WiSR: Wireless Domain Generalization Based on Style Randomization

Shijia Liu, Zhenghua Chen, Min Wu, Chang Liu, and Liangyin Chen

Abstract—Current wireless cross-domain solutions are limited to cross-one-factor tasks, requiring target domain data participation for training or position-independent feature extraction using multiple transceivers. Therefore, this paper aims to demonstrate cross-domain wireless sensing in a more challenging domain generalization (DG) setting without multiple transceivers or target domain data. Specifically, we propose a style-randomized cross-domain wireless sensing model called WiSR, which extracts domain-invariant features from multiple source domains. It quantifies Channel State Information (CSI) differences in the subcarrier dimensions as subcarrier-domain styles and instructs the feature extractor to gradually bias the gesture signals by randomizing the subcarrier-domain styles at the feature level. Meanwhile, a domain classifier that shares the same feature extractor is instructed to gradually bias the domain signals by randomizing the gesture features. Then, the adversarial training framework enables the domain classifier to reduce the influence of domain signals on the feature extractor. Extensive experiments have been performed on three gesture datasets with varying amounts of subcarriers from devices with different NICs, including cross-one-factor (such as room, user, location, and orientation) and cross-multi-factor sensing tasks. The results demonstrate that our method considerably increases performance on wireless DG tasks. Our code is available at: https://github.com/LiuSjia/WiSR.

Index Terms—Domain Generalization, Channel State Information, Wireless Sensing, Gesture Recognition

1 Introduction

Combining wireless sensing with artificial intelligence enables various ubiquitous, contactless, and intelligent sensing applications, such as activity recognition [1], [2], human sensing [3], gesture recognition [4], human identification [5], and motion tracking [6]. Nevertheless, they are often trained based on identically distributed data assumptions, i.e., the training and test data follow similar distributions. Hence, the model performance degrades dramatically when test data are from a different data distribution. As a result of the disparity in domain representation between the training and testing domains, this issue is referred to as the domain shift [7] problem. Cross-domain recognition is a prevalent problem in the real world [8], and it is more challenging for wireless sensing applications that utilize the environmental sensitivity of WiFi signals.

In recent years, numerous wireless cross-domain techniques have been developed [9], including domain adaptation (DA) and position-independent feature extraction. The DA-based technique optimizes the neural network for domain-invariant feature extraction and requires the target domain data for model training to enable advanced adaptation to the new distribution. In view of the fact that this technique requires the network to be retrained when presented with a new target domain, it is clearly restrictive. First, even for the same gesture in the same domain, the CSI of Wi-Fi signals may alter over time. As a matter of fact, new domain data always appears sequentially in real-world applications rather than being collected and labeled in advance [10]. Additionally, obtaining domain data prior to deployment is challenging for high-risk applications, as target data may be difficult to obtain or unknown [11]. Position-independent feature-based approaches, on the other hand, require data collection from multiple transceiver configurations and are restricted to a particular sensing area. In addition to increasing implementation costs, it defeats the advantages of wireless sensing in terms of ease of deployment.

To overcome the aforementioned disadvantages, we strive to achieve cross-domain wireless sensing in a more challenging setting that does not require multiple transceivers or target domain data for training. Domain generalization (DG) [12] offers excellent inspiration for the idea that domain-invariant features can be extracted via intrinsic connections between one or more distinct but related source domains [13]. In this paper, we present WiSR, a wireless DG model based on domain style randomization. The motivation it is that executions in different domains have varying impacts on the subcarrier dimension of channel state information (CSI). We consider quantifying the impact on the subcarrier dimension at the feature level and refer to it as *subcarrier-domain style*. During the model's optimization process, there are two different learning routes: *gesture-biased learning* and *adversarial domain-biased learning*. In *gesture-biased learning*, random interpolation continuously obfuscates subcarrier-domain styles to produce fuzzy subcarrier-

- *Shijia Liu and Chang Liu are with the College of Computer Science, Sichuan University, Chengdu, 610065, China. E-mail: shijialiu@stu.scu.edu.cn, 13541000972@163.com*
- *Zhenghua Chen is with the Institute for Infocomm Research (I2R) and Centre for Frontier AI Research (CFAR), A*STAR, Singapore. E-mail: chen0832@e.ntu.edu.sg*
- *Min Wu is with the Machine Intellection Department and Institute for Infocomm Research (I2R), A*STAR, Singapore. E-mail: wumin@i2r.a-star.edu.sg*
- *Liangyin Chen is with College of Computer Science and the Institute for Industrial Internet Research, Sichuan University, Chengdu, 610065, China. E-mail: chenliangyin@scu.edu.cn*

Corresponding author: Liangyin Chen
Manuscript revised December 18, 2022.

domain styles. Then, the fuzzy subcarrier-domain styles are transferred to gesture features using adaptive instance normalization (AdaIN) [14]. As a result, the feature extractor gradually becomes insensitive to the domain signal and is forced to optimize in the direction of the de-styled gesture signals. Similarly, *adversarial domain-biased learning* will continuously transfer subcarrier-domain styles to fuzzy gesture features, which will enable the domain classifier sharing the same feature extractor to gradually become insensitive to gesture signals and be forced to optimize in the direction of domain signals. Meanwhile, adversarial training is used to make the feature extractor further away from the domain signal. Using such a subcarrier-domain style randomization-based approach will improve the transferability of the learned gesture features across domains.

Using gesture recognition as an example, extensive experiments were conducted on three publicly available real-world gesture datasets, which included varying quantities of subcarriers and were collected using wireless devices with different network interface cards (NICs). Besides evaluating cross-one-factor tasks based on changes in only one environmental factor (i.e., room, user, location, and orientation) in current wireless cross-domain sensing approaches, we also evaluate cross-multi-factor tasks. Our results demonstrate that our method outperforms existing benchmark approaches.

The following highlights our contributions:

- In this paper, cross-domain wireless sensing was achieved in the more challenging wireless DG scenario. To the best of our knowledge, wireless DG contributions remain relatively limited.
- In this paper, we present a novel *subcarrier-domain style* and investigate the effect of executions in different domains on gesture signals in CSI.
- This paper presents WiSR, a wireless DG model based on *subcarrier-domain style* randomization. It enhances the transferability of the learned gesture features across domains through *gesture-biased learning* and *adversarial domain-biased learning*.
- We conducted extensive experiments on three datasets containing different amounts of subcarriers from devices with different NICs, encompassing both cross-one-factor and cross-multi-factor tasks. The experimental results demonstrate the superiority of our method.

2 Related work

2.1 Cross-Domain Wireless Sensing

Existing cross-domain wireless sensing methods fall into two categories: DA-based and position-independent feature-based.

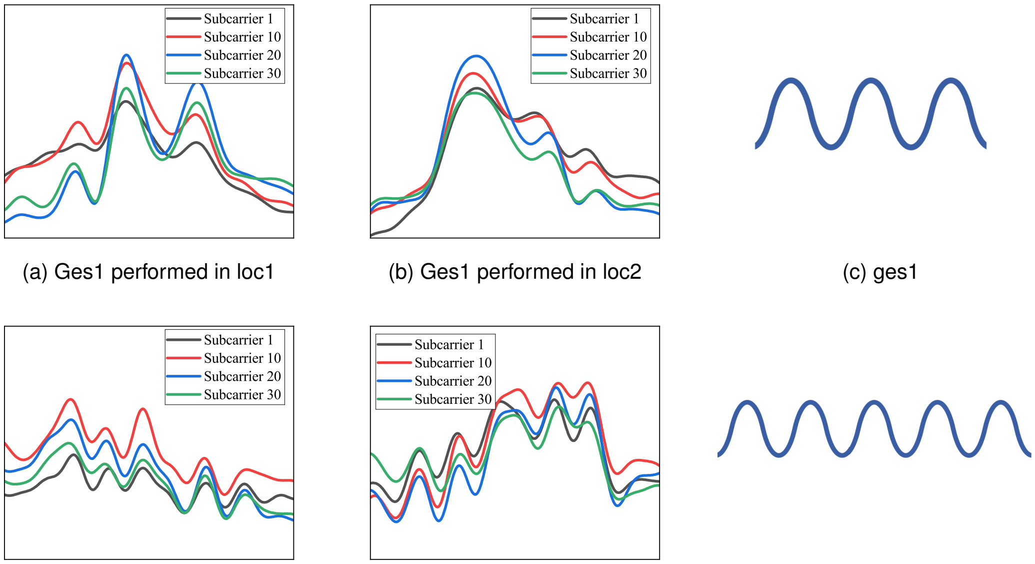
2.1.1 DA-based approaches

There have been numerous recent works that utilize transfer learning [15], [16] and DA methods such as domain adversarial training [17], [18], [19], data augmentation combined with pseudo-labeling and consistency regularization [20] for the purpose of achieving cross-domain wireless sensing. They must acquire enough target domain data to train or fine-tune the model in order to accomplish source-to-target domain adaptation. Nevertheless, such a method is obviously quite limited and labor-intensive. Few-shot learning [21], which takes only a few of the target domain data to map the source domain to the target domain, is gaining popularity. Shi et al. [22] developed a novel scheme using a matching network with enhanced CSI to facilitate one-shot learning of human activity recognition. Zhang et al. [23] proposed a graph-based few-shot learning method with a dual attention mechanism to perform CSI-based activity recognition. Zhang et al. [24] presented a dual-path prototypical network to achieve domain-transferable mapping in wireless gesture recognition. Although they considerably minimize the burden of data collection in the target domain, the cost of training a model for each target domain is expensive, and it is not feasible to gather all potential domains. Moreover, even for the same gesture in the same domain, the CSI of Wi-Fi signals may alter over time. Obtaining target domain data in advance may also be problematic for some applications with high risks. Therefore, this strategy remains quite restrictive.

2.1.2 Position-independent feature-based approaches

Several previous studies have used specific techniques to derive position-independent features before training the model. Widar3.0 [25] adopted Doppler Frequency Shift (DFS) to extract cross-domain Body-coordinate Velocity Profiles (BVP) for gesture recognition with at least three receivers and one transmitter device setup. WiHF [26] developed a dual-task DNN framework utilizing motion patterns as position-independent features to identify gestures and users collaboratively, with the same device requirements as Widar3.0. WiGesture [27] recognized gestures using the hand-orient view-based strategy and the position-independent Motion Navigation Primitive (MNP), requiring one transmitter and two receiver device setups. Niu et al. [28] recognized a predefined set of position-independent gestures based on the gesture fragment number and series of motion direction changes in a $2{\times}2$m sensing area with one transmitter and two receivers. DPSense [29] designed a quality-oriented signal processing framework to use the sensing quality of signal segments of each gesture for cross-domain recognition. The aforementioned techniques require at least two receivers and region-specific recognition to ensure position-independent feature extraction, which is contrary to the simplicity and portability of wireless sensing.

Current wireless works necessitate target domain data for training or position-independent feature extraction using several transceiver devices. In addition, most aforementioned studies only demonstrate their cross-domain performance on cross-one-factor tasks while ignoring cross-multi-factor recognition. Therefore, we develop WiSR, a wireless DG model based on subcarrier-domain style randomization that attempts to generalize the model to the unseen test domain with the more challenging setting of no multiple receiver devices and target domain data. WiSR also exhibits its effectiveness on cross-one-factor and cross-multi-factor tasks.



(a) Ges1 performed in loc1 (b) Ges1 performed in loc2 (c) ges1

(d) Ges2 performed in loc1 (e) Ges2 performed in loc2 (f) ges2

Fig. 1. For two gestures performed at two different locations, (a), (b), (d), and (e) illustrate the CSI amplitudes of subcarriers 1, 10, 20, and 30. (c) and (f) conceptualize the similarity between subcarriers in gestures 1 and 2 across different domains.

2.2 Domain Generalization (DG)

DG differs from DA, which has been widely used for cross-domain wireless sensing, in that it can extract invariants from several related but different source domains without utilizing data from the target domain. The current DG is primarily used for a variety of computer vision applications, such as person re-identification [30], semantic segmentation [31], image compression [32], and street view recognition [33]. DG is also useful in fault diagnosis [34], natural language processing [35], medical analysis [36], and speech utterance [37]. Unlike the above application fields, wireless sensing involves more complex domains. A wireless domain may include a variety of environmental factors such as user, room, location, and orientation. In the complex wireless domain in which multiple environmental factors interact, wireless gesture features are complex. Meanwhile, wireless sensing is implemented based on environmental sensitivity, and it is therefore challenging to be robust to the variations of environmental factors in cross-domain issues. We have found that it is difficult to obtain good generalization performance for general DG methods when applied to wireless data; this calls for developing a DG method tailored to wireless data.

3 OBSERVATION AND MOTIVATION

Multiple-input multiple-output (MIMO) and orthogonal frequency division multiplexing (OFDM) have increased the link capacity of modern wireless communications. Subcarriers of various frequencies generated by a transmitter with N_t antennas are reflected by various objects and received by a receiver with N_r antennas, with K subcarriers contained between each antenna pair. Thus, we can acquire CSI data with the dimensions (N_tN_rK, T), where N_tN_rK represents the number of subcarriers and T represents the length of the time series. They also represent the subcarrier and time dimension of the CSI matrix, respectively. Numerous applications utilize statistical values in the CSI time dimension to achieve in-domain wireless sensing [38]. As a result, it is extremely difficult to analyze the domain differences in the CSI time dimension. The differences in the CSI subcarrier dimension, however, have received less attention.

We attempt to investigate how domain differences affect CSI in the subcarrier dimension. Fig. 1 illustrates the CSI amplitudes of four subcarriers for two gestures executed at two different locations. For the same gesture performed at different locations in Fig. 1a and Fig. 1b, as well as Fig. 1d and Fig. 1e, significant differences between them illustrate the interference of domain shifts in gesture recognition. In addition, the same subcarrier exhibits different sensitivities to gestures executed at different locations, i.e., different domains exert a differential effect on the subcarriers.

After carefully inspecting Fig. 1a and Fig. 1b, as well as Fig. 1d and Fig. 1e, the intrinsic relationship can be faintly perceived. For ease of understanding, Fig. 1c and Fig. 1f abstract the commonalities between Fig. 1a and Fig. 1b, as well as between Fig. 1d and Fig. 1e, respectively. They represent the similarity of gesture features across domains, despite the complexity of these similarities in practice. In other words, the gestural information is comparable to an iron wire symbol. The iron wire symbol will be deformed after being rubbed with various strengths and directions (interference from different domains). The goal of cross-domain gesture recognition is to identify the original gesture information contained within this deformed wire. Correspondingly, Fig. 1c and Fig. 1f can be viewed as the original iron wire symbol, i.e., the gesture signal without domain signal participation under ideal conditions. Fig. 1a and Fig. 1b, as well as Fig. 1d and Fig. 1e can be viewed as the deformed iron wire symbol, i.e., different degrees of deformation of the overview waveforms in Fig. 1c and Fig. 1f. Hence, the following conclusions can be drawn:

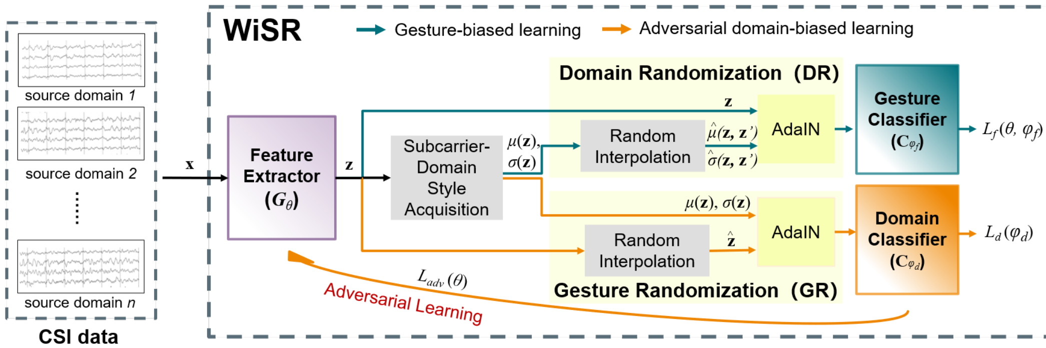


Fig. 2. The structure of WiSR. Gesture-biased learning makes the Feature Extractor and Gesture Classifier gradually bias the gesture signal. Adversarial domain-biased learning makes the Domain Classifier gradually bias the domain signal and the Feature Extractor away from the domain signal.

Domain differences result in different variations at the subcarrier level, which in turn leads to varying degrees of deformation of the gesture signal, resulting in inter-domain differences in the CSI.

Therefore, we consider quantifying such subcarrier-level variation as a distinctive domain style and refer to it as *subcarrier-domain style*. Performing the same gesture in different domains results in different subcarrier-domain styles, which will distort the gesture signal. The purpose of cross-domain wireless sensing is to minimize the influence of subcarrier-domain style and extract domain-independent gesture signals. In other words, the objective is to find the de-styled Fig. 1c and Fig. 1f from the differently styled Fig. 1a and Fig. 1b, as well as Fig. 1d and Fig. 1e. We consider obfuscating the subcarrier-domain style by random interpolation and transferring it to the original features using AdaIN [14]. Such an operation makes the feature extractor increasingly insensitive to the subcarrier-domain style but biased towards de-styled gestural signals.

4 PROBLEM DEFINITION

It is challenging for a source-learned model to generalize well without having access to target data [11]. Consequently, DG [12] was introduced to overcome the domain shift problem in the absence of target data, which implies the availability of one or multiple distinct but relevant source domains and enables the model to discover stable patterns across them. These stable patterns independent of marginal distributions will be more generalizable to unseen domains. A significant distinction between DA and DG is that DA has access to target domain data during training, but DG does not. Therefore, DG is more challenging and realistic in practice.

Let X denote the input space and Y the output (label) space. A *source domain* can be defined as $S = \{(x_i, y_i)\}_{i=1}^{n} \sim P_{XY}$, where $x_i \in X, y_i \in Y$, and P_{XY} is the joint distribution on $X \times Y$. In DG, we assume to have access to m ($m \geq 1$) similar but distinct source domains for training, i.e., *training domain* is $\mathcal{S} = \{S_j\}_{j=1}^{m}$. Each S_j contains i.i.d. data-label pairs sampled from P_{XY}^{j}. The goal of DG is to learn a robust and generalizable predictive function $f : X \rightarrow Y$ from the m source domains and achieve a minimum prediction error on the unseen *target domain* $\mathcal{T}$. The corresponding joint distribution of $\mathcal{T}$ is denoted by $P_X^{\mathcal{T}}$. Also, $P_X^{\mathcal{T}} \neq P_{XY}^{j}, \forall j \in \{1, \ldots, m\}$, but all the source domains and target domains have the same gesture categories.

5 METHODOLOGY

WiSR's structure is illustrated in Fig. 2. CSI data collected from multiple source domains is processed by the *Feature Extractor* to produce the original feature $\mathbf{z}$. Subsequently, the mean $\mu(\mathbf{z})$ and variance $\sigma(\mathbf{z})$ of the original feature $\mathbf{z}$ are calculated in the *Subcarrier-Domain Style Acquisition* and considered as subcarrier-domain style. The subcarrier-domain style $\mu(\mathbf{z})$, $\sigma(\mathbf{z})$ and the original features $\mathbf{z}$ are input to the *Domain Randomization (DR)* and *Gesture Randomization (GR)* modules, whose outputs are then input to the *Gesture Classifier* and the *Domain Classifier*, which share the same *Feature Extractor*. These two processes are referred to as *Gesture-biased learning* and *Adversarial domain-biased learning*, respectively. It is the goal of *Gesture-biased learning* to make *Feature Extractor* and *Gesture Classifier* increasingly insensitive to domain signals and biased toward gesture signals. Accordingly, DR randomly interpolates the subcarrier-domain style $\mu(\mathbf{z})$, $\sigma(\mathbf{z})$ to obtain the fuzzy subcarrier-domain style $\hat{\mu}(\mathbf{z}, \mathbf{z}'), \hat{\sigma}(\mathbf{z}, \mathbf{z}')$, which is then transferred via the AdaIN [14] to the original feature $\mathbf{z}$. In contrast to *Gesture-biased learning*, *Adversarial domain-biased learning* aims to make the *Domain Classifier* insensitive to gesture signals and biased toward domain signals. It will also bias the *Feature Extractor* away from domain signals. So, in order to reduce the impact of gesture features on the *Domain Classifier*, GR randomly interpolates gesture features $\mathbf{z}$ to obtain fuzzy gesture feature $\hat{\mathbf{z}}$. Then, the current subcarrier-domain style $\mu(\mathbf{z})$, $\sigma(\mathbf{z})$ is transferred into fuzzy gesture features $\hat{\mathbf{z}}$ via the AdaIN [14]. In a subsequent step, adversarial training is used in order to enable the *Feature Extractor* to be away from domain signals. Afterward, each module of WiSR will be described. The pseudocode of WiSR can be found in Algorithm 1.

5.1 Model Inputs

Benefiting from MIMO and OFDM, we will obtain N CSI complex matrices with dimensions of $(N_t N_r K, T)$, where

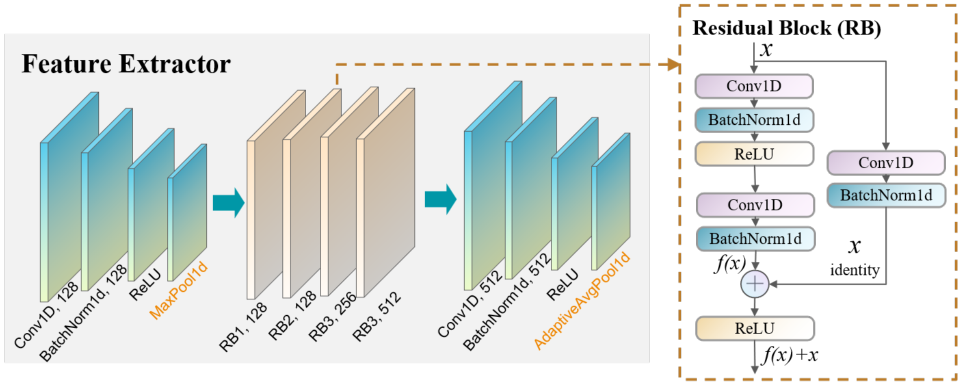


Fig. 3. The network structure of the feature extractor, which consists of Conv1D layer, BatchNorm1D layer, ReLU activation layer, MaxPool1D layer, four basic residual blocks (RB), and AdaptiveAvgPool1D layer.

N_tN_r is the number of transceiver antenna pairs, K is the number of subcarriers between each transceiver antenna pair, and T is the length of the time series. The CSI complex matrices are converted into amplitude and phase, respectively. Subsequently, merging the phase after phase calibration and the amplitude after Butterworth filtering, we will obtain a tensor with the dimensions $(2N_tN_rK, T)$ and use it as the input of WiSR.

5.2 Feature Extractor and Classifier

Feature Extractor was designed based on the skip connection of Residual Networks (ResNets) [39], and its network architecture is shown in Fig. 3. Conv1D layer, BatchNorm1D layer, Rectified Linear Unit (ReLU) activation layer, MaxPool1D layer, four basic residual blocks (RB), as well as Conv1D layer, BatchNorm1D layer, ReLU layer, and AdaptiveAvgPool1D layer, make up the Feature Extractor. An illustration of the RB structure is shown in the dashed box in Fig. 3. By adding a skip connection to forward propagation, RB enables intermediate activations to reach deeper layers of the network, thereby avoiding gradient disappearance and gradient explosion problems commonly associated with multilayer networks. The AdaptiveAvgPool layer [40], which is used before the final fully connected layer in Feature Extractor, automatically calculates the kernel size and stride for the average pool based upon the output and input sizes. The AdaptiveAvgPool layer is popular in neural networks because it reduces redundant information, enlarges the receptive field, and prohibits overfitting [41]. Additionally, each classifier consists of only one fully connected layer.

5.3 Subcarrier-Domain Style Acquisition

We consider quantifying subcarrier variation as *subcarrier-domain style*. However, statistical analysis of different subcarriers at the CSI data level may be error-prone due to the interleaving effects of multiple environmental factors and gesture signals. Therefore, in light of the powerful feature extraction capabilities of neural networks, we consider quantifying subcarrier-domain style at the feature level. Instead of interfering with the model's inputs, WiSR directs the model's learning process. An input $\mathbf{x}$ of dimension $(2N_tN_rK, T)$ is convolved in its time direction using Conv1D layers with an in-channel number of $2N_tN_rK$, and the feature extractor will generate a feature $\mathbf{z}$ of dimension $(1, 512)$. WiSR computes the mean $\mu(\mathbf{z})$ and variance $\sigma(\mathbf{z})$ of feature $\mathbf{z}$ to characterize the feature variations in CSI subcarrier dimensions. $\mu(\mathbf{z})$ and $\sigma(\mathbf{z})$ represent the subcarrier-domain style of this domain.

5.4 Adaptive Instance Normalization (AdaIN)

Style transfer is an interesting topic in computer vision. Historically, redrawing an image in a specific style required an experienced artist and a great deal of time. However, with the training of neural networks, it is possible to transfer the texture (style) of one image A to another image B, thereby quickly obtaining an image B in the style of image A, such as a photograph of the Great Wall in the style of Van Gogh. AdaIN [14] is a simple yet effective neural style transfer method that can achieve arbitrary style transfer in real-time. For a given input A and an input B, AdaIN adjusts the mean and variance of the content features of input B and aligns them with the mean and variance of the style features of input A to achieve style transfer from A to B. The calculation is as follows:

$$\mathrm{AdaIN}(A, B) = \sigma(A)\left(\frac{B - \mu(B)}{\sigma(B)}\right) + \mu(A), \tag{1}$$

where $\mu()$ and $\sigma()$ are the formulas for calculating the mean and variance, respectively.

5.5 Random Interpolation

Similar to mixup [42], linear interpolation can be performed on two different data to produce new data under the assumption of prior knowledge. Therefore, given data can be randomized by linear interpolation using randomly selected data, and the new mixed feature is called fuzzy data. This process is called random interpolation.

In two distinct training routes, we must acquire fuzzy subcarrier-domain styles and fuzzy gesture features, respectively. Then, WiSR can transfer fuzzy subcarrier-domain styles to gesture features and transfer the subcarrier-domain styles to fuzzy gesture features, thereby making the feature extractor and domain classifier insensitive to domain signals and gesture signals, respectively. For the fuzzy subcarrier-domain style and fuzzy gesture feature, we require as much randomness as possible in selecting the random sample, so we do not restrict the domain or gesture category of the randomly selected $\mathbf{z}'$. The fuzzy subcarrier-domain style $\hat{\mu}(\mathbf{z},\mathbf{z}'), \hat{\sigma}(\mathbf{z},\mathbf{z}')$ and fuzzy gesture feature $\hat{\mathbf{z}}$ are calculated as follows:

$$\begin{aligned}\hat{\mu}(\mathbf{z},\mathbf{z}') &= \alpha \cdot \mu(\mathbf{z}) + (1-\alpha)\mu\left(\mathbf{z}'\right), \\ \hat{\sigma}(\mathbf{z},\mathbf{z}') &= \alpha \cdot \sigma(\mathbf{z}) + (1-\alpha)\sigma\left(\mathbf{z}'\right),\end{aligned} \tag{2}$$

$$\hat{\mathbf{z}} = \alpha \cdot \mathbf{z} + (1-\alpha)\mathbf{z}', \tag{3}$$

where $\alpha \sim Uniform(0,1)$ is the random interpolation weight.

5.6 Gesture-biased Learning

A gesture classifier $\mathbf{C}_{\varphi_f}$ parameterized by φ_f should classify gesture categories independently of domain signals. Therefore, in gesture-biased learning, the DR module is introduced to randomize the domain signals and enforce the model to learn the gesture-biased features. DR maintains gesture features and obfuscates subcarrier-domain styles to obtain fuzzy subcarrier-domain styles. The fuzzy subcarrier-domain style $\hat{\mu}(\mathbf{z},\mathbf{z}'), \hat{\sigma}(\mathbf{z},\mathbf{z}')$ will replace the original subcarrier-domain style in feature $\mathbf{z}$ through AdaIN:

$$\mathrm{DR}\left(\mathbf{z},\mathbf{z}'\right) = \hat{\sigma}(\mathbf{z},\mathbf{z}')\left(\frac{\mathbf{z}-\mu(\mathbf{z})}{\sigma(\mathbf{z})}\right) + \hat{\mu}(\mathbf{z},\mathbf{z}'). \tag{4}$$

Then, the subcarrier-domain style-randomized gesture feature $\mathrm{DR}\left(\mathbf{z},\mathbf{z}'\right)$ will be fed into the gesture classifier $\mathbf{C}_{\varphi_f}$. The obtained gesture-biased loss L_{f} will be minimized to jointly optimize the feature extractor $\mathbf{G}_\theta$ and the gesture classifier $\mathbf{C}_{\varphi_f}$:

$$\begin{aligned}&L_{\mathrm{f}}(\theta,\varphi_f) = \\ &-\mathbb{E}_{(\mathbf{x},\mathbf{y},\mathbf{d})\in\mathcal{S}} \sum_{m_f=1}^{M_f} \mathbf{y}_{m_f} \log \mathbf{C}_{\varphi_f}\left(\mathrm{DR}\left(\mathbf{G}_\theta(\mathbf{x}), \mathbf{G}_\theta(\mathbf{x}')\right)\right),\end{aligned} \tag{5}$$

where $\mathbf{x}'$ is randomly selected from the batch, regardless of its gesture and domain category, M_f is the number of gesture categories, $\mathbf{y}_{m_f}$ is the ground truth of the gesture category for input $\mathbf{x}$, and $\mathcal{S}$ is the source domains.

As a result, the network will be encouraged to be insensitive to the continuously shifting fuzzy subcarrier-domain styles and favor the gesture signal.

Algorithm 1: The pseudocode of WiSR.

Input: The source domain data $\mathcal{S}$, batch size N, the feature extractor $\mathbf{G}_\theta$, gesture classifier $\mathbf{C}_{\varphi_f}$, the domain classifier $\mathbf{C}_{\varphi_d}$, the weight coefficient λ_{adv}, learning rate η, number of training steps n.

Output: Gesture model parameters θ, φ_f

. Initialize $\mathbf{G}_\theta$, $\mathbf{C}_{\varphi_f}$ and $\mathbf{C}_{\varphi_d}$;
for *n training steps* **do**
  Sample a data batch $(\mathbf{X},\mathbf{Y},\mathbf{D})$ from $\mathcal{S}$
  $\mathbf{Z} = \mathbf{G}_\theta(\mathbf{X})$
  Obtain the subcarrier-domain style of $\mathbf{Z}$, i.e., the mean $\mu(\mathbf{Z})$ and variance $\sigma(\mathbf{Z})$
  $\mathbf{Z}' = \mathrm{SHUFFLE}\ (\mathbf{Z})$
  Gesture-biased learning
  $\alpha = Random(0,1)$
  $\hat{\mu}(\mathbf{Z},\mathbf{Z}') = \alpha \cdot \mu(\mathbf{Z}) + (1-\alpha)\mu\left(\mathbf{Z}'\right)$
  $\hat{\sigma}(\mathbf{Z},\mathbf{Z}') = \alpha \cdot \sigma(\mathbf{Z}) + (1-\alpha)\sigma\left(\mathbf{Z}'\right)$
  $\mathrm{DR}\left(\mathbf{Z},\mathbf{Z}'\right) = \hat{\sigma}(\mathbf{Z},\mathbf{Z}')\left(\frac{\mathbf{Z}-\mu(\mathbf{Z})}{\sigma(\mathbf{Z})}\right) + \hat{\mu}(\mathbf{Z},\mathbf{Z}')$
  $L_{\mathrm{f}} = -\frac{1}{N}\sum_{j=1}^{N}\sum_{m_f=1}^{M_f} \mathbf{Y} \log \mathbf{C}_{\varphi_f}\left(\mathrm{DR}\left(\mathbf{Z},\mathbf{Z}'\right)\right)$ `// Calculate gesture-biased loss.`
  $\theta \leftarrow \theta - \eta\nabla_\theta L_{\mathrm{f}},\ \varphi_f \leftarrow \varphi_f - \eta\nabla_{\varphi_f} L_{\mathrm{f}}$ `// Update feature extractor and gesture classifier.`
  Adversarial domain-biased learning
  $\alpha = Random(0,1)$
  $\hat{\mathbf{Z}} = \alpha \cdot \mathbf{Z} + (1-\alpha)\mathbf{Z}'$
  $\mathrm{GR}\left(\mathbf{Z},\mathbf{Z}'\right) = \sigma(\mathbf{Z})\left(\frac{\hat{\mathbf{Z}}-\mu(\hat{\mathbf{Z}})}{\sigma(\hat{\mathbf{Z}})}\right) + \mu(\mathbf{Z})$
  $L_{\mathrm{d}} = -\frac{1}{N}\sum_{j=1}^{N}\sum_{m_d=1}^{M_d} \mathbf{D} \log \mathbf{C}_{\varphi_d}\left(\mathrm{GR}\left(\mathbf{Z},\mathbf{Z}'\right)\right)$ `// Calculate domain-biased loss.`
  $\varphi_d \leftarrow \varphi_d - \eta\nabla_{\varphi_d} L_{\mathrm{d}}$ `// Update domain classifier.`
  $L_{\mathrm{adv}} = -\lambda_{\mathrm{adv}}\frac{1}{N}\sum_{j=1}^{N}\sum_{m_d=1}^{M_d} \frac{1}{M_d} \log \mathbf{C}_{\varphi_d}\left(\mathrm{GR}\left(\mathbf{Z},\mathbf{Z}'\right)\right)$ `// Calculate adversarial loss.`
  $\theta \leftarrow \theta - \eta\nabla_\theta L_{\mathrm{adv}}$ `// Update feature extractor.`
end

5.7 Adversarial Domain-biased Learning

In addition to gesture-biased learning, we also design adversarial domain-biased learning in order to direct the network away from the domain signal. Thus, the domain classifier, which shares the same feature extractor as the gesture classifier, is constructed independently and focuses on domain information. We also add an adversarial learning framework to the feature extractor to continuously weaken domain signals.

Ideally, a domain classifier $\mathbf{C}_{\varphi_d}$ parameterized by φ_d should classify domain categories independently of gesture categories. Therefore, GR is introduced to perform operations similar to DR but with opposing objectives. GR maintains the subcarrier-domain styles and obfuscates gesture features to obtain fuzzy gesture features. The subcarrier-domain style $\mu(\mathbf{z})$, $\sigma(\mathbf{z})$ will transfer to the fuzzy gesture feature $\hat{\mathbf{z}}$ through AdaIN:

$$\mathrm{GR}\left(\mathbf{z},\mathbf{z}'\right) = \sigma(\mathbf{z})\left(\frac{\hat{\mathbf{z}}-\mu\left(\hat{\mathbf{z}}\right)}{\sigma\left(\hat{\mathbf{z}}\right)}\right) + \mu(\mathbf{z}). \tag{6}$$

Then, the gesture-randomized feature $\mathrm{GR}\left(\mathbf{z},\mathbf{z}'\right)$ will be fed into the domain classifier $\mathbf{C}_{\varphi_d}$. The obtained domain-biased loss L_{d} will be minimized to optimize the domain classifier $\mathbf{C}_{\varphi_d}$:

$$\begin{aligned}&L_{\mathrm{d}}(\varphi_d) = \\ &-\mathbb{E}_{(\mathbf{x},\mathbf{y},\mathbf{d})\in\mathcal{S}} \sum_{m_d=1}^{M_d} \mathbf{d}_{m_d} \log \mathbf{C}_{\varphi_d}\left(\mathrm{GR}\left(\mathbf{G}_\theta(\mathbf{x}), \mathbf{G}_\theta(\mathbf{x}')\right)\right),\end{aligned} \tag{7}$$

where $\mathbf{x}'$ is randomly selected from the batch, regardless of its gesture and domain category, M_d is the number of

TABLE 1
The details of three datasets. The "CSI shape" column has a format of $(N, N_t N_r K, T)$, in which the three dimensions represent the total number of samples, the number of subcarriers, and the length of the time series, respectively. The "Average number of samples" column represents the average sample number of each gesture in each domain.

Dataset	Wireless NIC	CSI shape	Domain number	Gesture number	Average number of samples	Room number	Loc number	User number	Ori number
Widar3.0 [25]	Intel 5300	(12875, 90, 2500)	200	6	10.73	3	5	6	5
CSIDA [24]	Atheros NIC	(2844, 342, 1800)	25	6	18.96	2	2/3	5	/
ARIL [43]	Etuss N210s	(1392, 52, 192)	16	6	14.5	/	16	/	/

domain categories, and $\mathbf{d}_{m_d}$ is the ground truth of the domain category for input $\mathbf{x}$.

As a result, the domain classifier $\mathbf{C}_{\varphi_d}$ will be encouraged to be insensitive to the continuously shifting fuzzy gesture features and favor the domain signal.

As a final step, the adversarial loss L_{adv} is computed by the cross-entropy between the domain-biased prediction and the uniform distribution. The expectation is over inputs from all domains and all gestures. The uniform distribution used in the cross-entropy loss will maximize the entropy of domain-biased predictions. Then, the adversarial loss L_{adv} is optimised to minimize the influence of the domain signal on the feature extractor by applying adversarial training:

$$L_{\text{adv}}(\theta) = \\ -\lambda_{\text{adv}}\mathbb{E}_{(\mathbf{x},\mathbf{y},\mathbf{d})\in S}\sum_{m_d=1}^{M_d}\frac{1}{M_d}\log\mathbf{C}_{\varphi_d}\left(\text{GR}\left(\mathbf{G}_\theta(\mathbf{x}),\mathbf{G}_\theta(\mathbf{x}')\right)\right). \tag{8}$$

where $\mathbf{x}'$ is randomly selected from the batch, regardless of its gesture and domain category, and λ_{adv} is a weight coefficient, which is set at 0.2 empirically.

Based on Eq. 5, Eq. 7 and Eq. 8, we will obtain three different training objectives, which will be used to optimize different model components. $L_{\text{f}}(\theta, \varphi_f)$ will be minimized to optimize the feature extractor $\mathbf{G}_\theta$ and gesture classifier $\mathbf{C}_{\varphi_f}$, $L_{\text{d}}(\varphi_d)$ will be minimized to optimize the domain classifier $\mathbf{C}_{\varphi_d}$, and $L_{\text{adv}}(\theta)$ will be minimized to weaken the influence of the domain signal on the feature extractor $\mathbf{G}_\theta$.

In the test phase, WiSR will eliminate the other modules and connect the feature extractor $\mathbf{G}_\theta$ and the gesture classifier $\mathbf{C}_{\varphi_f}$ directly, that is $\mathbf{y}_{pre} = \mathbf{G}_\theta(\mathbf{C}_{\varphi_f}(\mathbf{x}))$, thereby ensuring independent prediction and reducing the computational effort.

6 EXPERIMENTAL EVALUATION

6.1 Datasets

Extensive experiments were conducted on three different public and self-collected gesture datasets from the real world. They were collected by different CSI extraction tools on different wireless NICs and devices, thus containing different numbers of antennas and subcarriers. Table 1 describes the details of these datasets.

6.1.1 Widar3.0 dataset

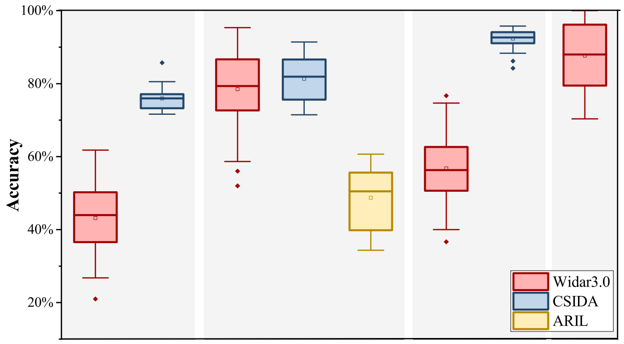
In the current state of wireless gesture analysis, Widar3.0 [25] is the most widely used cross-domain gesture dataset, from which six gestures were performed by six users in five locations and five orientations in three rooms, including Push&Pull, Sweep, Clap, Slide, Draw-O (horizontal), and Draw-Zigzag (horizontal). There was only one user who collected data in all rooms, and most domains in room 2 and 3 had only five samples per gesture. Data from Widar3.0 dataset was collected using six receivers with one antenna each and one transmitter with three antennas, but only one receiver was used in our experiments. In our experiments with Widar3.0 dataset, we consider four factors related to the environment: room, location, user, and orientation. In total, 12875 samples were collected across 200 domains, with an average of 10.73 samples per gesture per domain. For training and testing, we use all data from one receiver in a single room, with the exception of cross-room tasks that require the use of data from multiple rooms. All wireless devices were equipped with Intel 5300 wireless NICs and Linux 802.11n CSI Tool [44] collected compressed CSI of 30 subcarriers per link and 1000 packets per second collected. Accordingly, the input data shape of Widar3.0 dataset is $(12875, 1 \times 3 \times 30, T)$, where T is the length of each CSI sample. There is an inconsistent value for T in Widar3.0 dataset, however, it has been set to 2500 by zero-padding.

6.1.2 CSIDA dataset

CSIDA [24] dataset contains six gestures (pull left, pull right, lift up, press down, draw circle, and draw zigzag) performed by five users in two or three different locations in two rooms over a fixed time period of 1.8 seconds. One room contains three locations, while another contains only two. The dataset consists of 2844 samples and 25 domains, with an average of 18.96 samples per gesture per domain. In our experiments on CSIDA dataset, three environmental factors were considered: room, location, and user. CSIDA dataset was collected using a transmitter with one antenna and a receiver with three antennas, with all wireless devices equipped with Atheros wireless NICs. The Atheros CSI tool [45] provides uncompressed CSI of 114 subcarriers per link and sends 1000 packets per second. Accordingly, the input data shape of CSIDA dataset is $(2844, 1\times3\times114, 1.8\times1000)$. It is important to note that the Intel 5300 NIC used in Widar 3.0 dataset is obsolete [46], while the NIC and CSI tool used in the collection of CSIDA dataset are more appropriate for modern COTS wireless devices. WiSR's accuracy based on CSIDA dataset is therefore more representative.

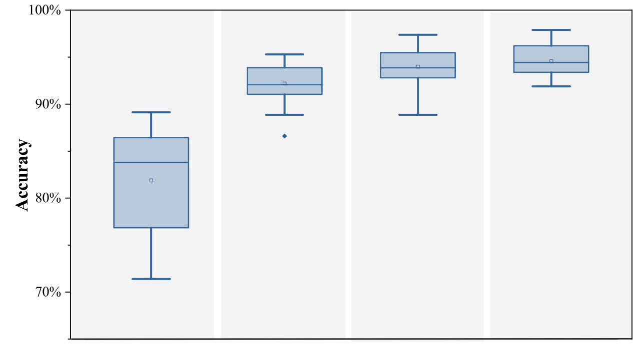
6.1.3 ARIL dataset

There are six gestures contained in ARIL [43] dataset (i.e., hand up, hand down, hand left, hand right, hand circle, and hand cross) that one individual performs at 16 locations within one room. The number of samples is 1392, and there are 16 domains, with an average of 14.5 samples per gesture per domain. Despite the fact that this dataset only contains



Datasets	Widar3.0	CSIDA	Widar3.0	CSIDA	ARIL	Widar3.0	CSIDA	Widar3.0
Mean Acc	43.17%	76.97%	78.49%	81.33%	48.74%	56.77%	92.35%	88.46%
Max Acc	61.78%	85.74%	95.33%	91.40%	60.70%	76.67%	95.79%	100%
Min Acc	21%	71.65%	52%	71.48%	34.35%	36.67%	84.21%	70.36%
Count	75	20	100	25	16	100	25	75
Cross Type	Cross-Room		Cross-Location			Cross-User		Cro-Ori

Fig. 4. The DG performance of cross-one-factor tasks on three different datasets.



Cross Type	Room-Loc	Room-User	User-Loc	Room-User-Loc
Mean Acc	81.89%	92.19%	93.98%	94.56%
Max Acc	89.12%	95.30%	97.37%	97.90%
Min Acc	71.40%	86.61%	88.87%	91.90%
Count	25	20	25	

Fig. 5. The DG performance of cross-multi-factor tasks on CSIDA dataset.

a single environmental variable (i.e., location), it utilizes a different universal software radio peripheral (USRP) device to collect CSI data, with the corresponding CSI sample being $(1 \times 52, 192)$. We use it therefore to evaluate the universality of our method for different wireless devices. The input data shape of ARIL dataset is $(1392, 52, 192)$.

6.2 Overall Performance

6.2.1 *Cross-One-Factor Evaluation*

Fig. 4 illustrates the average accuracy of the five leave-one-out cross-validations of the cross-one-factor tasks for three datasets across four types of environmental factors, including room, location, user, and orientation. The "Count" indicates the number of tasks. It is important to note that the training process did not include any test domain data, and only data from a single transceiver was used. On CSIDA, Widar3.0, and ARIL datasets, the average DG accuracy of cross-location recognition decreases sequentially from 81.33% to 78.49% to 48.74%. The most likely explanation is the sequential decrease in the number of subcarriers in the three datasets, which are 342 , 90 , and 52, respectively. A larger number of subcarriers will provide more information and facilitate domain-independent feature extraction. Additionally, discrepancies may arise due to the different NIC performance of the three datasets, where the NIC and CSI tool used in the collection of CSIDA dataset are more appropriate for modern COTS wireless devices. In terms of cross-room and cross-user recognition, Widar3.0 and CSIDA datasets have accuracy of 43.17% and 56.77%, 76.97% and 92.35% respectively. Besides the variations in subcarrier numbers and NICs mentioned above, it may also be due to the small number of samples for each gesture in Widar3.0 dataset, where the common user in room 2 and 3 collected five samples per gesture per domain. Furthermore, Widar3.0 dataset employs an unstable AP mode for data collection, which results in significant packet loss during the sampling procedure and varying sample lengths. The time dimension difference between samples can be as high as 1000 packets or more. It is necessary to conduct zero-padding on CSI samples in order to maintain a constant time length. CSIDA's dataset is free of this problem. Cross-orientation (Cro-Ori) DG accuracy is based solely on Widar3.0 dataset and reaches an accuracy of 88.46%. Thus, we recommend conducting wireless DG with sufficient training samples and subcarrier numbers, as well as a more stable data collection mode. Even so, WiSR's performance on the current datasets for a variety of cross-one-factor tasks is adequate to demonstrate the viability of wireless DG.

6.2.2 *Cross-Multi-Factor Evaluation*

Most current experimental evaluations of wireless cross-domain works are limited to cross-one-factor tasks, but we evaluated WiSR's performance on cross-multi-factor tasks. Due to the small average gesture numbers per gesture per domain for Widar3.0 dataset, we only used CSIDA dataset for cross-multi-factor validation. The results are shown in Fig. 5. WiSR's DG accuracy for cross-room-location (room-loc), room-user, user-location (user-loc), and room-user-location (room-user-loc) tasks is 81.89%, 92.19%, 93.98%, and 94.56%, respectively. Since CSIDA dataset only provides CSI data for two rooms, the number of source domains for cross-room-loc and cross-room-user tasks is identical to that of cross-location and cross-user tasks. With the addition of the room factor, however, the accuracy of cross-multi-factor tasks for the same number of source domains was comparable to that of cross-one-factor tasks. In this way, domain-invariant feature extraction is efficient. With domain-invariant features, precise gesture recognition is possible across multiple domains, regardless of the greater variance in training distributions. Compared to the cross-one-factor tasks, the cross user-loc and room-user-loc tasks show a substantial improvement in accuracy. The reason for this is that partial tasks within these two cross-multi-factor tasks contain more source domains than cross-one-factor tasks, and richer information is more favorable for feature extraction. These results illustrate WiSR's capacity to extract domain-invariant information from wireless environments, in which many environmental factors interact. Additionally, the recognition of the current domain can be considered as cross-multi-factor task when CSI changes over time. In the event that the gesture features remain unchanged, WiSR

TABLE 2
The DG performance with varying general DG methods, with bolded text representing the best results and underlined text representing the second-best results.

DataSet	ARIL	Widar3.0					CSIDA							
Type	C-L	C-R	C-L	C-U	C-O	Mean	C-R	C-L	C-U	C-RL	C-RU	C-UL	C-RUL	Mean
ERM [47]	46.80%	38.69%	76.36%	52.88%	87.93%	63.97%	72.23%	77.20%	89.43%	78.44%	89.38%	92.21%	92.23%	84.45%
IRM [48]	39.27%	36.15%	71.63%	52.10%	80.67%	60.14%	67.56%	73.08%	85.86%	75.41%	85.88%	86.50%	86.67%	80.14%
Mixup [42]	45.19%	40.67%	78.37%	**56.96%**	87.43%	65.86%	70.84%	72.29%	82.10%	73.60%	81.76%	83.49%	84.02%	78.30%
SelfReg [49]	44.45%	39.11%	76.71%	53.10%	86.67%	63.90%	69.08%	72.87%	82.91%	71.86%	83.10%	85.90%	86.36%	78.87%
RSC [50]	46.30%	36.54%	72.79%	50.23%	85.51%	61.27%	77.61%	81.21%	91.48%	83.03%	91.57%	92.84%	93.23%	87.28%
MMD-AAE [51]	34.16%	30.23%	28.82%	27.56%	42.47%	32.27%	**85.25%**	**88.75%**	84.24%	**90.12%**	81.16%	62.58%	62.84%	79.28%
DANN [52]	44.39%	40.36%	75.86%	54.77%	84.85%	63.96%	75.54%	80.52%	91.98%	80.83%	91.83%	93.26%	93.54%	86.79%
Proposed WiSR	**48.74%**	**43.17%**	**78.49%**	56.77%	**88.46%**	**66.72%**	76.97%	81.33%	**92.35%**	81.89%	**92.19%**	**93.98%**	**94.56%**	**87.61%**

TABLE 3
The DG performance of varying wireless cross-domain approaches, with bolded text representing the best results and underlined text representing the second-best results.

Type	C-R	C-L	C-U	C-O	C-RL	C-RU	C-UL	C-RUL	Mean
Proposed WiSR	76.97%	81.33%	**92.35%**	**88.46%**	81.89%	**92.19%**	**93.98%**	**94.56%**	**87.72%**
WiGR [24]	75.66%	79.95%	89.89%	73.20%	73.56%	89.87%	90.84%	90.20%	82.90%
AirFi [53]	**84.32%**	**87.13%**	82.96%	69.53%	**88.64%**	81.42%	70.35%	70.73%	79.39%
EI [17]	41.56%	45.87%	53.88%	41.98%	44.78%	52.04%	53.68%	52.98%	48.35%
Widar3.0 [25]	37.50%	46.27%	50.55%	42.32%	45.84%	49.89%	50.73%	50.55%	46.71%

trained on previous domains can be directly applied to the current domain without the need of additional data acquisition.

6.3 Comparison

6.3.1 Comparison with general DG methods

Seven state-of-the-art DG approaches are used for comparison. They all make use of the same feature extractor and classifier as WiSR. Among these, empirical source-risk minimizer (**ERM**) [47] utilizes only the base network to minimize the loss of all source domains; **IRM** [48] attempts to make the predictor Bayes optimal in all domains; **Mixup** [42] is one of the most popular techniques for data augmentation; **SelfReg** [49] employs only positive data pairs and designs a class-specific domain perturbation layer that is combined with Mixup; **MMD-AAE** [51] extends adversarial autoencoders by measuring features based on maximum mean difference (MMD); **RSC** [50] masks the key features in order to force the network to enable the remaining label-related features; and **DANN** [52] is a domain adversarial training approach. In Table 2, the average DG accuracy for all tasks is presented, with bold text denoting the best results and underlined text denoting the second best results. The "Type" column of Table 2 shows the various types of cross-domain tasks, where "C" indicates "cross", and "R", "L", "U", and "O" indicate the environmental factors of room, location, user, and orientation, respectively.

Using ERM as a benchmark, the accuracy of IRM on all tasks is lower than that of ERM. Mixup and SelfReg have improved the accuracy of most cross-domain tasks on Widar3.0 dataset, but they perform poorer on other datasets than ERM. In contrast, RSC is effective for cross-domain tasks in CSIDA dataset but less effective in other datasets. MMD-AAE achieves the highest accuracy for cross-room, cross-location, and cross-room-location tasks in CSIDA dataset, but the lowest accuracy for almost all other tasks. Even though DANN achieves higher accuracy than ERM on the majority of tasks in CSIDA and Widar3.0 datasets, the improvement is minimal or offset by a decrease in accuracy on other tasks. Furthermore, it is difficult to improve accuracy on ARIL dataset using all the aforementioned baseline methods. The performance of all baselines on the three datasets highlights the challenges of wireless DG, whereas our technique achieves not only significant improvements on all tasks in all datasets but also achieves the best or second-best performance in the majority of tasks, demonstrating its superiority. Taking into account the accuracy of all cross-domain tasks from CSIDA dataset as well as cross-orientation tasks from Widar3.0 dataset, WiSR achieved an average accuracy of 87.72% on the cross-one and cross-multiple environment factor tasks involving four environment factors, demonstrating its accuracy.

6.3.2 Comparison with wireless cross-domain methods

With the same DG setting (only denoised CSI from the source domains and collected by a single transceiver was used to train the model), some wireless cross-domain methods were reimplemented. **AirFi** [53] is a wireless DG method. **Widar3.0** [25] is a method that requires manual feature extraction from CSI collected by multiple transceivers. **EI** [17] and **WiGR** [24] are wireless DA methods that require target data for model training. **EI** [17] combined the correlation of the segment from the two receivers and FFT of each segment as the input to the deep learning model. **WiGR** [24] is a few-shot learning method that requires only a small amount of data from the target domain to facilitate the transfer of prototypes. Their accuracy is shown in Table. 3. As a result of denoised CSI being used instead of manual feature extraction, Widar3.0 [25] and EI [17] have obtained the lowest accuracy. AirFi [53] achieves the highest accuracy on the cross-room, cross-location, and cross room-loc tasks of CSIDA dataset, but low accuracy on other cross-domain tasks. Such a distribution of results is very similar to that of MMD-AAE [51]. Their similarity of results may be explained by the fact that they all employ feature alignment to minimize the differences in features extracted by adversarial autoencoder. WiGR [24] does not generalize

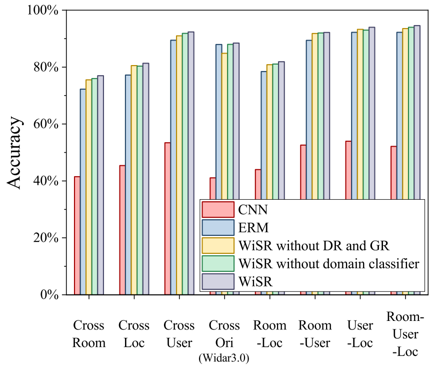


Fig. 6. The different DG performance of ablation study.

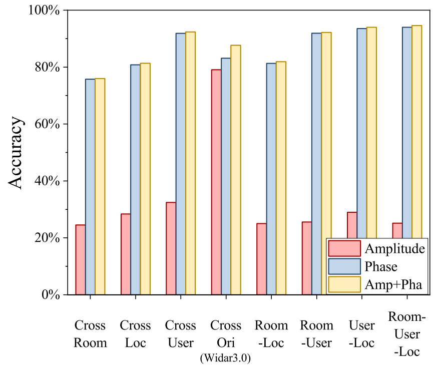


Fig. 7. The different DG performance with various input data.

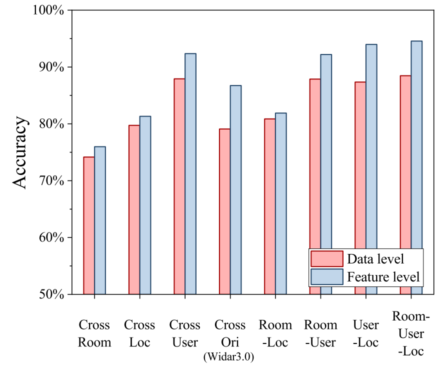


Fig. 8. The different DG performance with various operation objects.

well for unseen domains without participation from the target domain data. In all tasks, WiSR obtains maximum average accuracy, demonstrating its superior generalization capabilities.

6.4 Ablation Study

To investigate the contribution of each component to WiSR, ablation studies are conducted. Fig. 6 shows the performance of CNN, ERM, WiSR without DR and GR, WiSR without domain classifier (without domain-biased and adversarial loss), and complete WiSR on various cross-domain tasks. These experiments were conducted in the same manner as the previous experiments. Clearly, CNN performed the worst, demonstrating the effectiveness of our feature extractor. WiSR without DR and GR achieves slightly higher accuracy than ERM but lower than the complete WiSR, reflecting the importance of DR and GR. On the other hand, the difference between WiSR without domain classifier and ERM is the involvement of the DR module, i.e., whether the randomly interpolated subcarrier-domain style is transferred. WiSR without domain classifier has the second highest accuracy; however, it is smaller than the complete WiSR. This implies that both the DR module and adversarial training on the domain classifier play an important role.

6.5 Sensitivity Analysis

In this section, we perform sensitivity analyses on WiSR to evaluate the impact of various factors on WiSR performance.

6.5.1 Impact of input data

The input to WiSR is a tensor of amplitude and phase combinations (amp+pha). In Fig. 7, we depict the average accuracy of each task when only amplitude or phase information is used as model input. It is evident from Fig. 7 that amp+pha achieves the highest accuracy regardless of the tasks or datasets, as rich data facilitates the extraction of domain-independent features. In addition, we may deduce from the results that the extraction of domain-independent features of CSI primarily depends on phase information, as the cross-domain accuracy of the two datasets that depends solely on phase information is marginally lower than amp+pha. Nevertheless, the amplitude information provided by CSIDA dataset contributes little to the extraction of domain-independent features due to its low accuracy. In contrast, the amplitude information in Widar3.0 datasets contributes more to the domain-independent features.

6.5.2 Impact of operation object

In order to avoid confounding gesture signals in CSI data, which are interleaved by multiple environmental factors, WiSR calculates and operates at the feature level rather than at the data level. Thus, we evaluate WiSR performance when DR and GR are applied to CSI data and CSI features, respectively. Fig. 8 illustrates their DG accuracy and demonstrates that the generalization accuracy when DR and GR are applied to the data level is always inferior to that of CSI features, regardless of the tasks or datasets. Furthermore, the discrepancy between the accuracy of the two distinct operation objects is particularly apparent when performing cross-multi-factor tasks, which confirms the complexity of the interactions between multiple environmental factors and gesture signals. And the more environmental factors there are, the greater the impact on accuracy when CSI data is the operation object.

6.5.3 Impact of adversarial weight

The weight added to the adversarial loss during adversarial training are the only weight considered in WiSR. By varying the adversarial weight value, we examine the impact of adversarial training on generalization performance. An illustration of the average DG accuracy for various cross-domain tasks with adversarial weights ranging from 0.1 to 1 is shown in Fig. 9. In general, we can observe that tasks from different datasets respond differently to adversarial weights, but the variance is not excessive, particularly for the cross-room task from CSIDA dataset. Even though there is only a small difference in accuracy between various weights, it nevertheless demonstrates the effects of adversarial training on model performance. The best weight for WiSR is 0.2, since it results in the highest overall average accuracy.

6.5.4 Impact of sample number

As an illustration of the effect of the number of training samples on WiSR performance, Fig. 10 shows the average cross-domain accuracy of model training for varying numbers of training samples per domain and per gesture. As expected, cross-domain accuracy improves with an increase

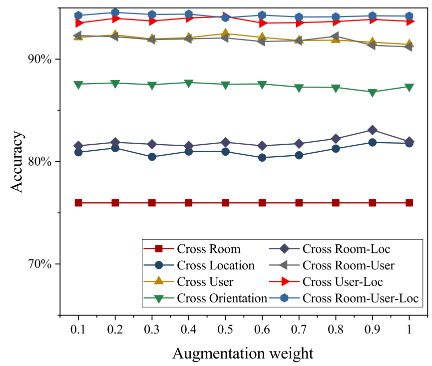


Fig. 9. The different DG performance with various adversarial weights.

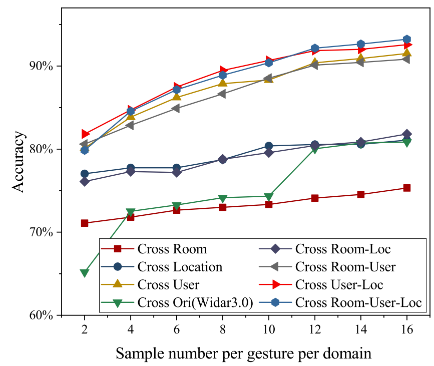


Fig. 10. The different DG performance with various sample numbers.

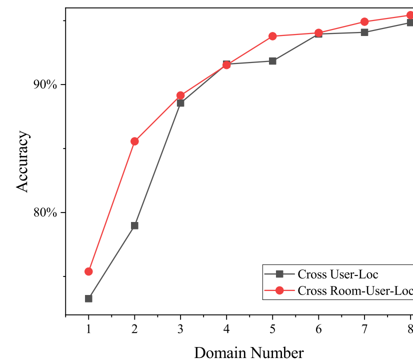


Fig. 11. The different DG performance with various domain numbers.

in gesture samples. In CSIDA dataset, whose average sample size is 18.93, accuracy is close to its best results when the number of training samples reaches 16. However, when the number of gesture training samples reaches 16, the accuracy of Widar3.0 dataset, which has an average sample size of 16.67-23.33, is 80.85%. It is less than the maximum accuracy of 88.46%. A possible reason for this difference could be that the richer subcarriers in CSIDA dataset allow for the extraction of domain-invariant features with fewer data while maintaining reliability. Thus, having sufficient gesture samples is essential when extracting domain-invariant features.

6.5.5 *Impact of domain number*

To better assess the practical applicability of the proposed technique, we investigate how the number of domains influences WiSR performance. We only evaluate cross user-loc and cross room-user-loc tasks since they contain more available source domains. There are 25 leave-one-out configurations for each task. According to CSIDA dataset, one room has two location labels and the other room has three location labels. The number of source domains for 15 of the 25 cross-domain tasks of each type is eight, and the number of source domains for the remaining 10 tasks is four, the same number as the number of source domains for the cross-user tasks. Thus, 15 cross-domain tasks with eight source domain numbers were evaluated in this experiment. Fig. 11 illustrates that WiSR's accuracy increases with an increase in the number of source domains. As a result of training with one source domain, the average accuracy of these two tasks is 78.20% and 72.70%, respectively. It can explain why the cross-room task on CSIDA dataset has a lower accuracy, since its training procedure utilizes only one source domain. The average accuracy of both tasks increases to about 90% when the number of source domains is increased to three. As a result of training with four source domain numbers, the average accuracy for the cross-user-loc and cross-room-user-loc tasks is 91.88% and 91.94%, respectively, which is comparable to the average accuracy for the cross-user tasks (92.35%). In the case of more than six source domains, accuracy gradually ceases to increase. When 8 source domains are used for training, the average accuracy of cross-user-loc and cross-room-user-loc tasks is 95.28% and 95.03%, respectively, which are greater than the 93.98% and 94.56% overall average accuracy for all 25 tasks, whose average domain number is $(15\times8+10\times4)/25 = 6.4$. Therefore, more source domains will facilitate the extraction of domain-independent feature information, even for tasks involving multiple environmental factors. Meanwhile, three distinct source domains are sufficient for WiSR to achieve excellent generalization performance.

7 Conclusion

In this paper, we present WiSR, a style-randomized wireless DG model for a more realistic setting where target-domain data and multiple transceivers are unavailable. It tries to creatively quantify the feature statistics in the subcarrier dimension of the CSI as subcarrier-domain styles so as to reduce their impact via style transfer and random interpolation. A gesture classifier and a domain classifier that share a feature extractor are constructed in order to continuously bias the extracted features towards gesture signals and away from domain signals, respectively, through gesture-biased learning and adversarial domain-biased learning. Through extensive evaluation on the cross-one-factor tasks and cross-multi-factor tasks of three datasets with varying amounts of subcarriers acquired from devices with varying NICs, we demonstrate the viability and robustness of WiSR.

Acknowledgments

This research is supported by National Natural Science Foundation of China under Grant 62072319, A*STAR under its Career Development Award (Grant No. C210112046), the Sichuan Science and Technology Program under Grant 2023YFQ0022 and 2022YFG0041, the Luzhou Science and Technology Innovation R&D Program (No. 2022CDLZ-6), (Sichuan University-Luzhou) Scientific and Technological Innovation R&D Project under Grant 2021CDLZ-11.

References

[1] 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, 2019.

[2] J. Yang, H. Zou, and L. Xie, "Securesense: Defending adversarial attack for secure device-free human activity recognition," *IEEE Transactions on Mobile Computing*, 2022.

[3] J. Yang, H. Zou, H. Jiang, and L. Xie, "Device-free occupant activity sensing using wifi-enabled iot devices for smart homes," *IEEE Internet of Things Journal*, vol. 5, no. 5, pp. 3991–4002, 2018.

[4] S. Tan, J. Yang, and Y. Chen, "Enabling fine-grained finger gesture recognition on commodity wifi devices," *IEEE Transactions on Mobile Computing*, vol. 21, no. 8, pp. 2789–2802, 2022.

[5] S. Liu, Y. Chen, H. Wang, H. Liang, and L. Chen, "A low-calculation contactless continuous authentication based on postural transition," *IEEE Transactions on Information Forensics and Security*, vol. 17, pp. 3077–3090, 2022.

[6] D. Wu, Y. Zeng, R. Gao, S. Li, Y. Li, R. C. Shah, H. Lu, and D. Zhang, "Witraj: Robust indoor motion tracking with wifi signals," *IEEE Transactions on Mobile Computing*, pp. 1–1, 2021.

[7] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 10, pp. 1345 – 1359, 2010.

[8] D. Zhang, D. Wu, K. Niu, X. Wang, F. Zhang, J. Yao, D. Jiang, and F. Qin, "Practical issues and challenges in csi-based integrated sensing and communication," *arXiv e-prints*, pp. arXiv–2204, 2022.

[9] J. Yang, X. Chen, D. Wang, H. Zou, C. X. Lu, S. Sun, and L. Xie, "Sensefi: A library and benchmark on deep-learning-empowered wifi human sensing," *Patterns*, vol. 4, no. 3, 2023.

[10] 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. 231:1–231:37, Feb. 2023. [Online]. Available: https://doi.org/10.1145/3570325

[11] K. Zhou, Z. Liu, Y. Qiao, T. Xiang, and C. C. Loy, "Domain generalization: A survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1–20, 2022.

[12] K. Muandet, D. Balduzzi, and B. Schölkopf, "Domain generalization via invariant feature representation," in *International Conference on Machine Learning*. PMLR, 2013, pp. 10–18.

[13] J. Cha, S. Chun, K. Lee, H.-C. Cho, S. Park, Y. Lee, and S. Park, "Swad: Domain generalization by seeking flat minima," *Advances in Neural Information Processing Systems*, vol. 34, pp. 22 405–22 418, 2021.

[14] X. Huang and S. Belongie, "Arbitrary style transfer in real-time with adaptive instance normalization," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, Oct 2017.

[15] B. Sheng, F. Xiao, L. Sha, and L. Sun, "Deep spatial–temporal model based cross-scene action recognition using commodity wifi," *IEEE Internet of Things Journal*, vol. 7, no. 4, pp. 3592–3601, 2020.

[16] Q. Bu, G. Yang, X. Ming, T. Zhang, J. Feng, and J. Zhang, "Deep transfer learning for gesture recognition with wifi signals," *Personal and Ubiquitous Computing*, pp. 1–12, 2020.

[17] 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.

[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] 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.

[20] B.-B. Zhang, D. Zhang, Y. Li, Y. Hu, and Y. Chen, "Unsupervised domain adaptation for device-free gesture recognition," *arXiv preprint arXiv:2111.10602*, 2021.

[21] J. Snell, K. Swersky, and R. Zemel, "Prototypical networks for few-shot learning," in *Advances in Neural Information Processing Systems*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, Eds., vol. 30. Curran Associates, Inc., 2017. [Online]. Available: https://proceedings.neurips.cc/paper/2017/file/cb8da6767461f2812ae4290eac7cbc42-Paper.pdf

[22] Z. Shi, J. A. Zhang, R. Y. Xu, and Q. Cheng, "Environment-robust device-free human activity recognition with channel-state-information enhancement and one-shot learning," *IEEE Transactions on Mobile Computing*, vol. 21, no. 2, pp. 540–554, 2022.

[23] Y. Zhang, Y. Chen, Y. Wang, Q. Liu, and A. Cheng, "Csi-based human activity recognition with graph few-shot learning," *IEEE Internet of Things Journal*, vol. 9, no. 6, pp. 4139–4151, 2022.

[24] 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, 2022.

[25] 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*, ser. MobiSys '19. New York, NY, USA: Association for Computing Machinery, 2019, p. 313–325.

[26] C. Li, M. Liu, and Z. Cao, "Wihf: Gesture and user recognition with wifi," *IEEE Transactions on Mobile Computing*, vol. 21, no. 2, pp. 757–768, 2020.

[27] 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.

[28] 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, 2022.

[29] 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.

[30] X. Jin, C. Lan, W. Zeng, and Z. Chen, "Feature alignment and restoration for domain generalization and adaptation," *arXiv preprint arXiv:2006.12009*, 2020.

[31] R. Gong, W. Li, Y. Chen, and L. V. Gool, "Dlow: Domain flow for adaptation and generalization," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019.

[32] M. Zhang, A. Zhang, and S. McDonagh, "On the out-of-distribution generalization of probabilistic image modelling," *Advances in Neural Information Processing Systems*, vol. 34, pp. 3811–3823, 2021.

[33] F. Qiao, L. Zhao, and X. Peng, "Learning to learn single domain generalization," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 12 556–12 565.

[34] M. Ragab, Z. Chen, W. Zhang, E. Eldele, M. Wu, C.-K. Kwoh, and X. Li, "Conditional contrastive domain generalization for fault diagnosis," *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1–12, 2022.

[35] B. Wang, M. Lapata, and I. Titov, "Meta-learning for domain generalization in semantic parsing," *arXiv preprint arXiv:2010.11988*, 2020.

[36] D. Mahajan, S. Tople, and A. Sharma, "Domain generalization using causal matching," in *International Conference on Machine Learning*. PMLR, 2021, pp. 7313–7324.

[37] V. Piratla, P. Netrapalli, and S. Sarawagi, "Efficient domain generalization via common-specific low-rank decomposition," in *International Conference on Machine Learning*. PMLR, 2020, pp. 7728–7738.

[38] J. Liu, G. Teng, and F. Hong, "Human activity sensing with wireless signals: A survey," *Sensors*, vol. 20, no. 4, 2020. [Online]. Available: https://www.mdpi.com/1424-8220/20/4/1210

[39] 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 (CVPR)*, June 2016.

[40] ——, "Spatial pyramid pooling in deep convolutional networks for visual recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 9, pp. 1904–1916, 2015.

[41] J. Liu, Z. Yao, S. Guo, H. Xie, and G. Yang, "Prirepvgg: Privacy-preserving 3-party inference framework for image-based defect detection," *Applied Sciences*, vol. 12, no. 19, 2022. [Online]. Available: https://www.mdpi.com/2076-3417/12/19/10168

[42] H. Zhang, M. Cisse, Y. N. Dauphin, and D. Lopez-Paz, "mixup: Beyond empirical risk minimization," in *International Conference on Learning Representations*, 2018.

[43] F. Wang, J. Feng, Y. Zhao, X. Zhang, S. Zhang, and J. Han, "Joint activity recognition and indoor localization with wifi fingerprints," *IEEE Access*, vol. 7, pp. 80 058–80 068, 2019.

[44] D. Halperin, W. Hu, A. Sheth, and D. Wetherall, "Tool release: Gathering 802.11n traces with channel state information," *SIGCOMM Comput. Commun. Rev.*, vol. 41, no. 1, p. 53, jan 2011. [Online]. Available: https://doi.org/10.1145/1925861.1925870

[45] Y. Xie, Z. Li, and M. Li, "Precise power delay profiling with commodity wifi," in *Proceedings of the 21st Annual International Conference on Mobile Computing and Networking*, ser. MobiCom '15. New York, NY, USA: Association for

Computing Machinery, 2015, p. 53–64. [Online]. Available: https://doi.org/10.1145/2789168.2790124

[46] A. Zhuravchak, O. Kapshii, and E. Pournaras, "Human activity recognition based on wi-fi csi data-a deep neural network approach," *Procedia Computer Science*, vol. 198, pp. 59–66, 2022.

[47] V. Vapnik, "Principles of risk minimization for learning theory," *Advances in neural information processing systems*, vol. 4, 1991.

[48] M. Arjovsky, L. Bottou, I. Gulrajani, and D. Lopez-Paz, "Invariant risk minimization," *arXiv preprint arXiv:1907.02893*, 2019.

[49] D. Kim, Y. Yoo, S. Park, J. Kim, and J. Lee, "Selfreg: Self-supervised contrastive regularization for domain generalization," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 9619–9628.

[50] Z. Huang, H. Wang, E. P. Xing, and D. Huang, "Self-challenging improves cross-domain generalization," in *European Conference on Computer Vision*. Springer, 2020, pp. 124–140.

[51] H. Li, S. J. Pan, S. Wang, and A. C. Kot, "Domain generalization with adversarial feature learning," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.

[52] Y. Ganin, E. Ustinova, H. Ajakan, P. Germain, H. Larochelle, F. Laviolette, M. March, and V. Lempitsky, "Domain-adversarial training of neural networks," *Journal of Machine Learning Research*, vol. 17, no. 59, pp. 1–35, 2016. [Online]. Available: http://jmlr.org/papers/v17/15-239.html

[53] D. Wang, J. Yang, W. Cui, L. Xie, and S. Sun, "Airfi: Empowering wifi-based passive human gesture recognition to unseen environment via domain generalization," *IEEE Transactions on Mobile Computing*, pp. 1–12, 2022.

Shijia Liu received the B.Eng. degree and is studying for her Ph.D. degree in the college of computer science from Sichuan University under the guidance of Professor Liangyin Chen. She has studied as a visiting Ph.D. student at Agency for Science, Technology and Research (A*STAR) in Singapore for one year, under the guidance of Dr. Zhenghua Chen. She is interested in wireless sensing.

Zhenghua Chen received the B.Eng. degree in mechatronics engineering from University of Electronic Science and Technology of China (UESTC), Chengdu, China, in 2011, and Ph.D. degree in electrical and electronic engineering from Nanyang Technological University (NTU), Singapore, in 2017. Currently, he is a Scientist and Lab Head at Institute for Infocomm Research, and an Early Career Investigator at Centre for Frontier AI Research (CFAR), Agency for Science, Technology and Research (A*STAR), Singapore. He has won several competitive awards, such as First Place Winner for CVPR 2021 UG2+ Challenge, A*STAR Career Development Award, First Runner-Up Award for Grand Challenge at IEEE VCIP 2020, Finalist Academic Paper Award at IEEE ICPHM 2020, etc. He serves as Associate Editor for Elsevier Neurocomputing and IEEE Transactions on Instrumentation and Measurement. He is currently the Vice Chair of IEEE Sensors Council Singapore Chapter and IEEE Senior Member. His research interests include smart sensing, data analytics, machine learning, transfer learning and related applications.

Min Wu is currently a senior scientist in Machine Intellection Department, Institute for Infocomm Research (I2R), Agency for Science, Technology and Research (A*STAR), Singapore. He received his Ph.D. degree in Computer Science from Nanyang Technological University (NTU), Singapore, in 2011 and B.S. degree in Computer Science from University of Science and Technology of China (USTC) in 2006. He received the best paper awards in InCoB 2016 and DASFAA 2015, and the finalist academic paper award in IEEE PHM 2020. He also won the CVPR UG2+ challenge in 2021 and the IJCAI competition on repeated buyers prediction in 2015. His current research interests include machine learning, data mining and bioinformatics.

Chang Liu received the M.S. degree in Management science and engineering from SiChuan University, Chengdu, China, in 2015, he is currently pursuing the Ph.D. degree with the College of computer science from Sichuan University, under the guidance of Prof. LiangYin Chen. He joined in the Second Research Institute, Civil Aviation Administration of China, where he is an Engineer. His major research interests include machine learning and airport operation

Liangyin Chen received the Ph.D. degree from the College of Computer Science, Sichuan University, China, in 2008. He joined the College of Computer Science at Sichuan University and has been a Professor in 2014. He works as the director of Sichuan Big Data Analysis and Fusion Application Technology Engineering Laboratory. From 2009 to 2010, he was a Visiting Researcher with the University of Minnesota, under the supervision of Prof. Tian He. His research interests include wireless sensor networks, Intelligent Internet of Things and IOT security, Industrial Internet, blockchains, and big data. He has authored or co-authored more than 100 papers, many of which were published in premier network journals and conferences such as IEEE Communications Surveys & Tutorials (COMST), IEEE Internet of Things Journal (JIOT), IEEE Transactions on Mobile Computing (TMC), IEEE Transactions on Information Forensics and Security (TIFS), IEEE Transactions on Industrial Informatics (TII), Knowledge-Based Systems (KBS), IEEE International Conference on Computer Communications (INFOCOM), ACM International Conference on Multimedia (MM), IEEE Communications Society Conference on Sensor and Ad Hoc Communications and Networks (Secon), and ACM Conference on Embedded Networked Sensor Systems (SenSys).