IMPERIAL

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   Julie A. McCann

Imperial College London

ECCV

EUROPEAN CONFERENCE ON COMPUTER VISION

MILANO
2024

IMPERIAL

AESE
Adaptive Emergent Systems Engineering

WiMANS, *the first dataset* for multi-user sensing based on WiFi. 9.4 hours of dual-band WiFi Channel State Information (CSI). Synchronized videos. Simultaneous activities of 0-5 users. The first benchmarks for WiFi-based multi-user activity sensing.

ECCV

EUROPEAN CONFERENCE ON COMPUTER VISION

MILANO
2024

Background

- WiFi-based Human Sensing
  - Non-intrusive
  - Environmental robust
  - Device-free
- Sensing Tasks and Models
  - Human identification: MLPs, LSTM, CNNs, CNN-LSTM hybrids
  - Human localization: Naive Bayes, Auto-encoders, LSTM, CNNs
  - Human activity recognition: MLPs, LSTM, CNNs, CNN-LSTM hybrids, GANs, attention-based bidirectional LSTM, Transformers

Drawbacks

- Single-user Limitation
  - A lack of public datasets that enable WiFi-based multi-user sensing
- Insufficient Modalities and Annotations
- Lack of Comprehensive Benchmarks

Challenges

- Comparison of Existing Datasets

| 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. [73]	1	557	7	90	1000	5	-	-	-	✓
SignFi [39]	1	14280	276	90	12.5~200	5	-	-	-	✓
FallDeFi [44]	1	1070	19	90	1000	5	-	-	-	✓
WiAR [22]	1	4161	16	90	30	5	-	✓	-	✓
ARIL [57]	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 [27]	1	2844	6	342	1000	5	-	✓	✓	✓
NTU-Fi [67]	1	2040	6	342	500	5	-	✓	-	✓
CPAR [66]	1	560	7	64	1000	2.4	-	✓	-	✓
Widar [45]	1	54	2	90	2000	5	-	-	✓	✓
Widar 2.0 [46]	1	24	2	90	1000	5	-	-	✓	✓
Widar 3.0 [77]	1	271050	22	90	1000	5	-	✓	✓	✓
Yang et al. [71]	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 [70]	1	1080	27	342	1000	5	✓	✓	-	✓
SHARPax [40]	0~1	108	3	242~996	133	5	-	-	-	✓
WiMANS (Ours)	0~5	11286	9	270	1000	2.4/5	✓	✓	✓	✓

WiMANS

- Each sample includes 0 to 5 users performing identical/different activities simultaneously

- 11286 CSI samples of dual WiFi bands (2.4 / 5 GHz)

- Synchronized video samples

- Annotated with (anonymized) user identities, locations, and activities

- The first benchmark for the multi-user sensing performance of WiFi-based models
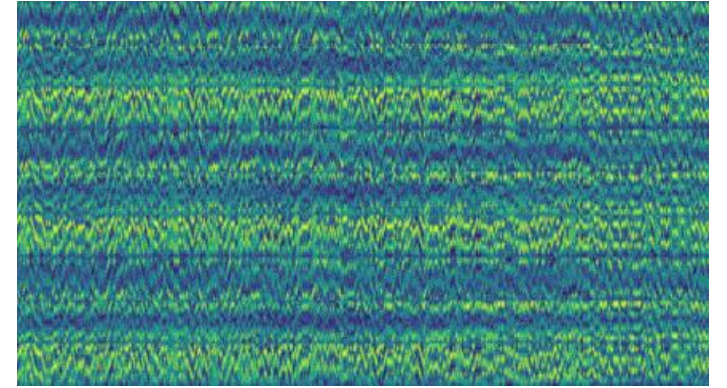
WiMANS
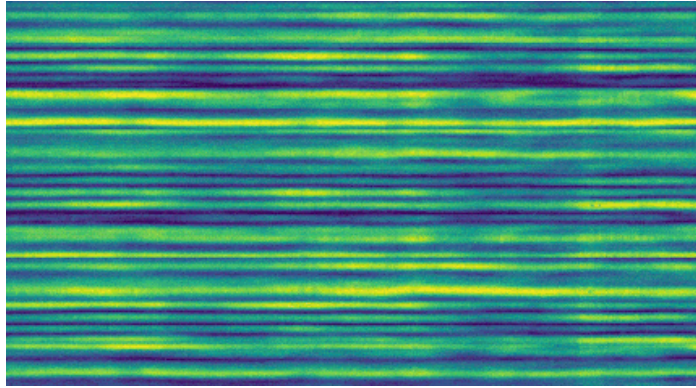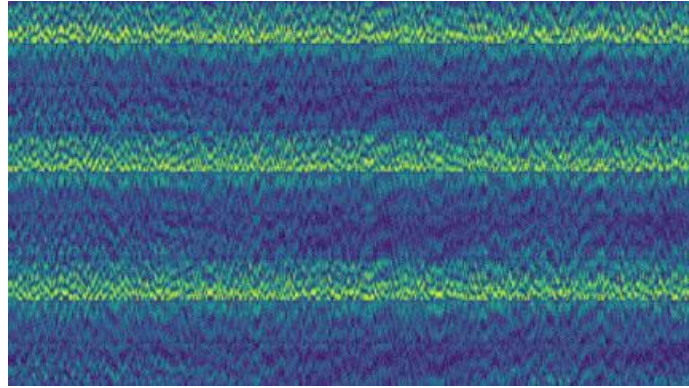
5 GHz   WiFi CSI   2.4 GHz

3 Users
Classroom

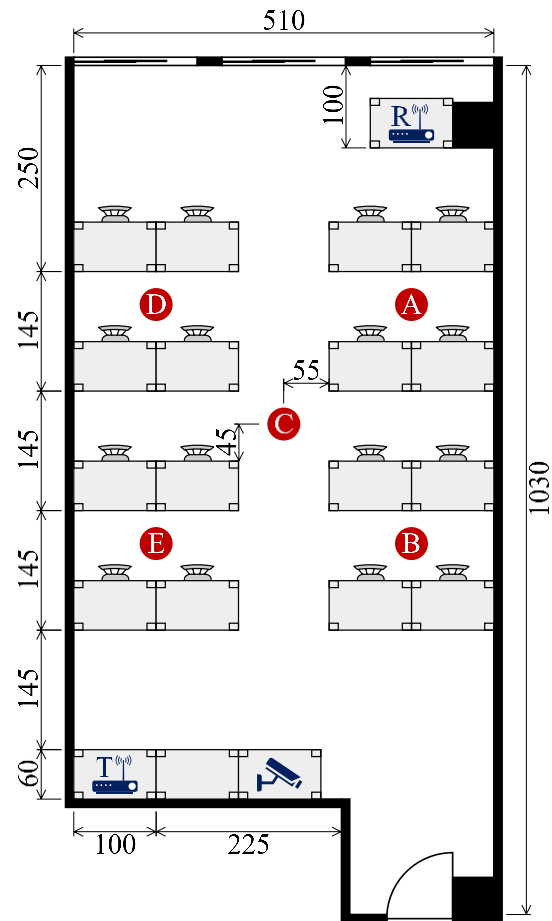4 Users
Meeting Room

5 Users
Empty Room

WiMANS

- Each CSI Sample
  - 3 seconds under a rate of 1000 Hz: 3000 time steps
  - A transmitter with 3 antennas
  - A receiver with 3 antennas
  - Each pair of antennas with 30 subcarriers
  - Dimension: 3000 x 3 x 3 x 30

- Each Video Sample
  - 3 seconds under a frame rate of 30 Hz: 90 time steps
  - RGB channels: 3
  - Frame resolution: 1920 x 1080
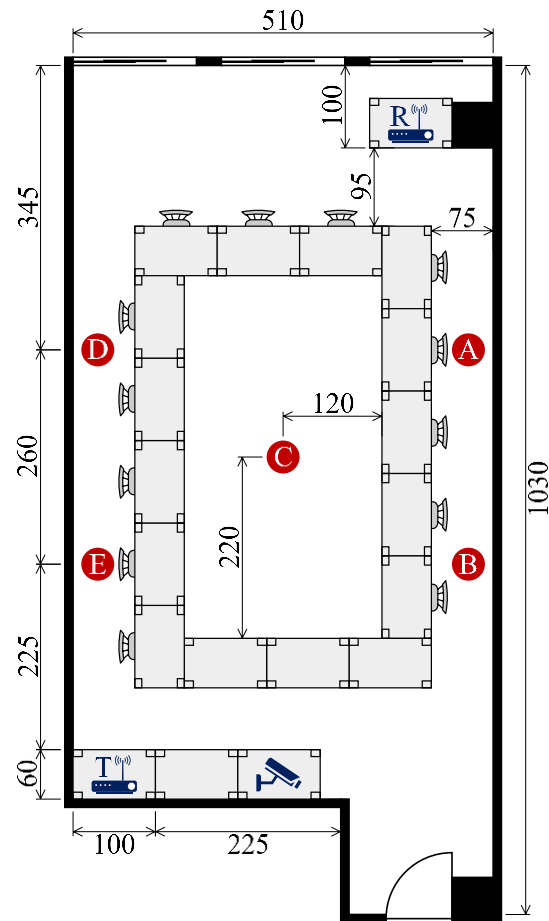 - Dimension: 90 x 3 x 1920 x 1080

WiMANS

- 9 Daily Activities
  - Nothing, Walking, Rotation, Jumping, Waving, Lying Down, Picking Up, Sitting Down, Standing Up
- 3 Daily Environments
  - Classroom, Meeting Room, Empty Room
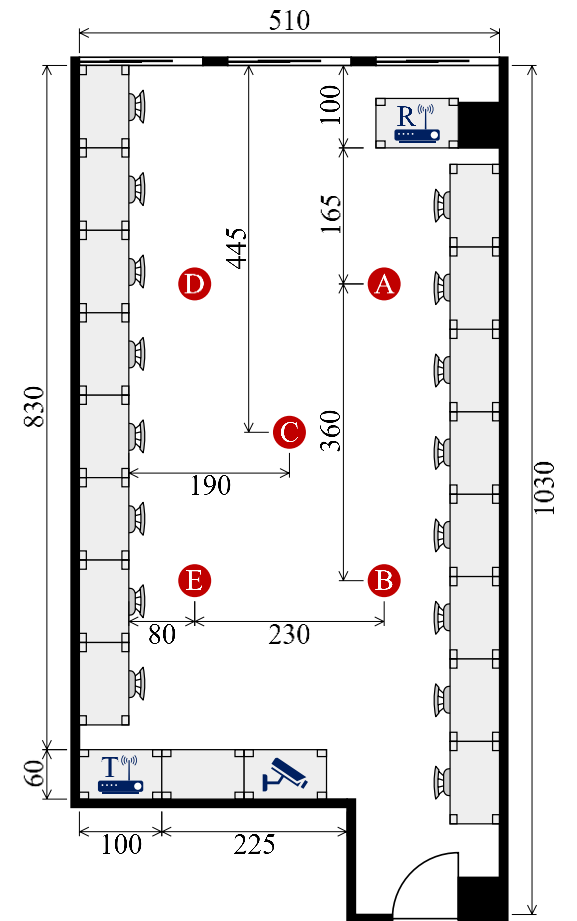
WiMANS

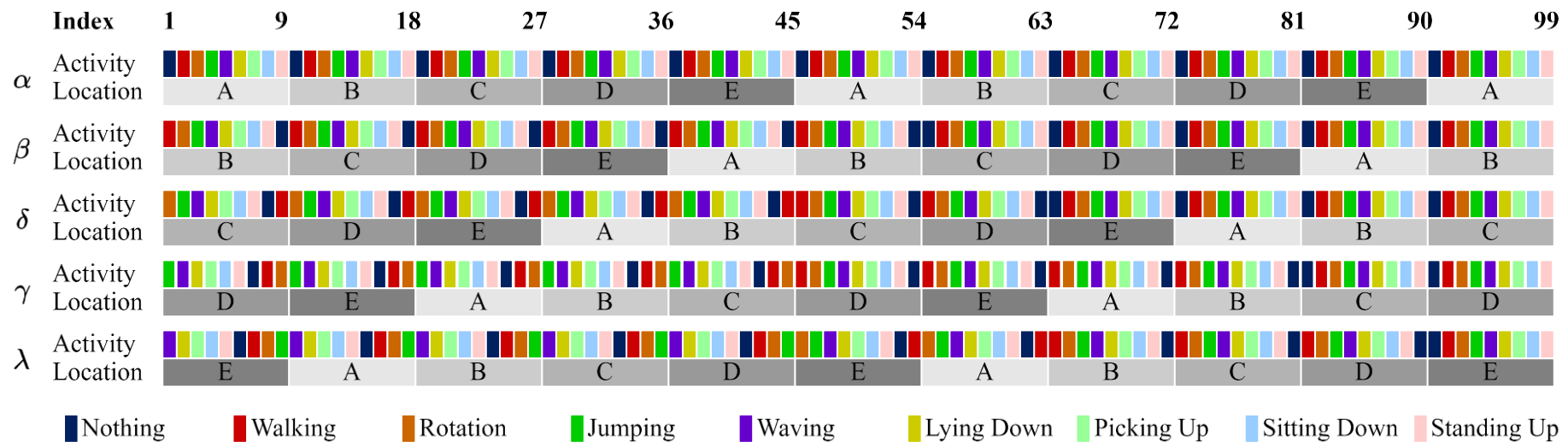Classroom   Meeting Room   Empty Room

WiMANS

- Simultaneous Scripts
  - Instruct users to perform identical/different activities at varying locations independently yet simultaneously
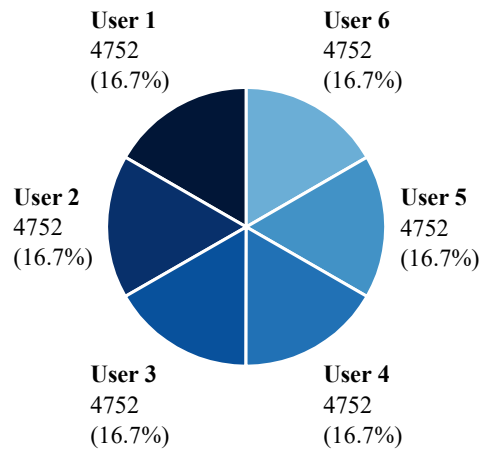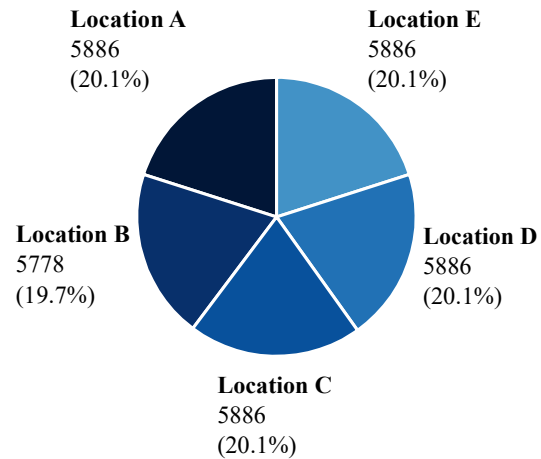
WiMANS

- Data Annotation
  - "act_<group>_<sample>"
  - Each group: a specific number of users and environment
  - Each label: environment, WiFi band, number of users, identities, locations, and activities
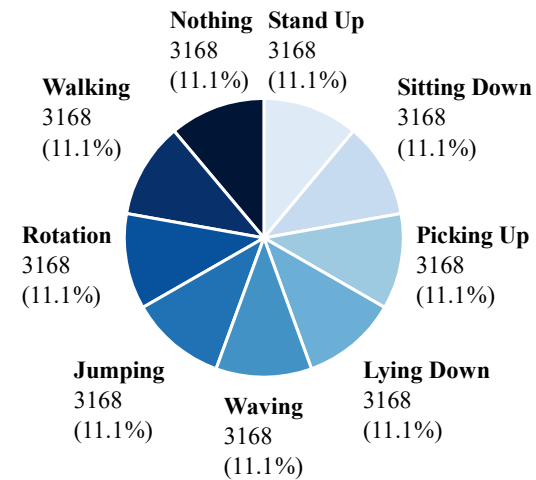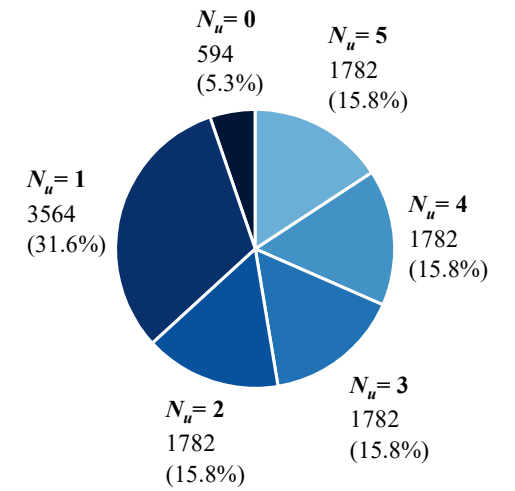
WiMANS

- Dataset Statistics

Identity Distribution

Location Distribution

Activity Distribution

Numbers of Users per Sample

Experiments

- WiFi-based Models
  - Short-time Fourier Transform Based Random Forest (ST-RF)
  - MLP
  - LSTM
  - CNN-1D
  - CNN-2D
  - CLSTM
  - ABLSTM
  - THAT

- Video-based Models
  - ResNet
  - S3D
  - MViT-v1
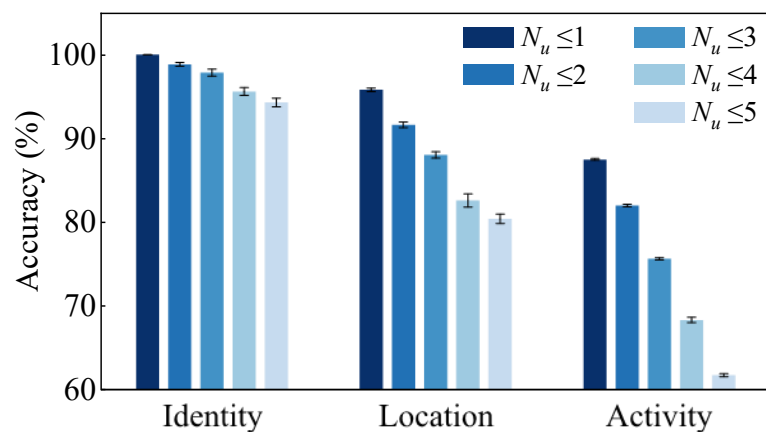 - MViT-v2
  - Swin-T
  - Swin-S

Experiments

- WiFi-based Models

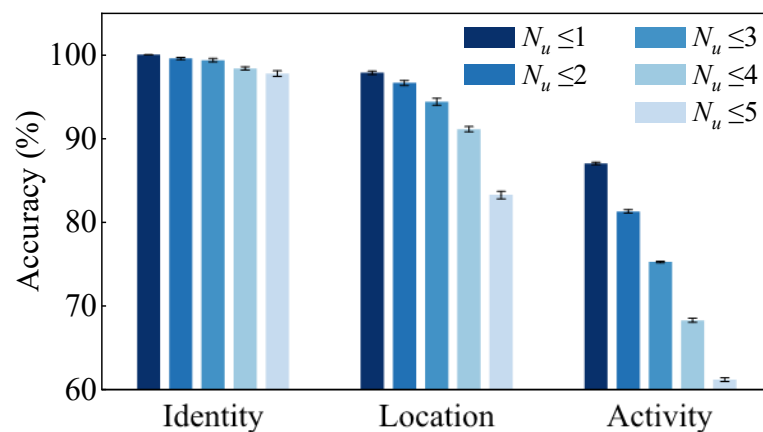| Model	Classroom			Meeting Room			Empty Room		
	Identity	Location	Activity	Identity	Location	Activity	Identity	Location	Activity
2.4 GHz									
ST-RF [73]	66.3±1.33	62.9±0.36	56.8±0.14	79.0±1.32	65.7±0.55	56.8±0.09	71.0±2.20	61.5±0.54	56.7±0.09
MLP [67]	65.7±0.91	60.5±0.47	57.6±0.13	75.9±0.62	64.2±0.47	57.8±0.16	75.3±0.59	58.8±0.26	57.0±0.12
LSTM [73]	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 [58]	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 [42]	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 [41]	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 [34]	**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 [73]	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 [67]	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 [73]	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 [58]	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 [42]	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 [41]	**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 [34]	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 [73]	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 [67]	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 [73]	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 [58]	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 [42]	**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 [41]	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 [34]	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

Experiments

- Number of Users

(a) Classroom   (b) Meeting Room   (c) Empty Room
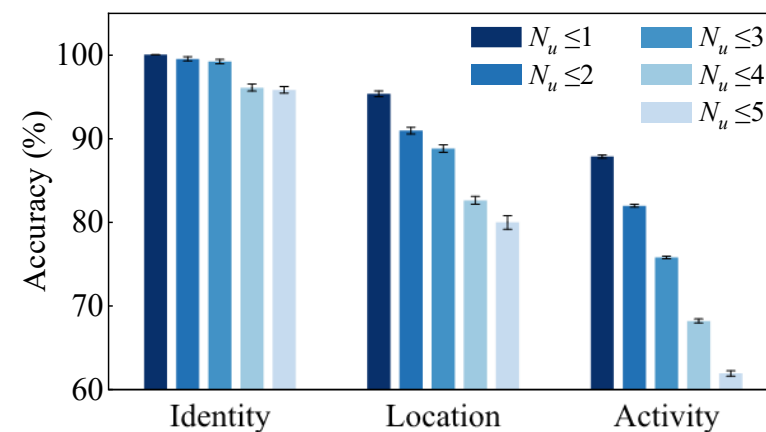
Experiments

- Video-based Models

Model	Classroom			Meeting Room			Empty Room		
	Identity	Location	Activity	Identity	Location	Activity	Identity	Location	Activity
ResNet [52]	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 [64]	**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 [36]	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 [38]	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 [38]	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

Experiments

- Model Complexity and Time Efficiency

Data	Model	Input Size	Parameters (M)	FLOPs (G)	Throughput (Recs/s)		
					Identity	Location	Activity
WiFi CSI	MLP [67]	810000	209.020	**0.418**	2918.97	**3399.19**	**3385.97**
	LSTM [73]	3000×270	1.609	0.971	**3047.69**	3045.22	3058.85
	CNN-1D [58]	3000×270	1.916	0.516	2685.73	2679.96	2670.38
	CNN-2D [42]	3000×270	**0.893**	1.691	2132.54	2183.23	2137.92
	CLSTM [41]	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 [34]	3000×270	4.900	1.650	1937.48	1971.58	1955.32
Video	ResNet [52]	90×3×112×112	33.393	**17.670**	50.81	51.10	51.10
	S3D [64]	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 [36]	90×3×224×224	34.559	577.742	38.24	38.17	38.16
	Swin-T [38]	90×3×224×224	28.180	268.797	44.90	44.80	44.83
	Swin-S [38]	90×3×224×224	49.838	518.993	27.40	27.43	27.41

Discussion

- Limitations
  - Daily activities
  - Challenging conditions
  - WiFi devices
- Future Work
  - Multi-user pose estimation
  - Dual-band augmented sensing
  - Cross-domain sensing

WiMANS

The First Dataset for WiFi-based Multi-user Activity Sensing

https://github.com/huangshk/WiMANS