PALM: Personalized Active Learning for mmWave-Based Activity Recognition

Hsin-Che Chiang
hsinchechiang@gmail.com

Network and Multimedia Systems Lab
Department of Computer Science
National Tsing Hua University

OUTLINE

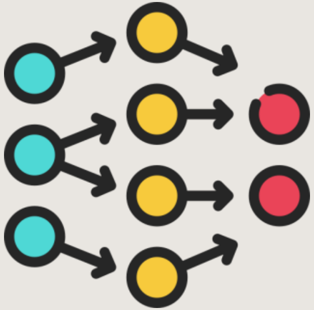
1. **H. Chiang**, Y. Wu, G. Li, S. Shirmohammadi, and C. Hsu, "PALM: Personalized active learning for mmWave-based activity recognition," submitted to IEEE Transactions on Instrumentation and Measurement (TIM), 2024.
2. **H. Chiang**, Y. Wu, S. Shirmohammadi, and C. Hsu, "Memory-efficient high-accuracy food intake activity recognition with 3D mmWave radars," in Proc. of the ACM International Workshop on Multimedia Assisted Dietary Management (MADiMa), 2023, pp. 33–41, Best Paper Award.
3. Y. Wu, **H. Chiang**, S. Shirmohammadi, and C. Hsu, "A dataset of food intake activities using sensors with heterogeneous privacy sensitivity levels," in Proc. of the ACM Multimedia Systems Conference (MMSys), 2023, pp. 416–422.
4. G. Li, **H. Chiang**, Y. Li, S. Shirmohammadi, and C. Hsu, "A driver activity dataset with multiple RGB-D cameras and mmWave radars," in Proc. of the ACM Multimedia Systems Conference (MMSys), 2024, pp. 360–366.

Human Activity Recognition (HAR)

01 **Motivation**
02 Methodology
03 Evaluations
04 Conclusion

⋮

HAR System



Driving

Drinking

Texting

...

Choosing the Sensor (1/2)

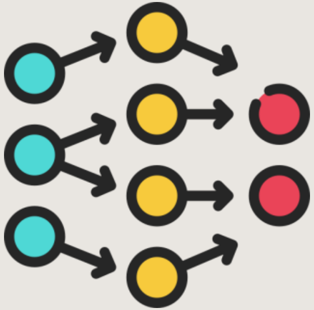
Wearable Sensors

- Accelerometer
- Gyroscope
- EMG

In-situ Sensors

- RGB Camera
- Depth Camera
- mmWave Radar

HAR System



Choosing the Sensor (2/2)

01 **Motivation**
02 Methodology
03 Evaluations
04 Conclusion

Wearable Sensors

- Accelerometer
- Gyroscope
- EMG

→ Troublesome for Daily Use

In-situ Sensors

- RGB Camera
- Depth Camera

→ Privacy Concerns

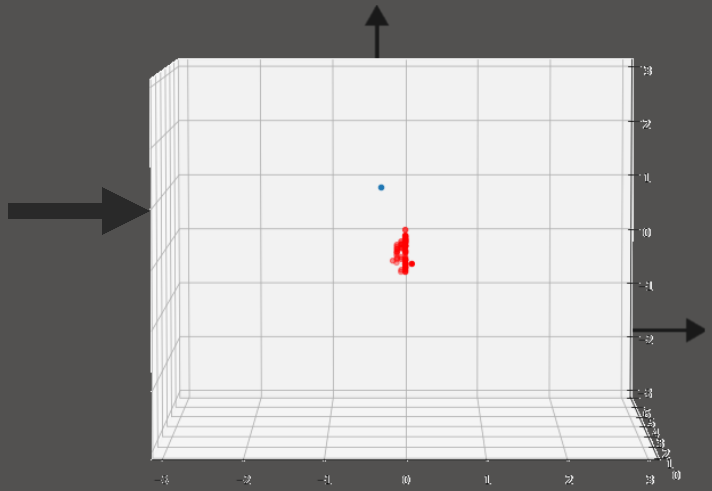
- **mmWave Radar**

→ **Non-intrusive**
Privacy Preserving

⋮

mmWave Radar

01 Motiva
02 Metho
03 Evalua
04 Conclu



mmWave Radar

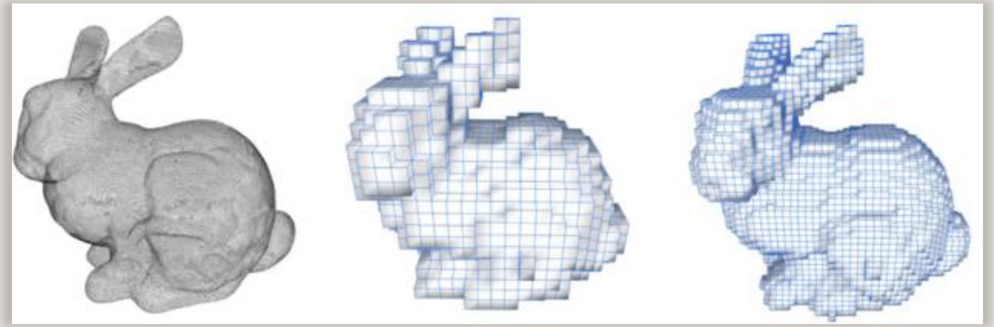
Sparse Dynamic Point Cloud

[1] Texas Instrument IWR1443BOOST, https://www.mouser.tw/ProductDetail/Texas-Instruments/IWR1443BOOST?qs=5aG0NVq1C4wT7gyvvDbMRw%3D%3D

Problem 1

Resource Inefficiency of Voxelization

- Voxelization is a common technique for point cloud preprocessing
- Using finer voxels leads to
  - Higher accuracy
  - **Higher memory consumption**



[1] B. Guan, S. Lin, R. Wang, F. Zhou, X. Luo, and Y. Zheng, "Voxel-based quadrilateral mesh generation from point cloud," Multimedia Tools and Applications 79, 2020.

Problem 2

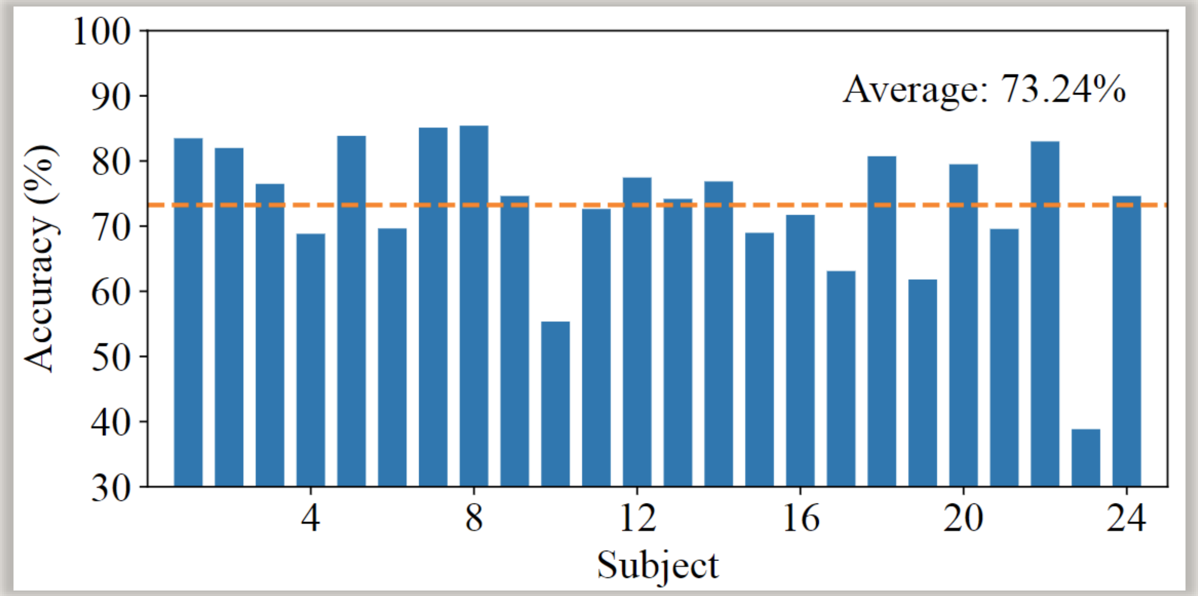
Cold Start Problem for New Users (1/2)

- Cold start (recommender systems)
  - The system cannot draw any inferences for users or items about which it has not yet gathered sufficient information
- Solution
  - Huge amount of training data to generalize to new users
  - **Personalized training data**
- Lack of large-scale public datasets for mmWave point clouds
  - ImageNet: 18 million labeled images, 20,000+ classes
  - Food Intake Activity Dataset: 24 subjects, 12 activities
  - Driver Activity Dataset: 15 subjects, 11 activities

Problem 2

Cold Start Problem for New Users (2/2)

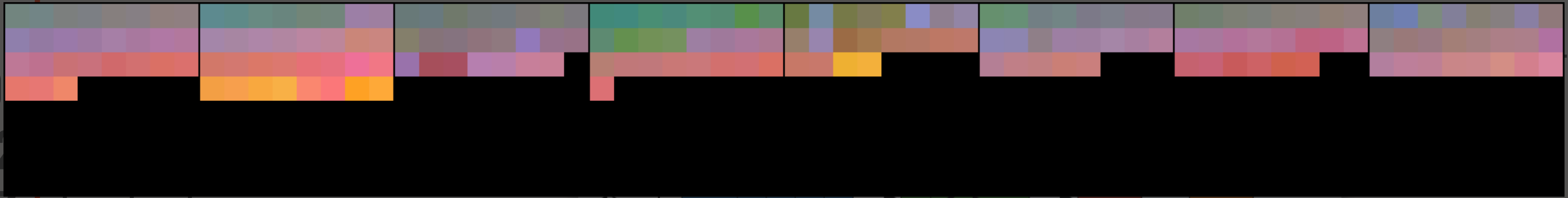
- Global test (80/20 train-test split): >95% accuracy
- Leave-one-(subject)-out test: ~73% accuracy

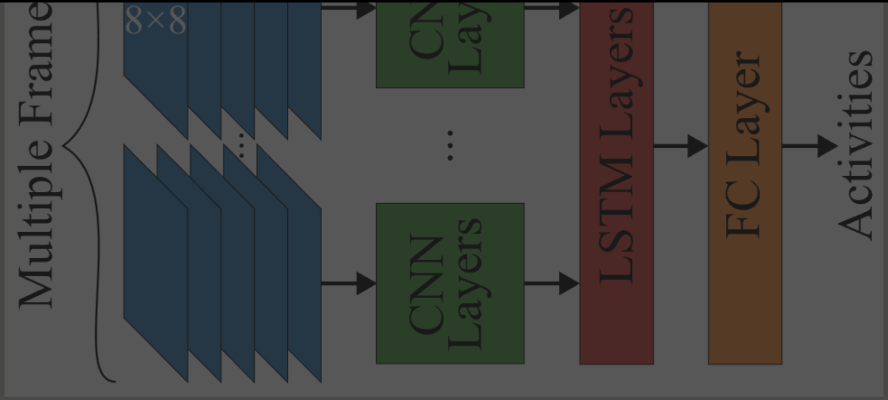


Proposal 1 : DPR

Preprocessing: Feature Map

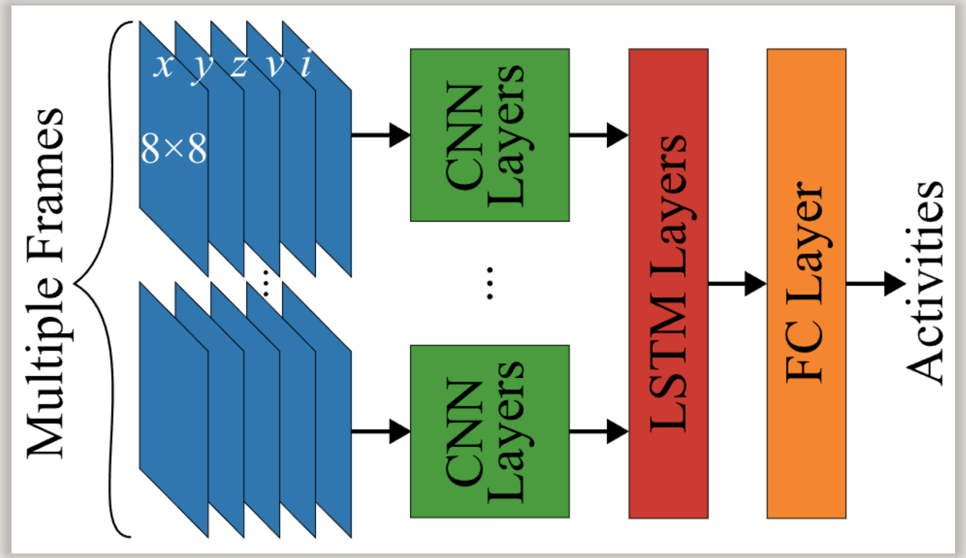
- Arrange each point cloud frame into an 8×8 “image”
- 5 channels (features): coordinates **x**, **y**, **z**, velocity **v**, intensity **i**





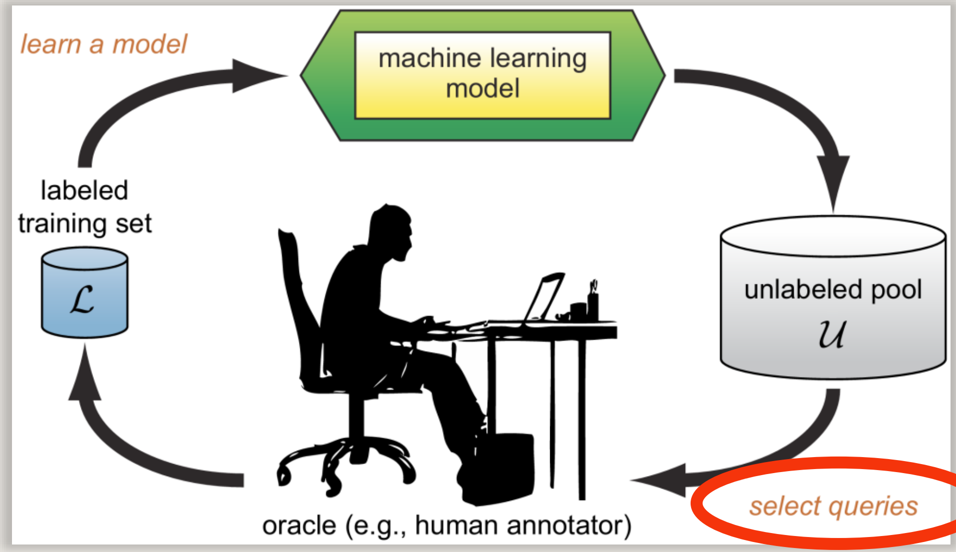
Neural Network Structure: CNN-LSTM

- CNN: **spatial information** within each frame
- LSTM: **temporal information** across nearby frames



What is Active Learning?

- A subfield of machine learning (and, more generally, AI)
- Make the algorithm **choose the data from which it learns**

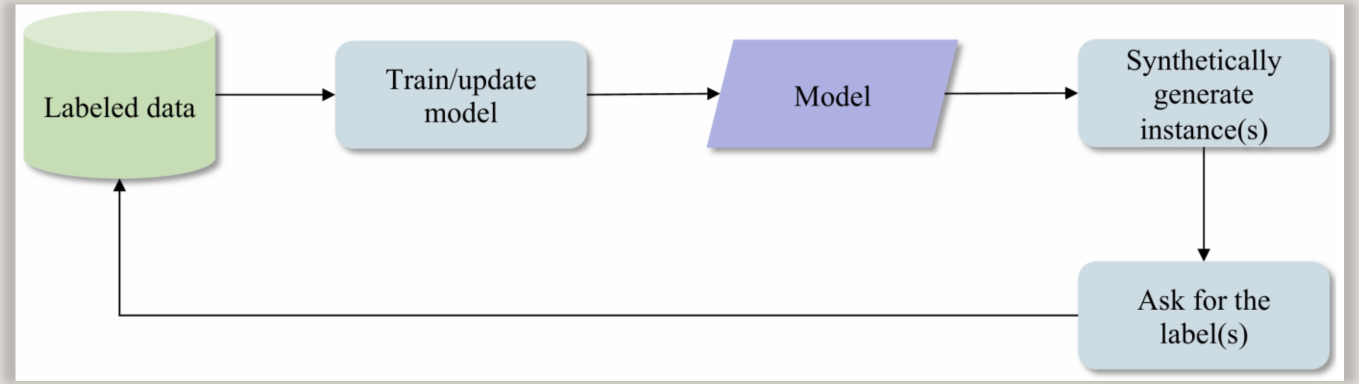


How?

[1] B. Settles, "Active learning literature survey," University of Wisconsin-Madison Department of Computer Science, Tech. Rep., 2009

Active Learning Scenarios (1/3)

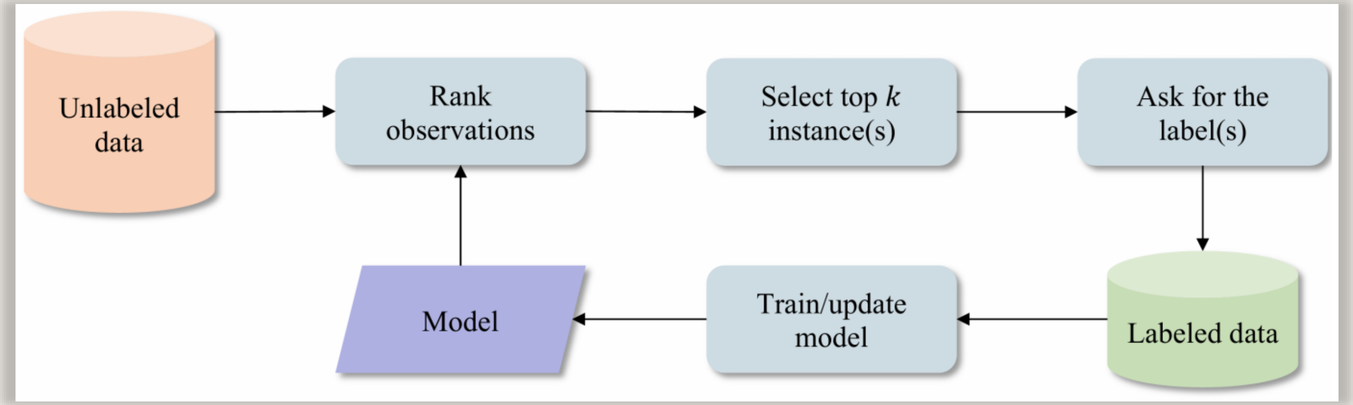
1. Membership Query Synthesis
   - Allows **synthetically generated** samples
   - Can generate nonsensical samples that human annotators cannot adequately label



[1] D. Cacciarelli and M. Kulahci, "Active learning for data streams: a survey," Machine Learning, vol. 113, no. 1, pp. 185–239, 2024.

Active Learning Scenarios (2/3)

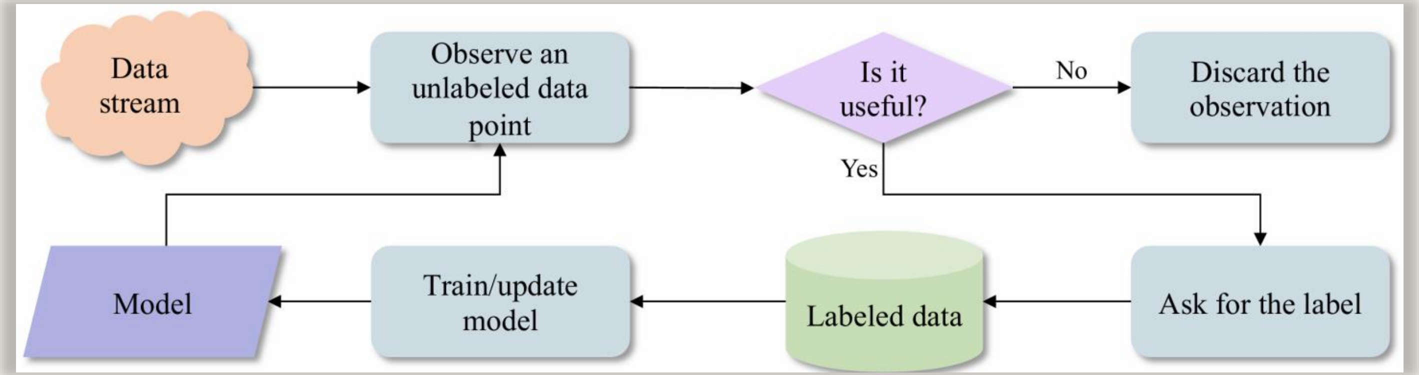
2. Pool-based Active Learning
    - Select from a **static pool** of unlabeled data
    - Suitable for tasks where large volumes of unlabeled data can be gathered simultaneously



[1] D. Cacciarelli and M. Kulahci, “Active learning for data streams: a survey,” Machine Learning, vol. 113, no. 1, pp. 185–239, 2024.

Active Learning Scenarios (3/3)

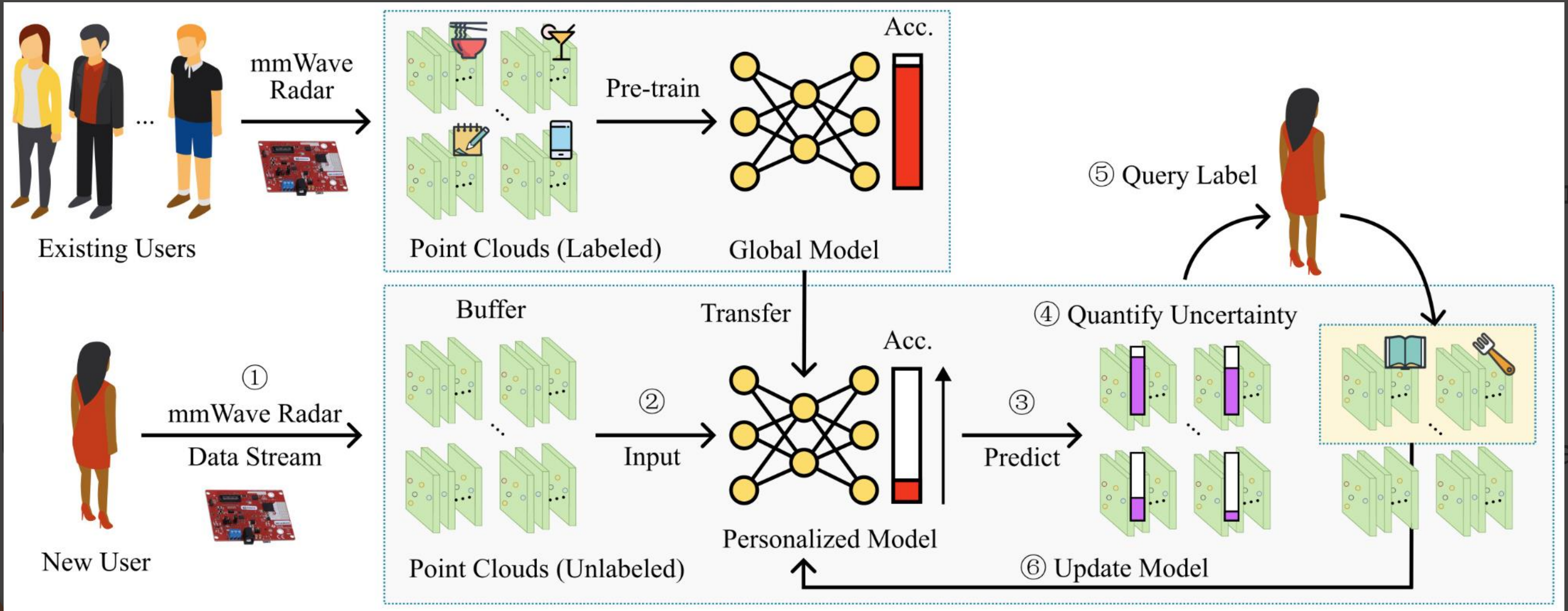
3. Stream-based (Online) Active Learning
   - Process data that arrives continuously
   - Example: **spam filtering**



[1] D. Cacciarelli and M. Kulahci, "Active learning for data streams: a survey," Machine Learning, vol. 113, no. 1, pp. 185–239, 2024.

Proposal 2 : PALM

Personalized Active Learning for mmWave



01
02
03
04

Multiple Predictions

- Required for some uncertainty quantification methods
- Can be achieved through approaches such as:
  - Monte Carlo (MC) dropout
  - MC batch normalization
  - Deep Ensembles
- We employ **MC dropout**
  - Requires minimum modification to the model
  - Dropout is already used for regularization during training
  - We use the same dropout rate during inference

Uncertainty Quantification (1/3)

- Regression task: ISO GUM defines Type A standard uncertainty as
  - The **experimental standard deviation of the mean** of multiple observations

$$\sigma_{\text{mean}} = \frac{\sigma}{\sqrt{n}}$$

- Classification task
  - No unified definition for uncertainty quantification of ordinal quantities or nominal properties

Uncertainty Quantification (2/3)

- Deviation of Predictions

$$D(x) = \sqrt{\frac{1}{M}\sum_{m=1}^{M}(p_{best}^{m} - \overline{p}_{best})^2}$$

- Least Confidence

$$LC(x) = 1 - p_{best}$$

- Margin Sampling

$$M(x) = p_{best} - p_{second}$$

- Information Entropy

$$H(x) = -\sum_{i=1}^{C} p_i \log(p_i)$$

Proposal 2 : PALM

Uncertainty Quantification (3/3)

Class	1	2	3	4	5	6	7	8	9	10	LC↑	M↓	H↑
Sample 1	1	0	0	0	0	0	0	0	0	0			
Sample 2	.46	.06	.06	.06	.06	.06	.06	.06	.06	.06			
Sample 3	.46	.46	.01	.01	.01	.01	.01	.01	.01	.01			
Sample 4	.1	.1	.1	.1	.1	.1	.1	.1	.1	.1			

Multiple Predictions with Entropy

- Multiple predictions can also be used to calculate Information Entropy
  - Entropy Mean
    - Entropy of the **mean prediction**
  - Max Entropy
    - **Maximum entropy** among predictions
  - BALD (Bayesian Active Learning by Disagreement)
    - Difference between **Entropy Mean** and **Mean Entropy**
    - Higher BALD score when the predictions disagree

BALD Example

	Class 1	Class 2	H
Prediction 1	100%	0%	0
Prediction 2	0%	100%	0
Mean Prediction	50%	50%	Entropy Mean = 1 Mean Entropy = 0

BALD Score = 1

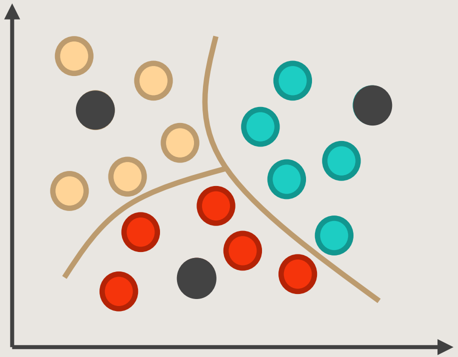
	Class 1	Class 2	H
Prediction 1	50%	50%	1
Prediction 2	50%	50%	1
Mean Prediction	50%	50%	Entropy Mean = 1 Mean Entropy = 1

BALD Score = 0

Proposal 2 : PALM

Uncertainty vs. Diversity

- Ensure the model is exposed to a wide range of scenarios
- Uncertainty vs. Diversity
  - Uncertainty: focuses on **ambiguous** predictions
  - Diversity: emphasizes **variability** in selected samples



Labeling Budget

- Labeling Budget (B): number of queries allowed each day
- Higher B
  - Accelerates model improvement
  - May **overwhelm the user** with frequent requests
- Lower B
  - Reduces user burden
  - May **slow down model learning**
- Balance user engagement and model enhancement

Model Update Strategy (1/2)

- Training Approaches
  - **Complete Re-Training**:
    - Used when a large number of labels is obtained, making the previous model obsolete
  - **Incremental Training**:
    - Fine-tunes the model to preserve existing knowledge without starting from scratch
- Epochs (E)
  - The number of epochs for updating the model affects the trade-off between computational efficiency and model refinement

Model Update Strategy (2/2)

- Freezing Layers (F): whether feature extraction layers are frozen
  - F = True: only classification layers are fine-tuned
  - F = False: the entire neural network is updated
- Resource vs. Accuracy
  - Freezing layers **conserves computational resources**, but may **lower accuracy**

Experimental Setup (1/3)

01 Motivation
02 Methodology
03 **Evaluations**
04 Conclusion

Food Intake Activity Dataset (FIAD)

Act.	Description	Time/Rep
a01	Drinking tea with a cup	4 sec
a02	Drinking tea with a bottle	4 sec
a03	Drinking tea with a straw	4 sec
a04	Eating a burger with hands	4 sec
a05	Eating fruit with a fork	4 sec
a06	Eating noodles with chopsticks	4 sec
a07	Sitting still	Continuous
a08	Picking up a call	4 sec
a09	Wiping mouth with a tissue	4 sec
a10	Writing on a piece of paper	4 sec
a11	Reading a book	4 sec
a12	Scrolling a smartphone	Continuous

Driver Activity Dataset (DAD)

Act.	Description	Time/Rep
a01	Waiting	Continuous
a02	Driving safely	Continuous
a03	Changing gears	4 sec
a04	Checking mirrors	4 sec
a05	Drinking water	4 sec
a06	Touching hair	4 sec
a07	Talking to passengers	4 sec
a08	Checking the phone	4 sec
a09	Picking up a call	4 sec
a10	Nodding	Continuous
a11	Reaching sideways	4 sec

Experimental Setup (2/3)

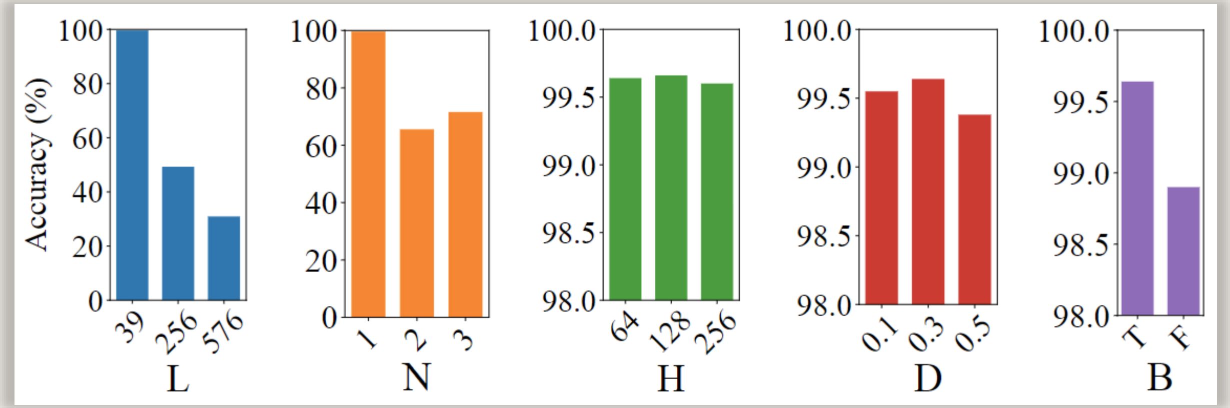
- Divide the mmWave data stream into individual samples:
  - Activities were recorded at a tempo of 4 sec / repetition (exceptions noted in the tables)
  - Employ a **sliding window**
    - Window size: 4 sec
    - Stride: 1 sec
- FIAD dataset: 34,560 samples (24 subjects × 12 activities × 120 sec)
- DAD dataset: 9,900 samples (15 subjects × 11 activities × 60 sec)

Experimental Setup (3/3)

- Hardware & Software
  - OS: Ubuntu 20.04
  - CPU: Intel Xeon-E5 2678 V3, 48 Cores @ 2.5 GHz
  - GPU: NVidia GeForce GTX 1080 Ti
  - Library: PyTorch 1.10.2 & Torch Vision 0.11.3
- Model Training
  - Loss function: Cross-Entropy Loss
  - Optimizer: Adam
  - Learning Rate: 10^{-3}

DPR Parameters

- L ∈ {**39**, 256, 576}, the input length of the LSTM layer
- N ∈ {**1**, 2, 3}, the number of LSTM layers
- H ∈ {64, **128**, 256}, the number of hidden LSTM states
- D ∈ {0.1, **0.3**, 0.5}, the dropout rate
- B ∈ {**True**, False}, a Boolean indicating the use of Bidirectional LSTM



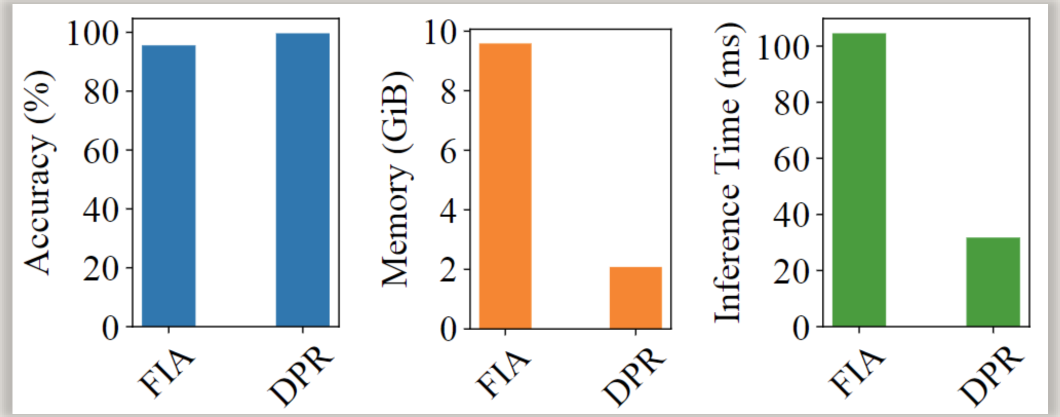
DPR Results (1/4)

- Compare against FIA[1], a state-of-the-art voxelization-based food intake activity recognition method
- Preprocessing:
  - DPR: feature maps
  - FIA: voxelization
- Neural Network:
  - Both use of CNN-LSTM
  - DPR: 2D-CNN (ResNet-34)
  - FIA: 3D-CNN

[1] Y. Wu, Y. Chen, S. Shirmohammadi, and C. Hsu, "AI-assisted food intake activity recognition using 3D mmWave radars," in Proc. of the ACM International Workshop on Multimedia Assisted Dietary Management (MADiMa), 2022, pp. 81–89.

DPR Results (2/4)

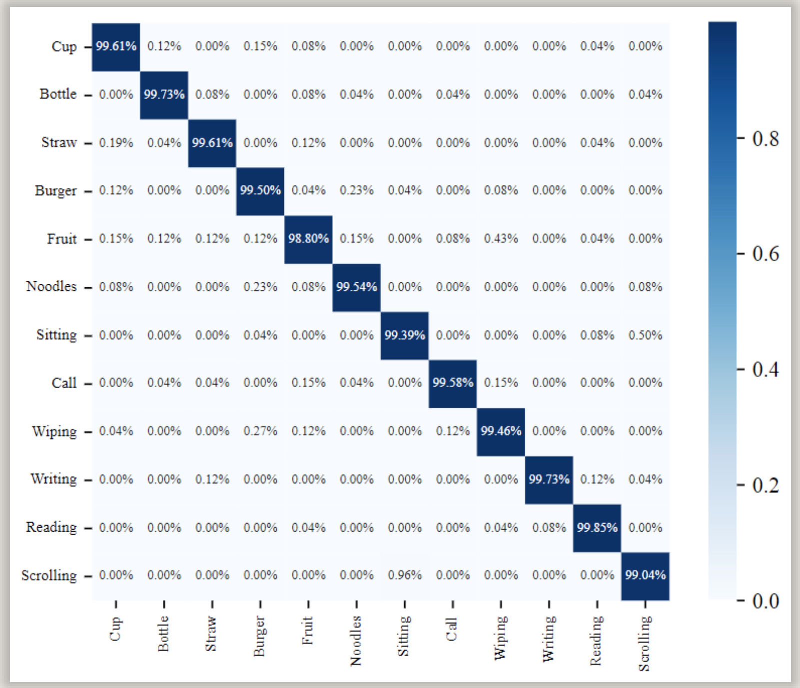
- DPR achieves 99.66% accuracy
  - **4.10% improvement** over FIA's 95.56%
- DPR consumes 2131 MiB of memory
  - **78.29% reduction** compared to FIA's 9817 MiB
- DPR has an average inference time of 31.75 ms / sample
  - **69.64% reduction** compared to FIA's 104.58 ms



Global Model

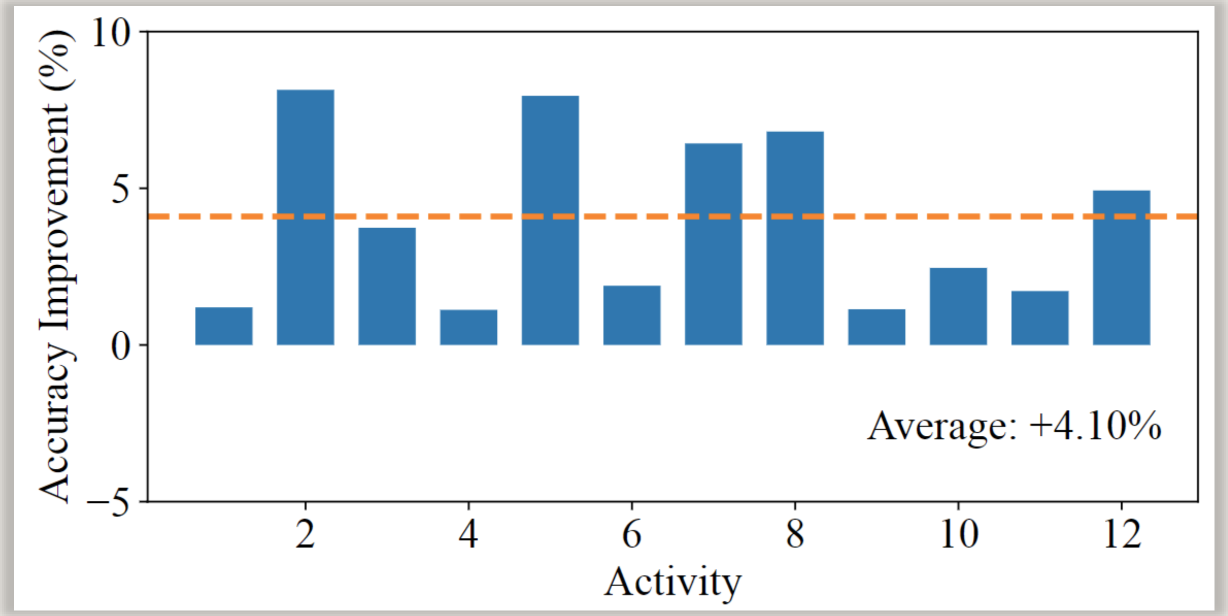
DPR Results (3/4)

- Confusion matrix show high accuracy across all activities



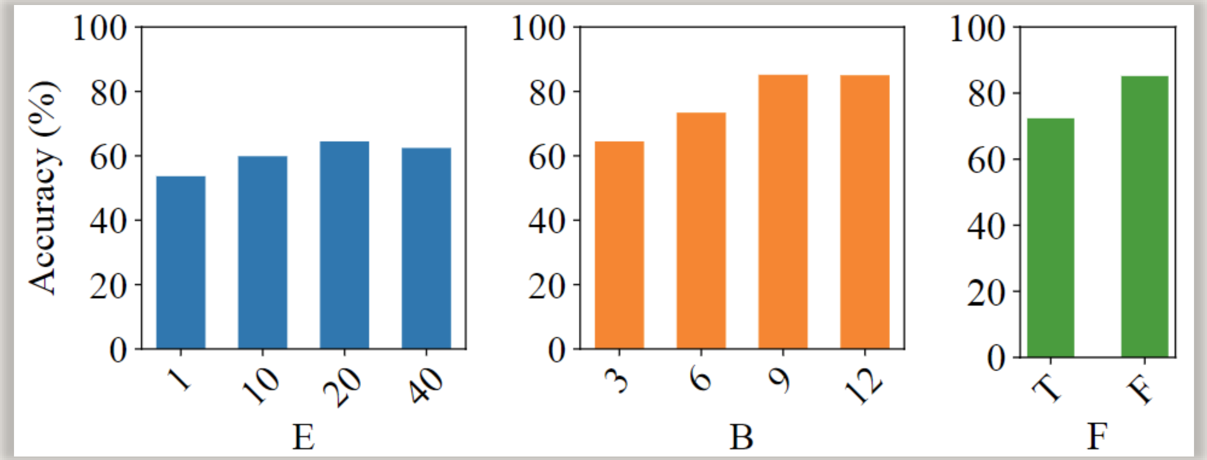
DPR Results (4/4)

- Improvements over FIA ranging from 1.12% to 8.14%



PALM Parameters

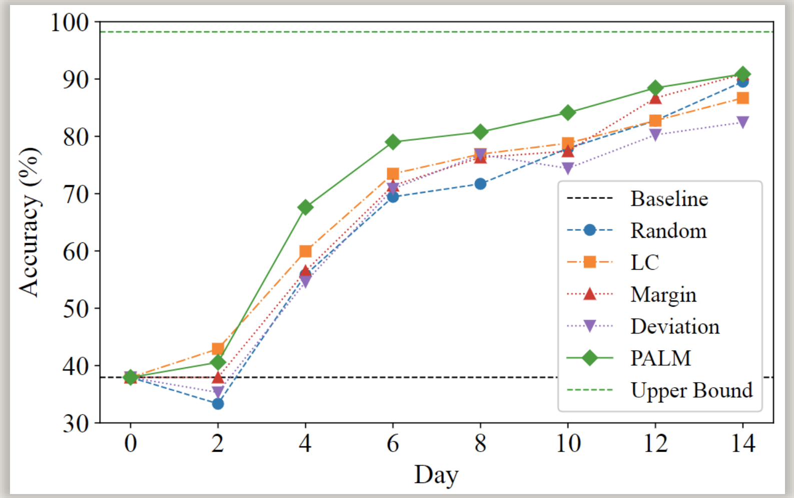
- E ∈ {1, 10, **20**, 40}, number of epochs each time labeled data is obtained
- B ∈ {3, 6, **9**, 12}, the labeling budget
- F ∈ {True, **False**}, where the feature extraction layers are frozen



Personalized Model

PALM Results (1/4): Active Learning

01 Motivation
02 Methodology
03 Evaluations
04 Conclusion

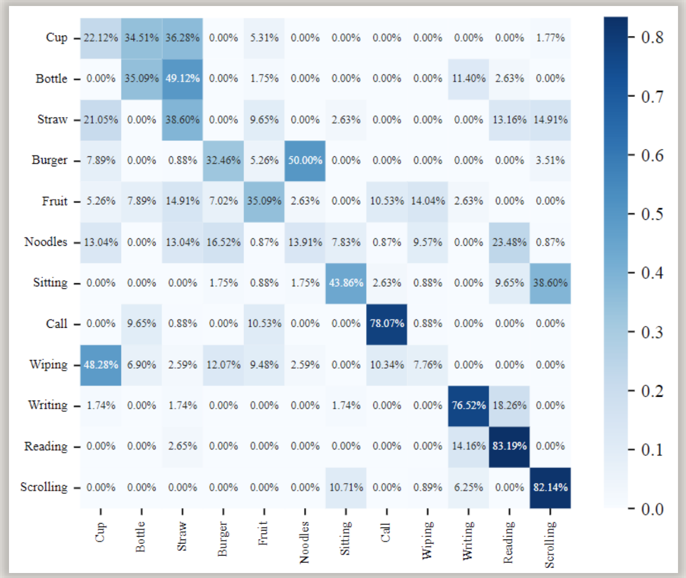


Method	Acc.	AUC vs. Baseline
Baseline (Without Active Learning)	37.94%	–
Random Query	89.55%	+66.93%
Least Confidence	87.21%	+79.02%
Margin Sampling	90.86%	+71.46%
Deviation of Predictions	83.04%	+68.27%
PALM (Entropy) (Single Prediction)	**91.08%**	**+87.66%**
PALM (Upper Bound)	**98.25%**	**+137.86%**

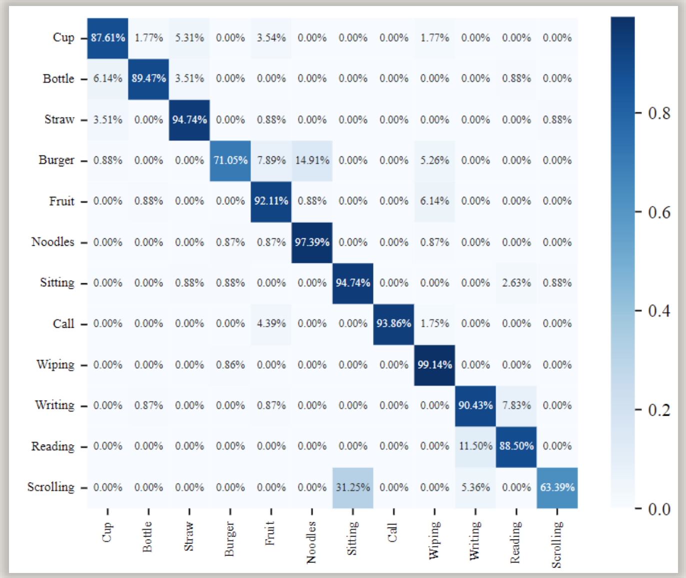
- Using Information Entropy, PALM outperformed all other active learning methods
- Not only the final accuracy is higher, but the green curve is constantly above other ones

PALM Results (1/4): Active Learning

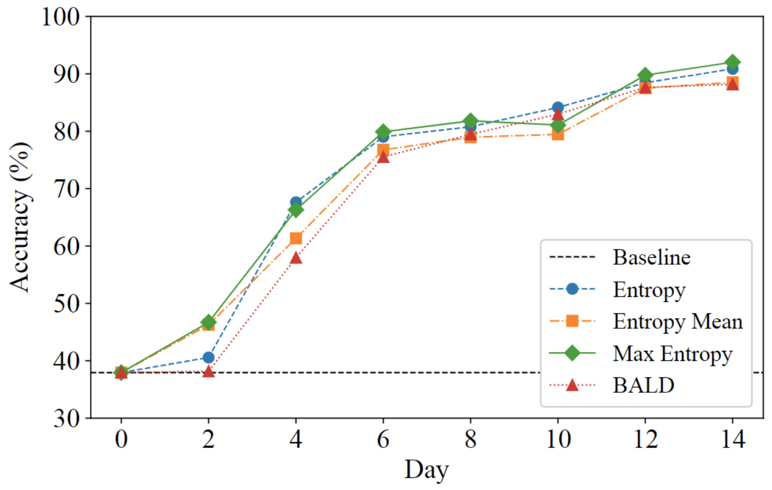
Before Active Learning



After Active Learning



PALM Results (2/4): Entropy-based

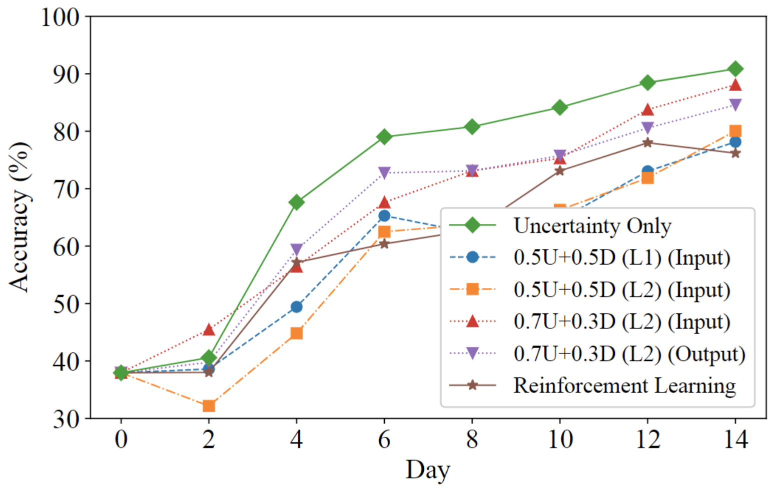


Method	Acc.	AUC vs. Baseline
Baseline (Without Active Learning)	37.94%	–
Entropy (Single Prediction)	91.08%	+87.66%
Entropy Mean (50 Predictions)	89.99%	+86.83%
Max Entropy (50 Predictions)	**92.03%**	**+90.85%**
BALD (50 Predictions)	89.25%	+80.49%

- Max Entropy slightly outperformed Entropy with single prediction
- Notably, these methods require 50x computational resources

Personalized Model

PALM Results (3/4): Diversity-based

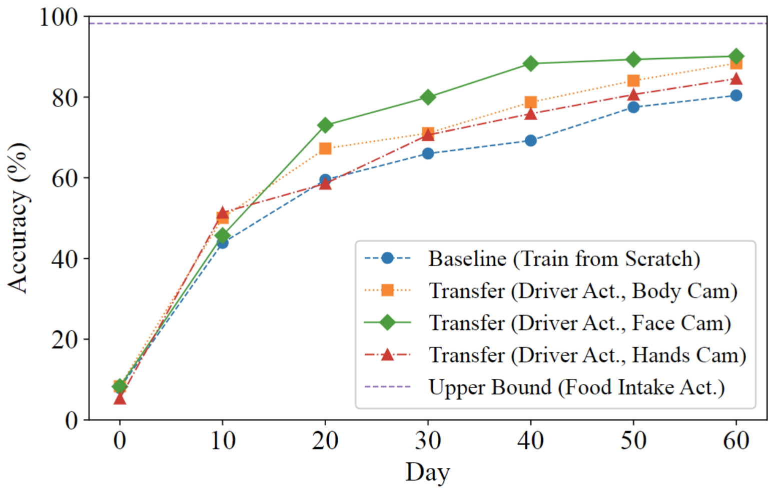


Method	Acc.	AUC
Uncertainty Only (Entropy, Single Prediction)	**91.08%**	–
0.5 Uncertainty + 0.5 Diversity (L1-norm) (Raw Input)	79.09%	-18.85%
0.5 Uncertainty + 0.5 Diversity (L2-norm) (Raw Input)	80.04%	-21.32%
0.7 Uncertainty + 0.3 Diversity (L2-norm) (Raw Input)	88.23%	-7.90%
0.7 Uncertainty + 0.3 Diversity (L2-norm) (LSTM Output)	85.23%	-8.14%
Reinforcement Learning (Highest Uncertainty or Diversity)	79.61%	-14.92%

- Different weights
- L1- vs. L2-norm
- Raw Input vs. LSTM Output

- None of the diversity-based methods outperform the uncertainty-based one

PALM Results (4/4): Transfer Learning



Method	Acc.	AUC vs. Baseline
Baseline (Train from Scratch)	80.99%	–
Transfer (Driver Activities, Body Cam)	90.06%	+10.29%
Transfer (Driver Activities, Face Cam)	**90.86%**	**+19.48%**
Transfer (Driver Activities, Hands Cam)	86.04%	+8.06%
Upper Bound (Food Intake Activities)	98.25%	+50.75%

- All three camera angles achieved higher accuracy than training from scratch
- Data with a similar camera angle performed better

Concluding Remark (1/2)

- We explored the potential of using **mmWave radars for HAR**
- We propose methods to train **resource-efficient personalized models**

1. DPR outperforms previous state-of-the-art voxelization-based methods
   - Increased accuracy by **4.10%**, achieving **99.66%**
   - Reduced memory consumption by **78.29%**
   - Reduced inference time by **69.64%**
2. PALM outperforms other active learning methods
   - Achieved an accuracy of **91.08%** over a two-week active learning period
   - Achieved an upper bound of **98.25%** over an extended period
   - AUC improvement of **+87.66%**, with an upper bound of **+137.86%**

Concluding Remark (2/2)

3. Among four entropy-based methods:
   - Max Entropy achieved the highest accuracy of **92.03%** and an AUC improvement of **+90.85%**
   - Higher computational cost due to multiple predictions
4. None of the diversity-based methods outperforms uncertainty-based
   - Shows the efficacy of the uncertainty-based methods
5. PALM can benefit from cross-application transfer learning
   - Increased accuracy by **9.87%**, achieving **90.86%**
   - Increased AUC by **+19.48%**

Future Directions

1. Enhance the **sample selection strategy** in active learning by exploring alternative methods of quantifying uncertainty and diversity
2. Apply PALM to **other domains** beyond food intake and driver activity recognition to identify domain-specific adaptations
3. Integrate PALM with **other sensors/data type** to address some limitations inherent to mmWave radar-based approaches
4. Explore **federated learning** for PALM to allow for privacy-preserving personalized models by enabling decentralized learning

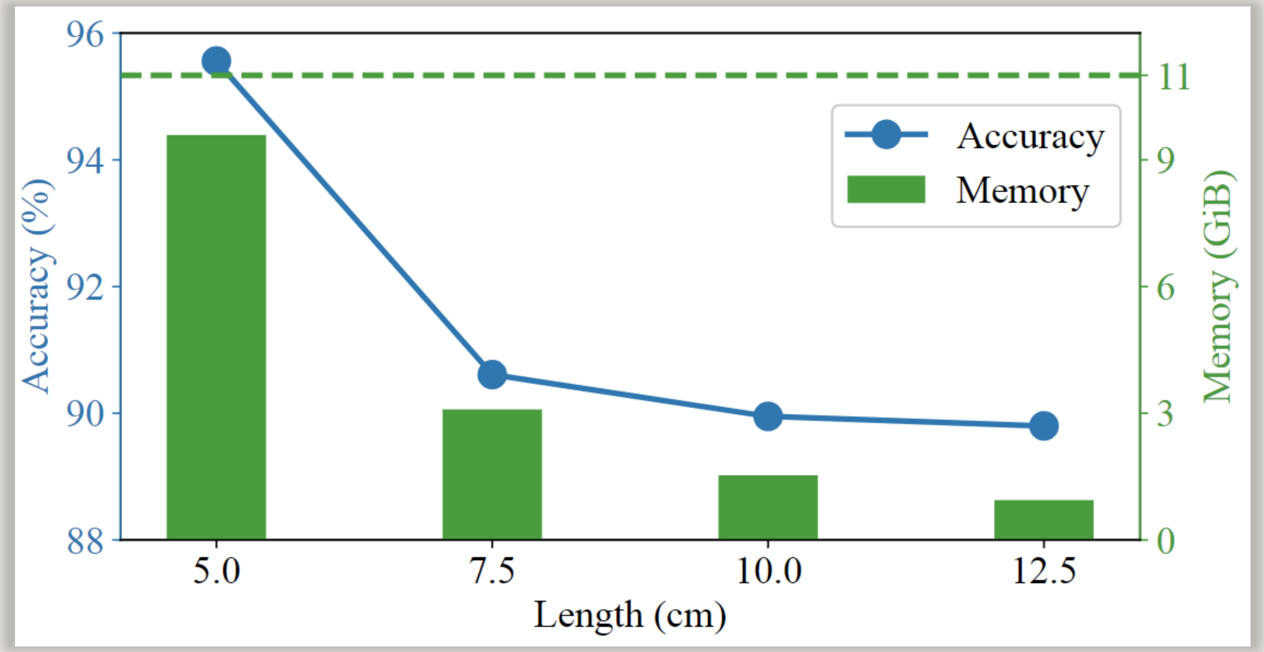
Thanks for Listening

Q&A Time

Problem 1

Resource Inefficiency of Voxelization

- Sparse Point Clouds → **Most voxels are empty** → Waste!



Query Strategy

- Select the **highest uncertainty** sample(s) for label querying
- Selecting the **top K uncertain samples** is possible
  - Labeling multiple samples at once can be intrusive and time-consuming, risking decreased user engagement
- We choose **K = 1** to reduce user burden
  - Keeps the process manageable and less disruptive
  - Promotes long-term participation and data quality