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

Rui Xiao, Jianwei Liu, Jinsong Han, Kui Ren

SenSys 2021

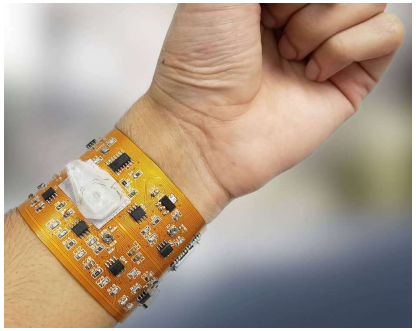
Human Gesture Recognition (HGR)

Virtual Reality

Medical Control

Smart Home

- **WiFi**-based solution:
  - No need to wear sensors
  - Less intrusive to user privacy
  - And also ubiquitous



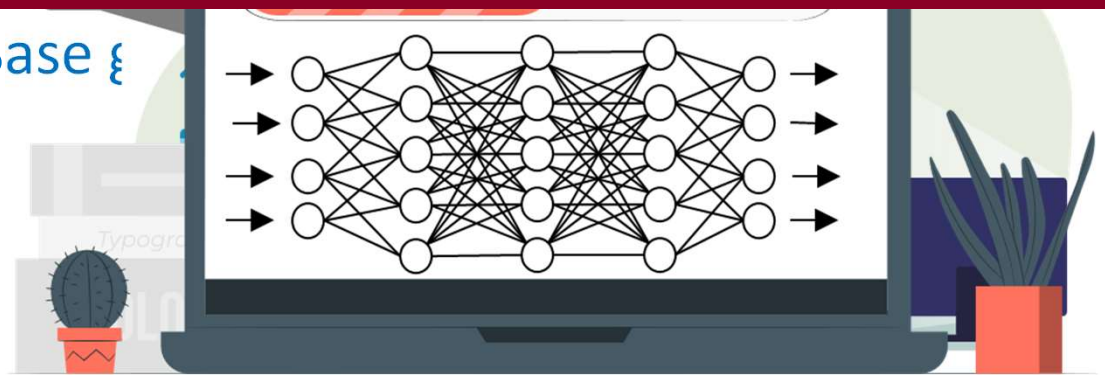
Uncomfortable

Privacy Leakage

WiFi-based HGR

- Works under supervised learning scheme: Three phases
  - Predefine base gestures

Question: What if now we want to recognize ***unseen gestures*** which is ***not included*** in base gestures?



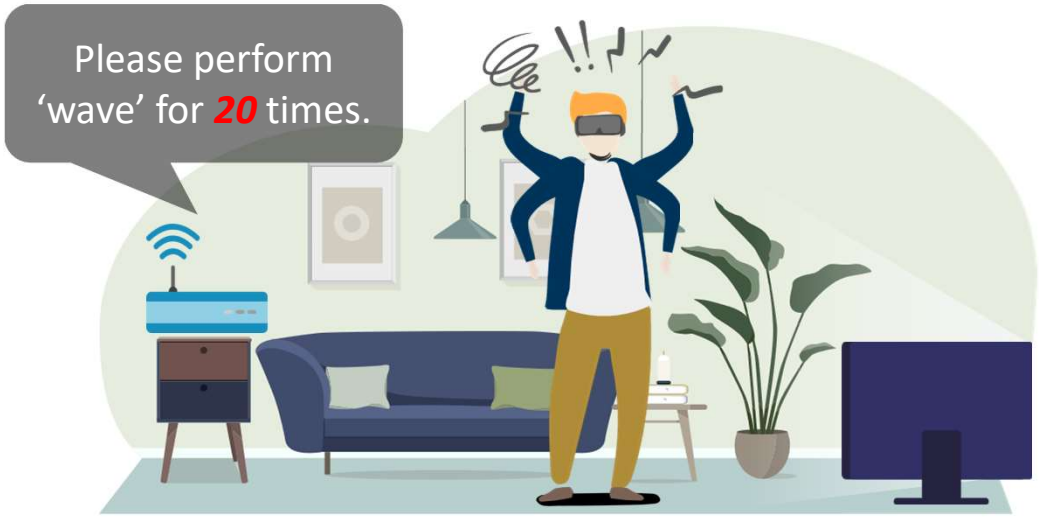
Unseen gestures are important.

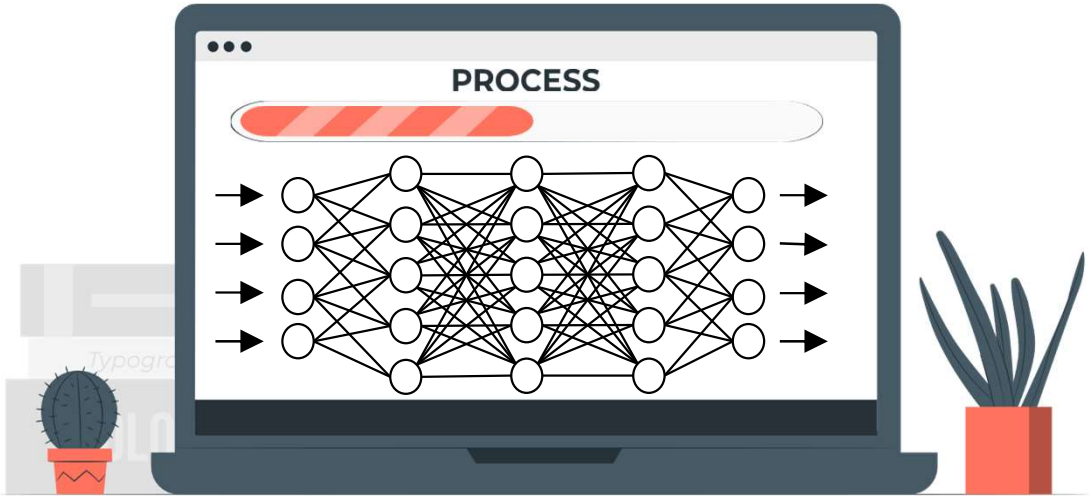
Limited Scalability

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

However, in existing systems, to recognize unseen gestures …

- Collect unseen gesture samples — Data Collection Overhead
- Retrain the whole model — Training Overhead





Problem Definition

Assume that our system can recognize a few base gestures.

When introducing unseen gestures, is it possible that:

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

In the literature, this problem falls into the category of ***few shot learning*** -- Learning with a few samples.

Few shot learning

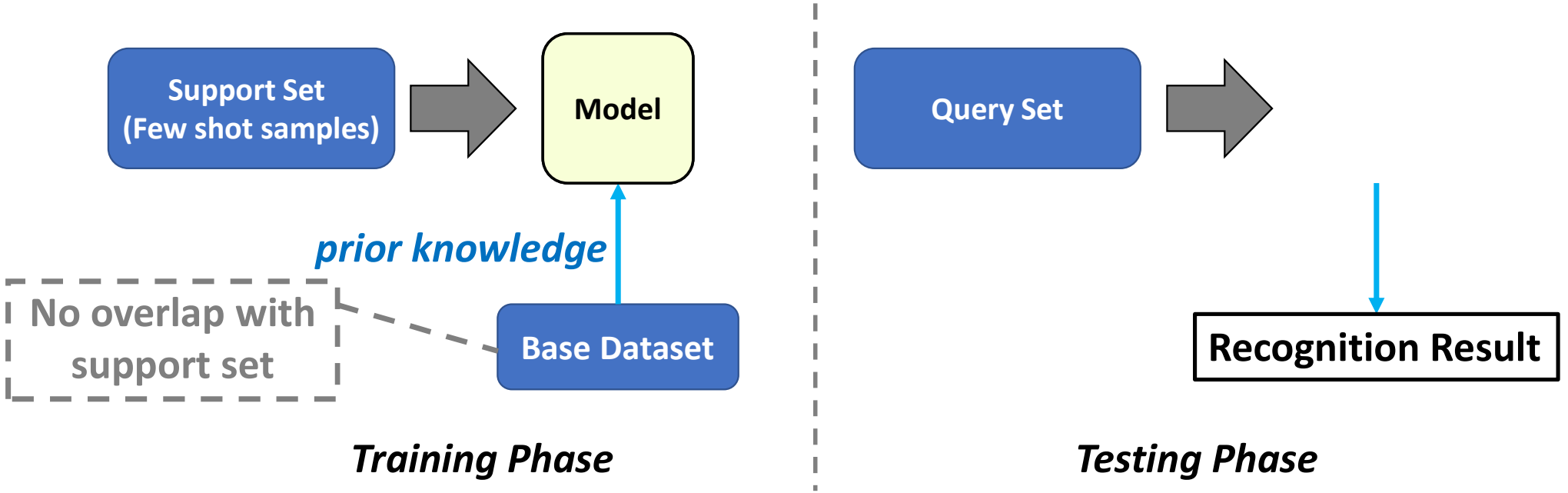
One shot is enough.

Few shot learning is easy for human. Why?

Because human beings have a large amount of ***prior knowledge***.

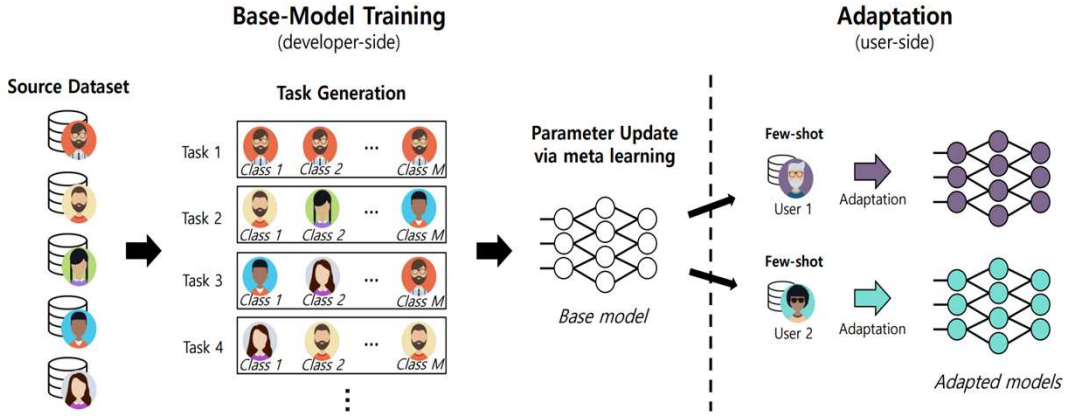
Few shot learning: Basics

- In few shot learning context
  - Training set is also called **support set**
  - Testing set is also called **query set**

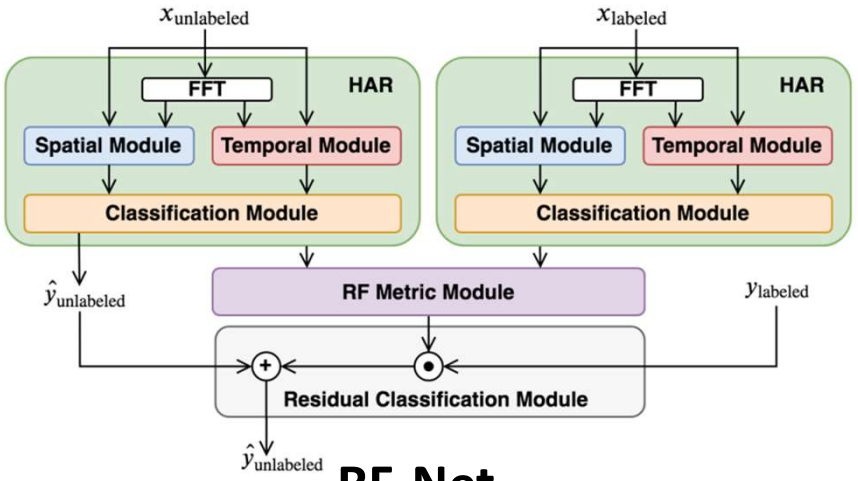


Existing works: *Meta-learning* as few shot solution

- MetaSense (Gong et al, SenSys'19)
  - New user and new device.
- RF-Net (Ding et al, SenSys'20)
  - New environments.



MetaSense

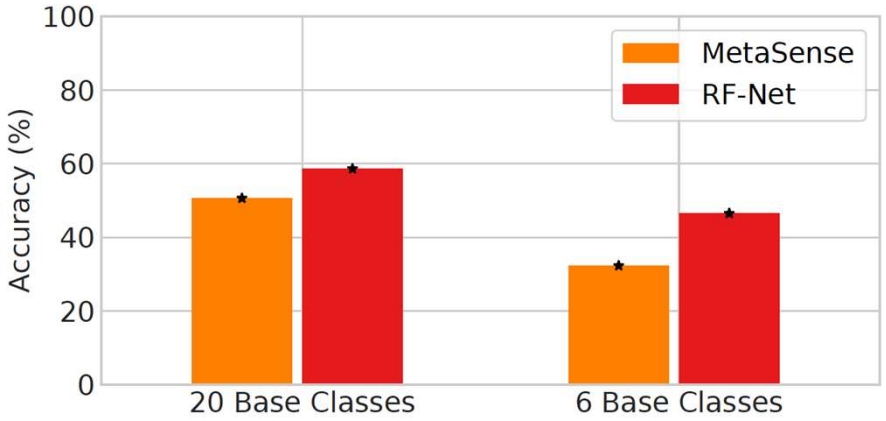


RF-Net

Challenge 1: Insufficient Prior Knowledge

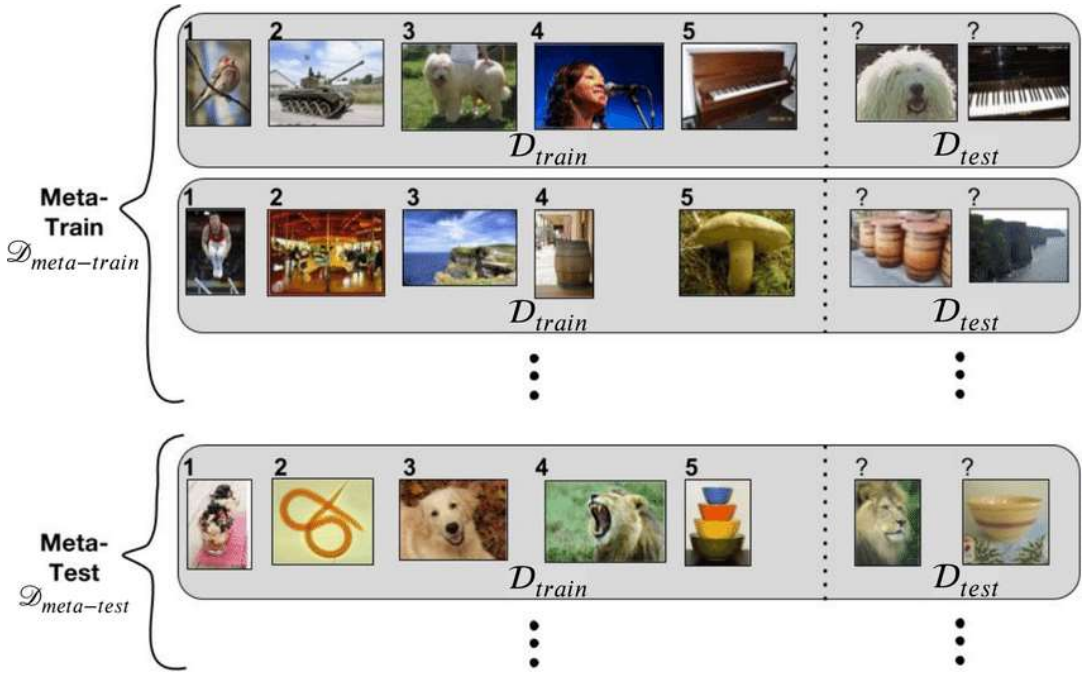
- The **base dataset** in WiFi HGR system is usually **small**.
- Therefore, the **prior knowledge** is usually **insufficient**.

We use 20 gestures as base classes and apply meta-learning scheme, the result of one shot recognition for unseen gestures is unacceptable:



Challenge 2: Complicated Training Process

- Meta-learning, known as 'learning to learn', **multiplies the complexity of the regular training process**.



Unseen Gesture Training

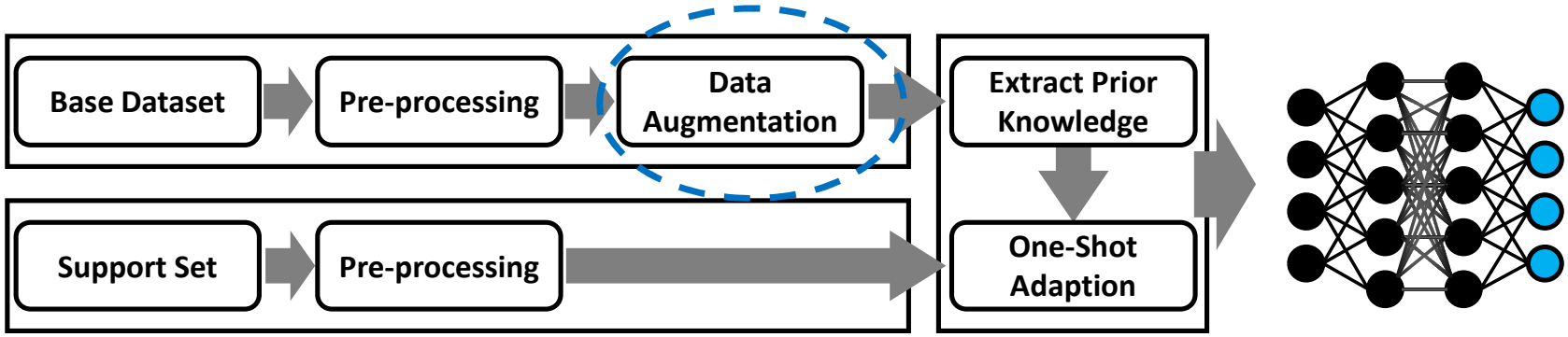
Training Overhead

Extract Prior Knowledge from Base Classes

Our Solution -- OneFi

➢ Overview:

- A data augmentation method based on signal modeling → **Enrich the prior knowledge**

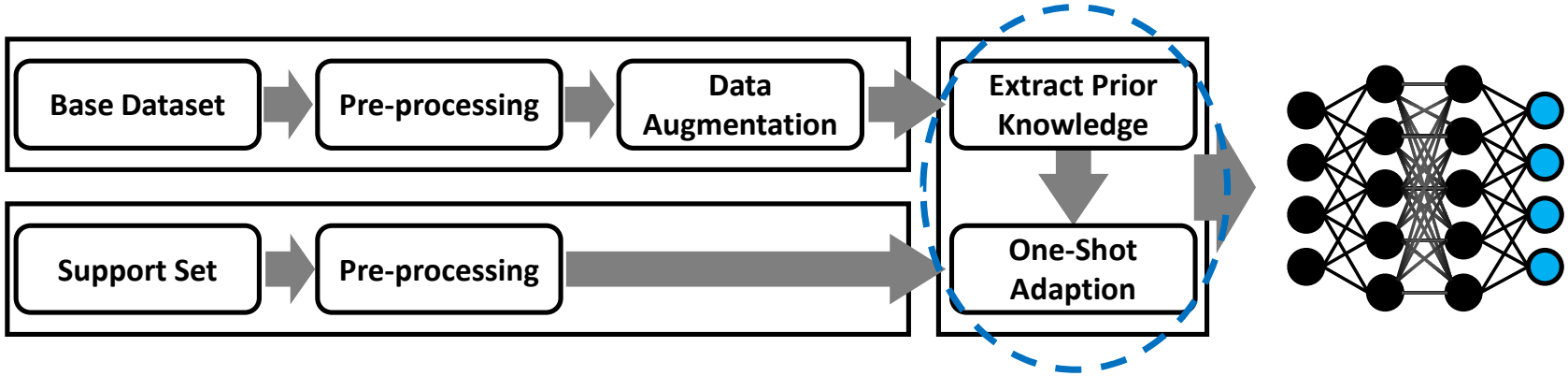


Our Solution -- OneFi

➢ Overview:

- A data augmentation method based on signal modeling → **Enrich the prior knowledge**
- A similarity-based one-shot learning framework
- A self-attention-based backbone

→ **Alleviate Training Overhead**

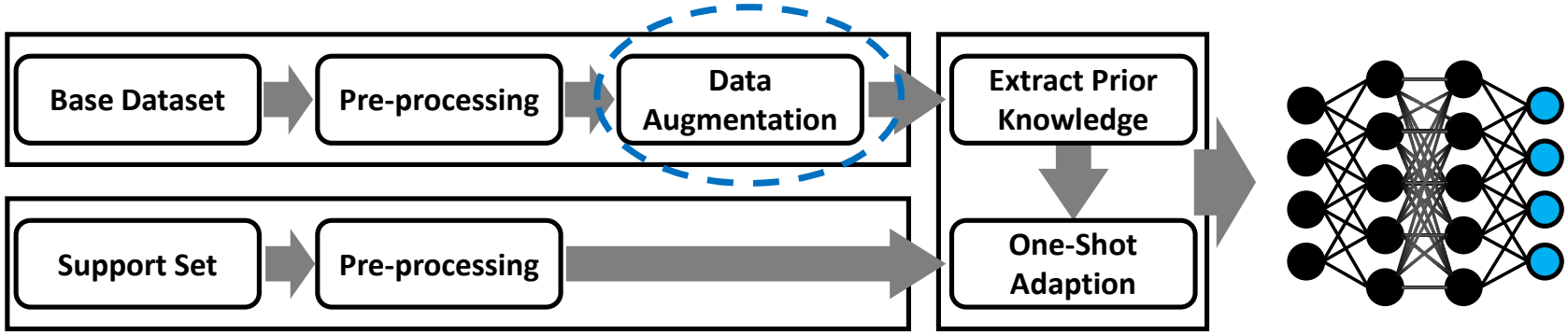


Our Solution -- OneFi

➢ Overview:

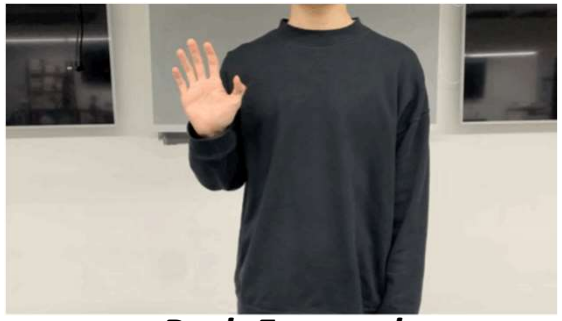
- A data augmentation method based on signal modeling → **Enrich the prior knowledge**
- A similarity-based one-shot learning framework
- A self-attention-based backbone

→ **Alleviate Training Overhead**

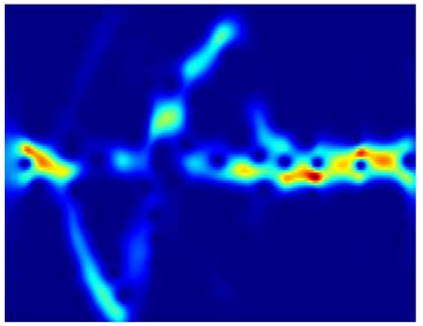


Data Augmentation – Generate Virtual Gestures

- Generates *additional, synthetic data* by signal modeling
- The intuition here is that we could generate a ‘push left’ gesture by rotating a ‘push forward’ gesture.



Push Forward



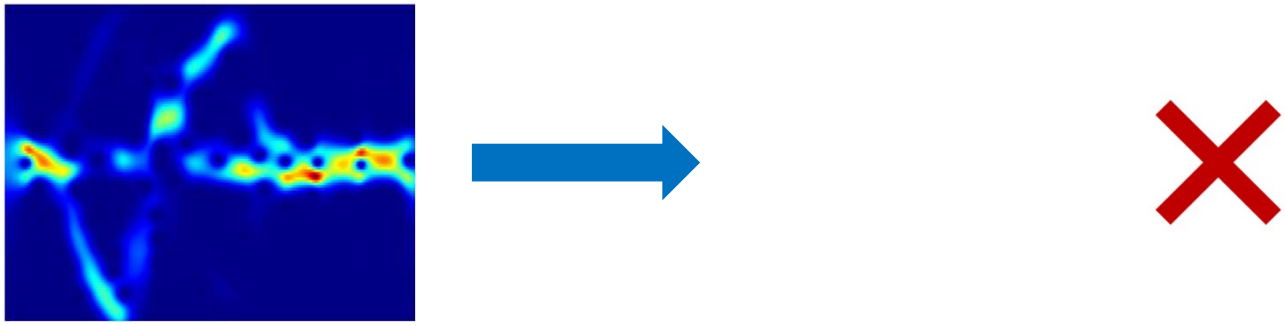
Doppler Spectrogram

Push Left

?

Data Augmentation – Generate Virtual Gestures

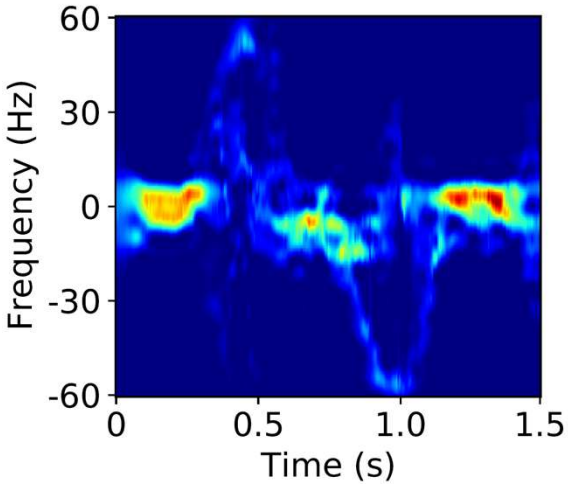
- Question: How to rotate a gesture?



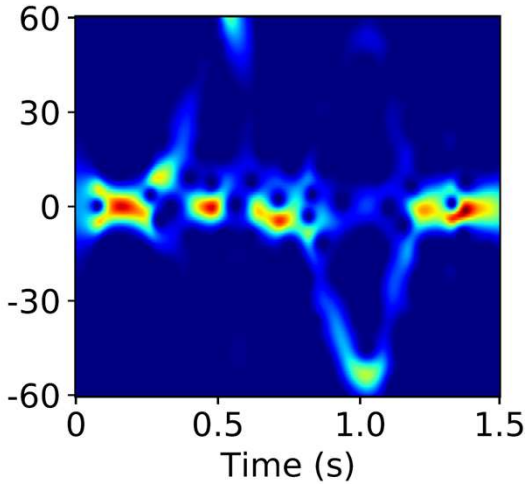
- **Signal modeling** is required!
- We compute the **velocity distribution** of the gesture, and apply **rotation** on the **velocity domain**.

Data Augmentation – Generate Virtual Gestures

- We compute the **velocity distribution** of the gesture, and apply **rotation** on the **velocity domain**.



virtual gesture



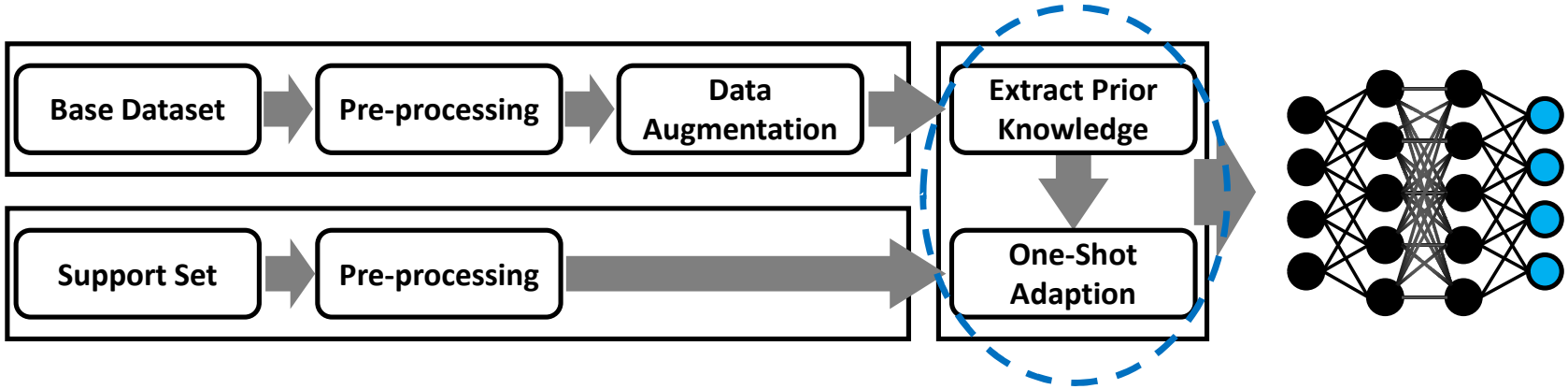
real gesture

- Virtual gesture is nearly identical to real gesture.
- Creating virtual gesture can enrich the prior knowledge.

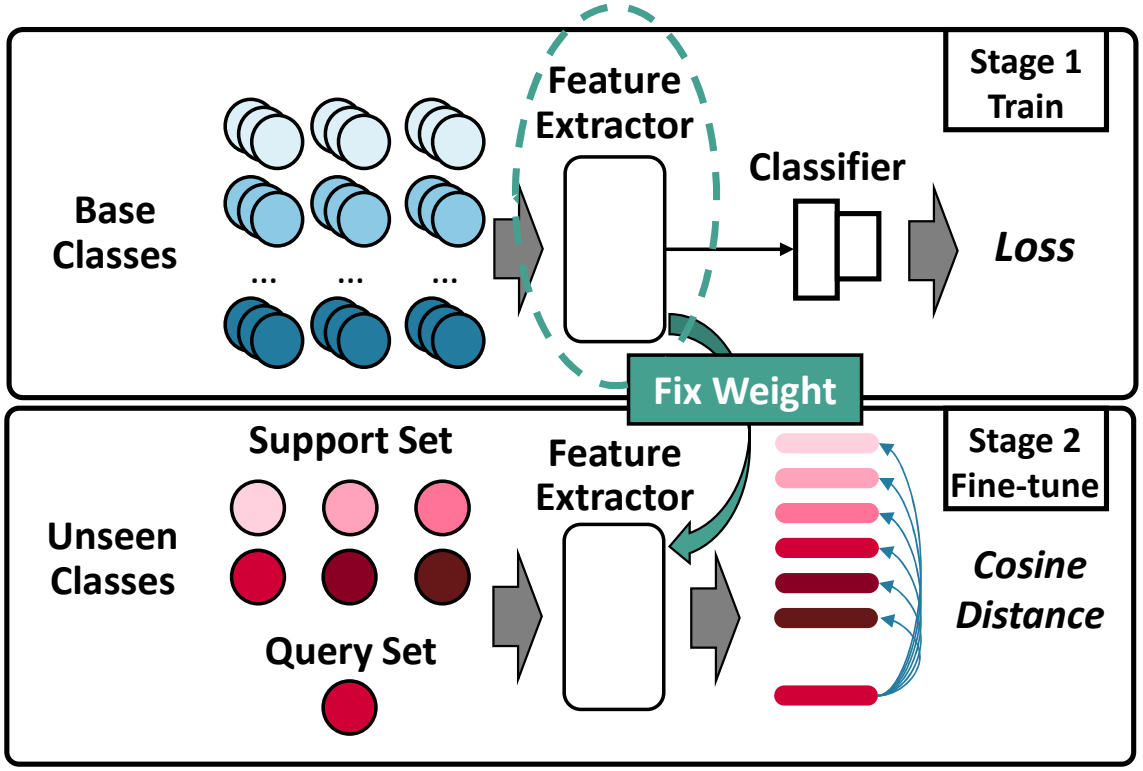
Our Solution -- OneFi

➢ Overview:

- A data augmentation method based on signal modeling
- A similarity-based one-shot learning framework → **Alleviate Training Overhead**
- A self-attention-based backbone



Similarity-based one-shot learning framework



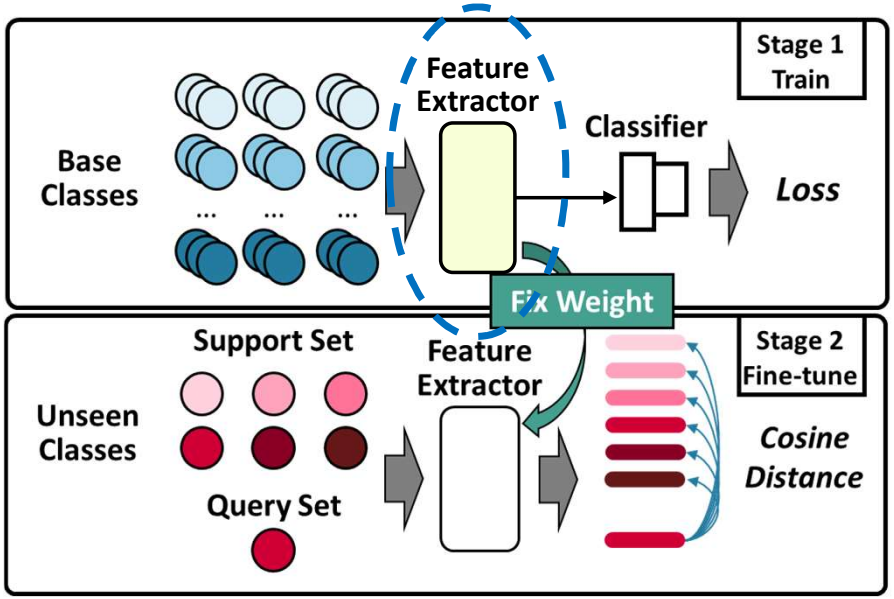
1. Encode Prior knowledge into the feature extractor
2. Finetune a new classifier based on the support set

- ***Lightweight*** one-shot learning framework
- Alleviating the training overhead

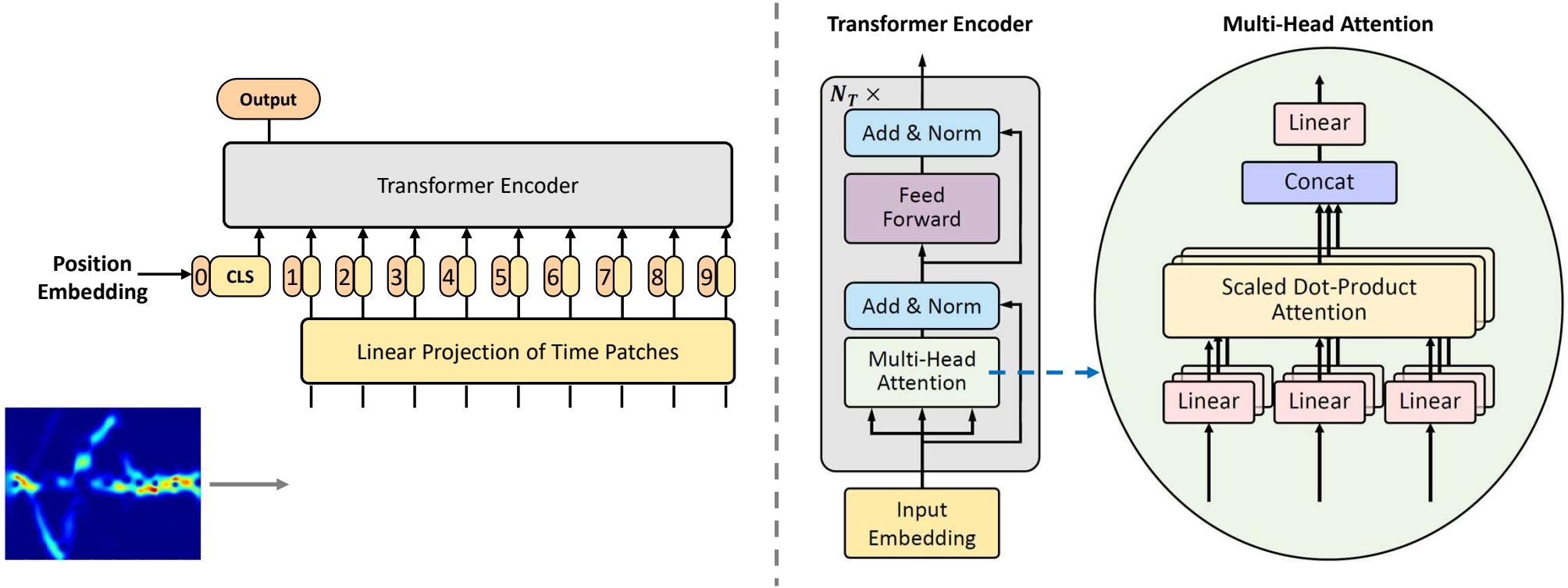
Our Solution -- OneFi

➢ Overview:

- A data augmentation method based on signal modeling
- A similarity-based one-shot learning framework
- A self-attention-based backbone → **Alleviate Training Overhead**



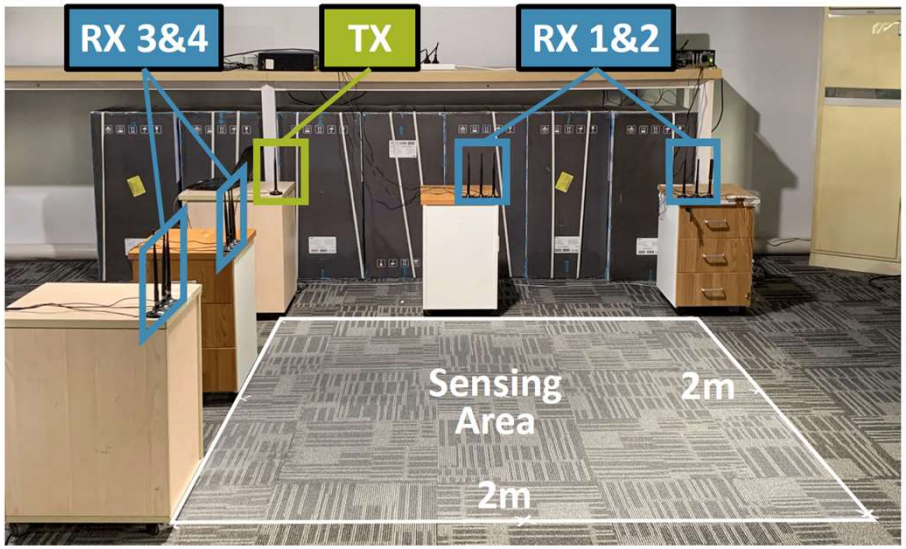
WiFi Transformer: A self-attention-based backbone



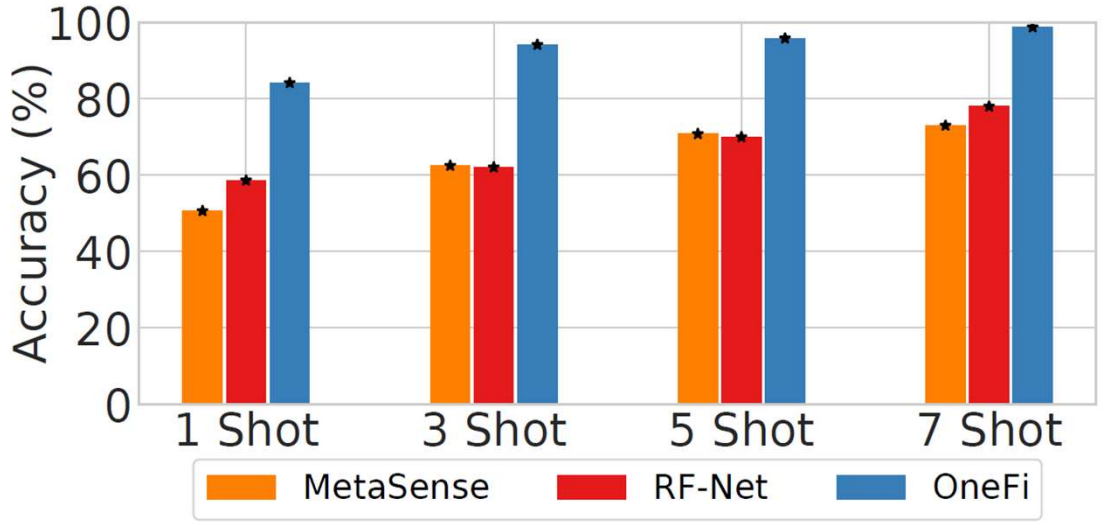
- Only needs ***O(1) sequential operations***
- Reduces the overhead of training the feature extractor

Evaluation Setup

- Implementation
  - Desktops with Intel 5300 NIC
- Collect 40 difference gesture classes.
  - 20 as base dataset
  - 20 as unseen dataset

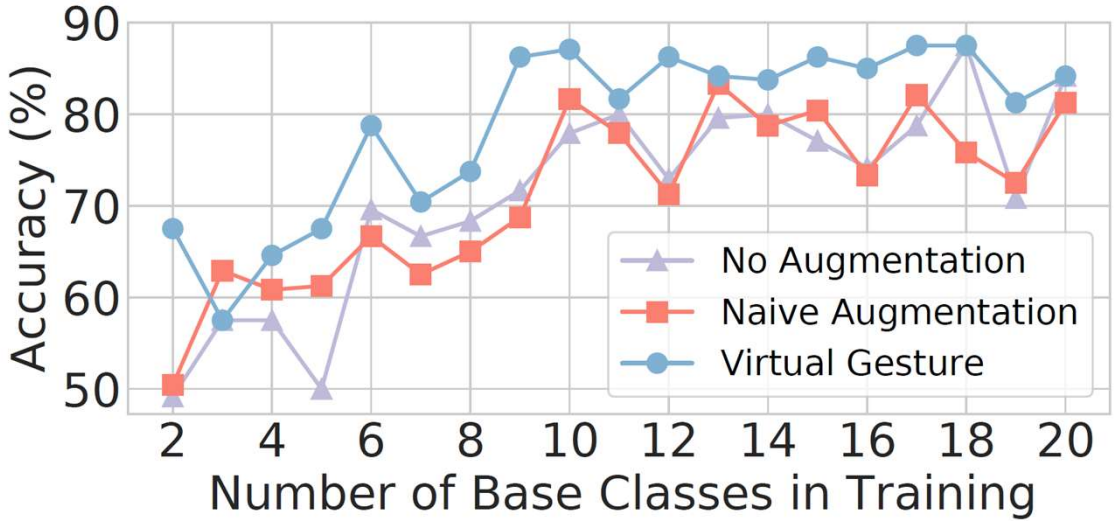


Overall Accuracy



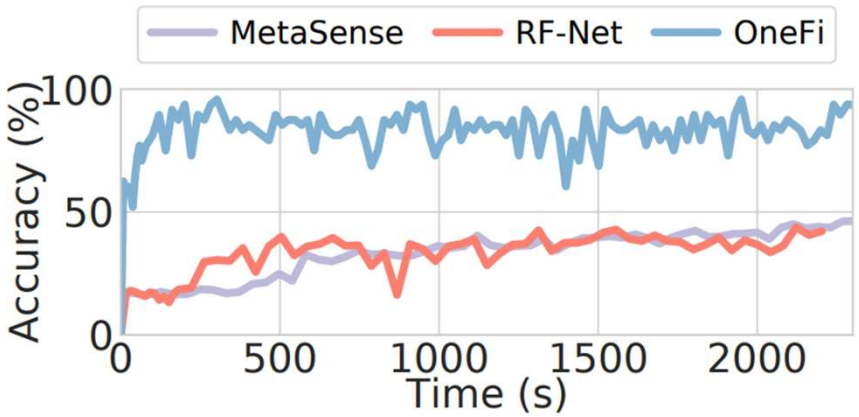
- Six-unseen-gesture classification
- Accuracy for *OneFi*: 84.2%, 94.2%, 95.8% and 98.8% with 1/3/5/7 shots
- Outperform *MetaSense* and *RF-Net* greatly

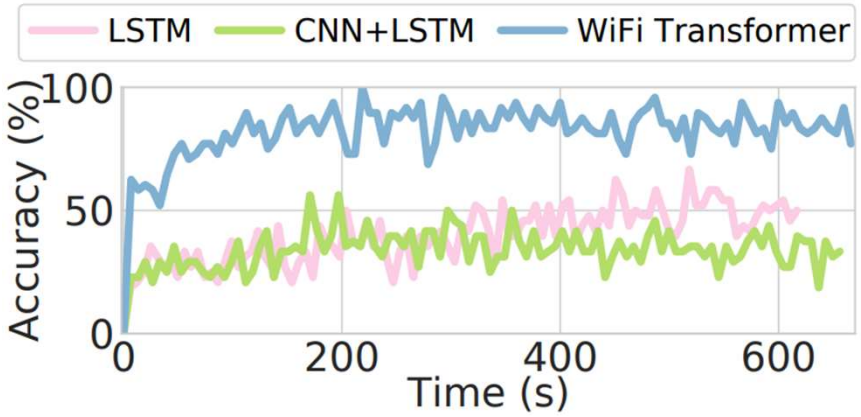
Effect of Virtual Gestures



- Virtual gesture generation is effective in enriching the prior knowledge and improving few-shot recognition accuracy.

Training Overhead

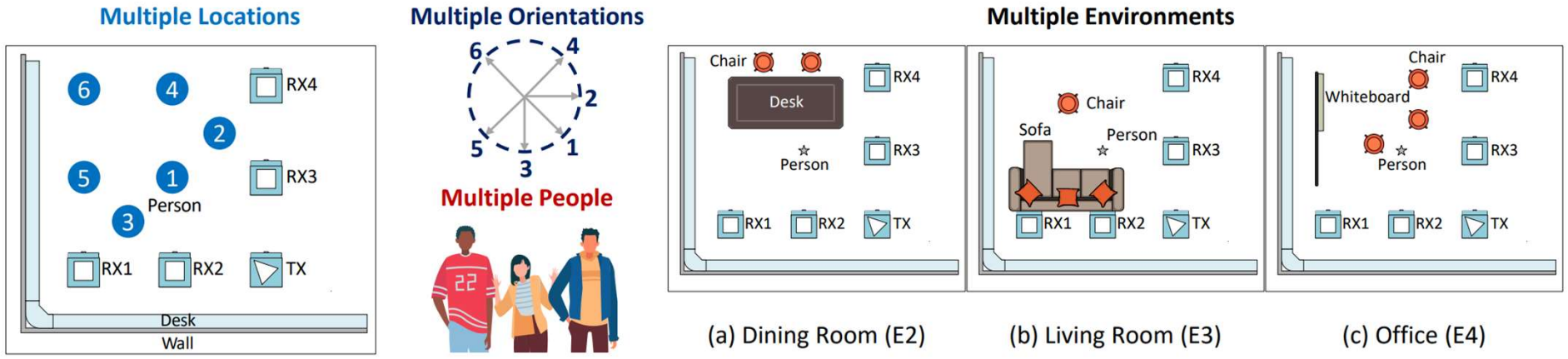




- *OneFi* entails significantly less training overhead.
- WiFi Transformer help to relieve the training overhead.

Summary of other evaluation results

- By **adding more shots**, *OneFi* can achieve high accuracy in **20-unseen-gesture** classification.
- *OneFi* works well with **small number of WiFi receivers**.
- *OneFi* generalizes in **cross-domain** experiments.



Conclusion

- We design OneFi, a **one-shot** WiFi HGR system. Extensive experiments demonstrate its strong ability to **recognize unseen gestures**.
- We present a novel **virtual gesture generation** technique that significantly **enriches the prior knowledge**.
- We propose a lightweight **one-shot learning framework**, together with a **self-attention based backbone** to **alleviate training overhead**.

Thanks! & Questions?