FDSense: Prior-Guided Cross-Room Domain Generalization in WiFi Sensing

Zhiyong Zheng[1,2,3], Ke Xu[1,2,3,*], Shijia Liu[4], and Jiangtao Wang[1,2,3,5,*]
[1]Suzhou Institute for Advanced Research of USTC, Suzhou, China
[2]School of AI and Data Science, University of Science and Technology of China, Hefei, China
[3]Suzhou Big Data & AI Research and Engineering Center, Suzhou, China
[4]Southwest Jiaotong University, The Institute of Smart City and Intelligent Transportation, Chengdu, China
[5]Shanghai Key Laboratory of Data Science, Shanghai, China
Email: zhiyongzheng@mail.ustc.edu.cn, {nick_xuke, wangjiangtao}@ustc.edu.cn, liushijia@swjtu.edu.cn

***Abstract*—WiFi sensing has become increasingly valuable in daily life by enabling the perception of human activities through the analysis of Channel State Information (CSI). However, its effectiveness is often hindered by domain shift. Existing approaches to mitigate domain shift either rely on target domain data for domain adaptation or overlook prior knowledge inherent to WiFi sensing. To address this gap, we propose FDSense, a novel framework that integrates signal processing with domain generalization guided by CSI priors. Specifically, we first employ signal processing techniques to denoise and standardize CSI signals. Next, we introduce a Selective Fourier Domain Mixing (SFDM) module, which encourages the model to capture domain-invariant features. Finally, deep neural networks are used for feature extraction and classification. Extensive cross-room experiments on a public dataset demonstrate that FDSense consistently outperforms state-of-the-art methods, highlighting its effectiveness and robustness.**

***Index Terms*—WiFi Sensing, Domain Generalization, Channel State Information, Fourier Domain**

I. Introduction

Enabled by IEEE 802.11n/ac standards, WiFi has evolved beyond communication to pervasive sensing via Channel State Information (CSI) [1]–[3]. Grounded in Orthogonal Frequency-Division Multiplexing (OFDM) and Multiple-Input Multiple-Output (MIMO) technologies, CSI forms a matrix of complex variables characterizing the channel frequency response. Human movements modulate these signals through multipath effects (e.g. reflection and scattering), embedding specific kinematic patterns into CSI measurements. Compared to vision-based modalities [4], WiFi sensing ensures privacy protection and robustness against illumination variations, and compared to wearable approaches [5], [6], it offers a non-invasive, device-free paradigm.

Benefiting from the powerful feature representation capabilities of deep neural networks, a series of deep learning-based approaches [7]–[11] have been proposed for WiFi sensing, achieving remarkable success. However, these methods frequently overlook the critical challenge of domain shift. Due to the inherent sensitivity of CSI, variations in the sensing environment inevitably lead to significant differences in data distribution.

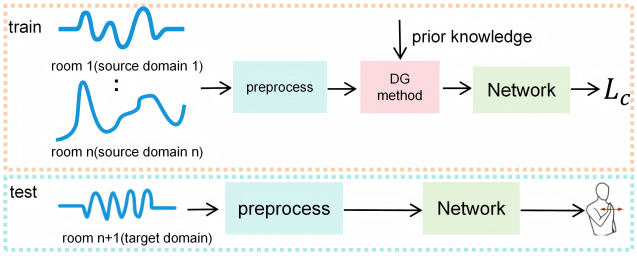

Fig. 1. Overview of our cross-domain setting and proposed solution.

Such distributional discrepancies are formally modeled as domain shift in machine learning contexts [12]–[14]. Consequently, this phenomenon causes severe performance degradation when a model trained in one scenario (source domain) is transferred to another (target domain), thereby substantially limiting the generalization capability and real-world scalability of WiFi sensing systems.

To mitigate the adverse effects of domain shift, numerous cross-domain studies have been conducted. Based on the availability of target domain data, these approaches are generally categorized into two paradigms: Unsupervised Domain Adaptation (UDA) and Domain Generalization (DG). UDA methods [15], [16] typically require access to unlabeled data from the target domain during the training phase. Consequently, they fail to fundamentally eliminate the dependence on target-specific information, restricting their applicability to completely unseen environments. Conversely, DG methods [13], [17] aims to learn domain-invariant sensing models leveraging exclusively source domain data, thereby removing the reliance on the target domain and offering a more practical solution for real-world deployment.

Capitalizing on the independence from target domain data, recent research has focused on applying DG paradigms to WiFi sensing. Generally, existing approaches can be categorized into two streams: model-based and data-based methods. Model-based methods [18] typically leverage physical domain

*The corresponding authors are Ke Xu (email: nick_xuke@ustc.edu.cn) and Jiangtao Wang (email: wangjiangtao@ustc.edu.cn)

knowledge to transform raw CSI into domain-invariant high-level features, subsequently employing deep neural networks as classifiers. However, these methods often impose stringent hardware prerequisites, thereby hindering their ubiquitous deployment in real-world scenarios. Conversely, data-based methods [12], [13], [19] utilize advanced machine learning paradigms to address domain shift by learning adaptive representations directly from data. Nevertheless, the features learned by these approaches often exhibit limited interpretability, which restricts their reliability in complex environments.

To address the aforementioned limitations, this work aims to design a novel data-based DG framework that effectively integrates domain-specific prior knowledge into the learning process. Our approach is grounded in a fundamental physical observation of WiFi sensing: domain-invariant action features are predominantly concentrated in low-frequency spectral regions, whereas domain-specific environmental noise resides primarily in high-frequency regions [20]. Leveraging this insight, we employ a spectral mixing strategy on the high-frequency components of samples that share the same action category but originate from distinct domains. By synthesizing virtual representations through high-frequency perturbation, this strategy simulates diverse sensing environments while strictly preserving semantic consistency. Consequently, the model is compelled to suppress domain-specific discrepancies and focus on robust, category-relevant features. Although inspired by Fourier Domain Adaptation (FDA) [21], our method distinctively targets generalization to unseen domains rather than adaptation to a known target. Furthermore, we introduce specific architectural refinements to accommodate the unique spectral distribution differences between CSI signals and visual images.

Beyond existing DG methods, we argue that domain generalization in WiFi sensing is not merely a machine learning optimization problem, but fundamentally intersects with signal processing principles. Therefore, in this paper, we propose a novel framework, FDSense, which synergizes signal processing priors with data-based machine learning techniques. Specifically, we first preprocess the raw CSI measurements to mitigate noise and standardize temporal-amplitude dimensions. Subsequently, we introduce a Selective Fourier Domain Mixing (SFDM) module, which strategically mixes spectral components of cross-domain samples within the same category to synthesize domain-invariant representations. Finally, a deep neural network is employed for robust feature extraction and classification.

Our main contributions are summarized as follows:

- We propose a novel framework, FDSense, which integrates signal processing techniques with domain generalization methods.
- Based on the prior knowledge of CSI in the frequency domain, we design an SFDM module that, while preserving the semantic information of the samples, encourages the model to focus more on domain-invariant factors, thereby improving its performance on unseen domains.
- Experiments on public WiFi sensing datasets demonstrate that FDSense outperforms existing WiFi sensing methods in cross-room tasks.

II. Related Work

A. WiFi Sensing Basics

In indoor environments, wireless signals interact with surrounding objects through reflection, diffraction, and scattering, resulting in multipath propagation. Modern WiFi devices, grounded in MIMO and OFDM technologies, capture these multipath characteristics in the form of CSI. Physically, the received signal is a superposition of multipath components, which can be principally decomposed into two parts: a static component derived from the Line-of-Sight (LoS) path and stationary objects, and a dynamic component modulated by moving human subjects. However, due to hardware imperfections, the measurements are inevitably distorted by random phase noise. Specifically, modern Network Interface Cards (NICs) report CSI as a matrix of complex numbers representing the Channel Frequency Response (CFR) across multiple orthogonal subcarriers. For a WiFi system equipped with N_{tx} transmit antennas and N_{rx} receive antennas operating over N_{sc} subcarriers, the collected CSI at a specific timestamp is typically organized as a complex tensor of dimensions $N_{tx} \times N_{rx} \times N_{sc}$. Each entry in this matrix captures the combined effect of signal attenuation and phase shift for a specific antenna pair and frequency, providing a fine-grained description of the scattering environment.

B. Cross Domain WiFi Sensing

Existing cross-domain WiFi sensing approaches can be broadly categorized into two paradigms: Unsupervised Domain Adaptation (UDA) and Domain Generalization (DG).

UDA-based methods typically operate on the assumption that unlabeled data from the target domain is accessible during the training phase to facilitate domain adaptation. WiSDA [15] employs weighted cosine similarity to align the feature distributions of corresponding subdomains across source and target domains. However, UDA-based methods fundamentally rely on the availability of target domain data. Consequently, they are restricted to adapting to specific, known environments and struggle to generalize to truly unseen domains, which severely limits their practical application.

In contrast to UDA, DG-based methods eliminate the reliance on target domain data, enabling generalization to completely unseen domains. Broadly, existing DG methods can be categorized into model-based and data-based approaches. Model-based methods typically leverage domain-specific physical knowledge to transform raw CSI data into domain-invariant high-level features, subsequently employing deep neural networks as classifiers. For instance, Widar3.0 [18] constructs a physical model to extract domain-invariant Body-Coordinate Velocity Profile (BVP) features, achieving strong cross-domain performance. However, the extraction of BVP features necessitates a hardware configuration with at least three receivers and one transmitter. This requirement

contradicts the fundamental principles of low-cost and easy deployment in WiFi sensing. Conversely, data-based methods primarily leverage theoretical advancements [22]–[24] in deep learning-based domain generalization to tackle the cross-domain challenge. These approaches advocate addressing domain shift by learning adaptive representations directly from data, rather than relying on explicit physical modeling. For instance, WiGRUNT [13] achieves cross-domain generalization in gesture recognition by training the network to dynamically focus on domain-invariant features via a spatio-temporal dual-attention mechanism. The underlying objective is to enable the network to adaptively determine which parts of the input data are crucial for cross-domain generalization. However, since these methods rely implicitly on statistical patterns learned from data, the resulting features often exhibit limited interpretability. This lack of physical grounding restricts their reliability and limits their broader applicability in practical WiFi sensing scenarios.

To summarize, existing cross-domain WiFi sensing approaches face significant limitations in practical deployment. UDA-based methods are inherently constrained by their dependence on target domain data, rendering them ineffective for completely unseen environments. Within the realm of DG, model-based methods impose stringent hardware requirements that often exceed the capabilities of standard commodity WiFi devices. Conversely, while data-based DG methods offer easier deployment, they typically lack the integration of domain-specific WiFi prior knowledge, leading to limited interpretability. To address these challenges, this paper aims to bridge the gap by proposing a novel prior-guided data-based DG framework. This approach is designed to synergize the flexibility of data-driven learning with signal processing priors, thereby achieving robust generalization across unseen domains without requiring complex hardware setups.

III. METHODOLOGY

A. Problem Formulation

In this section, we formally formulate the DG problem. Suppose S_{train} denotes the training set (source domain set) and S_{test} denotes the testing set (target domain set). In WiFi sensing, a domain represents variations in the scenario, e.g. differences in rooms. S_{train} can be expressed as:

$$S_{\text{train}} = \{S^1_{\text{train}}, S^2_{\text{train}}, \cdots, S^N_{\text{train}}\} \tag{1}$$

where N represents the number of source domains, and each source domain contains multiple samples along with their corresponding labels, which can be expressed as:

$$S^i_{\text{train}} = \{(x^i_j, y^i_j)\}^{n_i}_{j=1} \tag{2}$$

where x^i_j denotes the j-th sample in domain i, y^i_j denotes its corresponding label, and n_i represents the number of samples in domain i. The domain generalization task can be described as training a model on S_{train} that can directly generalize to unseen domains S_{test}. It should be noted that during training, samples from the target domain are not involved, while the class labels and corresponding domain labels of the source domains are fully available.

B. Overview

Fig. 2 illustrates the overall framework of the proposed FDSense, which consists of three main components: data preprocessing, the SFDM module, and deep neural network classification. Specifically, assuming a batch size of 1, for a given pair of source domain samples (x^i_j, y^i_j) and (x^k_t, y^k_t), where $x^i_j \in \mathbb{R}^{A\times C\times T_j}$ and $x^k_t \in \mathbb{R}^{A\times C\times T_t}$ are both complex matrices, and $y^i_j = y^k_t$. Here, A represents the number of antennas, C represents the number of subcarriers, and T_j and T_t represent the length of the time dimension. First, data preprocessing is applied to x^i_j and x^k_t, including denoising, time dimension resampling and amplitude normalization, resulting in $x^{i'}_j \in \mathbb{R}^{(A-1)\times M\times C\times T}$ and $x^{k'}_t \in \mathbb{R}^{(A-1)\times M\times C\times T}$, where M represents the combination of amplitude and phase. Then, $x^{i'}_j$ and $x^{k'}_t$ are fed into the SFDM module, where the high-frequency components of both $x^{i'}_j$ and $x^{k'}_t$ are bidirectionally mixed using mixup, generating domain-mixed samples $x^{ik'}_j \in \mathbb{R}^{(A-1)\times M\times C\times T}$ and $x^{ki'}_t \in \mathbb{R}^{(A-1)\times M\times C\times T}$. Finally, $x^{ik'}_j$ and $x^{ki'}_t$ are concatenated along the sample dimension and passed into a deep neural network for model training and inferencing.

C. Data Preprocessing

Considering that the original CSI data is often corrupted by noise, and that different samples frequently exhibit misalignment in their temporal dimensions as well as in the scales of amplitude and phase, preprocessing is required. Specifically, the preprocessing in this paper includes four steps: CSI ratio [25], Time-domain Difference of CSI (TD-CSI) [26], time dimension resampling and amplitude normalization. Typically, according to [19], CSI data can be modeled using the following formula:

$$H(f,t) = e^{-j\theta_n}\left(H_s(f,t) + H_d(f,t)\right) \tag{3}$$

where $e^{-j\theta_n}$ represents the phase offset, $H_s(f,t)$ denotes the static component in the environment, and $H_d(f,t)$ denotes the dynamic component in the environment.

The phase offset is a time-varying random noise, and its introduction increases the uncertainty of the signal. CSI-ratio method can eliminate the phase offset. Specifically, the CSI-ratio method selects one reference antenna, divides the signals of all other antennas by that of the reference antenna, and subsequently discards the reference antenna.

$H_s(f,t)$ usually originates from reflections caused by static objects such as walls, while $H_d(f,t)$ is induced by human motion in the environment. In WiFi sensing tasks, the primary interest lies in human motion. Therefore, we employ the TD-CSI method to remove the static component from the environment. Specifically, TD-CSI takes the difference between two samples taken within a short interval Δt. In such a short interval, the static path can be regarded as constant.

To enable batch processing, we resample CSI data based on timestamps to ensure a consistent temporal length across

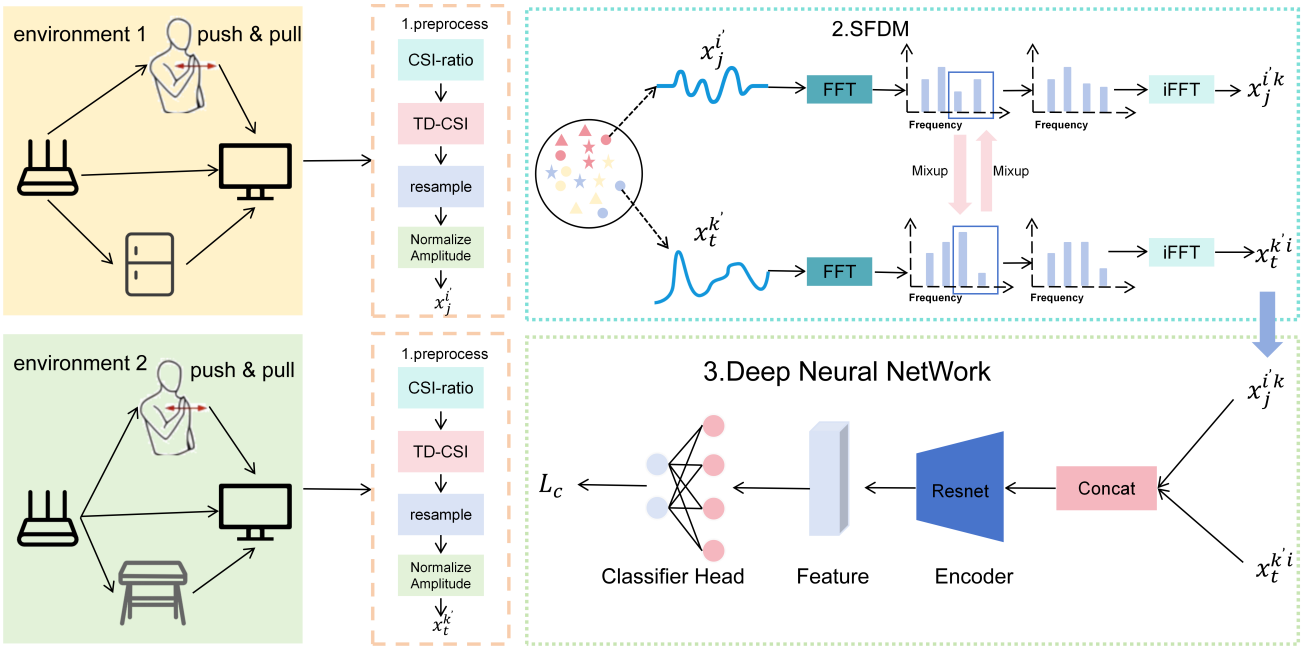

Fig. 2. The training pipeline of FDSense. It consists of three key components: preprocessing, the SFDM module, and a deep neural network. In the SFDM illustration, shapes denote action categories, while colors represent distinct domains. We specifically select samples from different domains that belong to the same category for mixing.

samples. Subsequently, we decompose the CSI complex matrix into amplitude and phase. Considering that amplitude and phase are on different scales, where the amplitude can be over 100 while the phase ranges from $-\pi$ to π, we perform amplitude normalization. Specifically, we first normalize the amplitude to the range $[0, 1]$ using min–max normalization, and then multiply it by π to map it into the range $(0, \pi)$.

D. Selective Fourier Domain Mixing(SFDM)

After preprocessing the CSI samples , we select $x_j^{i'}$ and $x_t^{k'}$ from different domains but with the same action label for SFDM module. This introduces variations in sensing scenarios, encouraging the model to learn domain-invariant features while preserving the correct action category and avoiding label noise.

Specifically, we first map the samples from the time domain to the frequency domain using the Fast Fourier Transform (FFT):

$$\begin{aligned} X_j^{i'}(f) &= \mathcal{F}\{x_j^{i'}(t)\} \\ X_t^{k'}(f) &= \mathcal{F}\{x_t^{k'}(t)\} \end{aligned} \tag{4}$$

where $\mathcal{F}$ denotes the Fast Fourier Transform (FFT). Subsequently, a linear interpolation [27] operation is applied to the high-frequency components, while the low-frequency components remain unchanged:

$$\begin{aligned} X_j^{ik'} &= \lambda\, X_j^{i'}(f_{\text{high}}) + (1-\lambda)\, X_t^{k'}(f_{\text{high}}), \\ X_t^{ki'} &= \lambda\, X_t^{k'}(f_{\text{high}}) + (1-\lambda)\, X_j^{i'}(f_{\text{high}}) \end{aligned} \tag{5}$$

Here, $\lambda \sim \text{Beta}(\alpha, \beta)$, and f_{high} denotes the high-frequency components. In our experiments, f_{high} is set to the latter half of the spectrum. Since the selected samples share the same label, there is no need to perform linear interpolation on the labels. Finally, an inverse Fast Fourier Transform (iFFT) is applied to $X_j^{i'k}$ and $X_t^{k'i}$ to map them back to the time domain:

$$\begin{aligned} x_j^{ik'} &= \mathcal{F}^{-1}\left(X_j^{ik'}\right), \\ x_t^{ki'} &= \mathcal{F}^{-1}\left(X_t^{ki'}\right) \end{aligned} \tag{6}$$

E. Deep Neural Network and Optimization

After passing through the SFDM module, we obtain the time-domain mixed samples $x_j^{ik'}$ and $x_t^{ki'}$. To fully exploit the fused cross-domain information, we concatenate these generated samples with the original samples along the batch dimension to construct an augmented training batch. These samples are then fed into a deep neural network for representation learning. We employ ResNet-18 [28] as the backbone feature extractor, denoted as $\mathcal{F}(\cdot;\theta_{enc})$, where θ_{enc} represents the learnable parameters. The high-level semantic feature vector z is extracted as:

$$z = \mathcal{F}(x;\theta_{enc}) \in \mathbb{R}^d \tag{7}$$

where x represents the input sample (either original or mixed) and d denotes the dimension of the feature embedding. Subsequently, a linear classification head (classifier), denoted as $\mathcal{G}(\cdot;\theta_{cls})$, maps the feature representations to the label space. The probability distribution over the action classes is computed via the Softmax function:

$$\hat{y} = \text{Softmax}(\mathcal{G}(z;\theta_{cls})) = \text{Softmax}(Wz + b) \tag{8}$$

where W and b are the weights and bias of the fully connected layer, respectively.

Since the SFDM module strictly selects samples from the same action category for mixing, the semantic labels of the mixed samples $x_j^{ik'}$ and $x_t^{ki'}$ remain consistent with the original labels y_j^i (which implies $y_j^i = y_t^k$). Therefore, we do not need to perform label interpolation. The entire framework is trained in an end-to-end manner by minimizing the standard Cross-Entropy (CE) loss:

$$\mathcal{L}_c = -\frac{1}{N_{batch}} \sum_{m=1}^{N_{batch}} \sum_{c=1}^{C} y_{m,c} \log(\hat{y}_{m,c}) \tag{9}$$

where N_{batch} is the total size of the augmented batch, C is the number of action classes, $y_{m,c}$ is the ground-truth binary indicator for class c, and $\hat{y}_{m,c}$ is the predicted probability. By optimizing this objective, the network learns to extract robust, domain-invariant features that are discriminative across different sensing environments. It is crucial to emphasize that the primary contribution of this work lies not in developing a novel network architecture for WiFi sensing, but rather in designing an effective inductive bias tailored for the domain shift challenge. By introducing this lightweight inductive bias, our framework achieves superior domain generalization without the need for additional learnable parameters or intricate backbone modifications.

IV. Experiment

A. Dataset and setup

In this paper, we use the Widar [18] dataset. Widar contains six transmitter–receiver pairs. In our experiments, we only use one transmitter–receiver pair. Each pair is equipped with three antennas, and each antenna provides 30 subcarriers. Since not all domains contain all action classes, we select six actions for our experiments: Push&Pull, Sweep, Clap, Slide, Draw-O (horizontal), and Draw-Zigzag (horizontal). In total, 25,020 samples are used. We conduct experiments under the cross-room setting. In the cross-room experiments, the target-domain data is used as the test set, while the source-domain data is split into training and validation sets with an 8:2 ratio. The model achieving the best performance on the validation set is then evaluated on the target domain.

The proposed FDSense framework is implemented in PyTorch and runs on an NVIDIA A6000 GPU. All CSI data are resampled to a temporal length of 800. The learning rate and weight decay of the model are optimized via Bayesian hyperparameter search. The batch size is set to 128, and the model is trained for 80 epochs.

B. Cross room performance

To validate the effectiveness of our framework, we conduct comparative experiments with state of the art WiFi sensing cross domain approaches, including WiGRUNT (DG method) [13], WiOpen (DG method) [19], WiSDA (UDA method) [15], and time-series UDA method AdvSKM [29]. It is worth noting that WiGRUNT, WiOpen, and WiSDA transform CSI data into images for action recognition, and we adopt their preprocessing methods when reproducing their approaches. AdvSKM uses the same preprocessing method as ours.

Table I reports the experimental results under the cross-room setting. All results are obtained by running each experiment four times with different seeds and reporting the mean and variance. The bold numbers indicate the highest mean accuracy. From the results, it can be seen that our method outperforms these cross-domain baselines, demonstrating its superiority.

TABLE I
Cross domain comparative experiments

Cross-domain	**Target Domain Accuracy (%)**		
Methods	room1	room2	room3
Wisda [15]	47.67 ± 0.22	56.60 ± 0.03	65.68 ± 0.27
WiGRUNT [13]	67.60 ± 1.74	80.47 ± 0.94	84.86 ± 0.67
Wiopen [19]	62.38 ± 0.41	74.78 ± 0.75	83.59 ± 0.36
AdvSKM [29]	73.57 ± 0.38	86.13 ± 0.01	88.94 ± 0.18
FDSense	**74.03 ± 0.62**	**87.05 ± 0.44**	**90.86 ± 0.22**

C. Ablation Study

We conduct an ablation study on the proposed SFDM module, and the results are presented in Table II. Here, 'w/o SFDM' denotes not using the SFDM module, 'w SFDM' denotes using the SFDM module, 'w STM' refers to applying our selective strategy while performing mixup in the time domain, and 'w FDM' refers to performing mixup in the high-frequency domain without the selective strategy. The results demonstrate that the SFDM module effectively improves the model's generalization under the cross-room setting. Moreover, both high-frequency mixup and the selective strategy are crucial, as removing either one results in reduced performance.

We also evaluated the SFDM module under other cross-domain settings, such as cross-user scenarios, where its performance did not surpass the baseline without SFDM. We attribute this limitation to the specificity of our physical prior. The underlying assumption that high-frequency components primarily encode domain-specific environmental noise is fundamentally designed for cross-room scenarios where multipath scattering from static objects dominates the high-frequency spectrum. In contrast, cross-user domain shifts arise from variations in physiological characteristics and motion styles, which significantly alter the low-frequency dynamic components of the CSI signal. Consequently, our spectral mixing strategy, which targets high-frequency perturbations to simulate environmental changes, is less applicable to user-induced domain shifts. While our frequency-domain prior is less effective for cross-user shifts where high-frequency variations are minimal, our design strategically prioritizes cross-room robustness to address critical single-user applications, such as monitoring a specific individual across different indoor spaces.

V. Conclusion

In this paper, we propose the FDSense framework to address domain shift in WiFi sensing. It combines signal processing

TABLE II
ABLATION STUDY ON SFDM

	room1	room2	room3
w/o SFDM	72.84 ± 1.98	86.62 ± 0.06	90.36 ± 0.34
w STM	70.02 ± 1.00	83.37 ± 1.56	87.80 ± 0.92
w FDM	55.74 ± 4.49	71.27 ± 0.84	78.33 ± 0.08
w SFDM	**74.03 ± 0.62**	**87.05 ± 0.44**	**90.86 ± 0.22**

with domain generalization guided by CSI priors. We first denoise and standardize the CSI signals, then use a Selective Fourier Domain Mixing (SFDM) module to encourage learning of domain-invariant representations, and finally apply deep neural networks for feature extraction and classification. Cross-room experiments on the Widar dataset demonstrate the effectiveness of our approach.

VI. ACKNOWLEDGMENT

This research is supported by the Open Project Program of Shanghai Key Laboratory of Data Science (No.2025090600006).

REFERENCES

[1] S. Yousefi, H. Narui, S. Dayal, S. Ermon, and S. Valaee, "A survey on behavior recognition using wifi channel state information," *IEEE Communications Magazine*, vol. 55, no. 10, pp. 98–104, 2017.

[2] Y. Ma, G. Zhou, and S. Wang, "Wifi sensing with channel state information: A survey," *ACM Computing Surveys*, vol. 52, no. 3, pp. 1–36, 2019.

[3] S. Tan, Y. Ren, J. Yang, and Y. Chen, "Commodity wifi sensing in ten years: Status, challenges, and opportunities," *IEEE Internet of Things Journal*, vol. 9, no. 18, pp. 17 832–17 843, 2022.

[4] D. Ahn, S. Kim, H. Hong, and B. C. Ko, "Star-transformer: a spatio-temporal cross attention transformer for human action recognition," in *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, 2023, pp. 3330–3339.

[5] N. Ahmad, H.-f. Leung, and F. Farnia, "Virtualhar: Virtual sensing device and correlation-based learning approach for multi-wearable sensing device-based human activity recognition," *IEEE Internet of Things Journal*, 2025.

[6] D. Anguita, A. Ghio, L. Oneto, X. Parra, J. L. Reyes-Ortiz *et al.*, "A public domain dataset for human activity recognition using smartphones." in *Esann*, vol. 3, no. 1, 2013, pp. 3–4.

[7] Z. Chen, L. Zhang, C. Jiang, Z. Cao, and W. Cui, "Wifi csi based passive human activity recognition using attention based blstm," *IEEE Transactions on Mobile Computing*, vol. 18, no. 11, pp. 2714–2724, 2018.

[8] J. Zhang, F. Wu, B. Wei, Q. Zhang, H. Huang, S. W. Shah, and J. Cheng, "Data augmentation and dense-lstm for human activity recognition using wifi signal," *IEEE Internet of Things Journal*, vol. 8, no. 6, pp. 4628–4641, 2020.

[9] B. Li, W. Cui, W. Wang, L. Zhang, Z. Chen, and M. Wu, "Two-stream convolution augmented transformer for human activity recognition," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 35, no. 1, 2021, pp. 286–293.

[10] F. Luo, S. Khan, B. Jiang, and K. Wu, "Vision transformers for human activity recognition using wifi channel state information," *IEEE Internet of Things Journal*, vol. 11, no. 17, pp. 28 111–28 122, 2024.

[11] F. Luo, A. Li, B. Jiang, S. Khan, K. Wu, and L. Wang, "Activitymamba: a cnn-mamba hybrid neural network for efficient human activity recognition," *IEEE Transactions on Mobile Computing*, 2025.

[12] S. Liu, Z. Chen, M. Wu, C. Liu, and L. Chen, "Wisr: Wireless domain generalization based on style randomization," *IEEE Transactions on Mobile Computing*, vol. 23, no. 5, pp. 4520–4532, 2023.

[13] Y. Gu, X. Zhang, Y. Wang, M. Wang, H. Yan, Y. Ji, Z. Liu, J. Li, and M. Dong, "Wigrunt: Wifi-enabled gesture recognition using dual-attention network," *IEEE transactions on human-machine systems*, vol. 52, no. 4, pp. 736–746, 2022.

[14] S. Liu, Z. Chen, M. Wu, H. Wang, B. Xing, and L. Chen, "Generalizing wireless cross-multiple-factor gesture recognition to unseen domains," *IEEE Transactions on Mobile Computing*, vol. 23, no. 5, pp. 5083–5096, 2023.

[15] W. Jiao, C. Zhang, W. Du, and S. Ma, "Wisda: Subdomain adaptation human activity recognition method using wi-fi signals," *IEEE Transactions on Mobile Computing*, 2024.

[16] B.-B. Zhang, D. Zhang, Y. Li, Y. Hu, and Y. Chen, "Unsupervised domain adaptation for rf-based gesture recognition," *IEEE Internet of Things Journal*, vol. 10, no. 23, pp. 21 026–21 038, 2023.

[17] B. Sheng, R. Han, F. Xiao, Z. Guo, and L. Gui, "Metaformer: Domain-adaptive wifi sensing with only one labelled target sample," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 8, no. 1, pp. 1–27, 2024.

[18] Y. Zhang, Y. Zheng, K. Qian, G. Zhang, Y. Liu, C. Wu, and Z. Yang, "Widar3. 0: Zero-effort cross-domain gesture recognition with wi-fi," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 11, pp. 8671–8688, 2021.

[19] X. Zhang, J. Huang, H. Yan, Y. Feng, P. Zhao, G. Zhuang, Z. Liu, and B. Liu, "Wiopen: A robust wi-fi-based open-set gesture recognition framework," *IEEE Transactions on Human-Machine Systems*, 2025.

[20] Y. He, Y. Chen, Y. Hu, and B. Zeng, "Wifi vision: Sensing, recognition, and detection with commodity mimo-ofdm wifi," *IEEE Internet of Things Journal*, vol. 7, no. 9, pp. 8296–8317, 2020.

[21] Y. Yang and S. Soatto, "Fda: Fourier domain adaptation for semantic segmentation," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 4085–4095.

[22] Y. Li, X. Tian, M. Gong, Y. Liu, T. Liu, K. Zhang, and D. Tao, "Deep domain generalization via conditional invariant adversarial networks," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 624–639.

[23] H. Yao, Y. Wang, S. Li, L. Zhang, W. Liang, J. Zou, and C. Finn, "Improving out-of-distribution robustness via selective augmentation," in *International Conference on Machine Learning*. PMLR, 2022, pp. 25 407–25 437.

[24] D. Li, Y. Yang, Y.-Z. Song, and T. Hospedales, "Learning to generalize: Meta-learning for domain generalization," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 32, no. 1, 2018.

[25] D. Wu, R. Gao, Y. Zeng, J. Liu, L. Wang, T. Gu, and D. Zhang, "Fingerdraw: Sub-wavelength level finger motion tracking with wifi signals," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 4, no. 1, pp. 1–27, 2020.

[26] W. Li, R. Gao, J. Xiong, J. Zhou, L. Wang, X. Mao, E. Yi, and D. Zhang, "Wifi-csi difference paradigm: Achieving efficient doppler speed estimation for passive tracking," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 8, no. 2, pp. 1–29, 2024.

[27] H. Zhang, M. Cisse, Y. N. Dauphin, and D. Lopez-Paz, "mixup: Beyond empirical risk minimization," *arXiv preprint arXiv:1710.09412*, 2017.

[28] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[29] Q. Liu and H. Xue, "Adversarial spectral kernel matching for unsupervised time series domain adaptation." in *IJCAI*, 2021, pp. 2744–2750.