MFFALoc: CSI-Based Multi-features Fusion Adaptive Device-free Passive Indoor Fingerprinting Localization

Xinping Rao, Zhenzhen Luo, Yong Luo, Yugen Yi, Gang Lei and Yuanlong Cao

***Abstract*—In recent years, the rise of location-based service applications such as cashier-less shopping, mobile advertisement targeting, and geo-based augmented reality (AR) has been remarkable. These applications offer convenient and interactive experiences by utilizing indoor localization technology. One popular research area in indoor localization is passive fingerprinting localization based on Channel State Information (CSI), which uses general-purpose Wi-Fi platforms and "unconscious cooperative sensing" to achieve device-free localization. However, existing studies face challenges related to inadequate fingerprint richness, limited distinguishability, and inconsistent fingerprint features in real-world dynamic environments. To address these challenges, we prpose MFFLoc in this paper. MFFLoc extracts and processes amplitude and phase information from CSI in a 2D manner. It then fuses the amplitude and phase information using multimodal fusion representation, resulting in rich and distinguishable fused fingerprint features. This approach allows MFFLoc to achieve satisfactory accuracy with just one communication link, reducing deployment costs. To overcome the issue of inconsistent fingerprint features in dynamic environments, MFFLoc proposes an unsupervised domain adaptation method. It employs a dual-flow structure, with one flow operating in the source domain and the other in the target domain. The adaptation layer, with correlated weights, remains unshared between the two flows. Meta-learning is also used to automatically determine the most suitable adaptation layer. Through extensive 6-day experiments conducted in a dynamic indoor environment, MFFLoc showcases superior performance compared to state-of-the-art systems. It demonstrates higher localization accuracy and robustness, making it a promising solution for indoor localization applications.**

***Index Terms*—Channel State Information (CSI), Device-free Passive Indoor Localization, Fingerprinting, Unsupervised Domain Adaptation.**

I. Introduction

In recent years, as society advances and technology rapidly evolves, there's a growing demand for safe, intelligent indoor environments. Real-time indoor localization and sensing, combined with smart technology, can enhance personalized services, automation, safety, and comfort [1]. Indoor people location awareness technology is now essential, shaping work, personal life, and new industries. Wi-Fi-based Device-free Passive Localization (WDFL) detects and locates targets by analyzing Wi-Fi signal changes, offering cost-effective, wide-reaching applications without needing specific devices [2]–[4]. While Received Signal Strength (RSS) is widely available and adaptable, it lacks precision due to multipath fading and device diversity, hindering its effectiveness in cost, granularity, and accuracy for localization [5]–[8].

The release of the IEEE 802.11n standard has facilitated the accessibility of Channel State Information (CSI) [9]. CSI offers fine-grained signal data, with each set encompassing amplitude and phase details of individual Orthogonal Frequency Division Multiplexing (OFDM) subcarriers [10]. In contrast to RSS, CSI presents a more comprehensive portrayal of the wireless environment, enabling effective characterization of multipath propagation in indoor wireless signals and the sensitive detection of multipath signal variations. Consequently, CSI holds the potential to significantly enhance wireless sensing sensitivity, extend operational range, and augment overall reliability [11], [12]. In recent times, CSI-based device-free passive localization has garnered prominence as one of the most practical and convenient approaches for addressing indoor personnel localization and sensing challenges. It mainly can be classified into two main categories: radio signal propagation model-based solutions [13]–[15], and fingerprint-based solutions [16]–[22]. Radio signal model-based solutions employ statistical models to estimate a user's location using CSI measurements, eliminating the need for fingerprint databases and reducing site surveys. However, accurate modeling is challenging due to theoretical limits and indoor multipath interference. In contrast, fingerprint-based solutions passively detect a target's presence through WiFi propagation changes, offering flexibility and ease of deployment without pre-measuring AP locations. The CSI-based device-free passive fingerprinting localization technique is recognized for its accuracy and feasibility, gaining interest in academia and industry. Nonetheless, it still confronts numerous challenges in practical applications.

Challenge 1: Many existing CSI-based device-free passive fingerprinting localization systems struggle to attain high localization accuracy in dynamic environments and are unsuitable for large-scale deployment, thereby limiting their scalability. These systems often rely on amplitude or phase information extracted from CSI measurement samples as fingerprint features, typically employing simple preprocessing

This paper was produced by Xinping Rao, Zhenzhen Luo, Yong Luo, Yugen Yi, Gang Lei and Yuanlong Cao (Corresponding author: Zhenzhen Luo). They are with the School of Software, Jiangxi Normal University, Nanchang 330027, China. (email: raoxp@jxnu.edu.cn; zhenzhenluo@jxnu.edu.cn; luoyong1020@jxnu.edu.cn; yiyg510@jxnu.edu.cn; leigang@jxnu.edu.cn; ylcao@jxnu.edu.cn)

Manuscript received July 18, 2023;

or phase calibration. Unfortunately, they frequently overlook the correlations between multiple CSI measurement samples over time and across multiple channels, resulting in the loss of valuable channel-related information. Consequently, CSI-based location fingerprint features lack richness, necessitating the deployment of numerous Monitor Points (MPs) and Wireless Access Points (APs) to achieve a specific level of positioning accuracy. This substantial increase in deployment and maintenance costs significantly hinders the widespread adoption of device-free passive localization technology for indoor applications.

Challenge 2: Many existing CSI-based device-free passive fingerprinting localization systems neglect the dynamic nature of fingerprint features. These systems often assume that acquired CSI location fingerprints remain valid for an extended period, an assumption that doesn't hold in practical applications. Furthermore, post-deployment, the process of associating new data (i.e., correlating newly recorded WiFi CSI readings with location) can be cumbersome and occasionally infeasible, compromising system robustness and adaptability to changing environments. While some research focuses on dynamic fingerprint localization techniques to mitigate the need for manual recalibration, they still face challenges, including high algorithm complexity, manual intervention requirements, and the inability to simultaneously handle various time scales, etc. To address these challenges, in this paper, we introduce MFFALoc, a novel CSI-based device-free passive fingerprinting localization system. For Challenge 1, MFFALoc employs multi-feature fusion, leveraging a single transmitter-receiver device (monitoring point) for cost-effective communication. It analyzes the multi-antenna CSI measurements, utilizes techniques like wavelet domain denoising, and generates a high-resolution fused fingerprint feature matrix by merging amplitude and phase time-frequency feature matrices. For Challenge 2, MFFALoc incorporates model domain adaptation. It treats the original environment as the source domain and the changed environment as the unlabeled target domain. Dual-flow processing is adopted for source and target data, with controlled weight differences between corresponding layers. MFFALoc uses Maximum Mean Discrepancy (MMD) to measure the dissimilarity between source and target representations at the output layer, allowing the localization model to adapt to environmental changes with updated, unlabeled data.

It is also worth noticng that MFFALoc is different from our previous work LTLoc [23]. While LTLoc can identify domain-invariant features using meta-learning and small labeled datasets after environmental changes, it relies solely on independent single CSI features, neglecting the correlation among CSI samples over time. Additionally, the labeling process for new data is labor-intensive and sometimes unfeasible in practical applications. In contrast, MFFALoc takes a distinct approach by incorporating both amplitude and phase information from multi-channel CSI measurements to construct the multi-antenna time-frequency feature matrix. It then fuses this matrix with the phase time-frequency feature matrix to create the final feature fingerprint. By preserving correlation information across multi-channel and time domains, MFFALoc enhances the richness of CSI location fingerprint features. Moreover, MFFALoc achieves domain adaptation using untagged data through an explicit analog domain offset within its dual-flow structure. This approach enables MFFALoc to adapt to unforeseen environmental changes autonomously, while maintaining satisfactory localization accuracy.

In summary, the main contribution of this paper can be described as follows:

1) MFFALoc employs techniques like wavelet domain denoising (WDD) to eliminate background features and environmental noise, enhancing target location features and generating multi-antenna time-frequency feature matrices. It uses a feature fusion framework to combine the multi-antenna amplitude-time frequency feature matrix with the phase-time frequency feature matrix, resulting in high-resolution fused fingerprint features that enhance localization accuracy and quality.
2) MFFALoc includes unsupervised domain adaptation, ideal for real-world deployments. It utilizes a dual-flow structure where the first flow processes source domain data, and the second handles target domain data. While layers in each flow may have distinct weights, constraints prevent significant divergence. MFFALoc employs the Maximum Mean Discrepancy (MMD) metric to measure the dissimilarity between source and target representations at the output layer, effectively capturing the relationship between the domains. Consequently, the localization model adapts to environmental changes using updated unlabeled data.
3) We conducted comprehensive testing of MFFALoc through six days of real-world trials in a congested indoor dynamic setting. The results demonstrate that MFFALoc exhibits a low deployment overhead and achieves consistent and satisfactory localization accuracy in dynamic scenarios, using just a pair of transceivers.

The remainder of this paper is structured as follows. Section II provides an overview of the existing literature, examining relevant works in the field. In Section III, we present preliminary aspects of CSI and discuss the inherent inconsistency problem associated with WiFi fingerprinting. Moving forward, Section IV outlines the system design and introduces key definitions, while also presenting detailed information about the functional modules. In Section V, we conduct real-world experiments to assess the effectiveness of MFFALoc in various environmental conditions. Finally, in Section VI, we conclude this paper and explore potential future directions.

II. Related Works

In this section, we present a comprehensive review of the state-of-the-art in device-free passive indoor localization. We focus on discussing the most relevant works to our research, specifically the CSI-based Device-free Passive Fingerprinting Localization and Dynamic Fingerprint Localization. These works provide valuable insights and serve as the foundation for our proposed methodology in this study.

A. CSI-based Device-free Passive Fingerprinting Localization

As previously mentioned, CSI-based Device-free Passive Fingerprinting Localization technology stands as the most

prevalent Wi-Fi positioning method, representing a prominent trend in the field, with many researchers carrying out research on it. Liu et al. [16] introduced a device-free indoor localization method using CSI and a visibility map algorithm with finite penetrability levels. Their approach involved modeling frequency correlation between subcarriers, constructing a complex network based on CSI, and using network topology to analyze relationships among subcarriers. They employed support vector regression (SVR) for indoor localization in various environments. Wei et al. [17] implemented a phase calibration method to mitigate phase shifts caused by imperfect synchronization and used a structural similarity-based enhancement method to extend datasets and obtain additional fingerprint feature information, enhancing localization accuracy. Han et al. [18] developed a defense method using Convolutional Neural Networks (CNN) to counter device-free localization attacks. They transformed the localization problem into an image classification problem and employed CNN for anomaly detection, enhancing system robustness and security. Zhang et al. [19] introduced the Integrated Multicore Extreme Learning Machine (ELM) to optimize spatio-temporal CSI information for improved localization accuracy in complex indoor environments. Yan et al. [20] presented a novel method for constructing radio images based on CSI, using amplitude measurements along with temporal, spatial, and frequency domain information to enhance fingerprint features. Zhang et al. [21] devised an integrated hierarchical framework, the Data and Knowledge Twin-Driven Large-Scale Device-free Localization (DFL) framework. It addresses complex surveillance region partitioning and employs a class-specific cost-adjusted limit learning machine classifier to improve localization. Wang et al. [22] introduced CiFi, a method that estimates the Angle of Arrival (AoA) using phase data from CSI. The estimated AoA image serves as input for an offline deep CNN, enabling network weight training.

The aforementioned research studies have undoubtedly made substantial contributions to CSI-based device-free passive fingerprinting localization techniques. *However, they still fail to effectively address the issue of fingerprint inconsistency. This limitation can significantly hamper performance when dealing with dynamic environments, and it also leads to elevated deployment costs due to the necessity of additional transmitting and receiving devices*.

B. Dynamic Fingerprint Localization

A fundamental assumption in device-free passive fingerprinting localization is that the training data distribution aligns with testing conditions, allowing for direct application of offline databases over time. However, real-world environments change, challenging this assumption. This dynamic fingerprinting issue has prompted research on reducing manual database updates. The researchers first thought of calibration-free schemes that combat dynamic fingerprint changes without manual intervention. Zou et al. [24] used a custom base station to collect tagged Received Signal Strength (RSS) measurements and applied Gaussian process regression for calibration-free localization. Wu et al. [25] reduced RSS fingerprint uncertainty by analyzing spatial gradients across multiple locations, enhancing localization reliability. Guo et al. [26] introduced the concept of fingerprint quality to assess candidate location reliability and prioritized the highest-quality location. *However, these schemes depend on dedicated devices and knowledge of base station or access point locations, limiting their applicability in device-free passive localization scenarios and posing scalability challenges*.

Subsequently, the researchers endeavored to construct and update the offline database by leveraging user-contributed fingerprints, i.e., Crowdsourcing shcemes. Rai et al. [27] used smartphone inertial sensors for a PDR-assisted crowdsourcing approach. Jung et al. [28] proposed an optimization scheme relying on untagged crowdsourced fingerprints to estimate positions. Zhao et al. [29] used multi-dimensional scaling to transform raw RSS data into a spatial representation for a comprehensive fingerprint database. Wan et al. [30] introduced a self-calibrating indoor positioning framework combining Wi-Fi ranging, crowdsourced fingerprinting, and low-cost sensors, achieving accurate multi-source fusion. *However, these schemes require manual user device intervention, making them unsuitable for device-free passive localization. Privacy concerns also arise as user-provided samples may disclose sensitive location information, limiting their applicability in privacy-sensitive scenarios*.

Clearly, neither calibration-free nor crowdsourcing schemes find applicability in the domain of device-free passive positioning due to the aforementioned drawbacks. Consequently, researchers have shifted their focus to calibration-reduction schemes and have done numerous relevant studies. For example, Chen et al. [31] introduced Fidora, a WiFi-based system addressing fingerprint inconsistency challenges through domain adaptation and clustering. Li et al. [32] proposed a dynamic fingerprint adaptation framework, requiring minimal human intervention for indoor localization. Chen et al. [33] presented a few-sample migration learning system, reducing data collection and labeling costs while achieving accurate localization. Chen and Chang [34] used Generative Adversarial Networks (GANs) to augment their dataset. Khatab et al. [35] trained autoencoders with unlabeled data to update the fingerprint database. Zhou et al. [36] introduced the AdapLoc method, employing domain adaptation and semantic alignment to reduce recalibration needs while maintaining accuracy. *However, these efforts often involve complex feature homogeneity handling, requiring training of feature spaces or encoders before localization, adding complexity to the algorithms. Moreover, they may struggle with addressing challenges across different time scales effectively*.

In summary, due to the rapid advancement of wireless communication technology and the widespread adoption of mobile smart devices, CSI-based device-free passive fingerprinting localization technology has emerged as a prominent solution among indoor localization field. It is considered a viable approach to meet current universal requirements. However, challenges, as outlined in Section I, persist for CSI-based device-free passive fingerprint localization techniques, particularly in the areas of fingerprint characterization, fingerprint inconsistency and deployment consumption.

III. PRELIMINARIES

A. Channel State Information

Wi-Fi signals utilizing the IEEE 802.11n communication protocol employ OFDM modulation. OFDM serves as a technique for encoding digital data onto multiple carriers at varying frequencies, and it has gained extensive adoption in wireless communications. This method partitions the channel into several orthogonal sub-channels, converting the high-speed data stream into a low-speed parallel data stream at the transmitter. Subsequently, it modulates the corresponding sub-channel for transmission. Upon reception, the signal is demodulated at the receiver and restored to its original data form. Recent advcances in wireless communications have facilitated the acquisition of a sampled version of the Channel Frequency Response (CFR) from a readily available commercial off-the-shelf (COTS) Wi-Fi Network Interface Card (NIC). The extracted sampled CFR is commonly represented as a sequence of complex numbers, containing information about the amplitude and phase of individual subcarriers:

$$H(f_i) = ||H(f_i)||e^{\angle H(f_i)} \tag{1}$$

where $H(f_i)$ is a CFR sample of the i^{th} subcarrier, with $||H(f_i)||$ and $\angle H(f_i)$ denote its amplitude and phase response, respectively. CFRs sampled at different subcarriers are generally treated as one CSI sample, which can be expressed as $H = \{H(f_i) \mid i \in [1, N]\}$, where N is the total number of subcarriers.

It is noteworthy that the CSI data varies depending on the specific tool utilized. Two commonly used tools are the Intel 5300 NIC CSI Tool [9] and the Atheros CSI tool [37]. The Intel 5300 NIC CSI Tool is designed around the Intel WiFi Wireless Link 5300 802.11n MIMO radio, incorporating modified firmware and open-source Linux wireless drivers. This tool leverages the capabilities of the IWL5300 to provide 802.11n CSI with 30 subcarrier groups. A CSI record generated by this tool is represented as an $M \times 30$ matrix, where M denotes the number of transmit and receive antenna pairs. The Atheros CSI tool, on the other hand, is an open-source 802.11n measurement and experimentation tool. It enables the extraction of detailed Physical Layer (PHY) wireless communication information from Atheros WiFi NICs, encompassing CSI, received packet payload, and other relevant data (such as timestamp, RSSI per antenna, data rate, etc.). The tool is theoretically compatible with all types of Atheros 802.11n WiFi chipsets. The CSI record is represented as an $M \times 56$ matrix when operating with 20 MHz subcarriers and an $M \times 144$ matrix when operating with 40 MHz subcarriers. Here, M represents the number of transmit and receive antenna pairs. In this study, we employed the Intel 5300 NIC CSI Tool for CSI data collection. The setup consisted of one transmitting antenna and three receiving antennas. As a result, the CSI sample matrix for a single packet is represented as a $1 \times 3 \times 30$ matrix, taking into account the number of transmitting antennas, receiving antennas, and subcarrier groups.

When the target is located within the localization area, it influences the propagation of the radio signal, resulting in phenomena like diffraction and refraction. These alterations subsequently cause changes in the CSI. Our previous studies [38], [39] have confirmed that both the amplitude and phase information obtained from CSI exhibit variations depending on the target's location. Based on this fact, the CSI-based device-free passive fingerprinting localization solutions extract the amplitude and phase information from CSI, which are used as fingerprint features after certain processing, and predicts the target's location by comparing the fingerprint features derived from recent CSI measurements with previously collected fingerprints through the matching algorithm, as depicted in Fig. 1 . These solutions generally involve two distinct phases: offline and online. In the offline phase, CSI measurements are collected while the target is located at a known position within the localization area. Then, the fingerprint features are extracted from the recorded CSI readings, and these features, along with their corresponding known locations, are consolidated to establish a fingerprint database. Moving to the online phase, a matching algorithm is employed to compare the fingerprint features extrcted from the current CSI measurements obtained from unknown locations with the fingerprint features stored in the fingerprint database. The algorithm determines the closest match between the current fingerprint features and the existing fingerprints, ultimately returning the predicted position of the user.

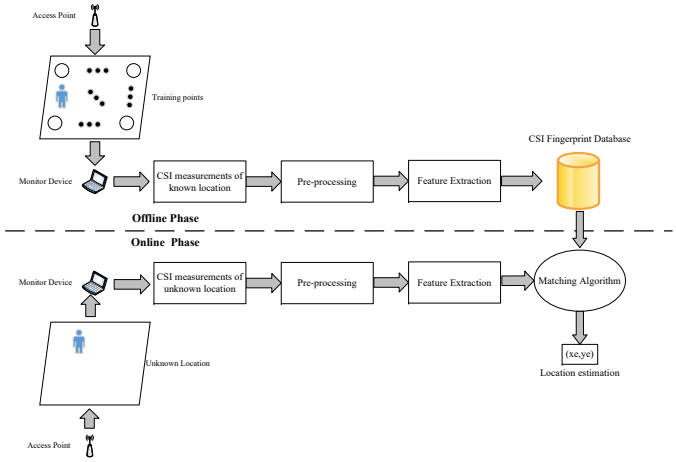

Fig. 1. Illustration of CSI-based Device-free passive fingerprinting localization solutions.

B. CSI Fingerprint Inconsistency Problem

Previous studies conducted [23], [38], [39] have provided evidence that CSI fingerprints exhibit sensitivity not only to the presence and location of the target but also to variations in the surrounding physical environment. Even if the target remains in a stationary position, the CSI fingerprints can still undergo variations over time, which are primarily attributed to changes in the surrounding environment, such as Open and close doors, move tables and chairs, and other environmental conditions. In this paper, we refer to this problem as WiFi fingerprint inconsistency. At the first blush, it may appear that the solution to this problem is straightforward. One possible approach would be to gather a large number of fingerprints

and utilize them to train a system capable of encompassing all possible fingerprint variations. Indeed, collecting an exhaustive set of fingerprints at a (sub)metre resolution to account for all possible environmental changes is a daunting and time-consuming task. The sheer number of potential variations in the wireless environment makes it impractical and resource-intensive to collect and include every single fingerprint in the database. Another plausible approach involves marking new CSI fingerprints whenever there is a change in the environment and utilizing these marked data to update or retrain the deployed localization system. However, this approach also presents challenges. To mark data for environmental changes, it is necessary to first confirm that a change has indeed occurred. This process can be cumbersome and impractical to implement in real-world scenarios. Additionally, marking new CSI fingerprints requires manual entry or collection of real position information from auxiliary systems, which can ntroduce further difficulties such as ensuring the accuracy and fidelity of the position information, as well as addressing synchronization issues between different systems. Therefore, while this approach may seem reasonable in theory, its practical implementation poses significant difficulties.

IV. MFFALoc Design Overview

To overcome the above challenges, in this section, we propose the MFFALoc system, and elaborate on the system design and implementation. The system overview is represented with an illustration in Fig. 2.

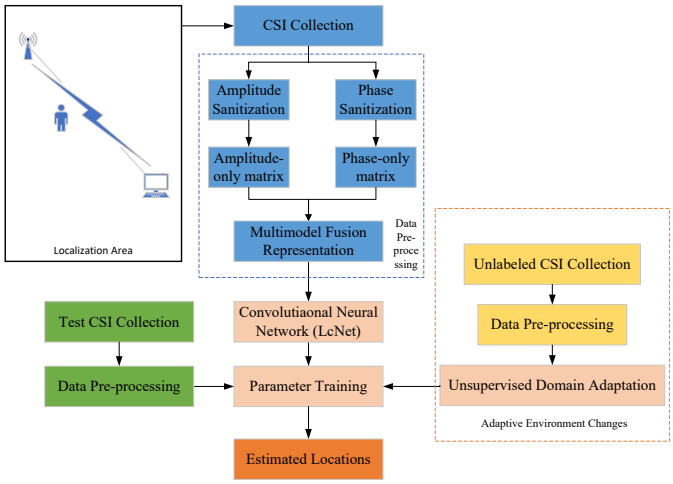

Fig. 2. Overview of the MFFALoc system.

It can be seen, the MFFALoc system comprises a singular router (AP), a receiver equipped with an Intel NIC 5300 board. *These components collectively establish a solitary communication link, forming the deployment of a minimal localization system.* The key implementation factor of the system lies in the comprehensive utilization of signal reflections from multiple antennas on the receiver, which facilitates a more profound understanding of channel fluctuations caused by objects within the sensing area. Specifically, it performs in-depth analysis of the amplitude and phase information extracted from the multi-antenna CSI measurements, applies wavelet domain denoising (WDD) and other processing techniques to eliminate background features and environmental noise, emphasizes the target location features, and generates the multi-antenna time-frequency amplitude feature matrix and multi-antenna time-frequency phase feature matrix, respectively. Subsequently, it fuse the above two matrices through a multimodel fusion representation framework to convert the potential spatial features within them into a feature map in the spatial domain ultimately yielding high-resolution fused fingerprint features. Finally, it is correlated with known locations and utilized to train the localization model, which in MFFALoc refers to a CNN model. Another crucial factor in the implementation of the MFFALoc system is the utilization of an unsupervised domain adaptation algorithm to solve the fingerprint inconsistency problem so that it can automatically adapt to environmental changes, ultimately achieving high granularity and robust indoor device-free passive localization.

A. Data Pre-processing

1) Amplitude and Phase Sanitization: Affected by the complex multipath environment indoors, the raw CSI measurements contain anomalous samples with multipath superposition and noise effects, which can affect the training of the localization model and lead to significant degradation of the localization performance. Hence, it becomes imperative to preprocess the acquired CSI data. Specifically, the amplitude data extracted from the CSI measurements undergo denoising firstly. While ordinary filters can eliminate some noise, they tend to compromise signal integrity. Therefore, in order to address this issue effectively, Wavelet Domain Denoising (WDD) is employed in the MFFALoc system for processing the collected CSI amplitude. Upon wavelet transformation of the signal, the resulting wavelet coefficients encapsulate crucial information. By selecting an appropriate threshold, wavelet coefficients exceeding the threshold are deemed as signal and retained, while those below the threshold are regarded as noise and eliminated by setting them to zero for denoising purposes. Although wavelet domain denoising can be largely considered as a form of low-pass filtering, it surpasses traditional low-pass filtering by effectively preserving the characteristics of the denoised signal [40]. It is evident that wavelet domain denoising is essentially a fusion of feature extraction and low-pass filtering, as illustrated in Fig. 3.After the noise-containing signal is input, it undergoes another round of low-pass filtering in the feature extraction stage. The processed signal is then reconstructed with the feature signal to obtain the final result. The noisy signal can be represented as equation 2:

$$S(k) = f(k) + \varepsilon \cdot e(k) \tag{2}$$

where $S(k)$ denotes the noisy signal, $f(k)$ is the useful signal we want, $e(k)$ denotes the noise, and $\varepsilon\cdot$ denotes standard deviation of the noise cofficient.

Subsequently, we proceed to process the raw phase data extracted from the CSI measurements. Due to the lack of time synchronization, the raw phase information may contain random errors and environmental noise. As a result, it cannot

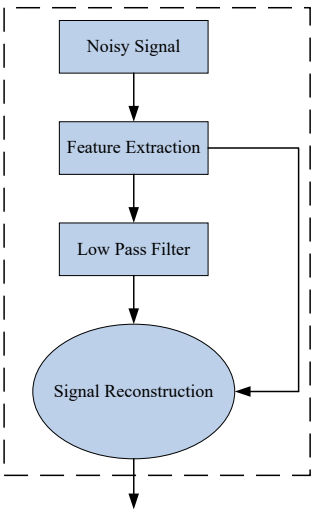

Fig. 3. Illustration of Wavelet Domain Denoising (WWD).

accurately describe the corresponding position information and is therefore unuseable. We need to calibrate the raw phases using the phase calibration algorithm proposed in [41] to obtain stable and usable calibrated phase data. The calbrated phase can be expressed as equation 3 below:

$$\hat{\varphi}_i = \widetilde{\varphi}_i - km_i - b = \widetilde{\varphi}_i - \frac{\widetilde{\varphi}_{30} - \widetilde{\varphi}_1}{m_{30} - m_1} m_i - \frac{1}{30}\sum_{j=1}^{30} \widetilde{\varphi}_j \quad (3)$$

where $\hat{\varphi}$ is a calibrated phase sample we want, $\widetilde{\varphi}$ is a raw phase sample, i, j are the subcarrier number, k, b are the slope and variance of the linear transformation, m_i is the index of subcarrier (see Table $7-25$f on page 50 of IEEE $802.11n-2009$ standard [42]). The instability of the phase data is effectively reduced by the linear transformation and the raw phase data is converted into analysable calibrated phase data.

After the above processing, the denoised amplitude and the calibrated phase information can better describe the information of each training location and improve the fingerprint richness.

2) Multimodel Fusion Representation: After the amplitude and phase Sanitization, they are used to construct the multi-antenna time-frequency amplitude feature matrix and multi-antenna time-frequency phase feature matrix. This construction aims to preserve their correlation not only in terms of timing but also across multiple channels. The raw CSI data are complex values of 30 subcarrier frequencies and transmitted between 1 transmitter antenna and 3 receiver antennas (as described in Section III-A above), and the multi-antenna amplitude and phase time-frequency feature matrices constructed by combining 30 consecutive amplitued and phase samples at the 30 subcarrier frequencies are a $3 \times 30 \times 30$ amplitude tensor and a $3 \times 30 \times 30$ phase tensor. These tensors have 3 channels just like images, which are facilitate the subsequent processing of the multimodal fusion representation model. To enhance the learned representations and capitalize on the complementary and redundant nature of the amplitude and phase features, we employ pixel-level multimodal fusion. This multimodal fusion representation combines multi-antenna time-frequency amplitude feature matrices and multi-antenna time-frequency phase feature matrices of a common training point. The objective is to leverage the multi-feature information contained in the CSI, thereby obtaining the final high-resolution fused fingerprint feature representations. It enables us to make optimal use of the available CSI data and exploit the synergistic benefits of integrating both amplitude and phase features. This comprehensive process is illustrated in Fig. 4, providing a detailed explanation and visualization for better understanding.

The fusion process commences by extracting potential CSI space features through two separate encoders. One encoder processes the amplitude tensor, while the other manages the phase tensor. Both tensors possess dimensions of $3 \times 30 \times 30$ (3 receiving antennas, 30 consecutive samples, 30 subcarrier frequencies). The amplitude and phase tensors undergo expansion and are individually subjected to Multiple Layer Perceptrons (MLPs) to acquire their respective CSI potential space features. The one-dimensional features from these encoding branches merge and enter another MLP for feature fusion. Subsequently, these CSI potential space features undergo transformation into a spatial domain feature map. The fused one-dimensional features reshape into a 24×24 two-dimensional feature map. Spatial information extraction involves two convolutional blocks with a kernel size of 9×9, a stride of 1, and no padding, yielding a more compressed feature map with spatial dimensions of 6×6. To match the input dimensions commonly used in CNN networks, considering computational complexity and feature richness, four inverse convolutional layers with a kernel size of 6×6, a stride of 1, and no padding, are employed to upsample the low-dimensional encoded feature maps to dimensions of $3 \times 30 \times 30$. This prepares them for input into the localization CNN model, ultimately producing the desired high-resolution fused fingerprint features.

B. Localization Model Structure

In the MFFALoc system, the localization task is treated as a regression task. The objective is to train the parameters of a CNN network using the high-resolution fused fingerprints obtained from the training points, along with their corresponding coordinates. The aim is to enable accurate prediction of the target location when provided with an unknown location fingerprint. This regression-based approach allows the CNN network to learn the spatial relationships and patterns present in the high-resolution fused fingerprints, enabling accurate localization predictions.

Fig. 5 depicts the architecture of LC-Net, the CNN utilized for localization in our MFFALoc. The LC-Net primarily comprises three Convolutional (Conv) blocks. Each Conv block is composed of convolutional layers followed by LeakyReLU activation layers. These Conv blocks play a crucial role in extracting relevant features from the input data by applying convolutional operations and incorporating non-linearity

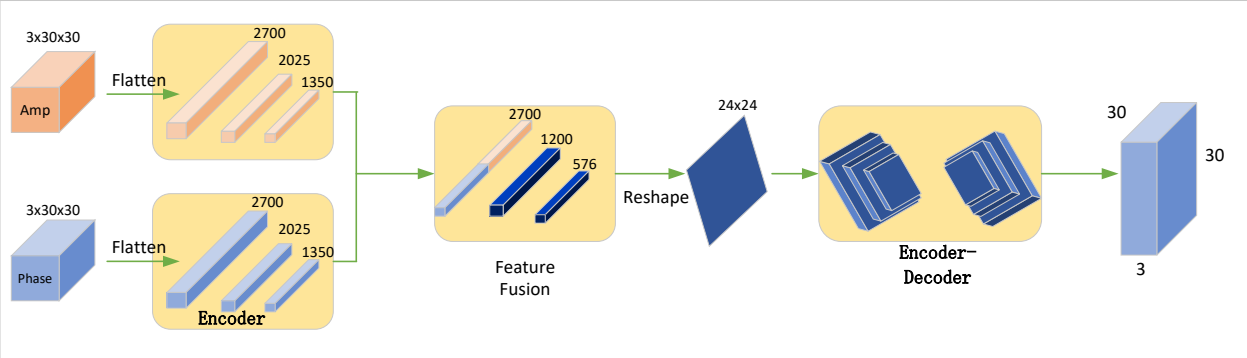

Fig. 4. Multimodal Fusion Representation Process. Two encoders extract the features from the amplitude and phase in the CSI domain. Then the features are fused and reshaped before going through an encoder-decoder network. The output is a $3 \times 30 \times 30$ feature map in the image domain.

through the LeakyReLU activation function. The combination of these Conv blocks enables LC-Net to learn and capture meaningful spatial patterns and relationships within the fused fingerprints, facilitating accurate localization predictions. Following the execution of a series of Conv blocks, LC-Net employs two fully connected layers to obtain the position coordinates. It is important to highlight that our LC-Net does not incorporate pooling operations. The decision to exclude pooling operations stems from the fact that each pixel in the obtained high-resolution fused fingerprint features contains detailed positional information. Applying pooling operations would potentially blur or distort these fine-grained positional features. Moreover, pooling operations are downsampling, which can result in the loss of valuable information [43]. By omitting pooling operations, LC-Net preserves the intricate positional details within the fused fingerprints, allowing for more precise localization. This design choice ensures that the network can leverage the full potential of the positional features present in the high-resolution fused fingerprints.

In the LC-Net, the fully connected layers are Gaussian distribution initialization ($std = 0.005$), bias are constant initialization (0.1). During training, the learning rate is initialized to 0.0001, the batch size N is 25, the training epochs $N_{epochs} = 100$, the Optimization is SGD and the loss fuction is the MSELoss (Mean Square Error Loss), which can be expressed as:

$$l = \frac{1}{N}\sum_{n=1}^{N}(x_n - \hat{x}_n)^2, \tag{4}$$

where x_n and $\hat{x}_n$ are the label and output of the LC-Net, respectively.

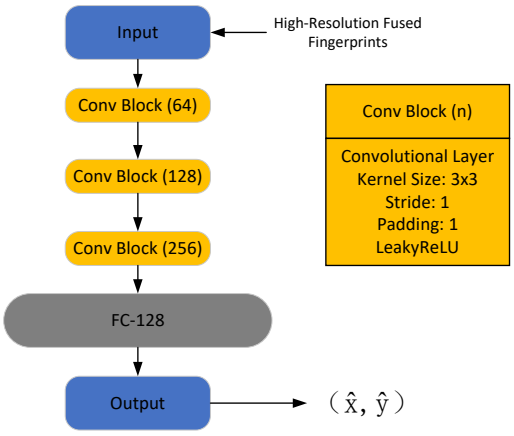

Fig. 5. The architecture of LC-Net.

C. Unsupervised Domain Adaptation

Once the LC-Net is trained with the fingerprint database, it establishes a mapping relationship between the fingerprint features and coordinates at each training point. Changes in the fingerprint features can significantly degrade the model's localization performance. However, the knowledge relevant to the task can still be extracted from the feature space of its middle-hidden layer. Therefore, the core concept of the MFFALoc system is to ensure that the deep network adapts to the localization task when the fingerprints change. The weights should be correlated but different for each of the two networks, one for the localization task before the fingerprint change and the other after. Essentially, this represents a domain adaptation problem within the field of transfer learning. Consequently, we can selectively share the parameters of the corresponding layers in the CNN model through the unlabeled data after fingerprints changed, which not only reduces the training complexity but also improves the transfer results compared to other similar approaches.

To achieve our objective, we propose a novel unsupervised domain adaptation method that incorporates a dual-flow structure, as illustrated in Fig. 6. One flow operates on the source-domain training set, referred to as the Fingerprint Database, while the other is applied to the target-domain training set, consisting of unlabeled samples. Both networks are trained jointly. We introduce meta-networks for transfer learning, which serve the purpose of automatically determining the source layers of the source model that are valuable and pertinent for learning the target task. While allowing the weights of the relevant layers to differ between the two networks, we also aim to ensure that they do not diverge excessively. Moreover, we leverage the Maximum Mean Discrepancy (MMD) [44]

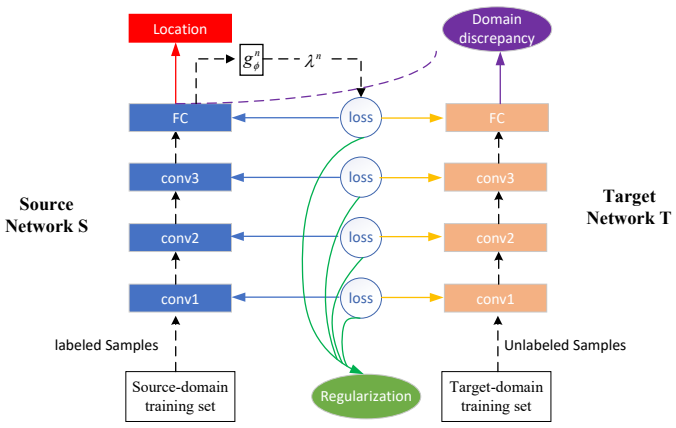

Fig. 6. Our dual-flow structure in the unsupervised domain adaptation method.

between the representations of the corresponding layers between two networks to enable us to reach the fact that, the two domains are related, despite their differences. In this context, we refrain from employing a domain classifier [45] despite its potential for superior results. This decision is motivated by the hard practical challenge of obtaining accurate domain labels, particularly those indicating environmental changes. MFFALoc relies solely on unlabeled data, devoid of class labels and domain labels, enabling MFFALoc to effectively handle unpredictable environmental changes without human intervention.

Let $X^S = \left\{X_i^S\right\}_{i=1}^{N^S}$ and $X^T = \left\{X_i^T\right\}_{i=1}^{N^T}$ denote the datasets from the source and target domains, respectively, $Y^S = \left\{y_i^S\right\}_{i=1}^{N^S}$ denote the source domain labels. Additionally, we use $\Theta^S = \left\{\theta_i^S\right\}$ and $\Theta^T = \left\{\theta_i^T\right\}$ to indicate the weights and bias parameters of all layers in the source and target networks.

1) Weights Regularization: Although we allow the weights of the relevant layers to differ between the two networks, we also want to prevent them from being too far apart in advance. This reflects the fact that the source and target domains are correlated, and prevents overfitting in the target flow when only a few samples are available. Therefore, we design a weight regularize r_w that represents the distance between the weights of two flows in a given layer. We can apply it directly to the difference between these weights, so it can be written as:

$$r_w\left(\theta_j^S, \theta_j^T\right) = \left\|\theta_j^S - \theta_j^T\right\|_2^2, j \in \Omega \tag{5}$$

where r_w denotes the loss between the corresponding layers of two flows, and it works only on the set Ω of layers whose parameters are not shared. *However, it does not take into account the different ranges of means and values in the two types of data, which affects the representation differences.*

To better represent this difference and introduce more flexibility, we propose to take into account the linear transformation between the weights. Then, we set our regularize as the exponential form, and write as:

$$r_w\left(\theta_j^S, \theta_j^T\right) = \exp(\left\|a_j\theta_j^S + b_j - \theta_j^T\right\|^2) - 1, j \in \Omega \tag{6}$$

a_j and b_j are hyperparameters of each layer $j \in \Omega$ and learned with all other network parameters during training.

In the domain adaptation task at hand, it is important to note that not all layer pairs of the source model may contribute significantly to learning the target task. To prioritize the utilization of useful layer pairs, we employ a weighted feature matching loss. This loss function enables us to emphasize specific layer pairs based on their relevance to the target task, the regularization can be expressed as:

$$L_w = \sum_{\Omega} \lambda^j r_w(\theta_j^S, \theta_j^T), j \in \Omega \tag{7}$$

where $\lambda^j \geq 0$ is the weight of transfer layer i with $\sum_{\Omega} \lambda^i = 1, i \in \Omega$. As the significance of the layers to be transferred can vary for each sample, we introduce a dynamic way by setting the layer weight as a function. Specifically, we utilize the softmax output of a small meta-network, which takes the characteristics of the source model as input:

$$\lambda = \{\lambda^i\} = g_\phi^i(\theta_i^S), i \in \Omega \tag{8}$$

where ϕ are the parameters of meta-network.

2) Domain Discrepancy: In addition to constraining the difference in weights of the corresponding layers in the two flows, we need to learn a domain-invariant final representation, *which is the feature before the output layer*. In this way, we can try to ensure that the estimated coordinates of the final output are as similar as possible in both flows. Our goal is to minimize the distance between the representations of the source and target domain by minimizing the Maximum Mean Discrepancy (MMD). It is one of the most widely used loss functions in Domain Adaptation and is primarily used to measure the distance between two different but related distributions. By mapping each sample to a Reproducing Kernel Hilbert Space (RKHS) and calculating the distance between the two means, the difference between the higher-order moments of the distributions is obtained and its maximum value is used as the distance measure of the two distributions. In our case, let $w_i^S = f(\Theta^T, x_i^S)$ and $w_i^T = f(\Theta^T, x_i^T)$ be the feature representations of the last layer in the source and target streams, respectively. Then, the squared MMD can be written as:

$$MMD^2(\Phi, S, T) = \left\| \frac{1}{N^S}\sum_{i=1}^{N^S}\varphi(w_i^S) - \frac{1}{N^T}\sum_{i=1}^{N^T}\varphi(w_i^T) \right\|_H^2 \tag{9}$$

where $\varphi(\cdot)$ is the mapping function. Simplifying Eq. (4) and replacing the inner product using the standard RBF kernel $k(u, v) = e^{-\|u-v\|^2/\sigma}, \sigma = 1$, finally obtain the rewriting squared MMD, and thus our domain discrepancy regularization is:

$$\begin{aligned} L_D = r_d(\Theta^S, \Theta^T | X^S, X^T) = \sum_{i,i'} \frac{k(w_i^S, w_{i'}^S)}{N^S * N^S} - \\ 2\sum_{i,j} \frac{k(w_i^S, w_j^T)}{N^S * N^T} + \sum_{j,j'} \frac{k(w_j^T, w_{j'}^T)}{N^T * N^T} \end{aligned} \tag{10}$$

Finally, we can obtain a total loss function L to train the parameters Θ^S and Θ^T of the source and target flow, and it can be expressed as:

$$\left\{\Theta_*^S, \Theta_*^T\right\} \leftarrow argmin_{\Theta^S, \Theta^T} L \tag{11}$$

$$L = L_S + L_W + L_D \tag{12}$$

where L_S is the standard MSE losses of the source flow.

In order to learn the target flow model parameters, *we first initialize the weights of the target flow using the source weights trained in the offline training stage.* Specifically, the parameters a_j and b_j for each layer $j \in \Omega$ are initialized to $a_j = 1$ and $b_j = 0$, representing a constant transformation. These parameters are then learned jointly via backpropagation with the target flow network. The meta-networks consist of 1-layer fully-connected networks, specifically designed for each layer pair between the source and target network. They take the intermediate features of the i-th layer of the source network as input and aim to learn the corresponding λ^i as outputs. To ensure the validity of λ^i, softmax functions are employed, satisfying the constraint $\sum_\Omega \lambda^i = 1, i \in \Omega$. Subsequently, all parameters of both flows are optimized jointly by minimizing the loss defined in Eqs. (11), (12) via backpropagation using Adam with a learning rate of 0.001. Considering that our meta-network has a limited influence on L, primarily through the regularization term L_W, therefore, updating ϕ using the gradient $\nabla_\theta L$ can be challenging. To overcome this limitation, we suggest the following training scheme:

1) Minimize L_W by updating Θ^S, Θ^T once.
2) Minimize L_D by updating Θ^S, Θ^T once.
3) Calculate L and minimize it by updating ϕ.

Algorithm 1 shows the specific process of the training programme for the two-stream structure.

Algorithm 1 Dual-flow Structure Training Process

Require: Training set: The fingerprint database D_T, The unlabelled test set D_S, The set Ω of layers whose parameters are not shared, Learning rate η.

Initialize hyperparameters a_j and b_j for each layer $j \in \Omega$ to 1 and 0, respectively.

repeat

Random choose a batch $\mathcal{B}_S \subset D_S$ and another batch $\mathcal{B}_T \subset D_T$ both with quantity B, and input them into the dual-flow structure Θ^S for $\mathcal{B}_S$, and Θ^T for $\mathcal{B}_T$, respectively.

Update θ^S by using $\nabla_{\theta^S} \frac{1}{B} \sum_{(x,y)\in\mathcal{B}_S,\mathcal{B}_T} \mathcal{L}_w \left(\theta^S | x, y, \phi\right)$

Update θ^T by using $\nabla_{\theta^T} \frac{1}{B} \sum_{(x,y)\in\mathcal{B}_S,\mathcal{B}_T} \mathcal{L}_w \left(\theta^T | x, y, \phi\right)$

Update a_j and b_j by using $\nabla_{a_j,b_j} \frac{1}{B} \sum_{(x,y)\in\mathcal{B}_S,\mathcal{B}_T} \mathcal{L}_w \left(\theta^T | x, y, \phi\right)$

Update θ^S by using $\nabla_{\theta^S} \frac{1}{B} \sum_{(x,y)\in\mathcal{B}_S,\mathcal{B}_T} \mathcal{L}_D \left(\theta^S | x, y, \phi\right)$

Update θ^T by using $\nabla_{\theta^T} \frac{1}{B} \sum_{(x,y)\in\mathcal{B}_S,\mathcal{B}_T} \mathcal{L}_D \left(\theta^T | x, y, \phi\right)$

Update ϕ by using $\nabla_{\phi} \frac{1}{B} \sum_{(x,y)\in\mathcal{B}_S,\mathcal{B}_T} \mathcal{L} \left(\phi | x, y, \theta^S, \theta^T\right)$

until done

V. Performance and Evaluation

This section presents the experimental settings and methodology employed in this study, followed by a detailed discussion on the performance evaluation and analysis of the proposed MFFALoc system. Furthermore, we implemented a prototype of the MFFALoc system using commercial WiFi devices and conducted extensive experiments in a dynamic indoor environment. To ensure thorough evaluation, we continuously assessed the localization performance for a period exceeding 6 days.

A. Experiment Evaluation

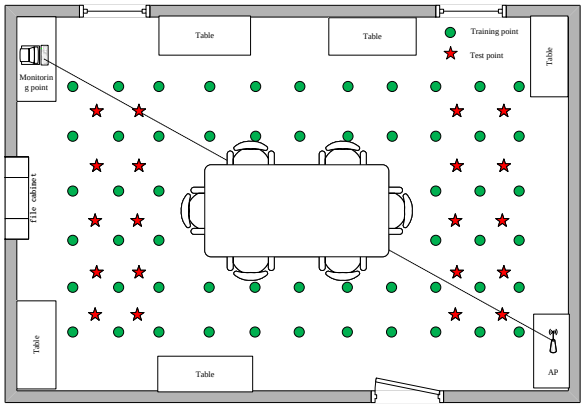

Fig. 7. Layout of the meeting room for training/test positions

1) Environment Devices and Data Collection: For evaluating the MFFALoc system, we utilized a TP-LINK WR841N router as a transmitter to continuously transmit wireless signals, while a laptop running Ubuntu 14.04 OS and equipped with an Intel 5300 NIC served as the receiver for capturing raw CSI measurements. The experiments were conducted in a typical conference room, which served as the indoor environment. This scenario was consistent with our previous work [29]. The conference room had a total area of 120 m^2 and featured a large conference table at the center, surrounded by stools and cabinets. In this setup, we evenly distributed 56 training points (RPs) and 20 test points (TPs) throughout the open space, as depicted in Fig. 7. To collect the CSI information, we used CSI Tools to continuously capture CSI measurements while the target was positioned at each training point. The TP-LINK router transmitted data at a rate of 100 packets per second, utilizing a single antenna. On the other hand, the Intel 5300 NIC had three receive antennas, each capable of capturing 30 subcarriers. Consequently, it was possible to obtain 30×3 CSI measurements per packet. During the data collection phase, we gathered approximately 15000 CSI measurements, equivalent to 150 seconds of CSI information per training point and 1500 CSI measurements, equivalent to 15 seconds of CSI information per test point. During the data acquisition process, the volunteer had the freedom to move naturally, engaging in actions such as turning, stretching, and crossing arms, as long as they remained at one of the designated training or testing points. The hardware processing speed limited the transmit rate to 100 packets per second. Due to hardware limitations, it is important to note that attempting to transmit data at higher rates may result in packets capturing fewer than 30 subcarriers of CSI measurements. Upon conversion into high-resolution

fingerprint features, we obtained 500 fingerprint matrices for each training point and 50 fingerprint matrices for each test point.

2) Environment Changes: To investigate the impact of environmental changes, we conducted a 6-day experiment. On the first day, we collected CSI measurements from all training and test points according to the aforementioned data collection setup. Subsequently, we intentionally introduced various common daily environmental changes, including actions such as opening and closing doors, repositioning chairs, and even moving cabinets. These changes were implemented as part of the normal usage of the conference room, and we collected CSI measurements from all test points accordingly. It is crucial to emphasize that we did not label these environmental changes. Therefore, all the collected data, including the CSI measurements, lack both class labels and domain labels.

To comprehensively assess the performance of our MFFLoc system in real-world scenarios, we conducted numerous performance evaluation tests on its design. The training dataset employed for each experiment, the target-domain unlabelled training dataset utilized for the unsupervised domain adaptation approach, and the corresponding test dataset are meticulously presented and described in detail at the outset of each subsection.

B. Comparative Evaluated Systems

To avoid be mixing apples and oranges, we exclusively evaluated device-free passive fingerprinting localization systems that do not require any labels. In order to provide a fair comparison with state-of-the-art work, we made appropriate modifications to some of the existing systems.

1) MFFALoc is our system proposed in this work.
2) Fidora [31] is the latest work that builds upon existing methods utilizing unsupervised domain adaptation structures. It incorporates both a data enhancer and a domain-adaptive classifier, enabling it to adapt to new data samples and consider variations in user body size. To ensure a fair comparison, we adjusted the final output position of Fidora from classification to regression, aligning it with the same approach used in MFFALoc.
3) MSDFL [38] is a notable example of existing fingerprint update methods. It does not rely on class labels or domain labels as input. Instead, it utilizes an empty localization region as an indication of a domain change. Once an empty room state is detected, all subsequent data collected is automatically labeled with the new domain label. To update the CSI fingerprint database, MSDFL employs a polynomial mapping function.
4) CiFi [22] is a representative example of a CNN-based localization system. It leverages the phase data from CSI to estimate the Angle of Arrival (AoA)of wireless signals. The estimated AoA image is subsequently utilized as input to an offline deep CNN, where the network weights are trained. To ensure a fair performance comparison, we employ the same CNN architecture with the MFFALoc. CiFi serves as a benchmark to evaluate and compare the performance of various techniques in processing and addressing the challenges posed by fingerprint variations.

C. Localization Performance Evaluation and Discussion

To assess the effectiveness of the proposed system in terms of robust localization under daily environmental variations, we conducted thorough evaluations of the localization performance of MFFALoc and other comparative systems. Additionally, we evaluated the robustness of MFFALoc under various indoor environment variations. Through extensive experimentation and analysis, MFFALoc has demonstrated superior performance compared to state-of-the-art works in predicting target locations, particularly in complex indoor scenarios. The detailed performance analysis of the MFFALoc system is provided below.

1) Overall Performance Evaluation: Initially, we evaluate the performance of the proposed MFFALoc system using the CSI measurements collected from all the test points over a period of 6 days. In this experiment, we utilized the Channel State Information (CSI) measurements of the training points collected on the first day as the source domain training data. Additionally, the CSI measurements of the test points collected over the course of the six days were employed as the target domain unlabelled training data in the domain adaptation method, as well as the test data. The ratio between the target domain training data and test data was 3:7 respectively.

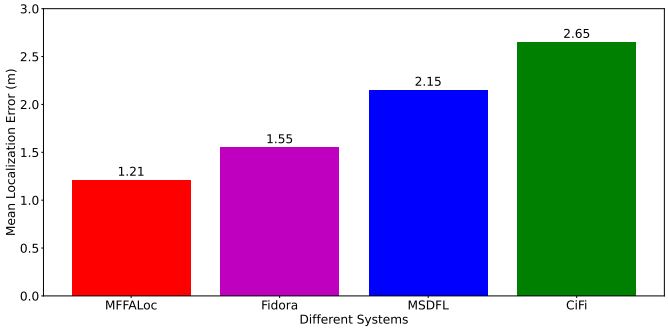

Fig. 8. The Mean Localization Error for each system

Fig . 8 illustrates the mean localization error for each system using the CSI measurements collected from all the test points over a period of 6 days. The MFFALoc system achieves the mean localization error of 1.21m under various indoor environmental variations. This represents an significantly improvement of 21.9% and 43.7% in accuracy compared to the mean localization errors of 1.55m for Fidora and 2.15m for MSDFL, respectively. And the CiFi system struggles to handle numerous environmental changes, resulting in the worst mean localization error of 2.65m.

Fig . 9 presents the cumulative distribution functions (CDFs) of the location estimation errors, which represent the bias between the localization result and the ground truth. It is evident from the plot that the 90th percentile localization accuracy of MFFALoc is approximately 2.1m, which is 19.2% higher than Fidora's 90th percentile accuracy of around 2.6m, and 40% higher than MSDFL's 90th percentile accuracy of 3.5m.

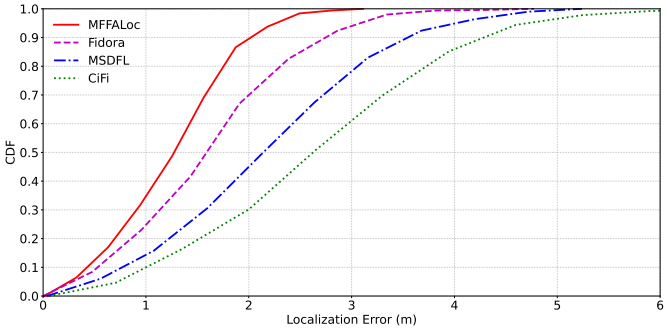

Fig. 9. The CDFs for each system

Furthermore, the max localization error of the MFFALoc system is also much lower than other systems.

The localization results of the MFFALoc system can be attributed to the following key factors:

1) Data Preprocessing: The MFFALoc system utilizes a data preprocessing method that combines both amplitude and phase information from CSI. This approach allows for the generation of high-resolution fused fingerprint features, enhancing the richness and separability of the fingerprints. In contrast, both MSDFL and Fidora only utilize the amplitude information of CSI, limiting the information available for localization.
2) Regression Model: In the current experiment, MFFALoc employs a regression model to establish the relationship between the fingerprint features and the position coordinates for estimating the target position. On the other hand, MSDFL and Fidora utilize a classification model for target position estimation at the initial stage of their designs. The regression model used in MFFALoc offers flexibility and adaptability in estimating target locations when the test and training points are not identical.
3) Unsupervised Domain Adaptation: The unsupervised domain adaptation method employed by MFFALoc incorporates explicit simulation of domain offsets through a dual-stream structure and a meta-learning network. This approach reduces training complexity and enhances the migration effect compared to other similar methods. By effectively adapting the localization model to changes in fingerprint features, MFFALoc achieves improved knowledge transfer-assisted localization.

2) Short-term Performance Evaluation: To assess the robustness of the MFFALoc system under varying indoor environmental dynamics, we conducted an investigation using CSI data from test points at different times on six consecutive days: day one, day two, day three, day four, day five, and day six. Specifically, in this experiment, we utilize the CSI measurements obtained from all training points on the initial day as the source domain training data. Simultaneously, the CSI measurements from the test points, collected each day, serve as both the target domain unlabeled training data and the test data for the domain adaptation method. This allocation is established in a 3:7 ratio, respectively. For short-term localization performance comparison, we evaluate the location errors of the MFFALoc system in Fig. 10.

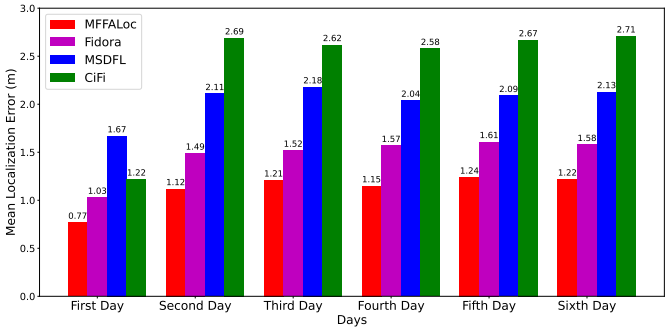

Fig. 10. Short-term Performance Comparison

Upon analyzing the results, we observe that when the environment remains unchanged on the first day, and the fingerprint features of the training points and test points still belong to the same distribution, the mean localization error of MFFALoc is 0.77m. This represents an improvement of approximately 25.2% compared to Fidora's 1.03m, about 53.9% compared to MSDFL's 1.67m, and about 36.9% compared to CiFi's 1.22m. In the subsequent days when the indoor environment undergoes changes, MFFALoc, Fidora, and MSDFL achieve mean localization errors of approximately 1.2m, 1.5m, and 2.1m, respectively. This indicates a relatively stable performance for these systems. However, CiFi exhibits a rapid decline to about 2.7m in localization accuracy due to its inability to handle changes in fingerprint features effectively. These experimental results can verify that the MFFALoc system obtains better localization performance with the robustness of short-term environmental changes.

D. Impact Analysis

1) Impact of Data Pre-processing on System Performance: To assess the efficacy of the data pre-processing in the MFFALoc system, we carried out an evaluation of the system's localization performance using solely amplitude information and solely phase information, respectively. This assessment was conducted independently of the multimodal fusion representation. In this experiment, we utilized the Channel State Information (CSI) measurements of the training points collected on the first day as the source domain training data. Additionally, the CSI measurements of the test points collected over the course of the six days were employed as the target domain unlabelled training data in the domain adaptation method, as well as the test data. The ratio between the target domain training data and test data was 3:7 respectively. Fig. 11 illustrates the overall localization performance of the MFFALoc system under different pre-processing schemes, namely the 50th percentile, 90th percentile, and mean localization accuracies.

Upon examination, it is evident that the MFFALoc system, when using a pure amplitude data preprocessing scheme, exhibits an average positioning error of 1.49m, which is relatively coarser compared to the other schemes. It also demonstrates 50th percentile and 90th percentile accuracies of 1.5m

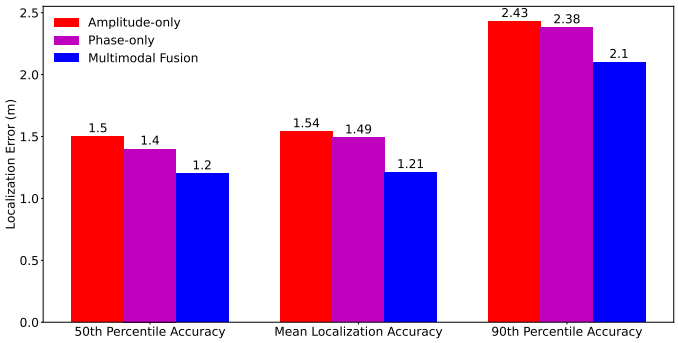

Fig. 11. Overall Localization Performance with Different Pre-processing Schemes

and 2.43m, respectively, which are the lowest among the three schemes. This result may be attributed to the fact that while the amlitude does provide information about the target's position, the calibrated phase is more sensitive to the location, as confirmed in our previous work [39]. Consequently, the MFFALoc system employs a pure phase data preprocessing scheme, resulting in an average positioning error of 1.49m, with 50th percentile and 90th percentile accuracies of 1.4m and 2.38m, respectively, slightly higher than the amplitude-only scheme. The comprehensive version of MFFALoc achieves the best performance, with a mean localization error of 1.21m and 50th percentile and 90th percentile accuracies of 1.2m and 2.1m, respectively. These results strongly validate the effectiveness of our data pre-processing approach in addressing Challenge 1 by combining temporal and spatial information from both amplitude and phase, thereby generating fused fingerprint features with enhanced richness and separability.

2) Impact of Localization Model Structure on System Performance: Since we utilize a deep convolutional neural network as the localization model, its structure has a significant impact on the performance of the MFFALoc system. Therefore, we designed five convolutional deep neural networks with different structures to study their effects on the system performance, and their respective structures are shown in Table I below. Given that the primary objective of the localization model is to accurately locate the same distribution in both training and test data, we employ the CSI measurements obtained from all training points on the initial day as the training data. Correspondingly, the CSI measurements from all test points collected on the first day are used as the test data in this experiment. We started from the most basic **A** structure and gradually replaced the large 5×5 convolutional kernels with small 3×3 convolutional kernels and increased the number of channels per convolutional layer, use the leakyRelu activation function, and remove the pooling layer, as illustrated in Table I.

Fig. 12 presents the mean localization error of the MFFALoc system when employing five different structural localization models on the first day. This analysis was conducted when the CSI fingerprint features of both the training and test points remained unchanged. It can be seen that the localization performance of the MFFALoc system gradually increases

TABLE I
DIFFERENT LOCALIZATION MODEL STRUCTURE

ConvNet Configuration				
A	**B**	**C**	**D**	**E**
Input layer ($3 \times 30 \times 30$)				
Conv5-32	Conv3-32	Conv3-32	Conv3-32	Conv3-64
Conv5-64	Conv5-64	Conv3-64	Conv3-64	Conv3-128
Conv5-128	Conv5-128	Conv5-128	Conv3-128	Conv3-256
FC-120				
Output layer				

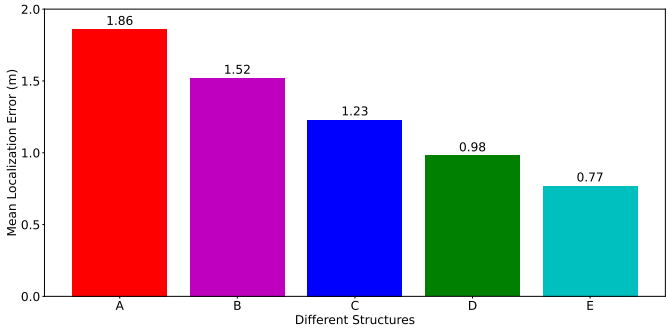

Fig. 12. Mean Localization Errors with Different CNN Structure

from structure **A** to structure **E**, and the MFFALoc system utilizing the localization model with structure **E** has the best performance and the most stable localization error. Therefore, we set the positioning model architecture to structure **E**, which is the most suitable architecture for the MFFALoc system.

3) Impact of Unsupervised Domain Adaptation Scheme on System Performance: While we have partially validated the effectiveness of our unsupervised domain adaptation scheme in the previous comparative analysis of system performance, we sought to conduct additional ablation experiments. These experiments aimed to provide further assessment of the effectiveness of the unsupervised domain adaptation scheme implemented in the MFFALoc system. In this experiment, we utilized the Channel State Information (CSI) measurements of the training points collected on the first day as the source domain training data. Additionally, the CSI measurements of the test points collected over the course of the six days were employed as the target domain unlabelled training data in the domain adaptation method, as well as the test data. The ratio between the target domain training data and test data was 3:7 respectively. Fig. 13 presents the overall localization performance, comparing the MFFALoc system with and without the unsupervised domain adaptation scheme.

The MFFALoc system without the unsupervised domain adaptation scheme demonstrated an average localization error of 2.11 m, a 50th percentile accuracy of 2.09 m, and a 90th percentile accuracy of 3.26 m. This represents a noteworthy reduction in localization error and performance compared to the full-fledged MFFALoc system. This outcome can be attributed to the utilization of all 6 days of test data, which exacerbates the fingerprint inconsistency issue, as discussed in Challenge 2. These findings provide strong validation for the effectiveness of the proposed unsupervised domain adaptation

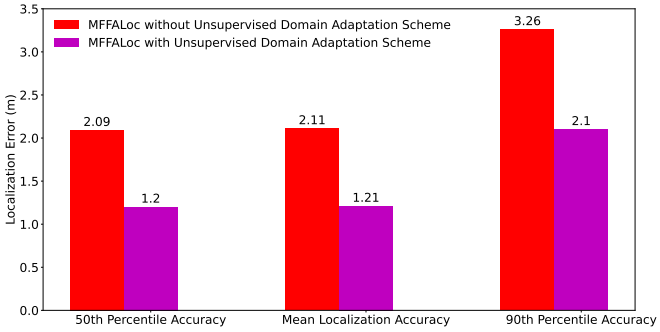

Fig. 13. Overall Localization Performance with and without Unsupervised Domain Adaptation Scheme

scheme in addressing Challenge 2. It can achieve satisfactory localization accuracy without the need for any human intervention, even without the need to be aware of the presence of fingerprint inconsistencies due to environmental changes.

4) Impact of Wavelet and Threshold Selection on System Performance in Wavelet Domain Denoising: To assess the influence of wavelet and threshold selection in wavelet domain denoising on the MFFALoc system's performance, we conducted an evaluation using various commonly employed wavelets and threshold coefficients. Given that their impact on system performance remains consistent across fingerprint variations, in this experiment, we utilized the Channel State Information (CSI) measurements from the training points collected on the first day as the training data and the CSI from the test points as the test data. The selection of wavelets and thresholds was easily implemented using the PyWavelets Python third-party extension library. Table II presents the average localization error of the MFFALoc system for different combinations of wavelet and threshold coefficients used in the Wavelet Domain Denoising (WDD) process.

TABLE II
THE MEAN LOCALIZATION ERRORS OF MFFALOC SYSTEM WITH DIFFERENT COMBINATIONS OF WAVELET AND THRESHOLD COEFFICIENTS

	Threshold coefficients			
Wavelets	VisuShrink	SureShrink	HeurSure	MinMax
Harr	0.81	0.83	0.82	0.80
Daubechies(db3)	0.78	0.79	0.81	0.77
Symlets(sym2)	0.78	0.80	0.82	0.79
Coiflets(coif3)	0.84	0.82	0.79	0.81

From Table II, it is easy to see that for different combinations of wavelet and threshold coefficients, the localisation performance of the MFFALoc system is extremely close to each other, and basically the difference is not very large. This may be due to the fact that the wavelet domain denoising is only the first step of our data preprocessing, and in the subsequent multimodal fusion characterisation module, the amplitude and phase data of the CSI are fused to characterise the fingerprints, and more discriminative fingerprint features are obtained. And this transformational fusion reduces the impact of wavelet domain denoising on the system performance. Of course, in practice, we still chose the combination of the Daubechies Wavelet base, which obtains the lowest average error, and the minimum-maximum threshold coefficients to be applied in the Wavelet Domain Denoising (WDD) process.

VI. CONCLUSION AND FUTURE WORK

In this paper, we introduced MFFALoc, a CSI-based multi-feature fusion adaptive device-free passive fingerprinting localization system that enables ubiquitous (sub)meter-level location information of device-free users with minimal hardware deployment requirements. MFFALoc addresses several challenges such as poor fingerprint richness, identifiability, and inconsistency of fingerprint features across different environments. It achieves this through a multimodal feature fusion representation and an unsupervised domain adaptor with a two-flow structure. Through extensive experiments, we demonstrated that MFFALoc surpasses existing techniques when dealing with unknown environmental variations. It not only provides sub-meter localization resolution with significantly improved accuracy but also exhibits robust performance under daily environmental variations. For future research, we aim to explore the following directions:

1) Developing a more automated method for collecting fingerprints of training points to reduce user effort and enhance the ease of use of fingerprint-based systems.
2) Addressing the challenge of serving multiple users simultaneously within the same region of interest in device-free passive localization systems. This can be achieved through deep fusion of multiple sensing anchors/modalities and multi-label learning approaches.
3) Exploring advanced machine learning tools such as Generative Adversarial Networks (GANs) for data augmentation and other domain adaptive and style-shifting algorithms customized for indoor localization scenarios.

ACKNOWLEDGMENTS

This work was supported in part by the National Natural Science Foundation of China (NSFC) under Grant (no.61967010, 62067003, 62062040, 62363015), in part by the Outstanding Youth Project of Jiangxi Natural Science Foundation under Grant No. 20212ACB212003, in part by the Jiangxi Province Key Subject Academic and Technical Leader Funding Project under Grant No. 20212BCJ23017, in part by the Science and Technology Research Project of Jiangxi Provincial Department of Education under Grant no. GJJ200333.

REFERENCES

[1] P. S. Farahsari, A. Farahzadi, J. Rezazadeh, and A. Bagheri, "A survey on indoor positioning systems for iot-based applications," *IEEE Internet Things J.*, vol. 9, no. 10, pp. 7680–7699, 2022.
[2] P. Roy and C. Chowdhury, "A survey on ubiquitous wifi-based indoor localization system for smartphone users from implementation perspectives," *ACM Comput. Surv.*, vol. 4, no. 3, pp. 298–318, 2022.
[3] Y. Ma, G. Zhou, and S. Wang, "Wifi sensing with channel state information: A survey," *ACM Comput. Surv.*, vol. 52, no. 3, pp. 1–36, 2019.
[4] K. Ngamakeur, S. Yongchareon, J. Yu, and S. U. Rehman, "A survey on device-free indoor localization and tracking in the multi-resident environment," *ACM Comput. Surv.*, vol. 53, no. 4, pp. 1–29, 2020.

[5] L. Zhao, H. Huang, X. Li, S. Ding, H. Zhao, and Z. Han, "An accurate and robust approach of device-free localization with convolutional autoencoder," *IEEE Internet Things J.*, vol. 6, no. 3, pp. 5825–5840, 2019.

[6] K. Wang, X. Yu, Q. Xiong, Q. Zhu, W. Lu, Y. Huang, and L. Zhao, "Learning to improve wlan indoor positioning accuracy based on dbscan-krf algorithm from rss fingerprint data," *IEEE Access*, vol. 7, pp. 72 308–72 315, 2019.

[7] Y. Duan, K.-Y. Lam, V. C. Lee, W. Nie, K. Liu, H. Li, and C. J. Xue, "Data rate fingerprinting: A wlan-based indoor positioning technique for passive localization," *IEEE Sens. J.*, vol. 19, no. 15, pp. 6517–6529, 2019.

[8] S. S. Moosavi and P. Fortier, "Fingerprinting localization method based on clustering and gaussian process regression in distributed massive mimo systems," in *IEEE Int. Symp. Person Indoor Mobile Radio Commun., PIMRC*. IEEE, 2020, pp. 1–7.

[9] D. Halperin, W. Hu, A. Sheth, and D. Wetherall, "Tool release: Gathering 802.11 n traces with channel state information," *ACM SIGCOMM Compu. Commun. Rev.*, vol. 41, no. 1, pp. 53–53, 2011.

[10] D. Chizhik, J. Ling, P. W. Wolniansky, R. A. Valenzuela, N. Costa, and K. Huber, "Multiple-input-multiple-output measurements and modeling in manhattan," *IEEE J. Sel. Areas Commun.*, vol. 21, no. 3, pp. 321–331, 2003.

[11] Z. Yang, Z. Zhou, and Y. Liu, "From rssi to csi: Indoor localization via channel response," *ACM Comput. Surv. (CSUR)*, vol. 46, no. 2, pp. 1–32, 2013.

[12] W. Liu, Q. Cheng, Z. Deng, H. Chen, X. Fu, X. Zheng, S. Zheng, C. Chen, and S. Wang, "Survey on csi-based indoor positioning systems and recent advances," in *Int. Conf. Indoor Position. Indoor Navig., IPIN*. IEEE, 2019, pp. 1–8.

[13] D. Wu, Y. Zeng, F. Zhang, and D. Zhang, "Wifi csi-based device-free sensing: from fresnel zone model to csi-ratio model," *CCF Trans. Pervasive Comput. Interact.*, pp. 1–15, 2022.

[14] Z. Tian, C. Ye, and Y. Jin, "Device-free indoor tracking via joint estimation of dfs and aoa using csi amplitude," in *Int. Conf. Microw. Millim. Wave Technol., ICMMT - Proc.* IEEE, 2021, pp. 1–3.

[15] D. Yu, Y. Guo, N. Li, and M. Wang, "Sa-m-sbl: An algorithm for csi-based device-free localization with faulty prior information," *IEEE Access*, vol. 7, pp. 61 831–61 839, 2019.

[16] Y. Liu and G. Li, "Device-free indoor localization of csi based on limited penetrable horizontal visibility graph," *IEEE Access*, vol. 10, pp. 71 120–71 132, 2022.

[17] W. Wei, J. Yan, X. Wu, C. Wang, and G. Zhang, "Csi fingerprinting for device-free localization: Phase calibration and ssim-based augmentation," *IEEE Wireless Commun. Lett.*, vol. 11, no. 6, pp. 1137–1141, 2022.

[18] Z. Han, L. Lin, Z. Wang, Z. Lian, C. Qiu, H. Huang, L. Zhao, and C. Su, "Cnn-based attack defense for device-free localization," *Mobile Inf. Syst.*, vol. 2022, 2022.

[19] J. Zhang, Y. Li, and W. Xiao, "Integrated multiple kernel learning for device-free localization in cluttered environments using spatiotemporal information," *IEEE Internet Things J.*, vol. 8, no. 6, pp. 4749–4761, 2020.

[20] J. Yan, L. Wan, W. Wei, X. Wu, W.-P. Zhu, and D. P.-K. Lun, "Device-free activity detection and wireless localization based on cnn using channel state information measurement," *IEEE Sens. J.*, vol. 21, no. 21, pp. 24 482–24 494, 2021.

[21] J. Zhang, W. Xiao, and Y. Li, "Data and knowledge twin driven integration for large-scale device-free localization," *IEEE Internet Things J.*, vol. 8, no. 1, pp. 320–331, 2020.

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

[23] Z. Li and X. Rao, "Toward long-term effective and robust device-free indoor localization via channel state information," *IEEE Internet Things J.*, vol. 9, no. 5, pp. 3599–3611, 2021.

[24] H. Zou, M. Jin, H. Jiang, L. Xie, and C. J. Spanos, "Winips: Wifi-based non-intrusive indoor positioning system with online radio map construction and adaptation," *IEEE Trans. Wireless Commun.*, vol. 16, no. 12, pp. 8118–8130, 2017.

[25] C. Wu, J. Xu, Z. Yang, N. D. Lane, and Z. Yin, "Gain without pain: Accurate wifi-based localization using fingerprint spatial gradient," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 1, no. 2, pp. 1–19, 2017.

[26] X. Guo, L. Li, F. Xu, and N. Ansari, "Expectation maximization indoor localization utilizing supporting set for internet of things," *IEEE Internet Things J.*, vol. 6, no. 2, pp. 2573–2582, 2018.

[27] A. Rai, K. K. Chintalapudi, V. N. Padmanabhan, and R. Sen, "Zee: Zero-effort crowdsourcing for indoor localization," in *Proc. Annu. Int. Conf. Mobile Comput. Networking, MobiCom'12*, 2012, pp. 293–304.

[28] S.-h. Jung, B.-c. Moon, and D. Han, "Unsupervised learning for crowdsourced indoor localization in wireless networks," *IEEE Trans. Mob. Comput.*, vol. 15, no. 11, pp. 2892–2906, 2015.

[29] Y. Zhao, W.-C. Wong, T. Feng, and H. K. Garg, "Calibration-free indoor positioning using crowdsourced data and multidimensional scaling," *IEEE Trans. Wireless Commun.*, vol. 19, no. 3, pp. 1770–1785, 2019.

[30] Q. Wan, X. Duan, Y. Yu, R. Chen, and L. Chen, "Self-calibrated multi-floor localization based on wi-fi ranging/crowdsourced fingerprinting and low-cost sensors," *Remote Sens.*, vol. 14, no. 21, p. 5376, 2022.

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

[32] L. Li, X. Guo, Y. Zhang, N. Ansari, and H. Li, "Long short-term indoor positioning system via evolving knowledge transfer," *IEEE Trans. Wireless Commun.*, vol. 21, no. 7, pp. 5556–5572, 2022.

[33] B.-J. Chen and R. Y. Chang, "Few-shot transfer learning for device-free fingerprinting indoor localization," in *ICC 2022-IEEE Int. Conf. Commun.* IEEE, 2022, pp. 4631–4636.

[34] K. M. Chen and R. Y. Chang, "Semi-supervised learning with gans for device-free fingerprinting indoor localization," in *IEEE Glob. Commun. Conf., GLOBECOM 2020*. IEEE, 2020, pp. 1–6.

[35] Z. E. Khatab, A. H. Gazestani, S. A. Ghorashi, and M. Ghavami, "A fingerprint technique for indoor localization using autoencoder based semi-supervised deep extreme learning machine," *Signal Process.*, vol. 181, p. 107915, 2021.

[36] R. Zhou, H. Hou, Z. Gong, Z. Chen, K. Tang, and B. Zhou, "Adaptive device-free localization in dynamic environments through adaptive neural networks," *IEEE Sens. J.*, vol. 21, no. 1, pp. 548–559, 2020.

[37] Y. Xie, Z. Li, and M. Li, "Precise power delay profiling with commodity wi-fi," *IEEE Trans. Mobile Comput.*, vol. 18, no. 6, pp. 1342–1355, 2019.

[38] X. Rao and Z. Li, "Msdfl: a robust minimal hardware low-cost device-free wlan localization system," *Neural Computing and Applications*, vol. 31, pp. 9261–9278, 2019.

[39] X. Rao, Z. Li, Y. Yang, and S. Wang, "Dfphasefl: a robust device-free passive fingerprinting wireless localization system using csi phase information," *Neural Computing and Applications*, vol. 32, pp. 14 909–14 927, 2020.

[40] X. Dang, X. Tang, Z. Hao, and Y. Liu, "A device-free indoor localization method using csi with wi-fi signals," *Sensors*, vol. 19, no. 14, p. 3233, 2019.

[41] Y. Zhuo, H. Zhu, and H. Xue, "Identifying a new non-linear csi phase measurement error with commodity wifi devices," in *Proc. Int. Conf. Parallel. Distrib. Syst. ICPADS*, 2016, pp. 72–79.

[42] W. G. of the 802 Committee *et al.*, "Ieee standard for information technology–telecommunications and information exchange between systems–local and metropolitan area networks–specific requirements part 11: Wireless lan medium access control (mac) and physical layer (phy) specifications amendment 5: Enhancements for higher throughput," *IEEE Std*, vol. 802, 2009.

[43] S. Sabour, N. Frosst, and G. E. Hinton, "Dynamic routing between capsules," in *Advances Neural Inf. Process Systems, NIPS*, vol. 30, 2017.

[44] M. Long, Y. Cao, J. Wang, and M. Jordan, "Learning transferable features with deep adaptation networks," in *Int. Conf. Mach. Learn., ICML*. PMLR, 2015, pp. 97–105.

[45] Y. Ganin, E. Ustinova, H. Ajakan, P. Germain, H. Larochelle, F. Laviolette, M. Marchand, and V. Lempitsky, "Domain-adversarial training of neural networks," *J. Mach. Learn. Res.*, vol. 17, no. 1, pp. 2096–2030, 2016.

Xinping Rao received the B.S., and M.S. degrees in electronic engineering from Xi an University of Posts and Telecommunications, Xi an, China, in 2012 and 2015, and the Ph.D. degree from the School of Mechano-electronic Engineering, Xidian University, in 2020, respectively. He is currently a Lecturer with the School of Software, Jiangxi Normal University, China. His research interests include device-free wireless localization and recognition, wireless networks, machine learning, and transfer learning. He has served as the Technical Reviewer for several journals, including the IEEE Internet of Things Journal, Neural Computing and Applications, IEEE Sensors Journal, Wireless Personal Communications, Computer Communications, and KSII Transactions on Internet and Information Systems.

Zhenzhen Luo received the the M.S. degree inSchool of Computer Science and Techonologyfrom Wuhan University of Techonology, China, in2013,and the Ph.D. degree in Department of National Engineering Research Center for E-Learning(NERCEL) from Central China Normal University(CCNU),China, in 2018. She is currently a associateprofessor in school of software at Jiangxi NormalUniversity. Her research interests include patternrecognition,deep learning and affective computing.

Yong Luo received the B.S. degree in Communication Engineering from Nanchang University, China, in 2003, the M.S. degree in Communication and Information System from Nanchang University, China, in 2006, and the Ph.D. degree in Signal and Information Processing from Shanghai University, China, in 2019. He is currently a lecturer in school of software at Jiangxi Normal University. His research interests include pattern recognition, image processing, computer vision, and machine learning.

Yugen Yi received his Ph.D. degree in applied mathematics from Northeast Normal University, Changchun, in 2016. He is an associate professor with the School of Software, Jiangxi Normal University, Nanchang. His research interests include artificial intelligence, computer vision, and machine learning.

Gang Lei is currently an Associate Professor and the Vice Chairman of the School of Software, Jiangxi Normal University, China, where he is also serving as the Associate Director of the Academic Committee. His research interests include big data technology, computer network management, and information systems.

Yuanlong Cao received the B.S. degree in computer science and technology from Nanchang University, China, in 2006, the M.S. degree in software engineering from the Beijing University of Posts and Telecommunications (BUPT), Beijing, China, in 2008, and the Ph.D. degree in communication and information system from the Institute of Network Technology, BUPT, in 2014. He was an Intern/Software Engineer with BEA TTC, IBM CDL, and DT Research, Beijing, China, from 2007 to 2011. He is currently a Professor with the School of Software, Jiangxi Normal University, Nanchang, China. His research interests include multimedia communications, network security and next-generation Internet technology. He has served as the Technical Reviewer for several journals, including the IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS, IEEE TRANSACTIONS ON COGNITIVE COMMUNICATIONS AND NETWORKING, COMPUTER COMMUNICATIONS.