

CSI-Bench: A Large-Scale In-the-Wild Dataset for Multi-task WiFi Sensing

Guozhen Zhu, Yuqian Hu, Weihang Gao, Wei-Hsiang Wang, Beibei Wang, K. J. Ray Liu
Origin Research

Presenter: Guozhen Zhu

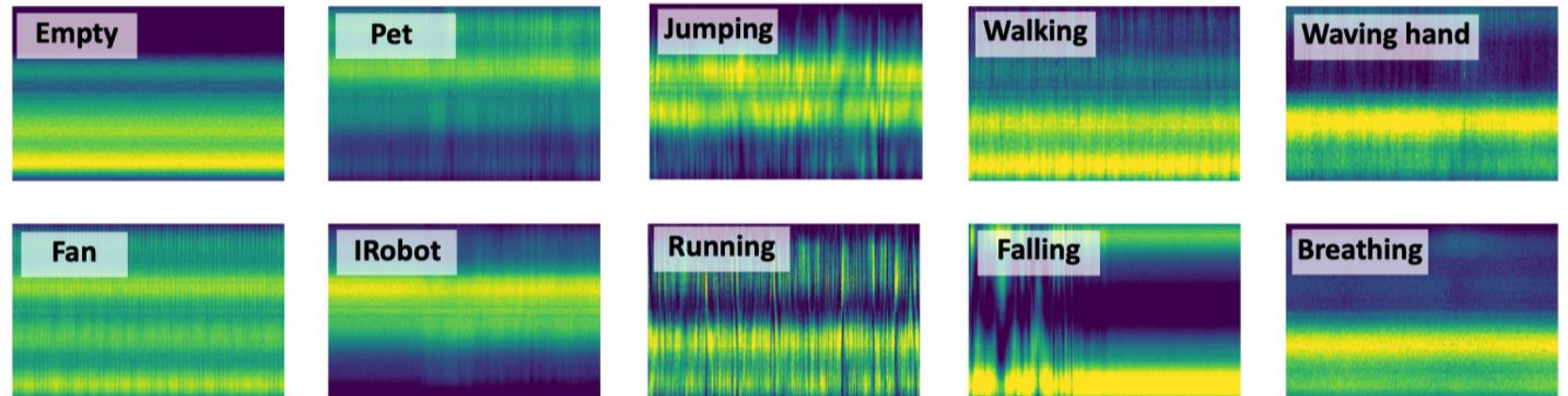
Email: gzzhu@terpmail.umd.edu

Project page:



Ubiquitous WiFi Sensing

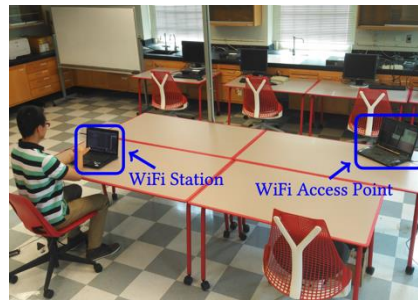
- Contactless
- Operates in the dark
- Works through walls
- Reuses existing infrastructure
- Privacy-preserving



Gap: Lab vs. Real-world

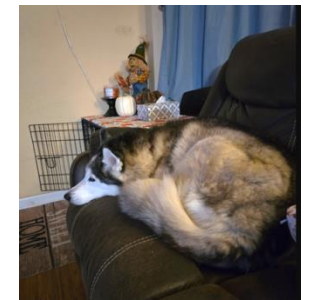
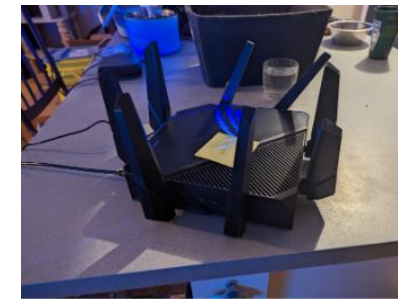
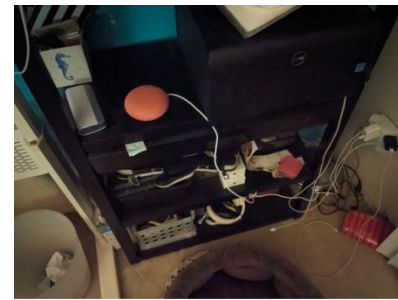
Lab

- Controlled environments
- Homogeneous hardware
- Scripted activities



Real-world

- Natural interference
- Diverse chipsets
- Continuous daily activities



Limitations of Existing Public Datasets

Dataset (Year)	Platform	# Edge Device Type	# Samples	# Tasks	# Envs	# Users	In-the-Wild
WiAR [14] (2019)	Intel 5300 NIC	1	4.8k	1	3	10	X
ARIL [35] (2019)	USRP	1	1.4k	2	1	1	X
Widar3.0 [47] (2021)	Intel 5300 NIC	1	271.1k	1	3	16	X
XRF5S [36] (2024)	Intel 5300 NIC	1	42.9k	1	4	39	X
SignFi [27] (2018)	Intel 5300 NIC	1	14.3k	1	2	5	X
WiMANs [21] (2024)	Intel 5300 NIC	1	11.3k	3	3	5	X
CSIDA [19] (2021)	Intel 5300 NIC	1	3k	1	2	5	X
MM-Fi [44] (2023)	Atheros CSI Tool	1	1.1k	1	4	40	X
CSI-Bench	Broadcom, Qualcomm, MediaTek, Espressif, NXP	16	231.6k	7	26	35	✓



Controlled environments



Fragmented recordings

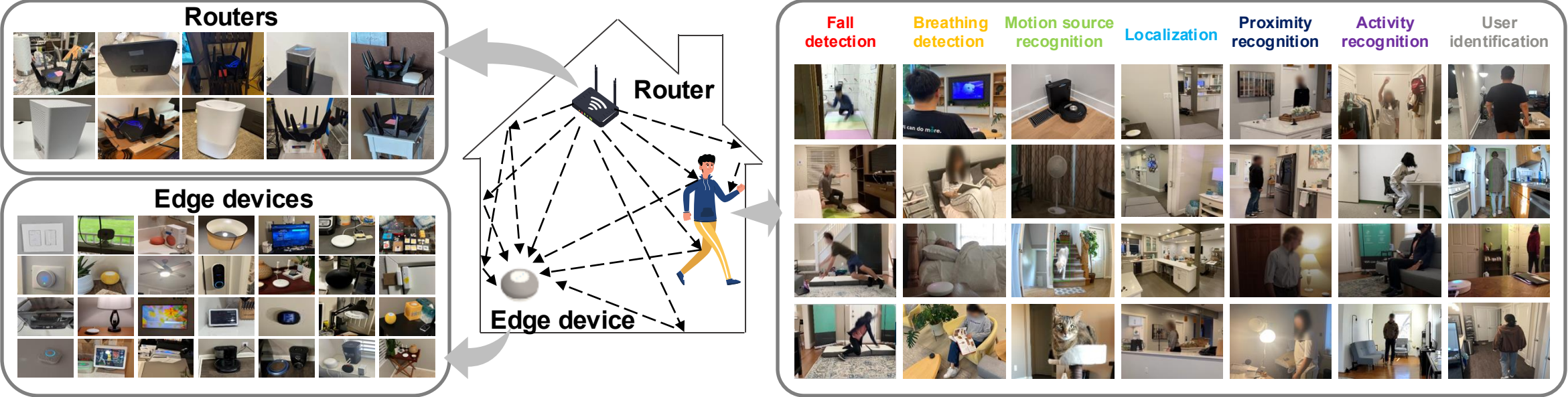


Task-specific focus



Low device diversity

CSI-Bench



Highlights

- **Scale:** 461 hours across 35 users, 16 device types, and 26 environments, with over 231k samples across 7 different tasks.
- **In-the-Wild:** Continuous CSI recorded in real users' homes
- **Multi-task:** Joint annotations enabling efficient multi-task learning
- **Standardization:** Rigorous evaluation protocols with difficulty tiers and OOD splits

Evaluation Setup

Single-task Specialized Datasets:

  Fall Detection

 Breathing

 Motion Source Recognition

 Localization

Multi-task Joint Datasets:

 Human Activity Recognition

 Human Identification

 Proximity Detection

Data Splits:

70 % Train | 15 % Val | 15 % Test

Difficulty Tiers:

Easy | Medium | Hard (Defined by signal quality and environment complexity)

OOD Evaluation:

Cross-Device | Cross-Environment | Cross-User

Metrics:

Accuracy & Weighted F1-Score

Baseline Evaluation

Strong baseline performance on in-distribution splits

Single-task

Model	Fall Detection		Breathing Detection		Room-Level Localization		Motion Source Recognition	
	Acc	F1	Acc	F1	Acc	F1	Acc	F1
MLP	92.16 ±0.91	92.17 ±0.92	97.59 ±0.08	97.59 ±0.08	87.14 ±0.80	86.90 ±0.83	98.86 ±0.07	98.86 ±0.07
ResNet-18	94.88 ±0.26	94.89 ±0.26	98.58 ±0.17	98.58 ±0.17	100.00 ±0.00	100.00 ±0.00	99.56 ±0.06	99.56 ±0.07
LSTM	94.93 ±0.51	94.92 ±0.50	98.62 ±0.16	98.62 ±0.16	99.12 ±0.27	99.12 ±0.26	98.42 ±0.19	98.42 ±0.19
Transformer	94.28 ±0.72	94.26 ±0.72	98.64 ±0.19	98.64 ±0.19	99.27 ±0.22	99.27 ±0.22	98.51 ±0.27	98.51 ±0.27
ViT	93.58 ±0.71	93.59 ±0.70	98.63 ±0.17	98.63 ±0.17	99.94 ±0.11	99.94 ±0.11	98.74 ±0.10	98.74 ±0.10
PatchTST	94.03 ±0.74	94.03 ±0.74	98.84 ±0.13	98.84 ±0.13	99.91 ±0.16	99.91 ±0.16	98.36 ±0.19	98.36 ±0.19
TimeSformer-1D	93.86 ±1.16	93.87 ±1.13	98.68 ±0.21	98.68 ±0.21	100.00 ±0.00	100.00 ±0.00	98.38 ±0.17	98.39 ±0.17

Multi-task learning

Task	Task-Specific Training		Multi-Task Joint Training		Improvement	
	Acc	F1	Acc	F1	ΔAcc	ΔF1
Human Activity Recognition	75.40 ±0.93	75.49 ±0.73	88.06 ±0.76	86.00 ±2.05	+12.66	+10.51
User Identification	99.51 ±0.32	99.51 ±0.32	99.55 ±0.06	99.70 ±0.27	+0.04	+0.19
Proximity Recognition	77.52 ±3.13	77.35 ±3.24	86.41 ±1.97	87.09 ±1.46	+8.89	+9.74

Multi-task learning improves accuracy while reducing model size by 60%

Out of Distribution Evaluation

Task	Cross-Device		Cross-Env		Cross-User	
	Acc	F1	Acc	F1	Acc	F1
Human Activity Recognition	61.82 $\pm$ 0.95	57.80 $\pm$ 0.78	54.92 $\pm$ 0.98	47.17 $\pm$ 1.12	54.72 $\pm$ 0.84	46.67 $\pm$ 1.00
Human Identification	59.94 $\pm$ 0.77	59.81 $\pm$ 0.96	/	/	/	/
Proximity Recognition	30.68 $\pm$ 3.11	28.76 $\pm$ 3.51	29.67 $\pm$ 1.63	27.12 $\pm$ 1.79	30.26 $\pm$ 1.94	25.97 $\pm$ 2.26

Significant performance drops reveal generalization challenges
Highlights need for domain adaptation techniques

Access Dataset, Code, Paper, and Documentation



<https://ai-iot-sensing.github.io/projects/project.html>
gzzhu@terpmail.umd.edu