Wi-SFDAGR: WiFi-based Cross-Domain Gesture Recognition via Source-free Domain Adaptation

Huan Yan, Xiang Zhang*, Jinyang Huang, Yuanhao Feng,
Meng Li, Anzhi Wang, Weihua Ou, Hongbing Wang, and Zhi Liu*

***Abstract*—WiFi Channel State Information (CSI)-based gesture recognition offers unique advantages, including cost-effectiveness and enhanced privacy protection, and has garnered significant attention in recent years. However, existing WiFi-based gesture recognition solutions exhibit poor generalization ability when deployed in new environment, orientation, or location. Although some methods combine labeled source domain and unlabeled target domain to learn domain-independent features, factors such as data privacy protection hinder access to source data during practical environment adaptation. Consequently, we consider realistic scenario where source data is unavailable during adaptation of unlabeled test data, and instead, a trained source domain model is used. In this paper, we propose Wi-SFDAGR, a WiFi-based Source-free Domain Adaptation Gesture Recognition framework. Specifically, we treat cross-domain as an unsupervised clustering problem, aiming to ensure that features within local neighborhoods exhibit similar prediction results while those farther apart display different prediction outcomes in the feature space. We theoretically analyze the effect of enhanced prediction consistency between neighbor points extracted from gestures on generalization error. Furthermore, we employ an attraction-dispersion network to strengthen prediction consistency among closely located features in the feature space while reducing it for distantly located features. To mitigate noise introduced during nearest neighbor sample selection in the feature space (where predictions may not align with the input sample's prediction), we progressively improve nearby sample feature aggregation by estimating uncertainty to reweight local neighborhood predictions. Finally, extensive experiments are conducted on the Widar 3.0 and XRF55 datasets and the results show our proposed framework outperforms most cross-domain methods.**

***Index Terms*—Gesture Recognition, Channel State Information, Source-free Domain Adaptation, Deep Learning.**

I. Introduction

GESTURE recognition, as a means for computers to understand human body language, plays an increasingly important role in the field of human-computer interaction [1]. Existing solutions based on visual devices and sensors face challenges such as privacy protection, limited line of sight, and high implementation costs [2]. With the continuous proliferation of WiFi devices and the ongoing development of WiFi sensing technology, WiFi-based gesture recognition garners more attention from researchers in the Internet of Things (IoT) systems [3], [4]. Its applications are diverse, including medical monitoring [5], smart homes [6], virtual reality [7], and other fields.

WiFi-based sensing technologies utilize Channel State Information (CSI) or Received Signal Strength Indicator (RSSI) to describe the characteristics of wireless channels. Due to the typically poor stability of collected RSSI, most WiFi-based sensing methods currently tend to utilize CSI from the underlying physical layer for implementation [8]–[11]. In indoor environments, reflections and diffractions of signals caused by the human body introduce additional paths. Therefore, the impact of human activities on the propagation of WiFi signals is characterized by the signals reaching the receiver. By establishing a mapping between these signal changes and different human activities, the basic concept of gesture recognition based on WiFi is established [12], [13], [13]–[15]. Currently, an increasing number of WiFi-based gesture recognition methods demonstrate promising performance. However, the inherent complexity of WiFi signals is susceptible to environmental deployment, resulting in differences in the distribution of training (source domain) and testing (target domain) data. This leads to significant performance degradation when these methods are applied in different indoor environments. The limitations in unseen domains restrict the model's generalization.

In recent years, some researchers address the domain/distribution shift problem through **model-based** and **learning-based** approaches. In model-based methods, efforts primarily focus on proposing signal processing algorithms to extract environment-independent features (such as Body-coordinate Velocity Profile (BVP) [16], Doppler Frequency Shift (DFS) [17], etc.) from the received signals. However, these model-based approaches also come with limitations. On one hand, they require prior knowledge to guide the manual design of domain-agnostic features, increasing the complexity of system implementation. On the other hand, due to limited fitting capability, they exhibit lower performance for complex gesture recognition tasks. Due to some limitations of the aforementioned model-based methods, and with the continuous increase in data volume and computing power, as well as the ongoing development of deep learning methods, learning-based methods are increasingly attracting the attention of researchers [18]–[21]. They achieve this by designing task-relevant neural network architectures to enable end-to-end

Huan Yan, Anzhi Wang, Weihua Ou and Hongbing Wang, School of Big Data and Computer Science, Guizhou Normal University, Guiyang, 550025, China.

Xiang Zhang and Yuanhao Feng, CAS Key Laboratory of Electromagnetic Space Information, University of Science and Technology of China, Hefei, 230026, China.

Jinyang Huang and Meng Li, School of Computer and Information, Hefei University of Technology, Hefei, 230601, China.

Zhi Liu, Department of Computer and Network Engineering, The University of Electro-Communications, Tokyo, 1828585, Japan.

* Corresponding authors.

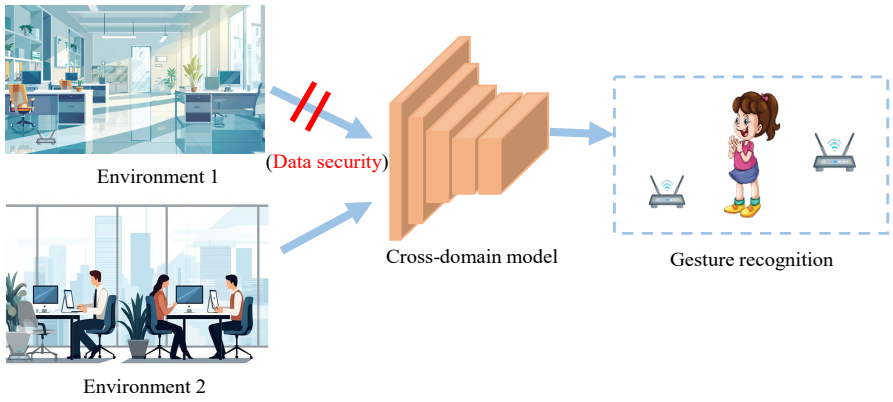


Fig. 1. Under data security restrictions, cross-domain (e.g., different environments) gesture recognition cannot access source domain data during model training.

adaptive learning, thereby extracting discriminative features. Currently, some WiFi-based gesture recognition methods trend toward using Unsupervised Domain Adaptation (UDA) techniques to address cross-domain issues [18], [19]. Traditional domain adaptation methods typically rely on the simultaneous use of labeled source domain data and unlabeled target domain data to learn domain-invariant features for gesture recognition tasks. However, this approach faces significant challenges in practical applications, particularly in terms of data security and privacy protection. The development of source-free domain adaptation technology aims to address these issues, and its core motivations can be elaborated from the following perspectives:

First, from the perspective of data security, many real-world application scenarios (such as healthcare and financial transactions) impose strict requirements on data privacy. Source domain data often contains sensitive information, and its transmission and use may violate relevant laws and regulations, posing risks of data breaches. Second, from a practical standpoint, even in scenarios where data sharing is permitted, storing and processing large-scale source domain data incurs substantial computational overhead and storage costs, which is particularly disadvantageous for resource-constrained edge devices [22]. Finally, from the viewpoint of technological trends, with the growing awareness of data sovereignty and advancements in privacy-preserving computation, the development of adaptation methods that do not rely on source domain data has become an important research direction in the field. As shown in Figure 1, Environment 1 and Environment 2 represent the source and target domains, respectively. Unlike traditional methods, source-free domain adaptation completely eliminates the need to access data from Environment 1 during training. Instead, it achieves cross-domain adaptation through innovative techniques such as feature distribution alignment and model knowledge transfer. This approach not only effectively addresses data security concerns but also significantly reduces computational resource requirements, offering greater convenience for practical deployment.

To overcome the limitations of existing learning-based methods and enable cross-domain models to adapt well to complex real-world scenarios where data privacy or transmission bandwidth is critical, this paper aims to extract domain-independent discriminative features from WiFi signals. It considers a more realistic scenario, where the source data is no longer available during the adaptation of unlabeled test data, and instead, a trained source domain model is used (i.e., source-free domain adaptation gesture recognition). We explore effective ways to improve the generalization performance of sensing models in this context. Specifically, we analyze two key issues that can be used to enhance the generalization ability of existing models:

How to theoretically analyze the impact of feature space optimization on cross-domain generalization boundary? This issue is currently one of the key challenges in cross-domain gesture recognition. Existing research methods primarily focus on intuitively designing network architectures for cross-domain learning and analyze their effectiveness in results, while neglecting theoretical analysis of the impact of feature space optimization on cross-domain generalization boundary. Even though these methods have proven effective on WiFi datasets, their applicability to other domain datasets cannot be discerned (i.e., lacking sufficient theoretical foundation).

How to design a deep neural network that can effectively extract domain-independent gesture features, given only the knowledge of a trained source domain model? Existing designs of deep neural networks for cross-domain gesture recognition primarily rely on joint input from source and target domain datasets. By analyzing the similarities and differences between these datasets, they extract domain-independent features. Previous methods overlook the fact that in practical scenarios, access to the source domain dataset is often restricted due to security or transmission limitations. Even if these models show effectiveness in results, their performance may significantly degrade when applied in real environments.

To address the issues identified in the above analysis. In this paper, we propose a WiFi-based source-free domain adaptation gesture recognition framework, namely Wi-SFDAGR. Firstly, we give a *theoretical analysis* of the source-free domain adaptation problem. Our theoretical analysis indicates that attracting predictions of local nearest neighbor points in the

feature space and dispersing predictions of features that are distant in the feature space have an impact on the generalization boundary. Therefore, we seek to optimize the upper bound of the objective and utilize an *attraction-dispersion network* to enhance the prediction consistency of features that are closer in the feature space while reducing the prediction consistency of features that are farther apart in the feature space. Furthermore, when selecting local nearest neighbor points, these samples may contain noise (i.e., the predictions of the nearest neighbor samples may not be consistent with the predictions of the input sample), inevitably affecting the cross-domain feature learning of attraction-dispersion network. To reduce the impact of noise introduced by nearest neighbor samples, we reweight the predictions of neighboring samples by estimating the *uncertainty weights* of the nearest neighbor samples, gradually enhancing the feature aggregation of nearby samples. We conduct extensive experiments on the publicly available gesture recognition dataset, Widar3.0, which is based on commercial WiFi devices. The results demonstrate that our proposed method, Wi-SFDAGR, achieves recognition rates of 97.30%, 97.17%, and 95.52% for cross-location, cross-orientation, and cross-environment scenarios, respectively. Moreover, we share all the source codes on https://github.com/snail-yh/Wi-SFDAGR to facilitate further validation and optimization.

The main contributions of this paper are summarized below:

- We theoretically analyze the impact of strengthening the prediction consistency of local neighboring points and weakening the prediction consistency of features that are far apart in the feature space on the source-free domain adaptation generalization boundary.
- We propose Wi-SFDAGR, a WiFi-based source-free domain adaptation gesture recognition framework, which utilizes an attraction-dispersion network to enhance the prediction consistency of features closer in the feature space while simultaneously weakening the prediction consistency of features farther away. To the best of our knowledge, this is the first work addressing source-free domain adaptation for gesture recognition based on WiFi.
- To mitigate the noise present in the nearest neighbor samples, we reweight the predictions of local neighborhoods by estimating uncertainty, gradually enhancing the feature aggregation of nearby samples.
- We conduct extensive experiments on the Widar3.0 and XRF55 datasets, including tasks across location, orientation, and environment. The experimental results demonstrate the superiority of this method.

The rest of this paper is organized as follows. Section II introduces related work, including WiFi-based gesture recognition and domain adaptation. Section III presents the preliminaries, including CSI-based gesture recognition, problem definition, and insights into the design of learning-based gesture recognition. Section IV describes the experimental methodology, with subsection IV-A detailing the system architecture, IV-B covering preprocessing, and IV-C outlining the design of the source-free domain adaptation framework. Subsequently, in Section IV, Wi-SFDAGR is evaluated. Section V concludes the paper.

II. RELATED WORK

A. WiFi-based Gesture Recognition

Due to the non-invasive, through-wall, and ubiquitous nature of WiFi sensing, WiFi-based gesture recognition has attracted increasing attention from researchers. Existing WiFi-based gesture recognition methods focus on modeling the correlation between WiFi signals and gestures to improve gesture recognition accuracy or enhance the generalization ability of gesture recognition model. In the past, a considerable amount of research work was conducted based on handcrafted features. Initially, data preprocessing was performed on raw CSI data, followed by the extraction of handcrafted features (such as statistical features), and finally, gesture recognition was carried out using machine learning methods (such as SVM) [20], [21], [23]–[28]. WiFinger [23] utilized detailed CSI to perceive and identify subtle finger gestures. It incorporated a mechanism for removing environmental noise to mitigate the impact of environmental changes on signal dynamics. Moreover, WiFinger addressed individual diversity and gesture inconsistency by capturing the intrinsic behavior of gestures. Gao et al. [24] characterized the perceived quality of the signal by building a mathematical model that relates the gesture signal to ambient noise, from which they further derived a unique metric, Error of Dynamic Phase index (EDP-index), to quantitatively describe the perceived quality of each gesture signal segment. Wi-NN [25] utilized time-domain-based feature extraction methods to extract clear motion information, obtaining gesture motion feature data. This collected data was then combined with a weighted K-Nearest Neighbors (KNN) classifier to complete the classification task.

With the continuous increase in data volume and the improvement of computational power, deep learning can handle more complex and detailed tasks. Many researchers are focusing on utilizing deep learning for wireless sensing. Tong et al. [20] developed a gesture recognition system based on CSI by constructing a dynamic CNN-GRU-Attention (CGA) model. They utilized data processing methods such as phase correction and unwrapping, along with a newly proposed adaptive gesture action truncation algorithm, to extract phase differences and remove redundant information, thereby ensuring the validity of the input data. Gu et al. [21] developed an attention-mechanism-based gesture recognition framework, which aims to appropriately track the importance of amplitude and phase information of CSI ratio to adaptively extract salient features associated with gestures. Yang et al. [26] proposed a new deep Siamese representation learning architecture for one-shot gesture recognition that extends the spatio-temporal pattern learning capabilities of standard Siamese structures by combining convolution and bidirectional recurrent neural networks. WiHGR [27] constructed the phase difference matrix based on the phase difference between two adjacent receiving antennas on the dominant path. An improved Attention-based Bi-directional Gate Recurrent Unit (ABGRU) network is proposed to automatically learn and extract discriminant features from phase difference matrices. Regani et al. [28] have

utilized rich multipaths in through-the-wall Settings to develop statistical models of channel changes caused by gestures. The model is used to derive the correspondence between the relative distance of the hand movement and the attenuation of the Time Reversal Resonating Strength (TRRS). Using this relationship and the geometric features of gesture shapes, a feature extraction module is designed to realize gesture classification.

Although the aforementioned research has modeled the correlation between WiFi signals and gesture actions, it overlooks the influence of different environments on gesture recognition. Performing the same gesture action in different environments can lead to significant variations in WiFi waveforms [29], [30]. Therefore, some researchers are focusing on the cross-domain recognition capability of gesture recognition systems [16], [31]–[35]. Widar 3.0 [16] introduced a novel domain-agnostic feature capable of capturing Body-coordinate Velocity Profile (BVP) of human gestures at lower signal levels. Theoretically, BVP is independent of any domain-specific information in the original WiFi signal, thus serving as a unique indicator of human gestures. Furthermore, Widar 3.0 developed a one-size-fits-all model that requires only one training session but can adapt to different data domains. WiGesture [31] switched from a traditional transceiver view to a hand-oriented view and extracted features independent of location-specific factors known as Motion Navigation Primitive (MNP), which capture patterns of changes in the direction of hand movement and share consistent patterns when users perform the same gesture at different location-specific factors. Wi-Learner [32] first achieved cross-domain gesture recognition by capturing gesture-induced Doppler Frequency Shift (DFS) from noise measurements using a well-designed signal processing scheme. The low-dimensional features are then extracted using an autoencoder based on convolutional neural networks and fed into a downstream model for gesture recognition. Su et al. [33] eliminated the random phase noise in CSI and performed phase calibration, and fed the processed phase sequence into a lightweight deep neural network for cross-domain gesture recognition. CROSSGR [34] has extracted user-independent but gesture-related features from WiFi channel measurements that do not rely on prior knowledge of the target (e.g., user ID/ tag), significantly enhancing the utility of the developed WiFi sensing system. WiGr [35] utilized the similarity between query sample representations and class prototypes in the embedding space for gesture classification, thereby avoiding the impact of cross-domain CSI pattern variations.

B. Domain Adaptation

Traditional deep learning assumes that training and testing samples come from the same distribution, where the model is trained on the training set and then tested on the test set. However, in real-world scenarios, there exist differences in the distributions of training and testing data [36], [37]. When a trained model is deployed in an environment with distributional differences, its performance may significantly deteriorate. Domain adaptation is a method to address such distributional differences between domains. Its essence lies in identifying the similarity between the source domain and the target domain and leveraging this similarity to transfer the knowledge acquired in the source domain to the target domain.

The current domain adaptation solutions mainly include methods based on domain distribution discrepancy, adversarial approaches, reconstruction-based methods, and sample generation techniques. Ge et al. [38] aligned the conditional distributions by minimizing the Conditional Maximum Mean Discrepancy (CMMD) and extracted discriminative information from the target domain by maximizing the mutual information between samples and predicted labels. The CMMD directly measures the difference between conditional distributions by embedding them in the Reproducing Kernel Hilbert Space (RKHS) of conditional distributions. Zhu et al. [39] effectively constructed an intermediate domain by learning to sample patches from two domains based on a game-theoretic model. It learned patches from both the source and target domains to maximize cross-entropy, while minimizing the cross-entropy using two semi-supervised mixed losses in the feature and label spaces. Na et al. [40] introduced a fixed ratio-based mixture to enhance multiple intermediate domains between the source and target domains. On the enhanced domains, source-dominant and target-dominant models with complementary features were trained. Du et al. [41] proposed a Cross-domain Gradient Discrepancy Minimization (CGDM) method, which explicitly minimizes the gradient differences generated by the source and target samples. Mekhazni et al. [42] proposed a novel Dissimilarity-based Maximum Mean Discrepancy (D-MMD) loss for aligning paired distances, which can be optimized using gradient descent with relatively small batch sizes.

III. PRELIMINARY

A. CSI-based Gesture Recognition

In WiFi-based sensing applications, Channel State Information (CSI) describes the channel properties from the transmitter to the receiver, including environmental conditions, signal attenuation, and other factors. When individuals are in environments with deployed WiFi devices, their movements introduce additional paths due to signal reflections and diffractions. Therefore, the impact of human activities on WiFi signal propagation is reflected in the signals reaching the receiver. Typically, due to multipath effects, the signal at the receiver is a combination of signals from all paths, primarily divided into static paths and dynamic paths. The former refers to the line-of-sight path and reflections from static indoor objects (such as furniture, walls, etc.), while the latter refers to reflections from moving objects (such as the human body). Therefore, CSI, represented as $H(f,t)$, can be expressed as:

$$H(f,t) = e^{-j2\pi\Delta f t}(H_s(f,t) + H_d(f,t)), \tag{1}$$

where $e^{-j2\pi\Delta f t}$ arises due to the non-time synchronization between the transmitter and receiver, resulting in time-varying random phase offsets. Additionally, Δf represents the difference in carrier frequencies. $H_s(f,t)$ denotes the channel frequency response for static paths, while $H_d(f,t)$ represents the

channel frequency response for dynamic paths. Specifically, $H_d(f,t)$ can be expressed as:

$$H_d(f,t) = \sum_{k=1}^{p_d} \alpha_k(f,t) e^{-j2\pi f \tau_k(t)}, \tag{2}$$

where p_d denotes the number of dynamic paths, $\alpha_k(f,t)$ represents the complex value indicating the initial phase and attenuation of the k-th path, and $\tau_k(t)$ signifies the propagation delay on the k-th path. In WiFi-based gesture recognition, different gestures cause changes in the propagation path length of the dynamic component. That is to say, the CSI records the correlation between the gesture movements and the changes in the propagation signal. By analyzing this correlation, the corresponding gestures can be recognized in the received WiFi signal.

B. Problem Definition

Definition 1: In-domain gesture recognition

We refer to environmental factors unrelated to the gesture itself as domain factors. In WiFi-based gesture recognition, we focus on four domain factors: user location, user orientation, environmental spatial layout, and user diversity. Previous in-domain gesture recognition methods assume that the training set and the test set have the same probability distribution and train corresponding models for different domains. However, when the model is adapted to different domains, the recognition performance is greatly reduced.

Definition 2: Cross-domain gesture recognition

For cross-domain gesture recognition. We are given a labelled source domain with N_s samples as $\mathcal{D}_s = \{(x_i^s, y_i^s)\}_{i=1}^{N_s}$ according to the source distribution Q defined on $\mathcal{X} \times \mathcal{Y}$, where $\mathcal{X}$ represents the input CSI dataset and $\mathcal{Y}$ represents the label set. Meanwhile, We are also given an unlabelled target domain with N_t samples as $\mathcal{D}_t = \{x_i^t\}_{i=1}^{N_t}$ according to the target distribution P that somewhat differs from Q. Both the source domain and the target domain contain the same C categories (only discusses the close-set setting in this paper). Due to the inherent complexity of WiFi signals, they are easily influenced by the deployment environment, leading to certain differences between the source distribution Q and the target distribution P. Therefore, the goal of cross-domain gesture recognition is to train a model on the source domain that can still achieve good performance on a target domain with a probability distribution different from that of the source domain.

Problem formulation: Source-free adaptation domain gesture recognition

The goal of source-free domain adaptation gesture recognition is to enable the model to adapt to the target domain without directly accessing the source domain data, but only receiving a pre-trained model from the source domain during adaptation. Therefore, the labeled source domain dataset $\mathcal{D}_s$ is used to train the source domain model before the adaptation phase. Our model consists of two components: a feature extractor f and a classifier g. They are parameterized by parameters β and θ respectively. The output feature representation of the feature extractor is denoted as $z_i = f(x_i;\beta) \in \mathbb{R}^h$, where h represents the output dimension of the feature extractor. The output class representation of the classifier is denoted as $p_i = \delta(g(f(x_i;\beta);\theta)) \in \mathbb{R}^C$, where δ is the softmax function. We regard source-free domain adaptation as an unsupervised clustering problem, where a pre-trained model from the source domain is utilized to initialize the adaptation phase in the target domain. By considering the neighborhood information, this intrinsic structure of the target data is used for source-free domain adaptation gesture recognition.

C. Insights into the Design Learning-based Gesture Recognition

In the context of the above problem, we consider neighborhood information and utilize the inherent structure of the target data to solve unsupervised domain adaptation for gesture recognition. In this section, we provide insights into the design of learning-based gesture recognition. We analyze how promoting similar predictive results within local neighborhoods in the feature space while having different predictive results for more distant features in the feature space can help reduce the expected error of the generalization boundary in the target domain. We randomly select a sample x' from the target domain dataset $\mathcal{D}_t$ as an example. We then divide the remaining dataset into nearest neighbors $\mathcal{D}_r = \{x_i^r\}_{i=1}^{N_r}$, and non-neighbors $\mathcal{D}_l = \{x_i^l\}_{i=1}^{N_l}$.

Theorem 1. (Rademacher Generalization Bound [43], [44]). *Let $\mathcal{G}$ be a family of functions mapping from $\mathcal{X}$ to $[0,1]$ and $\hat{D}$ a fixed sample of size n drawn from the distribution D. Then, for any $\delta > 0$, with probability $1-\delta$, the following holds for all $g \in \mathcal{G}$:*

$$|\mathbb{E}_D(g) - \mathbb{E}_{\hat{D}}(g)| \leq 2\mathfrak{R}_{n,G}(\mathcal{G}) + \sqrt{\frac{log\frac{2}{\delta}}{2n}}, \tag{3}$$

where $\mathbb{E}_D(g)$ and $\mathbb{E}_{\hat{D}}(g)$ respectively represent the generalization risk and empirical risk of g. $\mathfrak{R}_{n,G}(\mathcal{G})$ represents Rademacher complexity.

Lemma 2. *For $\mathcal{D}_r$ and $\mathcal{D}_l$, with the label function represented as g_t and $\mathcal{D}_r \cap \mathcal{D}_l = \emptyset$, we can obtain the following equation:*

$$\begin{aligned}
\xi_{\hat{\mathcal{D}}_t}(h) &= \frac{1}{N_t}\sum_{i=0}^{N_t}\mathcal{L}(x_i, f_t(x_i)) \\
&\approx \frac{1}{N_t}\sum_{i=0}^{N_r}\mathcal{L}(x_i, f_t(x_i)) + \frac{1}{N_t}\sum_{i=0}^{N_l}\mathcal{L}(x_i, f_t(x_i)) \\
&\approx \frac{N_r}{N_t}\frac{1}{N_r}\sum_{i=0}^{N_r}\mathcal{L}(x_i, f_t(x_i)) + \frac{N_l}{N_t}\frac{1}{N_t}\sum_{i=0}^{N_l}\mathcal{L}(x_i, f_t(x_i)) \\
&\approx R_r\xi_{\hat{\mathcal{D}}_r}(h) + R_l\xi_{\hat{\mathcal{D}}_l}(h),
\end{aligned} \tag{4}$$

where R_r and R_l represent the proportion of $\mathcal{D}_r$ and $\mathcal{D}_l$ to the target domain dataset $\mathcal{D}_t$, respectively, and $R_r + R_l = \frac{N_t-1}{N_t}$.

From Theorem 1, we can conclude the following:

$$\xi_{\mathcal{D}_t}(h) \leq \xi_{\hat{\mathcal{D}}_t}(h) + 2\mathfrak{R}_{n,G}(\mathcal{G}) + \sqrt{\frac{log\frac{2}{\delta}}{2n}}. \tag{5}$$

Combining Lemma 2, we can derive the following inequality:

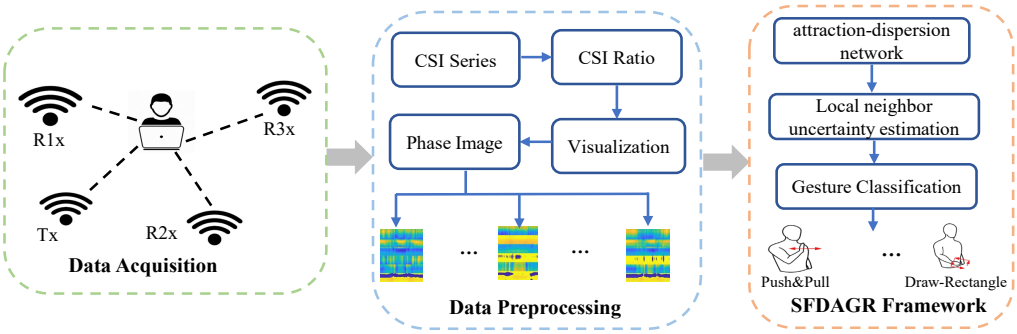


Fig. 2. The System framework of Wi-SFDAGR. Wi-SFDAGR is primarily divided into two parts: data preprocessing and SFDAGR. In the data preprocessing, CSI ratios are generated by utilizing CSI readings from different antennas, which not only eliminates a significant amount of noise but also extends the sensing range. Subsequently, the phase of the CSI ratios is visualized as CSI images suitable for deep learning processing. Finally, a source-free domain adaptation model is trained through the Attraction-Dispersion Network and Local Neighbor Uncertainty Estimation modules, enabling the extraction of domain-independent features.

$$\xi_{\mathcal{D}_t}(h) \leq R_r \xi_{\hat{\mathcal{D}}_r}(h) + R_l \xi_{\hat{\mathcal{D}}_l}(h) + 2\mathfrak{R}_{n,G}(\mathcal{G}) + \sqrt{\frac{log\frac{2}{\delta}}{2n}}, \quad (6)$$

where $2\mathfrak{R}_{n,G}(\mathcal{G}) + \sqrt{\frac{log\frac{2}{\delta}}{2n}}$ is a constant term. From the generalization bound of the target domain, we can observe that the generalization error of the target hypothesis h is constrained by two terms: the training error $\xi_{\hat{\mathcal{D}}_r}(h)$ on the nearest neighbor set and the training error $\xi_{\hat{\mathcal{D}}_l}(h)$ on the non-neighbor set. When searching for the optimal hypothesis space, the number of nearest neighbors selected in each iteration is fixed, namely $N_r = K$, where K is a constant term. This implies that, in addressing source-free domain adaptation for WiFi-based gesture recognition, minimizing the training errors of the first three terms of Equation 6 separately can reduce the generalization boundary of the target domain.

IV. IMPLEMENTATION

A. System Architecture

This section mainly describes the system framework of the proposed Wi-SFDAGR, as shown in Figure 2. Specifically, to address practical scenarios where data privacy or transmission bandwidth becomes critical issues in gesture recognition, we propose a method for gesture recognition via source-free domain adaptation. In Figure 2, Wi-SFDAGR consists primarily of data preprocessing and Source-free Domain Adaptation Gesture Recognition (SFDAGR) framework. The detailed processing procedures of each component in Wi-SFDAGR are introduced below.

B. Data Preprocessing

Due to the limitations of spectrum resources, researchers have turned to multiple antennas and Multiple Input Multiple Output (MIMO) techniques [45] to enhance the throughput and extend the transmission distance of communication systems. As a result, during Channel State Information (CSI) collection, multiple transmitter-receiver antenna pairs exist between the transmitter and receiver, capturing a more granular time and frequency structure. For the widely used Intel 5300 commercial WiFi card, different antennas share the same radio frequency oscillator, meaning their time-varying phase offsets are identical. Consequently, the term $e^{-j2\pi\Delta ft}$ remains consistent across different antennas. To address sensing challenges, many researchers employ the ratio of CSI readings from two antennas, known as the CSI ratio. This approach not only eliminates significant noise but also extends the sensing range [46]–[49].

$$\begin{aligned} H_r(f,t) &= \frac{H_1(f,t)}{H_2(f,t)} \\ &= \frac{e^{-j2\pi\Delta ft}(H_{s,1}(f,t) + H_{d,1}(f,t))}{e^{-j2\pi\Delta ft}(H_{s,2}(f,t) + H_{d,2}(f,t))} \\ &= \frac{H_{s,1}(f,t) + \sum_{k=1}^{p_d} \alpha_{k,1}(f,t) e^{-j2\pi f \tau_k(t)}}{H_{s,2}(f,t) + \sum_{k=1}^{p_d} \alpha_{k,2}(f,t) e^{-j2\pi f \tau_k(t)}}, \end{aligned} \quad (7)$$

where $H_1(f,t)$ and $H_2(f,t)$ denote the CSI measurements from two antennas, respectively. As highlighted in [46], the division operation effectively eliminates random phase shifts and pulse amplitude noise because the power ratio across antennas on the same receiver remains consistent, and the clocks are synchronized. Orthogonal Frequency Division Multiplexing (OFDM), a widely adopted Multi-Carrier Modulation (MCM) technology, offers efficient spectrum utilization and robust resistance to multipath fading. CSI, derived from OFDM decoding, captures detailed propagation path states for each subcarrier, including amplitude and phase variations across individual subcarriers. Consequently, the dimensions of the CSI ratio received by the receiver can be expressed as $N_t \times N_r \times N_s \times T$, where N_t and N_r represent the number of transmitting and receiving antennas, respectively, N_s denotes the number of subcarriers, and T corresponds to the time dimension. To design a deep learning model for adaptively extracting gesture-related feature representations, CSI signals, which differ significantly from conventional data types such as images or text, must first be transformed into a format suitable for deep learning architectures. Building on the proven success of convolutional neural networks (CNNs) in computer vision tasks, we follow the methodology of prior research [10], [21] by converting the phase information of the CSI ratio

into an image format during data visualization. First, extract the phase values (ranging from $-\pi$ to $+\pi$) of 30 subcarriers at consecutive time-domain sampling points, and construct a two-dimensional phase matrix with the time dimension as the horizontal axis and the subcarrier index as the vertical axis. Then, use a gradient color coding (where blue and orange correspond to $-\pi$ and $+\pi$ phase values, respectively) to generate a single-device heatmap. Finally, concatenate the data from six sets of devices along the vertical axis to form a three-dimensional tensor [C, 6*N, T], where the dimension C=3 represents the channels of the generated image, N is the number of subcarriers for a single device, and T is the time duration. The specific CSI ratio phase image is shown in Figure 3.

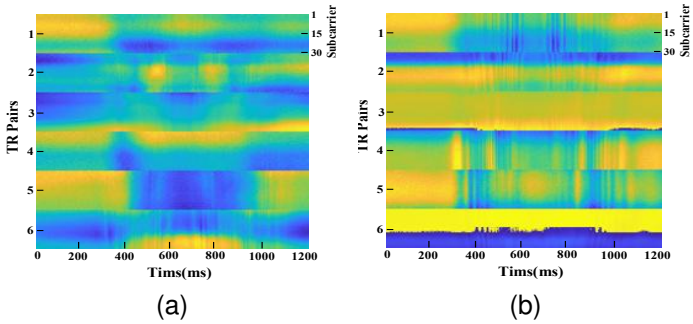


Fig. 3. Phase images for different gestures: (a) represents sweep, and (b) represents clap.

C. SFDAGR Framework

In this section, we detail the proposed SFDAGR framework, which enables the extraction of discriminative gesture-related features in cross-domain settings using only a model trained on the source domain, thereby achieving accurate gesture recognition. First, we employ a backbone network (e.g., a CNN) to extract rough semantics from the phase images of the CSI ratio. These semantics, which are not yet refined to be environment-independent, encompass both domain-related and gesture-related features. Next, the extracted rough semantics are processed through an attraction-dispersion network, which encourages the network to focus on domain-independent features—those solely related to gesture actions. Finally, we apply local neighbor uncertainty estimation to enhance the aggregation of nearest neighbors in the feature space, further strengthening the network's ability to learn domain-independent features. Below, we provide a detailed description of the two key components within the SFDAGR framework.

[Attraction-Dispersion Network] Unlike cross-domain learning methods that have access to source domain data, real-world cross-domain scenarios often only provide a model trained on the source domain due to constraints such as data transmission security. To train a model capable of extracting domain-independent features in the target domain, the source domain model serves as the initialization for the target domain model, leveraging its existing feature extraction capabilities [50]. This implies that the structure and initial parameters of the backbone network in the cross-domain model are identical to those of the source domain model. Let x_i represent the i-th phase image of the CSI ratio in the target domain. After passing through the backbone network, the rough feature representation is denoted as f_i:

$$f_i = B(x_i), \tag{8}$$

where B denotes the backbone network. In Section III-C, we provide a theoretical analysis of how to enhance prediction consistency within the local neighborhood of the feature space while ensuring distinct predictions for distant features. This approach helps reduce the expected error of the generalization boundary in the target domain. To address the cross-domain problem, we focus on both local neighborhood features and distant features after obtaining the rough features in the target domain. First, we employ cosine similarity to select the K nearest neighbor samples of f_i from the target features. The cosine similarity is calculated as follows:

$$s_{i,j} = \frac{< f_i, f_j >}{\|f_i\|\|f_j\|}, j \in \mathcal{N}, \tag{9}$$

$$\mathcal{N} = \{n | n \neq i, \forall n \in 1, 2, ..., N_t\}, \tag{10}$$

where $s_{i,j}$ denotes the similarity between input f_i and f_j. We select the top K samples based on similarity (from high to low) to form the nearest neighbor sample set $\mathcal{S}$ for the input sample x_i. Samples not included in the nearest neighbor set are grouped into a non-nearest neighbor sample set, denoted as $\mathcal{T}$. Following the approach in previous work [51], we exclude only the input sample within the mini-batch as $\mathcal{T}$. To reduce computational complexity and memory consumption during network training, two memory systems are implemented to store all target features and their corresponding prediction results. The features and predictions computed in each mini-batch are used to update these memory systems. To enhance prediction consistency among nearest neighbors in the feature space while ensuring inconsistent predictions for distant samples, we construct the following losses based on $\mathcal{S}$ and $\mathcal{T}$:

$$\mathcal{L} = -log\frac{P(\mathcal{S}_i)}{P(\mathcal{T}_i)} = -log(\frac{\prod\limits_{j \in \mathcal{S}_i} \frac{e^{p_i^T p_j}}{\sum_{k=1}^{N_t} e^{p_i^T p_k}}}{\prod\limits_{j \in \mathcal{T}_i} \frac{e^{p_i^T p_j}}{\sum_{k=1}^{N_t} e^{p_i^T p_k}}}), \tag{11}$$

where p represents the output prediction of the corresponding sample. From Equation 11, it is evident that minimizing $\mathcal{L}$ enhances prediction consistency among nearest neighbor samples in the feature space while ensuring inconsistent predictions for distant samples, thereby reducing generalization error in the target domain. However, calculating $P(\mathcal{S}_i)$ and $P(\mathcal{T}_i)$ involves all target samples, which increases computational complexity.

This issue aligns with previous work [52], requiring the optimization of Equation 11 to derive an upper bound target.

$$\begin{aligned}\mathcal{L} &= -\sum_{j\in\mathcal{S}_i}[p_i^T p_j - log\sum_{k=1}^{N_t} e^{p_i^T p_k}] + \sum_{j\in\mathcal{T}_i}[p_i^T p_j - log\sum_{k=1}^{N_t} e^{p_i^T p_k}] \\ &= -\sum_{j\in\mathcal{S}_i} p_i^T p_j + \sum_{j\in\mathcal{T}_i} p_i^T p_j + N_{d_i} log(\sum_{k=1}^{N_t} e^{p_i^T p_k}) \\ &\leq -\sum_{j\in\mathcal{S}_i} p_i^T p_j + \sum_{j\in\mathcal{T}_i} p_i^T p_j + N_{d_i}(\sum_{k=1}^{N_t}\frac{p_i^T p_k}{N_t} + logN_t) \\ &\simeq -\sum_{j\in\mathcal{S}_i} p_i^T p_j + \sum_{j\in\mathcal{T}_i} p_i^T p_j + N_{d_i}(\sum_{k\in N_{\mathcal{T}_i}}\frac{p_i^T p_k}{N_{\mathcal{T}_i}} + logN_t) \\ &= -\sum_{j\in\mathcal{S}_i} p_i^T p_j + \frac{N_{\mathcal{S}_i}}{N_{\mathcal{T}_i}}\sum_{j\in\mathcal{T}_i} p_i^T p_j + N_{d_i} logN_t,\end{aligned} \tag{12}$$

where $N_{\mathcal{S}_i}$ and $N_{\mathcal{T}_i}$ represent the number of nearest neighbor samples and the number of remaining samples in the mini-batch, respectively, and N_{d_i} denotes the difference between $N_{\mathcal{S}_i}$ and $N_{\mathcal{T}_i}$. Consistent with [52], since $N_{\mathcal{S}_i}$ is smaller than $N_{\mathcal{T}_i}$, the third step of $\mathcal{L}$ calculation can be derived using Jensen's inequality. Thus, we obtain the following loss function:

$$\mathcal{L} = -\sum_{j\in\mathcal{S}_i} p_i^T p_j + \lambda\sum_{j\in\mathcal{T}_i} p_i^T p_j. \tag{13}$$

By minimizing $\mathcal{L}$, we enhance the prediction consistency of nearest neighbor samples and the prediction inconsistency of distant samples in the feature space, thereby reducing the generalization error in the target domain. However, when selecting nearest neighbor samples, the uncertainty in model predictions can lead to inconsistencies between the selected samples and the input sample predictions. To address this issue, we reweight the local neighborhood predictions by learning uncertainty weights for the nearest neighbors, gradually improving feature aggregation for nearby samples.

[Local Neighbor Uncertainty Estimation] To enhance the uncertainty weighting (i.e., selection confidence) for the chosen nearest neighbor samples, we employ an entropy-based uncertainty estimation approach. The underlying principle can be summarized as follows: When the prediction of an input sample aligns with its nearest neighbor, the system yields low entropy, indicating high confidence (low uncertainty). Conversely, prediction discrepancies between the input sample and its neighbors result in high entropy, reflecting greater uncertainty.

Based on this principle, we develop a weighting mechanism that assigns uncertainty weights to nearest neighbor samples through entropy calculation of their combined probability distributions. Specifically, for a given input sample x_i with prediction probability p_i, we first compute its average probability distribution with each of the K nearest neighbors:

$$p'_{ij} = (p_i + p_{ij})/2, j = 1, 2, ..., K, \tag{14}$$

where p_{ij} represents the prediction probability of the j-th nearest neighbor. The entropy of this averaged distribution p'_{ij} is then calculated as:

$$H(p'_{ij}) = -\sum_{c=1}^{C} p'_{ijc} log p'_{ijc}, \tag{15}$$

where p'_{ijc} represents the output probability corresponding to class c. This entropy value serves as the quantitative measure for determining the uncertainty weight, effectively balancing the contribution of each nearest neighbor sample in the final prediction. Then we can obtain the uncertainty weight for each nearest neighbor sample:

$$\omega_j = e^{-(\frac{H(p'_{ij})}{logC})}, \tag{16}$$

where $\frac{H(p'_{ij})}{logC}$ represents rescaling the entropy by its maximum value, the purpose of calculating the negative exponential weight is consistent with [51], aiming to place more weight on low entropy values and reduce the weight on high entropy values. Combining Equation 13 and 16, we can obtain the final objective loss function as follows:

$$\mathcal{L} = -\sum_{j\in\mathcal{S}_i} p_i^T(\omega_j p_j) + \lambda\sum_{j\in\mathcal{T}_i} p_i^T p_j. \tag{17}$$

We summarize the process of the SFDAGR framework in Algorithm 1, wherein the resolution of the entire network parameters can be achieved in an end-to-end manner through the standard backpropagation algorithm [53].

Algorithm 1: SFDAGR Framework

Input: /* The trained source domain model and the target domain data $\mathcal{D}_t$ */
Output: /* Optimized model */
Build memory bank storing all target features and predictions;
while *Adaptation* **do**
  Sample mini-batch data from $\mathcal{D}_t$ and Update memory bank;
  For each feature f_i in mini-batch data, retrieve K-nearest neighbors ($\mathcal{S}_i$) and their predictions from memory bank;
  Calculate the entropy of the average probability distribution of p_i with each of its K nearest neighbors;//Eq.(15)
  Compute the uncertainty weight for each nearest neighbor sample;//Eq.(16)
  Update model by minimizing Eq.(17).

V. EXPERIMENTAL EVALUATION

A. Experiment Setting

1) Experimental Datasets: To evaluate the effectiveness of our model for source-free domain adaptation gesture recognition, we conduct extensive experiments on typical publicly available WiFi-based datasets (Widar3.0 and XRF55).

Widar3.0 [16] is one of the most widely used datasets for WiFi-based gesture recognition. The dataset collection

TABLE I
The detailed descriptions of Widar3.0 dataset

Environments	No. of Users	Gestures	No. of Locations	No. of Orientations	No. of Samples
1st (Classroom)	9	1: Push Pull; 2: Sweep; 3: Clap; 4:Slide; 5: Draw-O(Horizontal); 6: Draw-Zigzag(Horizontal); 7: Draw-N(Horizontal); 8: Draw-Triangle(Horizontal); 9: Draw-Rectangle(Horizontal);	5	5	10125
2nd (Hall)	4	1: Push Pull; 2: Sweep; 3: Clap; 4:Slide; 5: Draw-O(Horizontal); 6: Draw-Zigzag(Horizontal);	5	5	3000
3rd (Office)	4	1: Push Pull; 2: Sweep; 3: Clap; 4:Slide; 5: Draw-O(Horizontal); 6: Draw-Zigzag(Horizontal);	5	5	3000

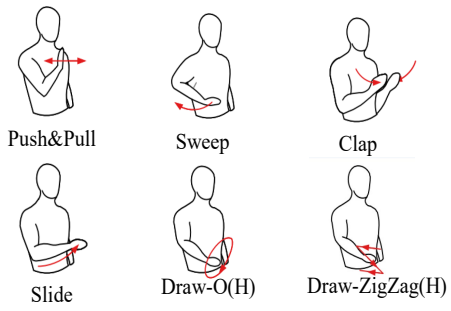


Fig. 4. The gesture sketches evaluated in the experiment (from Widar3.0 [16]).

TABLE II
DETAILS OF THE MODEL ARCHITECTURE FOR SOURCE-FREE DOMAIN ADAPTATION GESTURE RECOGNITION.

Layer	Output Size	Detailed Configuration
cov1	112×112	$7 \times 7, 64$
cov2	56×56	3×3maxpool, $\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 2$
cov3	28×28	$\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 2$
cov4	14×14	$\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 2$
cov5	7×7	$\begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \times 2$
fc1	512	average pool,fc
fc2	256	fc
fc3	6/8	fc

device consists of one transmitter and six receivers, each is equipped with an Intel 5300 wireless card. Each receiver has three antennas arranged in a line. The operating frequency of the device is 5.825 GHz, with a sampling rate of 1000 packets per second. Widar3.0 mainly consists of two types of datasets. Firstly, there are 12,000 gesture samples commonly used in human-computer interaction (16 users × 5 locations × 5 orientations × 6 gestures × 5 instances). Secondly, there are 5,000 gesture samples representing digits 0-9 on a horizontal plane (2 users × 5 locations × 5 orientations × 10 gestures × 5 instances). To ensure a fair comparison, we only utilized 4,500 gesture samples (9 users × 5 locations × 5 orientations × 6 gestures × 5 instances) from the Widar3.0 dataset for source-free domain adaptation gesture recognition. The sketches of the gestures are shown in Figure 4. The detailed descriptions of Widar3.0 dataset are provided in Table I.

XRF55 [54] is a large multi-radio frequency dataset for human action analysis, consisting of 42.9K samples of 55 action categories related to human-computer interaction collected from 39 users. To verify the effectiveness of the proposed Wi-SFDAGR, we selected only the samples related to gesture actions collected using WiFi devices (e.g., drawing a circle, drawing a cross, pushing, pulling, swiping left, swiping right, swiping up, and swiping down) for the experiments. Among them, 30 users in environment 1 repeated the gestures 20 times within 5 seconds, and the remaining 9 users repeated the gestures 20 times within 5 seconds in the other three environments, totaling 6,240 samples.

2) Implementation Details: In our implementation details, data preprocessing is done using Matlab, while model training is conducted using Python tools. In Wi-SFDAGR, we employ ResNet-18 [55] as the backbone network, which is pre-trained on the large-scale ImageNet dataset. This pre-training approach is a dominant paradigm for initializing the backbone of object detection and segmentation models [56]. Previous research has demonstrated that pre-training can significantly alleviate the issue of models being unable to learn parameters effectively due to limited training datasets [10], [57]. Specifically, we utilize the same network architecture as AaD [52], where the final part of the network consists of fully connected layers, batch normalization, and weighted normalization. The specific network parameter configuration is shown in Table II. In this table, the "Layer" column corresponds to the name of each layer and has no impact on the model architecture. "fc" indicates the execution of a fully connected layer, and "Output Size" specifies the dimension of the output features. The details of cov1 indicate that 64 convolutional kernels of size 7×7 are used for convolution. The first step of cov2 represents pooling with a 7×7 kernel, followed by the same notation. The final fc layer outputs a vector of length 6 or 8, corresponding to the six or eight gestures in Widar 3.0 or XRF55, respectively. We use SGD with a momentum of 0.9, a batch size of 20, a learning rate of 0.1, and train for 40 epochs. For λ, we set it as $\lambda = (1+10*\frac{iter}{max\_iter})^{-\beta}$, where the decay factor β controls the decay rate. In our implementation, β is set to 1.

B. Overall Performance

We first evaluate the overall performance of Wi-SFDAGR in cross-location, cross-orientation, and cross-environment scenarios. Table III shows the gesture recognition results on the Widar3.0 and XRF55 datasets, and we also compare

TABLE III
THE ACCURACY OF WI-SFDAGR UNDER CROSS-LOCATION, CROSS-ORIENTATION, AND CROSS-ENVIRONMENT SETTINGS IN THE WIDAR3.0 DATASET AND THE XRF55 DATASET.

Method	SF-UDA	Widar3.0			Mean	XRF55
		cross-location	cross-orientation	cross-environment		cross-environment
Widar3.0 [16]	✗	90.48%	81.58%	83.30%	85.12%	-
ImgFi [58]	✗	39.58%	38.12%	40.37%	39.36%	31.90%
EI [59]	✗	73.33%	79.70%	-	-	-
WiSR [60]	✗	67.73%	69.74%	52.77%	63.41%	26.66%
Recurrent ConFormer [61]	✗	73.84%	85.88%	50.38%	70.03%	16.54%
THAT [62]	✗	71.56%	81.76%	49.71%	67.68%	23.23%
WiHF [3]	✗	91.22%	80.64%	-	-	-
WiGNN [63]	✗	95.20%	93.30%	-	-	-
WiGRUNT [57]	✗	97.08%	93.39%	95.36%	95.28%	55.92%
WiDual [64]	✗	97.39%	94.87%	-	-	-
AaD [52]	✓	95.90%	95.38%	93.20%	94.83%	55.64%
Ours	✓	97.30%	97.17%	95.52%	96.66%	57.99%

them with several other domain adaptation methods (such as Widar3.0 [16], ImgFi [58], EI [59], WiSR [60], Recurrent ConFormer [61], THAT [62], WiHF [3], WiGNN [63], WiGRUNT [57], WiDual [64], and AaD [52]). All these results use open-source code implementations with the same dataset settings as in this paper (The EI [59] result is derived from [65], and the results of WiHF, WiGNN, and WiDual are derived from their respective reported results.). Among these comparative methods, only the AaD and Wi-SFDAGR methods do not require the assistance of source domain data during the adaptation phase, relying solely on the pre-trained model from the source domain. Other methods need the help of source domain data during the adaptation phase. As can be seen from Table III, in the Widar3.0 dataset, Wi-SFDAGR achieves gesture recognition accuracies of 97.30%, 97.17%, and 95.52% under cross-location, cross-orientation, and cross-environment settings, respectively, with an average performance of 96.66%. In the XRF55 dataset, Wi-SFDAGR achieves a gesture recognition accuracy of 57.99% under cross-environment settings, which is higher than the recognition rates of existing methods except for WiDual. It performs comparably to WiDual, but WiDual requires the use of source domain data during training. Additionally, we also observed an interesting phenomenon: regardless of the method used, the performance across environments is lower than that in cross-location and cross-orientation settings. We can attribute this phenomenon to the background knowledge of wireless perception, as different environments imply different indoor layouts, resulting in greater influence on the CSI received by the receivers, thereby making it more difficult for the network model to capture distinctive features related to gestures in cross-environment scenarios.

To observe the recognition accuracy of different gestures and the misjudgment rate between them, we present the confusion matrix for test location 1, test orientation 1, and test environment 3 in the Widar3.0 dataset, as shown in Figure 5. From Figure 5, it can be seen that in the cross-location test, the gestures “o” and “zigzag” achieve the highest recognition accuracy, while the gesture “clap” has the lowest accuracy. In the cross-orientation test, the gestures “push” and “o” show the highest recognition accuracy, while the gesture “clap” again has the lowest accuracy. In the cross-environment test, the gestures “o” and “zigzag” also achieve the highest recognition accuracy. Additionally, we observe an interesting phenomenon: the gesture “clap” is most frequently misjudged as the gesture “slide”. The reason for this can be attributed to human body dynamics, as the motion trajectories of “slide” and “clap” share more inclined and sliding characteristics, making it challenging for the model to distinguish between them.

C. Ablation Study

In Table IV, we report the results of the ablation experiments for each individual component in Wi-SFDAGR for cross-location experiments on the Widar3.0 dataset (all results are from tests on location 1, consistent with Figure 5). The first row in Table IV tests the baseline performance, which directly evaluates the target domain data using the pre-trained source domain model. The second row tests the performance of the attraction-dispersion network. The third row tests the performance of local neighbor-weighted uncertainty weighting. Initially, the baseline performance achieved the lowest accuracy of 95.78%. By incorporating the attraction-dispersion network, gesture recognition performance improved by 0.45% (as shown in the second row of Table IV). By further considering the uncertainty weighting of local neighbor points, the gesture recognition model’s performance improved by 1.1%.

Observations from the ablation experiments indicate that the attraction-dispersion network and local neighbor-weighted uncertainty play crucial roles in gesture recognition. Further

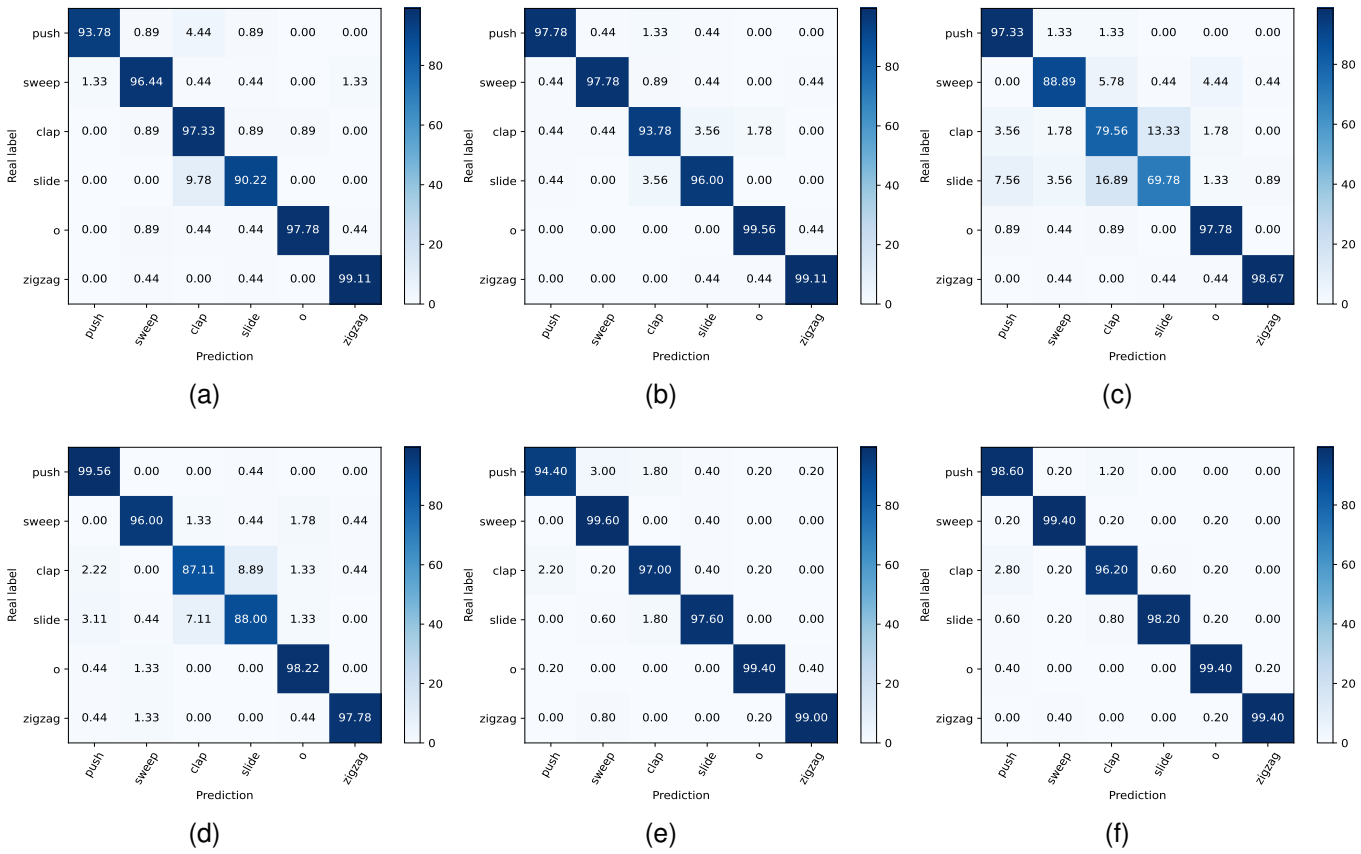


Fig. 5. The confusion matrices of our method and baseline method in the cross-location (tested at location 1), cross-orientation (tested in orientation 1), and cross-environment (tested in environment 3) settings: (a) Baseline method in the cross-location (95.78%); (b) Wi-SFDAGR in the cross-location (97.33%); (c) Baseline method in the cross-orientation (88.67%); (d) Wi-SFDAGR in the cross-orientation (94.44%); (e) Baseline method in the cross-environment (97.83%); (f) Wi-SFDAGR in the cross-environment (98.53%).

analysis reveals that the attraction-dispersion network significantly enhances the model's generalization capability by optimizing intra-class and inter-class distances in the feature space. Specifically, this network improves the model's ability to capture key gesture-related features in cross-domain scenarios by attracting samples of the same class and dispersing samples of different classes. Additionally, the local neighbor-weighted uncertainty mechanism dynamically adjusts sample weights, effectively reducing the impact of noisy data on model training and further improving recognition accuracy. These experimental results not only validate the effectiveness of each component but also reveal their synergistic effects in enhancing model performance. The combination of the attraction-dispersion network and the local neighbor-weighted uncertainty mechanism enables Wi-SFDAGR to excel in cross-domain gesture recognition tasks. Finally, through comparative experiments with different backbone networks, we also discover that regardless of whether ResNet-18 or ResNet-50 is employed, both the ADN and LNUE components consistently maintain their effectiveness.

D. t-SNE Feature Visualization

To intuitively demonstrate the effectiveness of our proposed method, i.e., Wi-SFDAGR ability to extract distinguishing features related to gestures, we employ t-SNE [66] to visualize the feature distribution of the test set in the Widar3.0 dataset. As shown in Figure 6, Figure 6a and 6b represent the feature distribution across locations for Wi-SFDAGR and the baseline method, Figure 6c and 6d represent the feature distribution across orientations for Wi-SFDAGR and the baseline method, and Figure 6e and 6f represent the feature distribution across environments for Wi-SFDAGR and the baseline method. Similar to the confusion matrix displayed in the overall performance, we only select data from location 1, orientation 1, and environment 3 as the target domain. The baseline method refers to directly testing the trained source domain model in the target domain. From Figure 6, we can observe some interesting phenomena: 1. Despite achieving the highest accuracy in the cross-orientation setting, Wi-SFDAGR

TABLE IV
ABLATION STUDIES OF THE SUB-COMPONENTS ARE CONDUCTED, IN WHICH WI-SFDAGR IS ASSESSED BASED ON THE CLASSIFICATION ACCURACY (%) ACROSS DIFFERENT LOCATIONS AND MEASURED USING VARIOUS BACKBONE NETWORKS. ADN:ATTRACTION-DISPERSION NETWORK, LNUE:LOCAL NEIGHBOR UNCERTAINTY ESTIMATION.

ADN	LNUE	Accuracy (ResNet-18)	Accuracy (ResNet-50)
✗	✗	95.78%	96.14%
✓	✗	96.23%	97.38%
✓	✓	97.33%	98.41%

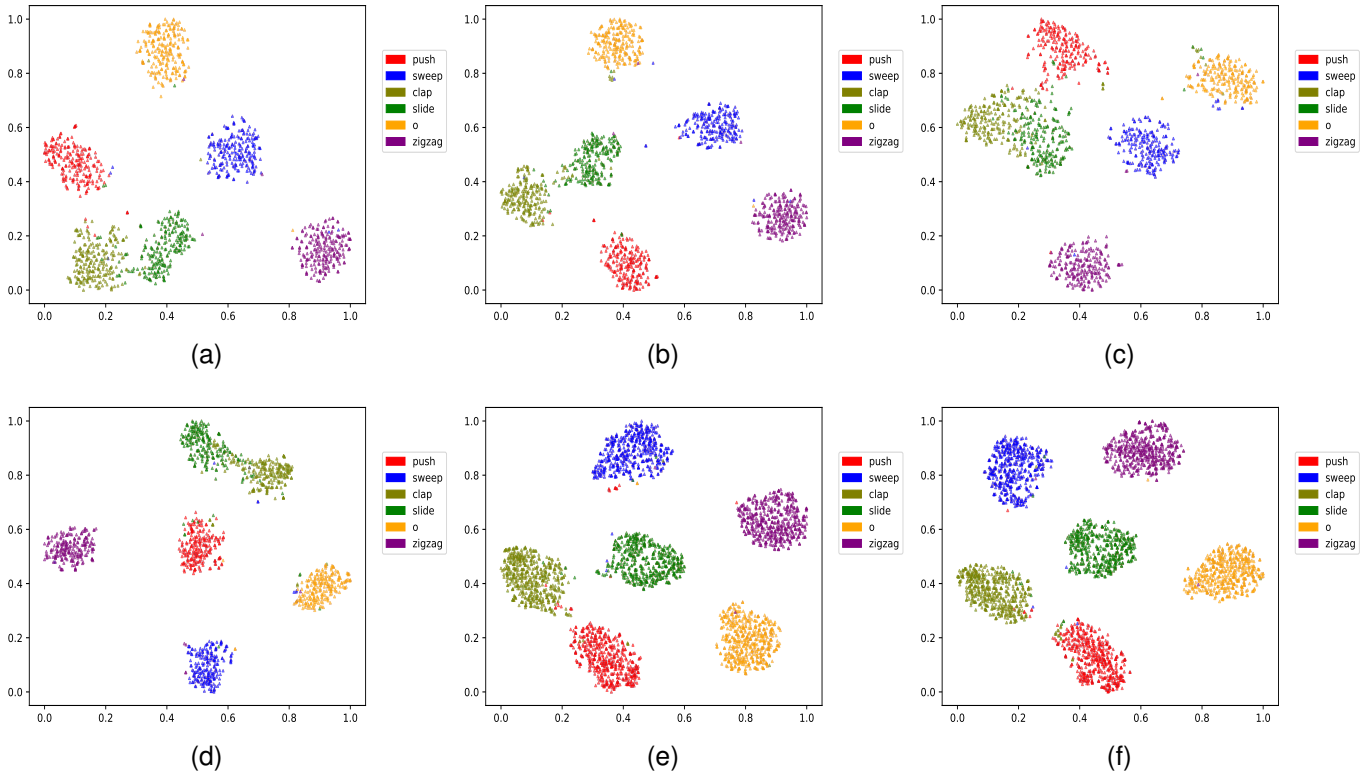


Fig. 6. The feature visualization of our method and baseline method in the cross-location (tested at location 1), cross-orientation (tested in orientation 1), and cross-environment (tested in environment 3) settings: (a) Baseline method in the cross-location; (b) Wi-SFDAGR in the cross-location; (c) Baseline method in the cross-orientation; (d) Wi-SFDAGR in the cross-orientation; (e) Baseline method in the cross-environment; (f) Wi-SFDAGR in the cross-environment.

exhibits a significantly larger intra-class variance compared to the cross-location and cross-environment settings. 2. Without domain adaptation, gesture ”clap” and ”slide” are prone to confusion, consistent with the findings from the confusion matrix in overall performance. This is because these two gestures are highly similar in terms of human behavioral kinetics. 3. Regardless of whether Wi-SFDAGR or the baseline method is used, gesture ”o” and ”zigzag” consistently achieve the highest accuracy. This is because these two gestures are intuitively easier to distinguish compared to other gestures.

VI. Conclusion and Discussion

In this paper, we design a source-free domain adaptation gesture recognition framework, Wi-SFDAGR, based on channel state information obtained from WiFi. During the adaptation to unlabeled test data, the source data is no longer available, instead, a pre-trained source domain model is used. To extract environment-independent robust gesture features, we first theoretically analyze the impact of enhancing the prediction consistency of local neighborhood samples and weakening the prediction consistency of features that are far apart in the feature space on the generalization boundary of source-free domain adaptation. Then, we employ an attraction-dispersion network to enhance the prediction consistency of features that are close in the feature space while weakening the prediction consistency of features that are far apart. To the best of our knowledge, this is the first work to address source-free domain adaptation for WiFi-based gesture recognition. Additionally, we reweight local neighborhood predictions by estimating the uncertainty of nearest neighbor samples, gradually enhancing the feature aggregation of nearby samples. Finally, extensive experiments on the publicly available Widar3.0 dataset verify the effectiveness of Wi-SFDAGR.

In the future, WiFi-based gesture recognition research can advance through two key directions. First, the current datasets are limited in number and often focus on single-device, simple scenarios, making it challenging to evaluate algorithms in complex environments. Developing a large-scale open-source dataset that integrates multiple devices and covers diverse complex scenarios, such as various room structures, interference conditions, and dynamic environments, is essential. Such a dataset will provide a unified benchmark and facilitate research on cross-device and cross-environment generalization. Second, exploring the combination of open-set learning and incremental learning strategies can enhance models’ ability to handle open categories and dynamic environments. Open-set learning helps identify unknown gestures, while incremental learning allows models to adapt to new categories or environments over time. These advancements will improve the practical deployment of systems and foster applications in smart homes, health monitoring, and beyond. By optimizing datasets and enhancing model generalization, future research will drive further progress in wireless sensing technology.

Acknowledgments

This work is supported by the Guizhou Provincial Basic Research Program(Natural Science) (Qiankehejichu-Youth[2024]345), the Guizhou Provincial Basic Research

Program(Natural Science) (Qiankehejichu-MS[2025]277), the National Natural Science Foundation of China (Grant No. 62462015, No. 62072097, and No. U22A2026), the Young Elite Scientist Sponsorship Program By Gast (Grant No. GASTYESS202429), the Anhui Province Science Foundation for Youths (Grant No. 2308085QF230), and the Qiankehe Platform Talents (Grant No. BQW[2024]015). We would like to thank the editors and anonymous reviewers for their insightful comments and constructive feedback.

REFERENCES

[1] H. Abdelnasser, M. Youssef, and K. A. Harras, "Wigest: A ubiquitous wifi-based gesture recognition system," in *2015 IEEE conference on computer communications (INFOCOM)*. IEEE, 2015, pp. 1472–1480.

[2] H. F. T. Ahmed, H. Ahmad, and C. Aravind, "Device free human gesture recognition using wi-fi csi: A survey," *Engineering Applications of Artificial Intelligence*, vol. 87, p. 103281, 2020.

[3] C. Li, M. Liu, and Z. Cao, "Wihf: Enable user identified gesture recognition with wifi," in *IEEE INFOCOM 2020-IEEE Conference on Computer Communications*. IEEE, 2020, pp. 586–595.

[4] A. Virmani and M. Shahzad, "Position and orientation agnostic gesture recognition using wifi," in *Proceedings of the 15th Annual International Conference on Mobile Systems, Applications, and Services*, 2017, pp. 252–264.

[5] N. M. Mahmoud, H. Fouad, and A. M. Soliman, "Smart healthcare solutions using the internet of medical things for hand gesture recognition system," *Complex & intelligent systems*, vol. 7, pp. 1253–1264, 2021.

[6] A. Li, E. Bodanese, S. Poslad, T. Hou, K. Wu, and F. Luo, "A trajectory-based gesture recognition in smart homes based on the ultrawideband communication system," *IEEE Internet of Things Journal*, vol. 9, no. 22, pp. 22 861–22 873, 2022.

[7] K. M. Sagayam and D. J. Hemanth, "Hand posture and gesture recognition techniques for virtual reality applications: a survey," *Virtual Reality*, vol. 21, pp. 91–107, 2017.

[8] X. Zheng, J. Wang, L. Shangguan, Z. Zhou, and Y. Liu, "Smokey: Ubiquitous smoking detection with commercial wifi infrastructures," in *IEEE INFOCOM 2016-The 35th Annual IEEE International Conference on Computer Communications*. IEEE, 2016, pp. 1–9.

[9] X. Zhang, Y. Gu, H. Yan, Y. Wang, M. Dong, K. Ota, F. Ren, and Y. Ji, "Wital: A cots wifi devices based vital signs monitoring system using nlos sensing model," *IEEE Transactions on Human-Machine Systems*, 2023.

[10] 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.

[11] X. Zheng, K. Yang, J. Xiong, L. Liu, and H. Ma, "Pushing the limits of wifi sensing with low transmission rates," *IEEE Transactions on Mobile Computing*, 2024.

[12] H. Yan, Y. Zhang, Y. Wang, and K. Xu, "Wiact: A passive wifi-based human activity recognition system," *IEEE Sensors Journal*, vol. 20, no. 1, pp. 296–305, 2019.

[13] J. Huang, B. Liu, C. Miao, X. Zhang, J. Liu, L. Su, Z. Liu, and Y. Gu, "Phyfinatt: An undetectable attack framework against phy layer fingerprint-based wifi authentication," *IEEE Transactions on Mobile Computing*, 2023.

[14] L. Xu, X. Zheng, X. Du, L. Liu, and H. Ma, "Wicamera: Vortex electromagnetic wave-based wifi imaging," *IEEE Transactions on Mobile Computing*, 2024.

[15] X. Leiyang, Z. Xiaolong, and L. Liang, "Vortex em wave-based rotation speed monitoring on commodity wifi," *Chinese Journal of Electronics*, 2024.

[16] Y. Zheng, Y. Zhang, K. Qian, G. Zhang, Y. Liu, C. Wu, and Z. Yang, "Zero-effort cross-domain gesture recognition with wi-fi," in *Proceedings of the 17th annual international conference on mobile systems, applications, and services*, 2019, pp. 313–325.

[17] K. Niu, F. Zhang, X. Wang, Q. Lv, H. Luo, and D. Zhang, "Understanding wifi signal frequency features for position-independent gesture sensing," *IEEE Transactions on Mobile Computing*, vol. 21, no. 11, pp. 4156–4171, 2021.

[18] H. Zou, J. Yang, Y. Zhou, and C. J. Spanos, "Joint adversarial domain adaptation for resilient wifi-enabled device-free gesture recognition," in *2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)*. IEEE, 2018, pp. 202–207.

[19] B.-B. Zhang, D. Zhang, Y. Hu, and Y. Chen, "Unsupervised domain adaptation for wifi gesture recognition," in *2023 IEEE Wireless Communications and Networking Conference (WCNC)*. IEEE, 2023, pp. 1–6.

[20] G. Tong, Y. Li, H. Zhang, and N. Xiong, "A fine-grained channel state information-based deep learning system for dynamic gesture recognition," *Information Sciences*, vol. 636, p. 118912, 2023.

[21] Y. Gu, H. Yan, X. Zhang, Y. Wang, J. Huang, Y. Ji, and F. Ren, "Attention-based gesture recognition using commodity wifi devices," *IEEE Sensors Journal*, vol. 23, no. 9, pp. 9685–9696, 2023.

[22] Y. Fang, P.-T. Yap, W. Lin, H. Zhu, and M. Liu, "Source-free unsupervised domain adaptation: A survey," *Neural Networks*, p. 106230, 2024.

[23] S. Tan and J. Yang, "Wifinger: Leveraging commodity wifi for fine-grained finger gesture recognition," in *Proceedings of the 17th ACM international symposium on mobile ad hoc networking and computing*, 2016, pp. 201–210.

[24] R. Gao, W. Li, Y. Xie, E. Yi, L. Wang, D. Wu, and D. Zhang, "Towards robust gesture recognition by characterizing the sensing quality of wifi signals," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 6, no. 1, pp. 1–26, 2022.

[25] Y. Zhang, B. Yuan, Z. Yang, Z. Li, and X. Liu, "Wi-nn: Human gesture recognition system based on weighted knn," *Applied Sciences*, vol. 13, no. 6, p. 3743, 2023.

[26] J. Yang, H. Zou, Y. Zhou, and L. Xie, "Learning gestures from wifi: A siamese recurrent convolutional architecture," *IEEE Internet of Things Journal*, vol. 6, no. 6, pp. 10 763–10 772, 2019.

[27] W. Meng, X. Chen, W. Cui, and J. Guo, "Wihgr: A robust wifi-based human gesture recognition system via sparse recovery and modified attention-based bgru," *IEEE Internet of Things Journal*, vol. 9, no. 12, pp. 10 272–10 282, 2021.

[28] S. D. Regani, B. Wang, and K. R. Liu, "Wifi-based device-free gesture recognition through-the-wall," in *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2021, pp. 8017–8021.

[29] C. Chen, G. Zhou, and Y. Lin, "Cross-domain wifi sensing with channel state information: A survey," *ACM Computing Surveys*, vol. 55, no. 11, pp. 1–37, 2023.

[30] H. Kang, Q. Zhang, and Q. Huang, "Context-aware wireless-based cross-domain gesture recognition," *IEEE Internet of Things Journal*, vol. 8, no. 17, pp. 13 503–13 515, 2021.

[31] R. Gao, M. Zhang, J. Zhang, Y. Li, E. Yi, D. Wu, L. Wang, and D. Zhang, "Towards position-independent sensing for gesture recognition with wi-fi," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 5, no. 2, pp. 1–28, 2021.

[32] C. Feng, N. Wang, Y. Jiang, X. Zheng, K. Li, Z. Wang, and X. Chen, "Wi-learner: Towards one-shot learning for cross-domain wi-fi based gesture recognition," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 6, no. 3, pp. 1–27, 2022.

[33] J. Su, Q. Mao, Z. Liao, Z. Sheng, C. Huang, and X. Zhang, "A real-time cross-domain wi-fi-based gesture recognition system for digital twins," *IEEE Journal on Selected Areas in Communications*, 2023.

[34] X. Li, L. Chang, F. Song, J. Wang, X. Chen, Z. Tang, and Z. Wang, "Crossgr: Accurate and low-cost cross-target gesture recognition using wi-fi," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 5, no. 1, pp. 1–23, 2021.

[35] X. Zhang, C. Tang, K. Yin, and Q. Ni, "Wifi-based cross-domain gesture recognition via modified prototypical networks," *IEEE Internet of Things Journal*, vol. 9, no. 11, pp. 8584–8596, 2021.

[36] P. Oza, V. A. Sindagi, V. V. Sharmini, and V. M. Patel, "Unsupervised domain adaptation of object detectors: A survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023.

[37] S. Zhao, X. Yue, S. Zhang, B. Li, H. Zhao, B. Wu, R. Krishna, J. E. Gonzalez, A. L. Sangiovanni-Vincentelli, S. A. Seshia *et al.*, "A review of single-source deep unsupervised visual domain adaptation," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 33, no. 2, pp. 473–493, 2020.

[38] P. Ge, C.-X. Ren, X.-L. Xu, and H. Yan, "Unsupervised domain adaptation via deep conditional adaptation network," *Pattern Recognition*, vol. 134, p. 109088, 2023.

[39] J. Zhu, H. Bai, and L. Wang, "Patch-mix transformer for unsupervised domain adaptation: A game perspective," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 3561–3571.

[40] J. Na, H. Jung, H. J. Chang, and W. Hwang, "Fixbi: Bridging domain spaces for unsupervised domain adaptation," in *Proceedings of the*

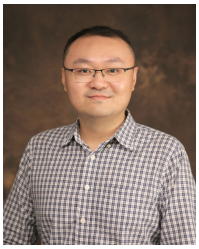
IEEE/CVF conference on computer vision and pattern recognition, 2021, pp. 1094–1103.
[41] Z. Du, J. Li, H. Su, L. Zhu, and K. Lu, "Cross-domain gradient discrepancy minimization for unsupervised domain adaptation," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2021, pp. 3937–3946.
[42] D. Mekhazni, A. Bhuiyan, G. Ekladious, and E. Granger, "Unsupervised domain adaptation in the dissimilarity space for person re-identification," in *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXVII 16*. Springer, 2020, pp. 159–174.
[43] M. Mohri, A. Rostamizadeh, and A. Talwalkar, *Foundations of machine learning*. MIT press, 2018.
[44] T. Chu, Y. Liu, J. Deng, W. Li, and L. Duan, "Denoised maximum classifier discrepancy for source-free unsupervised domain adaptation," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 36, no. 1, 2022, pp. 472–480.
[45] N. Costa and S. Haykin, *Multiple-input multiple-output channel models: theory and practice*. John Wiley & Sons, 2010.
[46] Y. Zeng, D. Wu, J. Xiong, E. Yi, R. Gao, and D. Zhang, "Farsense: Pushing the range limit of wifi-based respiration sensing with csi ratio of two antennas," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 3, no. 3, pp. 1–26, 2019.
[47] Y. Zeng, D. Wu, J. Xiong, and D. Zhang, "Boosting wifi sensing performance via csi ratio," *IEEE Pervasive Computing*, vol. 20, no. 1, pp. 62–70, 2020.
[48] Y. Zeng, D. Wu, J. Xiong, J. Liu, Z. Liu, and D. Zhang, "Multisense: Enabling multi-person respiration sensing with commodity wifi," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 4, no. 3, pp. 1–29, 2020.
[49] 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.
[50] Y. Zhang, Z. Wang, and W. He, "Class relationship embedded learning for source-free unsupervised domain adaptation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 7619–7629.
[51] M. Litrico, A. Del Bue, and P. Morerio, "Guiding pseudo-labels with uncertainty estimation for source-free unsupervised domain adaptation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 7640–7650.
[52] S. Yang, S. Jui, J. van de Weijer *et al.*, "Attracting and dispersing: A simple approach for source-free domain adaptation," *Advances in Neural Information Processing Systems*, vol. 35, pp. 5802–5815, 2022.
[53] Y. LeCun, D. Touresky, G. Hinton, and T. Sejnowski, "A theoretical framework for back-propagation," in *Proceedings of the 1988 connectionist models summer school*, vol. 1, 1988, pp. 21–28.
[54] F. Wang, Y. Lv, M. Zhu, H. Ding, and J. Han, "Xrf55: A radio frequency dataset for human indoor action analysis," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 8, no. 1, pp. 1–34, 2024.
[55] 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.
[56] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE conference on computer vision and pattern recognition*. Ieee, 2009, pp. 248–255.
[57] 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.
[58] C. Zhang and W. Jiao, "Imgfi: A high accuracy and lightweight human activity recognition framework using csi image," *IEEE Sensors Journal*, 2023.
[59] W. Jiang, C. Miao, F. Ma, S. Yao, Y. Wang, Y. Yuan, H. Xue, C. Song, X. Ma, D. Koutsonikolas *et al.*, "Towards environment independent device free human activity recognition," in *Proceedings of the 24th annual international conference on mobile computing and networking*, 2018, pp. 289–304.
[60] S. Liu, Z. Chen, M. Wu, C. Liu, and L. Chen, "Wisr: Wireless domain generalization based on style randomization," *IEEE Transactions on Mobile Computing*, 2023.
[61] M. Shang and X. Hong, "Recurrent conformer for wifi activity recognition," *IEEE/CAA Journal of Automatica Sinica*, vol. 10, no. 6, pp. 1491–1493, 2023.
[62] 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.
[63] Y. Chen and X. Huang, "Wignn: Wifi-based cross-domain gesture recognition inspired by dynamic topology structure," *IEEE Wireless Communications*, 2024.
[64] C. Cao, Y. Ding, M. Dai, W. Gong, and X. Zhao, "Real-time cross-domain gesture and user identification via cots wifi," *IEEE Transactions on Mobile Computing*, 2025.
[65] Y. Liu, A. Yu, L. Wang, B. Guo, Y. Li, E. Yi, and D. Zhang, "Unifi: A unified framework for generalizable gesture recognition with wi-fi signals using consistency-guided multi-view networks," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 7, no. 4, pp. 1–29, 2024.
[66] L. Van der Maaten and G. Hinton, "Visualizing data using t-sne." *Journal of machine learning research*, vol. 9, no. 11, 2008.

Huan Yan received the B.E. degree from Hefei University of Technology, China, in 2017, and his D.E. degree from the same university in 2023. He is currently a lecturer in the School of Big Data and Computer Science at Guizhou Normal University. His research interests include Wireless Security and Wireless Sensing.

Xiang Zhang received the B.E. degree from Hefei University of Technology, China, in 2017, and his D.E. degree from the same university in 2023. He was sponsored by China Scholarship Council (CSC) (from 2022.6 to 2023.6) for joint Ph.D. study supervised by Prof. Yusheng Ji (IEEE Fellow) at National Institute of Informatics, Japan. Currently, he is a postdoc with the School of Cyber Science and Technology, University of Science and Technology of China. His research interests include wireless sensing, affective computing and wireless security. In this area, he has published 20+ papers in international peer-reviewed journals and conferences, including IEEE S&P, TAFFC, TIFS, TMC, THMS, ACM Mobicom and MM. He is a TPC Member of ACM MM, IEEE ICME and Globecom. He is the recipient of IEEE SMC Society Andrew P. Sage Best Transactions Paper Award and IEEE HITC Distinguished Phd Dissertation Award.

Jinyang Huang is a lecturer at the School of Computer Science and Information Engineering, Hefei University of Technology (HFUT) and the Secretary-General of the Anhui Province Key Laboratory of Affective Computing and Advanced Intelligence Machine (led by Prof. Meng Wang (IEEE Fellow)). He obtained his Ph.D. in Computer Science and Technology from the School of Cyberspace Security, University of Science and Technology of China (USTC) in 2022. He was sponsored by China Scholarship Council (CSC) (from 2020.12.1 to 2021.11.31) for joint Ph.D. study supervised by Assoc. Prof. Lu Su from Purdue University and Chang Wen Chen from University at Buffalo, USA. His research interests include Multimodal Perception, Human-computer Interaction, Wireless Security, and Signal Processing. In this area, he has published 30 papers in international peer-reviewed journals and conferences, including ToN, TMC, TIFS, TAC, THMS, MobiCom, Infocom, ACM MM, and ECCV. He has served as a TPC member for conferences, including ACM MM, IEEE ICME, and Globecom, and has the honor of becoming ACM MM 2024 Outstanding Reviewers. He is a Guest Editor for Applied science. He is the recipient of IEEE HITC Distinguished PhD Dissertation Award.

Yuanhao Feng received the Ph.D. degree from the University of Science and Technology of China in 2023. He is currently a Post-Doctoral Fellow with the Department of Computing, The Hong Kong Polytechnic University. He has long-term cooperation with the Research Group of Dr. Wang Meng, Hefei University of Technology. He has published many top journals and conference papers, such as MobiCom, INFOCOM, and IEEE TRANSACTIONS ON MOBILE COMPUTING. His research interests include the AIoT, which includes wireless sensing and communication. He received the Outstanding Graduate Award from the School of Computer Science, USTC, in 2023.



Meng Li is an Associate Professor and Personnel Secretary at the School of Computer Science and Information Engineering, Hefei University of Technology (HFUT), China. He is also a Post-Doc Researcher at Department of Mathematics and HIT Center, University of Padua, Italy, where he is with the Security and PRIvacy Through Zeal (SPRITZ) research group led by Prof. Mauro Conti (IEEE Fellow). He obtained his Ph.D. in Computer Science and Technology from the School of Computer Science and Technology, Beijing Institute of Technology (BIT), China, in 2019. He was sponsored by ERCIM 'Alain Bensoussan' Fellowship Programme (from 2020.10.1 to 2021.3.31) to conduct Post-Doc research supervised by Prof. Fabio Martinelli at CNR, Italy. He was sponsored by China Scholarship Council (CSC) (from 2017.9.1 to 2018.8.31) for joint Ph.D. study supervised by Prof. Xiaodong Lin (IEEE Fellow) in the Broadband Communications Research (BBCR) Lab at University of Waterloo and Wilfrid Laurier University, Canada. His research interests include security, privacy, applied cryptography, blockchain, TEE, and Internet of Vehicles. In this area, he has published 83 papers in international peer-reviewed journals and conferences, including TIFS, TDSC, ToN, TMC, TKDE, TODS, TSC, COMST, ISSTA, MobiCom, ACISP, ICICS, SecureComm, TrustCom, ICC, and IPCCC. He is a Senior Member of IEEE, CIE, CIC, and CCF. He is an Associate Editor for IEEE TIFS, IEEE TNSM, and IEEE IoTJ. He has served as a TPC member for conference, including ICDCS, TrustCom, ICC, Globecom, HPCC, ICA3PP, and KSEM. He is the recipient of IEEE HITC Award for Excellence (Early Career Researcher).

Anzhi Wang received the Ph.D. degree in Computer Science and Technology from the College of Computer Science, Sichuan University, Chengdu, China, in 2017. From September 2021 to July 2022, he was a visiting scholar with the School of Informatics, Xiamen University, China. He is currently an associate professor with the School of Big Data and Computer Science, Guizhou Normal University, China. His current research interests mainly focus on the fields of computer vision, digital image processing and deep learning, more specifically, include salient object detection, camouflaged object detection, image dehazing, image super-resolution, and etc. He has led many national and provincial research projects, including those funded by the National Natural Science Foundation of China and the Natural Science Foundation of Guizhou Province. He has published over 40 papers in related international academic journals and conferences. He also serves as a member of CCF, CAAI, and CSIG, and serves as a reviewer for several international journals including IEEE-TIP, IEEE-TMM, Expert Systems With Applications, IEEE-SPL, EAAI, NCA, NPL, and for several various conferences including IJCAI, ICME, ICASSP, PRCV, ACAIT, and etc.

Weihua Ou received the PhD degree in Information and Communication Engineering from Huazhong University of Science and Technology, Wuhan, China in 2014. He worked as postdoc from 2016 to 2017 in University of Technology Sydney, Australia. His research interests include computer vision and pattern recognition. He has published more than 60 academic papers on top journals and conferences, including IEEE TMM, IEEE TNNLS, PR, and so on. He has served as Program Committee Member of top conferences, such as AAAI, CVPR.

Hongbing Wang received the PhD degree in computer science from Nanjing University, China. He is a professor from Guizhou Normal University. He served as a visiting scientist with CSIRO ICT Center, Australia, from Apr. 2009 to Mar. 2010. Prior to this, he has visited the Hongkong University and the Waterloo University during 2003-2008. His research interests include service computing, cloud computing, and big data. He published more than 50 refereed papers in international journals and conferences, e.g., the Journal of Web Semantics, the Journal of Systems and Software, the IEEE Transactions on Parallel and Distributed Systems, the IEEE Transactions on Services Computing, ICSOC, ICWS, SCC, etc.

Zhi Liu (Senior Member, IEEE) received the Ph.D. degree in informatics from the National Institute of Informatics. He is currently an Associate Professor with The University of Electro Communications. His research interests include video network transmission and mobile edge computing. He is an Editorial Board Member of Wireless Networks (Springer) and IEEE OPEN JOURNAL OF THE COMPUTER SOCIETY.