor scenarios are shown in Fig. 14(a) and Fig. 14(b). It also demonstrates that, at various distance errors, the proposed FM-Pnet exhibits better positioning accuracy and generalization performance compared to the three baseline methods.

2) Noise Injection Enhancement: We investigate the impact of noise injection enhancement on cross-date positioning performance by introducing three types of noise: AWGN, Additive Generalized Gaussian Noise (AGGN), and Pink Noise (PN). Each type of noise is applied individually to the received FM signals during the training process. Additionally, we experiment with mixing noise by adding multiple types of noise to the signals, referred to as mixed noise. To introduce random noise at different power levels, we randomly select the Signal-to-Noise Ratio (SNR) from a uniform distribution ranging from 0 dB to 50 dB, and then inject the noise into the signals accordingly. All other training parameters remain consistent with the non-noise enhancement scenario. The FM-Pnet, trained on Day1 data, is then utilized to evaluate positioning performance on Day1, Day2, and Day3 without introducing additional noise to the test signal samples. The results are illustrated in Table VII and Fig. 15.

As shown in Table VII, the noise-enhanced models significantly improve the positioning performance over three days compared to the non-enhanced model. Both the MDE and STD of the positioning are reduced. As shown in Fig. 15, the model with noise enhancement outperforms the non-enhancement model in positioning performance at different distance errors. The experiment results demonstrate that noise injection in the training process can further enhance the positioning performance. This is an effective solution to improve the generalization of FM-Pnet cross dates in indoor scenario.

When comparing different noise injection methods for data enhancement, AGGN shows superior performance in cross-date positioning tests when only one type of noise is added. It demonstrates lower MDE and STD values than AWGN and PN in samples from the second and third days. For example, on the second day, the MDE is 1.3557, and the STD is 2.2166, while AWGN injection yields an MDE of 1.5886 and STD of 2.4840, and PN injection results in an MDE of 1.9884 and STD of 2.6945. Furthermore, in cross-date experiments, mixed noise injection exhibits superior stability.

TABLE VIII
RESULTS OF ENRICHING TRAINING SAMPLES

Training Set	Method	Indoor		Outdoor	
		MDE	STD	MDE	STD
1-Day	KWNN	1.7259	2.7259	5.3261	4.7933
	Histogram	1.6104	2.6579	6.6603	5.1077
	Combined	1.5760	2.1528	5.4003	4.0354
	FM-Pnet	0.3880	1.2427	4.0460	4.7674
2-Day	KWNN	2.7362	3.1022	5.8132	4.6615
	Histogram	2.4047	2.9193	6.4044	4.7242
	Combined	2.4532	2.4344	5.5827	3.5744
	FM-Pnet	0.0531	0.3119	3.5093	4.2751
3-Day	KWNN	3.3221	3.2629	5.4122	4.2531
	Histogram	2.6625	2.8548	6.0170	4.5432
	Combined	2.8492	2.4397	5.1771	3.4199
	FM-Pnet	0.0240	0.2088	2.4504	3.7232

The MDE and STD for the second-day are 1.4134 and 2.4714, respectively, and for the third-day, they are 1.6348 and 2.3815, demonstrating strong stability and adaptability across different days.

3) Enriching Training Samples: We also evaluate the effectiveness of the enriching training sample strategy in improving the generalization performance of FM-Pnet. To this end, we create multiple training sets to train the localization model independently. The "1-Day" set consists solely of FM signal data collected on the first day. The "2-Day" set includes data from both the first and second days, while the "3-Day" set comprises data collected over a three-day period. Each training set comprises 8,000 samples per location per day. For testing, we employ additional data collected on a separate day not included in the training sets. Specifically, 2,000 samples per location are reserved for testing purposes.

The MDE and STD for indoor and outdoor scenario experimental results are presented in Table VIII. In indoor scenario, utilizing data from a single day, FM-Pnet demonstrates an MDE and STD of 0.3880 and 1.2427, respectively, in cross-date testing. However, with the dataset enriched with samples from three days, FM-Pnet's MDE and STD decrease significantly to 0.02240 and 0.2088, showcasing a remarkable improvement in FM-Pnet's localization performance due to dataset enrichment. Conversely, the best-performing baseline method, Combined, exhibits MDE values of 1.5760, 2.4532, and 2.8492 when employing three datasets, suggesting that the baseline methods do not benefit from the enriched dataset. Similarly, in outdoor scenario, when training with data from a single day, FM-Pnet's MDE and STD in cross-date testing are 4.0460 and 4.7674, respectively. When training with three days of FM data, MDE and STD decrease to 2.4504 and 3.7232, reflecting an enhancement in FM-Pnet's positioning performance. In contrast, using the enrichment strategy with 1, 2, and 3 days of FM signal data, the best-performing baseline method, Combined, has MDE values of 5.4003, 5.5827, and 5.1771, with little variation. These experimental results suggest that the proposed FM-Pnet achieves substantial enhancement in positioning generalization performance by utilizing more enriched data. CDF results depicted in Fig. 16 also illustrate that enriched training samples contribute to improved FM-Pnet positioning accuracy in cross-date experiments.

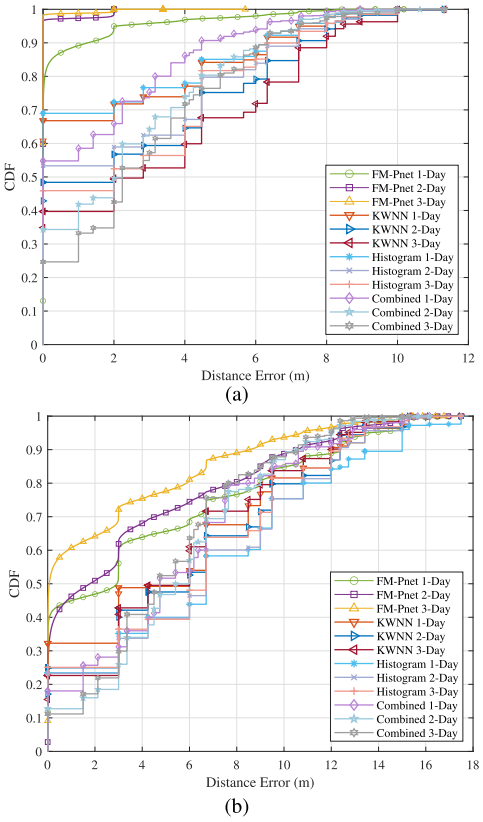
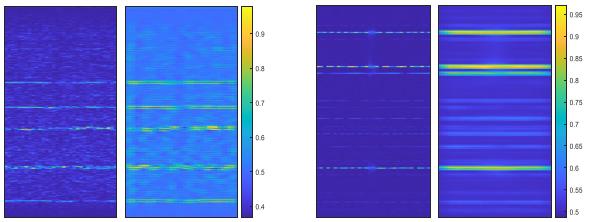


Fig. 16. Performance of enriching training samples. (a) Indoor scenario, (b) outdoor scenario.

F. Attention Mechanism Analysis

In order to demonstrate and analyze the role of the attention mechanism in FM-Pnet positioning, we conduct a visualization analysis of the input data and the corresponding attention weight matrix obtained through neural network layers. Firstly, for FM data with bandwidth of 4MHz, and length of 4,096 in the indoor environment, we randomly selected a signal sample from the test set and presented its related visualization in Fig. 17(a). The left matrix represents the real part of STFT transformation on the FM signal, and the right matrix represents the corresponding attention weight matrix. In the attention weight matrix, the brighter the pixel color, the closer the corresponding element value is to 1, indicating that the element value at that location of time-frequency representation is more important for positioning. By observing Fig. 17(a), we can see that the brighter pixels in the attention weight matrix correspond to five narrowband signals, while the attention weight values for the remaining noise parts are relatively small. This indicates that under the influence of the attention mechanism, FM-Pnet focuses more on the areas with signals, to a certain extent neglecting the areas dominated by environmental noise.

Similarly, for outdoor FM signals with a bandwidth of 20MHz and a length of 4,096, we also randomly select a signal sample and plot the corresponding input matrix (real part of STFT) and its associated attention weight matrix, as shown in Fig. 17(b). In the left figure, four distinct narrowband signals with higher SNR are evident in the input time-frequency



(a) (b)

Fig. 17. Attention mechanism. (a) Indoor scenario, (b) outdoor scenario.

representation, which is mirrored by pronounced pixel bars in the attention matrix shown in the right figure. Notably, the attention weight values for the less prominent signal areas are significantly smaller, and those for the background noise parts appear much darker. Among the four distinct signal segments, the third one from the top is relatively less pronounced. In the attention weight matrix on the right, it is evident that the attention weight values corresponding to this signal are slightly smaller.

These visualizations of both indoor and outdoor data validate the basic mechanism of attention, which is that through self-learning in neural networks, signal-rich regions can play a more significant role in subsequent positioning tasks. Generally, regions with more distinct signals have a larger effect on the positioning results.

G. Complexity Analysis

In practical applications, complexity plays a crucial role in deploying deep learning-based positioning systems. Deep learning complexity is usually assessed by quantifying the number of Floating-Point Operations (FLOPs) and the count of trainable Parameters (Params) in the network. The computation of FLOPs for a convolutional layer is calculated as follows:

$$\mathrm{FLOPs}_{\mathrm{CNN}} = 2 \times H_{in} \times W_{in} \times K_{ker} \times C_{in} \times C_{out}, \quad (20)$$

where C_{in} and C_{out} represent the number of input and output channels, H_{in} and W_{in} respectively represent the height and width of the input image, K_{ker} is the size of the convolutional kernel. For a single fully connected layer, the required FLOPs are given by the following formula:

$$\mathrm{FLOPs}_{\mathrm{FC}} = 2 \times F_{in} \times F_{out}, \quad (21)$$

where F_{in} and F_{out} represent the input and output feature dimensions, respectively. The formula for calculating the number of parameters required for a convolutional layer is:

$$\mathrm{Params}_{\mathrm{CNN}} = (K_{ker} \times C_{in} \times C_{out}) + C_{out}. \quad (22)$$

For a single fully connected layer, the required parameters can be calculated using the following formula:

$$\mathrm{Params}_{\mathrm{FC}} = (F_{in} + 1) \times F_{out}. \quad (23)$$

When considering a 25-class classification scenario, the FM-Pnet has a complexity load of 2,848M FLOPs and 14M Params. It's important to note that the computational complexity of STFT is significantly lower than that of network inference, rendering it negligible by comparison.

VI. Conclusion

In this paper, we have proposed an FM-based positioning method called FM-Pnet, which is implemented using deep learning techniques. This method enables positioning by leveraging the time-frequency representations of FM signals. In order to validate the performance of FM-Pnet, we have conducted extensive experiments in both indoor and outdoor scenarios. The experimental results have demonstrated that our proposed FM-Pnet outperforms three existing RSS-based positioning methods in terms of both positioning accuracy and stability. The performance of FM-Pnet can be further enhanced in the frequency domain or in the time domain by increasing the bandwidth of the collected signals or extending the length of each sample. Additionally, through injecting noise in the training process or enriching training samples with cross-date signals, the generalization of FM-Pnet over a large time span can be improved. Therefore, the proposed FM-Pnet is a competitive positioning method for both indoor and outdoor scenarios, holding promising prospects for future applications. Future research could focus on exploring additional features inherent in FM signals that contribute to positioning accuracy, while also investigating lightweight deep learning neural networks to minimize complexity and enhance scalability for real-world deployment.

References

[1] H. Huang, G. Gartner, J. M. Krisp, M. Raubal, and N. Van De Weghe, "Location based services: Ongoing evolution and research agenda," *J. Location Based Services*, vol. 12, no. 2, pp. 63–93, Apr. 2018.

[2] J. Carmigniani and B. Furht, "Augmented reality: An overview," in *Handbook of Augmented Reality*. New York, NY, USA: Springer, 2011, pp. 3–46.

[3] C. Anthes, R. J. García-Hernández, M. Wiedemann, and D. Kranzlmüller, "State of the art of virtual reality technology," in *Proc. IEEE Aerosp. Conf.*, Mar. 2016, pp. 1–19.

[4] K. Rose, S. Eldridge, and L. Chapin, "The Internet of Things: An overview," in *Proc. Internet Soc. (ISOC)*, vol. 80, 2015, pp. 1–50.

[5] T. H. Dixon, "An introduction to the global positioning system and some geological applications," *Rev. Geophys.*, vol. 29, no. 2, pp. 249–276, May 1991.

[6] R. Li et al., "Advances in BeiDou navigation satellite system (BDS) and satellite navigation augmentation technologies," *Satell. Navigat.*, vol. 1, no. 1, pp. 1–23, Dec. 2020.

[7] Z. M. Kassas, "Collaborative opportunistic navigation [student research highlight]," *IEEE Aerosp. Electron. Syst. Mag.*, vol. 28, no. 6, pp. 38–41, Jun. 2013.

[8] D. Jeong, B. Lee, D. G. Kim, and S. Sung, "Performance verification of AC magnetic positioning system for unmanned vehicle navigation and control," in *Proc. IEEE/ION Position, Location Navigat. Symp. (PLANS)*, Apr. 2023, pp. 189–195.

[9] V. Moghtadaiee and A. G. Dempster, "Indoor location fingerprinting using FM radio signals," *IEEE Trans. Broadcast.*, vol. 60, no. 2, pp. 336–346, Jun. 2014.

[10] X. Chen, Q. Xiang, L. Kong, H. Xu, and X. Liu, "Learning from FM communications: Toward accurate, efficient, all-terrain vehicle localization," *IEEE/ACM Trans. Netw.*, vol. 31, no. 1, pp. 42–57, Feb. 2023.

[11] M. Rabinowitz and J. J. Spilker, "A new positioning system using television synchronization signals," *IEEE Trans. Broadcast.*, vol. 51, no. 1, pp. 51–61, Mar. 2005.

[12] X. Liu et al., "A neuro-inspired positioning system integrating MEMS sensors and DTMB signals," *IEEE Trans. Broadcast.*, vol. 69, no. 3, pp. 823–831, Sep. 2023.

[13] K. Maaloul, B. Lejdel, E. Clementini, and N. M. Abdelhamid, "Bluetooth beacons based indoor positioning in a shopping malls using machine learning," *Bull. Electr. Eng. Informat.*, vol. 12, no. 2, pp. 911–921, Apr. 2023.

[14] Z. Zhao, W. Huang, Y. Liang, and Y. Zhang, "A NLOS detection method based on machine learning in UWB indoor location system," in *Proc. Int. Conf. Virtual Reality Intell. Syst. (ICVRIS)*, Jul. 2020, pp. 214–217.

[15] S. Sung, H. Kim, and J.-I. Jung, "Accurate indoor positioning for UWB-based personal devices using deep learning," *IEEE Access*, vol. 11, pp. 20095–20113, 2023.

[16] X. Tong, Y. Wan, Q. Li, X. Tian, and X. Wang, "CSI fingerprinting localization with low human efforts," *IEEE/ACM Trans. Netw.*, vol. 29, no. 1, pp. 372–385, Feb. 2021.

[17] I. Neupane, B. Alsinglawi, and K. Rabie, "Indoor positioning using Wi-Fi and machine learning for Industry 5.0," in *Proc. IEEE Int. Conf. Pervasive Comput. Commun. Workshops Affiliated Events (PerCom Workshops)*, Mar. 2023, pp. 359–362.

[18] J. Li, I.-T. Lu, and J. S. Lu, "Hybrid fingerprinting and ray extension localization in NLOS regions," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 12, pp. 23503–23516, Dec. 2022.

[19] Z. Liu, L. Chen, X. Zhou, Z. Jiao, G. Guo, and R. Chen., "Machine learning for time-of-arrival estimation with 5G signals in indoor positioning," *IEEE Internet Things J.*, vol. 10, no. 11, pp. 9782–9795, Jun. 2023.

[20] Z. Zhu, C. Guo, R. Bao, M. Chen, W. Saad, and Y. Yang, "Positioning using visible light communications: A perspective arcs approach," *IEEE Trans. Wireless Commun.*, vol. 22, no. 10, pp. 6962–6977, Oct. 2023.

[21] X. Wang, X. Wang, and S. Mao, "Indoor fingerprinting with bimodal CSI tensors: A deep residual sharing learning approach," *IEEE Internet Things J.*, vol. 8, no. 6, pp. 4498–4513, Mar. 2021.

[22] B. Wang et al., "A novel weighted KNN algorithm based on RSS similarity and position distance for Wi-Fi fingerprint positioning," *IEEE Access*, vol. 8, pp. 30591–30602, 2020.

[23] X. Wang, L. Gao, S. Mao, and S. Pandey, "CSI-based fingerprinting for indoor localization: A deep learning approach," *IEEE Trans. Veh. Technol.*, vol. 66, no. 1, pp. 763–776, Jan. 2017.

[24] X. Wang, X. Wang, and S. Mao, "Deep convolutional neural networks for indoor localization with CSI images," *IEEE Trans. Netw. Sci. Eng.*, vol. 7, no. 1, pp. 316–327, Jan. 2020.

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

[26] S.-H. Fang, J.-C. Chen, H.-R. Huang, and T.-N. Lin, "Is FM a RF-based positioning solution in a metropolitan-scale environment? A probabilistic approach with radio measurements analysis," *IEEE Trans. Broadcast.*, vol. 55, no. 3, pp. 577–588, Sep. 2009.

[27] A. Popleteev, V. Osmani, and O. Mayora, "Investigation of indoor localization with ambient FM radio stations," in *Proc. IEEE Int. Conf. Pervasive Comput. Commun.*, Mar. 2012, pp. 171–179.

[28] Y. Chen, D. Lymberopoulos, J. Liu, and B. Priyantha, "Indoor localization using FM signals," *IEEE Trans. Mobile Comput.*, vol. 12, no. 8, pp. 1502–1517, Aug. 2013.

[29] Y. Lin, Y. Tu, Z. Dou, L. Chen, and S. Mao, "Contour stella image and deep learning for signal recognition in the physical layer," *IEEE Trans. Cognit. Commun. Netw.*, vol. 7, no. 1, pp. 34–46, Mar. 2021.

[30] S. Zheng, S. Chen, and X. Yang, "DeepReceiver: A deep learning-based intelligent receiver for wireless communications in the physical layer," *IEEE Trans. Cogni. Commun. Netw.*, vol. 7, no. 1, pp. 5–20, Mar. 2021.

[31] S. Zheng et al., "Toward next-generation signal intelligence: A hybrid knowledge and data-driven deep learning framework for radio signal classification," *IEEE Trans. Cogn. Commun. Netw.*, vol. 9, no. 3, pp. 564–579, Jun. 2023.

[32] A. Youssef, J. Krumm, E. Miller, G. Cermak, and E. Horvitz, "Computing location from ambient FM radio signals [commercial radio station signals]," in *Proc. IEEE Wireless Commun. Netw. Conf.*, vol. 2, Mar. 2005, pp. 824–829.

[33] T. Mukherjee, P. Kumar, D. Pati, E. Blasch, E. Pasiliao, and L. Xu, "Losi: Large scale location inference through FM signal integration and estimation," *Big Data Min. Anal.*, vol. 2, no. 4, pp. 319–348, Dec. 2019.

[34] P. Kumar, T. Mukherjee, E. Pasiliao, and L. Xu, "Cheap approximate localization using FM radio," in *Proc. 23rd SIGSPATIAL Int. Conf. Adv. Geographic Inf. Syst.*, Nov. 2015, pp. 1–4.

[35] S. Yoon, K. Lee, Y. Yun, and I. Rhee, "ACMI: FM-based indoor localization via autonomous fingerprinting," *IEEE Trans. Mobile Comput.*, vol. 15, no. 6, pp. 1318–1332, Jun. 2016.

[36] A. Popleteev, "Indoor localization using ambient FM radio RSS fingerprinting: A 9-month study," in *Proc. IEEE Int. Conf. Comput. Inf. Technol. (CIT)*, Aug. 2017, pp. 128–134.

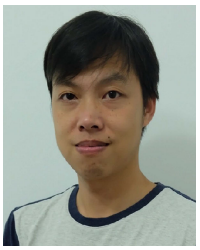
[37] A. Popleteev, "Improving ambient FM indoor localization using multipath-induced amplitude modulation effect: A year-long experiment," *Pervas. Mobile Comput.*, vol. 58, Aug. 2019, Art. no. 101022.

[38] L. Alzubaidi et al., "Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions," *J. Big Data*, vol. 8, no. 1, pp. 1–74, Mar. 2021.

[39] Z. Wu, C. Shen, and A. van den Hengel, "Wider or deeper: Revisiting the ResNet model for visual recognition," *Pattern Recognit.*, vol. 90, pp. 119–133, Jun. 2019.

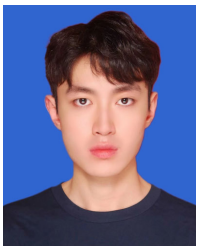
[40] S. Xie, R. Girshick, P. Dollár, Z. Tu, and K. He, "Aggregated residual transformations for deep neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 5987–5995.

[41] G. Pant, D. P. Yadav, and A. Gaur, "ResNeXt convolution neural network topology-based deep learning model for identification and classification of pediastrum," *Algal Res.*, vol. 48, Jun. 2020, Art. no. 101932.



Shilian Zheng received the B.S. degree in telecommunication engineering and the M.S. degree in signal and information processing from Hangzhou Dianzi University, Hangzhou, China, in 2005 and 2008, respectively, and the Ph.D. degree in communication and information systems from Xidian University, Xi'an, China, in 2014.

He is currently a Researcher with the National Key Laboratory of Electromagnetic Space Security, Jiaxing, China. His research interests include cognitive radio, deep learning-based radio signal processing, and electromagnetic space security.



Jiacheng Hu received the B.S. degree from Sichuan Normal University, Chengdu, China, in 2022. He is currently pursuing the M.S. degree with the School of Communications Engineering, Hangzhou Dianzi University, Hangzhou. His current research interests include deep learning and navigation via signals of opportunity.

Luxin Zhang received the M.S. degree in control science and engineering from Zhejiang University of Technology, Hangzhou, China, in 2021.

He is currently an Assistant Engineer with the National Key Laboratory of Electromagnetic Space Security, Jiaxing, China. His research interests include cognitive radio, radio signal processing, and learning-based radio signal recognition.

Kunfeng Qiu was born in Shaoxing, Zhejiang, China, in 1997. He received the M.S. degree in control engineering from Zhejiang University of Technology, Hangzhou, China, in 2022.

He is currently an Assistant Engineer with the National Key Laboratory of Electromagnetic Space Security, Jiaxing, China. His research interests include radio signal processing and radio signal recognition based on deep learning.

Jie Chen received the B.S. degree from Taizhou University, Zhejiang, China, in 2021. He is currently pursuing the M.S. degree with the School of Communications Engineering, Hangzhou Dianzi University, Hangzhou. His current research interests include deep learning and radio signal processing.

Peihan Qi (Member, IEEE) was born in Henan, China, in 1986. He received the B.S. degree in telecommunications engineering from Chang'an University, Xi'an, China, in 2006, and the M.S. and Ph.D. degrees in communication and information systems from Xidian University, Xi'an, in 2011 and 2014, respectively.

He has been a Professor with the School of Telecommunications Engineering, Xidian University, since July 2022. His research interests include compressed sensing, spectrum sensing in cognitive radio networks, and intelligent signal processing.

Zhijin Zhao received the M.S. and Ph.D. degrees from Xidian University, Xi'an, China, in 1984 and 2009, respectively. She was a Visiting Scholar of adaptive signal processing and blind signal processing with Darmstadt University of Technology and the University of Erlangen–Nürnberg in 1993 and 2003, respectively. She is currently a Professor with the School of Communication Engineering, Hangzhou Dianzi University, Hangzhou, China. Her research interests include communication signal processing, cognitive radio technology, intelligent signal processing, and other aspects of research. She was the President of the School of Communication Engineering, Hangzhou Dianzi University; a Senior Member of China Electronics Society; and a member of the National Signal Processing Society.

Xiaoniu Yang is currently a Chief Scientist with the National Key Laboratory of Electromagnetic Space Security, Jiaxing, China. He is also an Academician of Chinese Academy of Engineering and a fellow of Chinese Institute of Electronics. He has published the first software radio book in China [X. Yang, C. Lou, and J. Xu, (*Software Radio Principles and Applications*, Publishing House of Electronics Industry, 2001, in Chinese)]. He holds more than 40 patents. His current research interests include software-defined satellites, big data for radio signals, and deep learning-based signal processing.