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

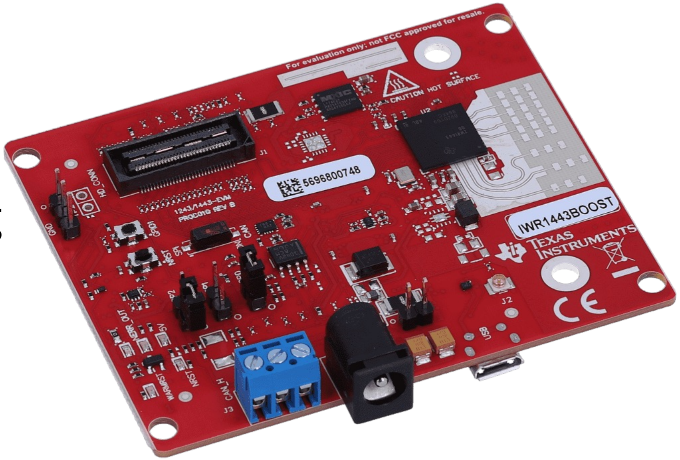
MADiMa 2023

FOOD INTAKE ACTIVITY RECOGNITION

- Motivation:
  - Rise of **lifestyle-related diseases**
  - → Monitoring food intake gains importance
- Problems with food logging/food diaries:
  - Error-prone, inconvenient, and time-consuming
  - → Need for an **automated** recognition system

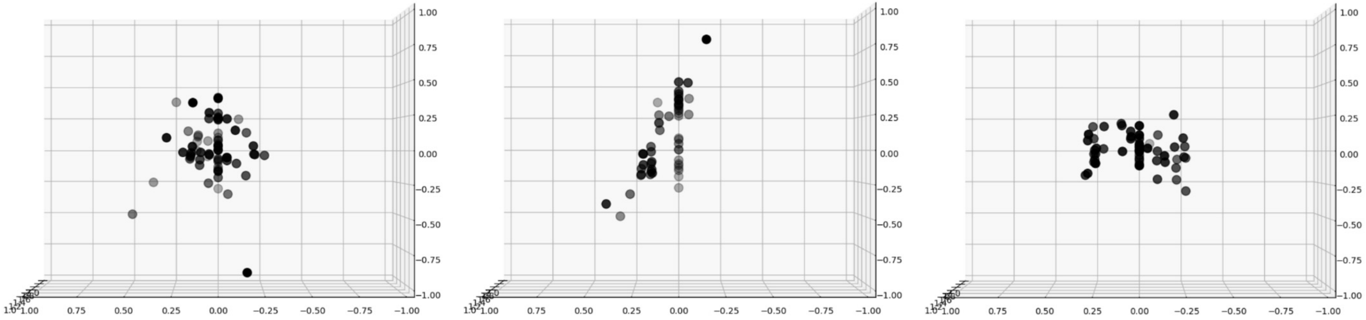
FOOD INTAKE ACTIVITY RECOGNITION

- Motivation:
  - Rise of **lifestyle-related diseases**
  - → Monitoring food intake gains importance
- Problems with food logging/food diaries:
  - Error-prone, inconvenient, and time-consuming
  - → Need for an **automated** recognition system
- Problems with RGB cameras:
  - Privacy concerns drive users away
  - Vulnerable to poor/fluctuating lighting
  - → Explore alternative sensors, e.g., the **mmWave radar**



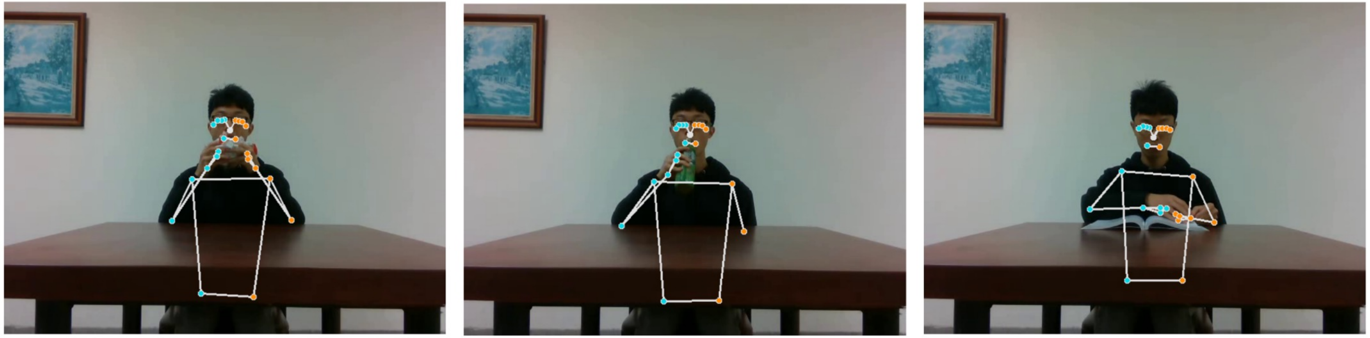
MMWAVE POINT CLOUD DATASET [MMSys '23]

Sparse Dynamic
3D Point Cloud



(a) (c) (e)

RGB videos
(w/ skeletons)

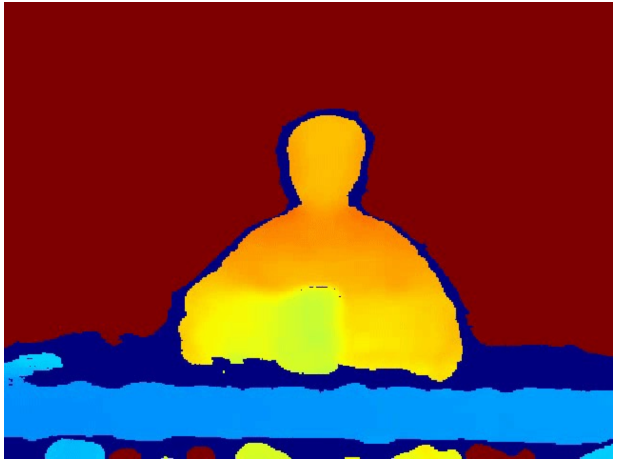


(b) (d) (f)

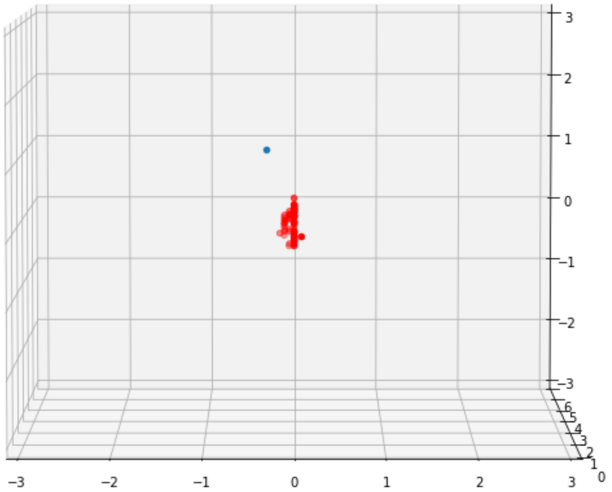
[MMSys '23] Y. Wu, H. Chiang, S. Shirmohammadi, and C. Hsu. 2023.
A Dataset of Food Intake Activities Using Sensors with Heterogeneous Privacy Sensitivity Levels. In Proc. of the ACM MMSys '23. 416–422.

SAMPLE SENSOR DATA

RGB Video



Depth Video

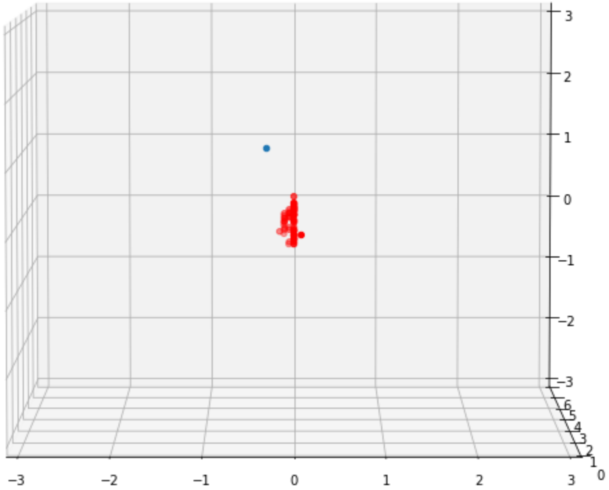


Sparse Dynamic
3D Point Cloud

SAMPLE SENSOR DATA

RGB Video

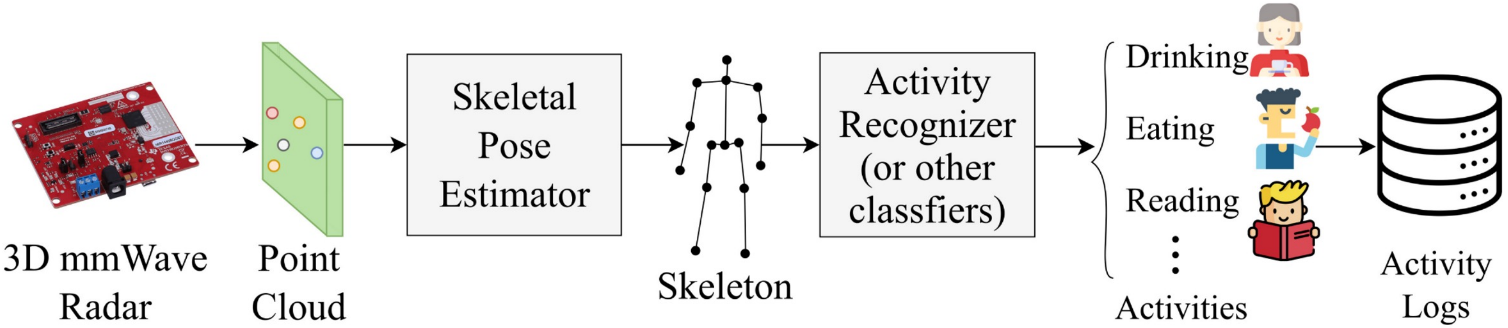
Depth Video



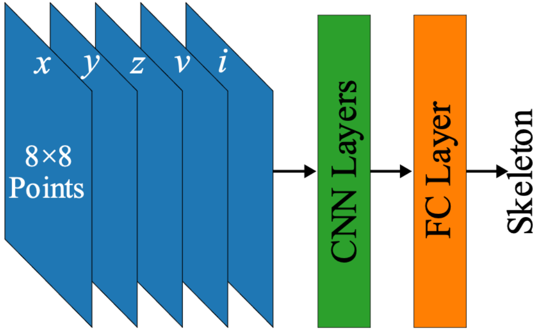
Sparse Dynamic
3D Point Cloud

SKELETAL POSE ESTIMATION

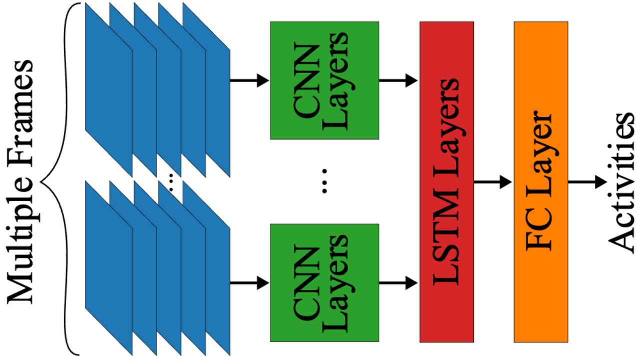
- Challenges:
    - Food intake activities involve **smaller movements**
    - Differentiating them requires higher precision
- Solution:
    - Skeletal Pose Estimation (a.k.a. Human Joint Estimation)
    - **Leverage the human body structure** to capture subtle nuances



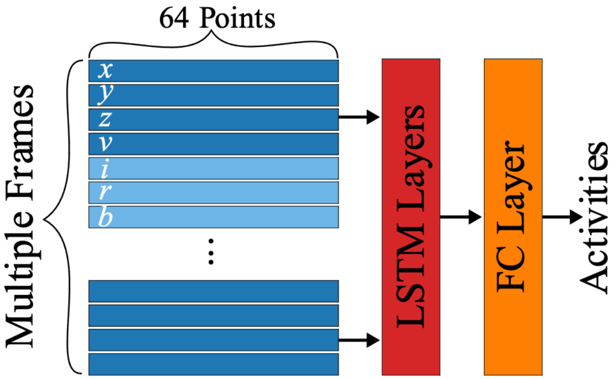
PROPOSED SOLUTIONS



① **Skeletal Pose Estimator (SPE)**

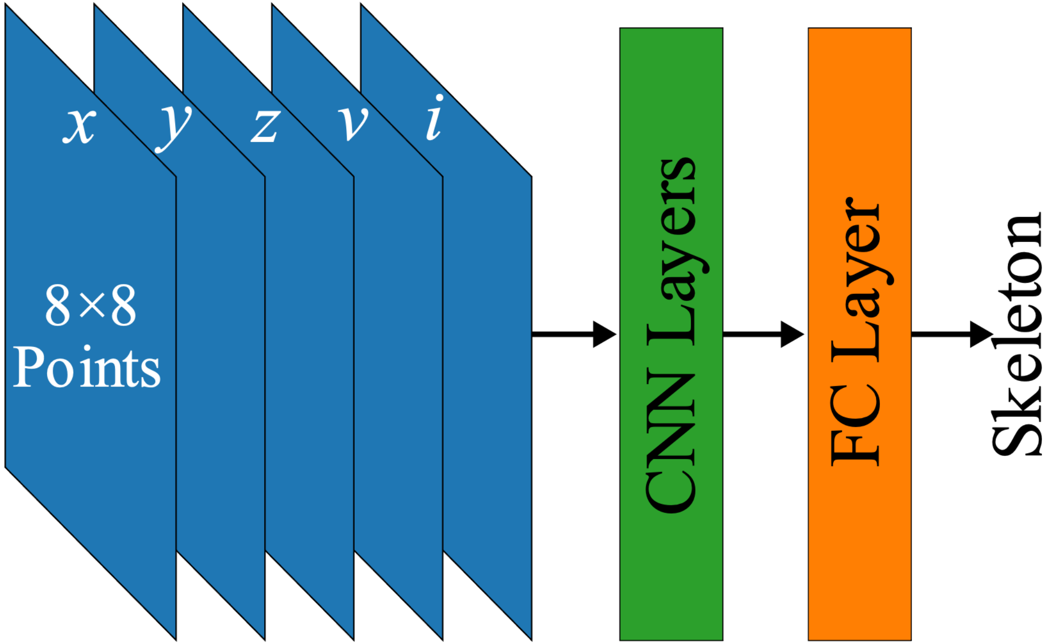


② **Dynamic Point Cloud Recognizer (DPR)**



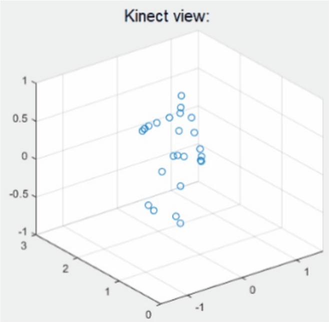
③ **Lightweight Dynamic Point Cloud Recognizer (LDPR)**

PROPOSED SOLUTIONS

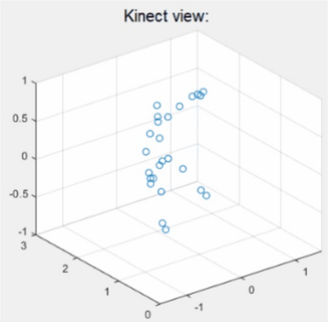


① **Skeletal Pose Estimator (SPE)**

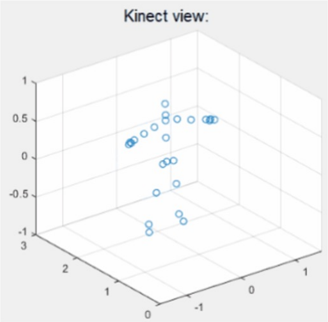
MARS [1] mmWave-Based Assistive Rehabilitation System for Smart Healthcare



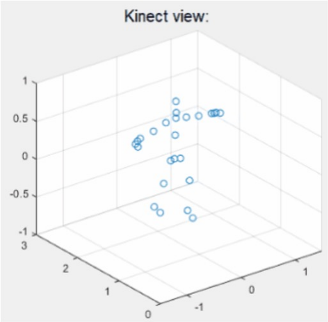
10) Right limb extension



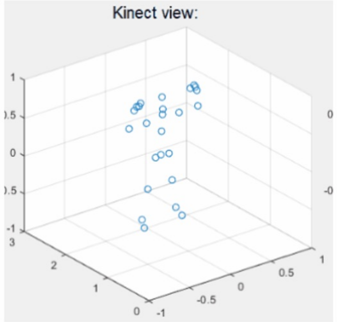
9) Left limb extension



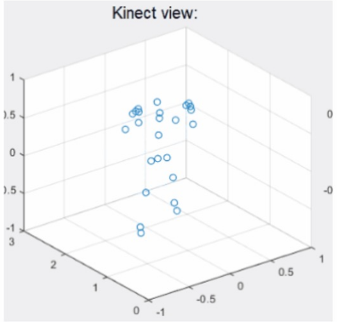
8) Right side lunge



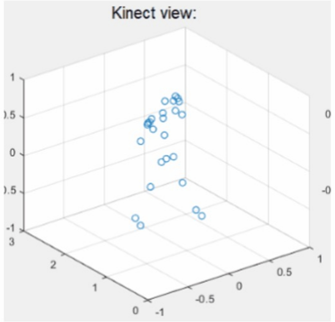
7) Left side lunge



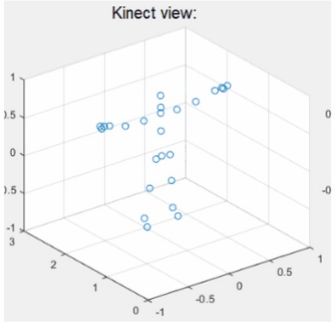
6) Squad



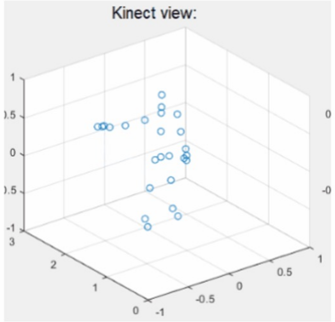
5) Right front lunge



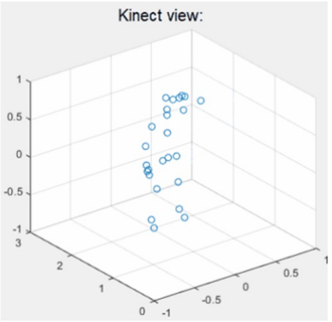
4) Left front lunge



3) Both upper limb extension



2) Right upper limb extension



1) Left upper limb extension

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

ERRORS OF THE ESTIMATED JOINTS IN SPE①

	MARS	AlexNet	GoogLeNet	ResNet-18	ResNet-34	ResNet-50
Nose	8.28 (±0.17)	6.91 (±0.06)	4.73 (±0.05)	4.41 (±0.05)	**4.22 (±0.05)**	4.75 (±0.08)
L. Shldr	6.66 (±0.10)	5.65 (±0.05)	3.91 (±0.04)	3.74 (±0.04)	**3.58 (±0.04)**	3.93 (±0.05)
R. Shldr	6.58 (±0.11)	5.54 (±0.05)	3.86 (±0.04)	3.67 (±0.04)	**3.52 (±0.04)**	3.91 (±0.06)
L. Elbow	8.68 (±0.15)	7.13 (±0.05)	4.84 (±0.04)	4.59 (±0.04)	**4.39 (±0.04)**	4.83 (±0.05)
R. Elbow	8.38 (±0.16)	7.05 (±0.05)	4.97 (±0.04)	4.76 (±0.04)	**4.57 (±0.04)**	4.94 (±0.04)
L. Wrist	11.52 (±0.13)	9.60 (±0.06)	6.53 (±0.05)	6.36 (±0.05)	**6.06 (±0.05)**	6.55 (±0.06)
R. Wrist	11.78 (±0.10)	10.19 (±0.06)	7.29 (±0.05)	7.15 (±0.05)	**6.85 (±0.05)**	7.33 (±0.06)
L. Pinky	12.51 (±0.13)	10.44 (±0.07)	7.14 (±0.05)	6.96 (±0.05)	**6.62 (±0.05)**	7.13 (±0.06)
R. Pinky	13.27 (±0.15)	11.47 (±0.07)	8.24 (±0.06)	8.08 (±0.06)	**7.74 (±0.06)**	8.28 (±0.06)
L. Index	12.54 (±0.13)	10.50 (±0.07)	7.20 (±0.06)	7.00 (±0.05)	**6.67 (±0.05)**	7.21 (±0.06)
R. Index	13.16 (±0.11)	11.41 (±0.07)	8.23 (±0.06)	8.06 (±0.06)	**7.73 (±0.06)**	8.26 (±0.06)
L. Thumb	11.63 (±0.13)	9.70 (±0.06)	6.62 (±0.05)	6.46 (±0.05)	**6.15 (±0.05)**	6.63 (±0.05)
R. Thumb	11.96 (±0.09)	10.36 (±0.06)	7.42 (±0.05)	7.29 (±0.05)	**6.99 (±0.05)**	7.47 (±0.06)
Average	10.54 (±1.33)	8.92 (±1.16)	6.23 (±0.85)	6.04 (±0.86)	**5.78 (±0.83)**	6.25 (±0.85)

ERRORS OF THE ESTIMATED JOINTS IN SPE ①

	MARS	AlexNet	GoogLeNet	ResNet-18	ResNet-34	ResNet-50
Nose	8.28 (±0.17)	6.91 (±0.06)	4.73 (±0.05)	4.41 (±0.05)	**4.22 (±0.05)**	4.75 (±0.08)
L. Shldr	6.66 (±0.10)	5.65 (±0.05)	3.91 (±0.04)	3.74 (±0.04)	**3.58 (±0.04)**	3.93 (±0.05)
R. Shldr	6.58 (±0.11)	5.54 (±0.05)	3.86 (±0.04)	3.67 (±0.04)	**3.52 (±0.04)**	3.91 (±0.06)
L. Elbow	8.68 (±0.15)	7.13 (±0.05)	4.84 (±0.04)	4.59 (±0.04)	**4.39 (±0.04)**	4.83 (±0.05)
R. Elbow	8.38 (±0.16)	7.05 (±0.05)	4.97 (±0.04)	4.76 (±0.04)	**4.57 (±0.04)**	4.94 (±0.04)
L. Wrist	11.52 (±0.13)	9.60 (±0.06)	6.53 (±0.05)	6.36 (±0.05)	**6.06 (±0.05)**	6.55 (±0.06)
R. Wrist	11.78 (±0.10)	10.19 (±0.06)	7.29 (±0.05)	7.15 (±0.05)	**6.85 (±0.05)**	7.33 (±0.06)
L. Pinky	12.51 (±0.13)	10.44 (±0.07)	7.14 (±0.05)	6.96 (±0.05)	**6.62 (±0.05)**	7.13 (±0.06)
R. Pinky	13.27 (±0.15)	11.47 (±0.07)	8.24 (±0.06)	8.08 (±0.06)	**7.74 (±0.06)**	8.28 (±0.06)
L. Index	12.54 (±0.13)	10.50 (±0.07)	7.20 (±0.06)	7.00 (±0.05)	**6.67 (±0.05)**	7.21 (±0.06)
R. Index	13.16 (±0.11)	11.41 (±0.07)	8.23 (±0.06)	8.06 (±0.06)	**7.73 (±0.06)**	8.26 (±0.06)
L. Thumb	11.63 (±0.13)	9.70 (±0.06)	6.62 (±0.05)	6.46 (±0.05)	**6.15 (±0.05)**	6.63 (±0.05)
R. Thumb	11.96 (±0.09)	10.36 (±0.06)	7.42 (±0.05)	7.29 (±0.05)	**6.99 (±0.05)**	7.47 (±0.06)
Average	$\boxed{10.54\ (\pm 1.33)}$	8.92 (±1.16)	6.23 (±0.85)	6.04 (±0.86)	**5.78 (±0.83)**	6.25 (±0.85)

ERRORS OF THE ESTIMATED JOINTS IN SPE①

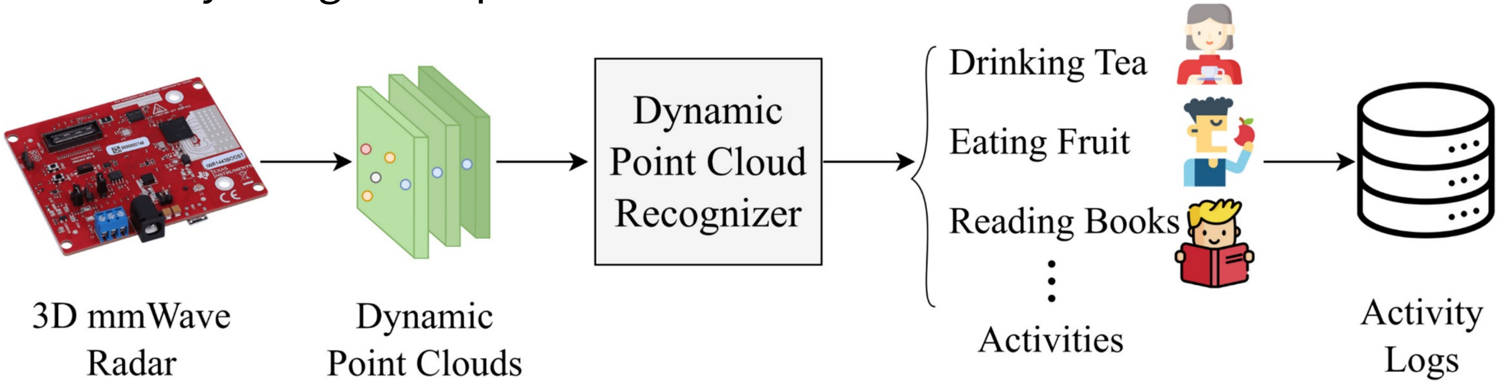
	MARS	AlexNet	GoogLeNet	ResNet-18	ResNet-34	ResNet-50
Nose	8.28 (±0.17)	6.91 (±0.06)	4.73 (±0.05)	4.41 (±0.05)	**4.22 (±0.05)**	4.75 (±0.08)
L. Shldr	6.66 (±0.10)	5.65 (±0.05)	3.91 (±0.04)	3.74 (±0.04)	**3.58 (±0.04)**	3.93 (±0.05)
R. Shldr	6.58 (±0.11)	5.54 (±0.05)	3.86 (±0.04)	3.67 (±0.04)	**3.52 (±0.04)**	3.91 (±0.06)
L. Elbow	8.68 (±0.15)	7.13 (±0.05)	4.84 (±0.04)	4.59 (±0.04)	**4.39 (±0.04)**	4.83 (±0.05)
R. Elbow	8.38 (±0.16)	7.05 (±0.05)	4.97 (±0.04)	4.76 (±0.04)	**4.57 (±0.04)**	4.94 (±0.04)
L. Wrist	11.52 (±0.13)	9.60 (±0.06)	6.53 (±0.05)	6.36 (±0.05)	**6.06 (±0.05)**	6.55 (±0.06)
R. Wrist	11.78 (±0.10)	10.19 (±0.06)	7.29 (±0.05)	7.15 (±0.05)	**6.85 (±0.05)**	7.33 (±0.06)
L. Pinky	12.51 (±0.13)	10.44 (±0.07)	7.14 (±0.05)	6.96 (±0.05)	**6.62 (±0.05)**	7.13 (±0.06)
R. Pinky	13.27 (±0.15)	11.47 (±0.07)	8.24 (±0.06)	8.08 (±0.06)	**7.74 (±0.06)**	8.28 (±0.06)
L. Index	12.54 (±0.13)	10.50 (±0.07)	7.20 (±0.06)	7.00 (±0.05)	**6.67 (±0.05)**	7.21 (±0.06)
R. Index	13.16 (±0.11)	11.41 (±0.07)	8.23 (±0.06)	8.06 (±0.06)	**7.73 (±0.06)**	8.26 (±0.06)
L. Thumb	11.63 (±0.13)	9.70 (±0.06)	6.62 (±0.05)	6.46 (±0.05)	**6.15 (±0.05)**	6.63 (±0.05)
R. Thumb	11.96 (±0.09)	10.36 (±0.06)	7.42 (±0.05)	7.29 (±0.05)	**6.99 (±0.05)**	7.47 (±0.06)
Average	10.54 (±1.33)	8.92 (±1.16)	6.23 (±0.85)	6.04 (±0.86)	**5.78 (±0.83)**	6.25 (±0.85)

ERRORS OF THE ESTIMATED JOINTS IN SPE①

	MARS	AlexNet	GoogLeNet	ResNet-18	ResNet-34	ResNet-50
Nose	8.28 (±0.17)	6.91 (±0.06)	4.73 (±0.05)	4.41 (±0.05)	**4.22 (±0.05)**	4.75 (±0.08)
L. Shldr	6.66 (±0.10)	5.65 (±0.05)	3.91 (±0.04)	3.74 (±0.04)	**3.58 (±0.04)**	3.93 (±0.05)
R. Shldr	6.58 (±0.11)	5.54 (±0.05)	3.86 (±0.04)	3.67 (±0.04)	**3.52 (±0.04)**	3.91 (±0.06)
L. Elbow	8.68 (±0.15)	7.13 (±0.05)	4.84 (±0.04)	4.59 (±0.04)	**4.39 (±0.04)**	4.83 (±0.05)
R. Elbow	8.38 (±0.16)	7.05 (±0.05)	4.97 (±0.04)	4.76 (±0.04)	**4.57 (±0.04)**	4.94 (±0.04)
L. Wrist	11.52 (±0.13)	9.60 (±0.06)	6.53 (±0.05)	6.36 (±0.05)	**6.06 (±0.05)**	6.55 (±0.06)
R. Wrist	11.78 (±0.10)	10.19 (±0.06)	7.29 (±0.05)	7.15 (±0.05)	**6.85 (±0.05)**	7.33 (±0.06)
L. Pinky	12.51 (±0.13)	10.44 (±0.07)	7.14 (±0.05)	6.96 (±0.05)	**6.62 (±0.05)**	7.13 (±0.06)
R. Pinky	13.27 (±0.15)	11.47 (±0.07)	8.24 (±0.06)	8.08 (±0.06)	**7.74 (±0.06)**	8.28 (±0.06)
L. Index	12.54 (±0.13)	10.50 (±0.07)	7.20 (±0.06)	7.00 (±0.05)	**6.67 (±0.05)**	7.21 (±0.06)
R. Index	13.16 (±0.11)	11.41 (±0.07)	8.23 (±0.06)	8.06 (±0.06)	**7.73 (±0.06)**	8.26 (±0.06)
L. Thumb	11.63 (±0.13)	9.70 (±0.06)	6.62 (±0.05)	6.46 (±0.05)	**6.15 (±0.05)**	6.63 (±0.05)
R. Thumb	11.96 (±0.09)	10.36 (±0.06)	7.42 (±0.05)	7.29 (±0.05)	**6.99 (±0.05)**	7.47 (±0.06)
Average	10.54 (±1.33)	8.92 (±1.16)	6.23 (±0.85)	6.04 (±0.86)	**5.78 (±0.83)**	6.25 (±0.85)

END-TO-END SOLUTION

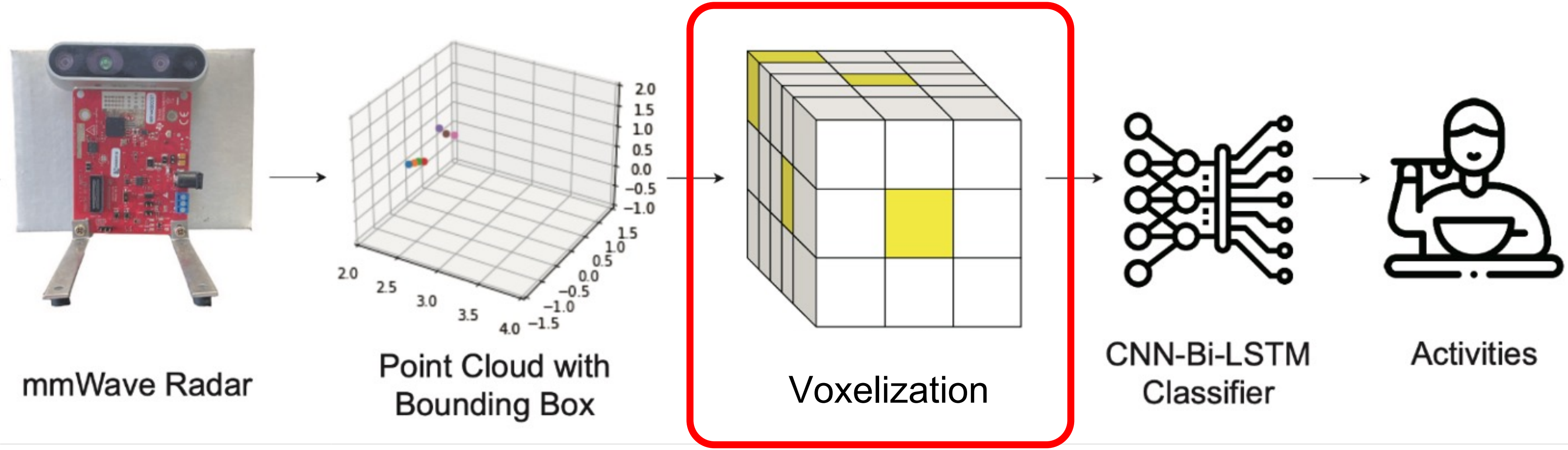
- End-to-end learning
  - An ML technique where we train a **single neural network** for complex tasks
  - Directly using raw input data without manual feature extraction



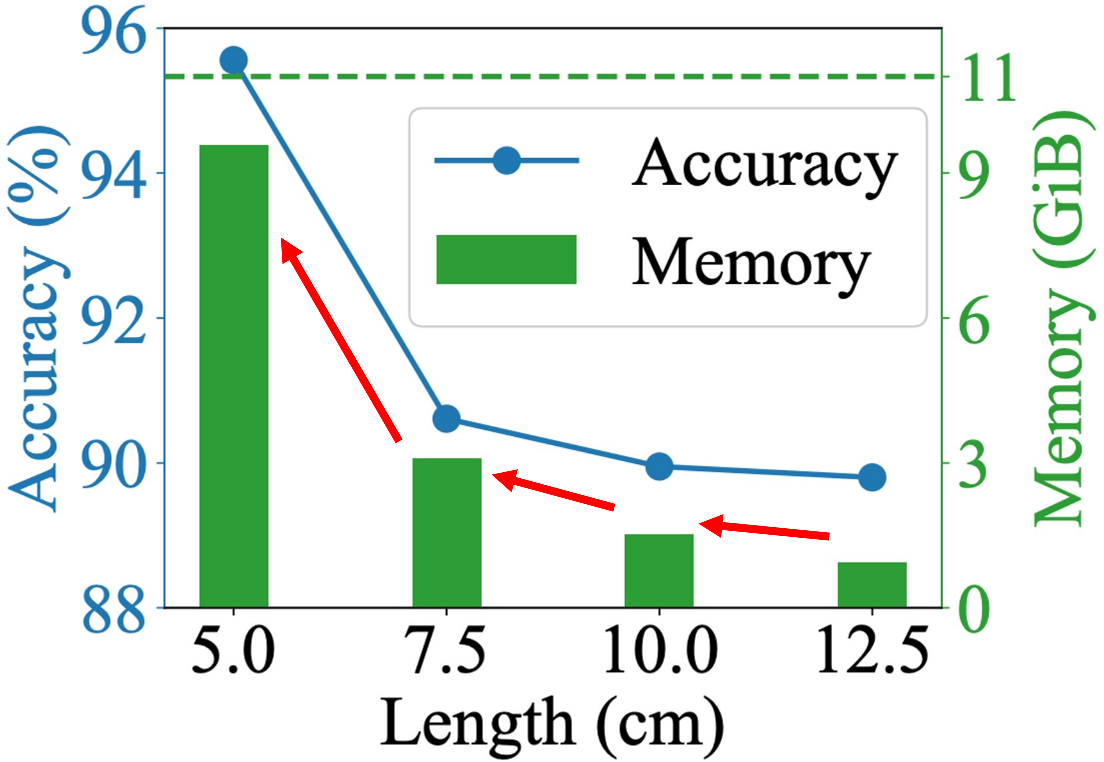
- Food Intake Activity (FIA) [MADiMa '22]
  → Uses **voxelization**, which is quite memory inefficient for sparse point clouds

[MADiMa ‘22] 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.

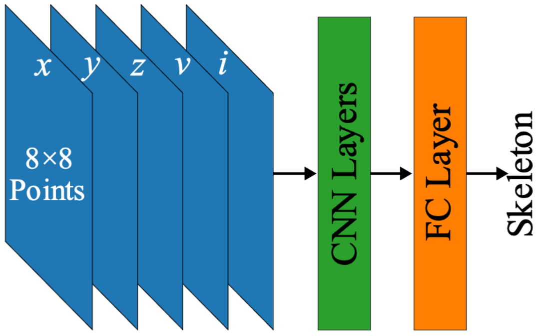
FIA PIPELINE

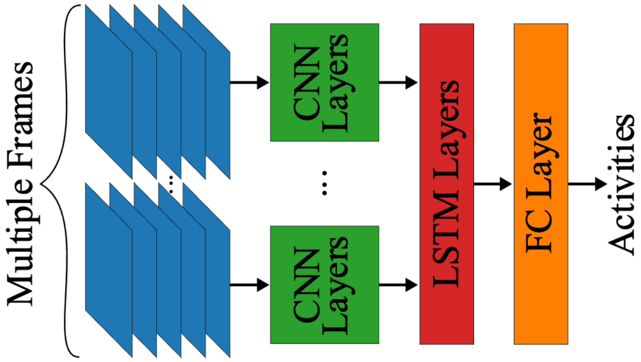


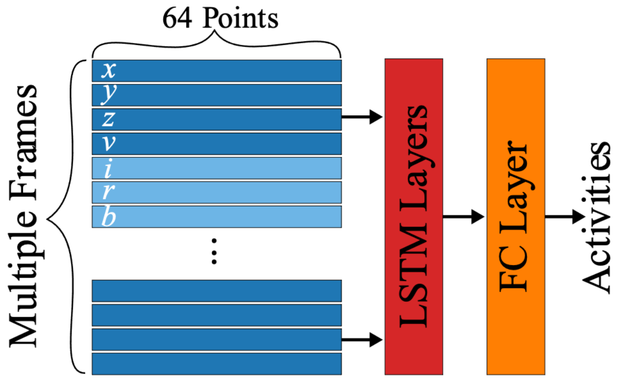
FIA Voxel Size↓ Accuracy↑ Memory↑



PROPOSED SOLUTIONS





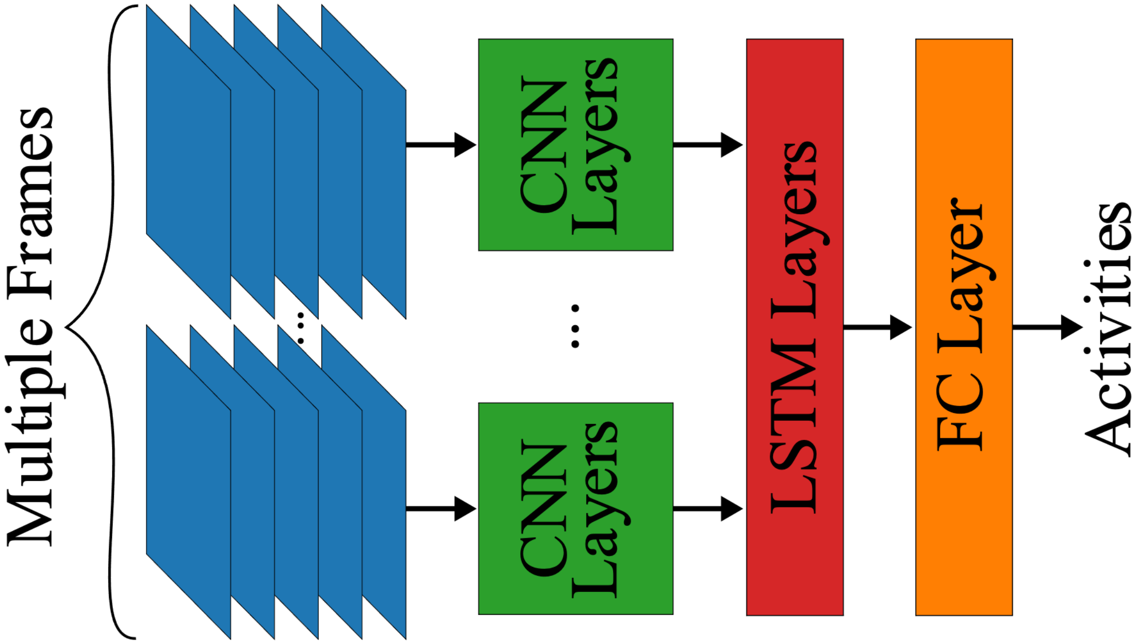


① **Skeletal Pose Estimator (SPE)**

② **Dynamic Point Cloud Recognizer (DPR)**

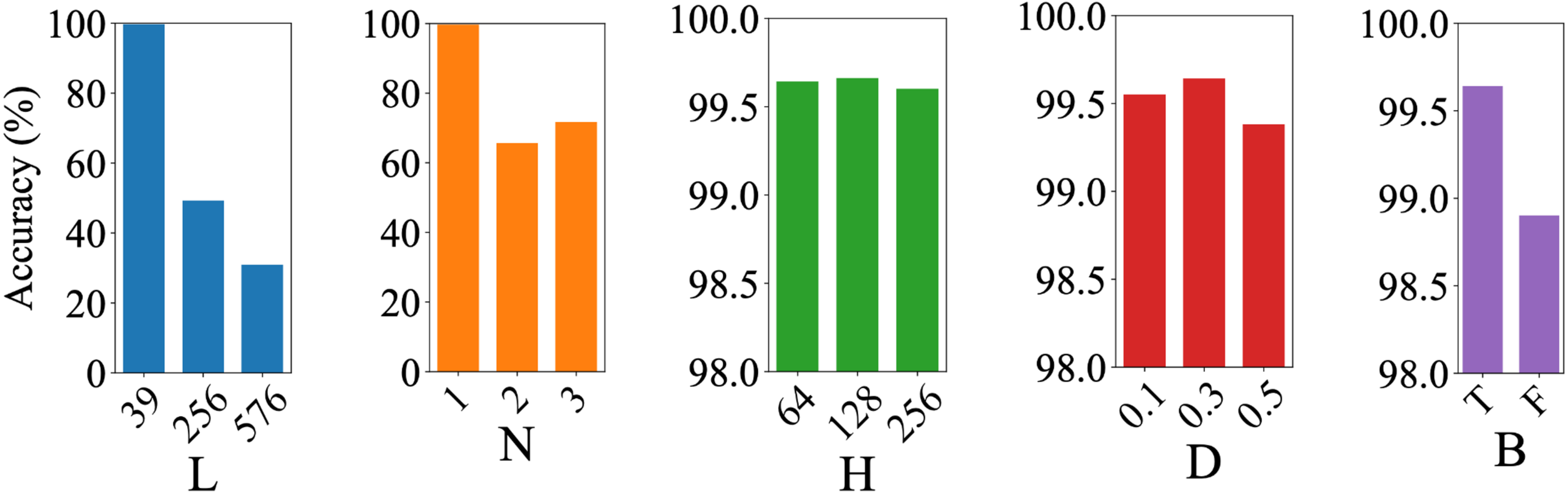
③ **Lightweight Dynamic Point Cloud Recognizer (LDPR)**

PROPOSED SOLUTIONS



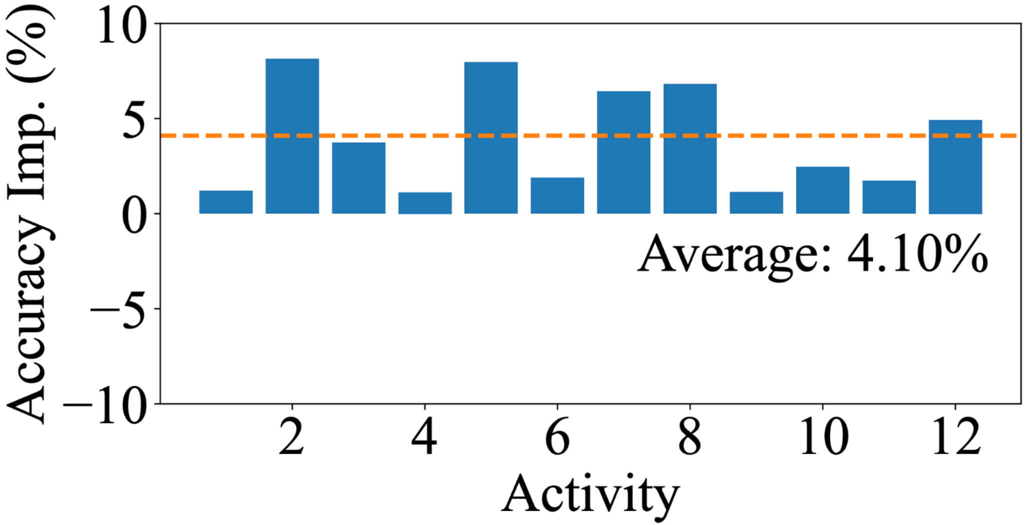
② Dynamic Point Cloud Recognizer (DPR)

FINDING THE OPTIMAL PARAMETERS

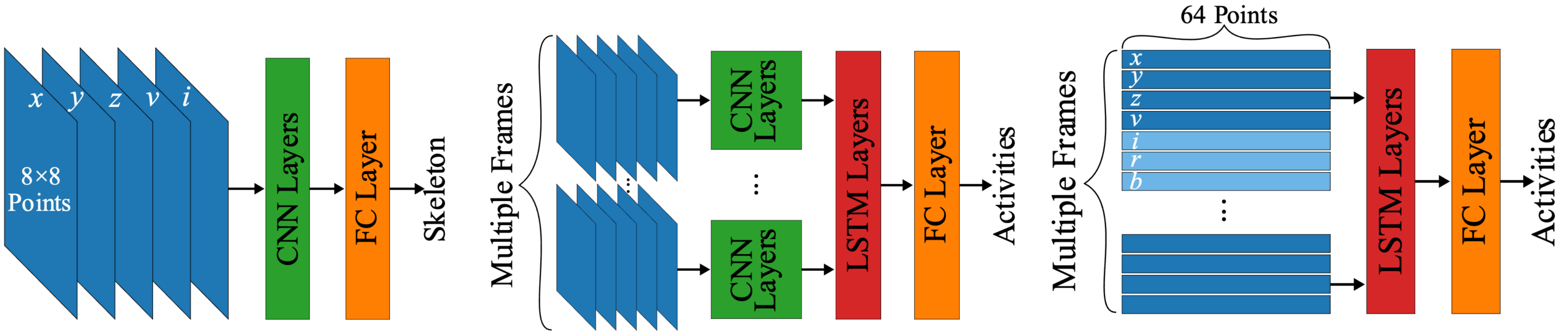


IMPROVEMENT OF DPR② OVER FIA

- Activity Recognition Accuracy
  - 95.56% → 99.66% (4.10% improvement)
- GPU Memory Consumption
  - 9817 MiB → 2131 MiB (**78.29% reduction**)



PROPOSED SOLUTIONS

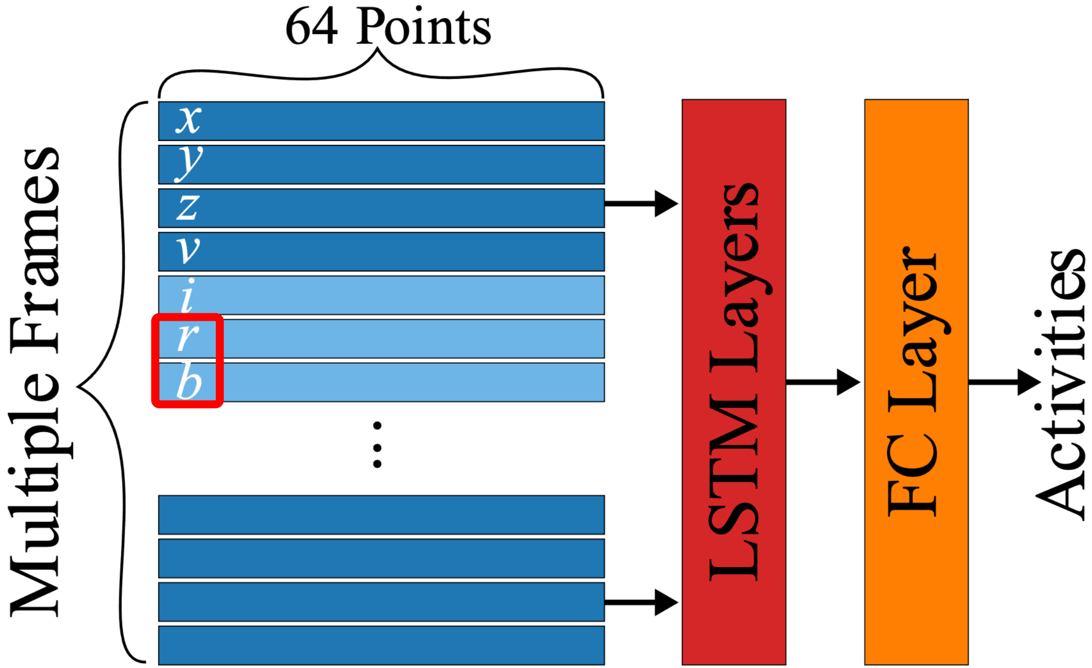


① Skeletal Pose Estimator (SPE)

② Dynamic Point Cloud Recognizer (DPR)

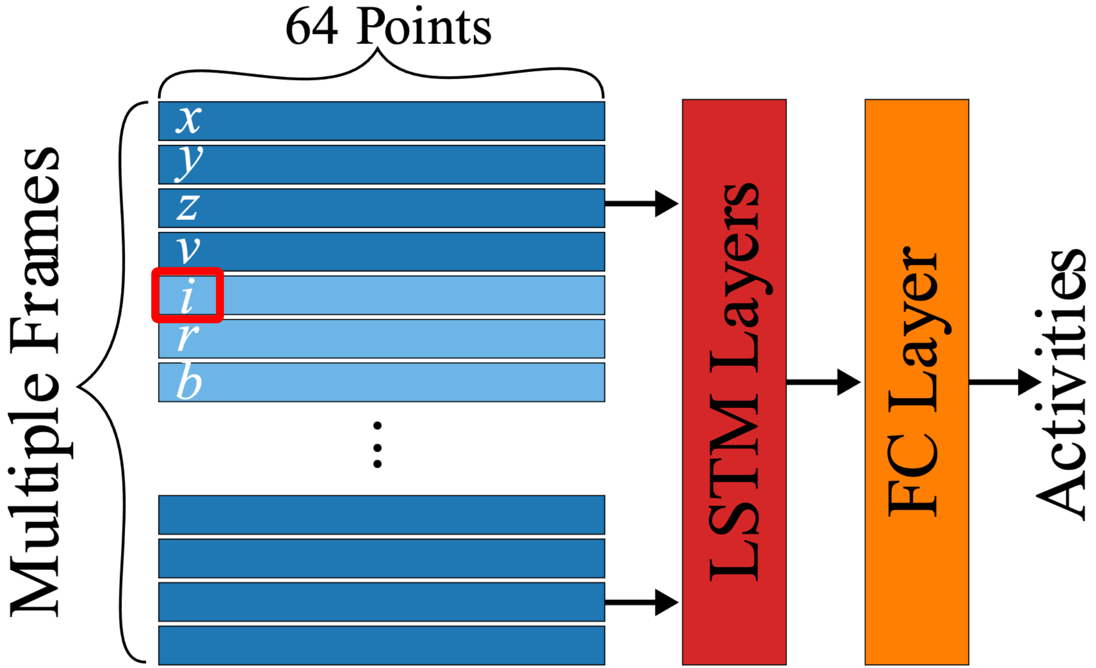
③ Lightweight Dynamic Point Cloud Recognizer (LDPR)

PROPOSED SOLUTIONS



③ **Lightweight Dynamic Point Cloud Recognizer (LDPR)**

PROPOSED SOLUTIONS



③ **Lightweight Dynamic Point Cloud Recognizer (LDPR)**

LDPR③ ACCURACY USING DIFFERENT FEATURES

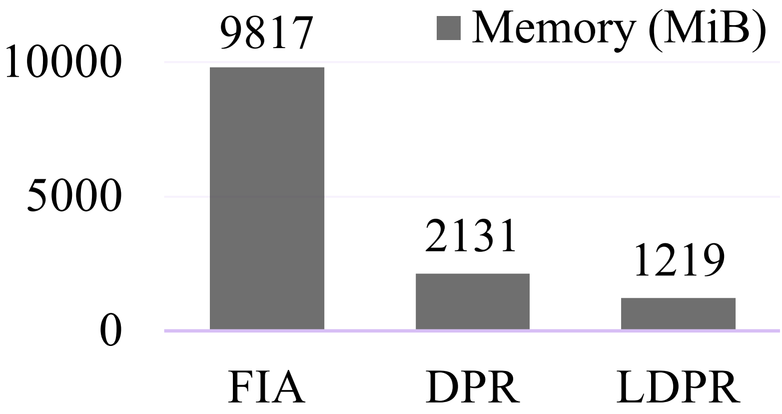
Accuracy	$\{x, y, z\}$	$\{r, b\}$	$\{x, y, z, r, b\}$
With i	98.51%	99.74%	99.78%
Without i	**99.81%**	99.74%	**99.81%**

LDPR③ ACCURACY USING DIFFERENT FEATURES

Accuracy	$\{x, y, z\}$	$\{r, b\}$	$\{x, y, z, r, b\}$
With i	98.51%	99.74%	99.78%
Without i	**99.81%**	99.74%	**99.81%**

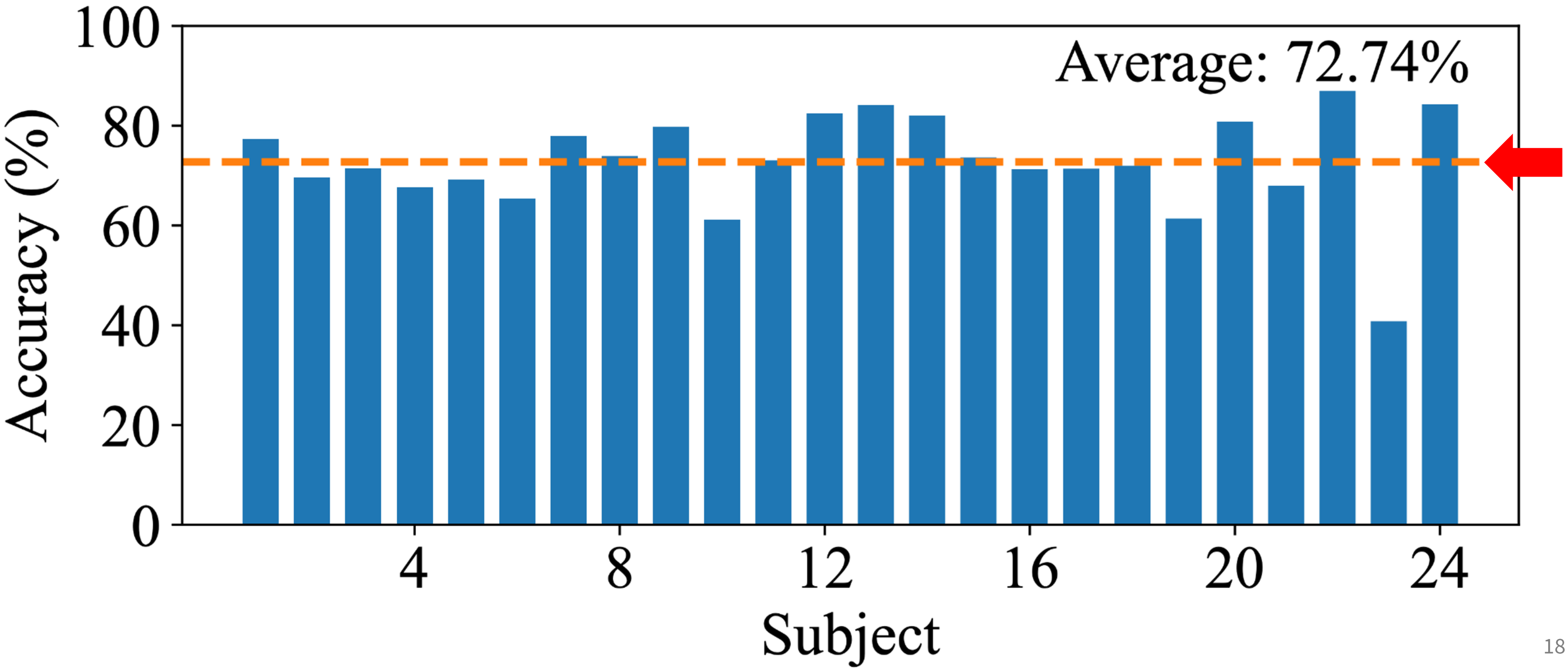
IMPROVEMENT OF LDPR③ OVER DPR②

- Activity Recognition Accuracy
  - 99.66% → 99.81 (0.15% improvement)
- GPU Memory Consumption
  - 2131 MiB → 1219 MiB (**42.80% reduction**)

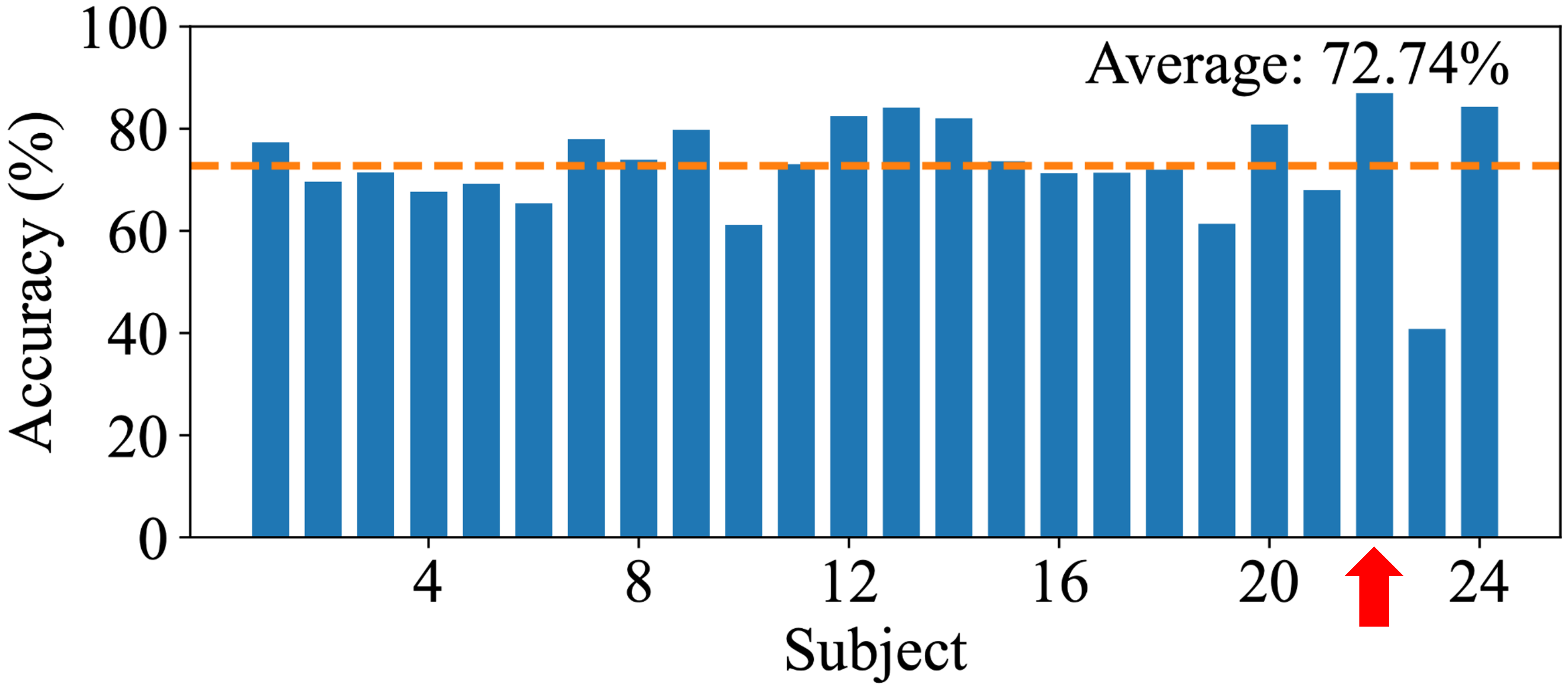


- Besides the standard 80-20 split, we also experimented with the **leave-one-out test**
  - Each subject is selected for testing
  - Remaining 23 subjects for training
  - Replicates the situation of **recognizing a new subject**

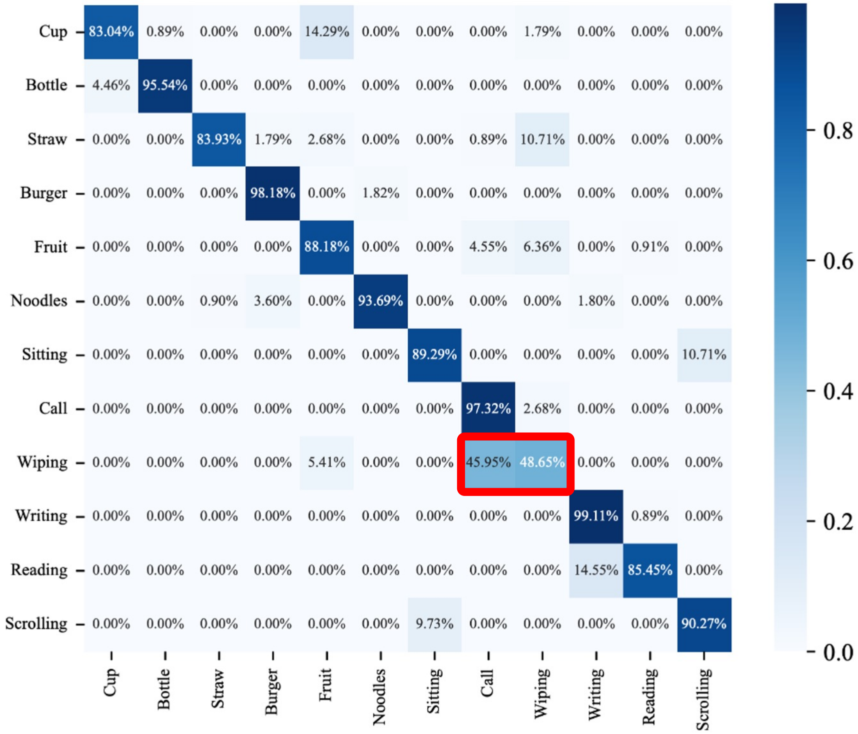
LDPR③ ACCURACY: LEAVE-ONE-OUT TEST

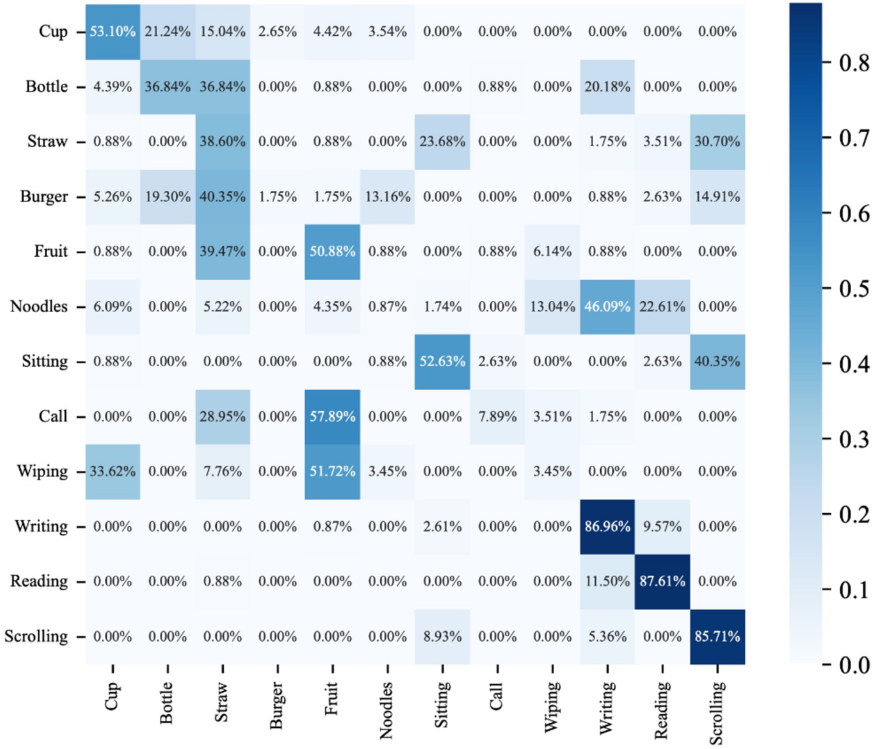


LDPR③ ACCURACY: LEAVE-ONE-OUT TEST

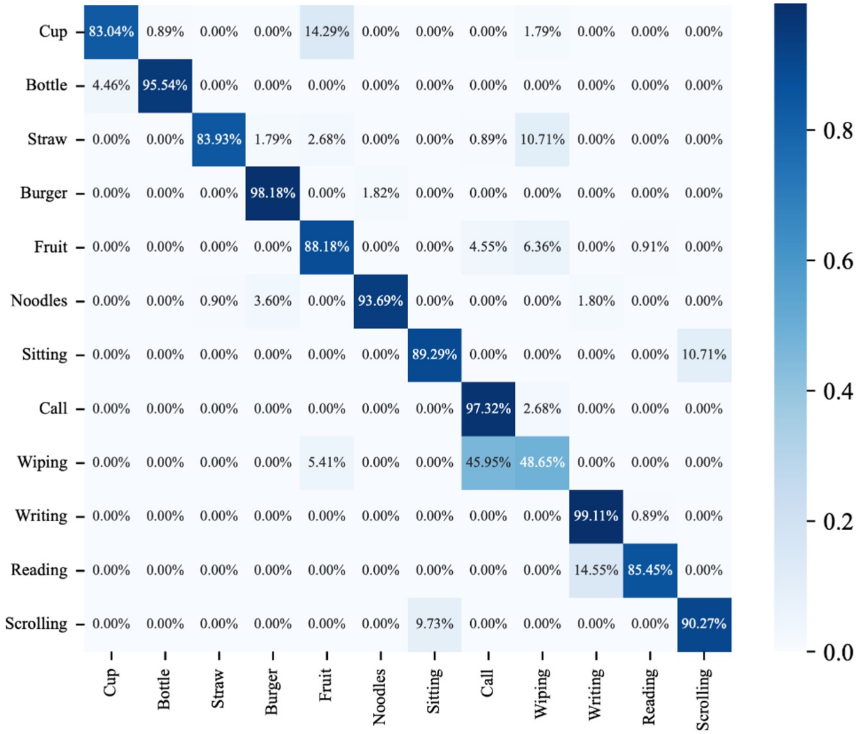


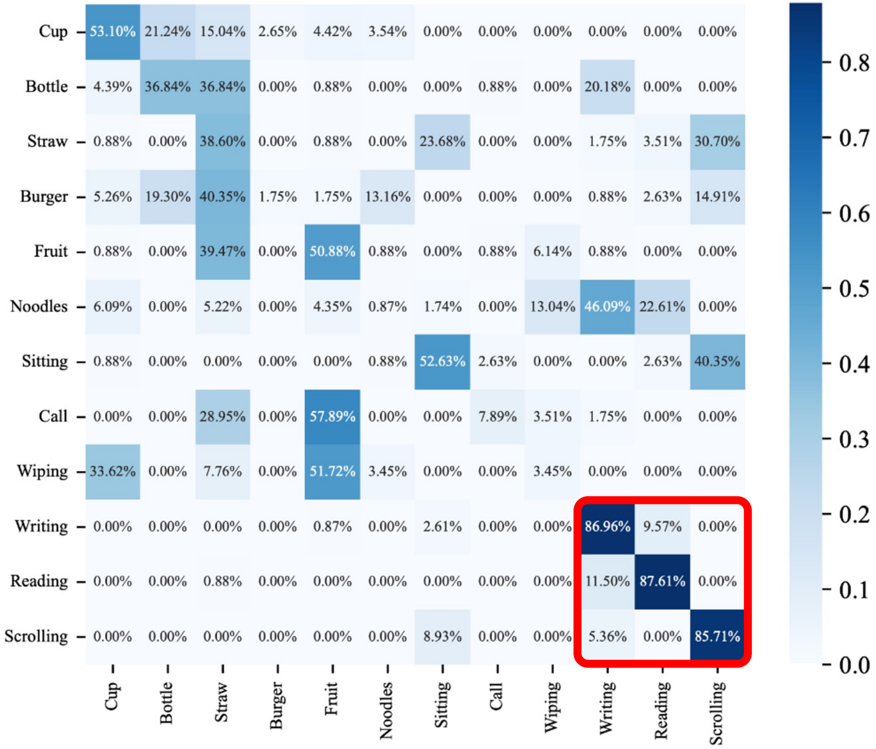
CONFUSION MATRIX: BEST/WORST SUBJECT



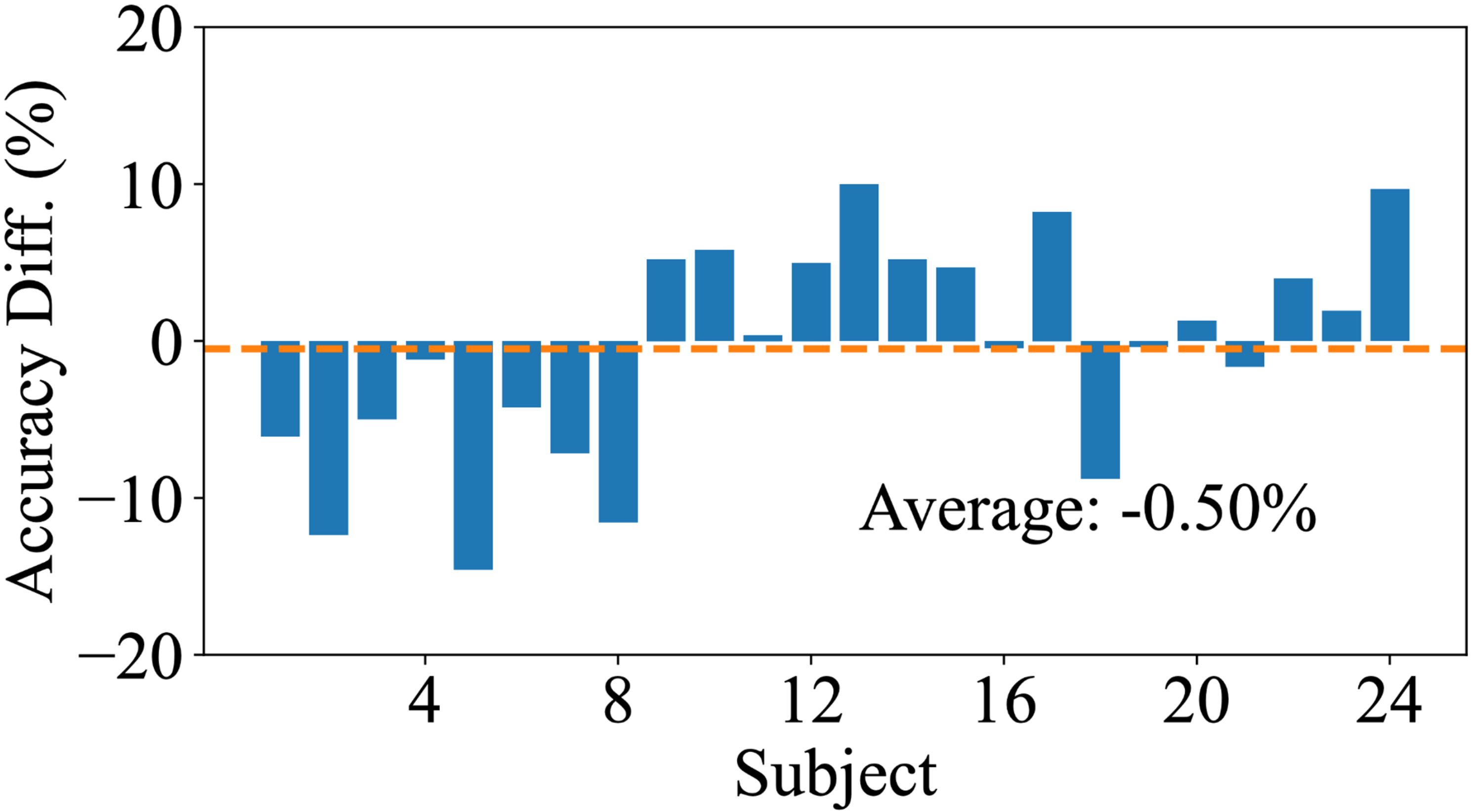


CONFUSION MATRIX: BEST/WORST SUBJECT





ACCURACY DIFF. BETWEEN LDPR③ AND DPR②



SUMMARY

Task:

- Food intake activity recognition using mmWave point clouds

Proposed solutions:

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

Ongoing works:

- Further enhancing the **precision of the estimated skeletons**
- Enhance the model's **generalization capability**
- Apply these solutions to other scenarios, e.g., **Driver Monitoring System** (DMS)

ANY QUESTIONS?