



# RFBoost: Understanding and Boosting Deep WiFi Sensing via Physical Data Augmentation

WEIYING HOU, Department of Computer Science, The University of Hong Kong, China  
CHENSHU WU, Department of Computer Science, The University of Hong Kong, China

Deep learning shows promising performance in wireless sensing. However, deep wireless sensing (DWS) heavily relies on large datasets. Unfortunately, building comprehensive datasets for DWS is difficult and costly, because wireless data depends on environmental factors and cannot be labeled offline. Despite recent advances in few-shot/cross-domain learning, DWS is still facing data scarcity issues. In this paper, we investigate a distinct perspective of radio data augmentation (RDA) for WiFi sensing and present a data-space solution. Our key insight is that wireless signals inherently exhibit data diversity, contributing more information to be extracted for DWS. We present RFBoost, a simple and effective RDA framework encompassing novel physical data augmentation techniques. We implement RFBoost as a plug-and-play module integrated with existing deep models and evaluate it on multiple datasets. Experimental results demonstrate that RFBoost achieves remarkable average accuracy improvements of 5.4% on existing models without additional data collection or model modifications, and the best-boosted performance outperforms 11 state-of-the-art baseline models without RDA. RFBoost pioneers the study of RDA, an important yet currently underexplored building block for DWS, which we expect to become a standard DWS component of WiFi sensing and beyond. RFBoost is released at <https://github.com/aiot-lab/RFBoost>.

CCS Concepts: • **Human-centered computing** → *Ubiquitous and mobile computing systems and tools*.

Additional Key Words and Phrases: Wireless Sensing, Deep Learning, Data Augmentation.

## ACM Reference Format:

Weiying Hou and Chenshu Wu. 2024. RFBoost: Understanding and Boosting Deep WiFi Sensing via Physical Data Augmentation. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 8, 2, Article 58 (June 2024), 26 pages. <https://doi.org/10.1145/3659620>

## 1 INTRODUCTION

The past decade has witnessed the conceptualization, development and commercialization of wireless sensing, especially WiFi sensing, which has revolutionized how sensing was practiced as well as how WiFi was used [28, 45, 57]. Many applications have been explored and enabled, including motion detection [46, 60], sleep monitoring [27, 61], fall detection [16, 32], gait recognition [41, 47], etc. In recent years, there has been a shift from using signal processing methods to leveraging deep learning methods for wireless sensing [2, 4, 13, 19, 20, 49, 62, 65, 69], which we term as *Deep Wireless Sensing* (DWS) for brevity. Thanks to the strong ability of feature learning, DWS promises impressive performance compared with conventional methods that usually need complicated handcrafted feature engineering.

One of the biggest challenges for DWS, however, is building comprehensive wireless datasets for training. Large-scale datasets play a critical role in deep learning techniques. The rapid development of Computer Vision (CV) benefits significantly from the wide availability of large-scale public datasets like ImageNet [7]. In DWS,

---

Authors' Contact Information: [Weiying Hou](#), [wyhou@cs.hku.hk](mailto:wyhou@cs.hku.hk), Department of Computer Science, The University of Hong Kong, Hong Kong, China; [Chenshu Wu](#), [chenshu@cs.hku.hk](mailto:chenshu@cs.hku.hk), Department of Computer Science, The University of Hong Kong, Hong Kong, China.

---

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

© 2024 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM 2474-9567/2024/6-ART58

<https://doi.org/10.1145/3659620>

however, it is cumbersome and costly to build a dataset alike, mainly because wireless data (e.g., WiFi signals) is difficult to collect and, more importantly, infeasible to label offline. Therefore, similar large-scale datasets are so far missing for DWS, which essentially limits the potential and adoption of these techniques.

It is an exhausting and expensive process to collect and label wireless sensing data. On the one hand, as is well recognized, wireless sensing, especially WiFi sensing, heavily depends on the environments, placements, subjects, devices, etc. To build a comprehensive dataset, one needs to exhaust all these domains and collect abundant data under every setting. On the other hand, wireless data cannot be labeled offline, *i.e.*, after the data has been collected. While unlabeled images and videos can be helpful by properly labeling them, it is infeasible to achieve the same for wireless data, even with manual efforts, as they are non-perceptual to human eyes. To mitigate the data issue, existing works attempt to train DWS models with limited data via techniques like few-shot learning [13, 49], transfer learning [2], generative learning [24], cross-modality training [4], etc.

In this paper, we consider a different perspective from the root of the problem, the training data, and investigate *data augmentation* to increase the diversity of data without collecting more. Data augmentation is a set of techniques that artificially inflate the training samples from existing data. It has been a common practice and proved effective in CV field for better learning performance. The key insight is that more information can be extracted from the original dataset via augmentation. Image data augmentation (IDA) for CV is commonly implemented via random transformations, such as flipping, cropping, erasing, rotation, translation, color space transformations, neural style transfer, etc [38]. However, as shown in Fig. 1, these IDA techniques originally designed for images may not be applicable to wireless data due to the fundamental differences in the inherent data structures (See §2). New techniques are therefore needed to realize effective data augmentation for wireless sensing data. To our best knowledge, no prior work has been dedicated to Radio Data Augmentation (RDA), although the idea has been primarily touched on ad-hoc in the literature [49, 70].

We target the RDA problem in this work and propose RFBoost, the first RDA framework that augments WiFi sensing data substantially and effectively. The design of RFBoost is grounded on the physical properties of WiFi signals. Particularly, we investigate physics-inspired data augmentation techniques for DWS by exploiting the inherent data diversity of WiFi signals. Signal diversity is well-perceived in wireless communication and sensing and can be observed in all the time, frequency, and space domains. Thanks to data diversity, a single sample (e.g., one measurement for a particular gesture/activity) can contribute to multiple independent observations, underpinning a promising opportunity for physical data augmentation (PDA). Take frequency diversity as an example. Different subcarriers observe diverse responses to human activities and environmental changes, as if an array of cameras capture the same scene with multiple images in one shot. While diversity has been well studied, especially for signal processing-based sensing (e.g., subcarrier selection has been a common practice in WiFi sensing [42, 60, 61]), it is surprisingly under-exploited in DWS.

To leverage data diversity for effective RDA, we investigate different policies and devise a set of novel techniques to boost the data size. We first explore different wireless data representations to understand the performance of DWS, and on this basis, present a DWS framework that takes time-frequency spectrograms as the deep neural networks (DNNs) inputs. Then we introduce RDA as a plug-and-play module that can be flexibly added to existing DWS structures. Specifically, it serves as an intermediate processing step between the raw wireless signals and the DNN inputs. To leverage frequency diversity, we generate spectrograms on multiple informative subcarriers, and furthermore, we design effective approaches for mixing the spectrograms on different subcarriers. Space diversity attributed to multiple antennas, if available, can also be utilized similarly. To exploit time diversity, we employ various time windows for spectrogram generation, producing multi-resolution spectrograms for training. With all these augmented spectrograms, which are all physically meaningful, RFBoost can inflate the training samples considerably and thus improve the learning performance significantly. Note that, focusing on data augmentation, RFBoost is orthogonal and complementary to research on more advanced DWS models that

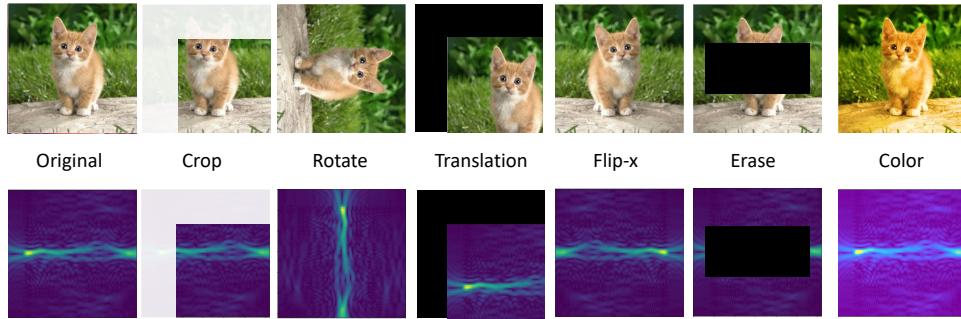


Fig. 1. Image data augmentation for images does not apply well to RF spectrograms.

promise better accuracy or need fewer data. It also differs from data synthesis that has been recently explored in the literature [2, 4, 39, 54].

We implement RFBoost using the PyTorch framework as an open library. We validate the RDA effectiveness on three public WiFi sensing datasets for gesture recognition [65], fall detection [56], and human activity recognition [3]. We experiment with 11 state-of-the-art models, among which 9 accept spectrogram input (See Tab. 1) and thus can be boosted with RFBoost, to examine the performance gains by RDA. Experimental results demonstrate remarkable performance improvements on all the datasets and models. With RDA only, we achieve an average accuracy gain of 5.4% on 9 existing models without collecting any new data. Evaluation of the fall detection dataset shows that RDA can empower small dataset to reach equivalent performance that was previously only achieved by over 4× of its training data to achieve the same accuracy of 93%. RDA also proves effective for various DNN models, including standard CV models and those customized for wireless/sensory sensing. By introducing RDA to DWS, we believe RFBoost will promote practical and deployable DWS and inspire many follow-up research opportunities.

**Contributions:** Our key contributions are summarized as follows:

- We introduce the problem of radio data augmentation to the community as a data-space solution to the data scarcity of DWS. We present an in-depth understanding of DWS by exploring wireless data representations and design RFBoost, a simple and effective RDA framework that incorporates a plug-and-play RDA module to boost the training of DWS models.
- We propose a set of physical data augmentation techniques by leveraging the inherent data diversity of WiFi signals, which boost training data substantially and effectively in an interpretable way.
- We implement RFBoost as an open library and evaluate it on multiple datasets and models. Results show that RFBoost augments the training significantly and achieves consistent performance gains across different datasets and models. RFBoost’s code is released at <https://github.com/aiot-lab/RFBoost>.

## 2 PRELIMINARIES

### 2.1 Deep Wireless Sensing

Most prior deep wireless sensing works build upon existing neural networks in the CV field, ranging from convolutional neural networks (CNNs) to Generative Adversarial Networks (GANs) and transformers, many of which aim to reduce the need of massive training data [2, 4, 13, 49]. In the context of WiFi sensing, the source input data is typically the complex Channel State Information (CSI). As illustrated in Fig. 2, without loss of generality, the data representations for the neural network inputs can be roughly categorized as three levels, depending on how much processing is done on the raw data:

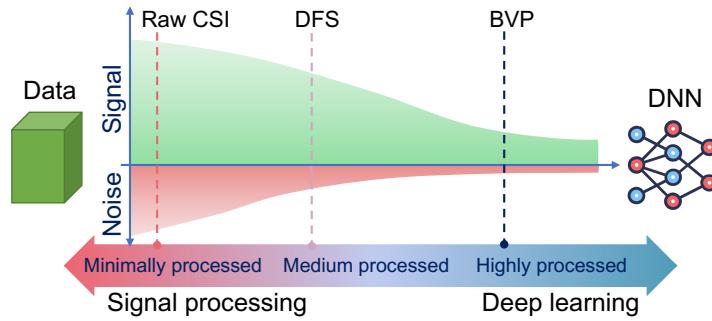


Fig. 2. Design choices for data preprocessing in DWS. Unprocessed or minimally processed data preserves the most information for learning, but suffers from severe noises; Highly processed data denoises the raw signals but also experiences information loss; Medium processed data seems to balance noise reduction and signal reserving for learning.

**Raw CSI:** Some designs argue that DNNs can learn the most from the least preprocessed CSI and thus directly feed in the raw CSI. As CSI is of complex values, the inputs are then usually represented as the CSI amplitude only [62], the amplitude concatenating the phase [29], and sometimes the I/Q-components [70]. Raw CSI, or minimally processed radio data, while containing the most information, however, may meanwhile suffer from severe noises.

**Time-frequency spectrograms:** Many works, either using signal processing or deep learning, employ time-frequency spectrograms typically generated by Short-Term Fourier Transform (STFT), which embody the Doppler Frequency Shifts (DFSs) for target motion and activities. Such spectrograms involve medium efforts of preprocessing and balance between denoising raw data and reserving signals for learning. Spectrograms can be represented as images in terms of data format, which can be easily fed into many standard CV models. Using time-frequency representations is becoming popular and demonstrated to be effective [25, 59]. Considering many subcarriers on WiFi CSI, many works perform Principal Component Analysis (PCA) to reduce the dimension and generate one or a few representative spectrograms [35, 42]. Alternatively, in DWS, one can also directly use all spectrograms generated on all subcarriers for learning.

**Highly-processed information:** CSI can be further processed to extract advance information that is supposed to be more explainable and environment-independent, such as *Body-coordinate Velocity Profiles* (BVPs) [65] or 2D/3D skeleton [20], etc. While noises can be reduced, a considerable amount of relevant information may be meanwhile lost, confining what DNNs can best learn. Furthermore, multiple links are usually compulsory in this case to calculate the desired information, and the computation for the processing can be high [65].

In RFBoost, we mainly focus on DWS for WiFi sensing with spectrograms as inputs, as the spectrogram representation achieves a good trade-off of the two key components in DWS, *i.e.*, the more explainable but handcrafted preprocessing and the less interpretable yet effective deep learning. While conventional wireless sensing frequently optimizes to denoise and reduce the dimension of raw signals, we argue that for DWS, reserving abundant information, even coming with noises, can be beneficial.

To justify our design choice, we carry out a preliminary experiment based on Widar3 dataset [65], one of the largest public WiFi sensing datasets thus far, to understand the DWS performance using raw CSI, DFS spectrograms, and BVPs as proposed in Widar3, respectively. As shown in Fig. 3, DFS spectrogram inputs significantly outperform raw CSI on most models as expected. Surprisingly, DFS inputs also produce comparable performance as the highly-processed and computation-intensive BVPs on the Widar3 model. As will be demonstrated later, it will even exceed BVPs when RDA is applied. While we agree that the models can be further optimized towards

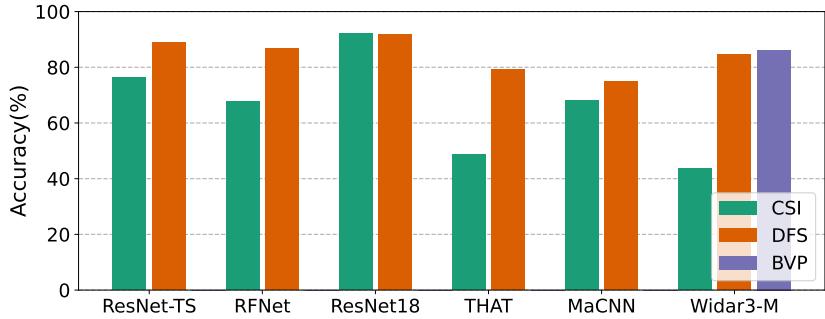


Fig. 3. Performance comparison using CSI, DFS, and BVPs as inputs on different models. DFS inputs outperform raw CSI for most models, and achieve comparable accuracy with BVPs. Note that the original Widar3 model (Widar3-M) only accepts BVPs, and is modified here to intake CSI and DFS. Model description can be found in Tab. 1.

different input formats, the preliminary results with common baselines prove DFS spectrograms as effective data representations for DWS.

## 2.2 Data Augmentation

It is generally recognized that larger datasets promise better deep learning models. Data augmentation is thus developed as a data-domain solution to the problem of limited data. The target is to enlarge the size and quality of a given training dataset. In CV, this is usually done by data warping augmentations which transform existing images with preserved labels by geometric or color transformations, such as cropping, flipping, random erasing, color channel/intensity changes, neural style transfer, etc. Fig. 1 illustrates some common operations for image data augmentation, and we refer to [38] for more information.

These well-studied IDA techniques for image data have performed well on many CV tasks. However, they do not perform as well when applied to radio data for augmentation. Radio data, even sometimes being represented as images for visualization, has its own underlying structures and physical meanings different from regular images. Take time-frequency spectrograms as an example. A spectrogram image indicates the frequency response at a certain time, with the horizontal and the vertical dimensions representing the frequency and time, respectively. Applying common IDA operations on a spectrogram may change the physical notion and produce unwanted distortions. For example, as in Fig. 1, rotating a spectrogram will reverse the time and frequency dimensions, flipping it will reverse the time series (e.g., a pull gesture can then become a push gesture), and changing the color space does not bring any new information, etc. Such operations make little sense to radio data, albeit some of them may still benefit DWS accuracy. Although IDA has been extended to other fields like audio and speech processing, data augmentation for DWS has barely been studied so far.

RDA is an important block of DWS as the problem of data scarcity is particularly critical in DWS due to the difficulties of collecting and labeling wireless data. Although there are a few open public datasets available nowadays, data collection for a particular application is usually still required, which renders as one of the biggest hurdles for building a DWS solution. Inspired by IDA, RFBoost pioneers the study of RDA for various DWS applications and delivers an effective solution. Similar to IDA in CV field, RDA is orthogonal to other solutions to handle the issue of limited training data, such as transfer learning [2], pretraining [69], one-shot or zero-shot learning [13, 49], cross-domain learning [65], cross-modality training [4], etc., which are out of the scope of this study. Also, note that data synthetic approaches like GANs can mitigate the data scarcity problem as well by

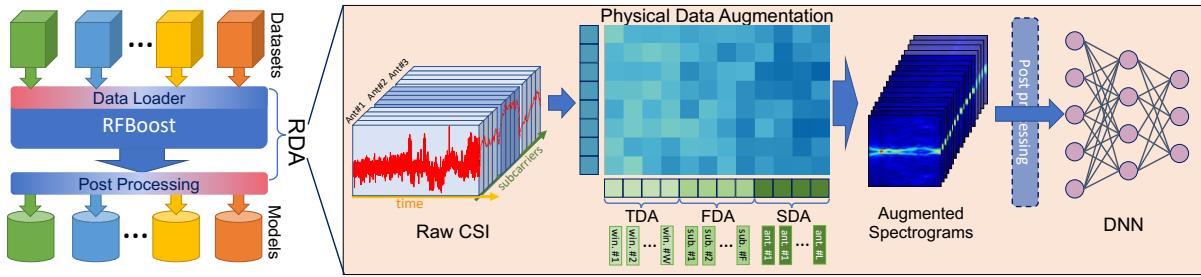


Fig. 4. Overview of RFBoost design. RFBoost is designed as a plug-and-play module that can be flexibly integrated with existing models (left). It performs RDA between the raw CSI data and the DNN inputs (right).

creating synthetic virtual instances (that are not physical) [64]. Yet in this work, we keep our focus on *physical data augmentation* (PDA) and do not target synthetic data generation.

### 3 RFBOOST DESIGN

#### 3.1 Overview

Fig. 4 illustrates an overview of the design of RFBoost. The core of RFBoost is the proposed RDA techniques that augment the input data by leveraging time-, frequency-, and space-diversity and output augmented time-frequency spectrograms for further learning. We aim to design RFBoost as a plug-and-play component that can be easily integrated into many existing and emerging DWS systems. Particularly, we envision RFBoost to be added into or partly replace the preprocessing pipeline of existing works between raw CSI signals and the DNN inputs. Depending on the specific DWS models, certain post-processing can further apply to the augmented outputs of RFBoost. By default, RFBoost only augments data for training, yet we also study the effect of test-time augmentation (§3.4).

Our key insight behind RFBoost is that the inherent data diversity of wireless signals potentially contributes more information that can be extracted to boost DWS training. Such diversity, e.g., frequency diversity, is a generally recognized notion that has been exploited to enhance signal processing performance for wireless sensing. This is usually done by selecting the best or the most representative signal(s) via methods like PCA [32, 35, 42, 57] or Maximal Ratio Combining (MRC) [47, 61]. While such processing enhances signal-noise-ratio (SNR) and reduces computation for signal processing, it may lead to information loss for deep learning in DWS. How to exploit data diversity for RDA is uncharted in DWS. Below, we first review the data diversity of WiFi signals and then propose PDA techniques that leverage such signal diversity.

#### 3.2 Data Diversity for Physical Data Augmentation

We first briefly review CSI and explain signal diversity before presenting the proposed RDA techniques.

CSI reflects the channel of the multipath propagation of wireless signals, which interact with the environments and humans and thus contain environmental information. Wireless sensing is then achieved by analyzing the CSI changes due to target movements in the environments. In WiFi OFDM systems, CSI is reported from the PHY layer and is estimated on multiple subcarriers, each corresponding to a carrier frequency  $f$ , as well as each antenna (if there is more than one). Therefore, from the data perspective, the time series of CSI measurements can be simply treated as a 3D array  $\mathbb{H}_{T \times F \times L}$ , where  $T$  denotes the total number of WiFi packets (CSI is measured discretely in time on a per packet base),  $F$  is the number of subcarriers, and  $L = N_{\text{Tx}} \times N_{\text{Rx}}$  denotes the number of links (i.e., the number of Tx antennas)  $N_{\text{Tx}}$  multiplying the number of Rx antennas  $N_{\text{Rx}}$ . Each element of  $\mathbb{H}_{T \times F \times L}$ ,

denoted as  $H(t, f, l)$ , represents the CSI measurement at time  $t$ , subcarrier  $f$ , and link  $l$ , which is a complex value with amplitude and phase denoted as follows.

$$H(t, f, l) = \sum_{m=1}^M a_m(t, f, l) \exp(-j2\pi f \tau_m(t, f, l)) + n(t, f, l), \quad (1)$$

where  $a_l(t, f, l)$  and  $\tau_l(t, f, l)$  are the complex amplitude and propagation delay of the  $m$ -th multipath component, and  $M$  denotes the total number of multipath components.  $n(t, f, l)$  stands for the additive thermal noise.

**3.2.1 Frequency Diversity.** CSI reported on different subcarriers exhibit different sensitivities to human activities due to the slightly different underlying carrier frequencies which feature different wavelengths. Depending on the specific environments and locations of the target and transceivers, such diversity can be very dominant. As in Fig. 5, we depict the spectrograms calculated on two subcarriers, a more sensitive subcarrier that captures strong Doppler responses to human motion and a less sensitive subcarrier that contains more noise. Although some subcarriers contain larger noise and produce lower SNR spectrograms, they may still contribute information gains that can be learned by DNNs and should not be just discarded.

**3.2.2 Time Diversity.** As mentioned above, to recognize an activity, the raw CSI time series are usually transformed as time-frequency spectrograms via STFT, and the resultant spectrograms basically represent the Doppler Frequency Shifts (DFS) caused by the activity. The STFT operation truncates the time series of signals using a sliding time window with a fixed length  $w$  and performs FFT for each window, which creates a trade-off between time- and frequency-domain resolutions. Using a shorter window reduces the frequency resolution but increases the time responsiveness, and applying a longer window results in the opposite effects. Therefore, although high frequency- and time-domain resolutions cannot be achieved simultaneously within the same window, employing multiple windows with diverse lengths will harness the time diversity. Spectrograms with shorter windows will better capture fast-changing DFS, while those using longer windows can offer more accurate averaged frequency responses. Fig. 5 portrays spectrograms using different STFT windows.

**3.2.3 Space Diversity.** Space diversity comes from different antennas, which in principle observe the channel independently at slightly shifted locations. Therefore, even on the same subcarrier with the same time window, the generated spectrograms can be considerably different. Space diversity is only available when there is more than one antenna forming multiple links (*i.e.*, Tx-Rx antenna pairs). Many previous works rely on an antenna array for phased-array processing [35, 65]. Differently, we exploit multiple antennas as opportunistic diversity for more independent observations.

### 3.3 Physical Data Augmentation

We now present how to leverage the discussed data diversity for physical data augmentation (PDA) in DWS systems.

**3.3.1 Frequency-domain Data Augmentation (FDA).** A straightforward way to utilize frequency diversity is to generate spectrograms on each subcarrier and add all of them to the training dataset. By doing so, the dataset size will be increased by  $F \times L$  times, where  $F$  is the total number of subcarriers and  $L$  is the number of links. This approach may suffer from two issues: 1) First, depending on the specific WiFi chipsets and driver, there can be tens of or hundreds of subcarriers<sup>1</sup>. Generating spectrograms on each individual subcarriers may not necessarily improve the accuracy, yet will significantly increase the computation. As will be shown in §5, using all 90 subcarriers on the Widar3 dataset may only produce as good performance as selecting 6 of them. 2) Second,

<sup>1</sup>Although a default subcarrier spacing of 312.5 kHz is specified in WiFi standards, different WiFi drivers will report different numbers of subcarriers (*e.g.*, they may downsample or skip some subcarriers in the reported CSI). In addition, the latest IEEE 802.11ax reduces the default subcarrier spacing to 78.125 kHz, adding 4 times more subcarriers than 802.11ac.

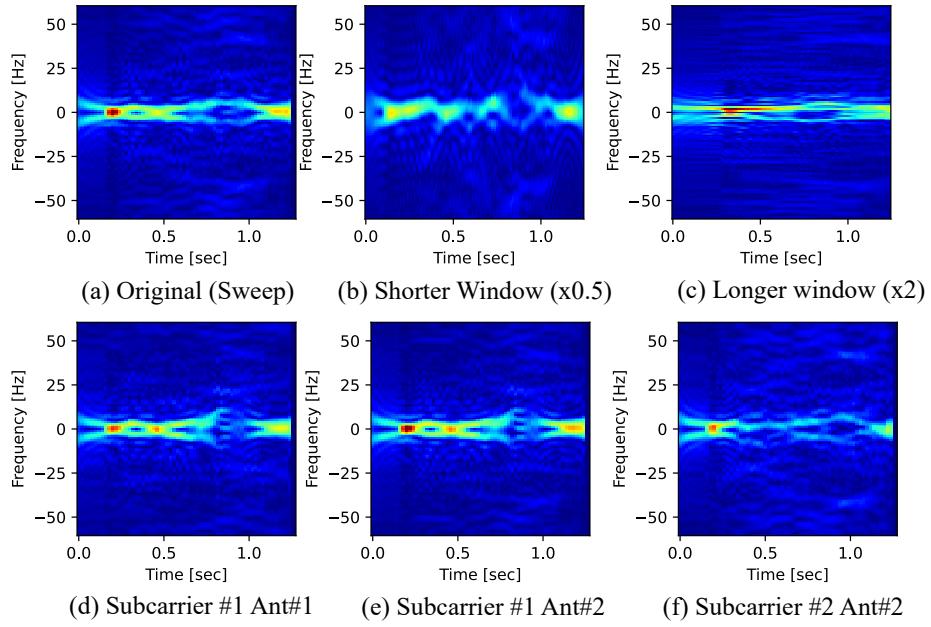


Fig. 5. Data diversity. For the same activity, the spectrograms using different windows ((b) and (c)), on different antennas ((d) and (e)), on different subcarriers ((e) and (f)), capture different DFS responses and background noise.

as subcarriers are continuous over the channel bandwidth, neighboring subcarriers might observe very similar signals. Therefore, spectrograms generated on all subcarriers may suffer from potential redundancy, which should be avoided.

To optimize FDA, we propose two different policies. For both cases, we consider all the  $F$  subcarriers on each link.

(1) *Individual Subcarrier Selection (ISS)*. Instead of using all the subcarriers, we select only a subset of the most representative subcarriers. To do so, we introduce a metric, namely *motion statistic*, as the core criteria for selection. Motion statistic is a metric for detecting motion and quantifying motion strength using WiFi signals, which is proposed in [60] based on statistical electromagnetic approaches. Specifically, motion statistic is defined as the first sample of the autocorrelation function (ACF) of CSI time series, which can be calculated efficiently. Not only being as a motion indicator, but it has also been demonstrated that motion statistic can serve as the equivalent channel gain on each subcarrier, which is useful for sensitive subcarrier selection and optimal subcarrier combining [47, 61].

Then our goal is to select a subset of  $K$  subcarriers for spectrogram generation such that they experience the highest motion statistics while spanning as broad and uniform as possible over the entire channel bandwidth, referred to as *ISS-K*. As shown in Fig. 6, we propose a heuristic to select such  $K$  subcarriers: Given a fixed  $K$ , we first divide the entire bandwidth into  $K$  uniform sub-bands. Then we select the top subcarrier of the highest motion statistic within each sub-band, resulting in  $K$  of them. Note the value  $K$  can be flexibly changed by user input or by certain automatic policy and, typically, is expected to be much smaller than  $F \times L$ .

(2) *Grouped Subcarrier Mixing (GSM)*. As  $K \ll F \times L$ , many subcarriers will not be used by the above ISS policy only. To fully exploit frequency diversity, we further propose to group all the subcarriers as  $G$  representative clusters and then combine the spectrograms within each cluster as a mixed version for augmentation. To make

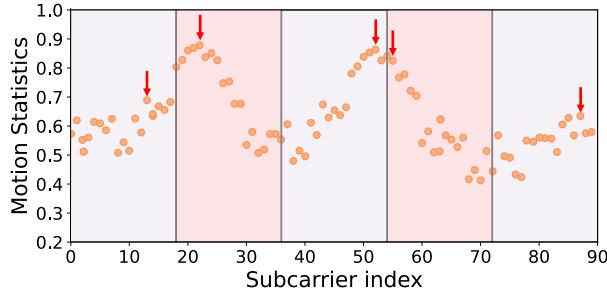


Fig. 6. Individual subcarrier selection when  $K = 5$ . The frequency band is divided into five sub-bands indicated by the color bars. Subcarriers with the highest motion statistic in each sub-band are selected and marked by red arrows.

sure the  $G$  mixed spectrograms contribute the most information for augmentation, we perform  $k$ -means clustering to minimize the inter-group similarity and maximize the intra-group similarity, where the similarity is measured as the Euclidean distance between two spectrograms. Once the  $G$  groups of subcarriers are obtained, we perform Maximal Ratio Combining (MRC) as in [61] on each group to optimally combine all the subcarriers therein. MRC is basically weighted combining, for which we use the motion statistics as the weights here. Different from [61] that performs MRC on the ACF signals, we directly use MRC to combine the generated spectrograms on different subcarriers. Formally, denote  $S(f, t)$  as the spectrogram generated on subcarrier  $f$ . The mixed spectrogram for the subcarriers in group  $g$  is then calculated as  $S_g(t) = \frac{1}{N_g} \sum_{f \in g} \omega(f, t) S(f, t)$ , where  $\omega(f, t)$  is the motion statistic for the corresponding subcarrier at time  $t$  and  $N_g$  denotes the number of subcarriers in the  $g$ -th group. As an alternative, in case of including noisy groups whose average motion statistic indicates relatively low motion among groups, we can select Top- $G$  groups by sorting the sum of motion statistics in each group instead of using every group as input. Similar to the value of  $K$ ,  $G$  can also be dynamically changed to fit specific datasets. While ISS is not necessarily data augmentation but more about data selection, GSM is true data augmentation since the augmented samples are mixed from multiple spectrograms on individual subcarriers and are thus different from any existing spectrograms generated from a specific subcarrier. Additionally, in case multiple antennas are available for space diversity, we can perform FDA separately for each individual antenna for further augmentation.

**3.3.2 Time-domain Data Augmentation (TDA).** To leverage time diversity for RDA, we apply  $W$  windows, each of a certain length  $w_i$ , for spectrogram generation. The number of windows and the corresponding window lengths can be given by the user or determined automatically by certain metrics (as will be discussed next). In RFBoost, by default, we will use  $W = 3$  windows, one default window, one lengthened window, and one shortened window. We use a fixed DFT point number ( $N_{DFT}$ ) used by STFT to maintain the uniformity of frequency dimension across different window sizes.

**3.3.3 Motion-aware Data Augmentation (MDA).** We employ motion statistics to detect the periods of motion as in [60]. With motion awareness, we devise two MDA methods below.

(1) *Motion-aware Random Erasing (MRE)*. Drawing inspiration from object-aware erasing strategies in CV [9, 72], our *Motion-aware Random Erasing* technique strategically places erasure masks at intersections of a randomly chosen window and detected motion-rich region, depicted in Fig. 7a. This masking process encourages DNNs to emphasize the holistic attributes of the data, thus alleviating overfitting to localized, motion-specific features. It is designed to enhance the model's ability to generalize by obscuring parts of the data that may lead to an excessive focus on minute, potentially non-generalizable details. Different from random time/frequency masking in [33],

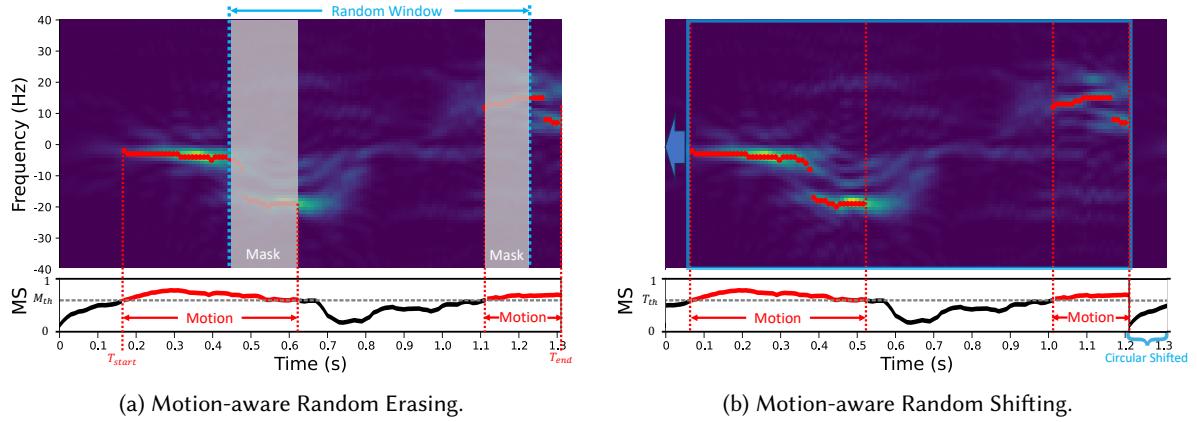


Fig. 7. Motion-aware Data Augmentation. (a) The region of intersection between the **motion period** ( $MS > T_{th}$ ) and the **random window** is selectively masked. Threshold ( $M_{th}$ ) for motion period detection is set to the 2nd quartile point of all measurements of MS in the same environment. (b) The integrity of the **motion period** is maintained in the **blue box** and the marginal "less motion" period is circularly shifted by  $-0.1sec$ .

our *Motion-aware Random Erasing* considers the physical semantics of the spectrogram, which is the motion period.

(2) *Motion-aware Random Shifting* (MRS). In contrast to the conventional shifting augmentation used in CV, which often resorts to *padding* to fill the voids created by shifting, our method acknowledges that motion captured in CSI data typically starts and ends with "less motion" period. Utilizing circular shifting in this context does not compromise the physical meaning of the CSI sample while preserving more information than *padding*. Through motion statistics, we detect periods of significant motion, setting the boundaries for circular shifting to maintain the integrity of the "motion" period. For instance, as shown in Fig. 7b, since significant motion period is detected within  $[T_{start}, T_{end}]$  in this sample with a length of  $T$ . Then we randomize the shift distance  $\tau$ , where  $\tau \in [-T_{start}, T - T_{end}]$ , ensuring the motion period is not circular shifted, and the temporal semantics are maintained while applying MRS. This approach, similar to how spatial shift in CV reduces overfitting by diversifying a dataset's spatial characteristics, enhances the temporal diversity within the CSI datasets.

### 3.4 Test-time Augmentation

RFBoost mainly targets training data augmentation to mitigate the data scarcity problem. Nevertheless, RDA can also be applied at test time, which has been shown to be effective as well. A more robust prediction can be expected by augmenting a test sample in the same way as the training data and then combining the augmented predictions. The combination is usually done via a mathematical function or even a learnable neural network [37]. In RFBoost, we implement test-time augmentation via a straightforward majority voting method to evaluate the effectiveness.

Test-time augmentation comes at the cost of more computation in the evaluation stage, which translates into an extra delay in model inference. Although we do not apply test-time augmentation by default, it can be promising for less time-sensitive tasks in practice.

## 4 IMPLEMENTATION

**Final Dataset Size:** In the current RFBoost, we take a ratio-aware design and keep the final dataset size as an input parameter that can be flexibly determined by users for considerations of specific datasets and different tasks. Once it is determined, *i.e.*, given  $K$  and  $G$  as in §3, RFBoost itself can find the best augmentation, *i.e.*, the most informative  $K$  subcarriers and the most representative  $G$  groups of subcarriers. Currently, we only evaluate the performance by empirical choices, but leave it as future work to find the optimal augmentation size and policy automatically. We define *Augmentation Ratio* (ARatio) as the factor by which an augmentation method increases the size of the original dataset.

**Data Length Alignment:** In most DNN models, the input tensor in the same mini-batch is required to be the same size due to the commonly used fixed size Fully Connected (FC) layer. However, in wireless sensing datasets, the length of each sample varies vastly, mainly because it takes various time for different users to repeat different activities. RFBoost aligns the spectrograms by first padding the CSI time series for STFT and then applying padding, trimming, and/or downsampling to the generated spectrograms, depending on information loss due to such operations.

**Online Augmentation:** Data augmentation can be done on the fly during training (*i.e.*, online augmentation) or be completed beforehand with augmented data stored (*i.e.*, offline augmentation). Online and offline augmentation basically raises a trade-off between training runtime delay and disk I/O performance. In our implementation, we design an online approach for augmentation, together with a cache-and-reload mechanism to avoid repeated calculation of temporary results, such as spectrograms for each subcarrier, K-means clustering group, etc. Nevertheless, the design choice can be flexibly based on the dataset size in practice.

**Software:** We have developed RFBoost as a PyTorch library, requiring only a customized data loader to handle various raw CSI formats. The enhanced datasets are fully compatible with existing DNN architectures, facilitating easy integration into any PyTorch-based models that process spectrograms or derived formats like *Body-coordinate Velocity Profiles*. Our implementation of the compared baseline methods is partly built upon the released code of UniTS [25]. RFBoost code can be found here: <https://github.com/aiot-lab/RFBoost>.

**Hardware:** We train and evaluate RFBoost both locally using an NVIDIA A100 and RTX4090 GPU. The local A100 GPU is divided into 4 vGPUs using NVIDIA MIG mechanism with a shared cache-base and codebase. All the servers share disk Volumes as a codebase and cache-base, respectively.

## 5 EXPERIMENTS

We now verify the effectiveness of RFBoost via experiments on real-world datasets that are publicly available and on state-of-the-art models. We mainly focus on how much improvement RDA can obtain given existing models and datasets.

### 5.1 Datasets

Several WiFi sensing datasets have been made publicly available recently. Our evaluation features three datasets, namely Widar3 dataset [65] for gesture recognition and FallDar [56] for fall detection, and OPERAnet [3] for activity recognition. We choose Widar3 as it is reported the largest WiFi sensing public dataset. We are also particularly interested in fall detection, as it is the most difficult to collect real fall data, among many wireless sensing applications. It would be exciting to see if RFBoost can reduce the need for massive data for fall detection via augmentation. More details about each dataset are described below.

**5.1.1 Widar3.** The Widar3 dataset, as released in [65], encompasses data from 17 participants. These participants are engaged in performing 22 different gestures, oriented in 5 different directions towards a transmitter, and stationed at five distinct locations within the range of six receivers, each equipped with one Intel 5300 network card and placed at different orientations, across three different environments (rooms). Since not all 22 gestures

Table 1. Summary of Compared Baseline Models

Models	Accepted Inputs	Description	RFBoost-able
Widar3-2D [65]	BVP	Designed for cross-domain gesture recognition using a 2D-CNN+GRU network, with the environment-independent 3D BVPs processed from DFS spectrograms as inputs.	Y
Widar3-1D [65]	CSI/DFS	Modified from Widar3-2D by replacing 2D-CNN with 1D-CNN to intake 2D DFS spectrograms.	Y
RF-Net [8]	CSI/DFS	RF-specific meta-learning framework with dual-path base network specially designed for DWS.	Y
THAT [23]	CSI/DFS	A two convolution augmented transformer model.	Y
ResNet [14]	CSI/DFS	One of the most used backbone models in CV.	Y
AlexNet [22]	CSI/DFS	One classic CV model.	Y
ResNet-TS [44]	CSI/DFS	ResNet extended for time series classification.	Y
MaCNN [36]	CSI/DFS	Multi-modal fusion structure based on CNNs.	Y
LaxCat [15]	CSI/DFS	An explainable DNN for time series classification.	Y
SLNet [55]	DFS	A polarized convolutional network designed for spectrograms learning.	Y
UniTS [25]	CSI	STFT-inspired network for time series sensor data including CSI that achieved the SOTA.	N

Table 2. Descriptions of Widar3-G6 and Widar3-G22 for evaluation.

	Room	Gestures	# of Users	Total
Widar3-G6	1	1:Push&Pull;2:Sweep;3:Clap;4:Slide;5:Draw-O(H);6:Draw-Zigzag(H)	8	6000
	2	1:Push&Pull;2:Sweep;3:Clap;4:Slide;5:Draw-O(H);6:Draw-Zigzag(H)	3	2250
	3	1:Push&Pull;2:Sweep;3:Clap;4:Slide;5:Draw-O(H);6:Draw-Zigzag(H)	4	3000
Widar3-G22	1&2	0:Slide. 1:Push&Pull. 2:Clap. 3:Sweep. 4:Draw-Zigzag(Horizontal). 5:Draw-O(Horizontal). 6:Draw-N(Horizontal). 7:Draw-Rectangle(Horizontal). 8:Draw-Triangle(Horizontal). 9:Draw-Zigzag(Vertical). 10:Draw-N(Vertical). 11:Draw-4. 12:Draw-5. 13:Draw-7. 14:Draw-3. 15:Draw-2. 16:Draw-9. 17:Draw-1. 18:Draw-0. 19:Draw-6. 20:Draw-8. 21:Draw-O(Vertical)	2	13732

are performed in every room, we focus on a common subset across all the rooms, which consists of the 6 most frequently performed gestures by 15 users, yielding a total of 11,250 samples, namely Widar3-G6 (or simply Widar3 for short). On Widar3-G6, we will evaluate how RFBOOST can improve generalizability across unseen devices (deployment), environments and users.

We also use another subset of data from two users who have performed all the 22 gestures (Note that gestures of "Drawing numbers", *i.e.*, Draw-0 to Draw-9, are performed only by two users and in Room#1 in Widar3 dataset), yielding a total of 13,732 samples, namely Widar3-G22. These subset choices are pivotal in ensuring a comprehensive dataset that captures a diverse range of environmental conditions and participants or gestures to facilitate further cross-domain studies. More details can be found in Tab. 2.

**5.1.2 FallDar.** The FallDar dataset [56] is collected over six months and includes data from a five-member family, comprising two seniors and three adults, as well as falls simulated with a dummy. The dataset features 442

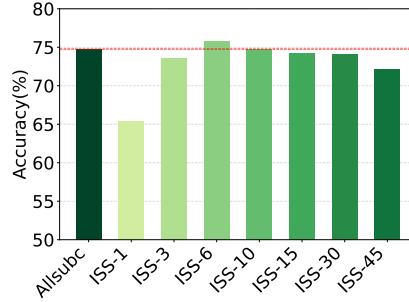


Fig. 8. Allsubc v.s. K-subcarriers using proposed Individual Subcarrier Selection (ISS-K, see §3.3)

instances of 6 fall types and approximately 2,000 samples for 21 normal activities. Among 442 instances of falls, 342 instances are performed by the dummy and the rest 100 instances are done by humans. CSI data is collected over 30 subcarriers with 3 antennas each. For evaluation, we classify the activities into two categories: fall and non-fall.

**5.1.3 OPERAnet.** The OPERAnet dataset [3] is a multi-modal dataset, and we focus on its CSI data collected from two Intel NUC devices for human activity recognition (HAR). This dataset features seven classes of indoor activities recorded over a span of 224 minutes. The activities include body-rotate, lie-down, sit, stand, stand-from-lie, walk, and a no-activity class. Note that OPERAnet dataset demonstrates a marked imbalance in distribution, as depicted in Fig. 15. For analytical convenience, we segment these samples into 3.2-second non-overlapping intervals, yielding a total of 5299 samples.

## 5.2 Baselines

Our focus is to evaluate how much relative performance improvements can be gained by RFBoost’s RDA, given existing datasets and models, rather than optimizing the absolute accuracy by, *e.g.*, pursuing sophisticated model design, which is out of our scope. Bearing this in mind, we consider the following categories of 11 baselines:

- Models adapted/customized for DWS, such as RF-Net [8], Widar3 [65] model and its variants, THAT [23], SLNet [55], etc. We will examine how much RDA can further improve models designed for wireless sensing.
- Common CV models, such as ResNet [14], AlexNet [7], etc. We want to see the performance, with RDA, on standard CV models that can be employed with little knowledge of wireless sensing and no model changes towards DWS.
- Advanced models optimized for spatial-temporal learning with sensory time series, including ResNet-TS [44], MaCNN [36], LaxCat [15], UniTS [25], SLNet [55], etc. These models are mostly designed for time series of general sensor data, including WiFi CSI data.

Tab. 1 summarizes the baseline models we compared. We integrate RFBoost into all these models for evaluation except for UniTS which only takes raw CSI as inputs. Some of these original models do not intake DFS spectrograms. Nevertheless, we can still feed DFS spectrograms as long as they accept 2D tensors as inputs. Again, we are particularly interested in understanding how RDA can enhance the performance of models that are not optimized for spectrograms or even not for wireless data.

To ensure fair comparisons between model performance w/ and w/o RDA on the dataset, uniform training configurations are employed for training on each model. This includes a consistent batch size of 256, learning rate of 0.001, and a standardized optimizer selection; AdamW is the default, with the exception of SLNet which utilizes RMSprop used in its original work. In alignment with our focus on assessing the data augmentation efficacy

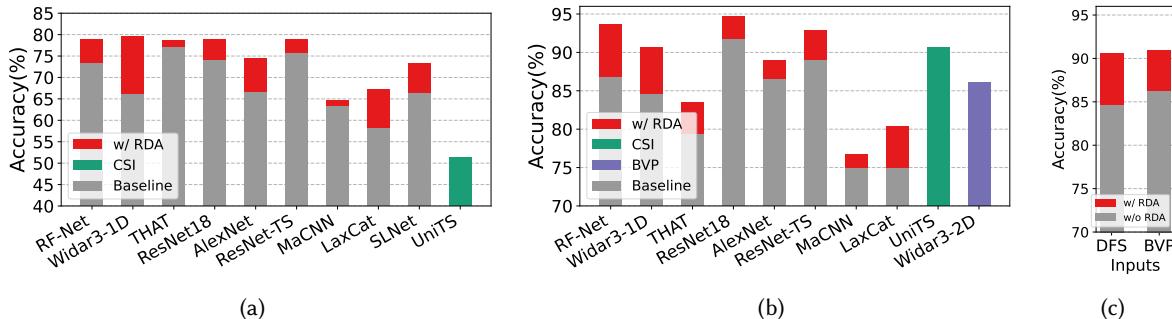


Fig. 9. RDA performance on Widar3 dataset in cross-Rx (a) and in-domain (b,c) evaluations and RDA on BVP compared with DFS using Widar3-2D (c).

of RFBoost, we guarantee convergence of each model under these training conditions by setting a maximum number of epochs and an early stopping criterion based on validation loss plateauing for a predetermined number of patience epochs.

### 5.3 RFBoost Effectiveness

We mainly focus our evaluation on the relative performance gains attributed to RFBoost, especially in terms of generalizability for cross-domain scenarios.

**5.3.1 Evaluation Settings.** In principle, all subcarriers should undergo augmentation. However, doing so is computationally expensive. For instance, training a single augmented epoch on the Widar3 dataset takes over 96 minutes on our NVIDIA A100 GPU server. To accelerate evaluation, we compare using all subcarriers against a selected subset. As shown in Fig. 8, a subset of just six selected subcarriers nearly matches the performance of using all. Implementing RDA on this subset saves 90 minutes per epoch and maintains comparable performance. Consequently, for computational efficiency, our evaluations use only six subcarriers selected by ISS-6 for augmentation, while keeping the baseline as using all subcarriers, *i.e.*, 90 subcarriers for Widar3 dataset. We first highlight the overall effectiveness by using an augmentation ratio (ARatio) of 9 (2× TDA and 6× MDA plus 1× FDA; ARatio calculated in reference to the size of the subset). For fair comparisons, all comparative studies are conducted in the same environment and evaluation settings.

**5.3.2 Gesture Recognition.** We mainly focus on *cross-Rx* evaluation on Widar3-G6 dataset in our effectiveness evaluation. Specifically, Widar3 dataset reports CSI from six Rx devices positioned in various orientations. As a result, *cross-Rx* evaluation assesses the model’s generalization capabilities across both different devices and their orientations/locations. For training, we select CSI from 3 of the 6 Rx while CSI from the other 3 Rx is discarded. In the test phase, we only test the model on the unseen Rx devices in the training phase. We use *cross-Rx* as the default setting unless otherwise specified.

We train and evaluate all the applicable baseline models on the Widar3 dataset with and without RDA, respectively. As shown in Fig. 9a, while the absolute accuracy is dependent on the specific model, RFBoost consistently improves the performance of all models by 5.4% on average, with the highest gain of 13.4% as well as the highest absolute accuracy of 79.63% being achieved on Widar3-1D. We also compare the original Widar3 model (denoted as Widar3-2D) [65], which uses a 2D-CNN+GRU structure and takes *Body-coordinate Velocity Profiles* (BVPs) as inputs.

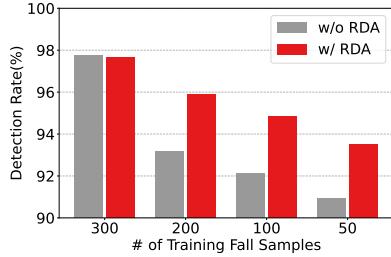


Fig. 10. Detection Rate on FallDar for fall detection.

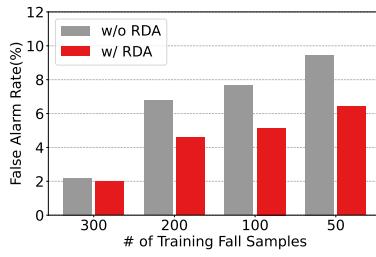


Fig. 11. False Alarm Rate on FallDar for fall detection.

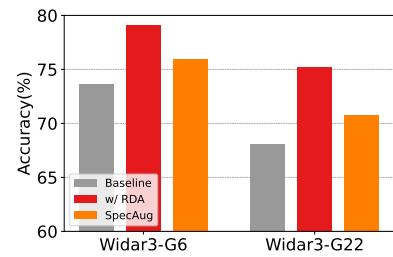


Fig. 12. Cross-Rx evaluation on Widar3-G6 and Widar3-G22 datasets.

We adapted the original Widar3-2D implementation to the PyTorch framework, noting that it processes 3D BVPs across the x-y spatial plane and time axis. To accommodate 2D DFS spectrograms and raw CSI, which encompass the frequency and time axes, we developed Widar3-1D. Comparing Fig. 9b and Fig. 9c, Widar3-2D, even using BVPs, does not yield the best performance compared with other models, especially those augmented by RFBoost. Although Widar3-2D using BVPs does perform better than the adapted Widar3-1D for DFS spectrograms (and much better than using raw CSI as in Fig. 3), the performance difference becomes negligible when both are augmented by RFBoost (As BVPs are calculated from DFS, they can also be augmented by RFBoost). The results validate our insights in §2 that the highly processed information such as BVPs may lose considerable information and thus prevent the best possible learning performance. Additionally, with the raw CSI as input, UniTS performs among the best in in-domain evaluation while dropping to 51% in cross-Rx scenarios, which validates our insights that raw CSI has bad generalization ability due to severe noises. Further considering the intensive overhead to compute BVPs, we argue that BVPs may not be the best choice for deep wireless sensing. Instead, DFS or other forms of spectrograms appear to be a superior representation, especially with RDA.

We also conduct experiments on Widar3-G22, which involves all 22 gestures, to further evaluate the performance of distinguishing similar gesture patterns. As the result shown in Fig. 12, RDA achieves a 7.16% accuracy gain compared with baseline using all subcarriers, compared with the gain of 5.52% on Widar-G6 dataset. As the number of gestures increases, there demands more training samples for models to learn the temporal features, thus, RDA introduces much more performance gain. As depicted in Fig. 13a, w/ RDA, most gestures are classified more accurately, *e.g.*, Gesture7 (Draw-Rectangle(Horizontal)), which is often mistaken as Gesture9 (Draw-Zigzag(Vertical)) or Gesture10 (Draw-N(Vertical)), receives an improvement of accuracy of 7% and reaches up to 85%. We also note that one pair of gestures become even harder to distinguish, which is Gesture18 (Draw-O) and Gesture5 (Draw-0) as they are extremely similar gestures according to the gesture illustration demos released on Widar3 website<sup>2</sup>. Although the error between Gesture5 and Gesture18 is increased, the overall performance of these two gestures are still improved by 4% and 1.2% w/ RDA since RDA improves model's ability to differentiate between these gestures with other patterns.

**5.3.3 Fall Detection.** For fall detection on the FallDar dataset, rather than repeating the evaluation on Widar3, we instead attempt to understand how many training samples for falls are (minimally) needed to achieve a reasonable performance. To answer this, we conduct a different experiment by gradually excluding the training samples for falls and comparing the performance by applying RDA. Specifically, there are around 442 fall samples in total. We use RF-Net as the baseline and train it with gradually reduced amounts of fall samples from 300 to 50 for training. Fig. 10 and Fig. 11 depict the results. As expected, the performance, in terms of detection rate and false alarm rate,

<sup>2</sup><https://ieee-dataport.org/open-access/widar-30-wifi-based-activity-recognition-dataset>

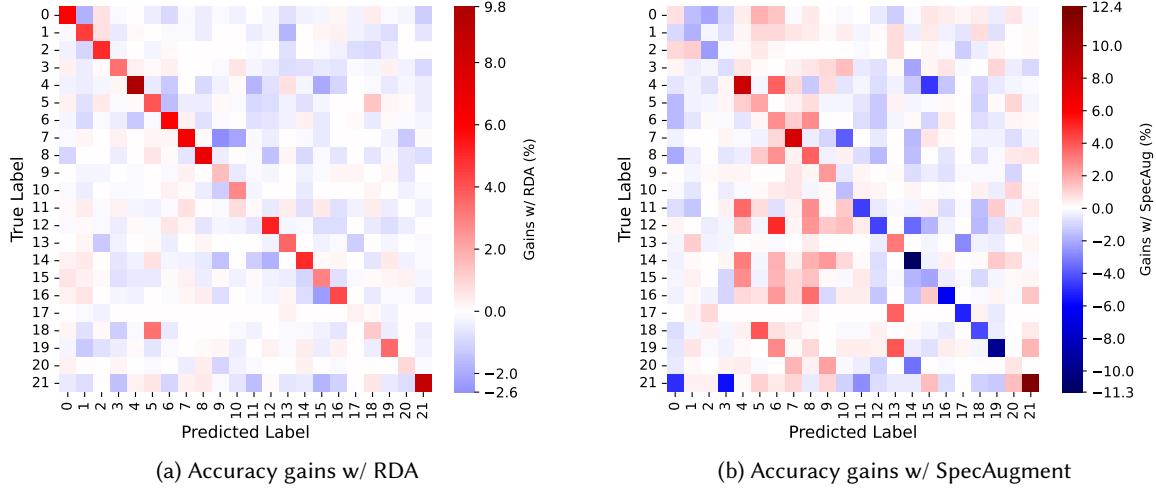


Fig. 13. Confusion matrix for performance gains (compared with confusion matrix of baseline) on Widar3-G22 dataset w/ RDA (a) or w/ SpecAugment (b). Both confusion matrices share the same color map with min&max values marked at color bars.

degrades with respect to shrunk training samples of falls, regardless of applying RDA or not. However, RFBoost improves the performance in any case, with about 2.7% gain in detection rate and on average 2.6% decrease in false alarm rate. More importantly, by applying RDA, we achieve the same performance using only 50 training samples of falls as that of using 200 samples without RDA, remarkably reducing the required amount by 4×.

**5.3.4 Human Activity Recognition.** In assessing RFBoost’s efficacy in addressing dataset imbalances, we turn our attention to the OPERAnet dataset, which is characterized by significant imbalanced label distribution (See Fig. 15). Such imbalances mirror the real-world scenarios in human activity recognition (HAR), where the duration and frequency of activities naturally vary. Typically, longer-duration activities like walking are over-represented, whereas critical but short-duration activities like falling are underrepresented. However, it is often these infrequent activities that are of greatest interest and importance. To investigate whether RFBoost can effectively mitigate the challenges posed by an imbalanced dataset, we employ RF-Net as the baseline model, which achieves an accuracy of 83.69% using the original dataset. Implementing RFBoost results in an enhanced overall accuracy of 96.26% the overall accuracy. As depicted in Fig. 14, the skewed distribution of the dataset adversely impacts the performance of less represented classes, such as ‘sit’, which constitutes a mere 6% of the total dataset. Upon the application of RDA, there is a notable improvement observed across all classes, averaging a 17% increase in performance. Significantly, the underrepresented classes exhibit a more substantial elevation in accuracy, effectively addressing the issue of limited data availability. For instance, the ‘sit’ class witnesses an exceptional increment in performance, surging by 40.6%.

#### 5.4 Cross-domain Generalization Ability

We conduct both in-domain and cross-domain evaluation on Widar3 dataset using RF-Net as a baseline model.

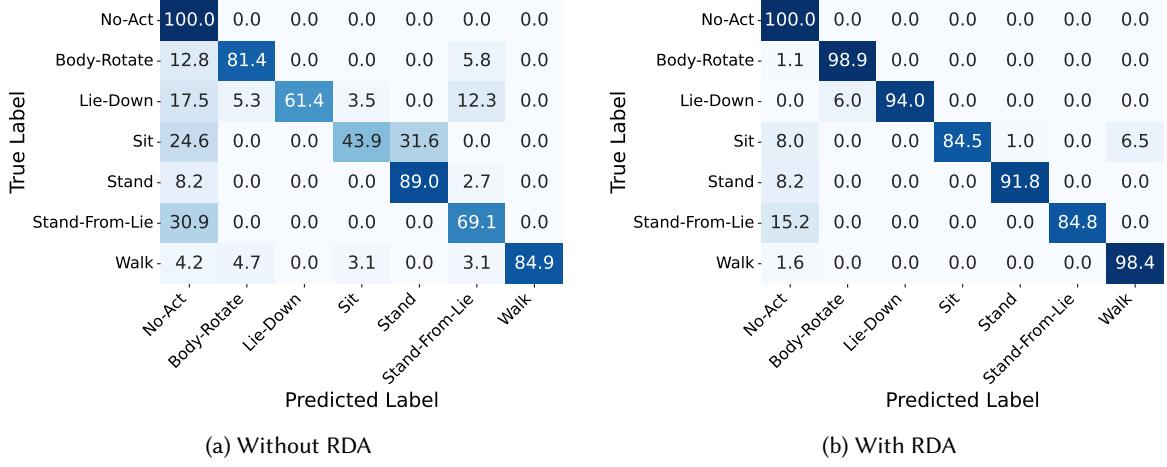


Fig. 14. The confusion matrices on OPERAnet dataset for human activity recognition.

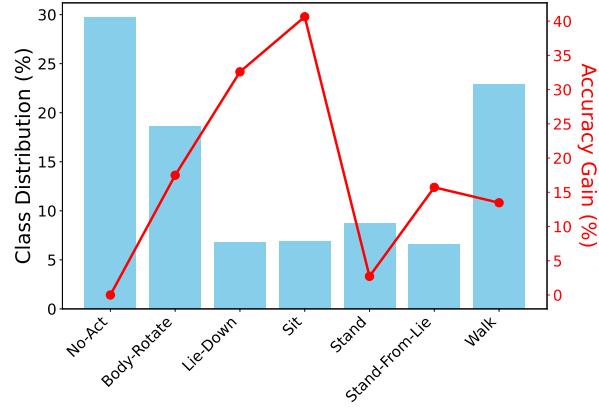


Fig. 15. Class distribution on OPERAnet dataset and accuracy gain after applying RDA.

**5.4.1 In-domain Performance.** Firstly, for in-domain evaluation, as in Fig. 9a, both training and test sets contain samples in the same domain. And RDA achieves an average performance gain of 4.23% over all models.

**5.4.2 Cross-Rx Evaluation.** As mentioned above, RFBoost improves cross-Rx performance by an average of 5.4% on 9 models. We now dive into the performance on each Rx device when training on Widar3-G6 and using RF-Net. Specifically, cross-Rx evaluation constructs a training set with 3 out of 6 Rx devices and evaluates on 3 unseen Rx devices. As depicted in Fig. 16, the result shows improvement for all Rx devices, with an average performance gain of 3.6% and peaking at 4.5% on Rx6.

**5.4.3 Cross-Room and Cross-Person Evaluation.** In addition, we conduct cross-room and cross-person evaluation on Widar3-G6 dataset. The training dataset comprises four individuals in Room#3, and the model's generalization ability is subsequently assessed by testing on other unseen rooms and individuals. As depicted in Fig. 17, the

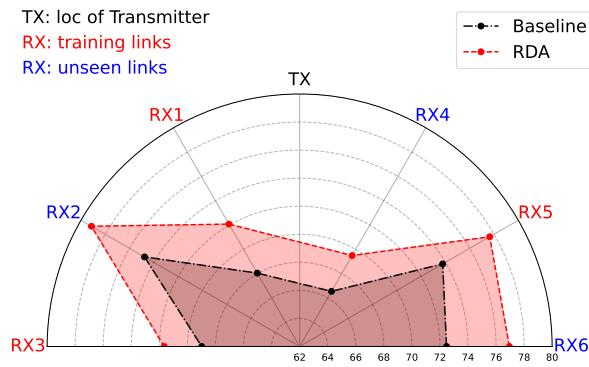


Fig. 16. Performance across links in their actual locations.

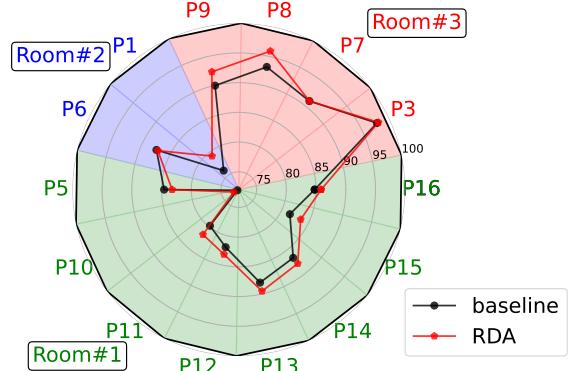


Fig. 17. Performance across Env.(Room) &amp; Persons (Pi).

evaluation highlights an average accuracy enhancement of 0.87% in cross-person scenarios and 1.1% in cross-room scenarios. Notably, Person#17 exhibited atypical performance, and thus, for the sake of clarity in the graphical representation, their data was excluded from the visualization (still included in performance analysis), where the test result of baseline is markedly lower at 46.45%, compared with the average performance of 85.36% among all other individuals.

Notably, we observed that the baseline performance in Cross-Room and Cross-Person evaluations is approximately 10% higher than in the Cross-RX evaluation. However, the improvement observed in these cross-domain evaluations is not as significant as that seen in the Cross-RX evaluation. This discrepancy underscores the need for continued research into robust and adaptable techniques.

## 5.5 Microbenchmarks

**5.5.1 Ablation Study.** Unless stated otherwise, we use RF-Net as the baseline model and focus on cross-Rx evaluation with the Widar3 dataset for the study hereafter. We conduct an ablation study to evaluate the effectiveness of TDA, FDA, and MDA, respectively. As depicted in Fig. 19, the overall performance gain reaches 5.52%, achieved by RDA (ARatio=9, 2×TDA, 6×MDA, 1×GSM). We further decompose RDA into three major modules of RDA, denoted as *TDA-only*, *FDA-only*, and *MDA only*, respectively. Moreover, their combinations are also studied, including *TDA×FDA* and *TDA×MDA*. The result shows standalone policies, *TDA only*, *FDA only*, and *MDA only*, can improve the performance by 1.54%, 2.59%, 3.1%, respectively, while their combination of *TDA×FDA* and *TDA×MDA*, which leave out MDA or FDA, achieve performance gain of 3.48% and 4.43%. Among all three major modules, TDA contributes the least while its combination with other modules still adds to gains by around 1%. We also observe that the performance improvement achieved through the introduction of GSM is not proportional to its high computational demand. By default, GSM requires the calculation of spectrograms for all subcarriers. Considering any budgetary constraints, GSM can be omitted with only a minimal impact on accuracy, specifically a decrease of 1.09%, thereby avoiding the computation of unnecessary spectrograms, saving both storage and computational resources.

**5.5.2 Comparison with Image Data Augmentation (IDA).** We undertake a comparative study to establish RFBoost's superiority over existing IDA techniques in DWS, adhering to the same augmentation ratio of 5 for a fair comparison. RF-Net serves as our baseline. Results in Fig. 18 indicate that while certain IDA methods, like masking, slightly boost accuracy, others such as rotation are detrimental. Duplicating samples yields no benefit. In contrast, RDA outperforms IDA, not only enhancing performance but also offering physical interpretability.

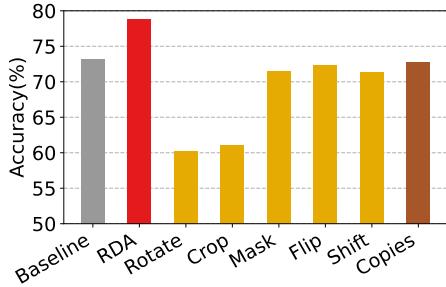


Fig. 18. Comparisons between IDA and RDA.

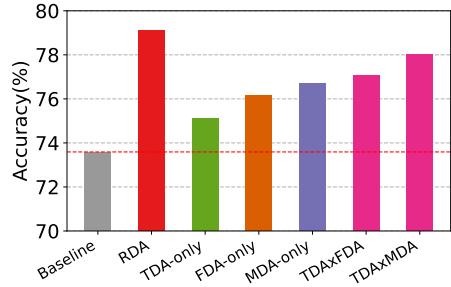


Fig. 19. Ablation study on different RDA policies.

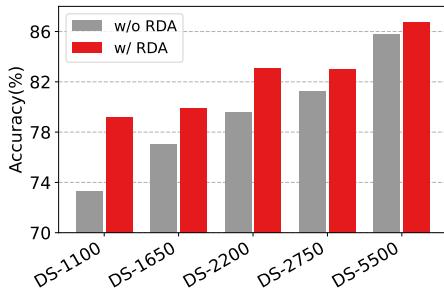


Fig. 20. Impact of data size.

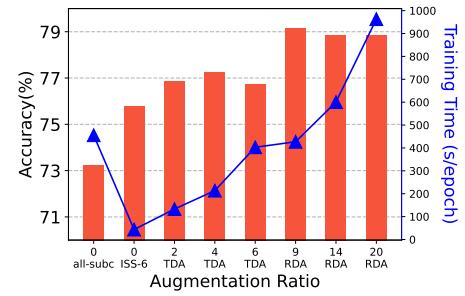


Fig. 21. Impact on Augmentation Ratio.

While this doesn't negate the applicability of all IDA techniques for DWS, it underscores the distinct advantage of RDA.

**5.5.3 Comparison with SpecAugment.** We refer to an open-source repository<sup>3</sup> for the implementation of SpecAugment[33], an augmentation method for mel spectrograms of audio data, featuring time warping, time masking and frequency masking. Thus, there are three parameters, namely, time warp parameter ( $W_{warp}$ ), frequency mask parameter ( $F_{freq}$ ), and time mask parameter ( $T_{time}$ ). All parameters set the boundaries for the random magnitude of data manipulations (warping or masking). For time and frequency masking, we pad the masked region with the mean value of the whole spectrogram. Note that, these parameters have to be searched manually and we select them such that the size of masking and range of warping closely match in magnitude with those used (size of masking and range of shifting) in our proposed *Motion-aware Data Augmentation*, where  $W_{warp} = 5$ ,  $F_{freq} = 20$ ,  $T_{time} = 20$ . The result shows SpecAugment achieves a performance gain of 2.39% in Cross-Rx evaluation on Widar3-G6 dataset, compared with 5.52% using RDA and 3.1% using *Motion-aware Data Augmentation-only*. The difference between *MDA* and the masking in SpecAugment is whether or not involving the information of motion in shifting distance decision and masking region selection. Additional gains when involving identified motion period also prove our design rationale in §3.3 that masking motion-specific region can alleviate overfitting towards local features and thus improve cross-Rx performance. We also evaluate SpecAugment on Widar3-G22 dataset, as

<sup>3</sup>[https://github.com/zcaceres/spec\\_augment](https://github.com/zcaceres/spec_augment)

depicted in Fig. 12. SpecAugment improves the performance by 2.67% while RDA achieves 7.16%. We further visualize the confusion matrices of both methods by analyzing how much gains they achieve compared with the baseline, shown in Fig. 13. We observe that although SpecAugment improves the performance, it seems to show a preference for the gesture types, such as Gesture4-9, and can even improve Gesture21 by 12.4%. This implies SpecAugment only suits for specific gesture patterns. Differently, our proposed RDA works positively for most of these 22 gestures.

**5.5.4 Impact of Original Data Size.** Unlike most generative models, RFBoost performs well with smaller training sets, but its performance varies depending on the data size. To examine the impact of training data size and assess the potential for reducing data collection efforts, we downsample the Widar3 training set to sizes of 1100, 1650, 2200, 2750, and 5500 and apply our RDA method while keeping the testing set consistent. The outcomes, as depicted in Fig. 20, reveal that RDA empowers smaller datasets to reach performance levels equivalent to those achieved with datasets up to twice their size. This finding illustrates the efficacy of RDA in enhancing model performance through efficient data utilization, thereby providing a substantial avenue for addressing data scarcity and optimizing performance even when data availability is limited.

**5.5.5 Augmentation Ratio.** In this experiment, we assess the influence of the augmentation ratio (ARatio), which is the ratio of the augmented to the original dataset size. Fig. 21 reveals that higher ARatios improve performance but also incur greater computational cost and extended training time. For the Widar3 dataset, an ARatio of 9 provides a balanced trade-off between accuracy and computational overhead.

**5.5.6 Subc-as-channel v.s. Subc-as-sample.** CV-based DNN models often support multi-channel input, exemplified by PyTorch ResNet implementations [34] that accept input shapes of  $(C, H, W)$ , where  $C$  signifies channel dimensions like RGB. For DWS models that consider spectrograms as input, certain DWS models draw inspiration from CV architectures, aligning subcarriers in the channel dimension. Conversely, the additional dimension in transitioning from CSI to spectrograms necessitates treating subcarriers as samples or aggregating them in models like RF-Net.

We assess these contrasting approaches using Widar3 trained on ResNet18. By consolidating all subcarriers into a single sample plus an augmented sample via RFBoost (ARatio = 1), we find that Subcarrier-as-channel outperforms Subcarrier-as-sample by 6%, with performance boosts of 1.83% and 3.64% when augmented with RDA, respectively. These findings advocate for treating subcarriers as channels in CV-inspired DNN models. Regardless of the input configuration, RDA enhances performance across the board.

**5.5.7 Comparative Analysis of Subcarrier Selection Methods.** Within RFBoost, Individual Subcarrier Selection (ISS) has emerged as a prudent approach for choosing representative subcarriers, significantly reducing training efforts. Although RFBoost's primary focus is on data augmentation rather than selection, it prompts an intriguing question: 'How well do existing subcarrier selection methods perform in deep wireless sensing?' We aim to address this question from an experimental perspective: Despite PCA's widespread adoption for effective dimension reduction while maximizing CSI variances in wireless sensing, its efficacy in deep wireless sensing is limited. In our study, we juxtapose Individual Subcarrier Selection (ISS) with PCA on the Widar3 datasets, employing the top-1, top-2, and top-3 principal components (PCA-1, PCA-2, PCA-3). Notably, PCA, when applied subcarrier-wise as in [32, 65], results in accuracies of 63.1%, 62.65%, and 61.14% respectively, which are approximately 9-10% lower than using three randomly selected subcarriers (Rand-3), and significantly lower than ISS-3. As depicted in Fig. 8 and Fig. 22a, we observed that PCA's performance is similar to a single subcarrier selected based on motion statistics (ISS-1), a criterion that provides physical context which is absent in PCA. In addition to PCA, we assess some traditional subcarrier selection methods like TopVar and TopMean in deep wireless sensing scenarios, which focus on maximum variance and mean amplitude, respectively. To better understand ISS, we examined a variant of ISS, termed TopMS, which selects subcarriers based on the highest motion statistics across the entire frequency

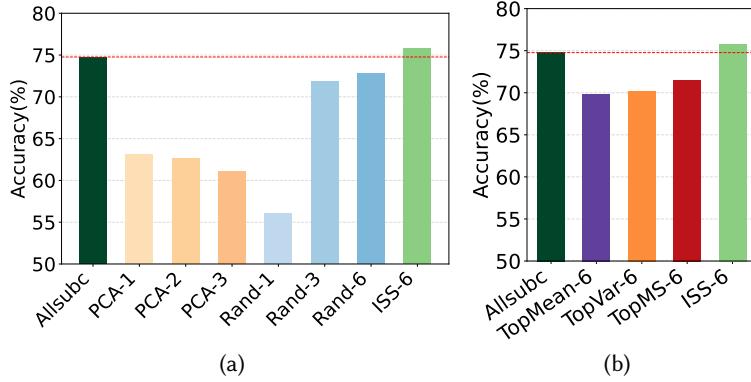


Fig. 22. Comparative analysis between using all subcarriers (Allsubc) and various metrics for subcarrier selection. (a) Allsubc v.s. top- $i$  principle components of all subcarriers (PCA- $i$ ), and randomly selected  $i$  subcarriers (rand- $i$ ), (b) Allsubc v.s. subcarriers with top-6 mean value (TopMean-6), subcarriers with top-6 variance (TopVar-6), subcarriers with top-6 motion statistics (TopMS-6), and ISS-6.

band instead of among sub-bands as ISS does (See §3.3). As illustrated in Fig. 22b, TopMS-6 outperforms all other intuitive methods but falls 4.32% short of ISS-6. This is because ISS-6 leverages both physical characteristics and the diversity inherent in distinct sub-bands, resulting in a more comprehensive and motion-aware subcarrier selection method.

**5.5.8 Test-Time Augmentation Performance.** To evaluate the effect of test-time augmentation, we re-run the test phase, using the previously trained RF-Net on Widar3, by augmenting each test sample in the same way as the training data. For the same test sample, majority voting is applied to aggregate multiple copies generated by augmentation methods. Our results show that test-time augmentation can bring an additional improvement of 0.7%, further increasing the best RDA accuracy from 78.84% to 79.54%. The computation overhead also increases marginally, adding an extra delay of 27 ms to each test sample with an ARatio of 9.

**5.5.9 Computational Complexity.** We mainly focus on both the time and space complexity of RFBoost. For time complexity, as mentioned in §5.3, we downsample all subcarriers before applying RDA, which greatly reduces the time for training the augmented dataset while achieving similar high performance as depicted in Fig. 21. We highlight that RFBoost cache-and-fetch mechanism also saves approx. 50% training time from the second epoch compared with 1st epoch for every augmentation method. We further analyze the space complexity, which apparently also increases as RDA creates more data. Note that, all the computation overhead only occurs in the training stage, which is usually done offline and therefore less time-sensitive. The raw CSI data for Widar3 dataset occupies 78 GB, and the extra space usage increases linearly with the augmentation ratio.

## 6 DISCUSSIONS AND FUTURE WORKS

RFBoost pioneers the study of radio data augmentation for deep wireless sensing. The experimental results show encouraging performance as well as inspire many follow-up research topics. There is certainly room for improvement and we discuss some limitations and future works here.

### 6.1 RDA Beyond Spectrograms

The current RFBoost is mainly designed within the scope of WiFi sensing using spectrogram learning [55]. We have already shown the proposed approaches can also be used for highly processed data representations like BVPs [20, 65], which are based on DFS spectrograms. The proposed methods can also be directly applied to other forms of spectrograms, *e.g.*, those generated by ACF calculation [47, 56]. Furthermore, we believe the idea of PDA can be extended to more data representations of WiFi. We are particularly interested in investigating and incorporating raw CSI augmentation in RFBoost.

### 6.2 RDA Beyond WiFi

Another exciting topic is to exploit RDA wireless sensing beyond WiFi, *e.g.*, millimeter wave (mmWave), UWB radars, etc. For example, in addition to time diversity, there are generally more antennas available for mmWave sensing, which underpin more space diversity. For FMCW mmWave radar, it is also possible to explore segmented frequency bands to increase frequency diversity [17]. Both can be explored for data augmentation for mmWave sensing. Despite increasing efforts on data synthesis for mmWave sensing recently [1, 2, 39], physical data augmentation is still lacking.

### 6.3 Auto Augment

Our experimental results show that there is seemingly an optimal size for post-augmented data. How to automatically determine the optimal final dataset size and select the best augmentation policies is thus another interesting question to answer. Auto augment has attracted numerous efforts with remarkable advances in image data augmentation [71], which can inspire auto augment for RDA.

A promising direction is to leverage diversity and affinity metrics [12] to automatically determine the policies and augmentation size that maximize the two values, without going through the heavy training process.

### 6.4 Synthetic Data Generation

We currently mainly focus on data warping augmentation, which transforms existing data for more samples. We note that RFBoost is true data augmentation, rather than data selection, as it generates new samples that do not exist in the original dataset. Synthetic data generation based on deep learning, *e.g.*, GANs, can also create synthetic virtual samples for training. There are some recent efforts in this direction by using adversarial learning [18], diffusion models [6], and/or resorting to a secondary modality like video data or IMU data [1, 2, 4]. The proposed approaches in RFBoost are complementary to these approaches and can be used jointly. Our future work will explore the combination of physical data augmentation and synthetic data generation, for which we are particularly interested in physics-inspired data generation.

## 7 RELATED WORKS

This work targets a new problem of radio data augmentation for deep wireless sensing. We briefly review the related literature on wireless sensing and data augmentation techniques.

### 7.1 Deep Wireless Sensing

Traditional approaches for wireless sensing mostly rely on signal processing and/or conventional machine learning, which have seen years of development and achieved great performance for some tasks [11, 16, 35, 42, 47, 50, 60, 61]. Recently, with the great success of deep learning in many applications, deep learning has been used in wireless sensing to achieve encouraging performance [5, 8, 24, 30, 43, 49, 53, 65, 70]. These solutions typically adopt existing neural networks and take either raw data as inputs or convert them into the frequency domain or certain feature space. For example, Widar3.0 [65] leverages CNN and RNN networks and estimates the body velocity

profiles as a location-independent representation. RFPose [66], RFPose3D [68], and RFAvatar [67] also use CNN models to capture body skeleton and mesh. EI [19], CrossSense [62], RF-Net [8] and Wi-Learner [10] use more sophisticated networks such as adversarial learning, transfer learning, meta-learning, mainly to handle the issue of environment dependency. WiSIA [24] employs conditional GAN to achieve WiFi imaging, which is also studied in [20, 40] with deep learning. Besides high accuracy, a particular focus of DWS solutions is to reduce the need for large-scale training data. For example, many of them aim to achieve one-shot learning or cross-domain learning [8, 13, 49, 65]. Some works take a further step to customize the networks in the hope of better fitting time series data. STFNets [59], which extends DeepSense [58], and UniTS [25] are both customized for temporal-spatial learning with STFT operators. STFNets also exploits time diversity with multi-resolution windows, but not for the purpose of data augmentation.

Different from these methods, RFBoost approaches the problem of limited data from the root, *i.e.*, the training data, and presents a data-space solution of radio data augmentation, which is orthogonal to the above solutions.

## 7.2 Data Augmentation

Some solutions explore beyond wireless and exploit cross-modality sensing and data synthesis to mitigate the data scarcity issue in DWS. IMU2Doppler [2] uses IMU datasets to train a network for mmWave signals with a small amount of labeled data. XModal-ID [21] considers the problem of WiFi-based person identification from a given video footage. A more interesting solution is proposed in [4] to translate online videos to instant simulated RF data for training DWS systems without massive RF data collection. Vid2Doppler [1] also synthesizes Doppler mmWave data from videos. Different from these methods, our proposed RDA aims to inflate existing training data without collecting more or resorting to a secondary modality.

Data augmentation is only primarily touched in the field. For example, MoRe-Fi [70] rotates the raw signals in the I/Q-plane for augmentation. OneFi [49] also augments data by rotating the velocity distribution, which is only applicable to specific tasks. Image data augmentation techniques like noise injection, scaling, cropping, and image mixture are applied to RF spectrograms in [63] for human activity recognition. Techniques like generative adversarial learning have been used to synthesize virtual samples for WiFi [18, 26, 48, 52, 56] or radar [31, 39, 51] signals. Differently, RFBoost systematically studies the problem of RDA and presents physical data augmentation, rather than generating virtual samples. Nevertheless, RDA is complementary to and can be stacked on top of existing cross-modality training or data synthesis methods.

Data augmentation is a well-studied area in CV and is extended to other fields such as speech processing [33]. A comprehensive overview is provided in [38] for literature review of image data augmentation. Although most image data augmentation techniques do not perform the best on wireless data, they inspire our study of a counterpart problem of radio data augmentation and motivate our design of RFBoost for deep wireless sensing.

## 8 CONCLUSIONS

This paper presents RFBoost, the first radio data augmentation framework for deep wireless sensing in the scope of WiFi sensing. We propose a set of physical data augmentation techniques that inflate existing training data significantly and effectively. We achieve so by leveraging the inherent data diversity of wireless signals. We implement RFBoost and integrate it with the state-of-the-art DWS models and evaluate it on three different WiFi sensing datasets. The results show that RFBoost achieves remarkable accuracy improvements on baseline models and datasets. Furthermore, by simply performing RDA with a common CV network, we can achieve comparable or even better performance than customized networks for RF signals with highly-processed inputs. RFBoost pioneers the study and inspires follow-up research on radio data augmentation. We also plan to extend RFBoost to more wireless data representations and other wireless sensing modalities, which we believe will become a standard component for DWS to overcome the problem of limited data.

## ACKNOWLEDGMENT

This work is supported in part by NSFC under grant No. 62222216 and Hong Kong RGC ECS under grant 27204522 and RIF under grant R5060-19.

## REFERENCES

- [1] Karan Ahuja, Yue Jiang, Mayank Goel, and Chris Harrison. 2021. Vid2Doppler: synthesizing Doppler radar data from videos for training privacy-preserving activity recognition. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–10.
- [2] Sejal Bhalla, Mayank Goel, and Rushil Khurana. 2021. IMU2Doppler: Cross-Modal Domain Adaptation for Doppler-based Activity Recognition Using IMU Data. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 5, 4 (2021), 1–20.
- [3] Mohammad Junaid Bocus, Wenda Li, Shelly Vishwakarma, Roget Kou, Chong Tang, Karl Woodbridge, Ian Craddock, Ryan McConville, Raúl Santos-Rodríguez, Kevin Chetty, and Robert J. Piechocki. 2022. OPERANet, a multimodal activity recognition dataset acquired from radio frequency and vision-based sensors. *Scientific Data* 9 (2022).
- [4] Hong Cai, Belal Korany, Chitra R Karanam, and Yasamin Mostofi. 2020. Teaching rf to sense without rf training measurements. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 4 (2020), 1–22.
- [5] Zhe Chen, Tianyue Zheng, Chao Cai, and Jun Luo. 2021. MoVi-Fi: Motion-robust vital signs waveform recovery via deep interpreted RF sensing. In *Proceedings of the 27th Annual International Conference on Mobile Computing and Networking*. 392–405.
- [6] Guoxuan Chi, Zheng Yang, Chenshu Wu, Jingao Xu, Yuchong Gao, and Tony Xiao Liu, Yunhao Han. 2024. RF-Diffusion: Radio Signal Generation via Time-Frequency Diffusion. In *The 30th Annual International Conference on Mobile Computing and Networking (ACM MobiCom'24)*. ACM.
- [7] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. 2009. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*. Ieee, 248–255.
- [8] Shuya Ding, Zhe Chen, Tianyue Zheng, and Jun Luo. 2020. RF-net: A unified meta-learning framework for RF-enabled one-shot human activity recognition. In *Proceedings of the 18th Conference on Embedded Networked Sensor Systems*. 517–530.
- [9] F. Dornaika, D. Sun, K. Hammoudi, J. Charafeddine, A. Cabani, and C. Zhang. 2023. Object-centric Contour-aware Data Augmentation Using Superpixels of Varying Granularity. *Pattern Recognition* 139 (2023), 109481. <https://doi.org/10.1016/j.patcog.2023.109481>
- [10] Chao Feng, Nan Wang, Yicheng Jiang, Xia Zheng, Kang Li, Zheng Wang, and Xiaojiang Chen. 2022. Wi-Learner: Towards One-Shot Learning for Cross-Domain Wi-Fi Based Gesture Recognition. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 6, 3, Article 114 (sep 2022), 27 pages. <https://doi.org/10.1145/3550318>
- [11] Yuda Feng, Yaxiong Xie, Deepak Ganeshan, and Jie Xiong. 2021. LTE-based Pervasive Sensing Across Indoor and Outdoor. In *Proceedings of the 19th ACM Conference on Embedded Networked Sensor Systems*. 138–151.
- [12] Raphael Gontijo-Lopes, Sylvia Smullin, Ekin Dogus Cubuk, and Ethan Dyer. 2020. Tradeoffs in data augmentation: An empirical study. In *Proceedings of International Conference on Learning Representations*.
- [13] Yu Gu, Huan Yan, Mianxiong Dong, Meng Wang, Xiang Zhang, Zhi Liu, and Fuji Ren. 2021. Wione: One-shot learning for environment-robust device-free user authentication via commodity wi-fi in man-machine system. *IEEE Transactions on Computational Social Systems* 8, 3 (2021), 630–642.
- [14] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 770–778.
- [15] Tsung-Yu Hsieh, Suhang Wang, Yiwei Sun, and Vasant Honavar. 2021. Explainable multivariate time series classification: a deep neural network which learns to attend to important variables as well as time intervals. In *Proceedings of the 14th ACM international conference on web search and data mining*. 607–615.
- [16] Yuqian Hu, Feng Zhang, Chenshu Wu, Beibei Wang, and KJ Ray Liu. 2021. DeFall: Environment-Independent Passive Fall Detection using WiFi. *IEEE Internet of Things Journal* (2021).
- [17] Chengkun Jiang, Junchen Guo, Yuan He, Meng Jin, Shuai Li, and Yunhao Liu. 2020. mmVib: micrometer-level vibration measurement with mmwave radar. In *Proceedings of the 26th Annual International Conference on Mobile Computing and Networking*. 1–13.
- [18] Dehao Jiang, Mingqi Li, and Chunling Xu. 2020. Wigan: A wifi based gesture recognition system with gans. *Sensors* 20, 17 (2020), 4757.
- [19] Wenjun Jiang, Chenglin Miao, Fenglong Ma, Shuochao Yao, Yaqing Wang, Ye Yuan, Hongfei Xue, Chen Song, Xin Ma, Dimitrios Koutsonikolas, et al. 2018. Towards environment independent device free human activity recognition. In *Proceedings of the 24th annual international conference on mobile computing and networking*. 289–304.
- [20] Wenjun Jiang, Hongfei Xue, Chenglin Miao, Shiyang Wang, Sen Lin, Chong Tian, Srinivasan Murali, Haochen Hu, Zhi Sun, and Lu Su. 2020. Towards 3D human pose construction using WiFi. In *Proceedings of the 26th Annual International Conference on Mobile Computing and Networking*. 1–14.
- [21] Belal Korany, Chitra R Karanam, Hong Cai, and Yasamin Mostofi. 2019. XModal-ID: Using WiFi for through-wall person identification from candidate video footage. In *The 25th Annual International Conference on Mobile Computing and Networking*. 1–15.

- [22] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. 2012. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems* 25 (2012).
- [23] Bing Li, Wei Cui, Wei Wang, Le Zhang, Zhenghua Chen, and Min Wu. 2021. Two-stream convolution augmented transformer for human activity recognition. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 35. 286–293.
- [24] Chenning Li, Zheng Liu, Yuguang Yao, Zhichao Cao, Mi Zhang, and Yunhao Liu. 2020. Wi-fi see it all: generative adversarial network-augmented versatile wi-fi imaging. In *Proceedings of the 18th Conference on Embedded Networked Sensor Systems*. 436–448.
- [25] Shuheng Li, Ranak Roy Chowdhury, Jingbo Shang, Rajesh K Gupta, and Dezhong Hong. 2021. UniTS: Short-Time Fourier Inspired Neural Networks for Sensory Time Series Classification. In *Proceedings of the 19th ACM Conference on Embedded Networked Sensor Systems*. 234–247.
- [26] Xinyi Li, Liqiong Chang, Fangfang Song, Ju Wang, Xiaojiang Chen, Zhanyong Tang, and Zheng Wang. 2021. CrossGR: Accurate and Low-cost Cross-target Gesture Recognition Using Wi-Fi. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 5, 1 (March 2021), 1–23. <https://doi.org/10.1145/3448100>
- [27] Xuefeng Liu, Jiannong Cao, Shaojie Tang, and Jiaqi Wen. 2014. Wi-sleep: Contactless sleep monitoring via wifi signals. In *2014 IEEE Real-Time Systems Symposium*. IEEE, 346–355.
- [28] Yongsen Ma, Gang Zhou, and Shuangquan Wang. 2019. WiFi sensing with channel state information: A survey. *ACM Computing Surveys (CSUR)* 52, 3 (2019), 1–36.
- [29] Yongsen Ma, Gang Zhou, Shuangquan Wang, Hongyang Zhao, and Woosub Jung. 2018. SignFi: Sign language recognition using WiFi. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 1 (2018), 1–21.
- [30] Muhammed Zahid Ozturk, Chenshu Wu, Beibei Wang, and KJ Ray Liu. 2021. GaitCube: Deep data cube learning for human recognition with millimeter-wave radio. *IEEE Internet of Things Journal* 9, 1 (2021), 546–557.
- [31] Muhammed Zahid Ozturk, Chenshu Wu, Beibei Wang, and KJ Ray Liu. 2022. Toward mmWave-Based Sound Enhancement and Separation. In *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 6852–6856.
- [32] Sameera Palipana, David Rojas, Piyush Agrawal, and Dirk Pesch. 2018. FallDeFi: Ubiquitous fall detection using commodity Wi-Fi devices. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 4 (2018), 1–25.
- [33] Daniel S. Park, William Chan, Yu Zhang, Chung-Cheng Chiu, Barret Zoph, Ekin D. Cubuk, and Quoc V. Le. 2019. SpecAugment: A Simple Data Augmentation Method for Automatic Speech Recognition. In *Proceedings of Interspeech 2019*. 2613–2617.
- [34] PyTorch. 2023. PyTorch ResNet. [https://pytorch.org/hub/pytorch\\_vision\\_resnet/](https://pytorch.org/hub/pytorch_vision_resnet/). Accessed: Aug 2023.
- [35] Kun Qian, Chenshu Wu, Yi Zhang, Guidong Zhang, Zheng Yang, and Yunhao Liu. 2018. Widar2. 0: Passive human tracking with a single Wi-Fi link. In *Proceedings of the 16th Annual International Conference on Mobile Systems, Applications, and Services*. 350–361.
- [36] Valentin Radu, Catherine Tong, Sourav Bhattacharya, Nicholas D Lane, Cecilia Mascolo, Mahesh K Marina, and Fahim Kawsar. 2018. Multimodal deep learning for activity and context recognition. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 4 (2018), 1–27.
- [37] Divya Shanmugam, Davis Blalock, Guha Balakrishnan, and John Guttag. 2021. Better aggregation in test-time augmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 1214–1223.
- [38] Connor Shorten and Taghi M Khoshgoftaar. 2019. A survey on image data augmentation for deep learning. *Journal of big data* 6, 1 (2019), 1–48.
- [39] Shelly Vishwakarma, Wenda Li, Chong Tang, Karl Woodbridge, Raviraj Adve, and Kevin Chetty. 2021. SimHumalator: An Open-Source End-to-End Radar Simulator for Human Activity Recognition. *IEEE Aerospace and Electronic Systems Magazine* 37, 3 (2021), 6–22.
- [40] Fei Wang, Sanping Zhou, Stanislav Panev, Jinsong Han, and Dong Huang. 2019. Person-in-WiFi: Fine-grained person perception using WiFi. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 5452–5461.
- [41] Wei Wang, Alex X Liu, and Muhammad Shahzad. 2016. Gait recognition using wifi signals. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. 363–373.
- [42] Wei Wang, Alex X Liu, Muhammad Shahzad, Kang Ling, and Sanglu Lu. 2015. Understanding and modeling of wifi signal based human activity recognition. In *Proceedings of the 21st annual international conference on mobile computing and networking*. 65–76.
- [43] Yanwen Wang, Jiaxing Shen, and Yuanqing Zheng. 2020. Push the limit of acoustic gesture recognition. *IEEE Transactions on Mobile Computing* (2020).
- [44] Zhiguang Wang, Weizhong Yan, and Tim Oates. 2017. Time series classification from scratch with deep neural networks: A strong baseline. In *2017 International Joint Conference on Neural Networks (IJCNN)*. 1578–1585.
- [45] Chenshu Wu, Beibei Wang, Oscar C. Au, and K. J. Ray Liu. 2022. Wi-Fi Can Do More: Towards Ubiquitous Wireless Sensing. *IEEE Communications Standards Magazine* (2022).
- [46] Chenshu Wu, Zheng Yang, Zimu Zhou, Xuefeng Liu, Yunhao Liu, and Jiannong Cao. 2015. Non-invasive detection of moving and stationary human with WiFi. *IEEE Journal on Selected Areas in Communications* 33, 11 (2015), 2329–2342.
- [47] Chenshu Wu, Feng Zhang, Yuqian Hu, and KJ Ray Liu. 2020. GaitWay: Monitoring and recognizing gait speed through the walls. *IEEE Transactions on Mobile Computing* 20, 6 (2020), 2186–2199.

- [48] Chunjing Xiao, Daojun Han, Yongsen Ma, and Zhiguang Qin. 2019. CsiGAN: Robust Channel State Information-Based Activity Recognition With GANs. *IEEE Internet of Things Journal* 6 (2019), 10191–10204.
- [49] Rui Xiao, Jianwei Liu, Jinsong Han, and Kui Ren. 2021. OneFi: One-Shot Recognition for Unseen Gesture via COTS WiFi. In *Proceedings of the 19th ACM Conference on Embedded Networked Sensor Systems*. 206–219.
- [50] Binbin Xie and Jie Xiong. 2020. Combating interference for long range LoRa sensing. In *Proceedings of the 18th Conference on Embedded Networked Sensor Systems*. 69–81.
- [51] Hongfei Xue, Qiming Cao, Chenglin Miao, Yan Ju, Haochen Hu, Aidong Zhang, and Lu Su. 2023. Towards Generalized mmWave-based Human Pose Estimation through Signal Augmentation. In *Proceedings of the 29th Annual International Conference on Mobile Computing and Networking*. ACM, Madrid Spain, 1–15. <https://doi.org/10.1145/3570361.3613302>
- [52] Kun Yang, Xiaolong Zheng, Jie Xiong, L. Liu, and Huadong Ma. 2022. WiImg: Pushing the Limit of WiFi Sensing with Low Transmission Rates. *2022 19th Annual IEEE International Conference on Sensing, Communication, and Networking (SECON)* (2022), 1–9.
- [53] Qiang Yang and Yuanqing Zheng. 2022. DeepEar: Sound Localization with Binaural Microphones. In *Proceedings of the IEEE INFOCOM*.
- [54] Y. Yang, Yutong Zhang, Yue Lang, Beichen Li, Shisheng Guo, and Qi Tan. 2023. GAN-Based Radar Spectrogram Augmentation via Diversity Injection Strategy. *IEEE Transactions on Instrumentation and Measurement* 72 (2023), 1–12.
- [55] Zheng Yang, Yi Zhang, Kun Qian, and Chenshu Wu. 2023. SLNet: A Spectrogram Learning Neural Network for Deep Wireless Sensing. In *20th USENIX Symposium on Networked Systems Design and Implementation (NSDI 23)*. USENIX Association, Boston, MA, 1221–1236. <https://www.usenix.org/conference/nsdi23/presentation/yang-zheng>
- [56] Zheng Yang, Yi Zhang, and Qian Zhang. 2022. Rethinking Fall Detection With Wi-Fi. *IEEE Transactions on Mobile Computing* (2022).
- [57] Zheng Yang, Zimu Zhou, and Yunhao Liu. 2013. From RSSI to CSI: Indoor localization via channel response. *ACM Computing Surveys (CSUR)* 46, 2 (2013), 1–32.
- [58] Shuochao Yao, Shaohan Hu, Yiran Zhao, Aston Zhang, and Tarek Abdelzaher. 2017. Deepsense: A unified deep learning framework for time-series mobile sensing data processing. In *Proceedings of the 26th ACM International Conference on World Wide Web*. 351–360.
- [59] Shuochao Yao, Ailing Piao, Wenjun Jiang, Yiran Zhao, Huajie Shao, Shengzhong Liu, Dongxin Liu, Jinyang Li, Tianshi Wang, Shaohan Hu, et al. 2019. Stfnets: Learning sensing signals from the time-frequency perspective with short-time fourier neural networks. In *The World Wide Web Conference*. 2192–2202.
- [60] Feng Zhang, Chenshu Wu, Beibei Wang, Hung-Quoc Lai, Yi Han, and KJ Ray Liu. 2019. WiDetect: Robust motion detection with a statistical electromagnetic model. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 3, 3 (2019), 1–24.
- [61] Feng Zhang, Chenshu Wu, Beibei Wang, Min Wu, Daniel Bugos, Hangfang Zhang, and KJ Ray Liu. 2019. SMARS: Sleep monitoring via ambient radio signals. *IEEE Transactions on Mobile Computing* 20, 1 (2019), 217–231.
- [62] Jie Zhang, Zhanyong Tang, Meng Li, Dingyi Fang, Petteri Nurmi, and Zheng Wang. 2018. CrossSense: Towards cross-site and large-scale WiFi sensing. In *Proceedings of the 24th Annual International Conference on Mobile Computing and Networking*. 305–320.
- [63] Jin Zhang, Fuxiang Wu, Bo Wei, Qieshi Zhang, Hui Huang, Syed W. Shah, and Jun Cheng. 2021. Data Augmentation and Dense-LSTM for Human Activity Recognition Using WiFi Signal. *IEEE Internet of Things Journal* 8, 6 (2021), 4628–4641. <https://doi.org/10.1109/JIOT.2020.3026732>
- [64] Xiaotong Zhang, Zhenjiang Li, and Jin Zhang. 2022. Synthesized Millimeter-Waves for Human Motion Sensing. (2022).
- [65] Yi Zhang, Yue Zheng, Kun Qian, Guidong Zhang, Yunhao Liu, Chenshu Wu, and Zheng Yang. 2021. Widar3.0: Zero-Effort Cross-Domain Gesture Recognition with Wi-Fi. *IEEE transactions on pattern analysis and machine intelligence* PP (2021).
- [66] Mingmin Zhao, Tianhong Li, Mohammad Abu Alsheikh, Yonglong Tian, Hang Zhao, Antonio Torralba, and Dina Katabi. 2018. Through-Wall Human Pose Estimation Using Radio Signals. In *Proceedings of IEEE/CVF CVPR*.
- [67] Mingmin Zhao, Yingcheng Liu, Aniruddh Raghu, Hang Zhao, Tianhong Li, Antonio Torralba, and Dina Katabi. 2019. Through-Wall Human Mesh Recovery Using Radio Signals. In *Proceedings of IEEE/CVF ICCV*.
- [68] Mingmin Zhao, Yonglong Tian, Hang Zhao, Mohammad Abu Alsheikh, Tianhong Li, Rumen Hristov, Zachary Kabelac, Dina Katabi, and Antonio Torralba. 2018. RF-Based 3D Skeletons. In *Proceedings of ACM SIGCOMM*.
- [69] Tianyue Zheng, Zhe Chen, Shuya Ding, and Jun Luo. 2021. Enhancing RF sensing with deep learning: A layered approach. *IEEE Communications Magazine* 59, 2 (2021), 70–76.
- [70] Tianyue Zheng, Zhe Chen, Shujie Zhang, Chao Cai, and Jun Luo. 2021. MoRe-Fi: Motion-robust and Fine-grained Respiration Monitoring via Deep-Learning UWB Radar. In *Proceedings of the 19th ACM Conference on Embedded Networked Sensor Systems*. 111–124.
- [71] Yu Zheng, Zhi Zhang, Shen Yan, and Mi Zhang. 2022. Deep autoaugment. In *Proceedings of International Conference on Learning Representations*.
- [72] Zhun Zhong, Liang Zheng, Guoliang Kang, Shaozi Li, and Yi Yang. 2017. Random Erasing Data Augmentation. arXiv:1708.04896 [cs.CV]