FM-Based Positioning via Deep Learning

Shilian Zheng $^1$, Jiachen Hu $^2$, Luxin Zhang $^2$, Kunfeng Qiu $^2$, Jie Chen $^2$, Peihan Qi $^2$, Zhijin Zhao $^2$, and Xiaoniu Yang $^2$

$^1$National Key Laboratory of Electromagnetic Space Security
$^2$Affiliation not available

December 7, 2023

Abstract

Frequency modulation (FM) broadcast signals, as opportunity 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 an end-to-end FM-based positioning method that leverages deep learning, known as FM-Pnet. This method utilizes the time-frequency representation of FM signals as the network input, allowing the network to automatically learn deep features for positioning. We further 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.
FM-Based Positioning via Deep Learning

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

Abstract—Frequency modulation (FM) broadcast signals, as opportunity 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 an end-to-end FM-based positioning method that leverages deep learning, known as FM-Pnet. This method utilizes the time-frequency representation of FM signals as the network input, allowing the network to automatically learn deep features for positioning. We further propose two strategies, noise injection and enriching training samples, to enhance the model’s generalization performance over long time spans. We 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 is a key technology widely used in modern society, and positioning navigation has become an indispensable part of human daily 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], as mainstream positioning technologies today, are widely applied in the field of positioning navigation. In outdoor areas with good signal reception, their positioning accuracy is high. For example, the positioning error of commercial GPS is usually no more than 10 meters (2D). However, GPS and BNSS both rely on space-based signals as navigation signals, which have relatively high construction costs. Moreover, they also require high timing accuracy for receiving devices. In addition, in densely built or obstructed areas (especially indoor areas), devices have low signal reception power for satellite signals, leading to a significant decrease in positioning accuracy. Therefore, it is necessary to develop technologies that do not rely on existing satellite positioning, in order to save positioning costs, expand the range of positioning applicability, and improve positioning robustness.

Navigation via Signals of Opportunity (NAVSOP) has been widely researched as a positioning method that does not rely on satellites [7]. Signals of opportunity refer to existing non-navigation radio frequency (RF) signals in the electromagnetic environment. Magnetic signals [8], analog modulated broadcast signals [9], Digital Terrestrial Multimedia Broadcast (DTMB) signals [10], Bluetooth signals [11], UWB signals [12], WiFi signals [13], mobile communication signals [14], visible light signals [15] can all be used for NAVSOP. Among them, WiFi signals have gained widespread attention in indoor positioning applications due to advantages such as the easy knowledge of the router’s location and wide indoor distribution [16]. In cases where the emission location of the WiFi signal is known, the position of the observation point is usually determined by measuring the energy of the WiFi signal using the Received Signal Strength (RSS) method [17]. Due to the complex propagation environment, the RSS values at a given location for continuously received WiFi packets are generally unstable, and RSS values can only provide rough information. Therefore, researchers have proposed to use deep learning methods to extract the CSI features of WiFi signals for position recognition, significantly improving the accuracy of WiFi indoor positioning [18–20]. However, WiFi signals have a small coverage area and weak ability to penetrate obstacles such as walls, making them practically effective only indoors. Compared to WiFi signals, frequency modulation (FM) broadcast signals (87MHz-108MHz) have higher transmission power, longer wavelength (around 3 m), stronger penetration, wider coverage range, and can be received well in various environments, whether outdoor with open spaces, indoor environments with obstructing building walls, or fully enclosed underground garages. Therefore, FM-based positioning is expected to become a universal positioning method applicable both indoors and outdoors [21].

Currently, FM-based positioning has been extensively studied in both indoor and outdoor 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. Probabilistic methods consider that each position has a corresponding RSS vector, which is a random variable. 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. [22], the observed RSS values at each position are modeled as a Gaussian distribution, and the location corresponding to each RSS value is estimated using 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 [23–25]. In outdoor scenarios, Youssef et al. [23] 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, the complexity of the channel resulted in a significant discrepancy between the energy fingerprint map and reality, leading to an error distance of 8 miles in the experiment. Kumar et al. [25] 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. In a large outdoor area, Mukherjee et al. [24] used the FM spectrum estimation model and probabilistic methods (Bayesian decision) to achieve a positioning error of 0.12 miles. In indoor scenarios, Yoon et al. [26] extended the modeling method to indoor scenes and established an indoor propagation model for FM signal RSS fingerprints, obtaining average errors of approximately 6 meters and 10 meters in two indoor scenarios at a campus and a downtown area.

Unlike probabilistic methods, deterministic methods estimate the location by considering the actual measured values. According to [27], in an indoor scenario, after normalizing all the collected RSS values, three methods, namely K-Nearest Neighbors (KNN), Support Vector Machine (SVM), and Gaussian Process (GP) Regression, were applied for positioning tests. Chen et al. [28] conducted experiments in three different indoor scenarios to compare the positioning using FM signal and WiFi signal RSS fingerprints, achieving similar accuracy. The combination of WiFi and FM RSS fingerprints was also attempted, resulting in improved positioning performance. In an indoor environment, Moghtadai et al. [9] used probabilistic methods (Histogram) and deterministic methods (KNN) separately, and then combined them (Combined) to achieve an average error distance of approximately 2.5 meters. Considering the variation of RSS energy fingerprints over time, Popleteev collected long-term FM data at fixed locations in an indoor environment to verify the performance of positioning [29 30].

We propose an FM-based positioning method using deep learning, namely FM-Pnet. Different from traditional FM-based positioning methods that mainly use RSS, our method utilizes the high-precision time-frequency representation of FM signals with a certain bandwidth, the maximum of which can cover the entire FM broadcast frequency band. By applying Short-Time Fourier Transform (STFT) to obtain the time-frequency representation of the signal, we use it as the input to a deep neural network to estimate the position. In order to improve the model’s generalization ability over a large time span, we also propose two strategies to enhance its robustness. Consistent with the deep learning paradigm, FM-Pnet consists of two stages: offline training and online inference. In the offline training stage, FM signals from known positions are collected to construct a dataset for training the model parameters. In the online inference stage, FM signals are received at the target location, undergo time-frequency transformation, and are fed to the well-trained model to estimate the position based on the output confidence. 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 will release these datasets after the publication of this paper.
- 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 formulate the problem. In Sec. III, we introduce our proposed FM-Pnet in detail. In Sec. IV, we compare the positioning performance of FM-Pnet with existing methods. Finally, we provide the concluding remarks in Sec. V.

II. 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, consider a flat area to be positioned, which can be divided into several grid blocks. The target point to be positioned is located within a certain grid. For a target point \((x, y)\), it can receive FM signals. Considering discrete time formulation, the received FM signal can be represented as

\[
r(n) = \sum_{i=1}^{M} h_i(n) \ast s_i(n) + w(n), n = 0, 1, \ldots, L - 1,
\]

where \(s_i(n)\) represents the narrowband FM signal in discrete time format, \(h_i(n)\) denotes the channel response, \(\ast\) denotes convolution, \(M\) represents the number of narrowband FM signals, \(L\) is the signal length, and \(w(n)\) is a additive white Gaussian noise (AWGN). From Eq. (1), it can be seen that the transmission paths differ among various points in relation to the radio tower, resulting in distinct channel responses and received signals. Consequently, different 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 is temporal variability in the 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 corresponding receiving location coordinates:

\[
\mathcal{M}_{FM} : r_\alpha(n) \rightarrow \text{Loc}_\alpha(x, y),
\]

where \(r_\alpha(n)\) represents the received FM signal in location \(\alpha\) and \(\text{Loc}_\alpha(x, y)\) represents the coordinate of location \(\alpha\). 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:

\[
\min ||\mathcal{M}_{FM}(r_\beta(n)) - \text{Loc}_\beta(x, y)||.
\]  

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

III. METHODOLOGY

A. Overall Framework

The method of using FM signals for positioning based on deep learning mainly relies on the powerful feature extraction ability of neural network models. With the assistance of a large amount of known position information of FM signal data, the network model continuously learns to map the input FM signal data to location coordinates through self-learning.

The proposed FM-Pnet in this paper enables end-to-end positioning by using FM signals, as illustrated in Fig. 2. Firstly, the framework selects the corresponding time-frequency representation computed from the raw FM signal as the 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 adopt the ResNeXt model, which performs excellently in the field of image classification, to process the input data of the dual-channel time-frequency representation. Unlike the original ResNeXt, we modify its stack depth and feature dimension to make it more suitable for handling the input.

Considering the impact of time span on FM-based positioning, FM-Pnet can choose to apply noise to the original signal before mapping the FM signal into a time-frequency representation. This enhances the robustness, adaptability, and generalization capability of the deep learning model trained, allowing it to better adapt to FM-based positioning over long time spans. Furthermore, considering 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 when using the trained classification model, we employ a weighted localization method similar as [13] 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 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 at different time points. 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. 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 comprises the partitioning of the signal into numerous brief, overlapping segments. Typically, each segment is subjected to windowing to mitigate spectral leakage, followed by individual application of a Fourier transform:

\[
\text{STFT}_r(m,k) = \sum_{n=0}^{N-1} [r(n)g(n-m)] e^{-j(2\pi(n-m)k/p)},
\]

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, \( g(\cdot) \) denotes the window function where a Hanning window is used in this paper. This transformation from the time domain to the frequency domain provides insights into the signal’s frequency content at different moments in time. Fig. 2 presents time-frequency representations of FM signals received at three distinct time instances. As shown in Fig. 2, 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 number is relevant to the FFT size and the overlap ratio. The overlap ratio typically refers to the ratio of the amount of overlapping samples between two adjacent time windows to the window size. These two factors also affect the resolution of the obtained time-frequency representation. The FFT size affects the frequency resolution of the time-frequency representation, where a larger value indicates a 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 ensure the integrity of the information, we extract the real and imaginary parts of this complex matrix separately and create a real matrix with dimensions of \( p \times q \times 2 \) as the input of the proposed FM positioning model:

\[
S(i,j,0) = (\text{Real}(\text{STFT}_r(\cdot)))_{i,j},
\]

\[
S(i,j,1) = (\text{Imag}(\text{STFT}_r(\cdot)))_{i,j},
\]

where \( S \) is the final time-frequency representation which is used as the model input, \( \text{Real}(\cdot) \) and \( \text{Imag}(\cdot) \) 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 \( 0 \leq i \leq p - 1, 0 \leq j \leq q - 1 \).

C. Adopted Network Structure

Convolutional Neural Network (CNN) relies on its hierarchical structure layout to have strong feature extraction capability for images and spatial structured data [34]. However, in CNNs, as the number of convolutional layers increases, degradation and gradient explosion phenomena occur. Residual Network (ResNet) effectively solves this problem by utilizing its skip connections, achieving deeper networks while maintaining stronger performance [35]. With the development of ResNet, ResNeXt [36] applies the concept of grouped convolution, significantly improving performance without increasing computational cost, and demonstrating outstanding robustness in computer vision tasks [37]. The ResNeXt network consists of a series of residual blocks, which have the same topology and follow two simple rules: 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 width of the module multiplies by 2. In ResNeXt, the form of splitting-transforming-merging can be represented as:

\[
R(x) = \sum_{i=1}^{C} T_i(x),
\]

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 $S$ 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 deepening the number of stacking layers. Table I presents the detailed structure of the network.

### 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. On the other hand, online positioning is the process of utilizing the trained model to estimate the location based on the received FM signal.

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 $3 \times 3$ positioning grid with a total of 9 positioning points, the positioning problem is converted into a 9-classification problem. FM-Pnet classifies FM signals received at location $i$ into the $i$-th category. Through this problem transformation, the deep learning training 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:

$$
L = - \sum_{i=1}^{G} \eta_i \log(p_i(S|f_\theta)),
$$

where $\eta_i$ represents the $i$-th element of the label of the time-frequency representation $S$, $p_i(S|f_\theta)$ represents the probability of the FM-Pnet network $f_\theta(\cdot)$ prediction $S$ 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 which have the same format of time-frequency input with the training data. To further improve the prediction performance of location estimation, we use the reevaluated results based on Bayes’law instead of the labeled location 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 [18]:

$$x_\rho = \sum_{i=1}^{G} \rho(i) \cdot x(i), \quad y_\rho = \sum_{i=1}^{G} \rho(i) \cdot y(i),$$

where $x_\rho$ and $y_\rho$ 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 of the target location inevitably undergoes changes over time, such as weather conditions, which may affect the propagation channel of FM signals. Therefore, at the same location, the received FM signals at different times will unavoidably have differences. Constructing an FM signal training set using FM signals collected over a short time span may have limited sample diversity, resulting in models trained based on this data being unable to adapt to changes in the channel environment caused by long time spans. 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 time-frequency representations. 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: Considering that the noise power may fluctuate during various moments of FM signal reception, the time-frequency representation tends to manifest differences at the same location. In real electromagnetic environments, the fluctuations in noise power are a common occurrence. 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:

\[ r'(n) = r(n) + N(n), \]  

where \( r(n) \) 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 simulating 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

\[ Q_{\text{max}} = \max_c(M_s), \quad Q_{\text{mean}} = \text{mean}_c(M_s), \]  

\[ Q_{\text{mix}} = \text{Cat}(Q_{\text{max}}, Q_{\text{mean}}), \]

where \( M_s \in \mathbb{R}^{2\times p \times q} \) represents the input dual-channel time-frequency representation, \( \max(\cdot) \) and \( \text{mean}(\cdot) \) respectively denote taking the maximum and average values in the channel dimension, and \( \text{Cat}(\cdot) \) 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 map. Finally, the attention weight matrix is obtained by applying the Sigmoid function, which restricts the element values between 0 and 1:

\[ Q_1 = \text{ReLU}(\text{Conv}2D1(Q_{\text{mix}})), \]  

\[ Q_2 = \text{Sigmoid}(\text{Conv}2D2(Q_1)), \]

where \( Q_1 \in \mathbb{R}^{p \times q} \) is the attention weight map, \( \text{Conv}2D1(\cdot) \) 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. \( \text{Conv}2D2(\cdot) \) is a 32-input and 1-output channel, 2D convolutional layer with a kernel size of 7, a stride of 1, and a 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

\[ Q_a = M_s \odot Q_2, \]

where \( \odot \) represents the Hadamard product.

IV. EXPERIMENTAL ANALYSIS

A. Experimental Setup

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

For the indoor scenario, we selected 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 selected a square area and placed 25 collection points which are marked as dot circles. These 25 collection points form an 8×8 m² square area, with a standardized lateral (vertical) distance of 2 m between adjacent collection positions. As for outdoor scenario, we chose 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 selected a square area measuring 15×12 m². Within this area, we set up 30 FM signal collection points. The horizontal (vertical) spacing between each collection point was set to 3 m.
As shown in Fig. 4(c) and Fig. 5(c), we abstracted the horizontal and vertical directions of both the indoor and outdoor 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 assigned a two-dimensional coordinate based on its actual physical position.

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

To examine the impact of signal bandwidth on positioning, we gathered FM signals with bandwidths of 320kHz (sampling rate of 400ksps) and 4MHz (sampling rate of 5Msps). Furthermore, in the outdoor scenario, we acquired data samples with an additional bandwidth of 20MHz (sampling rate of 25Msps) to assess the positioning performance with FM signals covering the entire frequency range. To assess the influence of FM signal time duration on positioning results, we extracted signal samples with lengths of 1,024, 4,096, and 16,384, respectively. To investigate the effect of the time span on the positioning method, we collected FM signals in the same location for multiple consecutive days. Distinct datasets were 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 were carried out on a computer equipped with an Intel Core i9-9900k CPU running at 3.6GHz and an NVIDIA GeForce RTX 2080. Model training was executed using the PyTorch framework. Throughout the training procedure, the Adam algorithm was employed for parameter updates. The batch size was set to 32. The initial learning rate was set to 0.001, and the learning rate was halved every two rounds during the training process with a total of 10 epochs.

4) Baselines: We will conduct a comparative analysis of our proposed method against existing techniques, namely deterministic KWNN [27], the probability-based histogram method (Histogram) [23], and the Combined method [9], which combines the KWNN and Histogram approaches. For the KWNN method, we iterate over all possible values of K and select the one that yields optimal positioning 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.

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:

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

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 predicted 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 outdoor and indoor scenarios are 7.8234 and 4.6795, respectively. Results exceeding these values are considered positioning failures. The unit of MDE is meters (m).

**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 taking the average of the standard deviation of the test distance errors for all samples in the test:

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

where \(D_i\) represents the distance error of sample \(S_i\). A higher value of STD suggests a broader distribution of errors, whereas a lower value indicates a more concentrated distribution of errors.

**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 Day 1. 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. 6. As shown in Fig. 6(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 outdoor scenario, as shown in Fig. 6(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. 7. As depicted in Fig. 7(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. 7(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.

**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 in Table [I] and Fig. [8].

As shown in Table [I] 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. 8(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 KNN 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. 8(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 KNN 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. 9.

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.
TABLE III
TEST RESULTS OF DIFFERENT SAMPLE LENGTHS

<table>
<thead>
<tr>
<th>Length</th>
<th>Methods</th>
<th>Indoor MDE</th>
<th>Indoor STD</th>
<th>Outdoor MDE</th>
<th>Outdoor STD</th>
</tr>
</thead>
<tbody>
<tr>
<td>1024</td>
<td>KNN</td>
<td>2.3403</td>
<td>2.9544</td>
<td>4.5737</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Histogram</td>
<td>2.1080</td>
<td>2.6170</td>
<td>4.4256</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Combined</td>
<td>2.0781</td>
<td>2.3594</td>
<td>4.2253</td>
<td></td>
</tr>
<tr>
<td></td>
<td>FM-Pnet</td>
<td>0.0968</td>
<td>0.4354</td>
<td>1.4118</td>
<td></td>
</tr>
<tr>
<td>4096</td>
<td>KNN</td>
<td>1.8370</td>
<td>2.9304</td>
<td>4.3577</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Histogram</td>
<td>1.9522</td>
<td>2.8504</td>
<td>5.0542</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Combined</td>
<td>1.9651</td>
<td>2.4805</td>
<td>4.9285</td>
<td></td>
</tr>
<tr>
<td></td>
<td>FM-Pnet</td>
<td>0.0689</td>
<td>0.3359</td>
<td>3.5516</td>
<td></td>
</tr>
<tr>
<td>16384</td>
<td>KNN</td>
<td>0.8254</td>
<td>1.1522</td>
<td>5.5042</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Histogram</td>
<td>0.8031</td>
<td>1.2850</td>
<td>4.0603</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Combined</td>
<td>0.7910</td>
<td>1.6340</td>
<td>5.8191</td>
<td></td>
</tr>
<tr>
<td></td>
<td>FM-Pnet</td>
<td>0.0028</td>
<td>0.1063</td>
<td>0.3669</td>
<td></td>
</tr>
</tbody>
</table>

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 KNN 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. 9(a). Compared to the other three methods, each CDF curve of FM-Pnet has a higher initial value and faster 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. 9(b). FM-Pnet achieves over 90% of the test samples with distance error lower than the random guessing value of 7.8234, while KNN 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.

In general, FM-Pnet has better positioning accuracy and stronger stability in samples of different lengths compared to RSS-based baseline methods. In the FM-Pnet method, as the sample length increases, the positioning performance keeps improving. This is because with the increase in sample length, the information contained in the time-frequency representation of each sample becomes richer, allowing for more accurate differentiation of their features during model training.

D. 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 IV, 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 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, KNN method exhibits the best performance, with MDE values reaching 6.3925 and 5.3261 on the test sets of Day2 and Day3, respectively.

Fig. 9. The influence of different sample lengths on the positioning performance. (a) Indoor scenario, (b) outdoor scenario.

TABLE IV
THE RESULTS OF CROSS-DATE TESTING

<table>
<thead>
<tr>
<th>Days</th>
<th>Methods</th>
<th>Indoor MDE</th>
<th>Indoor STD</th>
<th>Outdoor MDE</th>
<th>Outdoor STD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Day1</td>
<td>KNN</td>
<td>1.8370</td>
<td>2.9304</td>
<td>4.5737</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Histogram</td>
<td>1.9522</td>
<td>2.8504</td>
<td>5.0542</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Combined</td>
<td>1.9651</td>
<td>2.4805</td>
<td>4.9285</td>
<td></td>
</tr>
<tr>
<td></td>
<td>FM-Pnet</td>
<td>0.0689</td>
<td>0.3359</td>
<td>3.5516</td>
<td></td>
</tr>
<tr>
<td>Day2</td>
<td>KNN</td>
<td>4.5137</td>
<td>2.7612</td>
<td>6.3925</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Histogram</td>
<td>4.8040</td>
<td>2.8288</td>
<td>5.3261</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Combined</td>
<td>4.2250</td>
<td>2.2895</td>
<td>4.2188</td>
<td></td>
</tr>
<tr>
<td></td>
<td>FM-Pnet</td>
<td>2.6973</td>
<td>2.9491</td>
<td>5.6300</td>
<td></td>
</tr>
<tr>
<td>Day3</td>
<td>KNN</td>
<td>4.3361</td>
<td>3.1726</td>
<td>5.3261</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Histogram</td>
<td>4.8421</td>
<td>2.9123</td>
<td>5.0603</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Combined</td>
<td>4.1738</td>
<td>2.4121</td>
<td>4.0354</td>
<td></td>
</tr>
<tr>
<td></td>
<td>FM-Pnet</td>
<td>2.4397</td>
<td>2.9016</td>
<td>4.0460</td>
<td></td>
</tr>
</tbody>
</table>
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. 10(a) and Fig. 10(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 now evaluate the impact of noise injection enhancement on cross-date positioning performance by adding random white Gaussian noise to the received FM signals in the training process. To inject random noise with different power levels, we uniformly select the signal-to-noise ratio (SNR) from 0 dB to 50 dB and inject the corresponding noise into the signal. The remaining training parameters remained consistent with the non-noise enhancement case. We use the FM-PNet trained on the Day1 data to test its positioning performance on Day1, Day2, and Day3, without adding additional noise to the test signal samples. The results are shown in Fig. 11 and Table V.

As shown in Table V, the noise-enhanced model significantly improves the positioning performance over three days compared to the non-enhanced model. Both the MDE and STD of the positioning are reduced. Specifically, the MDE of Day1 decreased from 0.0689 to 0.0258, and STD decreased from 0.3595 to 0.2167. The MDE of Day2 decreased from 2.4397 to 1.7279, and STD decreased from 2.9016 to 2.4748. The MDE of Day3 decreased from 2.6973 to 1.7279, and STD decreased from 2.9491 to 2.4748. As shown in Fig. 11, 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.

3) Enriching Training Samples: We also evaluate the effectiveness of enriching the training samples strategy in improving the generalization performance of FM-Pnet based on outdoor FM signal data. Specifically, we create three datasets to train three positioning models. Dataset 1 comprises FM signal data from Day1, Dataset 2 includes data from both Day1 and Day2, and Dataset 3 incorporates data from Day1, Day2, and an additional day (excluding Day3). The training set consists of 8,000 samples per day, and 2,000 samples from Day3 are designated for testing.

Table VI shows the performance of three datasets. The MDE and STD of the experimental results are shown in Table VI. When trained with one day of data, the FM-Pnet method achieves an MDE of 4.0460 and an STD of 4.7674 in cross-date testing. Training with two days of data results in improved performance, with an MDE of 3.5093 and an STD of 4.2751, respectively. When training with FM data from three days, the MDE and the STD further decrease to 2.4504 and 3.7232. In comparison, the best-performing Combined method among the baseline methods achieves MDE of 5.4003, 5.5827, and 5.1771 when utilizing the enrichment strategy with one, two, and three days of FM signals, respectively. Little improvement has been observed with the baseline methods.
This suggests that the proposed FM-Pnet can achieve more substantial enhancements in positioning generalization performance through the use of more enriched data. The CDF result is shown in Fig. 12. It also illustrates that enriching the training samples enhances positioning accuracy across different dates in complex outdoor scenarios.

E. 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, we randomly selected a signal sample from the test set and plotted the corresponding input matrix (real part of STFT) and its associated attention weight matrix, as shown in Fig. 13(b). From the left figure, it can be observed that there are four distinct narrowband signals with higher SNR in the input time-frequency representation, and the corresponding pixel bars in the attention matrix are also quite pronounced. In addition, 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 relatively 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.

V. 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 enriching training samples with cross-date signals or injecting noise in the training process, 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 with promising prospects in the future.

REFERENCES

[1] H. Huang, G. Gartner, J. M. Krisp, M. Raubal, and N. Van de Weghe, “Location based services: ongoing evolution and re-