MEMORY-EFFICIENT HIGH-ACCURACY FOOD INTAKE ACTIVITY RECOGNITION WITH 3D MMWAVE RADARS

Hsin-Che Chiang[1], Yi-Hung Wu[1], Shervin Shirmohammadi[2], and Cheng-Hsin Hsu[1]

[1] Department of Computer Science, National Tsing-Hua University, Taiwan

[2] School of Electrical Engineering and Computer Science, University of Ottawa, Canada

MADiMa 2023

OVERVIEW

Task:

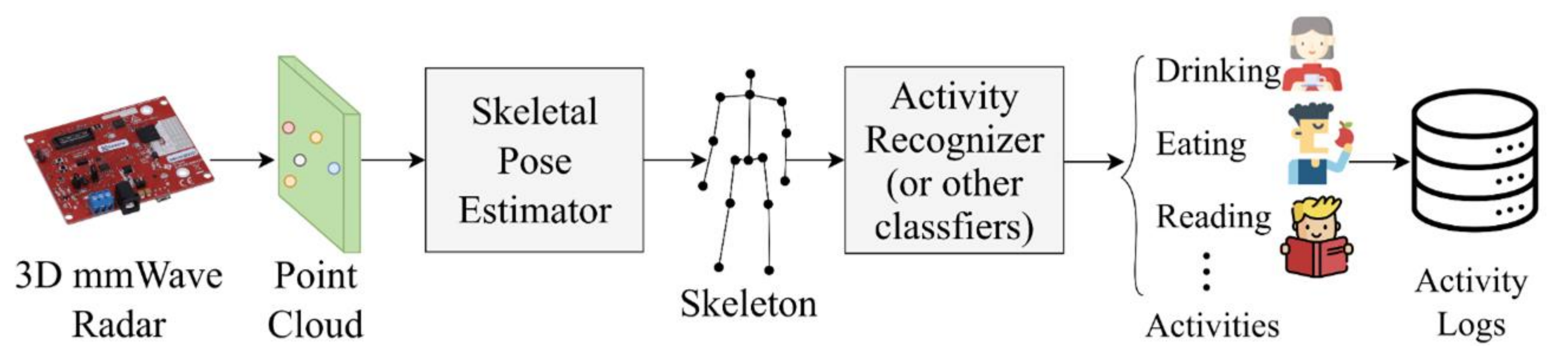
- Food intake activity recognition using a 3D mmWave radar

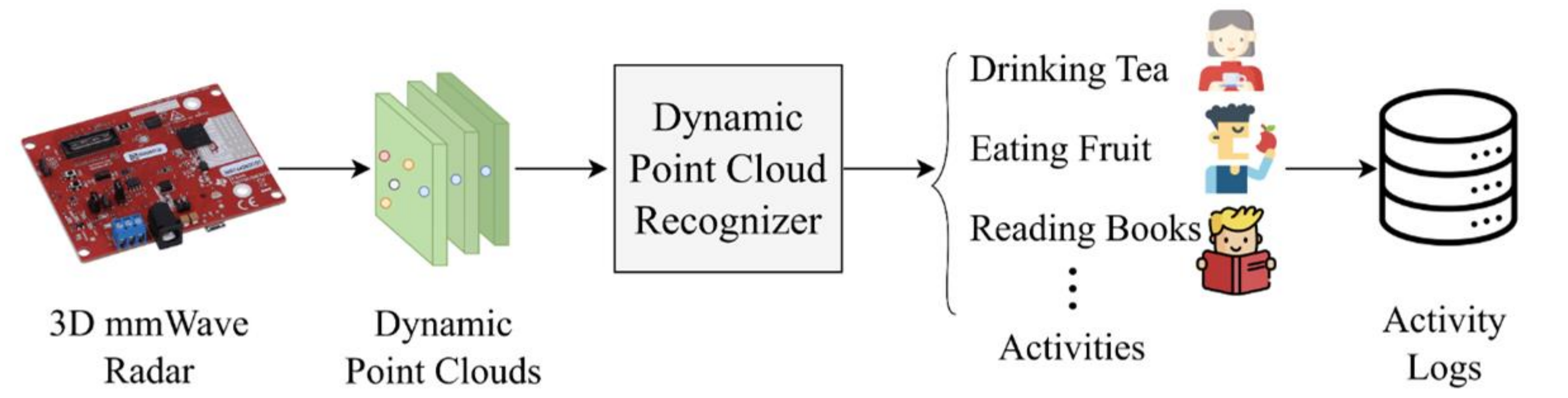
Two pipelines:

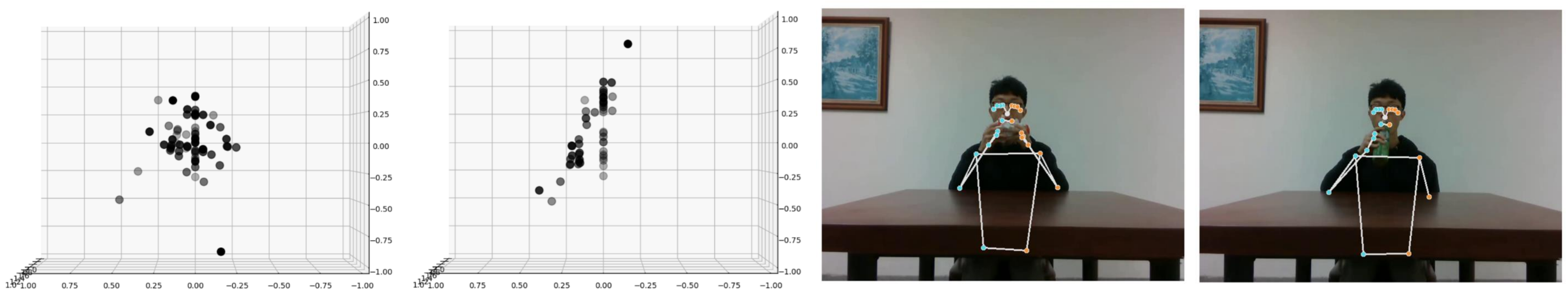
- Skeleton-based activity recognition
- End-to-end activity recognition

Three algorithms:

- **Skeletal Pose Estimator (SPE)**
- **Dynamic Point Cloud Recognizer (DPR)**
- **Lightweight Dynamic Point Cloud Recognizer (LDPR)**



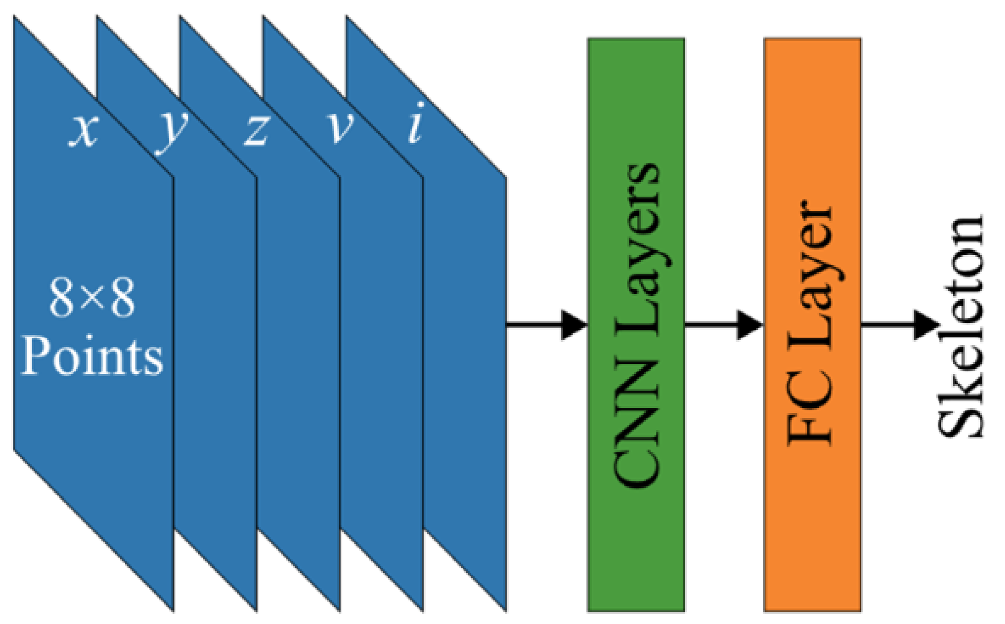




3D MMWAVE RADAR DATASET[1]

- Sensors: An RGB-D camera and a mmWave radar (RGB-D for evaluations only, not needed at runtime)
- Subjects: 24 (12 male, 12 female)
- Activities: 12 (3 eating, 3 drinking, and 6 other)

Skeletal Pose Estimator (SPE)



Motivation

- MARS[2]'s 2-layer CNN → higher error
- Too many layers → overfitting

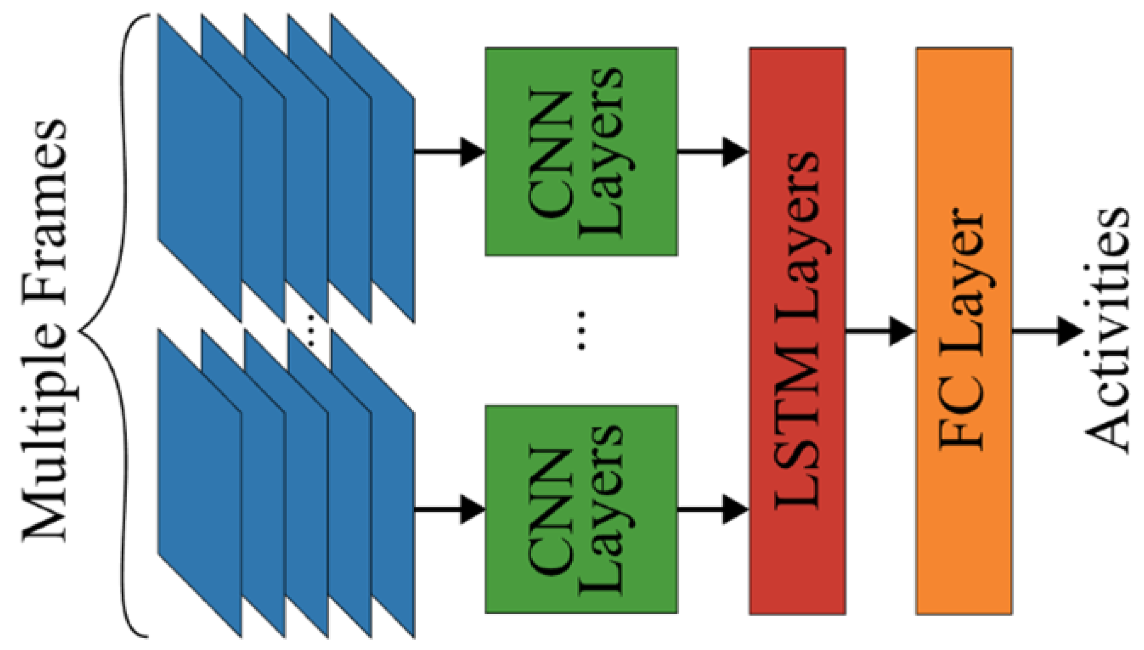
Proposed Solutions

- Test popular models (AlexNet, GoogLeNet, ResNet)
- Carefully decide the model depths

Evaluations

- ResNet-34 has the highest accuracy (45.16% reduction over MARS)
- ResNet-34 outperforms its 18- and 50-layer counterparts

Dynamic Point Cloud Recognizer (DPR)



Motivation

- FIA[3] struggles to strike a balance between voxel size and memory consumption

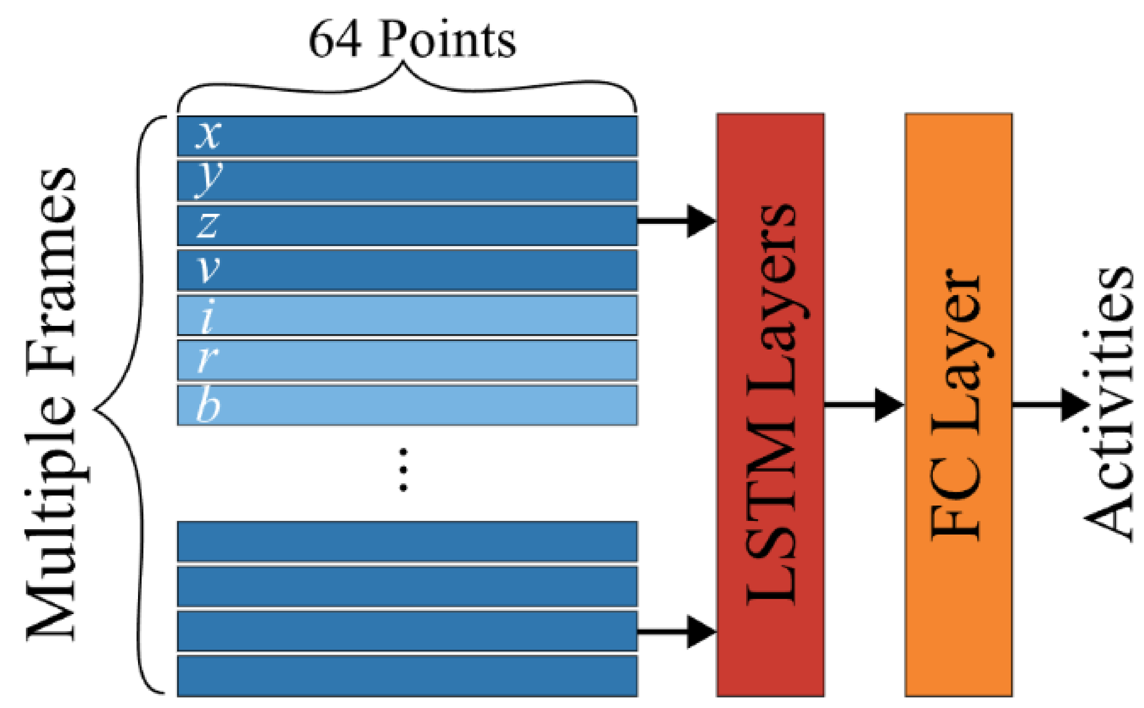
Proposed Solutions

- SPE's 8x8 grid + CNN → spatial features
- LSTM → temporal features

Evaluations

- Accuracy = 99.66% (4.10% improvement over FIA)
- Memory = 2131 MiB (78.29% reduction over FIA)

Lightweight DPR (LDPR)



Motivation

- Combining CNN and LSTM could be computationally heavy

Proposed Solutions

- Feature vectors → LSTM
- Eliminate the CNNs

Evaluations – 80-20 train-test split

- Accuracy = 99.81% (0.15% improvement over DPR)
- Memory = 1219 MiB (42.8% reduction over DPR)

Evaluations – Leave-one-out

- Accuracy = 40.77~86.97% (average 72.74%)
- Mimic the situations of new subjects

KEY FINDINGS

- mmWave point clouds can be used for fine-grained activity recognition, offering privacy advantages
- Model structures and depths need to be chosen properly to suit different tasks (e.g., rehabilitation vs. food intake)
- MARS's preprocessing can also be used for activity recognition without skeletons
- Using feature vectors without CNNs reduces overhead

ONGOING TASKS

- Further improving the precision of the estimated skeletons
- Enhancing model generalization to recognize activities from unseen subjects
- Adapting the proposed solutions for other applications like driver monitoring (for safety)

ANY QUESTIONS?

[1] Y. Wu, H. Chiang, S. Shirmohammadi, and C. Hsu. 2023. A Dataset of Food Intake Activities Using Sensors with Heterogeneous Privacy Sensitivity Levels. In Proc. of the ACM MMSys '23. 416–422.

[2] S. An and U. Ogras. 2021. MARS: mmWave-based Assistive Rehabilitation System for Smart Healthcare. ACM Transactions on Embedded Computing Systems 20, 5s, Article 72, 1–22.

[3] Y. Wu, Y. Chen, S. Shirmohammadi, and C. Hsu. 2022. AI-Assisted Food Intake Activity Recognition Using 3D mmWave Radars. In Proc. of the ACM MADiMa '22. 81–89.