UNIVERSITY of HOUSTON
CULLEN COLLEGE of ENGINEERING

OneFi: One-Shot Recognition for Unseen Gesture via COTS WiFi

Rui Xiao, Jianwei Liu, Jinsong Han, Kui Ren

Sensys 2021

Presented by Chenpei Huang

ANTS LAB

Outline

- Introduction
- System Design
  - Data Collection
  - Virtual Gesture Generation
  - Few-shot Recognition
- Evaluation
- Conclusion

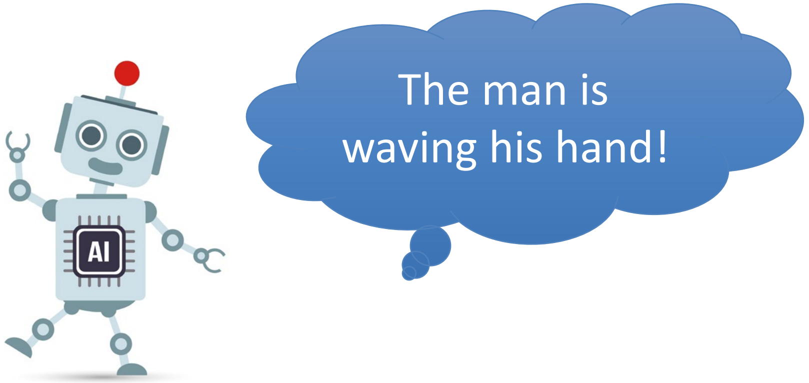
Introduction

Human Gesture Recognition

Virtual Reality

Medical Control

Smart Home Automation



Introduction

Traditional HGR solutions

- Use cameras to capture image or videos
- Wearable sensors

Drawbacks

- Leak user privacy (facial information)
- Inconvenient to user

WiFi-based HGR

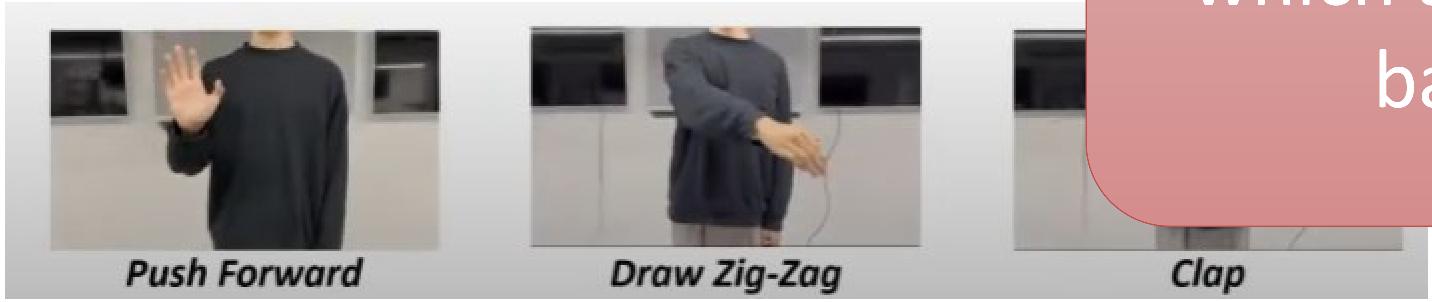
- No need to wear sensors
- Less intrusive to user privacy
- Ubiquitous

Introduction

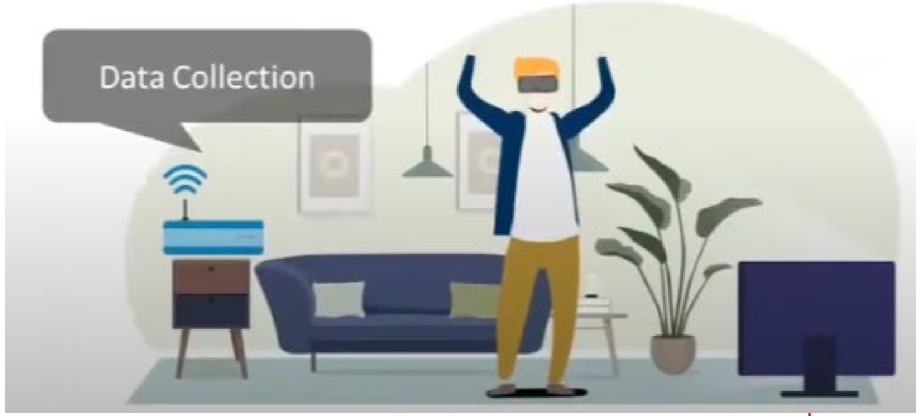
Supervised WiFi-based HGR

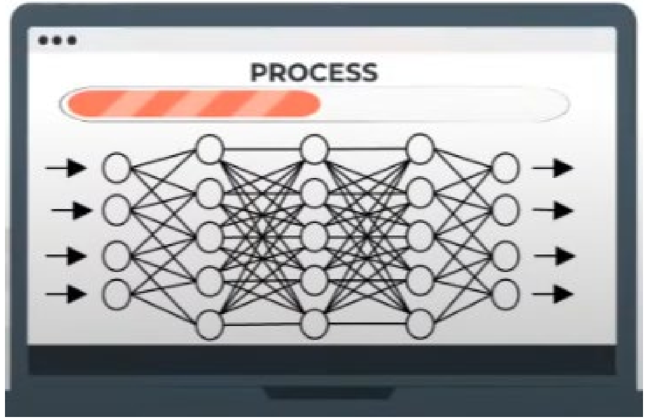
How about *unseen* gestures which are *not included* in base gestures?

- Predefine Base Gestures



- Collect data and train





Introduction

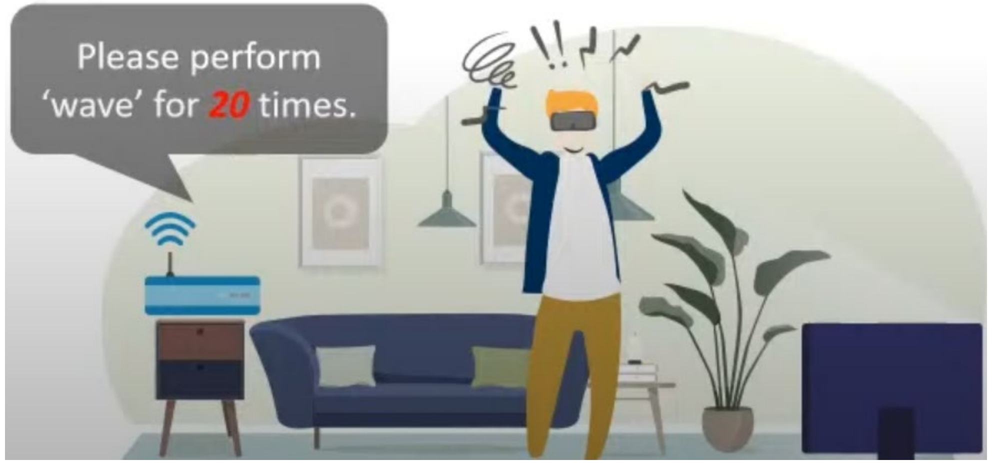
Unseen gestures are important.

- Predefined base gestures cannot keep up with ever-evolving demands.
- It's crucial to allow the user to adapt the system to their own preference.

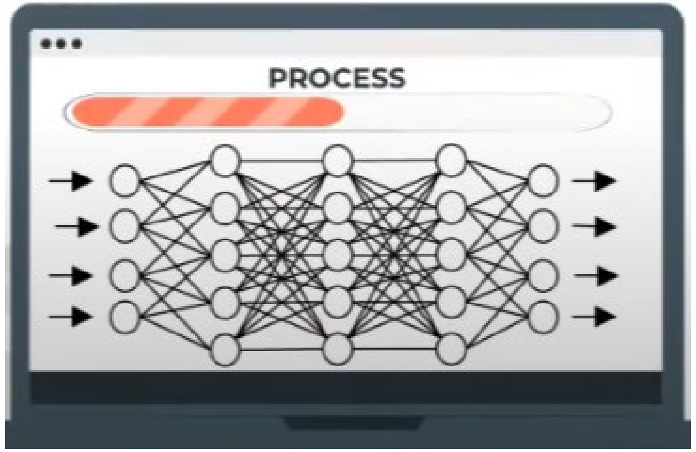
Overhead

Limited Scalability!

a) Data Collection Overhead.



b) Re-training Overhead.



Introduction

Problem Definition

1. Recognize a few base gestures.

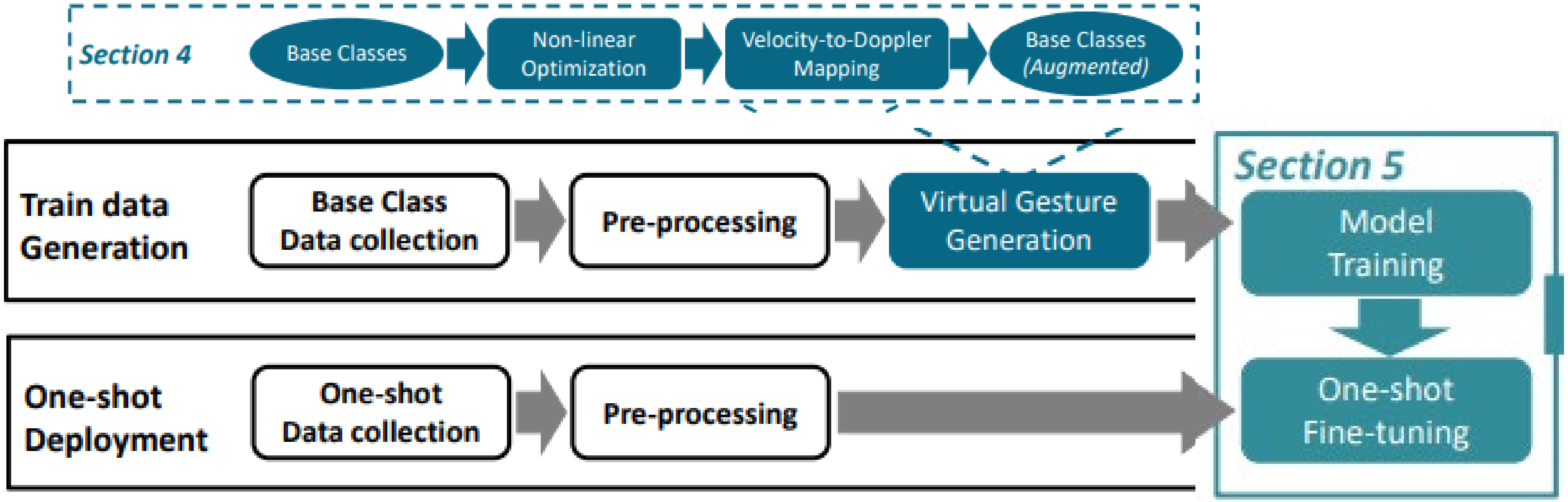
2. When introducing unseen gestures:
    - User only needs to collect one signal sample for any unseen gestures.
    - Model can fast adapt to new data without retraining the whole model.

Outline

- Introduction
- System Design
  - Data Collection/Pre-processing
  - Virtual Gesture Generation
  - Few-shot Recognition
- Evaluation
- Conclusion

System Design

Overview



1. Prepare training data
2. Prepare augmented training data
3. Prepare unseen one-shot data
4. Model Training & one-shot fine-tuning

System Design

Overview

1. Prepare training data
2. Prepare augmented training data
3. Prepare unseen one-shot data
4. Model Training & one-shot fine-tuning

+ self-attention-based backbone

Enrich the prior knowledge

Alleviate Training Overhead

System Design: Data Collection/Preprocessing

Data Collection: Channel State Information

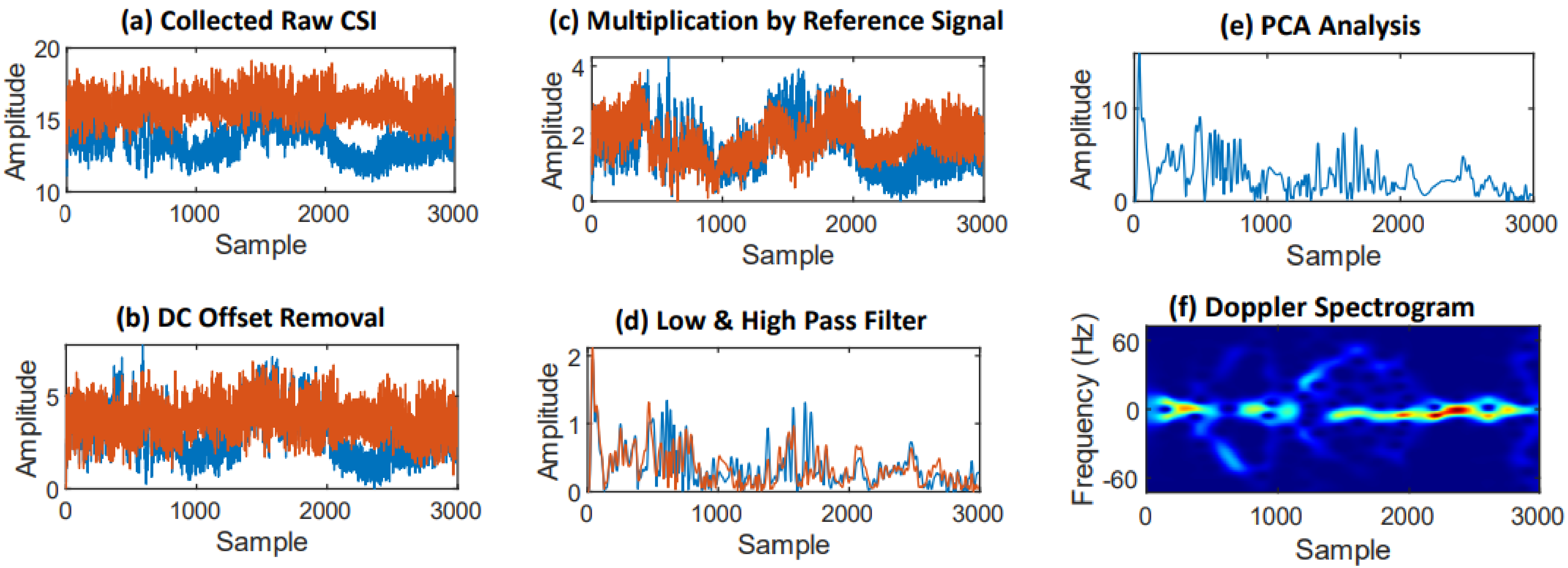
$$H(f_c, t) = \left(\sum_{k=1}^{K} \alpha_k(t) e^{-j2\pi f_c \tau_k(t)} \right) e^{j\epsilon(f_c, t)},$$

Doppler Effect

$$f_D(t) = -\frac{1}{\lambda} \frac{d}{dt} d(t), \quad \Longrightarrow \quad H(f_c, t) = \left(\overbrace{H_s(f_c)}^{\textbf{Static}} + \overbrace{\sum_{k \in P_{dn}} \alpha_k(f_c, t) e^{j2\pi \int_{-\infty}^{t} f_{D_k}(u) du}}^{\textbf{Dynamic: due to gesture}} \right) e^{j\epsilon(f_c, t)}$$

System Design: Data Collection/Preprocessing

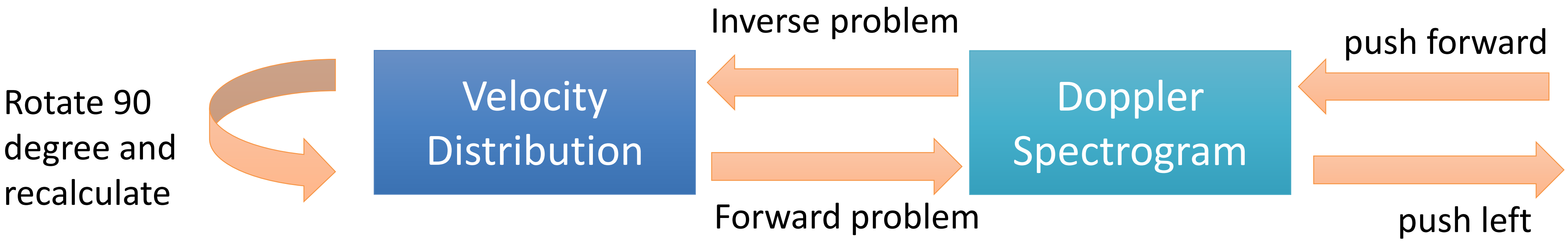
Workflow of data pre-processing



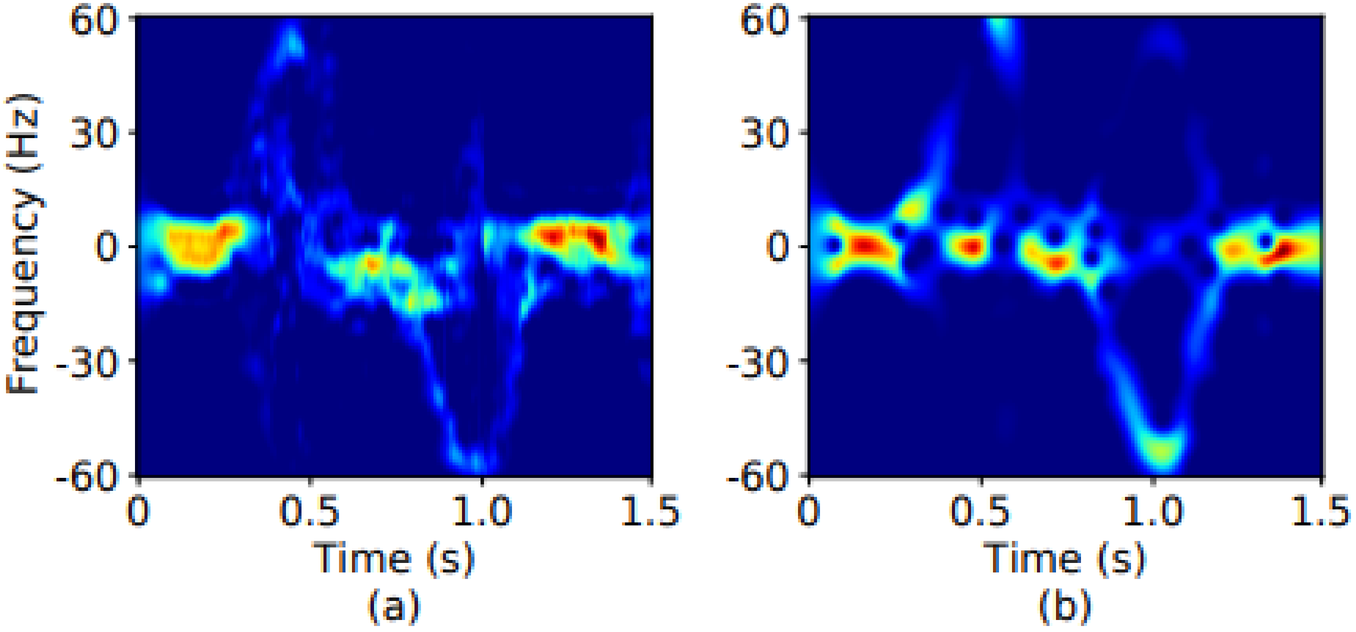
Virtual Gesture Generation

- Goal: use signal model to generate data from existing examples

Example: convert “push forward” to “push left”



Virtual Gesture Generation



(a) Virtual gesture (b) and real data

- Expand the base dataset by 12 times and improve the accuracy

Few-shot recognition

Basics

- In few-shot learning context
  - Base **training set**
  - Few-shot set is called **support set**
  - Testing set is called **query set**

$$\theta^*, W^*, b^* = \underset{\theta, W, b}{\arg\min}\, \mathcal{L}(\mathcal{D}; \theta, W, b)$$
$$= \underset{\theta, W, b}{\arg\min} \sum_{(\mathbf{x}, y) \in \mathcal{D}} -\log\left(W^T f_\theta(y|\mathbf{x}) + b\right).$$

$$s_j = \frac{f_\theta(\mathbf{x}) \cdot w_j}{\|f_\theta(\mathbf{x})\| \|w_j\|}.$$

Stages

- Train a feature extractor on **training set** (transformer backbone + cross-entropy loss)
- Fine-tune the classifier using **support samples**
- **Query samples x** are classified based on cosine similarity

WiFi Transformer

Why transformer?

- Doppler spectrogram is essentially a sequence data
- Self-attention can capture long-range interactions
- Parallel structure (no recurrent) saves training time

WiFi Transformer

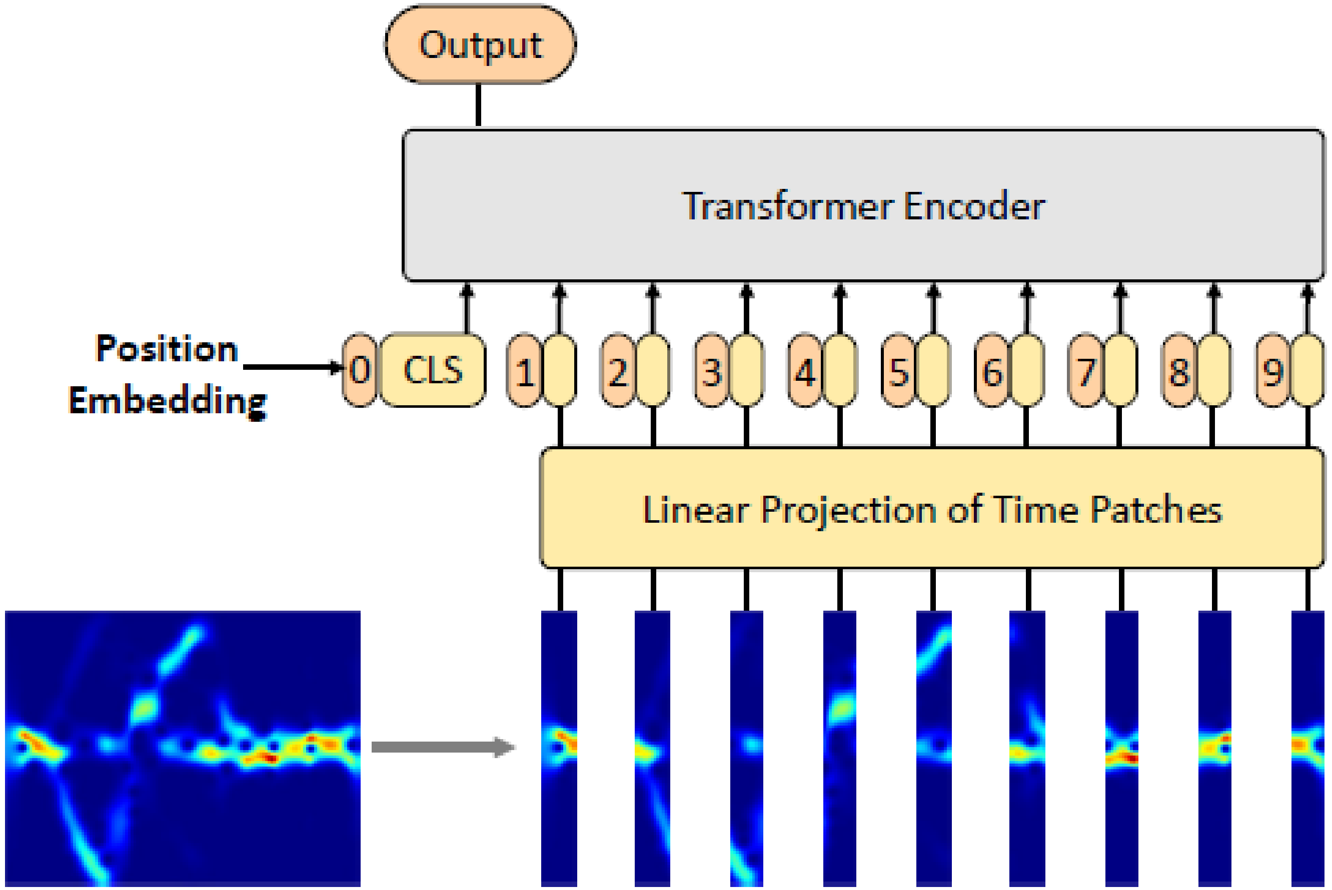
a) Model input: Doppler spectrum;

b) Position embedding: learnable vector to retain position information.

c) Multi-head self-attention block:

$$z = Attention(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d/h}}\right)V.$$

d) Model output: The first output learns the representation of whole sequence



Outline

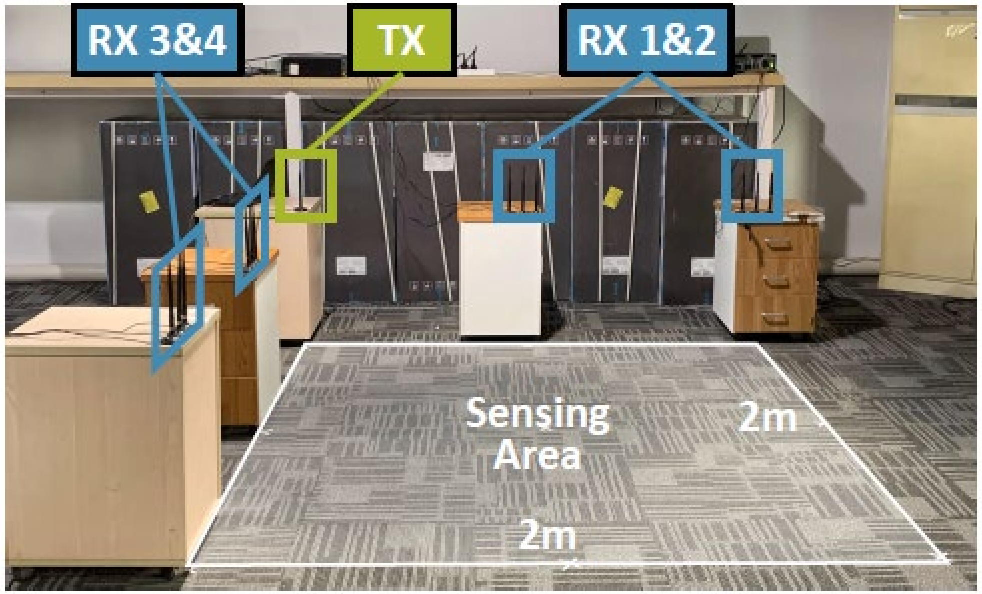
- Introduction
- System Design
  - Data Collection/Pre-processing
  - Virtual Gesture Generation
  - Few-shot Recognition
- Evaluation
- Conclusion

Evaluation

Experiment setup: COTS WiFi devices + 802.11 CSI Tool

Dataset: 2900 samples, 40 classes, 6 unseen classes

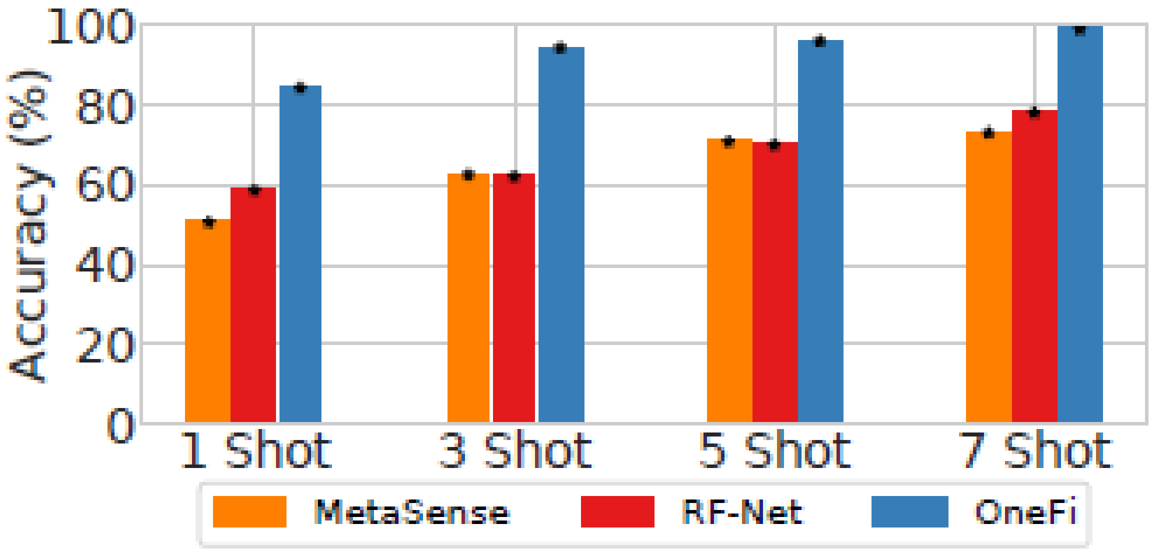
Metric: classification accuracy



Evaluation

Overall Accuracy

The accuracy of OneFi in 1/3/5/7 shot settings is 84.2%, 94.2%, 95.8%, and 98.8%, respectively.



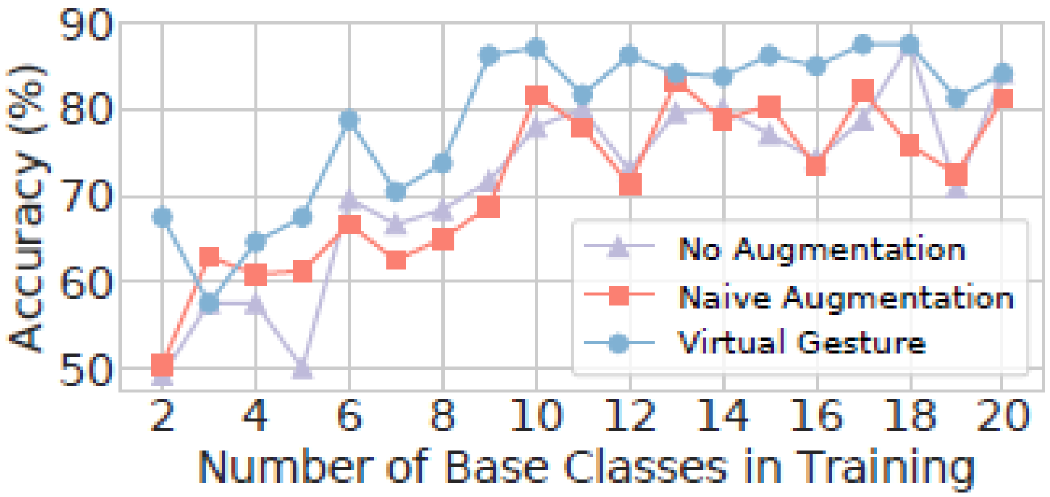
Reasons

- Few-shot framework
- Transformer backbone
- Data augmentation
- Hardware settings e.g., packet rates

Evaluation

Effect of Virtual Gestures (Data Augmentation)

Baseline: 1)without using data augmentation; 2) a naïve augmentation (add noise)



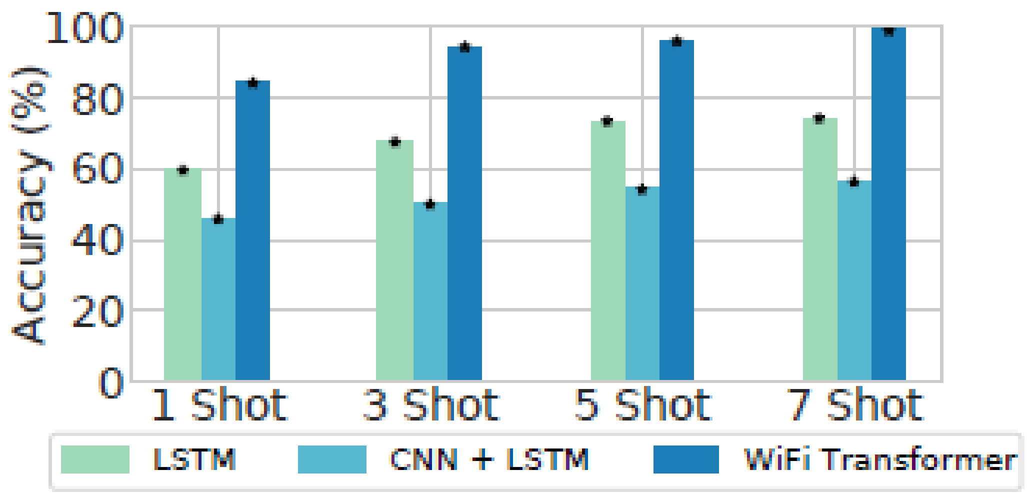
Result

- For gesture recognition, adding noise shows no remarkable improvement
- Virtual gestures is effective in improving the accuracy

Evaluation

Effect of Proposed Backbone

Compare with LSTM and LSTM+CNN



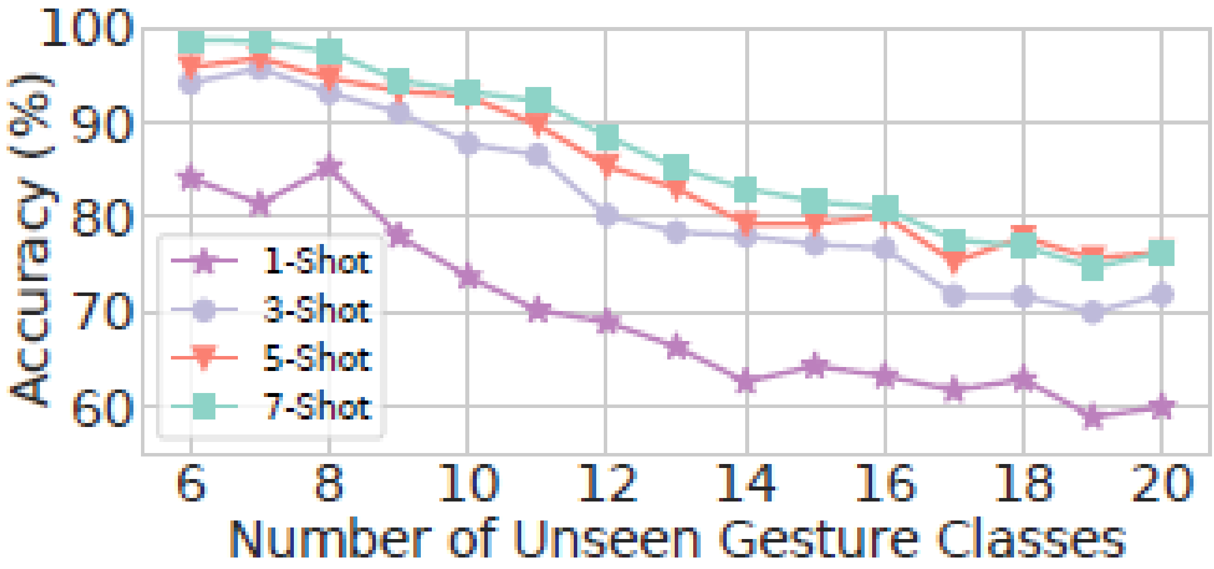
Result

- Transformer model outperforms in all settings
- CNN+LSTM can hardly converge

Evaluation

Impact of Number of Unseen Gestures

Vary the number of unseen classes from 6 to 20.



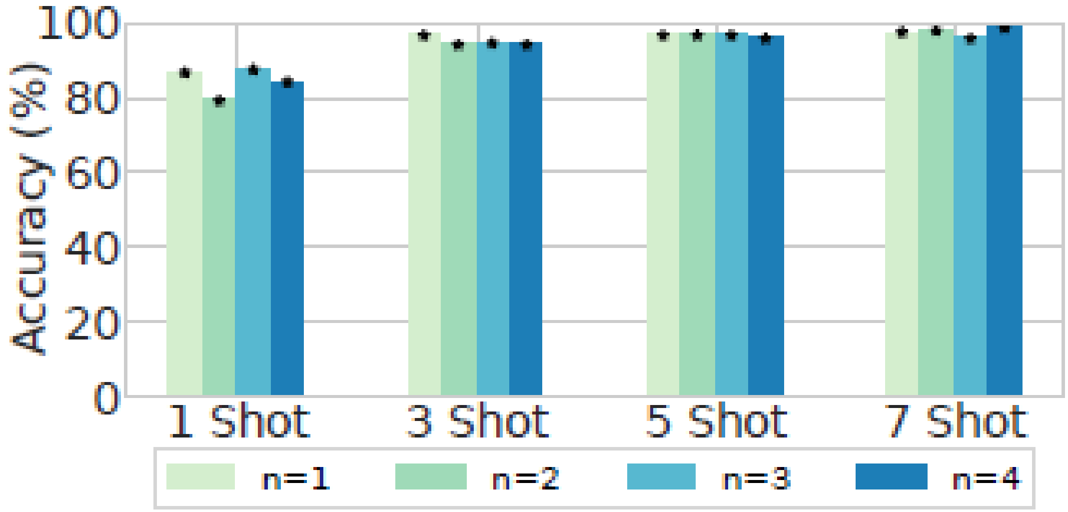
Result

- Performance decrease as the number of unseen classes
- Add more shots can achieve an accuracy >70%

Evaluation

Other evaluations

Impact of the number of receiver



Important in training, not important in inferring.

- Cross-environment
- Cross-orientation
- Cross-person
- Cross-location

Outline

- Introduction
- System Design
  - Data Collection/Pre-processing
  - Virtual Gesture Generation
  - Few-shot Recognition
- Evaluation
- Conclusion

Conclusion

- They propose OneFi, a one-shot HGR system to recognize unseen gestures using COTS WiFi.
- They propose virtual gestures and few-shot learning framework to mitigate extra effort in both data collection and model training.
- Such method achieve a high recognition accuracy in various settings, which is a promising step towards practical wireless human-computer interface.

THANK YOU

UNIVERSITY of HOUSTON | ENGINEERING