

SiMWiSense: Simultaneous Multi-Subject Activity Classification Through Wi-Fi Signals

Khandaker Foysal Haque, Milin Zhang, and Francesco Restuccia
Institute for the Wireless Internet of Things, Northeastern University, United States
{haque.k, zhang.mil, f.restuccia}@northeastern.edu

Abstract—Recent advances in Wi-Fi sensing have ushered in a plethora of pervasive applications in home surveillance, remote healthcare, road safety, and home entertainment, among others. Most of the existing works are limited to the activity classification of a single human subject at a given time. Conversely, a more realistic scenario is to achieve *simultaneous, multi-subject* activity classification. The first key challenge in that context is that the number of classes grows exponentially with the number of subjects and activities. Moreover, it is known that Wi-Fi sensing systems struggle to adapt to new environments and subjects. To address both issues, we propose *SiMWiSense*, the first framework for simultaneous multi-subject activity classification based on Wi-Fi that generalizes to multiple environments and subjects. We address the scalability issue by using the Channel State Information (CSI) computed from the device positioned closest to the subject. We experimentally prove this intuition by confirming that the best accuracy is experienced when the CSI computed by the transceiver positioned closest to the subject is used for classification. To address the generalization issue, we develop a brand-new few-shot learning algorithm named Feature Reusable Embedding Learning (FREL). Through an extensive data collection campaign in 3 different environments and 3 subjects performing 20 different activities simultaneously, we demonstrate that *SiMWiSense* achieves classification accuracy of up to 97%, while FREL improves the accuracy by 85% in comparison to a traditional Convolutional Neural Network (CNN) and up to 20% when compared to the state-of-the-art few-shot embedding learning (FSEL), by using only 15 seconds of additional data for each class. For reproducibility purposes, we share our 1TB dataset and code repository¹ [1].

I. INTRODUCTION

Wi-Fi is one of the most pervasive wireless technologies worldwide – it has been estimated that by 2025, the Wi-Fi economy will reach a value of \$4.9T [2]. Beyond ubiquitous indoor connectivity, Wi-Fi also allows to develop highly-pervasive device-free sensing applications. The latter are based on the intuition that the received Wi-Fi signals – in particular, the Channel State Information (CSI) computed to perform channel estimation and equalization – are affected by changes in the physical environment caused by any entity in between the source and the receiver. Among other applications, Wi-Fi sensing can be used for fine-grained indoor localization [3], activity recognition [4], [5], and health monitoring [6]. For an excellent survey on the topic, we refer the reader to [7].

Most of the relevant existing work – discussed in detail in Section V – focuses on performing classification of a single subject at a given time [8]–[11]. Even though achieving acceptable sensing performance, a significantly more relevant (and

realistic) problem is performing *simultaneous, multi-subject* Wi-Fi sensing. Moreover, it is well known that Wi-Fi sensing is highly-dependent of the considered subject and environment [12]. Although some attempts to address the issue have been made, they consider few activities – less than 5 [13] – or do not consider multi-subject classification [10].

In stark contrast to the existing works, we propose *SiMWiSense*, a completely novel approach for simultaneous multi-subject activity classification through Wi-Fi. Figure 1 shows a high-level overview of our approach. Beyond the generalization issue, the key challenge addressed by *SiMWiSense* is that by defining as n and m respectively the number of subjects and activities, the number of classes to distinguish becomes n^m . For example, 3 subjects and 10 activities correspond to more than 59,000 classes.

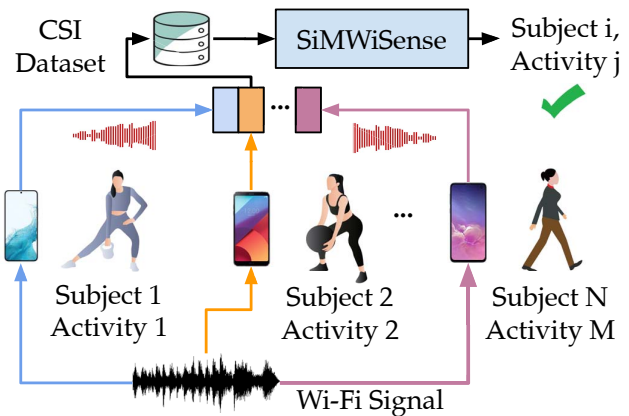


Fig. 1: High-level overview of *SiMWiSense*.

To address this critical issue, we utilize multiple Wi-Fi devices as CSI collectors, where the *closest* to a given subject will classify the activities conducted by that subject. We experimentally prove in Section II that the device closest to the subject will dominantly characterize the channel property between itself and the source of the Wi-Fi signal. Finding the closest device to a given subject falls under the Wi-Fi indoor localization and/or fingerprinting problem [14]–[16], which has been thoroughly investigated and thus considered out of our scope. Although assigning a device to a subject addresses the scalability issue – the classifier output becomes m sized – the overall performance may significantly degrade with new untrained environments and subjects. Thus, we developed a novel Few-Shot Learning (FSL) architecture which can adapt to any new environment, change

¹<https://github.com/kfoysalhaque/SiMWiSense>

in environment or any new subject with up to 15 seconds of new data for each class.

Summary of Novel Contributions

- We present *SiMWiSense*, the first framework for multi-subject simultaneous activity classification using Wi-Fi (Section III). Unlike existing approaches, *SiMWiSense* can distinctly classify among different human subjects performing multiple activities simultaneously by utilizing multiple CSI collectors, each associated to a given subject;

- To address the challenge of generalizing to new environments and subjects, we propose a novel FSL-based architecture called Feature Reusable Embedding Learning (FREL). In stark contrast to existing approaches, FREL can adapt to any new environment and subject through two main steps, namely meta-learning and fine-tuning. Moreover, in contrast to the traditional FSL, FREL combines both the embedding learning and meta-learning approaches to achieve better performance through fine-tuning the classifier with only a few additional samples (Section III);

- We extensively evaluate *SiMWiSense* through an exhaustive data collection campaign in 3 different environments and with 3 subjects performing 20 different activities simultaneously. We demonstrate that *SiMWiSense* achieves classification accuracy of up to 98%, while FREL improves the accuracy by 85% in comparison to a traditional convolutional neural network (CNN) and up to 30% when compared to the state of the art few-shot embedding learning (FSEL) [17], by only using 15 seconds of additional data for each class. **For reproducibility, we share our whole dataset, captured video streams of the activities as ground truth, and our code repository [1].**

II. SENSING PROXIMITY TEST

Wi-Fi sensing leverages tiny changes in the CSI computed through pilot symbols included in the physical layer (PHY) preamble. Although the CSI may be captured by monitoring a transmission link between Access Point (AP) and the stations (STAs) without any direct communication with the AP, a monitoring device captures the CSI of the propagation channel between itself and the AP. Thus, when the CSI monitors are spatially distant enough, they would monitor the independent propagation path between the corresponding antenna pair of the AP and itself [18]. Our key intuition is that the captured CSI is dominantly characterized by any physical change in the environment at spatially closer proximity.

To evaluate this, we perform the following extensive preliminary tests which demonstrate the viability of our key concept. We have performed the sensing proximity test in 3 different environments with 3 different subjects and 20 activities. We assign a CSI monitor to each of the subjects. The monitors are placed at a distance of 1.5m - 3.0m from each other, whereas one human subject performs activity at a distance of 1.5m - 2.0m from each of the sensing monitors for each of the environments. We considered three different environments as explained in Section IV-A. The experimental setup is shown

in Figure 2. We define the subjects with the closest proximity to the CSI Monitor 1, CSI Monitor 2 and CSI Monitor 3 as Subject 1, Subject 2 and Subject 3, respectively. From each environment, CSI is collected in three separate rounds where in every round a subject does 20 different activities and other subjects perform random activities.

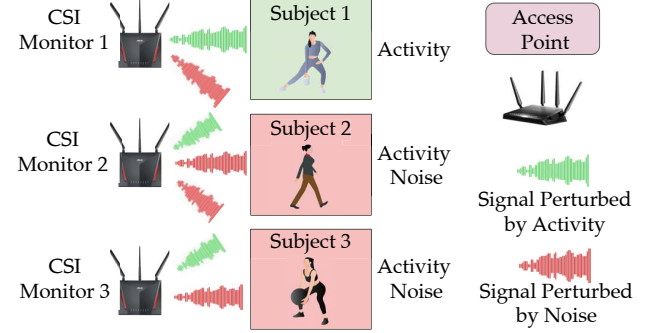


Fig. 2: *SiMWiSense* proximity test.

Figure 3 confirms our intuition. For example, it shows that in the classroom environment, the accuracy of Subject 1 is 95% from the CSI data of Monitor 1 whereas, with the exact same setup and tests, the accuracy of Subject 2 and Subject 3 decreases by 30% on an average with Monitor 1. This is because they are comparatively farther away from Subject 1 and more prone to the noises created by the other subjects at that instant. However, their performances improve drastically to 96% and 97% when we consider CSI Monitor 2 and CSI Monitor 3 for Subject 2 and Subject 3 respectively. The other two environments follow similar trends. **This clearly demonstrates that the CSI monitor closest to the subject performs better than other CSI monitors.**

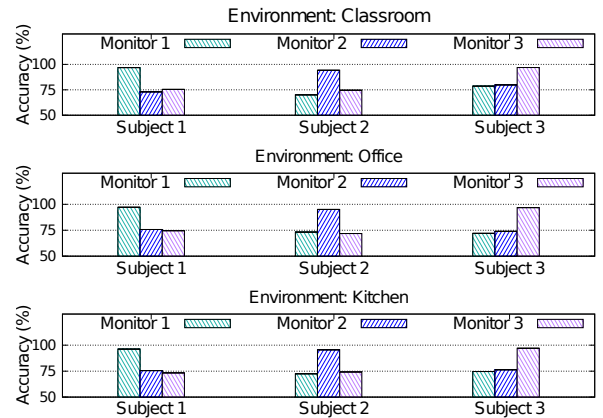


Fig. 3: Classification accuracy as function of sensing proximity.

III. OVERVIEW OF *SIMWISense*

We describe the *SiMWiSense* framework dividing it into three main task blocks: (i) sensing block (ii) preprocessing block and (iii) learning block as presented in Figure 4.

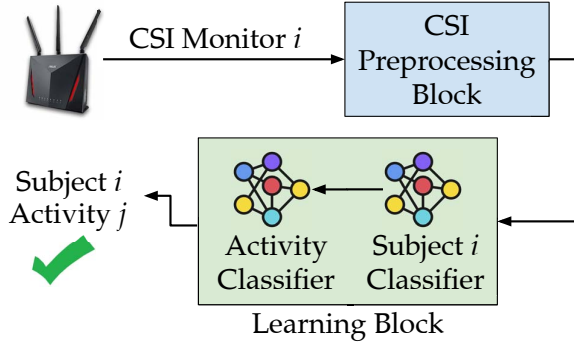


Fig. 4: Overview of SiMWiSense and Processing Blocks.

A. SiMWiSense Sensing and Preprocessing Blocks

The sensing block of SiMWiSense collects the CSI of Wi-Fi transmissions. Modern Wi-Fi systems are based on the Orthogonal Frequency Division Multiplexing (OFDM) modulation which processes multiple data streams in parallel over multiple orthogonal subcarriers. Each spatially diverse CSI monitor captures S samples during the time interval $T = t - t'$ with K orthogonal parallel subcarriers. Thus, the extracted CSI matrix of an $M \times N$ system is shown in Equation 1.

$$H_r^{m,n} = \begin{bmatrix} h_{1,1}^{m,n} & \dots & h_{1,s}^{m,n} & \dots & h_{1,K}^{m,n} \\ \vdots & & \vdots & & \vdots \\ h_{s,1}^{m,n} & \dots & h_{s,s}^{m,n} & \dots & h_{s,K}^{m,n} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ h_{S,1}^{m,n} & \dots & h_{S,s}^{m,n} & \dots & h_{S,K}^{m,n} \end{bmatrix} \quad (1)$$

Here, $H_r^{m,n}$ denotes the CSI matrix at receiver r of the transmit antenna m and receive antenna n where $1 \leq n \leq N$ and $1 \leq m \leq M$. The value $h_{k,s}^{m,n}$ denotes the CSI of the S -th sample at K -th subcarrier from transmit antenna m to the receive antenna n . For example, during the time interval $T = 0.2s$, if any CSI monitor captures $S = 600$ samples with a channel of 80 MHz bandwidth, $H_r^{m,n}$ will have $S \times K$ components where $K = 242$. It is worthwhile mentioning that even though the total number of subcarriers in 80 MHz channel is 256, we only consider data-transmitting subcarriers discarding the null and the guard ones.

After the collection of the CSI samples, we preprocess the captured data, as presented in Figure 5, before it is fed to the learning block. After S samples are collected, we align the data by discarding the missing and/or corrupted CSI measurements. Moreover, any abrupt amplification in the data is removed by normalizing with the mean CSI amplitude. Then, the captures are segmented with a fixed size non-overlapping window along the time domain. If the total number of samples captured during any time interval T is S , such that $T = T_1 + T_2 + T_3 + \dots + T_n$ where T is divided into n equal time windows, $S = S_1 + S_2 + S_3 + \dots + S_n$ are the corresponding sample captures of the n time segments. Thus, each window has the tensor dimension of $S_p \times K \times N$ where S_p is the number of samples in p -th time

window T_p , and $N = 2$ is the complex CSI measurement. This processed data is then fed to the input of the learning block.

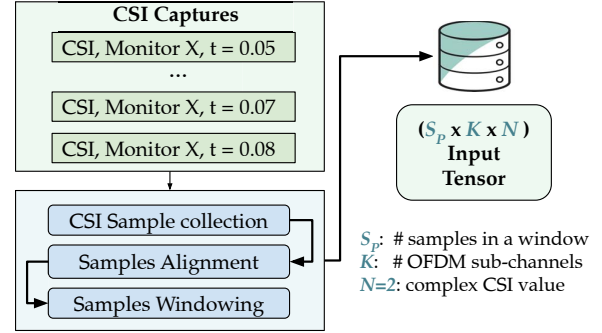


Fig. 5: CSI data processing in SiMWiSense

B. SiMWiSense Learning Block

One of the challenges in multi-subject detection is scalability. For P persons and Q activities, it has Q^P possible combinations, resulting in an exponential increase in the number of classes. One centralized model to classify multi-subject activities becomes difficult when P and Q are large. To tackle this problem, we propose a decentralized detection system for each subject. Specifically, a learning model is assigned to each device to sense the subject which is closest to it. Therefore, each subject only requires Q detection regions in hyperspace. For P subjects, the overall complexity reduces to $P \times Q$. This sensing system has two assumptions: (i) there will be at least the same number of CSI collectors as subjects; (ii) the subject closest to the device take the most significant part in shaping the channel property between the device and AP. Assumption (i) is reasonable since nowadays, almost everyone is inseparable from their smart devices, such as laptops or smartphones, in their work and daily lives. For (ii), we developed an experiment based on sensing proximity as shown in Section II.

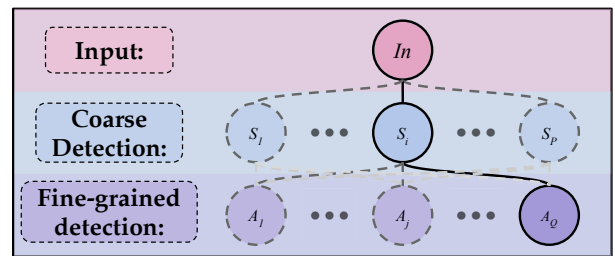


Fig. 6: Cascaded Learning Block.

To further decrease complexity, we propose a cascaded model for two-stage detection. As shown in Figure 6, in the first stage, a deep learning (DL) model is used to discriminate different subjects S coarsely. After the coarse detection, another fine-grained DL model is used to determine the activities A . Regardless of the output at the first stage, all the subjects will share the same fine-grained model at the second stage. Thus, the overall complexity becomes $P + Q$.

One challenging problem of the hierarchical detection model is that even if different persons do the same activities, their movements will have personal patterns and gestures. Furthermore, subjects may join or leave the detection system. Thus, it is impractical to have a universal classifier for activity detection. In addition, the performance of data-driven algorithms in wireless sensing usually will be downgraded by time-varying channel conditions. Thus, we need a model which can swiftly adapt to new subjects and channel features.

C. FREL Learning Algorithm

We propose a novel FSL architecture named FREL, which allows the DL model to adapt to new scenarios with only a few data. We utilize this algorithm in both subject detection and activity detection stages. Next, we will discuss the algorithm in detail. FSL aims at training models that can rapidly generalize to new tasks with only a limited number of labeled samples, is a strong candidate to tackle the data collection problem. One approach to FSL is to learn an embedding for multiple tasks [19], [20]. Specifically, a deep neural network (DNN) is used to learn a clustered mapping from input to the latent space. During the inference time, the embedding network does not need to be fine-tuned, and a few samples will be used as references to classify unobserved data. Another approach is meta-learning [21], [22] which involves two phases: (i) meta-training and (ii) fine-tuning. Meta-learning aims to learn shared features between different tasks during the meta-learning phase and quickly optimize the parameters with a few data points during the fine-tuning stage.

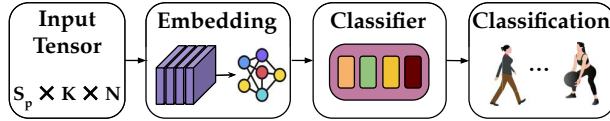


Fig. 7: Feature Reusable Embedding Learning (FREL).

For the first time, FREL combines embedding learning and meta-learning. Figure 7 demonstrates the structure of our FREL model. First, a DNN model is used to learn the embedding of the input and another classifier is used to decode the features in the latent space. Similar to meta-learning, the FREL consists of meta-learning and fine-tuning stages. We train the embedding network and classifier jointly with a mini-dataset during the meta-learning phase, which is the same as embedding learning. After obtaining the embedding, we further optimize the classifier with a few samples while testing. In contrast to embedding learning, we believe fine-tuning can provide better flexibility and granularity, which is more suitable for a dynamic system. Different from meta-learning, we only retrain the simple classifier instead of the whole structure, which can reduce the computation, enabling a faster adaption to new tasks. This design is inspired by [23], which shows that the effectiveness of meta-learning is mainly due to feature reuse. It is a simplified version of MAML [21] that fine-tunes the last few layers which can achieve comparable performance as the original algorithm.

Formally, we consider the embedding network as a function $E_\theta : X \rightarrow Z$, where Z denotes the latent vector. The classifier $C_\phi : Z \rightarrow Y$ is to find a mapping between encoded features Z and labels Y . θ and ϕ are the trainable parameters of the embedding network and classifier, respectively. Hence, The overall system $F_\psi(X) = Y$ can be written as $C_\phi(E_\theta(X)) = Y$, where $\psi = \{\theta, \phi\}$ is the total trainable parameters of the whole system. In FREL, models are trained on a set of mini-batches of data that only have N different classes (ways) and K samples (shots) of each class. Each batch of few-shot data can be considered as a new task $\tau_j = \{(x_i^j, y_i^j)\}_{i=1}^m$ in meta-learning. $m = N \times K$ denotes the total number of samples in one batch. The objective of meta-learning is to find a set of parameters ψ that minimize the expectation of the loss function $\mathcal{L}$ with respect to a group of meta-learning tasks $\mathcal{T} = \{\tau_j\}_{j=1}^n$, i.e.,

$$\min_{\{\theta, \phi\}} \frac{1}{n} \sum_{j=1}^n \left[\frac{1}{m} \sum_{i=1}^m \mathcal{L}(C_\phi(E_\theta(x_i^j)) = y_i^j) \right] \quad (2)$$

We merge the task set $\mathcal{T}$ into a single dataset $\mathbb{D}^{train}$ to get a better embedding, which is given by

$$\begin{aligned} \mathbb{D}^{train} &= \tau_1 \cup \dots \cup \tau_j \cup \dots \cup \tau_n \\ &= \{(x_i^1, y_i^1)\}_{i=1}^m \cup \dots \cup \{(x_i^n, y_i^n)\}_{i=1}^m \end{aligned} \quad (3)$$

We notice that by merging multiple tasks into single dataset, the optimization problem in Equation 2 can be reduced to a general DL problem, which can be solved by a gradient decent optimizer iteratively,

$$\{\theta, \phi\} = \{\theta, \phi\} - \alpha \frac{1}{mn} \sum_{i=1}^{mn} \nabla_{\{\theta, \phi\}} \mathcal{L}(C_\phi(E_\theta(x_i)), y_i) \quad (4)$$

where mn is the total number of data points in $\mathbb{D}^{train}$ and α is the learning rate. Once the optimal embedding θ^* is obtained, the classifier is fine-tuned on another small portion of data $\mathbb{D}^{tune}$. Unlike training on a combined set during the meta-learning, each iteration we randomly sample K shots from each of N ways in $\mathbb{D}^{tune}$ to build a new task τ and update classifier by gradient decent,

$$\phi = \phi - \beta \frac{1}{m} \sum_{i=1}^m \nabla_\phi \mathcal{L}(C_\phi(E_{\theta^*}(x_i)), y_i) \quad (5)$$

where β denotes the learning rate in the fine-tuning phase. Finally, performance is evaluated on the rest of unseen dataset $\mathbb{D}^{test}$.

We summarize FREL in Algorithm 1. It is worth pointing out that although FREL is proposed for WiFi sensing initially, the architecture is presented generally since it can be used for other FSL purposes. Next we discuss the specific setup that is used in the SiMWiSense learning block.

• **Embedding network:** Figure 8 shows the DNN architecture we use for embedding learning in FREL. It is composed of 4 convolutional layers followed by batch normalization and Rectified Linear Unit (ReLU) activation. Each convolutional

layer comprises 64 channels with a kernel size of 3×3 . After the first three convolutional layers, 2×2 Max pooling layers are used to down sample the previous layer's output. After the fourth convolutional layer, a global average pooling strategy is chosen to extract the feature to the latent space, resulting in a 64-dimensional feature space.

Algorithm 1: Feature Reusable Embedding Learning

Phase 1: FREL meta-learning

Require: learning rate α , dataset $\mathbb{D}^{train}$

Initialize: θ for embedding, ϕ for classifier

for $iteration = 1, 2, \dots$ **do**
 update θ and ϕ with $\mathbb{D}^{train}$ by Equation 4

end

Return: θ^* for embedding

Phase 2: FREL fine-tuning

Require: learning rate β , dataset $\mathbb{D}^{tune}$

Initialize: ϕ for classifier

for $epoch = 1, 2, \dots$ **do**
 for $episode = 1, 2, \dots$ **do**
 sample a task $\tau = \{(x_i, y_i)_{i=1}^m\}$ from $\mathbb{D}^{tune}$
 update ϕ with τ by Equation 5
 end

end

Return: ϕ^* for classifier

- **Classifier:** A fully-connected layer is used on top of the pre-trained embedding network as a linear decoder. Non-linearity functions such as ReLU are not applied since we aim to study the efficacy of the overall FREL's design rather than develop complicated DNN models. To investigate the effectiveness of fine-tuning in FREL, we also implement an untrainable K-Nearest Neighbor (K-NN) algorithm as a comparison after the meta-learning phase following the same procedure as another state-of-the-art FSEL model [17]. During the inference time of K-NN, K samples from each class are transformed into embedding as supports, and the queries are classified by a plurality vote of the K nearest supports.

- **Mini-dataset:** Usually in general FSL, the dataset such as Omniglot [24] and Mini-ImageNet [19] contains a large number of tasks and a few number of samples in each task. Algorithms are first pre-trained on multiple tasks and then fine-tuned and tested on single specific task. However, it is never feasible to get a dataset with comprehensive tasks in wireless sensing since the changing environment can always generate new tasks that models have never seen before. Thus, one significant difference in our implementation is the mini-dataset. We only utilize limited number of data collected in 15 seconds for $\mathbb{D}^{tune}$. The test set $\mathbb{D}^{test}$ is never exposed to the model during pre-training and fine-tuning. The mini-dataset makes the problem more challenging as models are learned from not only a few samples but a few tasks.

- **Learning strategy:** We evaluate our model with 5-shot learning, which means we have 5 samples for each class in

every mini-batch data. We choose Adam as the optimizer in both phases. The learning rate α and β are 0.01. Cross-entropy loss is used during the pre-training and fine-tuning stages for simplicity. Other metrics such as deep k-means [25] and prototypical loss [20] can be applied for different purpose.

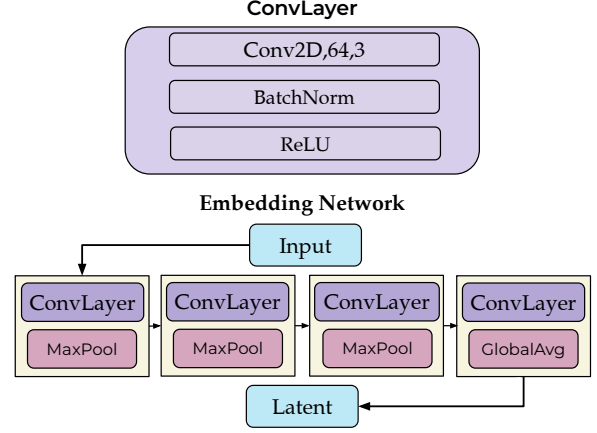


Fig. 8: CNN Embedding Network.

IV. EXPERIMENTAL EVALUATION

A. Experimental Setup

We evaluate the multi-subject sensing capability of SiMWiSense as well as the generalizing feature of FREL through an extensive data collection campaign in 3 different environments: classroom, office and kitchen with 3 human subjects performing 20 different activities in random order. We used off-the-shelf Netgear Nighthawk X4S AC2600 routers to set up the network, whereas IEEE 802.11ac compliant Asus RT-AC86U routers with Nexmon tool are used as the CSI extractor [26].

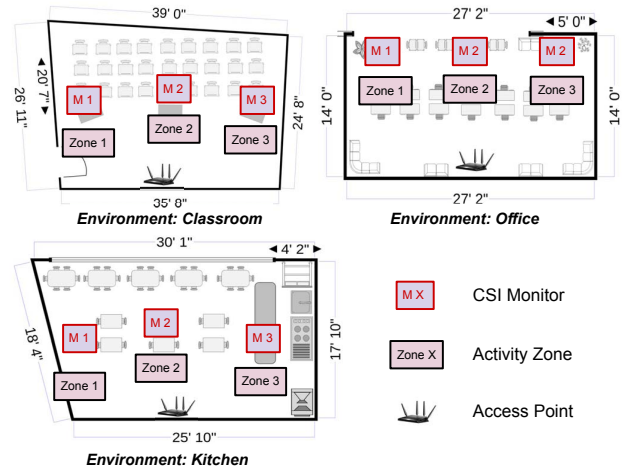


Fig. 9: Experimental Setup of SiMWiSense.

The AP and the STAs are configured with $M = 3$ antennas and $N_{ss} = 3$ spatial streams, while the CSI monitors are

configured with $M = 1$ antenna and $N_{ss} = 1$ spatial streams respectively. We send UDP packets from AP to the STAs to trigger the CSI monitors. The CSI has been collected at center frequency $f_c = 5.21GHz$ (i.e., channel 42) with signals having 80 MHz bandwidth. Three CSI monitors were placed in each of the environments at a distance of 1.5m - 2.0m from each other as shown in Figure 9. Three different subjects performed all the 20 activities: *push forward, rotate, hands up and down, waive, brush, clap, sit, eat, drink, kick, bend forward, wash hands, call, browsing phone, check wrist, read, waive while sitting, writing, side bend, and standing* at a distance of 1.5m - 2.0m from the CSI monitors.

Three separate models are trained for each of the monitors with the data of each corresponding subject performing the activities while the other subjects do random activities. For example, the model for Monitor 1 is trained with Subject 1 performing each of the 20 activities for 5 minutes while Subject 2 and Subject 3 perform different random activities in random order. Similarly, the data from each of the monitors are collected at the mentioned three different environments and for each subject. To evaluate the simultaneous multi-subject sensing, testing data has been collected while different subjects perform randomly chosen activities simultaneously. The experimental setup and the activity zone of the subjects at the mentioned three experimental sites are presented in Figure 10. To create the ground truth, we captured the video streams of the subjects performing different activities in synchronization with the data collection.

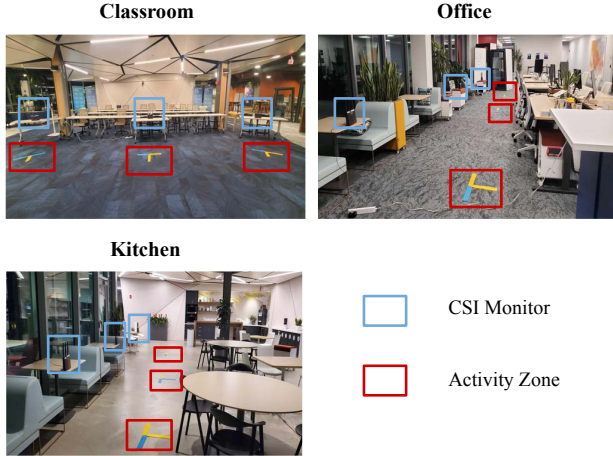


Fig. 10: Data collection locations.

B. Performance Evaluation of SiMWiSense

A time window size of 0.1s is considered for data segmentation, while each data window has 50 samples for all the tests performed in IV-B. Firstly, we do the performance evaluation of our two-stage detection system: (i) subject identification (ii) activity classification with baseline CNN as shown in Figure 8. Then, we demonstrate the generalization capability of FREL in both of the stages of the detection system and compare the

performance of FREL with both baseline CNN and state-of-the-art FSEL [17] model. In the rest of IV-B, Monitor 1, Monitor 2 and Monitor 3 are denoted as M1, M2, and M3 respectively whereas Subject 1, Subject 2 and Subject 3 are presented as Sub1, Sub2 and Sub3 respectively.

Baseline CNN: In the first step of the two-stage SiMWiSense detection, each CSI monitor classify Sub1, Sub2, Sub3 or "no activity". The subject identification performances of each of the monitors in three environments are presented in Figure 11.

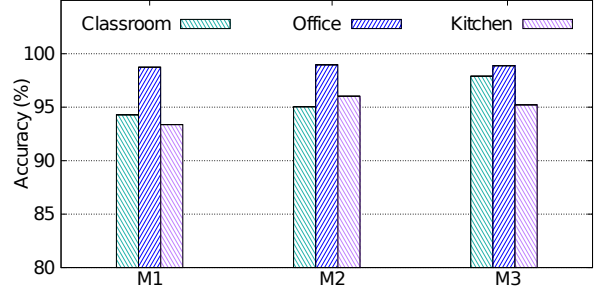


Fig. 11: Subject identification in SiMWiSense with baseline CNN

The results show that with the baseline CNN, the average accuracy of M1, M2 and M3 across all three environments are 95.47%, 96.68% and 97.34% respectively, which shows no significant performance discrepancy. On the other hand, the average performances in three different environments are 95.73%, 98.77% and 94.87% respectively which shows an average of 3.51% performance boost in office compared to the other environments. It is caused by the Non-line-of-sight propagation between the monitors and the distant subjects due to the presence of desks and computers, causing less noise in identifying the closest subject.

The simultaneous multi-subject activity classification performance of SiMWiSense with baseline CNN is presented in Figure 12. The average accuracy in the environments: office, classroom and kitchen are 98.51%, 97.37% and 97.49% respectively which follows the similar trend in performance depicting the stability and robustness of SiMWiSense. Moreover, the performance discrepancy of monitors M1, M2 and M3 are less than 2% achieving an average accuracy of 98.0%, 96.84% and 98.54% respectively with M1, M2 and M3.

Performance as a function of subcarrier resolution: It is known that Wi-Fi sensing performs worse with lower subcarrier resolutions [6], [27]. To compensate the lower subcarrier resolution, one can adopt extensive feature extraction techniques or higher sampling frequency which would increase the computation burden by intensifying the pre-processing steps and learning process dramatically. This stimulates us to study the trade-off of the subcarrier resolution and the SiMWiSense performance.

Figure 13 shows the performance of SiMWiSense as a function of the number of subcarriers. The first consecutive 20,

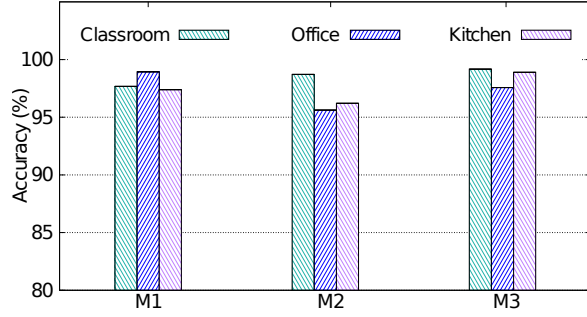


Fig. 12: Performance of SiMWiSense at three different environments with baseline CNN as presented in Figure 8

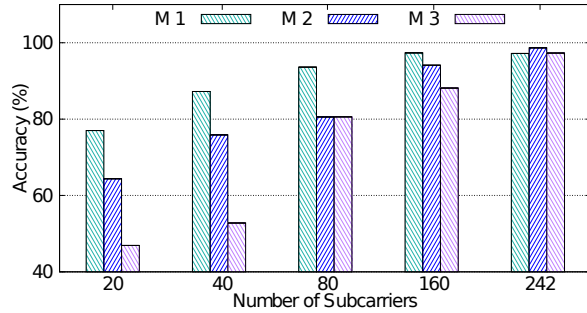
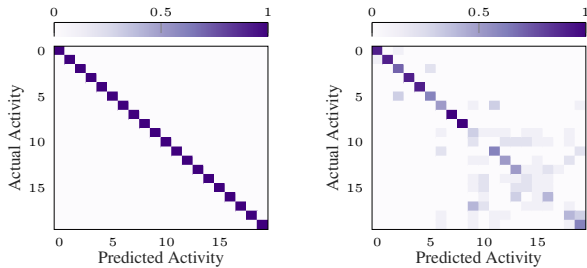


Fig. 13: Performance of SiMWiSense with baseline CNN as a function of number of subcarriers (environment: classroom).

40, 80 and 160 and 242 subcarriers are considered to emulate a sensing system with a lower bandwidth - thus, less number of consecutive subcarriers. The results show that the average performance of monitors decrease to 84.92% when the number of subcarriers decrease from 242 to 80, and it goes down to 62.76% with a percent decrease of 34.94 when we switch to only 20 subcarriers. Figure 14 presents the confusion matrices of baseline CNN when trained with 242 and 20 subcarriers respectively where they achieved an accuracy of 98.66% and 64.37% respectively. It is evident that, with only 20 subcarriers, the model gets confused with few activities whereas it performs



(a) 242 sub-carriers, accuracy: 98.66% (b) 20 sub-carriers, accuracy: 64.37%

Fig. 14: Confusion matrices of baseline CNN with 242 and 20 subcarriers respectively (in classroom, with monitor M2).

comparatively better in other activities. The top three classes which are hardest to distinguish at lower subcarrier resolutions are: wash hands (index 17), rotate (index 10), and brush (index 15). However, it is noticeable that when we switch to 20 subcarriers from 242, the input tensor dimension reduces by 12 times from $50 \times 242 \times 2 = 24200$ to $50 \times 20 \times 2 = 2000$ and still achieve around 64% accuracy on an average in the considered scenarios.

SiMWiSense performance with FREL: Even though the performance of the traditional CNN is quite good, they fail to generalize the environments or subjects. In such instances for generalizing to new environments and subjects, FREL excels in comparison to the traditional CNN. The performance of the FREL in new untrained environment are presented and compared with traditional CNN and FSEL [17] in Figure 15 and Figure 16 respectively. The results show that the overall performance of SiMWiSense with FREL in new untrained environments improves by 86.06%, 87.56% and 86.90% when the embedding network is trained in classroom, office and kitchen respectively in comparison to the baseline CNN. The highest performance achieved by FREL in any new untrained environment is 97.24% in the kitchen with monitor M1 whereas the lowest accuracy is 89.33% in the kitchen with monitor M2. Thus, it demonstrates the robustness, reliability and strong adaptive capability of FREL to new environments. As shown in Figure 16, FREL surpasses the FSEL by 25.71%, 17.43%, and 15.56% respectively when the FREL is trained in the classroom, office and kitchen and tested on other corresponding environments.

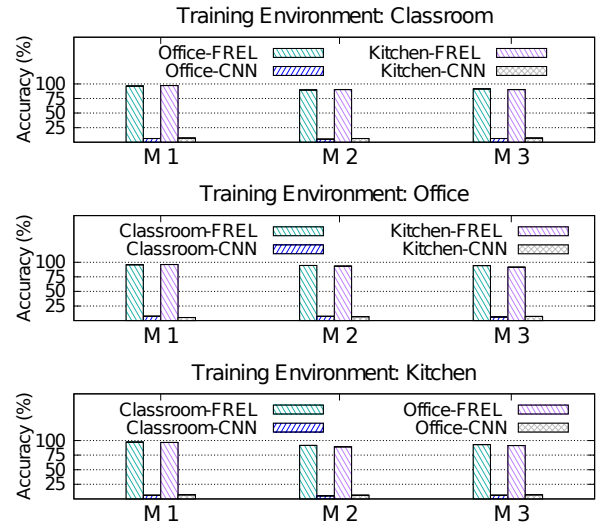


Fig. 15: Performance of FREL in simultaneous activity sensing with new untrained environments.

Figure 17 presents the confusion matrices of FREL and FSEL when trained with monitor M1 in the classroom and tested with monitor M1 in the office. It is evident from Figure 17b that the accuracy drop of FSEL is caused by a few activities which it finds difficult to distinguish. The top three activities

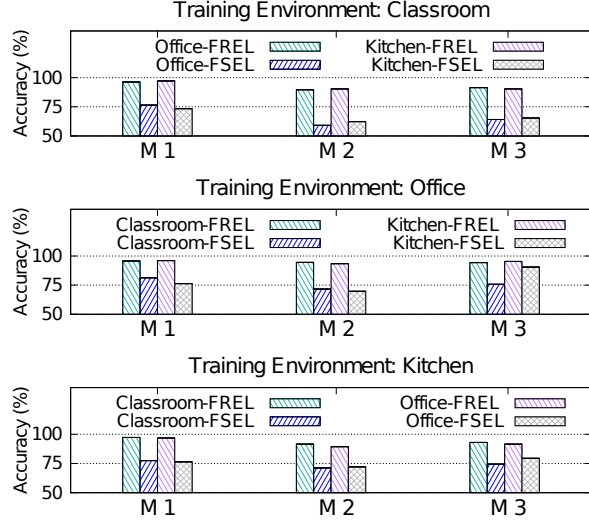


Fig. 16: Performance comparison of FREL with FSEL in new untrained environments.

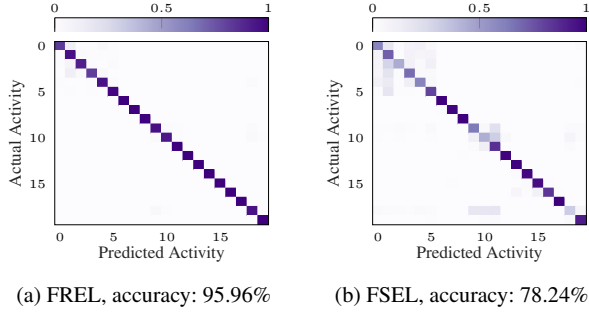


Fig. 17: Confusion matrices for FREL and FSEL when trained in classroom, with monitor M1 and tested in office, with monitor M1.

which FSEL finds most difficult to distinguish are waiving while sitting (index 18), rotating (index 10) and eating (index 3). On the other hand, the top three distinct activities for FSEL are drinking (index 6), waiving (index 8) and phone call (index 13).

FREL in generalizing subject identification across environments: FREL can also generalize the subject identification in new untrained environments, as presented in Figure 18. In fact, it can achieve up to an accuracy of 96.53% whereas the traditional CNN only limits to 6.19% compared to the 93.53% for FREL on an average across the monitors and the environments.

As depicted in Figure 19, when compared to FSEL, proposed FREL demonstrates an accuracy boost of 17.79%, 17.93%, and 17.55% in classroom, office and kitchen, respectively. Thus, FREL signifies stable and reliable performances across the environments in generalizing the 'subject identification' phase of the learning also.

FREL in generalizing the monitors across the environment: It would be interesting to see how a FREL model trained

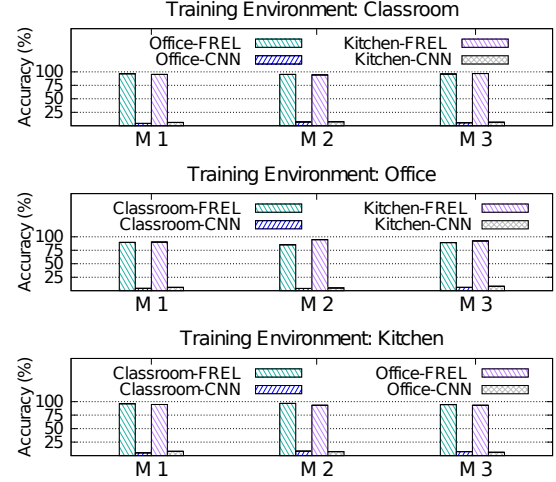


Fig. 18: Performance of FREL as the subject identifier in untrained environments

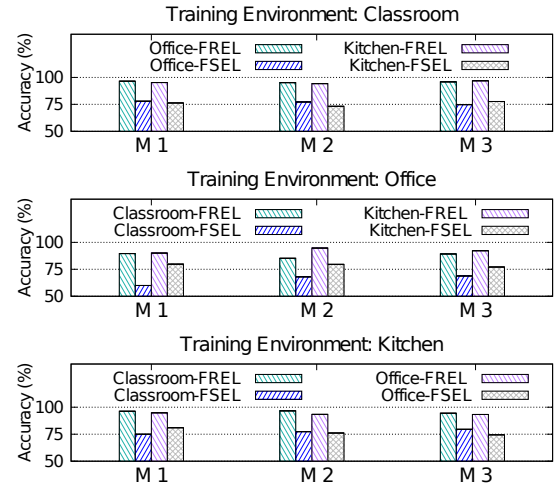


Fig. 19: Performance comparison of FREL and FSEL as the subject identifier in untrained environments

in one monitor can adapt and generalize to other monitors, thus generalizing new subjects of the same environment. It would further lessen the training time and drastically reduce the system deployment complexity. Figure 20 presents the performance of FREL with new untrained monitors across the three different environments. It can achieve an average accuracy of 91.71%, 94.51% and 94.78% in generalizing the other monitors while trained with Monitor 1, Monitor 2 and Monitor 3, respectively. Thus FREL enables any system to be trained only on one monitor and deployed with n number of monitors with only 15s new data samples from each monitor.

V. RELATED WORK

A significant amount of research in Wi-Fi sensing has been performed over the last few years. The reader may refer to the following surveys for a good compendium of the state of

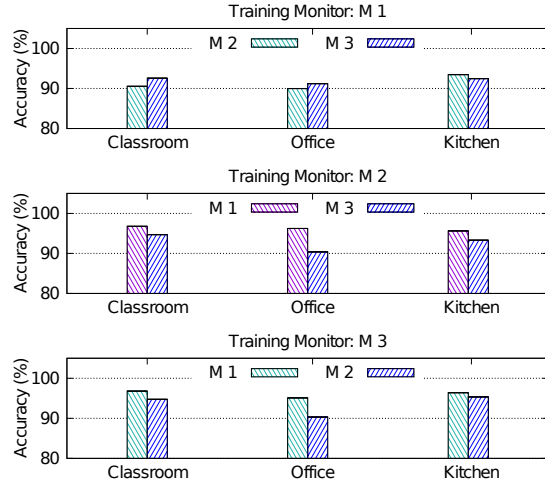


Fig. 20: Performance of FREL with new untrained monitors.

the art [7], [28], [29]. There have been several approaches to address the challenges of Wi-Fi sensing, which includes received signal strength indicator (RSSI) based approaches [30], [31], and passive Wi-Fi radar (PWR) [32], [33]. Wi-Fi sensing leveraging beamforming feedback information (BFI) which can be captured from any Wi-Fi network is the state-of-the-art approach [34]. However, CSI-based Wi-Fi sensing is by far the most popular approach.

DL has already been proven to be effective in various CSI-based Wi-Fi sensing applications, including human activity classification [35]–[37], gesture recognition [38], [39], health-monitoring [40]–[42], human counting [43], [44], indoor localization [15]. To explore other DL-based CSI sensing applications, we refer the readers to [7], [45]. A few of the interesting works on human activity classification are briefly discussed below. Shalaby et al. [35] proposed four different DL models, namely CNN with Gated Recurrent Unit (GRU), glscnn with GRU and attention, CNN with a GRU and a second CNN, and CNN with Long Short-Term Memory (LSTM) and a second CNN, achieved accuracy up to 99.46%. However, these models only consider a single subject performing six activities and can not generalize to new untrained environments. MCBCAR by Wang et al. [46] used generative adversarial network (GAN) and semi-supervised learning to address the challenge of the non-uniformly distributed data due to environmental dynamics. Even though this work considers the dynamic change in the environment, the framework is not designed to adapt to new untrained environments and simultaneous multi-subject sensing. AFSL-HAR framework by Wang et al. [47] achieves a performance gain in recognizing the new activities with a few samples of new data through few-shot learning.

Even though this work addressed the challenges of new scenario, activity or subject by fast adaptation with few new samples, the framework classify only one subject at a time in any environment. Ding et al. proposed WiLISensing [48] to address the challenge of variations in activity locations in the same environment with few new data samples through

Protonet. However, it is not evident how WiLISensing would adapt to a completely new scenario or new subject. Ding et al. proposed RF-Net [49], a meta-learning framework to adapt to new environments with few labeled data samples. However, the RF-Net’s CSI-based sensing performance ranges in (70-80)%. In contrast to the DL-based systems, Abdelnasser et al. [8] designed WiGest for gesture classification based on mutually-independent distinguishable application actions which does not need any training. **To the best of our knowledge, we are the first to propose a framework for simultaneous multi-subject activity recognition with Wi-Fi sensing.**

VI. CONCLUDING REMARKS

In this paper, we have proposed SiMWiSense, the first framework for simultaneous multi-subject sensing based on Wi-Fi CSI sensing. In contrast to the existing approaches, SiMWiSense can classify the activities of multiple subjects in the same environment independently and simultaneously. An FSL-based algorithm: FREL is also proposed for fast adaptation to changing data distributions, making the system robust for the dynamic environment, which can generalize new environments, subjects with only a few new samples of data achieving an accuracy up to 98.94%. We have evaluated the efficacy of the overall design using extensive data collected in three different scenarios: classroom, office and kitchen, with three subjects performing 20 activities simultaneously. We demonstrate that SiMWiSense surpasses the traditional CNN-based approach and a state-of-the-art FSL model by 85% and 20% on average.

VII. ACKNOWLEDGEMENTS

This material is based upon work supported in part by the National Science Foundation (NSF) under Grant No. CNS-2134973 and CNS-2120447. The views and opinions expressed in this work are those of the authors and do not necessarily reflect those of the NSF.

REFERENCES

- [1] K. F. Haque, M. Zhang, and F. Restuccia, “SiMWiSense Dataset and Code,” <https://github.com/kfoysalhaque/SiMWiSense>, 2023.
- [2] Wi-Fi Alliance, “The Economic Value of Wi-Fi: A Global View (2018 and 2023),” <https://tinyurl.com/EconWiFi>, 2021.
- [3] M. Abbas, M. Elhamshary, H. Rizk, M. Torki, and M. Youssef, “WiDeep: WiFi-based Accurate and Robust Indoor Localization System using Deep Learning,” in *2019 IEEE International Conference on Pervasive Computing and Communications (PerCom)*, 2019, pp. 1–10.
- [4] R. Presotto, G. Civitarese, and C. Bettini, “FedCLAR: Federated Clustering for Personalized Sensor-Based Human Activity Recognition,” in *2022 IEEE International Conference on Pervasive Computing and Communications (PerCom)*, 2022, pp. 227–236.
- [5] B. Khaerdinov, E. Ghaleb, and S. Asteriadis, “Deep Triplet Networks with Attention for Sensor-based Human Activity Recognition,” in *2021 IEEE International Conference on Pervasive Computing and Communications (PerCom)*, 2021, pp. 1–10.
- [6] Y. Zeng, D. Wu, J. Xiong, J. Liu, Z. Liu, and D. Zhang, “MultiSense: Enabling Multi-person Respiration Sensing with Commodity Wifi,” *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 4, no. 3, pp. 1–29, 2020.
- [7] Y. Ma, G. Zhou, and S. Wang, “WiFi Sensing with Channel State Information: A Survey,” *ACM Computing Surveys (CSUR)*, vol. 52, no. 3, pp. 1–36, 2019.
- [8] H. Abdelnasser, K. Harras, and M. Youssef, “A Ubiquitous WiFi-based Fine-Grained Gesture Recognition System,” *IEEE Transactions on Mobile Computing*, vol. 18, no. 11, pp. 2474–2487, 2018.

- [9] K. Qian, C. Wu, Z. Yang, Y. Liu, and K. Jamieson, "Widar: Decimeter-level Passive Tracking via Velocity Monitoring with Commodity Wi-Fi," in *Proc. of ACM MobiCom*, 2017.
- [10] F. Meneghello, D. Garlisi, N. Dal Fabbro, I. Tinnirello, and M. Rossi, "SHARP: Environment and Person Independent Activity Recognition with Commodity IEEE 802.11 Access Points," *IEEE Transactions on Mobile Computing*, pp. 1–16, 2022.
- [11] X. Wang, C. Yang, and S. Mao, "PhaseBeat: Exploiting CSI Phase Data for Vital Sign Monitoring with Commodity WiFi Devices," in *Proceedings of IEEE ICDCS*, 2017.
- [12] Y. Ma, S. Arshad, S. Muniraju, E. Torkildson, E. Rantala, K. Doppler, and G. Zhou, "Location- and Person-Independent Activity Recognition with WiFi, Deep Neural Networks, and Reinforcement Learning," *ACM Transactions on the Internet of Things*, vol. 2, no. 1, Jan. 2021. [Online]. Available: <https://doi-org.ezproxy.neu.edu/10.1145/3424739>
- [13] N. Bahadori, J. Ashdown, and F. Restuccia, "ReWiS: Reliable Wi-Fi Sensing Through Few-Shot Multi-Antenna Multi-Receiver CSI Learning," in *2022 IEEE 23rd International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM)*, 2022, pp. 50–59.
- [14] A. Shokry, M. Elhamshary, and M. Youssef, "The Tale of Two Localization Technologies: Enabling Accurate Low-Overhead WiFi-based Localization for Low-End Phones," in *Proceedings of the ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, 2017, pp. 1–10.
- [15] M. Abbas, M. Elhamshary, H. Rizk, M. Torki, and M. Youssef, "WiDeep: WiFi-based Accurate and Robust Indoor Localization System using Deep Learning," in *Proceedings of IEEE International Conference on Pervasive Computing and Communications (PerCom)*. IEEE, 2019, pp. 1–10.
- [16] F. Wang, J. Feng, Y. Zhao, X. Zhang, S. Zhang, and J. Han, "Joint Activity Recognition and Indoor Localization with WiFi Fingerprints," *IEEE Access*, vol. 7, pp. 80 058–80 068, 2019.
- [17] Y. Tian, Y. Wang, D. Krishnan, J. B. Tenenbaum, and P. Isola, "Rethinking Few-Shot Image Classification: A Good Embedding is All You Need?" in *Proceedings of European Conference on Computer Vision (ECCV)*. Springer, 2020, pp. 266–282.
- [18] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge university press, 2005.
- [19] O. Vinyals, C. Blundell, T. Lillicrap, D. Wierstra *et al.*, "Matching Networks for One Shot Learning," *Advances in neural information processing systems*, vol. 29, 2016.
- [20] J. Snell, K. Swersky, and R. Zemel, "Prototypical Networks for Few-shot Learning," *Advances in neural information processing systems*, vol. 30, 2017.
- [21] C. Finn, P. Abbeel, and S. Levine, "Model-agnostic Meta-learning for Fast Adaptation of Deep Networks," in *International conference on machine learning*. PMLR, 2017, pp. 1126–1135.
- [22] A. Nichol, J. Achiam, and J. Schulman, "On First-order Meta-learning Algorithms," *arXiv preprint arXiv:1803.02999*, 2018.
- [23] A. Raghu, M. Raghu, S. Bengio, and O. Vinyals, "Rapid Learning or Feature Reuse? Towards Understanding the Effectiveness of MAML," *arXiv preprint arXiv:1909.09157*, 2019.
- [24] B. Lake, R. Salakhutdinov, J. Gross, and J. Tenenbaum, "One Shot Learning of Simple Visual Concepts," in *Proceedings of the annual meeting of the cognitive science society*, vol. 33, no. 33, 2011.
- [25] M. M. Fard, T. Thonet, and E. Gaussier, "Deep k-means: Jointly clustering with k-means and learning representations," *Pattern Recognition Letters*, vol. 138, pp. 185–192, 2020.
- [26] F. Gringoli, M. Schulz, J. Link, and M. Hollick, "Free Your CSI: A Channel State Information Extraction Platform For Modern Wi-Fi Chipsets," in *Proceedings of the 13th International Workshop on Wireless Network Testbeds, Experimental Evaluation & Characterization*. New York, NY, USA: Association for Computing Machinery, 2019, p. 21–28.
- [27] S. Shi, Y. Xie, M. Li, A. X. Liu, and J. Zhao, "Synthesizing Wider WiFi Bandwidth for Respiration Rate Monitoring in Dynamic Environments," in *IEEE INFOCOM 2019-IEEE Conference on Computer Communications*. IEEE, 2019, pp. 181–189.
- [28] S. Tan, Y. Ren, J. Yang, and Y. Chen, "Commodity WiFi Sensing in 10 Years: Status, Challenges, and Opportunities," *IEEE Internet of Things Journal*, 2022.
- [29] S. M. Hernandez and E. Bulut, "WiFi Sensing on the Edge: Signal Processing Techniques and Challenges for Real-World Systems," *IEEE Communications Surveys & Tutorials*, 2022.
- [30] P. Ssekidde, O. Steven Eyobu, D. S. Han, and T. J. Oyana, "Augmented CWT Features for Deep Learning-based Indoor Localization using WiFi RSSI Data," *Applied Sciences*, vol. 11, no. 4, p. 1806, 2021.
- [31] A. Dubey, P. Sood, J. Santos, D. Ma, C.-Y. Chiu, and R. Murch, "An Enhanced Approach to Imaging the Indoor Environment using WiFi RSSI Measurements," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 9, pp. 8415–8430, 2021.
- [32] H. C. Yildirim, L. Storrer, P. De Doncker, J. Louveaux, and F. Horlin, "A Multi-antenna Super-resolution Passive Wi-Fi Radar Algorithm: Combined Model Order Selection and Parameter Estimation," *IET Radar, Sonar & Navigation*, vol. 16, no. 8, pp. 1376–1387, 2022.
- [33] W. Li, R. J. Piechocki, K. Woodbridge, C. Tang, and K. Chetty, "Passive wifi radar for human sensing using a stand-alone access point," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 3, pp. 1986–1998, 2020.
- [34] K. F. Haque, M. Zhang, F. Meneghello, and F. Restuccia, "BeamSense: Rethinking Wireless Sensing with MU-MIMO Wi-Fi Beamforming Feedback," *arXiv preprint arXiv:2303.09687*, 2023.
- [35] E. Shalaby, N. ElShennawy, and A. Sarhan, "Utilizing Deep Learning Models in CSI-based Human Activity Recognition," *Neural Computing and Applications*, vol. 34, no. 8, pp. 5993–6010, 2022.
- [36] H. Yan, Y. Zhang, Y. Wang, and K. Xu, "WiAct: A Passive WiFi-based Human Activity Recognition System," *IEEE Sensors Journal*, vol. 20, no. 1, pp. 296–305, 2019.
- [37] H. Ambalkar, X. Wang, and S. Mao, "Adversarial Human Activity Recognition Using Wi-Fi CSI," in *2021 IEEE Canadian Conference on Electrical and Computer Engineering (CCECE)*. IEEE, 2021, pp. 1–5.
- [38] M. T. Islam and S. Nirjon, "Wi-Fringe: Leveraging Text Semantics in WiFi CSI-Based Device-free Named Gesture Recognition," in *2020 16th International Conference on Distributed Computing in Sensor Systems (DCOSS)*. IEEE, 2020, pp. 35–42.
- [39] R. Gao, M. Zhang, J. Zhang, Y. Li, E. Yi, D. Wu, L. Wang, and D. Zhang, "Towards Position-independent Sensing for Gesture Recognition with Wi-Fi," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 5, no. 2, pp. 1–28, 2021.
- [40] A. Pandey, M. Zeeshan, and S. Kumar, "Csi-based joint location and activity monitoring for covid-19 quarantine environments," *IEEE Sensors Journal*, vol. 23, no. 2, pp. 969–976, 2023.
- [41] A. Kumar, S. Singh, V. Rawal, S. Garg, A. Agrawal, and S. Yadav, "CNN-based Device-free Health Monitoring and Prediction System using WiFi Signals," *International Journal of Information Technology*, pp. 1–13, 2022.
- [42] A. Sharma, J. Li, D. Mishra, G. Batista, and A. Seneviratne, "Passive WiFi CSI Sensing Based Machine Learning Framework for COVID-Safe Occupancy Monitoring," in *2021 IEEE International Conference on Communications Workshops (ICC Workshops)*. IEEE, 2021, pp. 1–6.
- [43] O. T. Ibrahim, W. Goma, and M. Youssef, "CrossCount: A Deep Learning System for Device-free Human Counting using WiFi," *IEEE Sensors Journal*, vol. 19, no. 21, pp. 9921–9928, 2019.
- [44] S. Depatla and Y. Mostofi, "Crowd Counting Through Walls Using WiFi," in *2018 IEEE international conference on pervasive computing and communications (PerCom)*. IEEE, 2018, pp. 1–10.
- [45] J. Yang, X. Chen, D. Wang, H. Zou, C. X. Lu, S. Sun, and L. Xie, "Deep Learning and its Applications to WiFi Human Sensing: A Benchmark and a Tutorial," *arXiv preprint arXiv:2207.07859*, 2022.
- [46] D. Wang, J. Yang, W. Cui, L. Xie, and S. Sun, "Multimodal CSI-based Human Activity Recognition Using GANs," *IEEE Internet of Things Journal*, vol. 8, no. 24, pp. 17 345–17 355, 2021.
- [47] Y. Wang, L. Yao, Y. Wang, and Y. Zhang, "Robust CSI-based Human Activity Recognition with Augment Few Shot Learning," *IEEE Sensors Journal*, vol. 21, no. 21, pp. 24 297–24 308, 2021.
- [48] X. Ding, T. Jiang, Y. Zhong, J. Yang, Y. Huang, and Z. Li, "Device-free Location-independent Human Activity Recognition via Few-shot Learning," in *2021 IEEE/CIC International Conference on Communications in China (ICCC Workshops)*. IEEE, 2021, pp. 106–111.
- [49] S. Ding, Z. Chen, T. Zheng, and J. Luo, "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*, 2020, pp. 517–530.