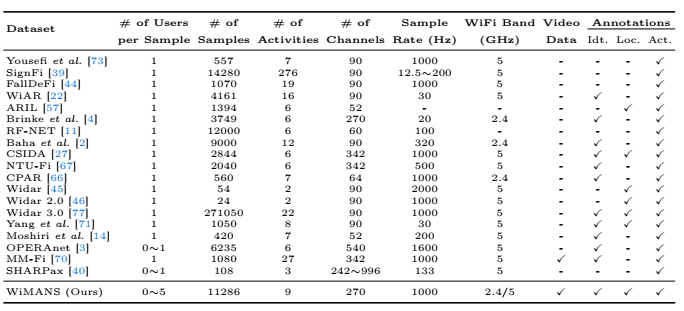
**WiMANS: A Benchmark Dataset for WiFi-based Multi-user Activity Sensing  
  
Abstract.** WiFi-based human sensing has exhibited remarkable potential to analyze user behaviors in a non-intrusive and device-free manner, benefiting applications as diverse as smart homes and healthcare. However, most previous 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 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**

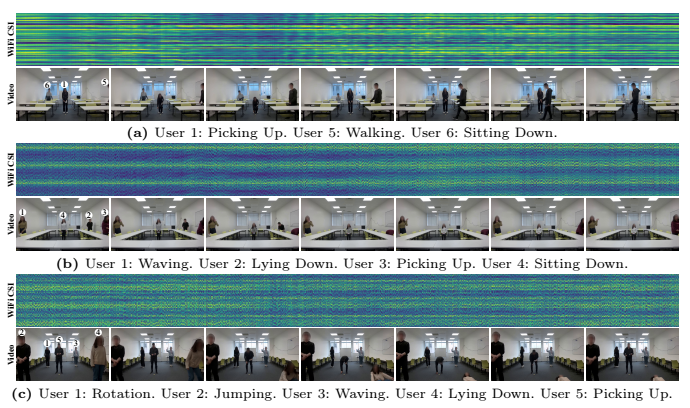
Recent years have witnessed the rapid progress of WiFi-based human sensing [6,26,49], which collects Channel State Information (CSI) from off-the-shelf WiFi devices to recognize human identities [56, 69, 76], locations [9, 12, 46], activities [7,21,34], *etc*. It plays an increasingly important role in differing applications such as security monitoring [37,51,84], smart homes [10,28], and healthcare [16,44,72]. Compared with cameras and on-body sensors, the use of WiFi CSI negates the necessity of filming users or attaching sensors to them [43]. Such non-intrusive approaches can make human sensing widely available [82], 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 [33]. More importantly, CSI can be gathered from ubiquitous existing WiFi devices, enabling device-free human sensing [29] 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.



The principle of WiFi-based human sensing is that human activities essentially interfere with WiFi signals and lead to signal variations [62]. Such variations are recorded in CSI, which thereby contains implicit features for human sensing. Since these human features typically intertwine with excessive noise [61], 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) [67], Long Short-Term Memory (LSTM) [8], Convolutional Neural Networks (CNNs) [56, 68, 69], CNN-LSTM hybrids [33, 41, 76], *etc*. (2) For human localization, prevalent models have been adopted, including Naive Bayes [63], Auto-encoders [15], LSTM [9], and CNNs [58]. (3) WiFi-based human activity recognition (HAR) has drawn the greatest attention from researchers, applying models such as MLPs [74], LSTM [73], CNNs [42, 75, 78], CNN-LSTM hybrids [23,48,85], Generative Adversarial Networks (GANs) [55], attention-based bidirectional LSTM (ABLSTM) [7], and Transformers [34,81].

iFi-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, 25, 32, 59] 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



**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.

annotate samples for specific tasks (*e.g.*, activity recognition), restricting the use 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 multiuser 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 [67] 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 the advanced performance than MLPs, LSTM is inefficient for lengthy sequences due to its step-by-step inputs. To boost both effectiveness and efficiency, CNNs [56, 68, 69] 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 [33,41,76] have proposed various CNN-LSTM hybrids, attaining the best performance for WiFi-based human identification.

*WiFi-based human localization* estimates user locations to support humancomputer interaction systems [49]. Initially, Naive Bayes [63], 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 [58] have been further explored for localization, where CNN-1D [58] 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 [73] based on handcrafted feature extraction have demonstrated inadequate ability to extract implicit features from CSI. Thereby, LSTM [73] and CNNs [42, 75, 78] have been extensively adopted and offer valuable insights on learning temporal and spatial features. CNN-LSTM hybrids [23, 48, 85] further leverage the advantages of CNNs and LSTM to become predominant models for HAR. To tackle dynamic environments, GANs [55] have been used to augment CNNs with adversarial learning. Attention-based models [53] also contribute to HAR [81], such as ABLSTM [7] which equips bidirectional LSTM with attention layers to enhance HAR performance. Recently, a two-stream convolution augmented transformer (THAT) [34] 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 features of different users for effective sensing. Recent works [12,25,32,50,54,59] 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.* [73] collected a dataset in an office for WiFi-based HAR, after which various CSI datasets were released for HAR in different environments [4, 14, 22, 71]. 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 [39] for sign language recognition, FallDeFi [44] for fall detection, and ARIL [57] for activity recognition and localization. For human tracking, Widar [45] and Widar 2.0 [46] were gathered and annotated with user walking traces. Some datasets contain changing scenarios, such as CPAR [66] for different users and RF-NET [11] for varying environments. Similarly, CSIDA [27] and Widar 3.0 [77] incorporate CSI of various domains (*i.e.*, locations, orientations, environments, and users). OPERAnet [3], a multimodal dataset, includes WiFi CSI as well as data from Passive WiFi Radar (PWR), Ultra-Wideband (UWB), and Kinect skeleton sensors. MM-Fi [70] 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.* [67] 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.* [40] 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.

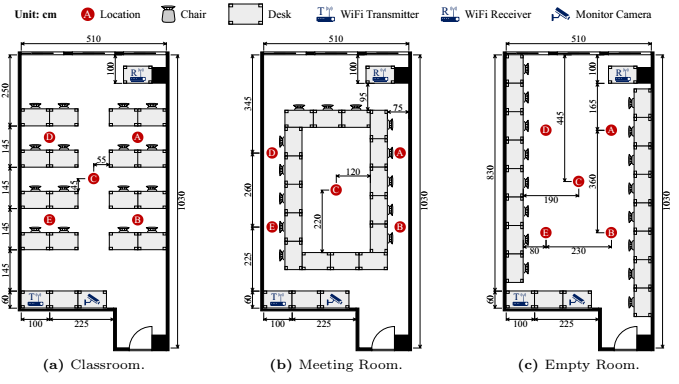
**3.1 Dataset Construction**

**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 [24] installed, following the previous works [3,73,77]. 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, 77], 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*×*3*×*30, and the dimension of each CSI sample is 3000*×*3*×*3*×*30. Along with each CSI sample, we also capture synchronized videos with a monitor camera for reference and unexplored tasks (*e.g.*, multiuser 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*×*3*×*1920*×*1080.

**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.

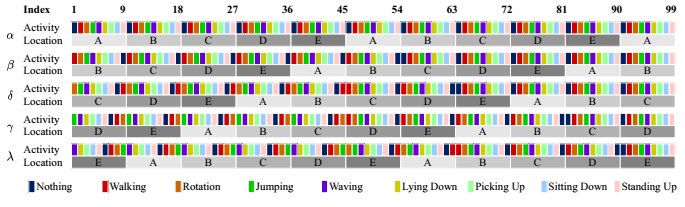
**Data Attributes.** WiMANS includes 9 activities which are representative in daily life, as discussed in the previous works [3,73,77]. These activities are: (1) Nothing, (2) Walking, (3) Rotation, (4) Jumping, (5) Waving, (6) Lying Down, (7) Picking Up, (8) Sitting Down, (9) Standing Up. To build up 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.



**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 {***α***, ***β***, ***δ***, ***γ***, ***λ***}, 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 ***α*** performs Picking Up, User 5 with ***λ*** performs Walking, and User 6 with ***β*** 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 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.



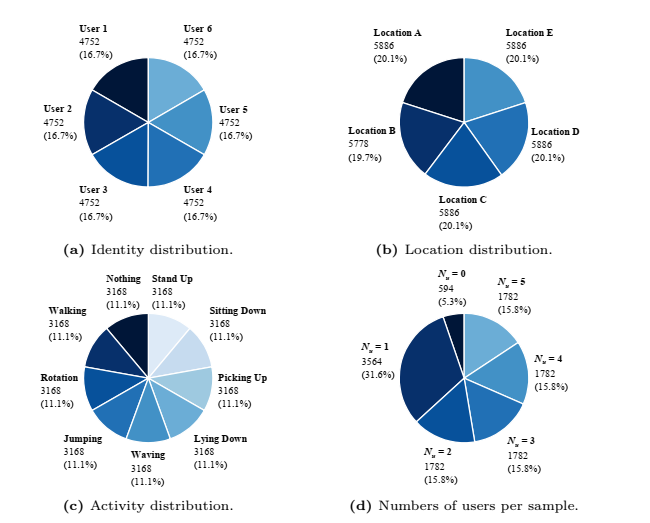
**Fig. 3:** Scripts that instruct users to perform identical/different activities at varying locations independently yet simultaneously.

**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 is more commonly used and more crowded than 5 GHz band, affected by more environmental noise [62,71]. Thus, most existing works [3,73,77] 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.



**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.

**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) [73], MLP [67], LSTM [73], CNN-1D [58], CNN-2D [42], CLSTM [41], ABLSTM [7], and THAT [34]. Particularly, CLSTM [41] is a hybrid of CNNs and LSTM, outperforming other models in WiFi-based identification. For WiFi-based localization, CNN-1D [58] has demonstrated the best performance by learning human features along the temporal dimension. THAT [34] equips two-stream Transformer encoders [53] 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 [52], S3D [64], MViT-v1 [13], MViT-v2 [36], Swin-T [38], and Swin-S [38]. 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 [7,34,73]. 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 [34,67].

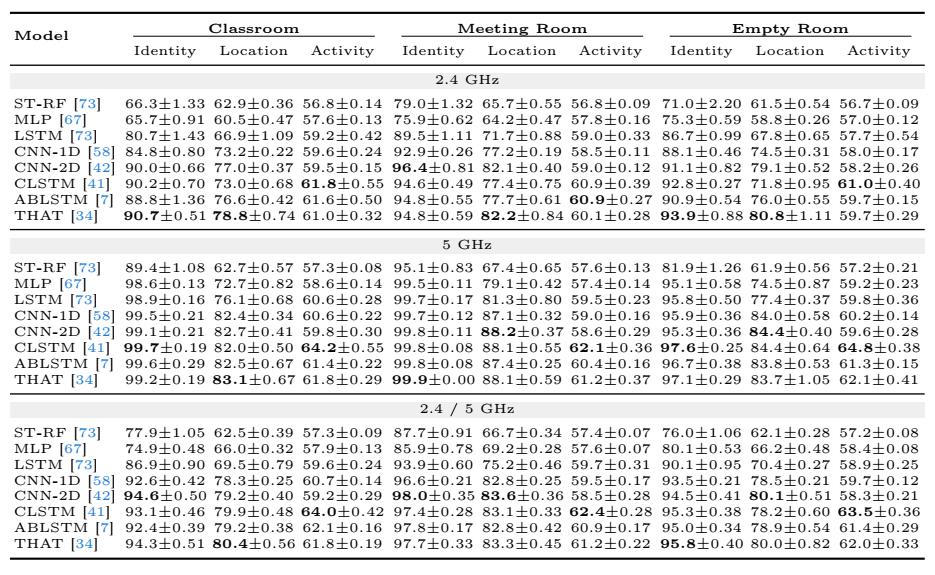
**4.3 Implementation Details**

**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 [7,34,73], 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*×*3*×*112*×*112 for ResNet [52]).

**Hyperparameters.** (1) For WiFi-based models, we leverage publicly accessible implementations as much as possible. Specifically, we initialize WiFi-based models with Xavier [19] 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.

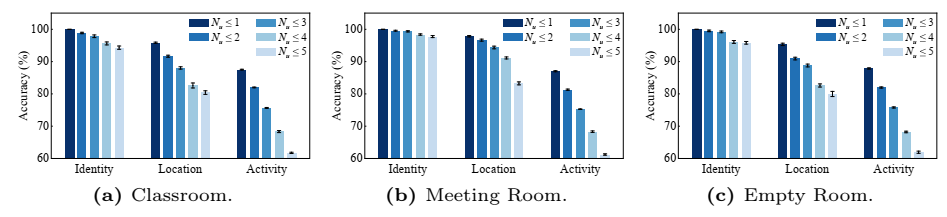
**Evaluation.** To recognize multiple labels, we connect each model to a linear layer followed by a sigmoid function for multi-label classification, using a fixed threshold of 0.5 to determine identities, locations, and activities. Both WiFibased and video-based models are optimized by Adam [31] on a 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.



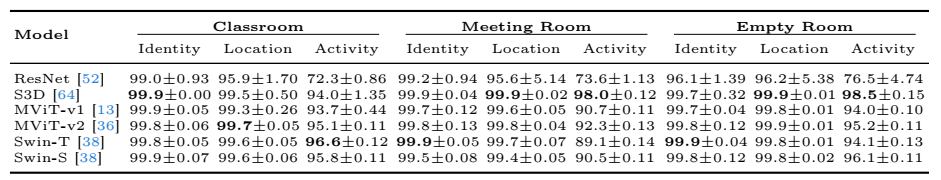
4.4 Results and Analysis

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.



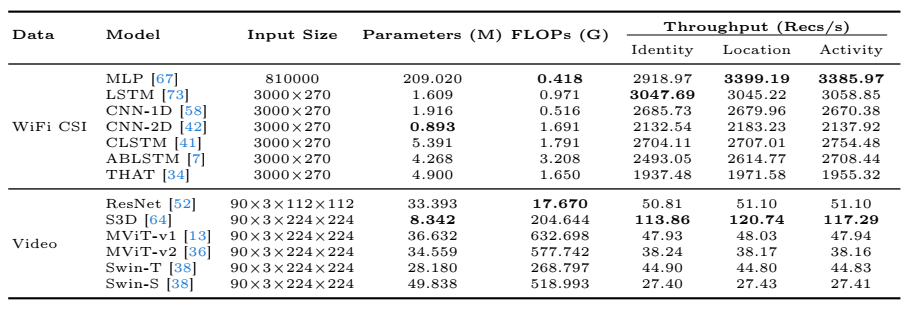
**Fig. 5:** Recognition performance of THAT [34] in terms of accuracy (%) regarding different numbers of users in each CSI sample (*Nu*). Sensing more users simultaneously results in lower performance. (“*Nu ≤ k*”: each sample includes 0 to *k* users.)

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



**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.)



**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 WiFibased 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 [3,73,77]. 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 [3,22,39,44,73,77]. 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 [65,70], the AX-CSI tool [20,40], *etc*.

1. **Genesis (2000):** The concept of indoor localization began to take shape around the turn of the millennium. Researchers recognized the limitations of GPS and other satellite-based technologies indoors, where signals could be blocked or distorted by walls, ceilings, and other structures. [The need for precise location tracking within multistory buildings, airports, and underground spaces drove the exploration of new solutions1](https://en.wikipedia.org/wiki/Indoor_positioning_system).
2. **Early Progress (2001-2010):** During this period, various techniques emerged:
   * **WiFi-Based Positioning System (WPS):** Leveraging existing WiFi infrastructure, WPS used signal strength measurements from WiFi access points to estimate a device’s location.
   * **Bluetooth Beacons:** Small Bluetooth devices (beacons) were strategically placed throughout indoor spaces to provide location context.
   * **Magnetic Positioning:** Some systems utilized Earth’s magnetic field for orientation and positioning.
   * **Dead Reckoning:** Devices estimated their position based on movement data (e.g., accelerometers and gyroscopes).
   * [**Optical and Acoustic Technologies:** Cameras, acoustic signals, and visual markers were explored for localization1](https://en.wikipedia.org/wiki/Indoor_positioning_system).
3. **Advancements (2010-2020):** Research and commercial efforts intensified:
   * **Ultra-Wideband (UWB):** UWB technology enabled high-precision ranging, achieving sub-meter accuracy.
   * **Fingerprinting:** Systems built databases of signal fingerprints (e.g., WiFi, Bluetooth) for specific locations.
   * **Sensor Fusion:** Combining data from multiple sensors (e.g., accelerometers, gyroscopes, magnetometers) improved accuracy.
   * **Machine Learning:** Algorithms learned patterns from sensor data to enhance localization.
   * **Beacon Networks:** Deployments of beacons expanded in retail, healthcare, and logistics.
   * [**Challenges:** Standardization remained elusive due to diverse building materials, spatial dimensions, and accuracy requirements1](https://en.wikipedia.org/wiki/Indoor_positioning_system).
4. **Recent Developments (2020-present):**
   * **5G and mmWave:** High-frequency 5G and millimeter-wave (mmWave) technologies promise better indoor positioning.
   * **Computer Vision:** Advances in image-based localization using cameras and visual landmarks.
   * **IoT Integration:** Smart buildings and IoT devices contribute to context-aware indoor positioning.
   * **Privacy Concerns:** Balancing accuracy with privacy protection remains a challenge.
   * [**Research and Industry Collaboration:** Efforts continue to refine algorithms, reduce error budgets, and create standards1](https://en.wikipedia.org/wiki/Indoor_positioning_system).

In summary, indoor localization has come a long way, evolving from early experiments to sophisticated systems that achieve centimeter-level accuracy. While challenges persist, the field continues to thrive, impacting industries such as retail, logistics, and emergency response. [For more detailed information, you can explore the NIST’s history of indoor localization research2](https://www.nist.gov/ctl/pscr/indoor-localization-nist). 🌐🔍