WiMANS: A Benchmark Dataset for WiFi-based Multi-user Activity Sensing

Shuokang Huang, Kaihan Li, Di You, Yichong Chen
Arvin Lin, Siying Liu, Xiaohui Li, and Julie A. McCann

Imperial College London, London SW7 2AZ, UK
{s.huang21, j.mccann}@imperial.ac.uk
https://github.com/huangshk/WiMANS

Abstract. WiFi-based human sensing has exhibited remarkable potential to analyze user behaviors in a non-intrusive and device-free manner, benefiting applications as diverse as smart homes and healthcare. However, most previous works focus on single-user sensing, which has limited practicability in scenarios involving multiple users. Although recent studies have begun to investigate WiFi-based multi-user sensing, there remains a lack of benchmark datasets to facilitate reproducible and comparable research. To bridge this gap, we present WiMANS, to our knowledge, the first dataset for multi-user sensing based on WiFi. WiMANS contains over 9.4 hours of dual-band WiFi Channel State Information (CSI), as well as synchronized videos, monitoring the simultaneous activities of multiple users. We exploit WiMANS to benchmark the performance of state-of-the-art WiFi-based human sensing models and video-based models, posing new challenges and opportunities for future work. We believe WiMANS can push the boundaries of current studies and catalyze the research on WiFi-based multi-user sensing.

Keywords: Human Sensing · Multi-user · WiFi · Benchmark Dataset

1 Introduction

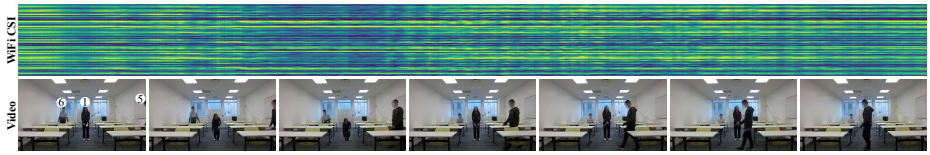
Recent years have witnessed the rapid progress of WiFi-based human sensing [6,25,48], which collects Channel State Information (CSI) from off-the-shelf WiFi devices to recognize human identities [55, 68, 75], locations [9, 12, 45], activities [7,20,33], *etc.* It plays an increasingly important role in differing applications such as security monitoring [36,50,83], smart homes [10,27], and healthcare [16,43,71]. Compared with cameras and on-body sensors, the use of WiFi CSI negates the necessity of filming users or attaching sensors to them [42]. Such non-intrusive approaches can make human sensing widely available [81], satisfying the need to monitor users who do not want to be filmed or wear sensors. WiFi-based human sensing is also robust to low-light or non-line-of-sight conditions, while using cameras for video-based analysis may be sensitive to low-light conditions and obstacles [32]. More importantly, CSI can be gathered from ubiquitous existing WiFi devices, enabling device-free human sensing [28] without the prerequisite of dedicated devices and particular deployments.

Table 1: Comparison of public WiFi-based human sensing datasets. WiMANS is *the first dataset* that involves multiple users performing different/identical activities simultaneously in each sample. "Idt.": Identity. "Loc.": Location. "Act.": Activity.

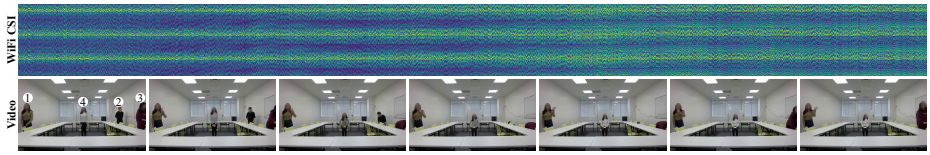
Dataset	# of Users per Sample	# of Samples	# of Activities	# of Channels	Sample Rate (Hz)	WiFi Band (GHz)	Video Data	Annotations		
								Idt.	Loc.	Act.
Yousefi *et al.* [72]	1	557	7	90	1000	5	-	-	-	✓
SignFi [38]	1	14280	276	90	12.5∼200	5	-	-	-	✓
FallDeFi [43]	1	1070	19	90	1000	5	-	-	-	✓
WiAR [21]	1	4161	16	90	30	5	-	✓	-	✓
ARIL [56]	1	1394	6	52	-	-	-	-	✓	✓
Brinke *et al.* [4]	1	3749	6	270	20	2.4	-	✓	-	✓
RF-NET [11]	1	12000	6	60	100	-	-	-	-	✓
Baha *et al.* [2]	1	9000	12	90	320	2.4	-	✓	-	✓
CSIDA [26]	1	2844	6	342	1000	5	-	✓	✓	✓
NTU-Fi [66]	1	2040	6	342	500	5	-	✓	-	✓
CPAR [65]	1	560	7	64	1000	2.4	-	✓	-	✓
Widar [44]	1	54	2	90	2000	5	-	-	✓	✓
Widar 2.0 [45]	1	24	2	90	1000	5	-	-	✓	✓
Widar 3.0 [76]	1	271050	22	90	1000	5	-	✓	✓	✓
Yang *et al.* [70]	1	1050	8	90	30	5	-	✓	✓	✓
Moshiri *et al.* [14]	1	420	7	52	200	5	-	✓	-	✓
OPERAnet [3]	0∼1	6235	6	540	1600	5	-	✓	-	✓
MM-Fi [69]	1	1080	27	342	1000	5	✓	✓	-	✓
SHARPax [39]	0∼1	108	3	242∼996	133	5	-	-	-	✓
WiMANS (Ours)	0∼5	11286	9	270	1000	2.4/5	✓	✓	✓	✓

The principle of WiFi-based human sensing is that human activities essentially interfere with WiFi signals and lead to signal variations [61]. Such variations are recorded in CSI, which thereby contains implicit features for human sensing. Since these human features typically intertwine with excessive noise [60], it is impossible to interpret WiFi CSI patterns for human sensing straightforwardly. Therefore, various models have been proposed to learn features from the CSI for different human sensing tasks. (1) For human identification, extensive literature has discussed the use of Multilayer Perceptrons (MLPs) [66], Long Short-Term Memory (LSTM) [8], Convolutional Neural Networks (CNNs) [55, 67, 68], CNN-LSTM hybrids [32, 40, 75], *etc.* (2) For human localization, prevalent models have been adopted, including Naive Bayes [62], Auto-encoders [15], LSTM [9], and CNNs [57]. (3) WiFi-based human activity recognition (HAR) has drawn the greatest attention from researchers, applying models such as MLPs [73], LSTM [72], CNNs [41, 74, 77], CNN-LSTM hybrids [22, 47, 84], Generative Adversarial Networks (GANs) [54], attention-based bidirectional LSTM (ABLSTM) [7], and Transformers [33, 80].

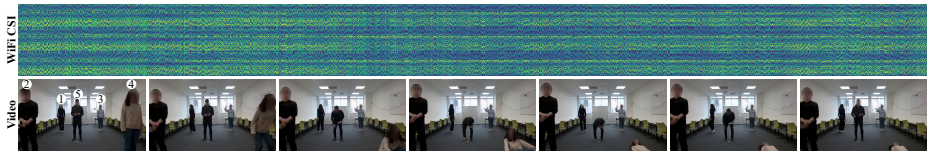
However, several critical drawbacks seriously hamper the practicability of WiFi-based human sensing. (1) *Single-user Limitation*: Existing public datasets only include a single user in each CSI sample, as shown in Table 1. Consequently, most WiFi-based models only recognize the identity/location/activity of a single user in each recognition, but many practical scenarios indeed consist of multiple users simultaneously. Although recent studies [12, 24, 31, 58] have attempted to sense multiple users simultaneously, there remains **a lack of public datasets that enable WiFi-based multi-user sensing**. (2) *Insufficient Modalities and Annotations*: Most datasets collect CSI of a single WiFi band (*i.e.*, 2.4 or 5 GHz) and do not incorporate synchronized videos. Such insufficiency disables the further study of unexplored tasks (*e.g.*, pose estimation). Moreover, many datasets annotate samples for specific tasks (*e.g.*, activity recognition), restricting the use



(a) User 1: Picking Up. User 5: Walking. User 6: Sitting Down.



(b) User 1: Waving. User 2: Lying Down. User 3: Picking Up. User 4: Sitting Down.



(c) User 1: Rotation. User 2: Jumping. User 3: Waving. User 4: Lying Down. User 5: Picking Up.

Fig. 1: Examples of WiFi CSI and videos in WiMANS, monitoring simultaneous activities performed by multiple users in various environments. (a) consists of 3 users in a classroom. (b) contains 4 users in a meeting room. (c) has 5 users in an empty room.

of WiFi CSI in diverse sensing tasks. (3) *Lack of Comprehensive Benchmarks*: Previous works mainly focus on novel models but few of them have provided benchmarks for these models. Specifically, no benchmark is available for multi-user activity sensing based on WiFi CSI.

To help resolve the above drawbacks, we present WiMANS, the first dataset that enables multi-user sensing based on WiFi CSI. WiMANS collects 11286 CSI samples of dual WiFi bands (2.4 / 5 GHz), along with synchronized videos for reference, as shown in Fig. 1. Each 3-second sample includes 0 to 5 users performing identical/different activities simultaneously, annotated with (anonymized) user identities, locations, and activities. Table 1 compares WiMANS with related datasets to highlight our novelty. Extensive experiments have been conducted to benchmark state-of-the-art WiFi-based models and video-based models. The main contributions and unique aspects of WiMANS are as follows:

- We construct a WiFi-based multi-user activity sensing dataset, where each sample monitors simultaneous activities of multiple users. To the best of our knowledge, WiMANS is the first dataset that collects dual-band WiFi CSI and videos for multiple users in each sample.
- WiMANS provides fine-grained annotations of user identities, locations, and activities to support various sensing tasks. The videos in WiMANS can further act as a reference for unexplored tasks (*e.g.*, multi-user pose estimation).
- Benchmark experiments have been conducted to analyze the multi-user sensing performance of WiFi-based models and video-based models. This work provides the first benchmarks for WiFi-based multi-user identification, localization, and activity recognition.

2 Related Work

2.1 WiFi-based Human Sensing

Human sensing with WiFi CSI is an increasingly promising alternative to traditional sensing technologies thanks to its non-intrusiveness, environmental robustness, and device-free merits. In WiFi-based human sensing, much effort has been devoted to three underlying yet distinct tasks: (1) human identification, (2) human localization, and (3) human activity recognition (HAR).

WiFi-based human identification serves to determine user identities by learning biometric patterns from CSI. To map CSI to identities, MLPs [66] simply feed all CSI values into fully connected layers, where excessive parameters lead to high complexity and poor generalization. In contrast, LSTM [8] regards CSI as sequences to learn temporal features. Despite its advanced performance compared to MLPs, LSTM is inefficient for lengthy sequences due to its step-by-step inputs. To boost both effectiveness and efficiency, CNNs [55, 67, 68] have been utilized by dividing the CSI into diverse receptive fields to learn spatial features with convolution filters. To combine the advantages of CNN and LSTM, recent works [32, 40, 75] have proposed various CNN-LSTM hybrids, attaining the best performance for WiFi-based human identification.

WiFi-based human localization estimates user locations to support human-computer interaction systems [48]. Initially, Naive Bayes [62], a basic statistic model, has been leveraged for localization, showing acceptable performance. Thereafter, Sparse Auto-encoder (SAE) [15] has been used to localize users based on CSI images, but its fully connected layers still lead to excessive parameters. LSTM [9] and CNNs [57] have been further explored for localization, where CNN-1D [57] shows its superior capability for WiFi-based human localization.

WiFi-based HAR is gaining popularity for user behavior analysis, which is of finer granularity than identification and localization. Initial methods [72] based on handcrafted feature extraction have demonstrated inadequate ability to extract implicit features from CSI. Thereby, LSTM [72] and CNNs [41, 74, 77] have been extensively adopted and offer valuable insights on learning temporal and spatial features. CNN-LSTM hybrids [22, 47, 84] further leverage the advantages of CNNs and LSTM to become predominant models for HAR. To tackle dynamic environments, GANs [54] have been used to augment CNNs with adversarial learning. Attention-based models [52] also contribute to HAR [80], such as ABLSTM [7] which equips bidirectional LSTM with attention layers to enhance HAR performance. Recently, a two-stream convolution augmented transformer (THAT) [33] has further integrated attention layers with multi-scale CNNs, achieving state-of-the-art performance for WiFi-based HAR.

However, all the above methods carry out single-user sensing only, while multi-user sensing based on WiFi CSI is more practical yet challenging. Multiple users in an environment may occlude each other, challenging the resolution of models. The mutual interference between users also requires models to disentangle the features of different users for effective sensing. Recent works [12, 24, 31, 49, 53, 58] have attempted to tackle these challenges and sense

multiple users simultaneously, but they have not published their datasets and have not provided any benchmark. Such an obvious need motivates WiMANS as the first benchmark dataset that involves multiple users in each CSI sample.

2.2 Datasets for WiFi-based Human Sensing

High-quality annotated CSI datasets are essential for advancing WiFi-based human sensing research. Table 1 summarizes existing CSI datasets in comparison to WiMANS. At the beginning, Yousefi *et al.* [72] collected a dataset in an office for WiFi-based HAR, after which various CSI datasets were released for HAR in different environments [4, 14, 21, 70]. Baha *et al.* [2] further constructed a dataset under line-of-sight and non-line-of-sight conditions. Some datasets were presented for specific tasks, such as SignFi [38] for sign language recognition, FallDeFi [43] for fall detection, and ARIL [56] for activity recognition and localization. For human tracking, Widar [44] and Widar 2.0 [45] were gathered and annotated with user walking traces. Some datasets contain changing scenarios, such as CPAR [65] for different users and RF-NET [11] for varying environments. Similarly, CSIDA [26] and Widar 3.0 [76] incorporate CSI of various domains (*i.e.*, locations, orientations, environments, and users). OPERAnet [3], a multi-modal dataset, includes WiFi CSI as well as data from Passive WiFi Radar (PWR), Ultra-Wideband (UWB), and Kinect skeleton sensors. MM-Fi [69] is also a multi-modal dataset, containing WiFi CSI, video frames, depth frames, LiDAR point cloud, and mmWave radar point cloud. For benchmarking, Yang *et al.* [66] collected the NTU-Fi dataset to compare WiFi-based human sensing models in terms of their performance, sizes, complexity, *etc.* To analyze the impact of wireless parameters (*e.g.,* bandwidth), Meneghello *et al.* [39] constructed the SHARPax dataset to explore WiFi-based HAR under varying settings.

Despite these datasets, none of them enables simultaneous multi-user sensing, which is crucial for practical WiFi-based human sensing advancement, since real-life scenarios typically involve multiple users simultaneously. To bridge this gap, we propose WiMANS, the first dataset for WiFi-based multi-user activity sensing, hoping to facilitate the next generation of human sensing based on WiFi.

3 WiMANS

WiMANS aims to gather WiFi CSI samples which monitor simultaneous activities of multiple users. Each CSI sample contains 0 to 5 users performing identical/different activities at the same time, as shown in Fig. 1. Fine-grained annotations are provided for all samples, including user identities, locations, and activities. Ultimately, WiMANS collects 11286 samples (over 9.4 hours) of dual-band WiFi CSI and synchronized videos.

3.1 Dataset Construction

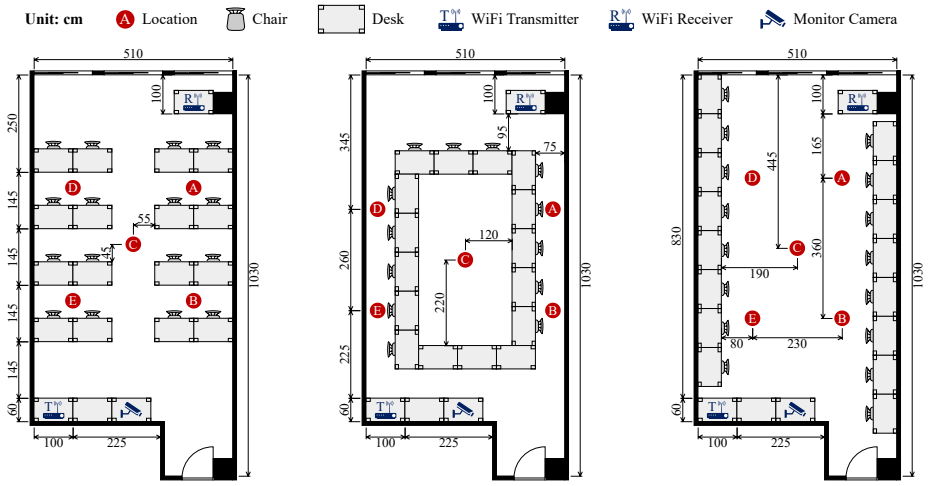
Hardware Setup. WiFi CSI monitors the variations of wireless signals by which a transmitter propagates packets to a receiver. To collect WiFi CSI, we deploy

two off-the-shelf computers (HP EliteDesk 800 G2 TWR) to serve as the transmitter and receiver. Each device is equipped with an Intel 5300 Network Interface Card and has the Linux 802.11n CSI tool [23] installed, following the previous works [3, 72, 76]. We set these devices to work in the monitor mode, enabling us to control the start and the end of each CSI sample. To exert the advantages of different WiFi bands, we collect dual-band WiFi CSI by setting devices to work on channel 12 for the 2.4 GHz band, and on channel 64 for the 5 GHz band. Meanwhile, we use a monitor camera to capture synchronized videos.

Data Collection. We employ the transmitter and receiver to collect a CSI sample in 3 steps: (1) The receiver listens to a WiFi channel and logs the CSI from all packets it receives; (2) The transmitter sends packets to the WiFi channel, and the users simultaneously perform designated activities; (3) The receiver stops logging and listening. Following the previous works [3, 76], we instruct users to perform activities in 3 seconds and thereby control the transmitter to send 3000 packets at a rate of 1000 packets per second. Regardless of packet loss, each 3-second CSI sample should have 3000 time steps under the sample rate of 1000 Hz. We will further analyze the issue of packet loss in Section 3.2. The transmitter and receiver individually have 3 antennas, and each pair of antennas uses 30 subcarriers for wireless communication. Thus, the dimension of CSI at each time step is $3\times3\times30$, and the dimension of each CSI sample is $3000\times3\times3\times30$. Along with each CSI sample, we also capture synchronized videos with a monitor camera for reference and unexplored tasks (*e.g.*, multi-user pose estimation). Each 3-second video has 90 frames under the frame rate of 30 Hz, with 3 RGB channels and the frame resolution of 1920×1080. Hence, the dimension of each video sample is $90\times3\times1920\times1080$.

Data Synchronization. During data collection, WiFi CSI is collected sample by sample, while long videos are recorded at the same time. Afterward, we synchronize WiFi CSI and videos by segmenting long videos into 3-second samples based on timestamps. Because the frame rate of videos is 30 Hz while the sample rate of CSI is 1000 Hz, we can synchronize WiFi CSI and video samples in 16.67 ms, which is unavoidable and tolerable for 3-second samples.

Data Attributes. WiMANS includes 9 activities which are representative in daily life, as discussed in the previous works [3, 72, 76]. These activities are: (1) Nothing, (2) Walking, (3) Rotation, (4) Jumping, (5) Waving, (6) Lying Down, (7) Picking Up, (8) Sitting Down, (9) Standing Up. To build up a comprehensive dataset, we collect data in 3 daily environments: (1) Classroom, (2) Meeting Room, (3) Empty Room. In each environment, we specify 5 locations and mark them as A, B, C, D, and E. Fig. 2 describes the layouts of three environments and the interior locations. We recruit 6 volunteers to act as users, including 3 females and 3 males, with an average age of 27.33 ± 0.94, height of 169.00 ± 6.83 cm, weight of 61.17 ± 10.04 kg, and Body Mass Index (BMI) of 21.25 ± 1.78. These users are assigned with identity labels (*e.g.*, User 1, User 2, ...) to protect their privacy and to support human identification. Regarding the human subject study, we have obtained an approval from institutional ethics committee (IRB) in advance and will provide an ethics statement in Section 5.3.



(a) Classroom. **(b)** Meeting Room. **(c)** Empty Room.

Fig. 2: Layouts of environments in WiMANS, where a transmitter and a receiver collect WiFi CSI, and a monitor camera captures synchronized videos for reference.

Collection Setup. We control the collection of each 3-second sample in three dimensions: (1) users, (2) locations, (3) activities. For the user dimension, we organize volunteers into user groups, where each group includes fixed users. For the location and activity dimensions, we devise simultaneous scripts for users in each group, so that they know where to stay and what to do independently yet simultaneously. User groups and simultaneous scripts are described as follows.

User Groups. We collect samples by user groups, where each group corresponds to a specific number of users and environment. For example, Fig. 1a shows a sample in the 30th user group, where 3 users (User 1, 5, and 6) are in the classroom. Specifically, there are 6 groups for 0 user, 36 groups for 1 user, 18 groups for 2 users, 18 groups for 3 users, 18 groups for 4 users, and 18 groups for 5 users. On the other hand, there are 38 groups for each environment. Each group contains 99 samples, and we totally obtain 11286 samples for 114 user groups in WiMANS. In each user group, we design simultaneous scripts for users to perform activities.

Simultaneous Scripts. In each group, we allocate scripts to fixed users, instructing them to perform activities at different locations independently yet simultaneously. To this end, we devise a script set $\{\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\delta}, \boldsymbol{\gamma}, \boldsymbol{\lambda}\}$, as shown in Fig. 3. Each script has 99 indexes, and each index corresponds to an activity at a location. Once we announce an index, users with different scripts realize where to stay and what to do. For example, in Fig. 1a, after we announce the index 25, User 1 with $\boldsymbol{\alpha}$ performs Picking Up, User 5 with $\boldsymbol{\lambda}$ performs Walking, and User 6 with $\boldsymbol{\beta}$ performs Sitting Down. Note that the scripts are put away before activities to minimize their impacts on data collection. Between the collection of each two samples, we first stop the collection of previous sample and then

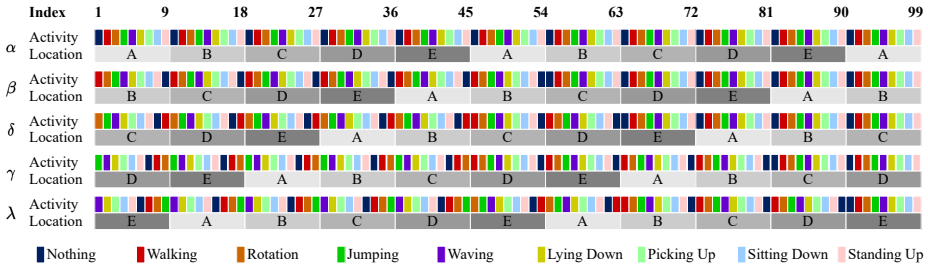


Fig. 3: Scripts that instruct users to perform identical/different activities at varying locations independently yet simultaneously.

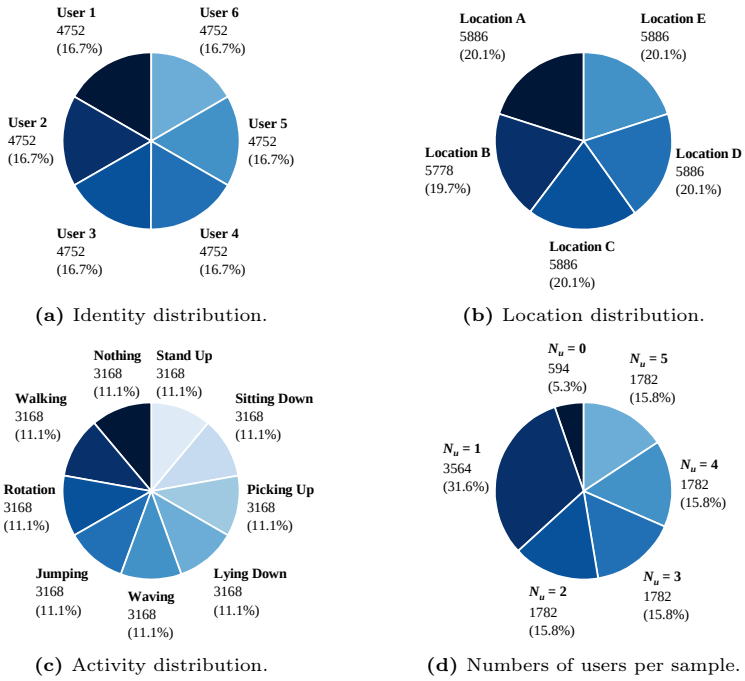
announce a new index to users, after which we start the collection of next sample. Therefore, we can ensure the purity of each sample, not containing nearby activities. More details of activities, users, user groups, and simultaneous scripts are provided in Appendix A.

Data Annotation. Since we apply the aforementioned scripts to instruct user activities, we have actually created data annotations before collection. We label each sample as "act_<group>_<sample>", where "<group>" is the user group index, and "<sample>" is the sample index. For example, Fig. 1a shows the sample labeled as "act_30_25", manifesting that it is the 25th sample of the 30th user group. We associate each label with the environment, WiFi band, number of users, identities, locations, and activities, to enable different sensing tasks. Note that the videos share the same labels with CSI and thereby can act as references to support unexplored tasks (*e.g.*, multi-user pose estimation).

3.2 Dataset Statistics

Data Distribution. WiMANS incorporates varying numbers of users in 11286 samples. 594 samples (5.3%) contain 0 user; 3564 samples (31.6%) contain 1 user; 7128 (63.2%) samples contain multiple users. Among the multi-user samples, there are 1782 (15.8%) samples for 2 users, 1782 (15.8%) samples for 3 users, 1782 (15.8%) samples for 4 users, and 1782 (15.8%) samples for 5 users. Fig. 4 presents the statistics of WiMANS. There are 4752 samples for each user, 5778∼5886 samples for each location, and 3168 samples for each activity, empowering models to learn representative features for different sensing tasks. Note that the sum of identities/locations/activities is not equal to the total sample number (11286), since each multi-user sample involves more than one identity, location, and activity.

Packet Loss. Ideally, each CSI sample consists of 3000 time steps, as mentioned in Section 3.1. However, packet loss inevitably exists in wireless communication, leading to missing time steps in CSI samples. The average packet loss rates are 4.52% and 2.31% for the 2.4 GHz band and the 5 GHz band, respectively. On average, 2.4 GHz samples have 2864.39 time steps, while 5 GHz samples have 2930.72 time steps. 2.4 GHz band suffers from more severe packet loss because it



(a) Identity distribution. **(b)** Location distribution.

(c) Activity distribution. **(d)** Numbers of users per sample.

Fig. 4: Statistics of WiMANS regarding the distributions of user identities, locations, activities, and numbers of users per sample. In each distribution, all categories have almost equivalent proportions to construct WiMANS as a relatively balanced dataset.

is more commonly used and more crowded than 5 GHz band, affected by more environmental noise [61,70]. Thus, most existing works [3,72,76] only apply the 5 GHz band for WiFi-based human sensing. However, compared to 5 GHz signals, 2.4 GHz signals can theoretically cover a larger area using longer wavelength to better penetrate and/or bypass obstacles. This inspires us to study dual-band augmented human sensing in the future, since WiMANS has gathered CSI of both 2.4 GHz and 5 GHz WiFi bands.

4 Experiments

This section utilizes WiMANS to benchmark the multi-user sensing performance of state-of-the-art WiFi-based and video-based models, with respect to human identification, localization, and activity recognition (HAR). We also compare these models in terms of model complexity and time efficiency.

4.1 Baselines

WiFi-based models. We evaluate 8 WiFi-based models on WiMANS, including Random Forest based on Short-time Fourier Transform (ST-RF) [72],

MLP [66], LSTM [72], CNN-1D [57], CNN-2D [41], CLSTM [40], ABLSTM [7], and THAT [33]. Particularly, CLSTM [40] is a hybrid of CNNs and LSTM, outperforming other models in WiFi-based identification. For WiFi-based localization, CNN-1D [57] has demonstrated the best performance by learning human features along the temporal dimension. THAT [33] equips two-stream Transformer encoders [52] with multi-scale convolutions and achieves state-of-the-art performance in WiFi-based HAR.

Video-based models. We apply 6 state-of-the-art video classification models for comparison, including ResNet [51], S3D [63], MViT-v1 [13], MViT-v2 [35], Swin-T [37], and Swin-S [37]. These models have demonstrated state-of-the-art performance in generic video classification.

4.2 Evaluation Metrics

We employ accuracy to measure the recognition performance of multi-user sensing, following the previous works [7,33,72]. Note that each multi-user sample includes multiple identity/location/activity labels corresponding to different users, respectively. For example, Fig. 1a shows a sample involving three identities: User 1, User 5, and User 6. Therefore, we measure the accuracy of recognizing these labels in each sensing task, rather than calculating the accuracy of classifying each sample. To evaluate model complexity and time efficiency, we adopt the number of parameters, floating point operations (FLOPs), and recognition throughput as metrics, following the previous works [33, 66].

4.3 Implementation Details

Data Preprocessing. We analyze model performance in different environments and accordingly split WiMANS into 3 subsets. Each subset is randomly split into a training set (80%) and a test set (20%) for evaluation. Following the previous works [7,33,72], we calculate and utilize the amplitudes of CSI for the evaluation of WiFi-based models. Inputs of all models are resized following the original papers (*e.g.*, $90\times3\times112\times112$ for ResNet [51]).

Hyperparameters. (1) For WiFi-based models, we leverage publicly accessible implementations as much as possible. Specifically, we initialize WiFi-based models with Xavier [18] and use a fixed learning rate of 10^{-3} to train them for 200 epochs with the batch size of 128. (2) For video-based models, we utilize the implementations provided by PyTorch and initialize them with the weights pre-trained by Kinetics-400 [5]. All video-based models are trained for 20 epochs with a fixed learning rate of 10^{-4} and the batch size of 8.

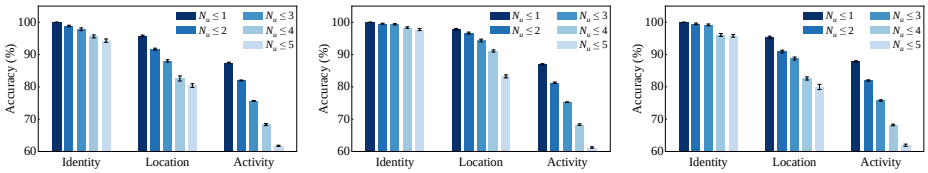
Evaluation. To recognize multiple labels, we connect each model to a linear layer followed by a sigmoid function for multi-label classification, using a fixed threshold of 0.5 to determine identities, locations, and activities. Both WiFi-based and video-based models are optimized by Adam [30] on a single Nvidia RTX A5000 GPU. We repeat each experiment 10 times with random seeds and report the means and standard deviations of results. More implementation details are provided in Appendix C.

Table 2: Recognition performance of WiFi-based models on WiMANS in terms of accuracy (%). Models show desirable performance for multi-user identification and localization, while there is still vast room for improvement in activity recognition.

Model	Classroom			Meeting Room			Empty Room		
	Identity	Location	Activity	Identity	Location	Activity	Identity	Location	Activity
2.4 GHz									
ST-RF [72]	66.3±1.33	62.9±0.36	56.8±0.14	79.0±1.32	65.7±0.55	56.8±0.09	71.0±2.20	61.5±0.54	56.7±0.09
MLP [66]	65.7±0.91	60.5±0.47	57.6±0.13	75.9±0.62	64.2±0.47	57.8±0.16	75.3±0.59	58.8±0.26	57.0±0.12
LSTM [72]	80.7±1.43	66.9±1.09	59.2±0.42	89.5±1.11	71.7±0.88	59.0±0.33	86.7±0.99	67.8±0.65	57.7±0.54
CNN-1D [57]	84.8±0.80	73.2±0.22	59.6±0.24	92.9±0.26	77.2±0.19	58.5±0.11	88.1±0.46	74.5±0.31	58.0±0.17
CNN-2D [41]	90.0±0.66	77.0±0.37	59.5±0.15	**96.4**±0.81	82.1±0.40	59.0±0.12	91.1±0.82	79.1±0.52	58.2±0.26
CLSTM [40]	90.2±0.70	73.0±0.68	**61.8**±0.55	94.6±0.49	77.4±0.75	60.9±0.39	92.8±0.27	71.8±0.95	**61.0**±0.40
ABLSTM [7]	88.8±1.36	76.6±0.42	61.6±0.50	94.8±0.55	77.7±0.61	**60.9**±0.27	90.9±0.54	76.0±0.55	59.7±0.15
THAT [33]	**90.7**±0.51	**78.8**±0.74	61.0±0.32	94.8±0.59	**82.2**±0.84	60.1±0.28	**93.9**±0.88	**80.8**±1.11	59.7±0.29
5 GHz									
ST-RF [72]	89.4±1.08	62.7±0.57	57.3±0.08	95.1±0.83	67.4±0.65	57.6±0.13	81.9±1.26	61.9±0.56	57.2±0.21
MLP [66]	98.6±0.13	72.7±0.82	58.6±0.14	99.5±0.11	79.1±0.42	57.4±0.14	95.1±0.58	74.5±0.87	59.2±0.23
LSTM [72]	98.9±0.16	76.1±0.68	60.6±0.28	99.7±0.17	81.3±0.80	59.5±0.23	95.8±0.50	77.4±0.37	59.8±0.36
CNN-1D [57]	99.5±0.21	82.4±0.34	60.6±0.22	99.7±0.12	87.1±0.32	59.0±0.16	95.9±0.36	84.0±0.58	60.2±0.14
CNN-2D [41]	99.1±0.21	82.7±0.41	59.8±0.30	99.8±0.11	**88.2**±0.37	58.6±0.29	95.3±0.36	**84.4**±0.40	59.6±0.28
CLSTM [40]	**99.7**±0.19	82.0±0.50	**64.2**±0.55	99.8±0.08	88.1±0.55	**62.1**±0.36	**97.6**±0.25	84.4±0.64	**64.8**±0.38
ABLSTM [7]	99.6±0.29	82.5±0.67	61.4±0.22	99.8±0.08	87.4±0.25	60.4±0.16	96.7±0.38	83.8±0.53	61.3±0.15
THAT [33]	99.2±0.19	**83.1**±0.67	61.8±0.29	**99.9**±0.00	88.1±0.59	61.2±0.37	97.1±0.29	83.7±1.05	62.1±0.41
2.4 / 5 GHz									
ST-RF [72]	77.9±1.05	62.5±0.39	57.3±0.09	87.7±0.91	66.7±0.34	57.4±0.07	76.0±1.06	62.1±0.28	57.2±0.08
MLP [66]	74.9±0.48	66.0±0.32	57.9±0.13	85.9±0.78	69.2±0.28	57.6±0.07	80.1±0.53	66.2±0.48	58.4±0.08
LSTM [72]	86.9±0.90	69.5±0.79	59.6±0.24	93.9±0.60	75.2±0.46	59.7±0.31	90.1±0.95	70.4±0.27	58.9±0.25
CNN-1D [57]	92.6±0.42	78.3±0.25	60.7±0.14	96.6±0.21	82.8±0.25	59.5±0.17	93.5±0.21	78.5±0.21	59.7±0.12
CNN-2D [41]	**94.6**±0.50	79.2±0.40	59.2±0.29	**98.0**±0.35	**83.6**±0.36	58.5±0.28	94.5±0.41	**80.1**±0.51	58.3±0.21
CLSTM [40]	93.1±0.46	79.9±0.48	**64.0**±0.42	97.4±0.28	83.1±0.33	**62.4**±0.28	95.3±0.38	78.2±0.60	**63.5**±0.36
ABLSTM [7]	92.4±0.39	79.2±0.38	62.1±0.16	97.8±0.17	82.8±0.42	60.9±0.17	95.0±0.34	78.9±0.54	61.4±0.29
THAT [33]	94.3±0.51	**80.4**±0.56	61.8±0.19	97.7±0.33	83.3±0.45	61.2±0.22	**95.8**±0.40	80.0±0.82	62.0±0.33

4.4 Results and Analysis

WiFi-based Multi-user Sensing. Table 2 presents the multi-user sensing performance of WiFi-based models. (1) Using the 2.4 GHz WiFi band, THAT and CNN-2D are the best performing models for identification, while CNN-2D outperforms other models in localization, and CLSTM achieves the best performance in HAR. (2) With the 5 GHz WiFi band, CLSTM and THAT show the highest accuracy in recognizing multiple identities, while THAT and CNN-2D demonstrate their superiority in localization, and CLSTM obtains better results than other models in HAR. (3) Exploiting both 2.4 and 5 GHz WiFi bands, CNN-2D and THAT have the best results in identifying and localizing multiple users, while CLSTM yields the highest accuracy for HAR. Generally, we can observe that the state-of-the-art models show better performance on identification (90.69∼99.99%) than localization (78.78∼88.20%), while having even lower performance on HAR (60.87∼64.82%). This is because HAR is more fine-grained than the other two tasks, causing higher difficulties to learn representative features. In contrast, multi-user identification is less fine-grained than localization and thereby results in better performance. Comparing different WiFi bands, using 5 GHz results in higher accuracy than 2.4 GHz, since the 2.4 GHz WiFi band suffers from more noise than 5 GHz, as mentioned in Section 3.2. Meanwhile, simply using the samples of two WiFi bands together cannot produce models of better performance. These results inspire us to study dual-band augmented multi-user sensing in the future.



(a) Classroom. **(b)** Meeting Room. **(c)** Empty Room.

Fig. 5: Recognition performance of THAT [33] in terms of accuracy (%) regarding different numbers of users in each CSI sample (N_u). Sensing more users simultaneously results in lower performance. ("$N_u \leq k$": each sample includes 0 to k users.)

Table 3: Recognition performance of video-based models on WiMANS in terms of accuracy (%). Video-based models show better performance than WiFi-based models.

Model	Classroom			Meeting Room			Empty Room		
	Identity	Location	Activity	Identity	Location	Activity	Identity	Location	Activity
ResNet [51]	99.0±0.93	95.9±1.70	72.3±0.86	99.2±0.94	95.6±5.14	73.6±1.13	96.1±1.39	96.2±5.38	76.5±4.74
S3D [63]	**99.9**±0.00	99.5±0.50	94.0±1.35	99.9±0.04	**99.9**±0.02	**98.0**±0.12	99.7±0.32	**99.9**±0.01	**98.5**±0.15
MViT-v1 [13]	99.9±0.05	99.3±0.26	93.7±0.44	99.7±0.12	99.6±0.05	90.7±0.11	99.7±0.04	99.8±0.01	94.0±0.10
MViT-v2 [35]	99.8±0.06	**99.7**±0.05	95.1±0.11	99.8±0.13	99.8±0.04	92.3±0.13	99.8±0.12	99.9±0.01	95.2±0.11
Swin-T [37]	99.8±0.05	99.6±0.05	**96.6**±0.12	**99.9**±0.05	99.7±0.07	89.1±0.14	**99.9**±0.04	99.8±0.01	94.1±0.13
Swin-S [37]	99.9±0.07	99.6±0.06	95.8±0.11	99.5±0.08	99.4±0.05	90.5±0.11	99.8±0.12	99.8±0.02	96.1±0.11

Numbers of Users. Fig. 5 analyzes the impacts of user numbers on the performance of THAT. As expected, sensing more users simultaneously results in lower recognition accuracy. (1) In the classroom, when the maximum user number increases from 1 to 5, the accuracy of identification, localization, and HAR decreases by 5.70%, 15.39%, and 25.74%, respectively. (2) In the meeting room, sensing up to 5 users simultaneously diminishes the accuracy of three tasks by 2.26%, 14.59%, and 25.78% compared to sensing a single user. (3) In the empty room, similar results are evident where the accuracy of three tasks declines by 4.19%, 15.36%, and 25.90% as the maximum user number raises from 1 to 5. We can observe that, compared to localization and identification, the performance of HAR is more sensitive to changes in the number of users, because HAR is of finer granularity, and increasing users leads to more occlusion and mutual interference. Meanwhile, in varying environments, the number of users exerts distinct impacts on the recognition accuracy, which is worthwhile to further study. These results illustrate the usefulness of WiMANS and highlight the challenges of sensing multiple users simultaneously based on WiFi CSI.

Video-based Multi-user Sensing. We provide the performance of video-based models in Table 3 for comparison. Overall, video-based models achieve better performance than WiFi-based models in three sensing tasks. Such results are expected since video-based analysis has been well studied in computer vision, which also demonstrate the quality of WiMANS, where users have performed distinguishing activities. Therefore, we can further exploit the videos in WiMANS as a reference for unexplored tasks in the future. Despite the remarkable performance of video-based models, they have much higher complexity and lower efficiency than WiFi-based models, as discussed below.

Table 4: Model complexity and time efficiency on WiMANS, where WiFi-based models show their superiority over video-based models. ("Recs/s": recognitions per second.)

Data	Model	Input Size	Parameters (M)	FLOPs (G)	Throughput (Recs/s)		
					Identity	Location	Activity
WiFi CSI	MLP [66]	810000	209.020	**0.418**	2918.97	**3399.19**	**3385.97**
	LSTM [72]	3000×270	1.609	0.971	**3047.69**	3045.22	3058.85
	CNN-1D [57]	3000×270	1.916	0.516	2685.73	2679.96	2670.38
	CNN-2D [41]	3000×270	**0.893**	1.691	2132.54	2183.23	2137.92
	CLSTM [40]	3000×270	5.391	1.791	2704.11	2707.01	2754.48
	ABLSTM [7]	3000×270	4.268	3.208	2493.05	2614.77	2708.44
	THAT [33]	3000×270	4.900	1.650	1937.48	1971.58	1955.32
Video	ResNet [51]	90×3×112×112	33.393	**17.670**	50.81	51.10	51.10
	S3D [63]	90×3×224×224	**8.342**	204.644	**113.86**	**120.74**	**117.29**
	MViT-v1 [13]	90×3×224×224	36.632	632.698	47.93	48.03	47.94
	MViT-v2 [35]	90×3×224×224	34.559	577.742	38.24	38.17	38.16
	Swin-T [37]	90×3×224×224	28.180	268.797	44.90	44.80	44.83
	Swin-S [37]	90×3×224×224	49.838	518.993	27.40	27.43	27.41

Model Complexity and Time Efficiency. Table 4 highlights the superiority of WiFi-based models compared with video-based models in terms of model complexity and time efficiency. Except for MLP which is rarely used in practice due to excessive parameters, the maximum number of parameters (5.391 M for CLSTM) in WiFi-based models is 1.55× less than the minimum number of parameters (8.342 M for S3D) in video-based models. Similarly, the maximum FLOPs (3.208 G for ABLSTM) in WiFi-based models is 5.51× lower than the minimum FLOPs (17.670 G for ResNet) in video-based models. Moreover, the throughput of THAT is 16.33∼17.02× higher than the throughput of S3D. These results signify that WiFi-based human sensing can achieve satisfactory performance with higher cost-effectiveness than video-based solutions. We further discuss the training time and testing time of models in Appendix C.

5 Discussion

To the best of our knowledge, WiMANS is the first benchmark dataset for WiFi-based multi-user activity sensing. In this section, we discuss the limitations and future work of WiMANS and provide an ethics statement.

5.1 Limitations

WiMANS aims to represent the most common multi-user scenarios in daily life. Therefore, we discuss the most general settings but not specific activities, conditions, dedicated devices, *etc.* Such general settings may lead to some limitations.
Daily Activities. WiMANS includes common daily activities only, not ethically risky activities (*e.g.*, falling, fighting), and the current activities in WiMANS are representative of daily life [3, 72, 76]. We will further consider other activities in our future work, given ethical approval to ensure the safety of volunteers.
Challenging Conditions. WiFi-based human sensing has the potential to solve the issues of obstacles. However, we do not intentionally set up obstacles since we use a camera to capture videos, though WiMANS does contain occlusion due to obstacles (*e.g.*, chairs, desks) and indeed people also overlap with each other.

WiFi Devices. We collect WiFi CSI using a single transmitter-receiver pair equipped with Intel 5300 Network Interface Cards, the most commonly used devices in the previous works [3,21,38,43,72,76]. More transmitter-receiver pairs may benefit WiFi-based multi-user sensing [1], while other dedicated devices can also collect CSI using the Atheros CSI tool [64,69], the AX-CSI tool [19,39], *etc.*

5.2 Future Work

Multi-user Pose Estimation. Recent works have discussed WiFi-based pose estimation [17, 29, 34, 46, 59, 78, 79], but they have not released their datasets. Since WiMANS collects both WiFi CSI and videos, we can extend our dataset by annotating videos with human body joints for multi-user pose estimation.
Dual-band Augmented Sensing. 2.4 GHz and 5 GHz WiFi bands have different advantages for human sensing, as mentioned in 3.2. We plan to investigate multi-user sensing augmented by dual WiFi bands for better performance.
Cross-domain Sensing. Recent studies [6, 82] have discussed cross-domain single-user sensing based on WiFi CSI. We plan to explore cross-domain multi-user sensing with varying numbers of users, unseen environments/users, *etc.*

5.3 Ethics Statement

Human subject study in WiMANS has been reviewed and approved by institutional ethics committee (IRB). Prior to data collection, each subject has been given the details of WiMANS and signed the consent form regarding safety, privacy, and releasing identifiable information. The activities in WiMANS are common in daily life (*e.g.*, Walking) to minimize participant risks. We create aliases for all subjects (*e.g.*, User 1) and blur their faces in videos to anonymize personally identifiable information. WiMANS should be used for academic research purposes only. We provide further discussions in Appendix D.

6 Conclusion

In this paper, we introduce WiMANS, the first WiFi-based human sensing dataset that incorporates multiple users in each CSI sample. WiMANS contains 11286 CSI samples of dual WiFi bands and synchronized videos, annotated with user identities, locations, and activities. We run benchmark experiments on WiMANS to evaluate the multi-user sensing performance of existing models based on WiFi and videos. WiFi-based models show desirable performance for multi-user identification and localization, while there is still vast room for improvement in activity recognition compared with video-based models. We make all data, source code, and documents publicly available, anticipating that WiMANS can advance WiFi-based human sensing research from single-user sensing towards more practical multi-user sensing. We hope that WiMANS can enable a variety of future works, such as multi-user pose estimation, dual-band augmented sensing, and cross-domain sensing.

References

1. Bahadori, N., Ashdown, J., Restuccia, F.: 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. pp. 50–59. IEEE (2022)
2. Baha'A, A., Almazari, M.M., Alazrai, R., Daoud, M.I.: A dataset for wi-fi-based human activity recognition in line-of-sight and non-line-of-sight indoor environments. Data in Brief **33**, 106534 (2020)
3. Bocus, M.J., Li, W., Vishwakarma, S., Kou, R., Tang, C., Woodbridge, K., Craddock, I., McConville, R., Santos-Rodriguez, R., Chetty, K., et al.: Operanet, a multimodal activity recognition dataset acquired from radio frequency and vision-based sensors. Scientific Data **9**(1), 474 (2022)
4. Brinke, J.K., Meratnia, N.: Dataset: Channel state information for different activities, participants and days. In: Proceedings of the 2nd Workshop on Data Acquisition to Analysis. pp. 61–64 (2019)
5. Carreira, J., Zisserman, A.: Quo vadis, action recognition? a new model and the kinetics dataset. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (July 2017)
6. Chen, C., Zhou, G., Lin, Y.: Cross-domain wifi sensing with channel state information: A survey. ACM Computing Surveys **55**(11), 1–37 (2023)
7. Chen, Z., Zhang, L., Jiang, C., Cao, Z., Cui, W.: Wifi csi based passive human activity recognition using attention based blstm. IEEE Transactions on Mobile Computing **18**(11), 2714–2724 (2018)
8. Ding, J., Wang, Y., Fu, X.: Wihi: Wifi based human identity identification using deep learning. IEEE Access **8**, 129246–129262 (2020)
9. Ding, J., Wang, Y., Si, H., Gao, S., Xing, J.: Three-dimensional indoor localization and tracking for mobile target based on wifi sensing. IEEE Internet of Things Journal **9**(21), 21687–21701 (2022)
10. Ding, J., Wang, Y., Si, H., Ma, J., He, J., Liang, K., Fu, S.: Multimodal fusion-gmm based gesture recognition for smart home by wifi sensing. In: 2022 IEEE 95th Vehicular Technology Conference. pp. 1–6. IEEE (2022)
11. Ding, S., Chen, Z., Zheng, T., Luo, J.: 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. pp. 517–530 (2020)
12. Duan, P., Li, C., Li, J., Chen, X., Wang, C., Wang, E.: Wisdom: Wi-fi based contactless multi-user activity recognition. IEEE Internet of Things Journal (2022)
13. Fan, H., Xiong, B., Mangalam, K., Li, Y., Yan, Z., Malik, J., Feichtenhofer, C.: Multiscale vision transformers. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 6824–6835 (2021)
14. Fard Moshiri, P., Shahbazian, R., Nabati, M., Ghorashi, S.A.: A csi-based human activity recognition using deep learning. Sensors **21**(21), 7225 (2021)
15. Gao, Q., Wang, J., Ma, X., Feng, X., Wang, H.: Csi-based device-free wireless localization and activity recognition using radio image features. IEEE Transactions on Vehicular Technology **66**(11), 10346–10356 (2017)
16. Ge, Y., Taha, A., Shah, S.A., Dashtipour, K., Zhu, S., Cooper, J., Abbasi, Q.H., Imran, M.A.: Contactless wifi sensing and monitoring for future healthcare-emerging trends, challenges, and opportunities. IEEE Reviews in Biomedical Engineering **16**, 171–191 (2022)

17. Geng, J., Huang, D., De la Torre, F.: Densepose from wifi. arXiv preprint arXiv:2301.00250 (2022)
18. Glorot, X., Bengio, Y.: Understanding the difficulty of training deep feedforward neural networks. In: Proceedings of the 13th International Conference on Artificial Intelligence and Statistics. pp. 249–256. JMLR Workshop and Conference Proceedings (2010)
19. Gringoli, F., Cominelli, M., Blanco, A., Widmer, J.: Ax-csi: Enabling csi extraction on commercial 802.11 ax wi-fi platforms. In: Proceedings of the 15th ACM Workshop on Wireless Network Testbeds, Experimental evaluation & CHaracterization. pp. 46–53 (2022)
20. Gu, Y., Zhang, X., Wang, Y., Wang, M., Yan, H., Ji, Y., Liu, Z., Li, J., Dong, M.: Wigrunt: Wifi-enabled gesture recognition using dual-attention network. IEEE Transactions on Human-Machine Systems **52**(4), 736–746 (2022)
21. Guo, L., Wang, L., Lin, C., Liu, J., Lu, B., Fang, J., Liu, Z., Shan, Z., Yang, J., Guo, S.: Wiar: A public dataset for wifi-based activity recognition. IEEE Access **7**, 154935–154945 (2019)
22. Guo, L., Zhang, H., Wang, C., Guo, W., Diao, G., Lu, B., Lin, C., Wang, L.: Towards csi-based diversity activity recognition via lstm-cnn encoder-decoder neural network. Neurocomputing **444**, 260–273 (2021)
23. Halperin, D., Hu, W., Sheth, A., Wetherall, D.: Tool release: Gathering 802.11 n traces with channel state information. ACM SIGCOMM Computer Communication Review **41**(1), 53–53 (2011)
24. He, J., Yang, W.: Imar: Multi-user continuous action recognition with wifi signals. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies **6**(3), 1–27 (2022)
25. Hernandez, S.M., Bulut, E.: Wifi sensing on the edge: Signal processing techniques and challenges for real-world systems. IEEE Communications Surveys & Tutorials (2022)
26. Hu, P., Tang, C., Yin, K., Zhang, X.: Wigr: a practical wi-fi-based gesture recognition system with a lightweight few-shot network. Applied Sciences **11**(8), 3329 (2021)
27. Huang, Q., Chen, H., Zhang, Q.: Joint design of sensing and communication systems for smart homes. IEEE Network **34**(6), 191–197 (2020)
28. Hussain, Z., Sheng, Q.Z., Zhang, W.E.: A review and categorization of techniques on device-free human activity recognition. Journal of Network and Computer Applications **167**, 102738 (2020)
29. Jiang, W., Xue, H., Miao, C., Wang, S., Lin, S., Tian, C., Murali, S., Hu, H., Sun, Z., Su, L.: Towards 3d human pose construction using wifi. In: Proceedings of the 26th Annual International Conference on Mobile Computing and Networking. pp. 1–14 (2020)
30. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. In: 3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings (2015)
31. Kong, H., Lu, L., Yu, J., Chen, Y., Xu, X., Tang, F., Chen, Y.C.: Multiauth: Enable multi-user authentication with single commodity wifi device. In: Proceedings of the Twenty-second International Symposium on Theory, Algorithmic Foundations, and Protocol Design for Mobile Networks and Mobile Computing. pp. 31–40 (2021)
32. Kong, H., Lu, L., Yu, J., Zhu, Y., Tang, F., Chen, Y.C., Kong, L., Lyu, F.: Push the limit of wifi-based user authentication towards undefined gestures. In: IEEE INFOCOM 2022-IEEE Conference on Computer Communications. pp. 410–419. IEEE (2022)

33. Li, B., Cui, W., Wang, W., Zhang, L., Chen, Z., Wu, M.: Two-stream convolution augmented transformer for human activity recognition. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 35, pp. 286–293 (2021)
34. Li, T., Fan, L., Zhao, M., Liu, Y., Katabi, D.: Making the invisible visible: Action recognition through walls and occlusions. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 872–881 (2019)
35. Li, Y., Wu, C.Y., Fan, H., Mangalam, K., Xiong, B., Malik, J., Feichtenhofer, C.: Mvitv2: Improved multiscale vision transformers for classification and detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 4804–4814 (2022)
36. Lin, Y., Gao, Y., Li, B., Dong, W.: Revisiting indoor intrusion detection with wifi signals: do not panic over a pet! IEEE Internet of Things Journal **7**(10), 10437–10449 (2020)
37. Liu, Z., Ning, J., Cao, Y., Wei, Y., Zhang, Z., Lin, S., Hu, H.: Video swin transformer. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 3202–3211 (2022)
38. Ma, Y., Zhou, G., Wang, S., Zhao, H., Jung, W.: Signfi: Sign language recognition using wifi. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies **2**(1), 1–21 (2018)
39. Meneghello, F., Chen, C., Cordeiro, C., Restuccia, F.: Toward integrated sensing and communications in ieee 802.11 bf wi-fi networks. IEEE Communications Magazine **61**(7), 128–133 (2023)
40. Mo, H., Kim, S.: A deep learning-based human identification system with wi-fi csi data augmentation. IEEE Access **9**, 91913–91920 (2021)
41. Moshiri, P.F., Nabati, M., Shahbazian, R., Ghorashi, S.A.: Csi-based human activity recognition using convolutional neural networks. In: 2021 11th International Conference on Computer Engineering and Knowledge. pp. 7–12. IEEE (2021)
42. Nirmal, I., Khamis, A., Hassan, M., Hu, W., Zhu, X.: Deep learning for radio-based human sensing: Recent advances and future directions. IEEE Communications Surveys & Tutorials **23**(2), 995–1019 (2021)
43. Palipana, S., Rojas, D., Agrawal, P., Pesch, D.: Falldefi: Ubiquitous fall detection using commodity wi-fi devices. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies **1**(4), 1–25 (2018)
44. Qian, K., Wu, C., Yang, Z., Liu, Y., Jamieson, K.: Widar: Decimeter-level passive tracking via velocity monitoring with commodity wi-fi. In: Proceedings of the 18th ACM International Symposium on Mobile Ad Hoc Networking and Computing. pp. 1–10 (2017)
45. Qian, K., Wu, C., Zhang, Y., Zhang, G., Yang, Z., Liu, Y.: 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. pp. 350–361 (2018)
46. Ren, Y., Wang, Z., Wang, Y., Tan, S., Chen, Y., Yang, J.: Gopose: 3d human pose estimation using wifi. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies **6**(2), 1–25 (2022)
47. Shalaby, E., ElShennawy, N., Sarhan, A.: Utilizing deep learning models in csi-based human activity recognition. Neural Computing and Applications pp. 1–18 (2022)
48. Tan, S., Ren, Y., Yang, J., Chen, Y.: Commodity wifi sensing in ten years: Status, challenges, and opportunities. IEEE Internet of Things Journal **9**(18), 17832–17843 (2022)

49. Tan, S., Zhang, L., Wang, Z., Yang, J.: Multitrack: Multi-user tracking and activity recognition using commodity wifi. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. pp. 1–12 (2019)
50. Tian, Z., Li, Y., Zhou, M., Li, Z.: Wifi-based adaptive indoor passive intrusion detection. In: 2018 IEEE 23rd International Conference on Digital Signal Processing. pp. 1–5. IEEE (2018)
51. Tran, D., Wang, H., Torresani, L., Ray, J., LeCun, Y., Paluri, M.: A closer look at spatiotemporal convolutions for action recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 6450–6459 (2018)
52. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I.: Attention is all you need. Advances in Neural Information Processing Systems **30** (2017)
53. Venkatnarayan, R.H., Page, G., Shahzad, M.: Multi-user gesture recognition using wifi. In: Proceedings of the 16th Annual International Conference on Mobile Systems, Applications, and Services. pp. 401–413 (2018)
54. Wang, D., Yang, J., Cui, W., Xie, L., Sun, S.: Multimodal csi-based human activity recognition using gans. IEEE Internet of Things Journal **8**(24), 17345–17355 (2021)
55. Wang, D., Yang, J., Cui, W., Xie, L., Sun, S.: Caution: A robust wifi-based human authentication system via few-shot open-set recognition. IEEE Internet of Things Journal **9**(18), 17323–17333 (2022)
56. Wang, F., Feng, J., Zhao, Y., Zhang, X., Zhang, S., Han, J.: Joint activity recognition and indoor localization with wifi fingerprints. IEEE Access **7**, 80058–80068 (2019)
57. Wang, F., Feng, J., Zhao, Y., Zhang, X., Zhang, S., Han, J.: Joint activity recognition and indoor localization with wifi fingerprints. IEEE Access **7**, 80058–80068 (2019)
58. Wang, F., Zhou, S., Panev, S., Han, J., Huang, D.: Person-in-wifi: Fine-grained person perception using wifi. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 5452–5461 (2019)
59. Wang, F., Zhou, S., Panev, S., Han, J., Huang, D.: Person-in-wifi: Fine-grained person perception using wifi. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 5452–5461 (2019)
60. Wang, W., Liu, A.X., Shahzad, M., Ling, K., Lu, S.: Understanding and modeling of wifi signal based human activity recognition. In: Proceedings of the 21st Annual International Conference on Mobile Computing and Networking. pp. 65–76 (2015)
61. Wei, B., Hu, W., Yang, M., Chou, C.T.: Radio-based device-free activity recognition with radio frequency interference. In: Proceedings of the 14th International Conference on Information Processing in Sensor Networks. pp. 154–165 (2015)
62. Wu, Z., Xu, Q., Li, J., Fu, C., Xuan, Q., Xiang, Y.: Passive indoor localization based on csi and naive bayes classification. IEEE Transactions on Systems, Man, and Cybernetics: Systems **48**(9), 1566–1577 (2017)
63. Xie, S., Sun, C., Huang, J., Tu, Z., Murphy, K.: Rethinking spatiotemporal feature learning: Speed-accuracy trade-offs in video classification. In: Proceedings of the European Conference on Computer Vision. pp. 305–321 (2018)
64. Xie, Y., Li, Z., Li, M.: Precise power delay profiling with commodity wifi. In: Proceedings of the 21st Annual international conference on Mobile Computing and Networking. pp. 53–64 (2015)
65. Xu, S., He, Z., Shi, W., Wang, Y., Ohtsuki, T., Guiy, G.: Cross-person activity recognition method using snapshot ensemble learning. In: 2022 IEEE 96th Vehicular Technology Conference. pp. 1–5. IEEE (2022)

66. Yang, J., Chen, X., Zou, H., Lu, C.X., Wang, D., Sun, S., Xie, L.: Sensefi: A library and benchmark on deep-learning-empowered wifi human sensing. Patterns **4**(3) (2023)
67. Yang, J., Chen, X., Zou, H., Wang, D., Xie, L.: Autofi: Towards automatic wifi human sensing via geometric self-supervised learning. IEEE Internet of Things Journal (2022)
68. Yang, J., Chen, X., Zou, H., Wang, D., Xu, Q., Xie, L.: Efficientfi: Toward large-scale lightweight wifi sensing via csi compression. IEEE Internet of Things Journal **9**(15), 13086–13095 (2022)
69. Yang, J., Huang, H., Zhou, Y., Chen, X., Xu, Y., Yuan, S., Zou, H., Lu, C.X., Xie, L.: Mm-fi: Multi-modal non-intrusive 4d human dataset for versatile wireless sensing. arXiv preprint arXiv:2305.10345 (2023)
70. Yang, J., Liu, Y., Liu, Z., Wu, Y., Li, T., Yang, Y.: A framework for human activity recognition based on wifi csi signal enhancement. International Journal of Antennas and Propagation **2021**, 1–18 (2021)
71. Yang, Z., Zhang, Y., Zhang, Q.: Rethinking fall detection with wi-fi. IEEE Transactions on Mobile Computing (2022)
72. Yousefi, S., Narui, H., Dayal, S., Ermon, S., Valaee, S.: A survey on behavior recognition using wifi channel state information. IEEE Communications Magazine **55**(10), 98–104 (2017)
73. Zhang, J., Tang, Z., Li, M., Fang, D., Nurmi, P., Wang, Z.: Crosssense: Towards cross-site and large-scale wifi sensing. In: Proceedings of the 24th Annual International Conference on Mobile Computing and Networking. pp. 305–320 (2018)
74. Zhang, R., Jiang, C., Wu, S., Zhou, Q., Jing, X., Mu, J.: Wi-fi sensing for joint gesture recognition and human identification from few samples in human-computer interaction. IEEE Journal on Selected Areas in Communications **40**(7), 2193–2205 (2022)
75. Zhang, R., Wu, S., Jiang, C., Cui, Y., Jing, X.: Wirelessid: Device-free human identification using gesture signatures in csi. In: 2021 IEEE 94th Vehicular Technology Conference. pp. 1–4. IEEE (2021)
76. Zhang, Y., Zheng, Y., Qian, K., Zhang, G., Liu, Y., Wu, C., Yang, Z.: Widar3.0: Zero-effort cross-domain gesture recognition with wi-fi. IEEE Transactions on Pattern Analysis and Machine Intelligence **44**(11), 8671–8688 (2021)
77. Zhang, Y., Yin, Y., Wang, Y., Ai, J., Wu, D.: Csi-based location-independent human activity recognition with parallel convolutional networks. Computer Communications **197**, 87–95 (2023)
78. Zhao, M., Li, T., Abu Alsheikh, M., Tian, Y., Zhao, H., Torralba, A., Katabi, D.: Through-wall human pose estimation using radio signals. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 7356–7365 (2018)
79. Zhao, M., Tian, Y., Zhao, H., Alsheikh, M.A., Li, T., Hristov, R., Kabelac, Z., Katabi, D., Torralba, A.: Rf-based 3d skeletons. In: Proceedings of the 2018 Conference of the ACM Special Interest Group on Data Communication. pp. 267–281 (2018)
80. Zhou, Y., Huang, H., Yuan, S., Zou, H., Xie, L., Yang, J.: Metafi++: Wifi-enabled transformer-based human pose estimation for metaverse avatar simulation. IEEE Internet of Things Journal (2023)
81. Zhou, Z., Wang, F., Yu, J., Ren, J., Wang, Z., Gong, W.: Target-oriented semi-supervised domain adaptation for wifi-based har. In: IEEE INFOCOM 2022-IEEE Conference on Computer Communications. pp. 420–429. IEEE (2022)

82. Zhou, Z., Wang, F., Yu, J., Ren, J., Wang, Z., Gong, W.: Target-oriented semi-supervised domain adaptation for wifi-based har. In: IEEE INFOCOM 2022-IEEE Conference on Computer Communications. pp. 420–429. IEEE (2022)
83. Zhu, G., Wu, C., Zeng, X., Wang, B., Liu, K.R.: Who moved my cheese? human and non-human motion recognition with wifi. In: 2022 IEEE 19th International Conference on Mobile Ad Hoc and Smart Systems. pp. 476–484. IEEE (2022)
84. Zou, H., Zhou, Y., Yang, J., Jiang, H., Xie, L., Spanos, C.J.: Deepsense: Device-free human activity recognition via autoencoder long-term recurrent convolutional network. In: 2018 IEEE International Conference on Communications. pp. 1–6. IEEE (2018)