COMP3516: Data Analytics for IoT

Lecture 09:
Deep Wireless Sensing

Chenshu Wu

Department of Computer Science

2025 Spring

香港大學
THE UNIVERSITY OF HONG KONG

HKU AIoT LAB
香港大學人工智能物联网實驗室

Contents

- Framework
- Data
  - Data Representation
  - Data Availability
  - Data Augmentation/Synthesis
- Feature
- Model

WiFi Sensing: A 10-year Journey

- SP-based: Geometrical Approaches (Reflection Model)
- SP-based: Statistical Approaches (Scattering Model)
- DL-based: Deep Wireless Sensing (Neural Network Model)
  - Wireless Data for Learning
  - Model Design for Wireless Data

Why (Not) Deep Learning?

- Why Not
    - Efficient
    - Explainable
    - Deployable (on IoT)
    - Data
- Why
    - Enabling more applications that are difficult to achieve with SP alone

Deep Wireless Sensing

- We will use WiFi signals as example in this lecture
- mmWave and other RF signals are similar

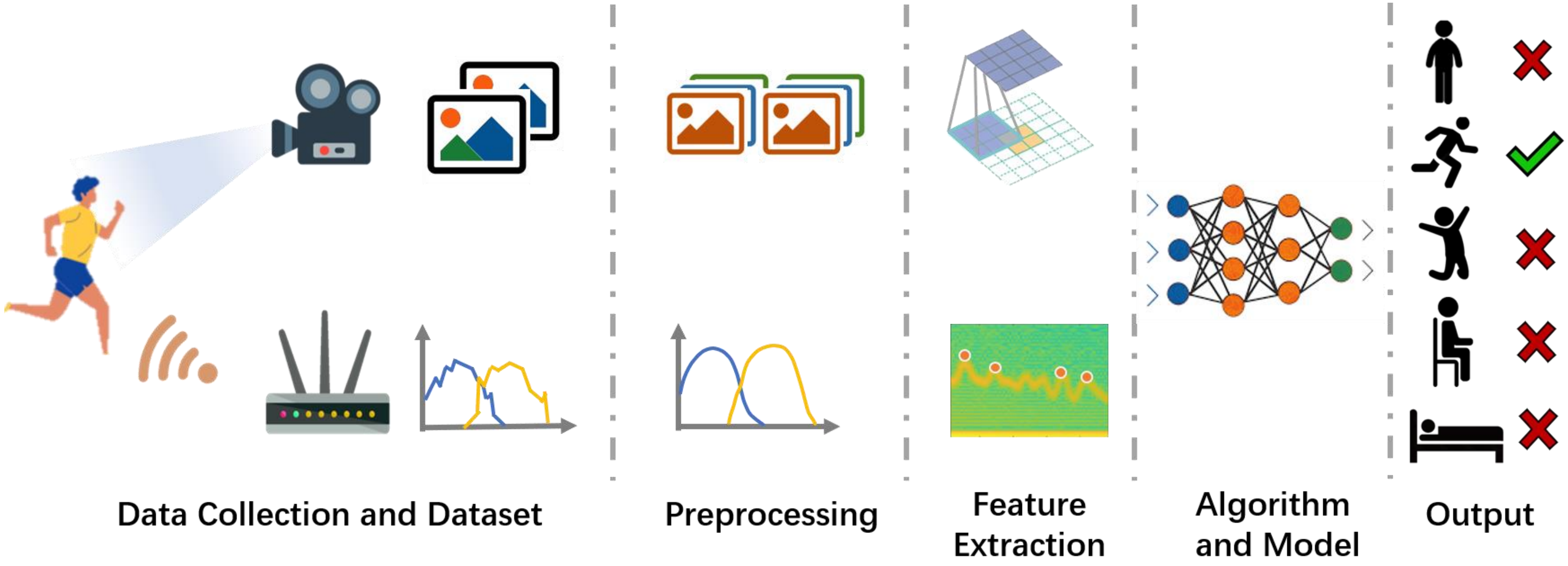


Diagram source: http://tns.thss.tsinghua.edu.cn/wst/docs/intro

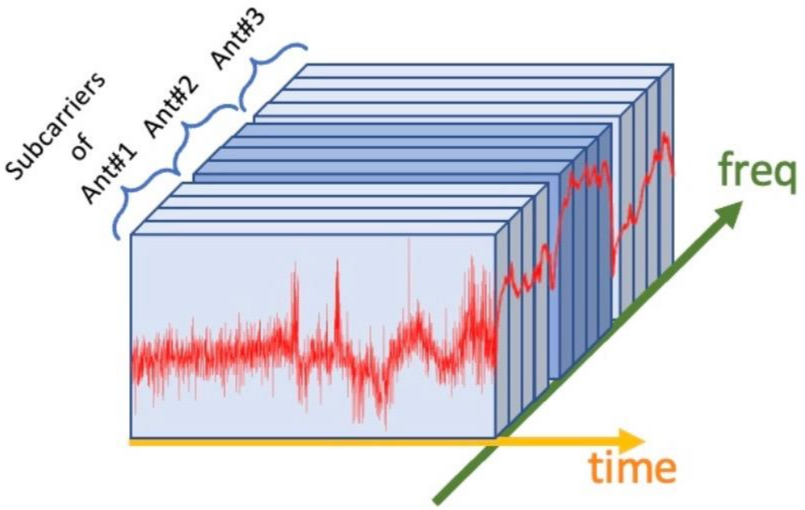
Questions to Ask

- Data: How can we collect (sufficient) data for training?
- Feature: What features should be used for learning?
- Model: What models should we use?

- *Deployment: How can the trained models be deployed?

CSI Data for Learning

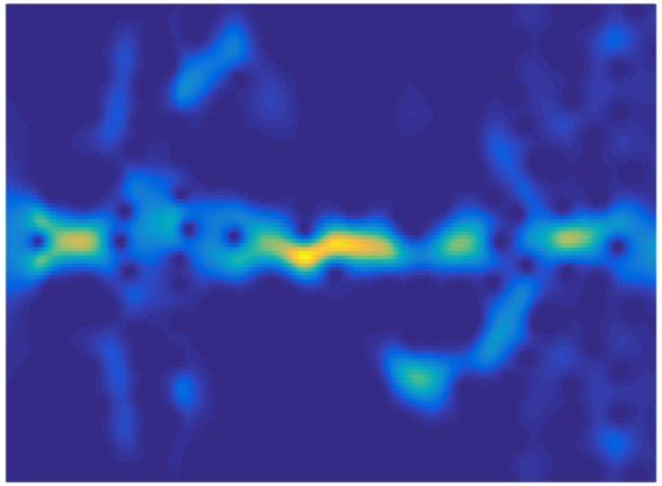
- Non-visual
  - Contain physical and geometric connotations in time, space, and frequency domains all non-visually intelligible (like images to human eyes)
  - (They can certainly be visualized)
  - Visual data are visual, of course
- Complex
  - Complex-valued tensors with amplitude and phase
  - Visual data are real-valued
- High-dimensional
  - time, subcarrier, antennas, transceivers
  - Visual data are usually 2D (image) or 3D (video)



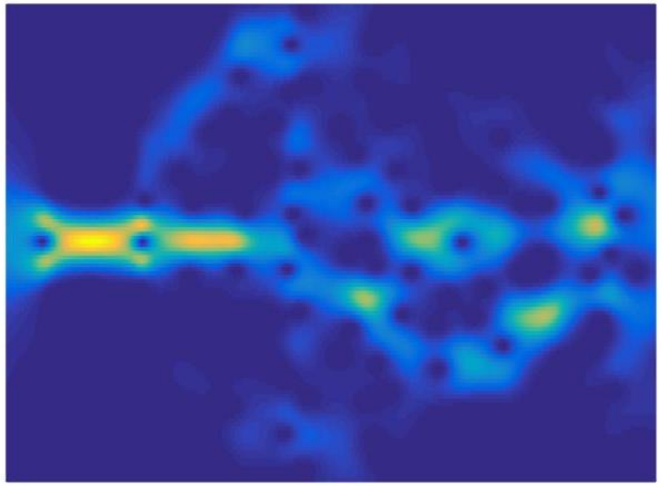
Data Format

CSI Datasets

- More difficult to collect
  - Much more complex setups
  - Less reliable platforms for data collection
    - Not always capture valid data
    - Not all captured data are useful/meaningful
  - Depends on many environmental factors
    - Users, device placements, user locations, user orientations, places/rooms, environmental factors like furniture, wireless configs, device heterogeneity…



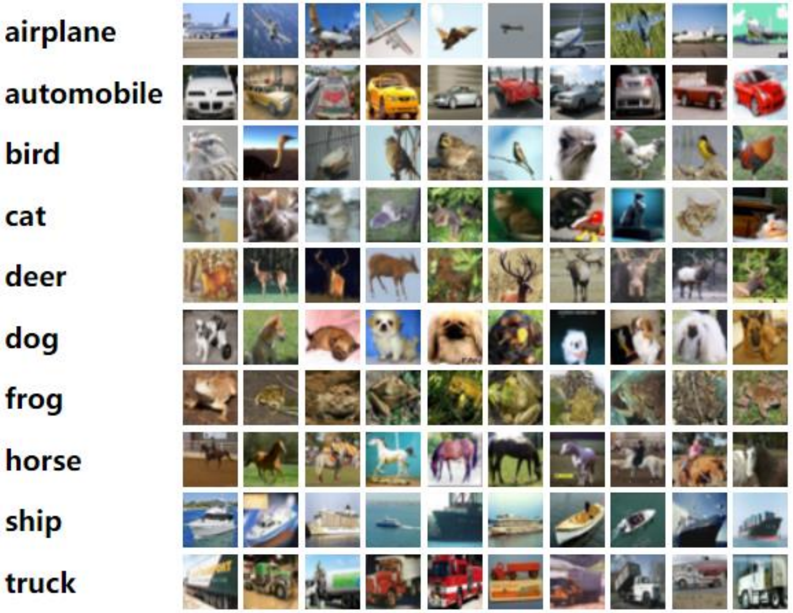
CSI in Setting 1

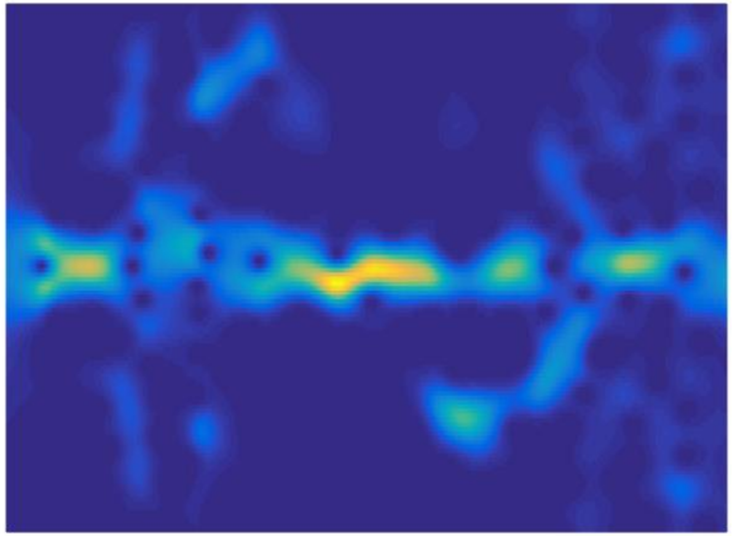


CSI in Setting 2

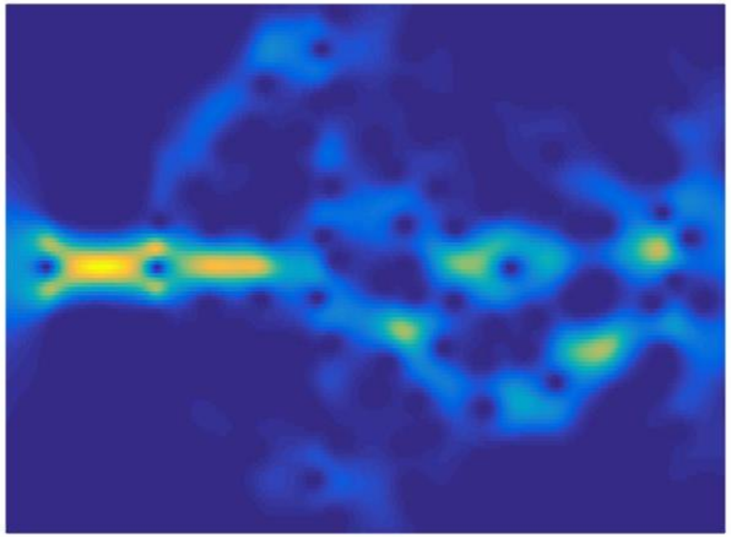
CSI Datasets

- More difficult to label
  - Cannot be labelled OFFLINE (unlike images!)





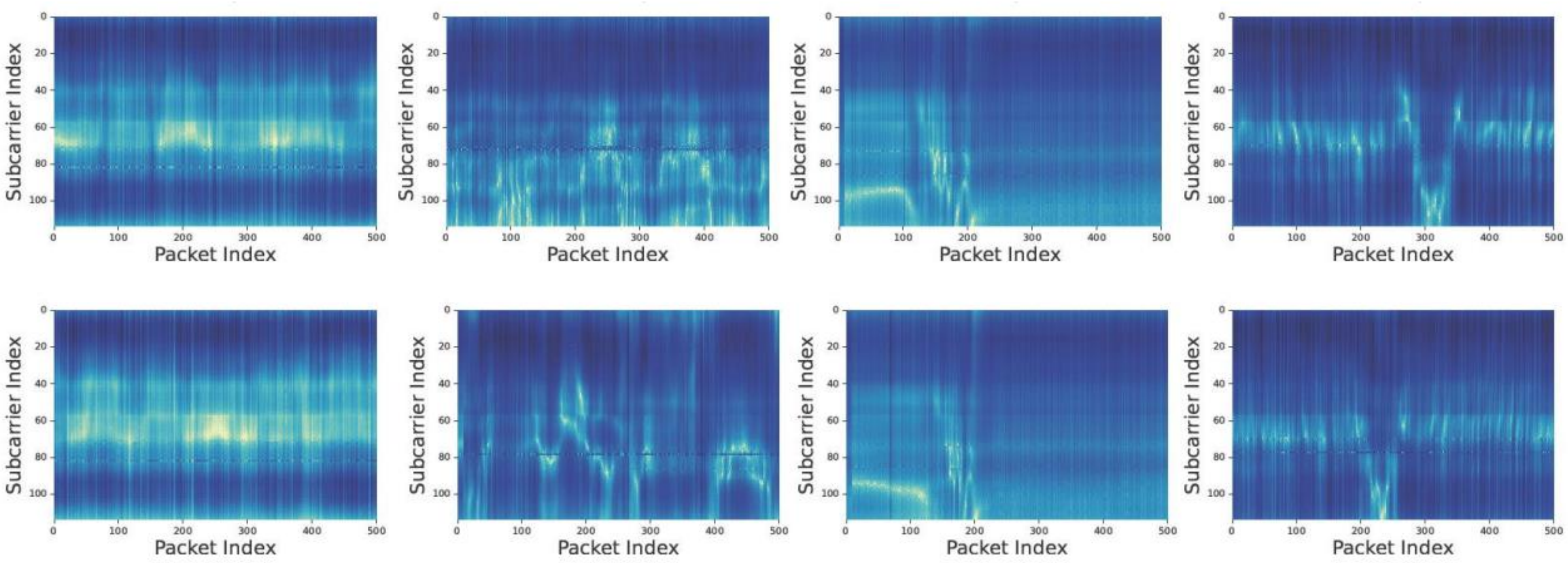
What is this?



And this??

CSI Datasets

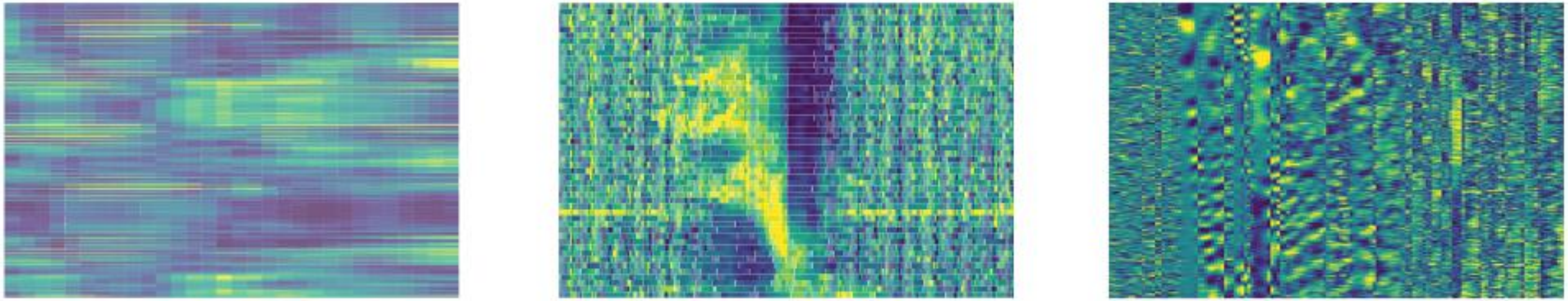
- More difficult to label
  - Cannot be labelled OFFLINE (unlike images!)



What are these?

CSI Datasets

- More difficult to label
  - Cannot be labelled OFFLINE (unlike images!)



(b) Wi-Fi matrix.

(c) FMCW matrix.

(d) IR matrix.

What is this?

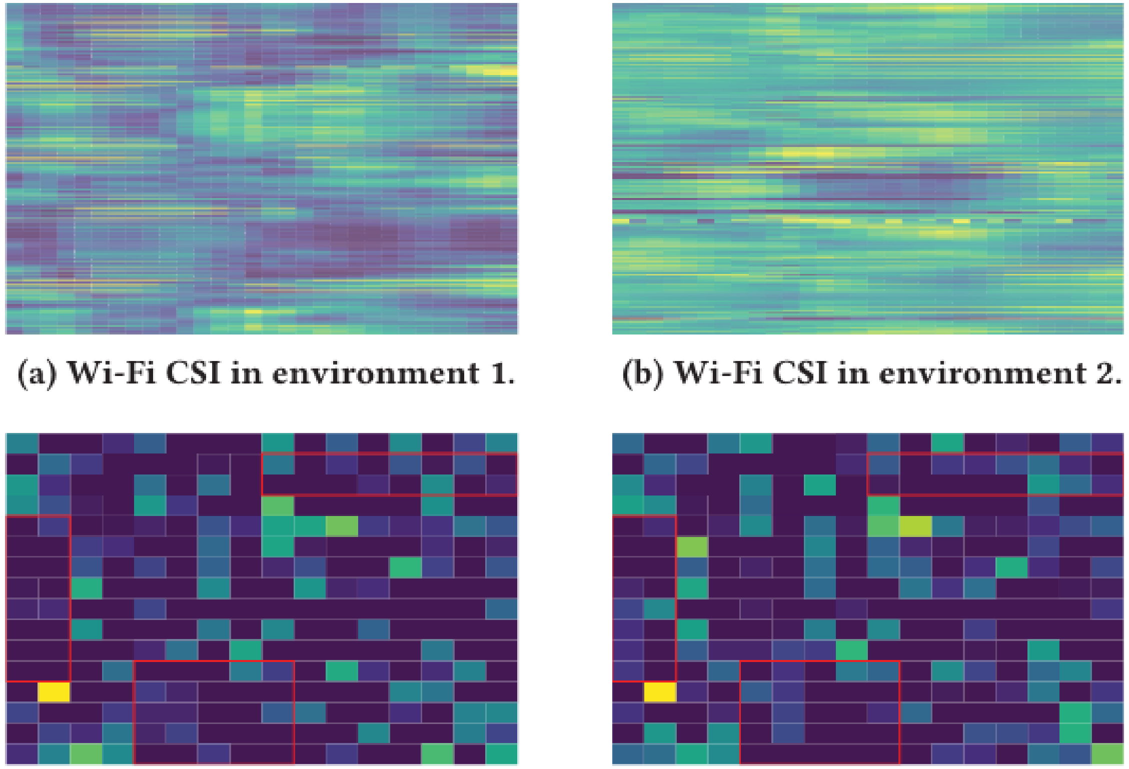
And this??

And this???

CSI Datasets

- Lack of large-scale datasets
  - **Widar3** dataset: 17 participants performing 22 gestures from 5 orientations towards one transmitter, standing at 5 different locations within the coverage of six receivers in 3 environments.
  - **300K** vs. **14M** (ImageNet)

- Difficult to learn
  - Too many impacting factors



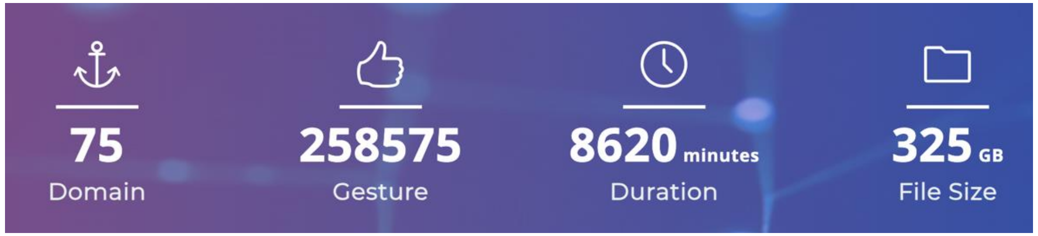
(a) Wi-Fi CSI in environment 1.

(b) Wi-Fi CSI in environment 2.

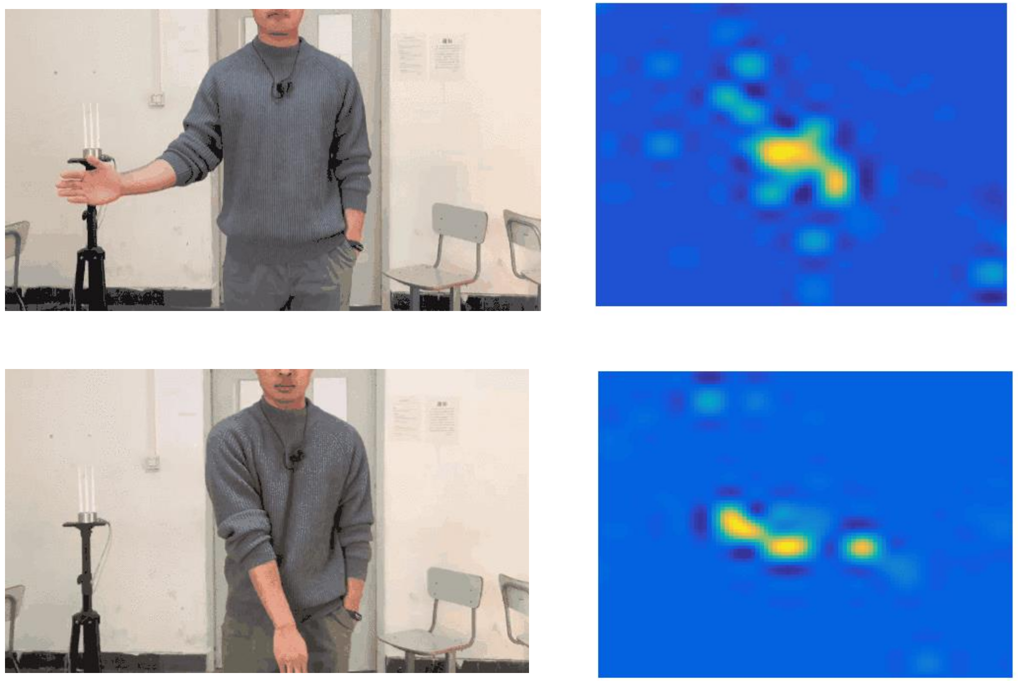
(c) Feature map of environment 1.

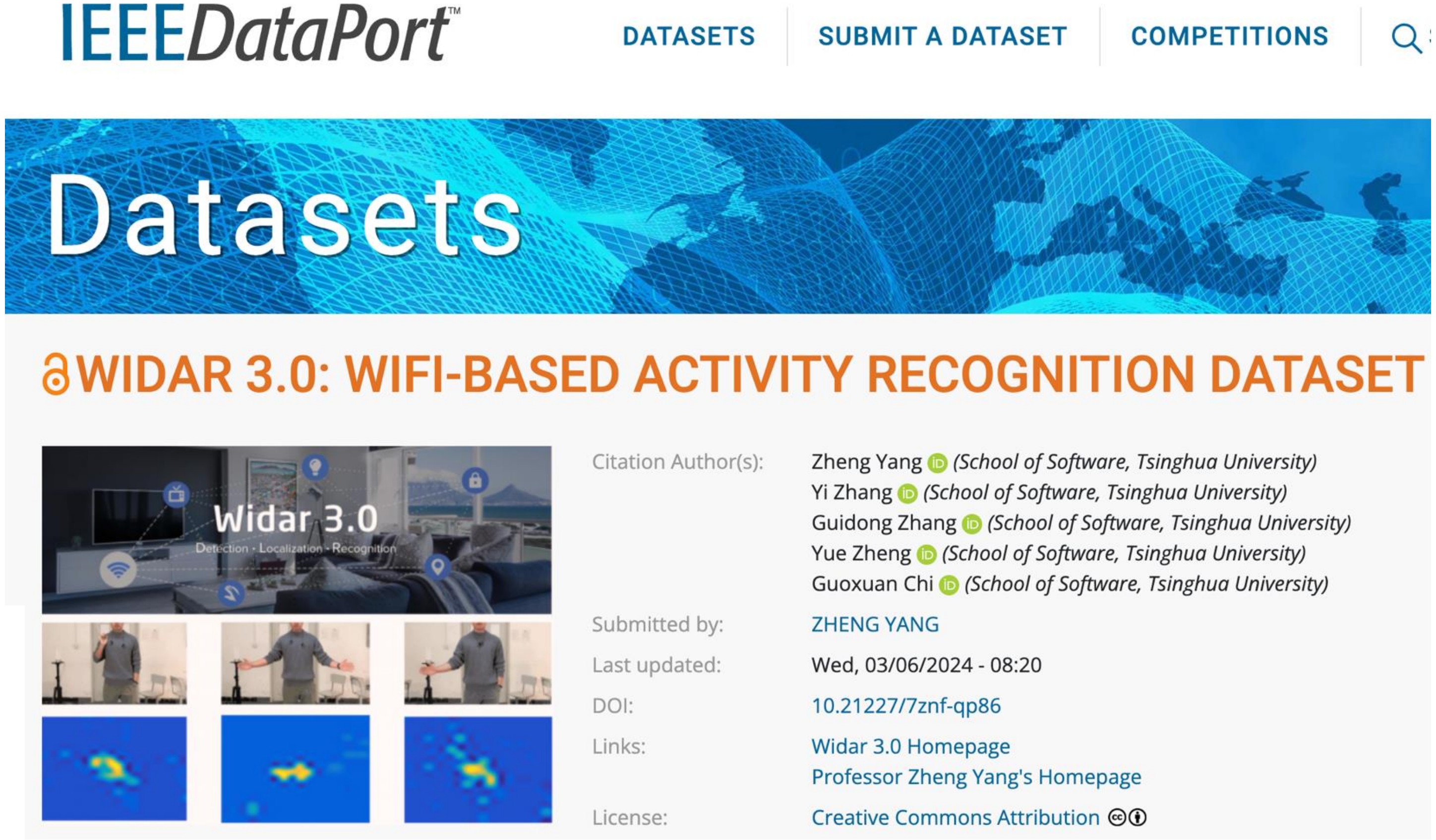
(d) Feature map of environment 2.

Widar3.0 dataset



75	258575	8620 minutes	325 GB
Domain	Gesture	Duration	File Size





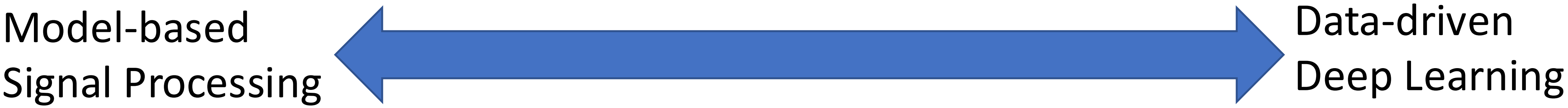
WIDAR 3.0: WIFI-BASED ACTIVITY RECOGNITION DATASET

Citation Author(s):	Zheng Yang (School of Software, Tsinghua University) Yi Zhang (School of Software, Tsinghua University) Guidong Zhang (School of Software, Tsinghua University) Yue Zheng (School of Software, Tsinghua University) Guoxuan Chi (School of Software, Tsinghua University)
Submitted by:	ZHENG YANG
Last updated:	Wed, 03/06/2024 - 08:20
DOI:	10.21227/7znf-qp86
Links:	Widar 3.0 Homepage Professor Zheng Yang's Homepage
License:	Creative Commons Attribution

https://ieee-dataport.org/open-access/widar-30-wifi-based-activity-recognition-dataset

From SP to DL

- How to perform signal processing so that the processed data would produce better learning performance?
- How to design neural networks so that they will better fit the unique wireless data (that are very different from visual data)?



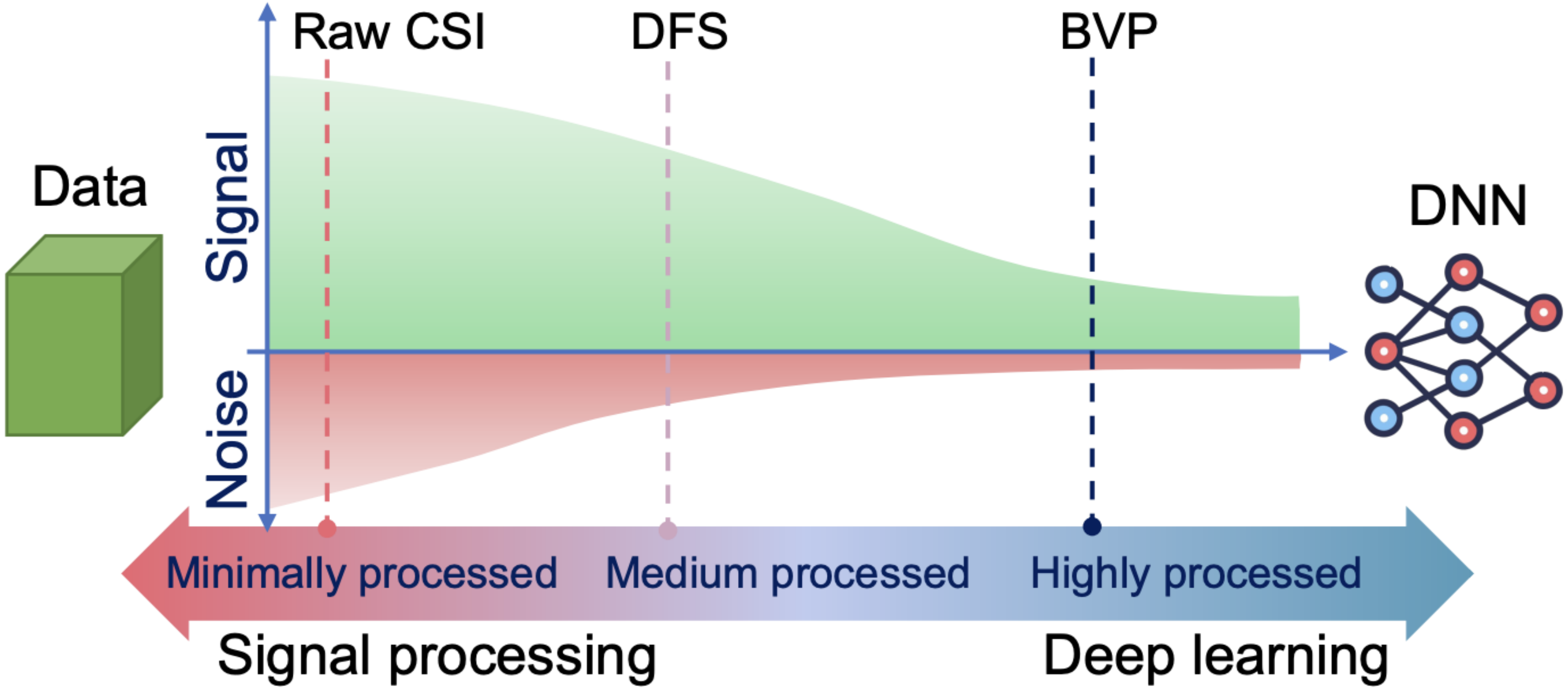
Model-based Signal Processing

Data-driven Deep Learning

model-guided,
data-driven
design?

Signal processing vs. Deep learning

- A “sweet point” of balancing SP and DL?

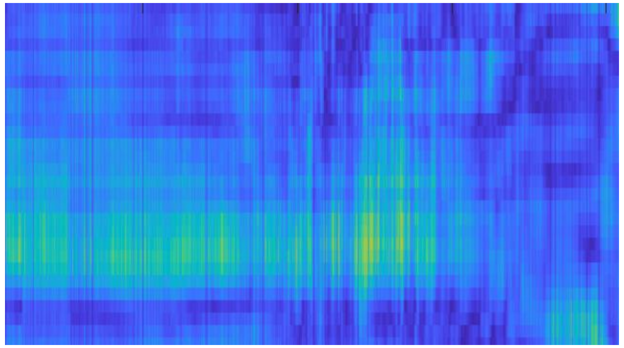


*BVP: Body-coordinate Velocity Profiles (BVPs) proposed in Widar3

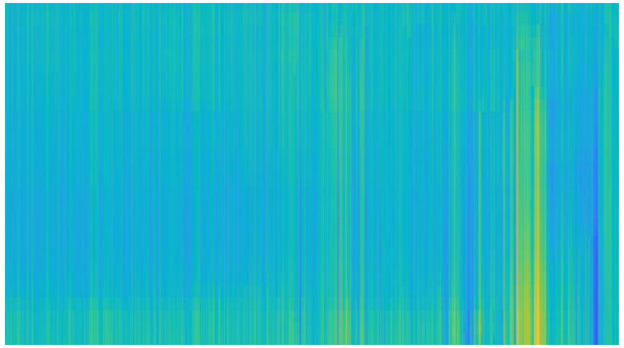
Hou, C. Wu, RFBoost: Physical Data Augmentation for Deep Wireless Sensing, ACM IMWUT 2024

Wireless Data Representation

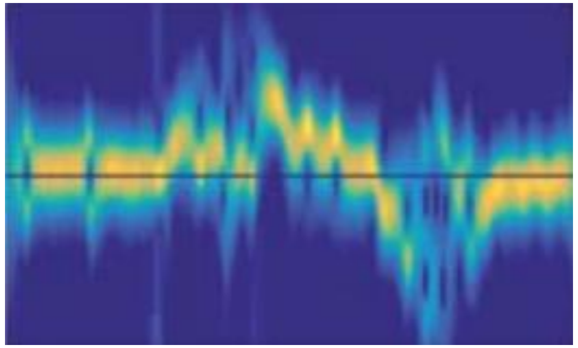
- WiFi CSI



Amplitude



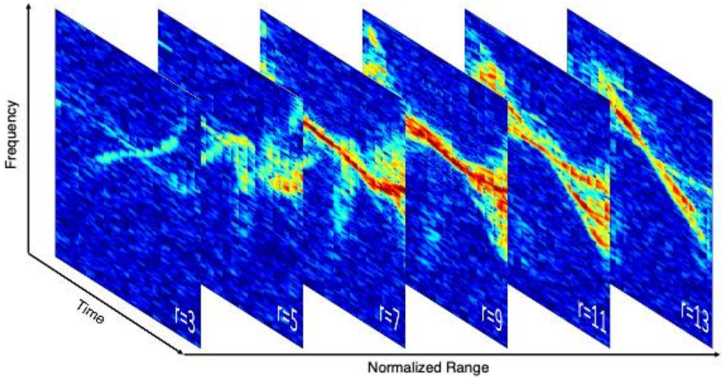
Phase



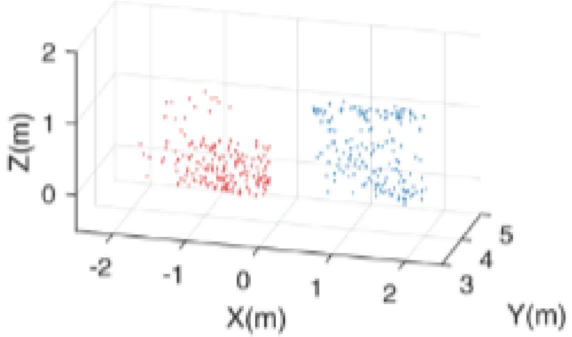
Spectrogram

…

- mmWave radar data



Doppler-Range Cube

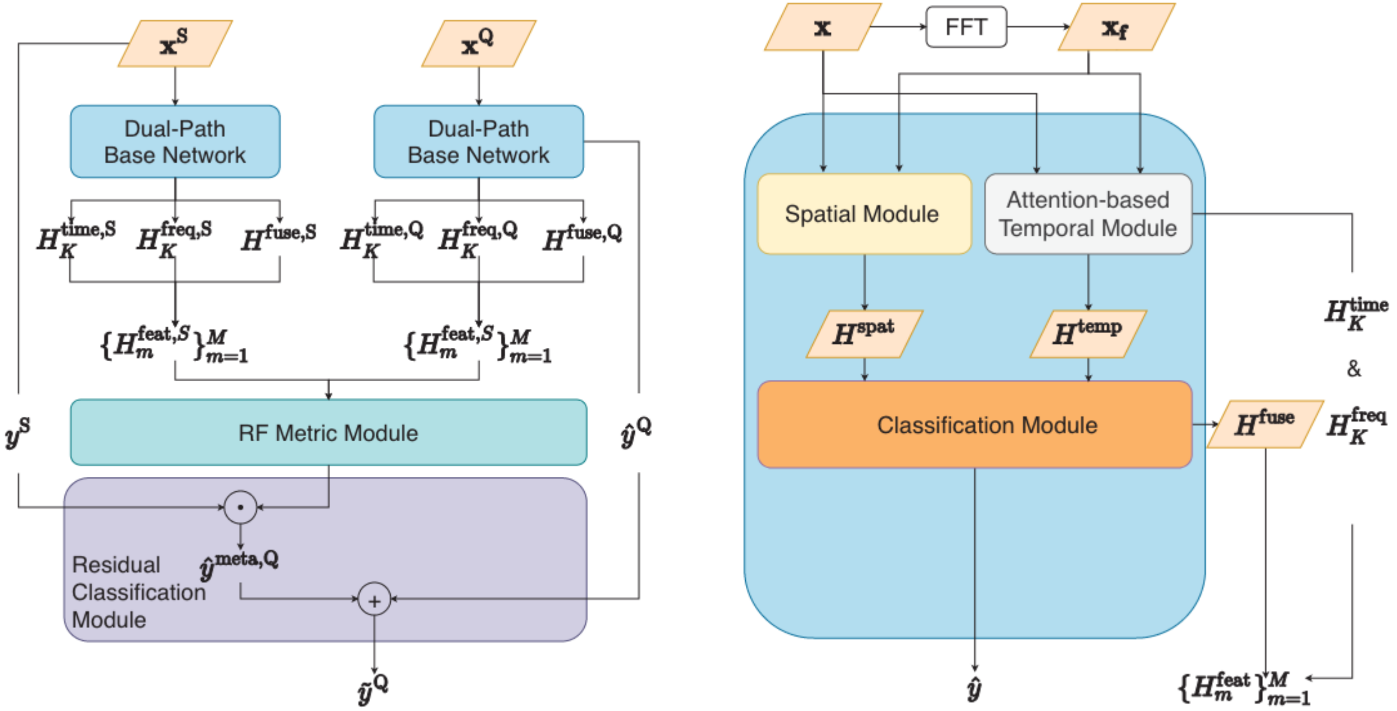


Point Clouds

…

Wireless Data Representation

- Raw CSI is all you need?
  - Complex values
  - Amplitude only
  - Amplitude + phase
  - I/Q-components

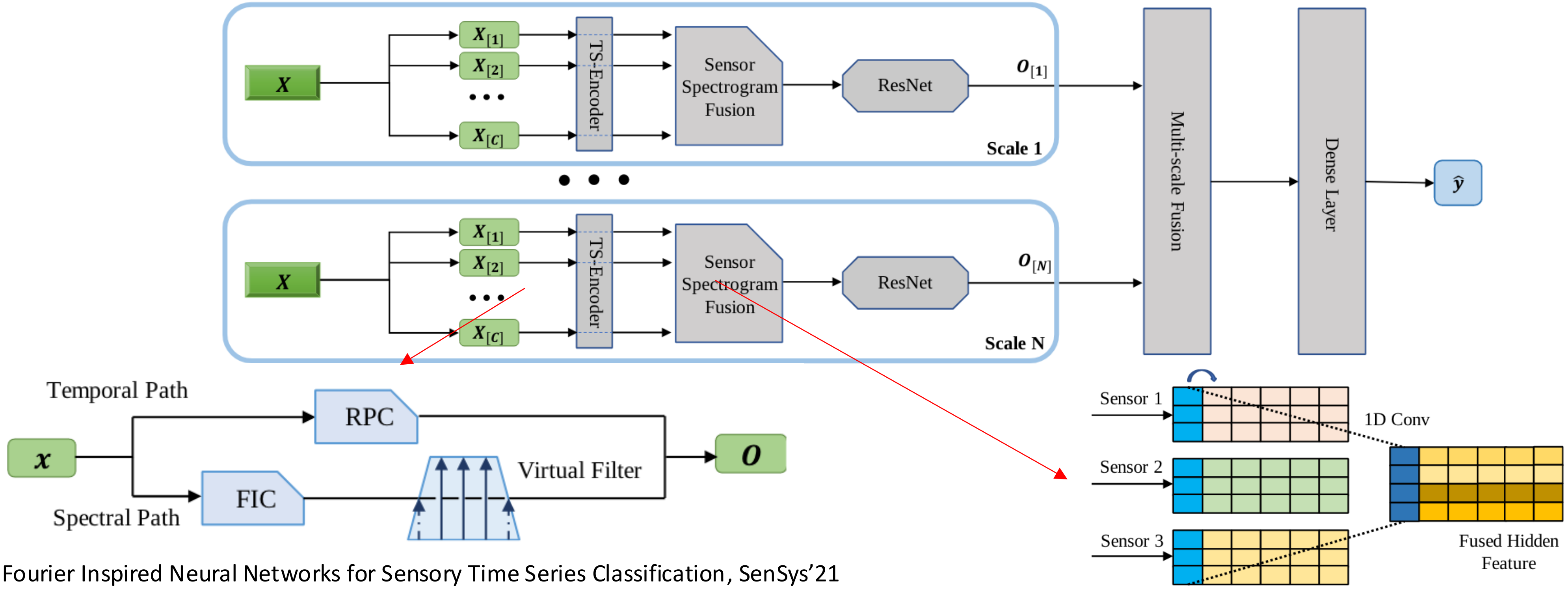


$\mathbf{x}^{S}$
$\mathbf{x}^{Q}$
Dual-Path Base Network
Dual-Path Base Network
$H_K^{\text{time},S}$ $H_K^{\text{freq},S}$ $H^{\text{fuse},S}$
$H_K^{\text{time},Q}$ $H_K^{\text{freq},Q}$ $H^{\text{fuse},Q}$
$\{H_m^{\text{feat},S}\}_{m=1}^{M}$
$\{H_m^{\text{feat},S}\}_{m=1}^{M}$
y^{S}
RF Metric Module
$\hat{y}^{Q}$
Residual Classification Module
$\hat{y}^{\text{meta},Q}$
$\tilde{y}^{Q}$
$\mathbf{x}$
FFT
$\mathbf{x_f}$
Spatial Module
Attention-based Temporal Module
H^{spat}
H^{temp}
H_K^{time}
&
H_K^{freq}
Classification Module
H^{fuse}
$\hat{y}$
$\{H_m^{\text{feat}}\}_{m=1}^{M}$

RF-Net: A Unified Meta-Learning Framework for RF-enabled One-Shot Human Activity Recognition, SenSys'20

Wireless Data Representation

- DFS: Time-Frequency Spectrograms
  - integrate STFT into neural networks by initializing convolutional filter weights as the Fourier coefficients.

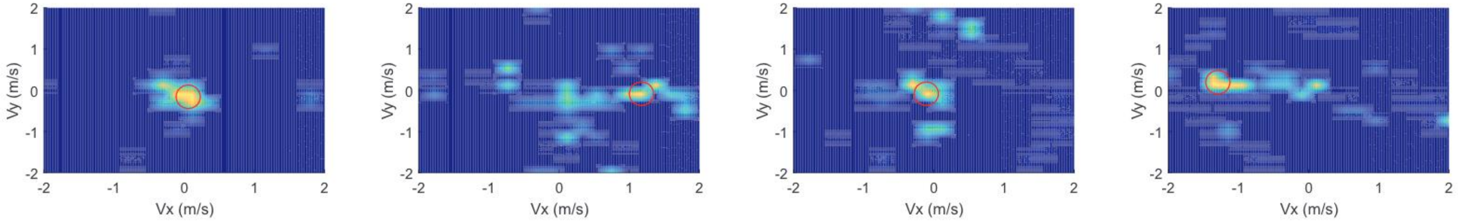


UniTS: Short-Time Fourier Inspired Neural Networks for Sensory Time Series Classification, SenSys’21

CSI Data Representation

- Widar3: Body-coordinate Velocity Profile (BVP)
  - Denoise via SP rather than completely relying on DL
  - Theoretically, domain-independent Body-coordinate Velocity Profile

For each activity,
obtain a BVP series:



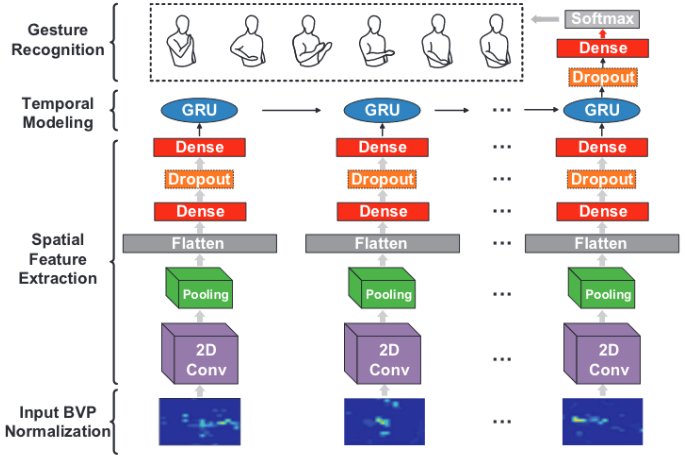
(a) Stage 1: Start

(b) Stage 2: Pushing

(c) Stage 3: Stop

(d) Stage 4: Pulling

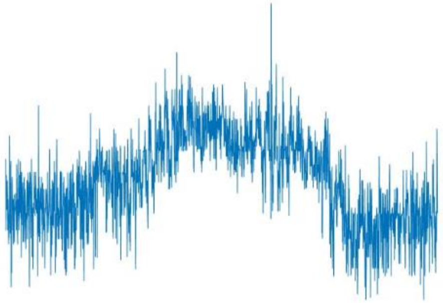
CNN+RNN for
classification



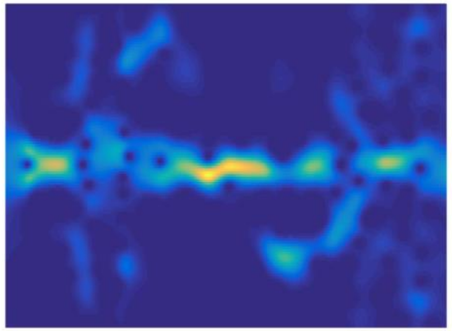
Zhang, Y., Zheng, Y., Qian, K., Zhang, G., Liu, Y., Wu, C., & Yang, Z. (2021). Widar3. 0: Zero-effort cross-domain gesture recognition with Wi-Fi. IEEE Transactions on PAMI

Widar3: BVP for Learning

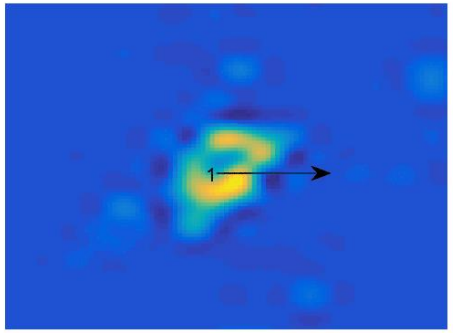
Domain-1
orientation #1
position #1
environment #1



CSI in Domain-1

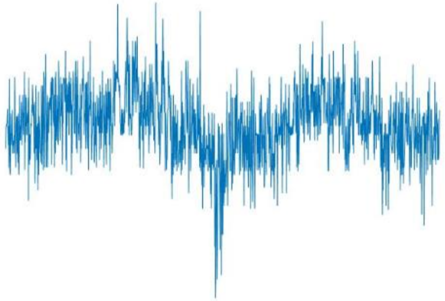


DFS in Domain-1

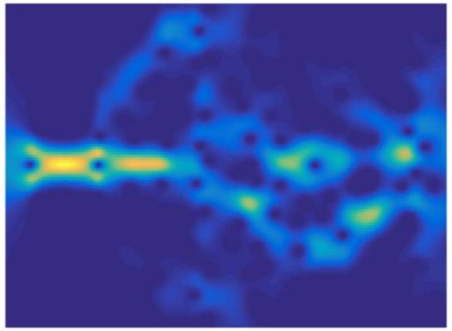


BVP in Domain-1

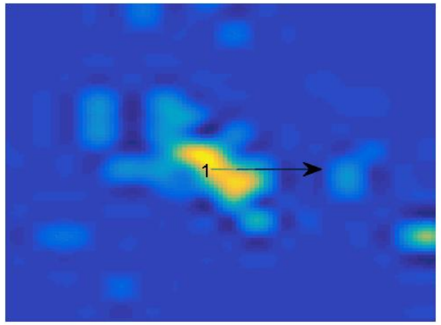
Domain-2
orientation #2
position #2
environment #2



CSI in Domain-2



DFS in Domain-2

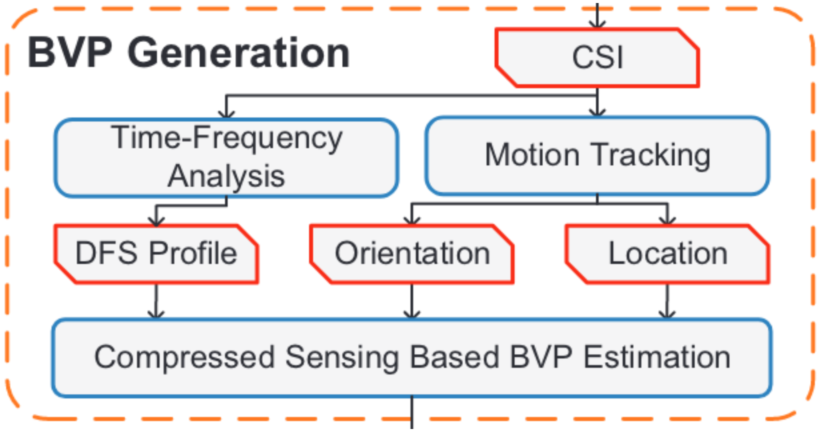


BVP in Domain-2

Zhang, Y., Zheng, Y., Qian, K., Zhang, G., Liu, Y., Wu, C., & Yang, Z. (2021). Widar3. 0: Zero -effort cross-domain gesture recognition with Wi-Fi. IEEE Transactions on PAMI

Wireless Data Representation

- BVP
  - Theoretically, domain-independent Body-coordinate Velocity Profile
  - Denoise via SP rather than completely relying on DL
  - However, BVP relies on multiple links and imposes heavy computation



BVP Generation

CSI

Time-Frequency Analysis

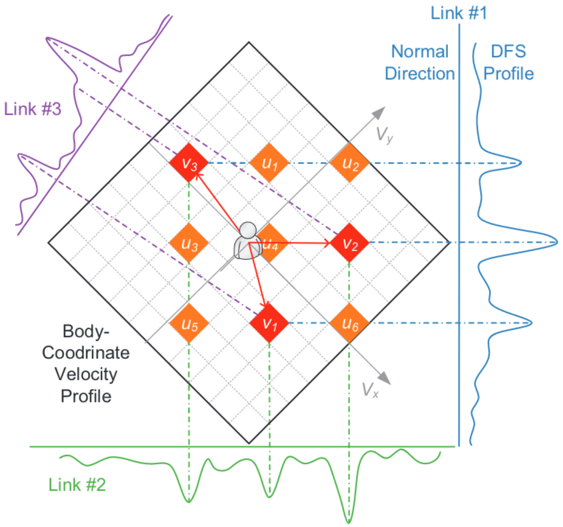
Motion Tracking

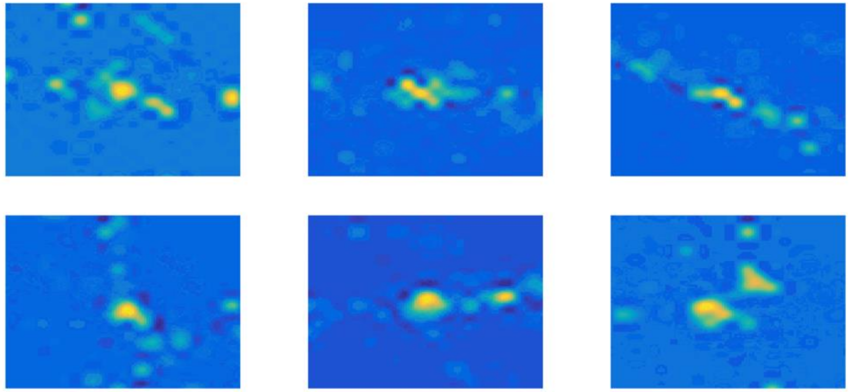
DFS Profile

Orientation

Location

Compressed Sensing Based BVP Estimation

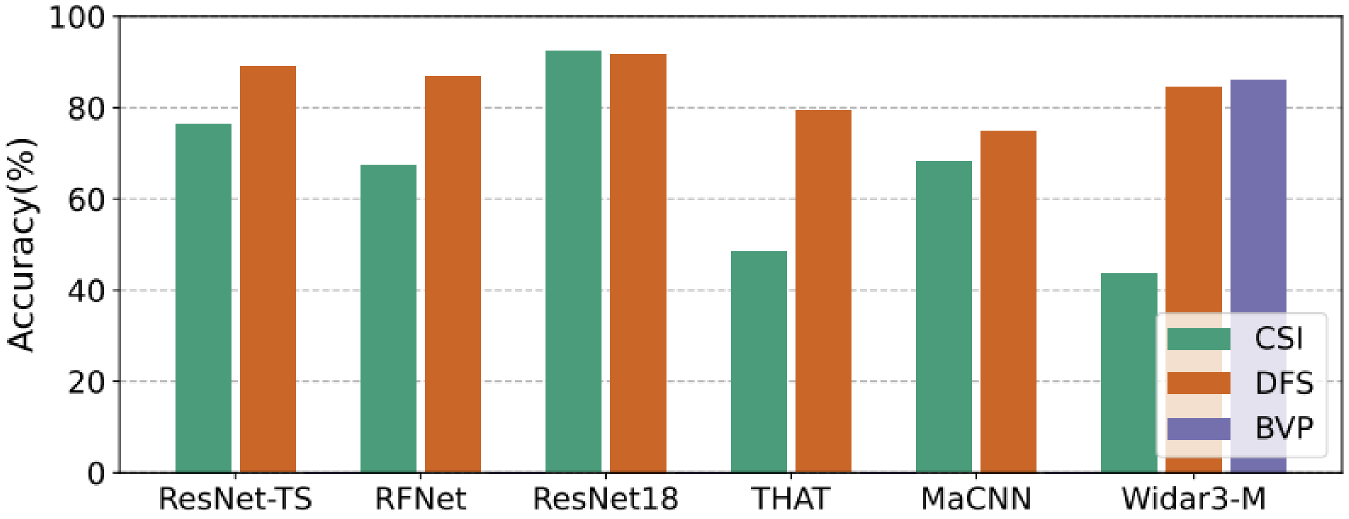




Zero-Effort Cross-Domain Gesture Recognition with Wi-Fi, MobiSys’19

Wireless Data Representation

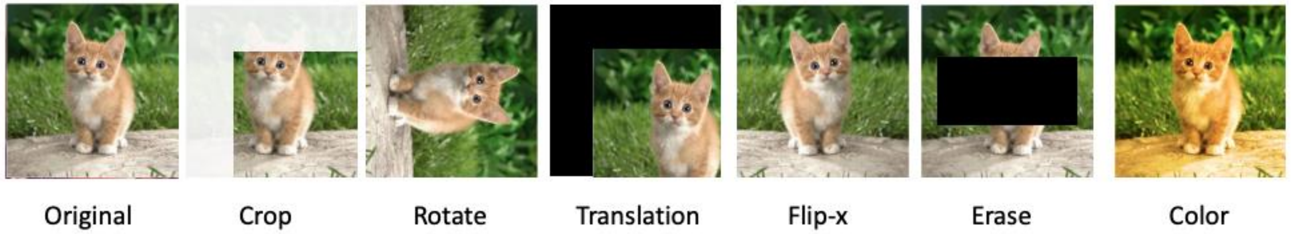
- What data representation should we use?
  - Still ad-hoc, no universal choice today
- Results of a preliminary comparison study



Performance comparison using CSI, DFS, and BVP as inputs on different models on Widar3 dataset.

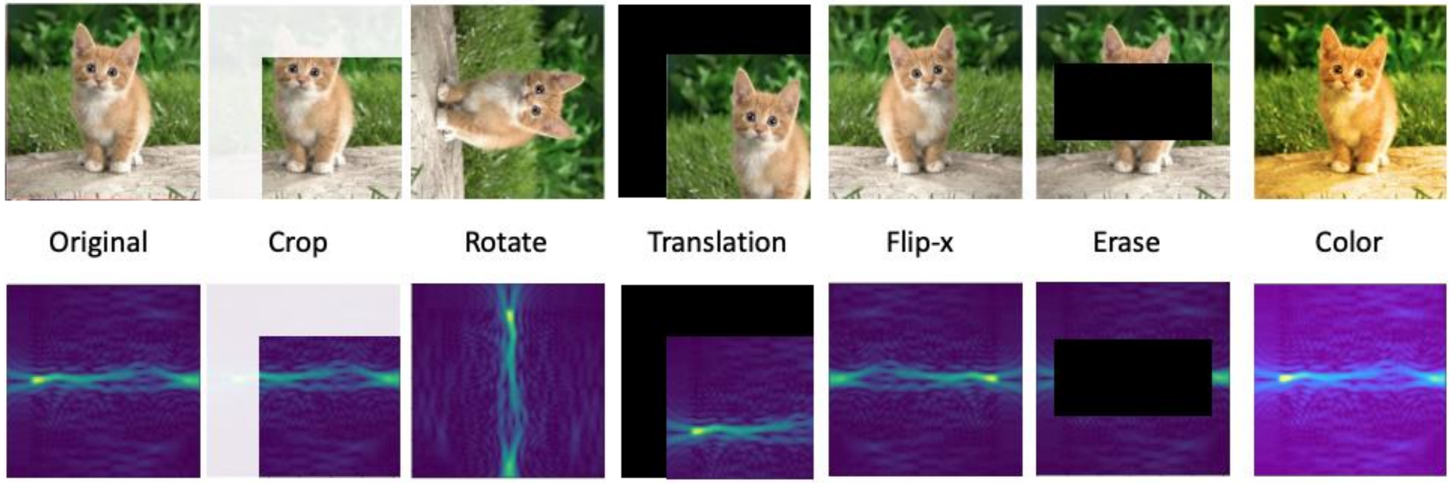
Data Augmentation

- Data augmentation
  - A set of techniques that artificially inflate the training samples from existing data.
  - Increase the amount of data without collecting more
  - Has been a common practice and proved effective in CV field
- Image data augmentation
  - Random transformation: flipping, cropping, erasing, rotation, translation, color space transformation, neural style transfer, etc.



Data Augmentation

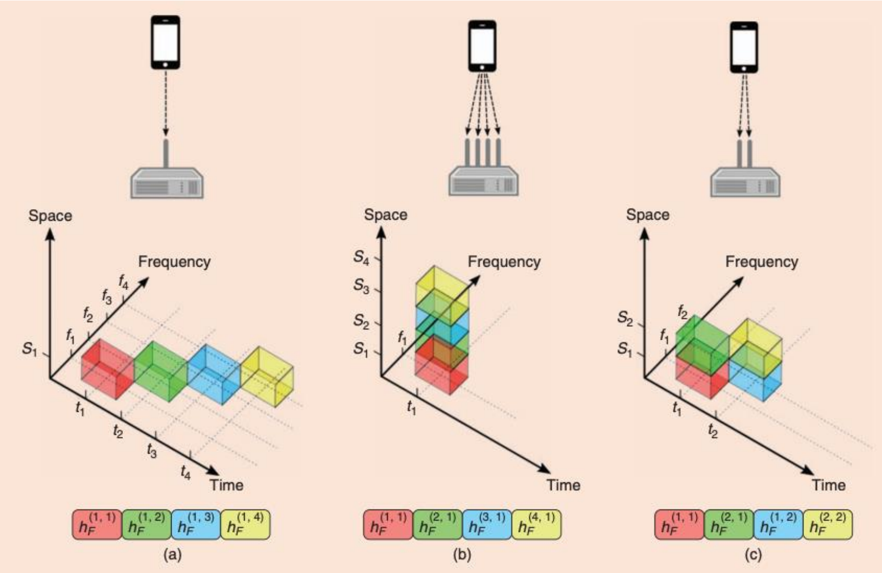
- Apply image data augmentation to radio data?
  - Radio data, even in the format of images, has different physical meanings
  - * rotating a spectrogram will reverse the time and frequency dimensions
  - * flipping it will reverse the time series (e.g., a pull gesture can become a push)
  - * random erasing can remove the critical part that has frequency responses
  - * changing the color space does not bring any new information

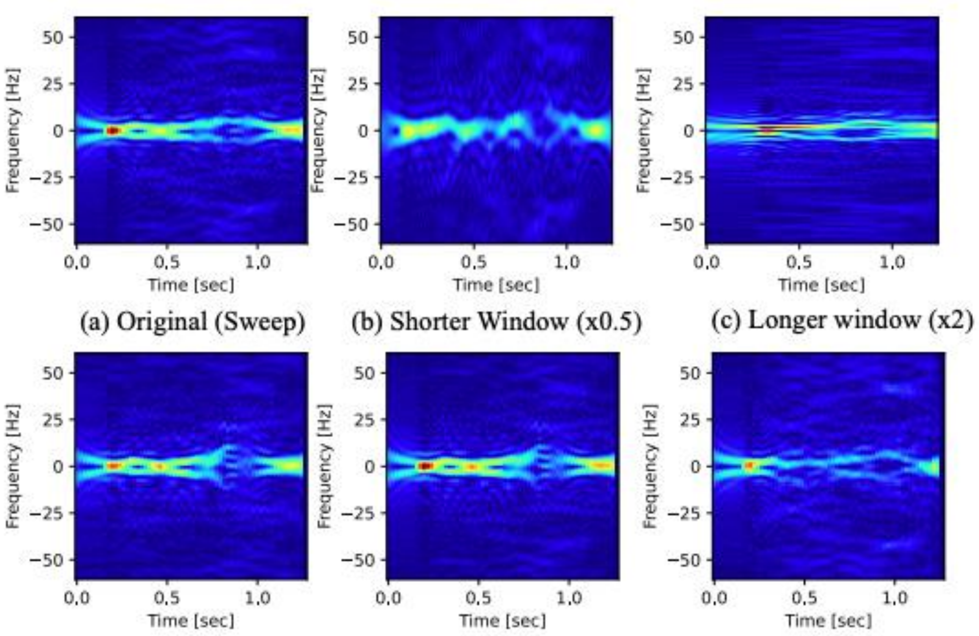


Make little sense, but may still benefit accuracy (given certain model and available benchmarks)!

Data Augmentation

- Physical Data Augmentation
  - Previously: PCA for dimension reduction → one single representative spectrograms per tx-rx link
  - Recall diversity





(a) Original (Sweep)

(b) Shorter Window (x0.5)

(c) Longer window (x2)

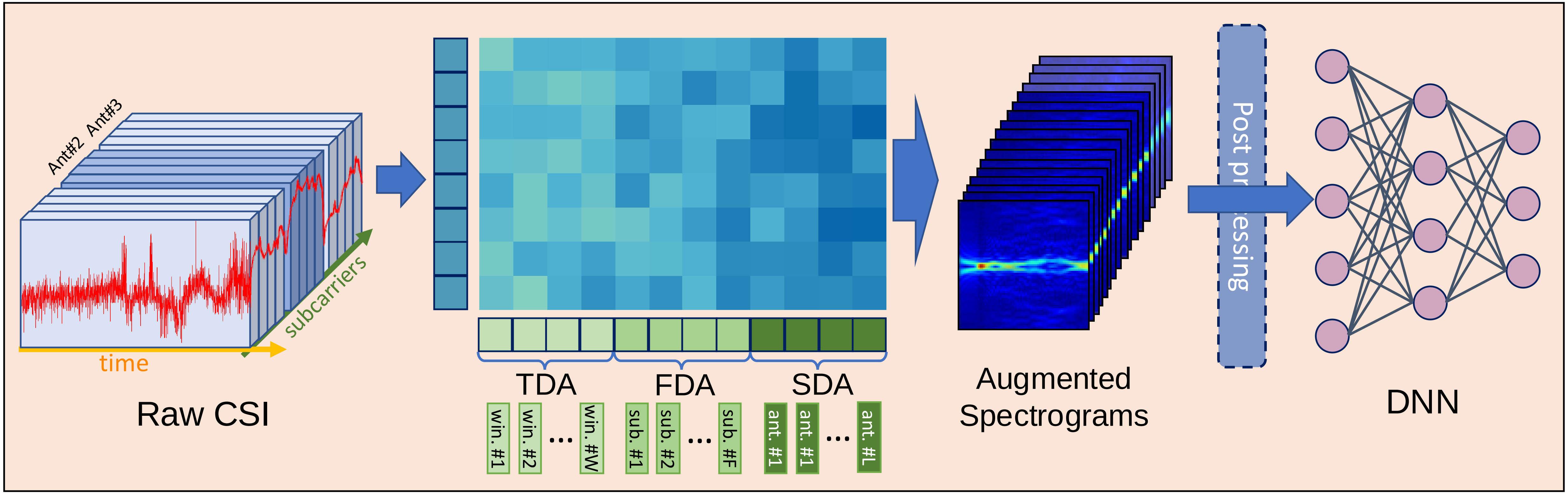
(d) Subcarrier #1 Ant#1

(e) Subcarrier #1 Ant#2

(f) Subcarrier #2 Ant#2

Data Augmentation

- Physical Data Augmentation
  - Leverage data diversity to augment spectrograms!

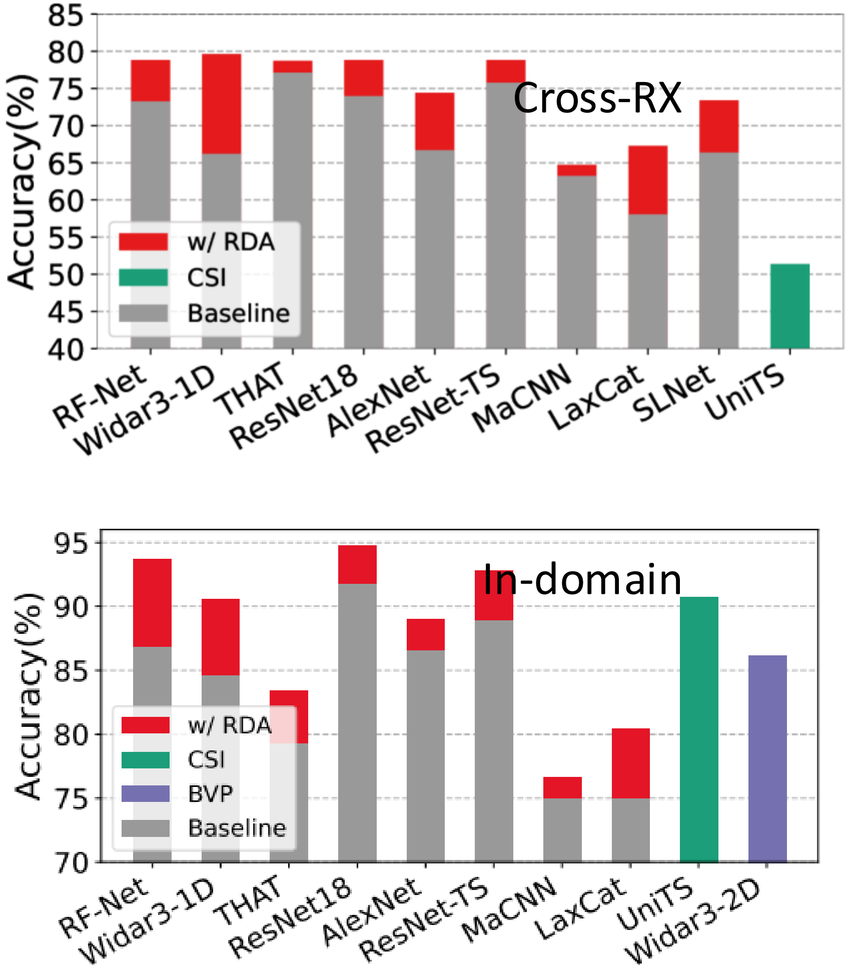


Hou, C. Wu, RFBoost: Physical Data Augmentation for Deep Wireless Sensing, ACM IMWUT 2024

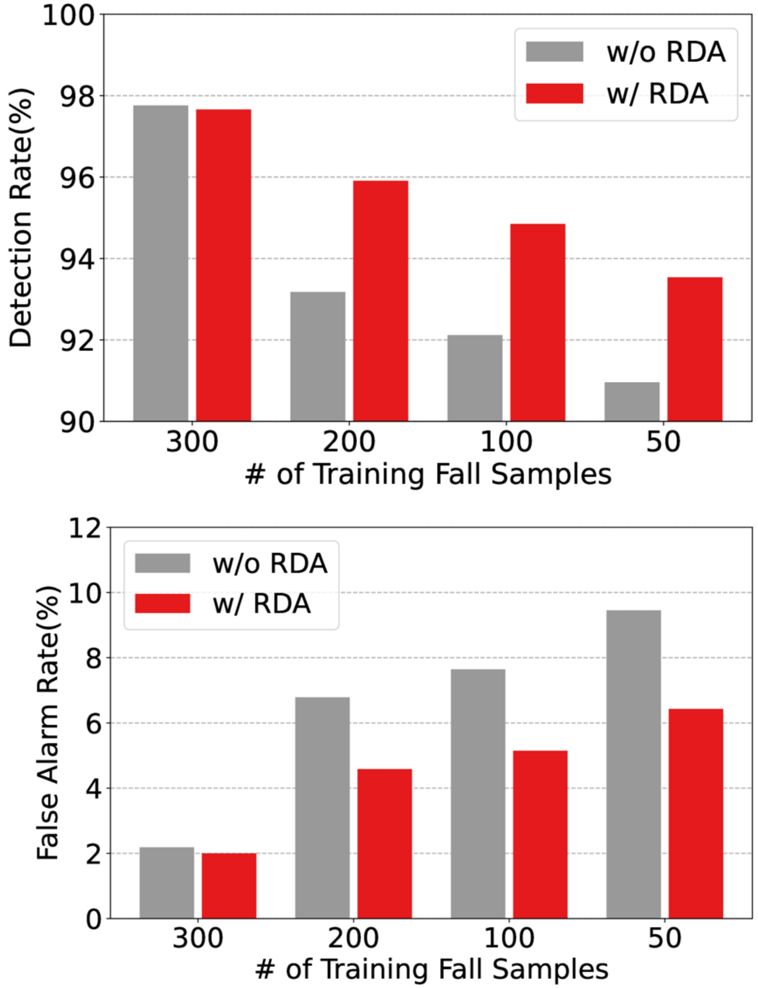
RFBoost: Physical Data Augmentation

- Physical Data Augmentation
  - **Time-domain Data Augmentation (TDA):** Using different windows for spectrogram generation
  - **Frequency/Space-domain Data Augmentation (FDA/SDA):** Combinations of different subcarriers/antennas
  - **Motion-aware Data Augmentation (MDA)**: Motion-aware Random Erasing and Shifting
- All augmented samples are from physical observations

RFBoost: Performance



Gesture recognition



Fall detection

RFBoost: Performance

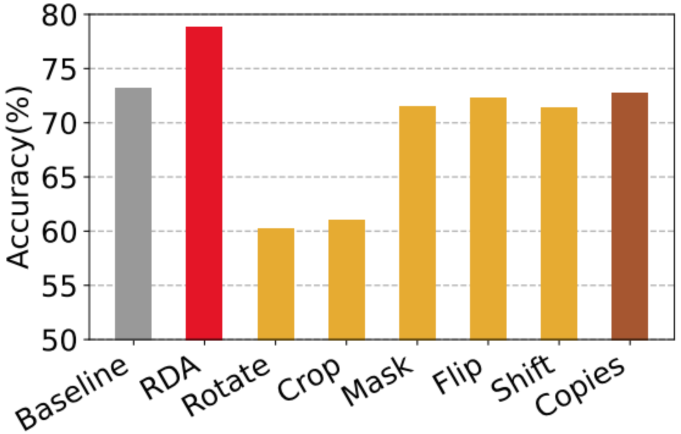


Fig. 18. Comparisons between IDA and RDA.

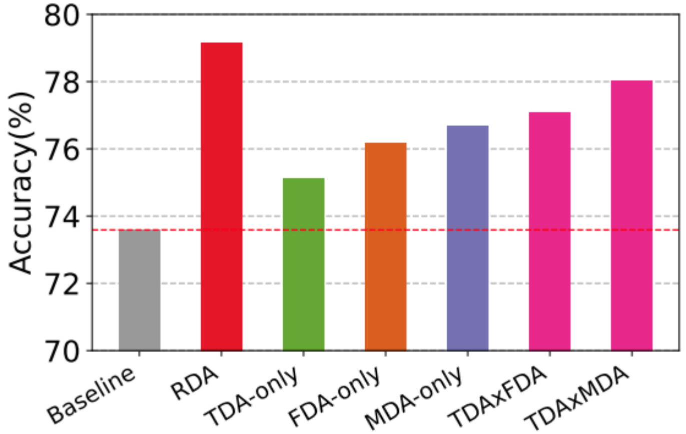
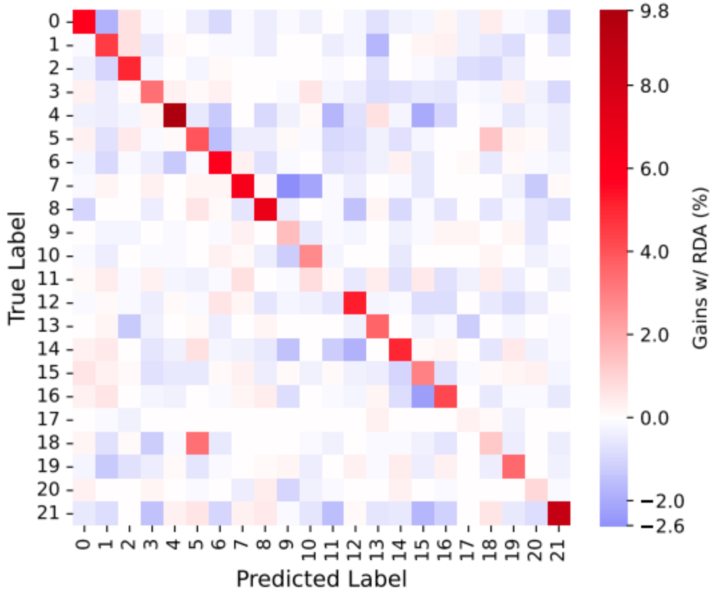
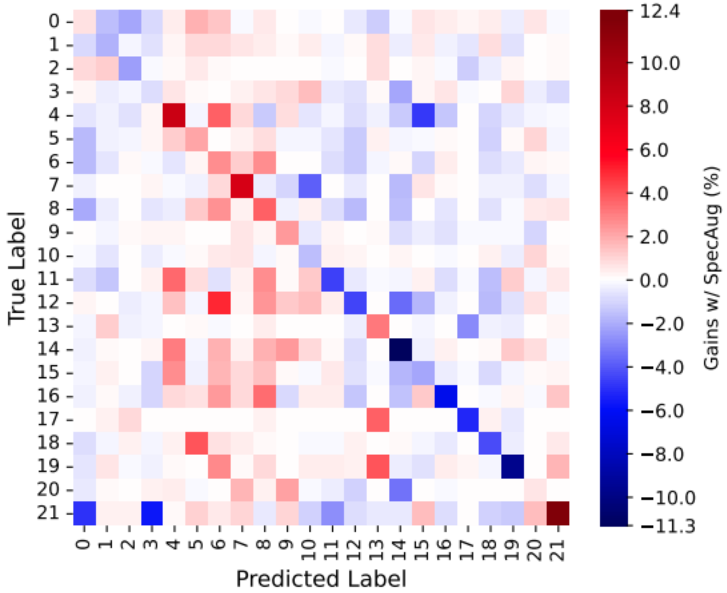


Fig. 19. Ablation study on different RDA policies.



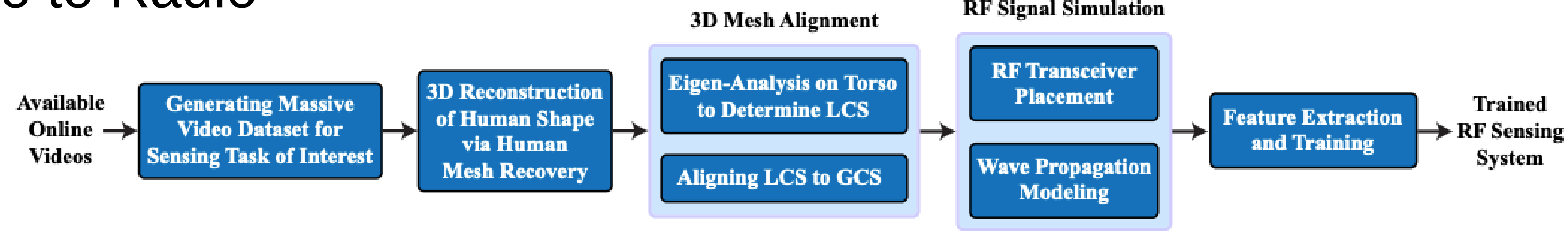
(a) Accuracy gains w/ RDA



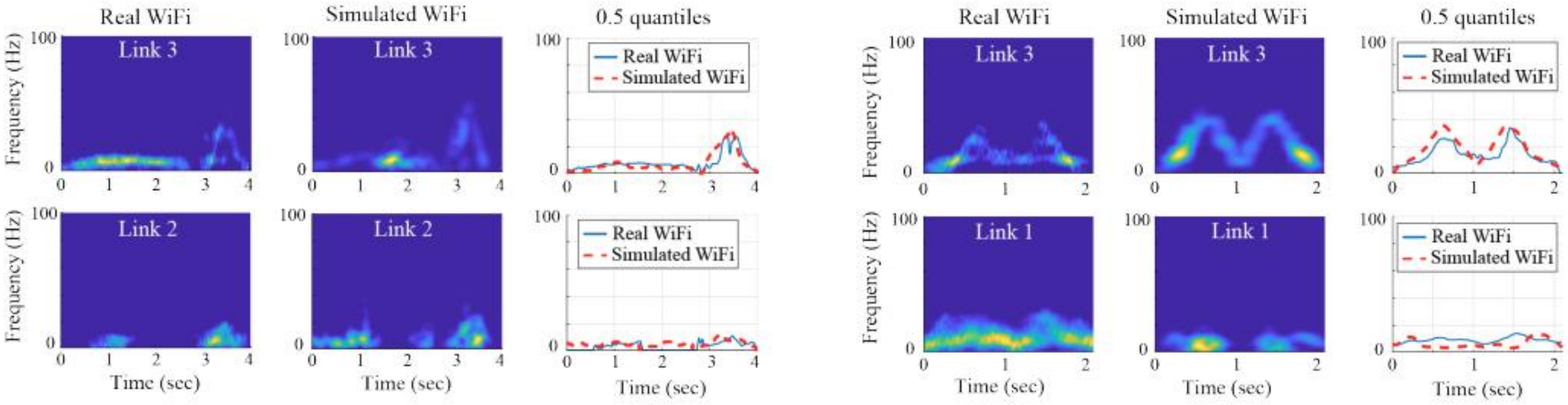
(b) Accuracy gains w/ SpecAugment

Data Synthesis/Generation

- Cross-modality training
  - Video to Radio



Available Online Videos → Generating Massive Video Dataset for Sensing Task of Interest → 3D Reconstruction of Human Shape via Human Mesh Recovery → 3D Mesh Alignment: Eigen-Analysis on Torso to Determine LCS; Aligning LCS to GCS → RF Signal Simulation: RF Transceiver Placement; Wave Propagation Modeling → Feature Extraction and Training → Trained RF Sensing System



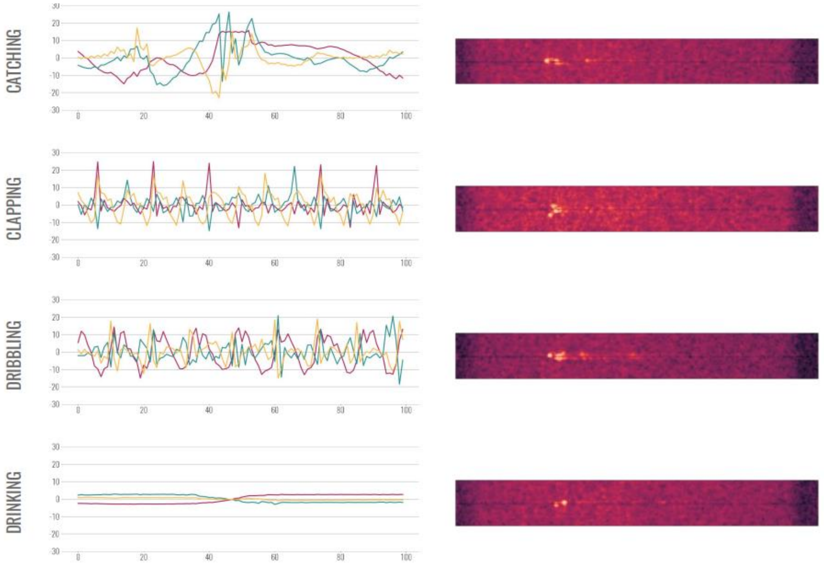
(a) Stiff-leg deadlift

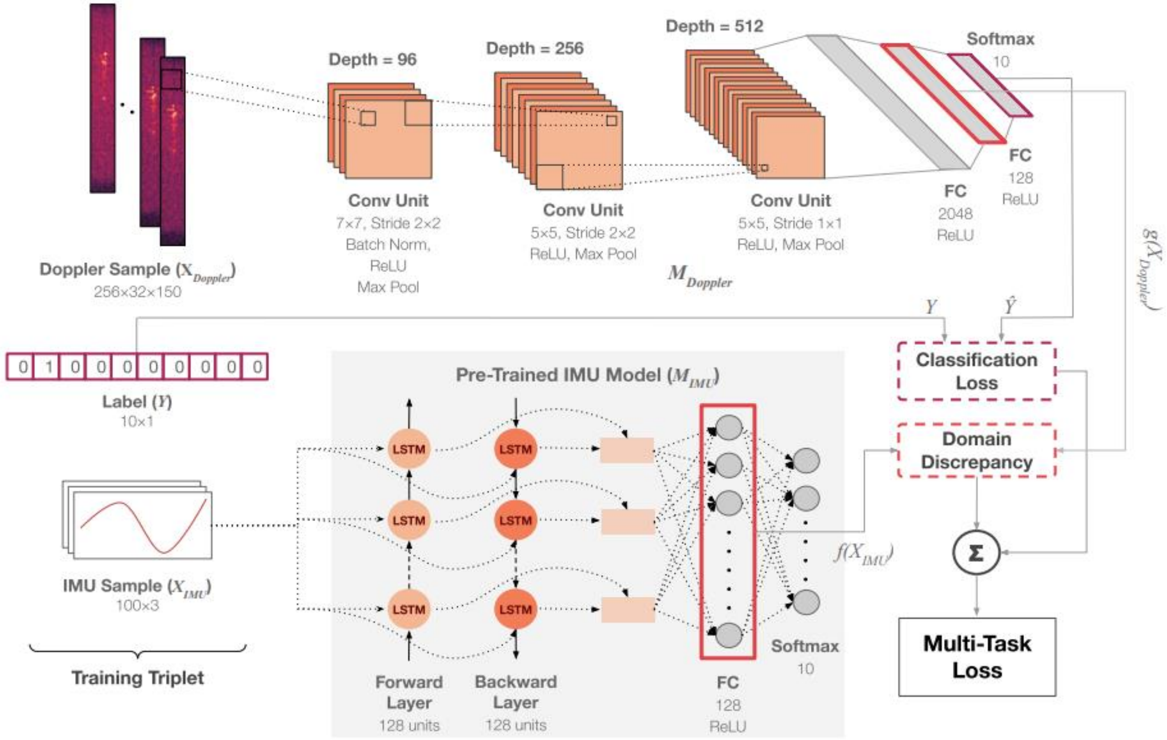
(b) Forward Lunge

Teaching RF to Sense without RF Training Measurements, IMWUT'20

Data Synthesis/Generation

- Cross-modality training
  - IMU to Radio





IMU2Doppler: Cross-Modal Domain Adaptation for Doppler-based Activity Recognition Using IMU Data, IMWUT’21

RF-Diffusion: Radio Signal Generation via Time-Frequency Diffusion

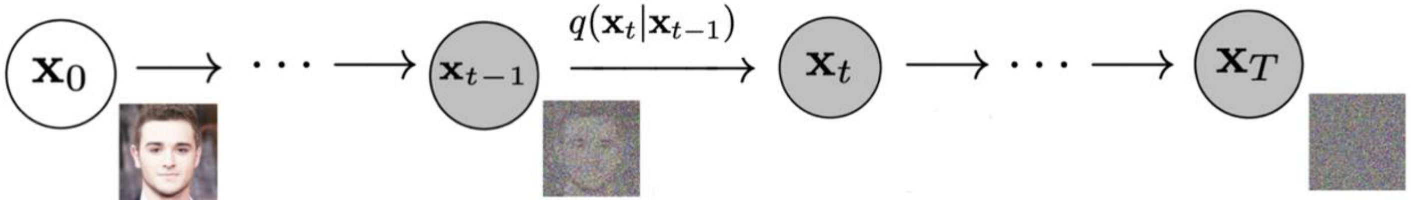
- The AIGC era has arrived, encompassing various modalities.
  - Text: ChatGPT, LLaMA, ChatGLM, Claude, …
  - Image: Stable Diffusion, DALL·E, Midjourney, …
  - Video: Sora, Imagen Video, CogVideo, …
  - Audio: Suno, AudioCraft, WaveNet, …
- Generative AI has mastered most modalities, except the RF signal.

How about building an RF Generative Model?

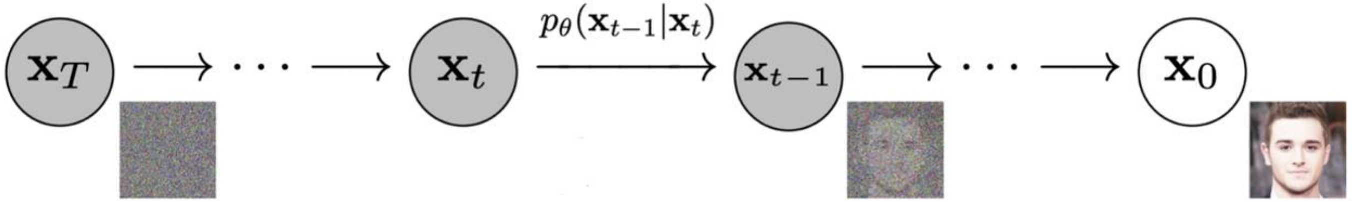
What is the diffusion process?

Traditional Diffusion Model performs the following two steps:

- **Forward Process:** gradually inject Gaussian noise to destruct the original data.



- **Reverse Process:** reconstruct the original distribution from noised data step by step.

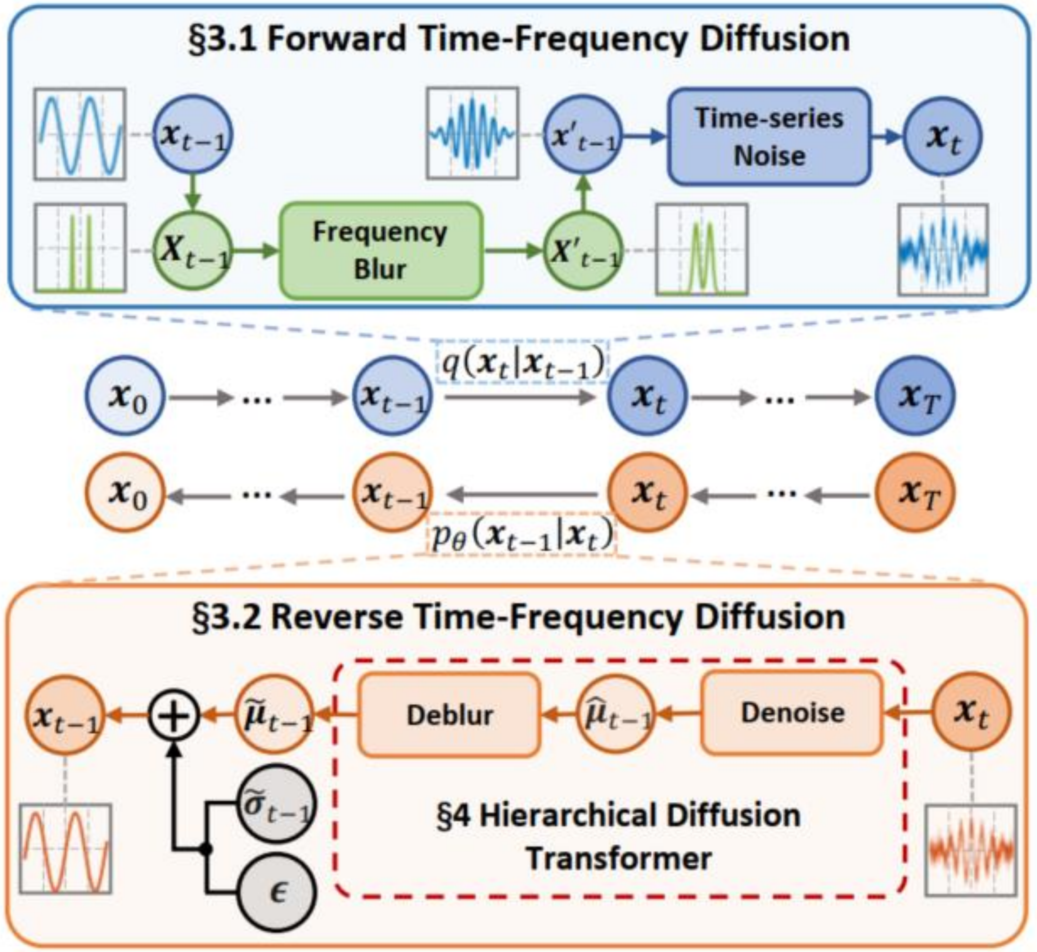


💡 **A well-trained Diffusion Model recover a learned data distribution from a random Gaussian sample, thereby achieving data generation.**

How to adapt Diffusion to RF?

From the **theoretical** perspective, we propose the **Time-Frequency Diffusion**:

- **Forward Process:** iteratively add Gaussian noise in the time domain, while blurs the spectrum in the frequency domain.
- **Reverse Process:** reconstruct x_{t-1} by denoising x_t in the time domain, while deblurring in the frequency domain.

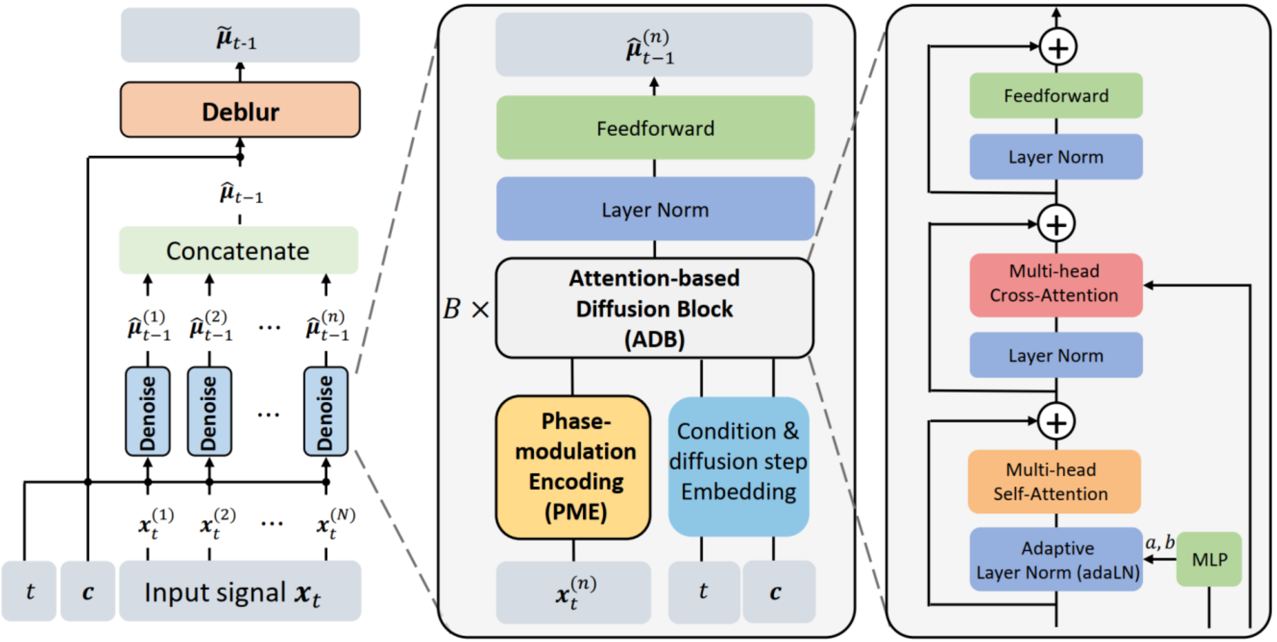


Time-Frequency Diffusion emphasizes the time-domain amplitude accuracy and the frequency-domain continuity

How to adapt Diffusion to RF?

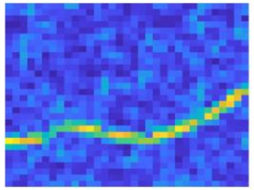
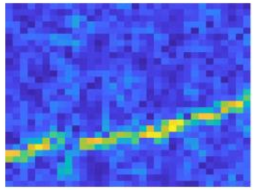
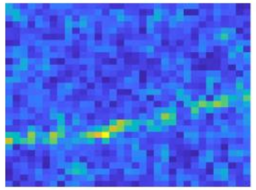
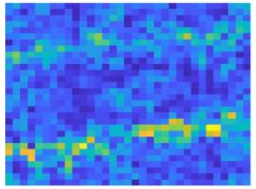
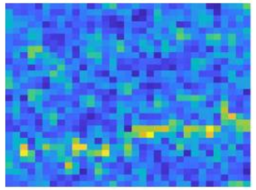
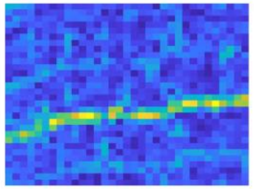
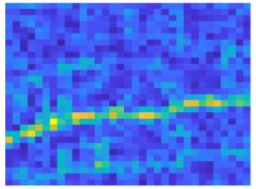
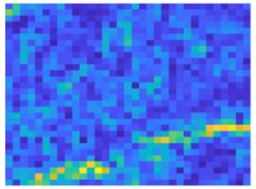
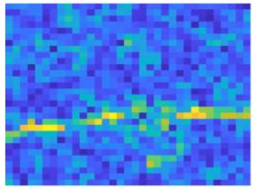
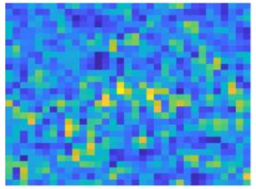
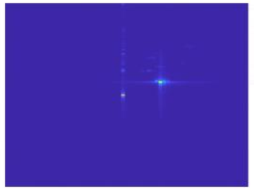
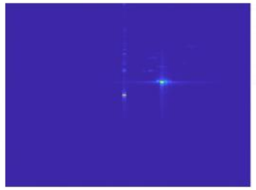
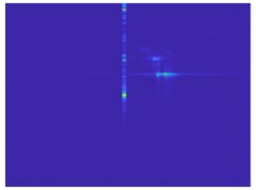
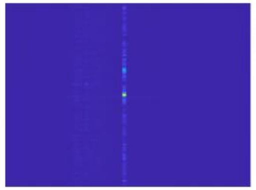
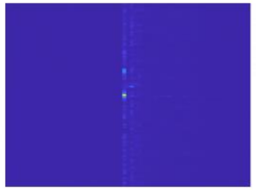
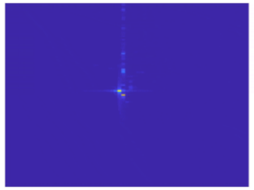
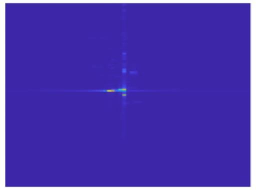
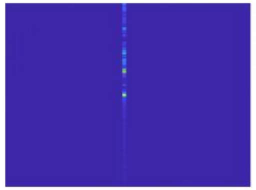
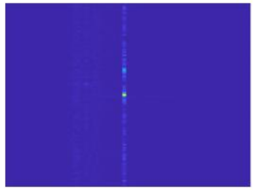
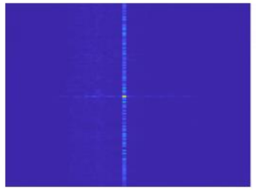
From the **implementation** perspective, we propose **Hierarchical Diffusion Transformer**:

- **Hierarchicy:** HDT is divided into 2 stages, the denoising and deblurring stages.
- **Complex-valued design:** modify the attention module and feed-forward module to adapt to complex signals.
- **Phase modulation:** a positional encoding scheme tailed for complex-valued signals.


$\tilde{\mu}_{t-1}$
Deblur
$\hat{\mu}_{t-1}$
Concatenate
$\hat{\mu}^{(1)}_{t-1}$ $\hat{\mu}^{(2)}_{t-1}$ … $\hat{\mu}^{(n)}_{t-1}$
Denoise
Denoise
…
Denoise
$x^{(1)}_t$ $x^{(2)}_t$ … $x^{(N)}_t$
t
c
Input signal x_t
$\hat{\mu}^{(n)}_{t-1}$
Feedforward
Layer Norm
B ×
Attention-based Diffusion Block (ADB)
Phase-modulation Encoding (PME)
Condition & diffusion step Embedding
$x^{(n)}_t$
t
c
Feedforward
Layer Norm
Multi-head Cross-Attention
Layer Norm
Multi-head Self-Attention
Adaptive Layer Norm (adaLN)
a, b
MLP

HDT adopts a **hierarchical architecture** to decouple noise and spectrum blur.

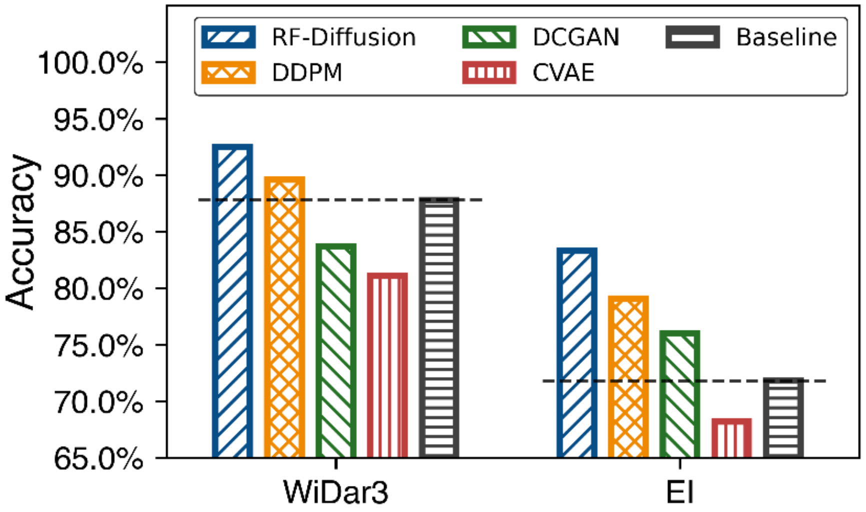
Evaluation Results

		Authentic	RF-Diffusion	DDPM	DCGAN	CVAE
Wi-Fi CSI (Doppler Frequency Spectrum)	Sample ①					
	Sample ②					
FMCW (Range-Velocity Doppler)	Sample ①					
	Sample ②					

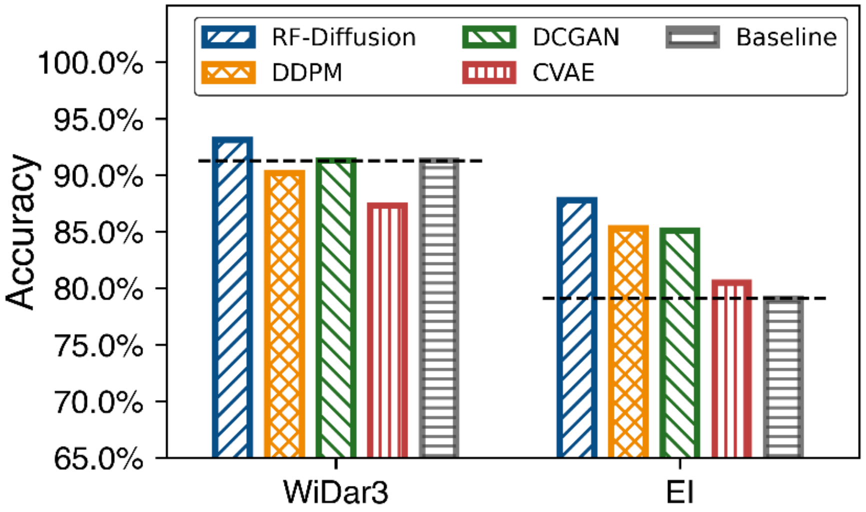
Case Study: Sensing data augmentation

Mixing synthesized wireless data with the original training set to jointly train gesture recognition models results in performance improvements:

- In cross-domain scenarios, Widar and EI saw increases of 4.7% and 11.5%, respectively;
- In in-domain situations, Widar and EI experienced gains of 1.8% and 7.5%, respectively;



Cross Domain

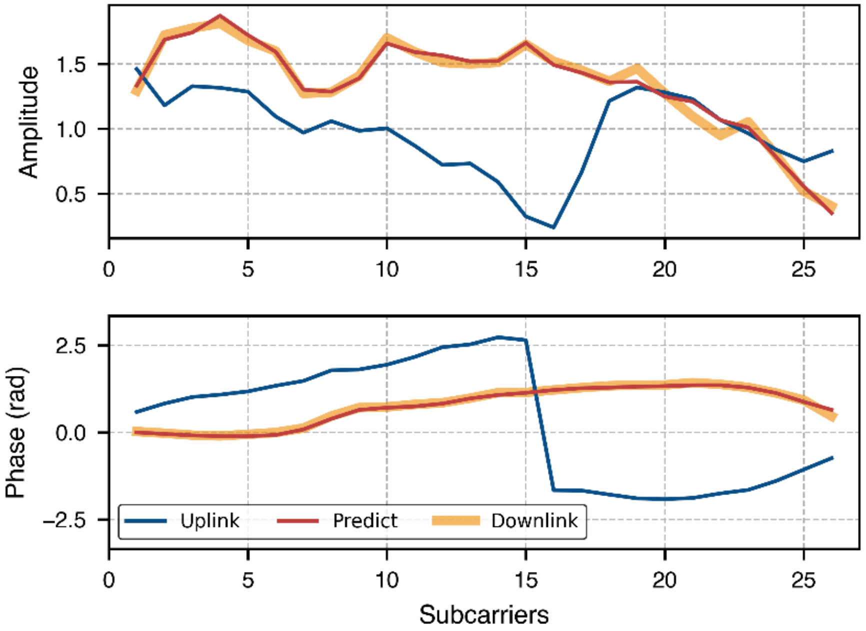


In Domain

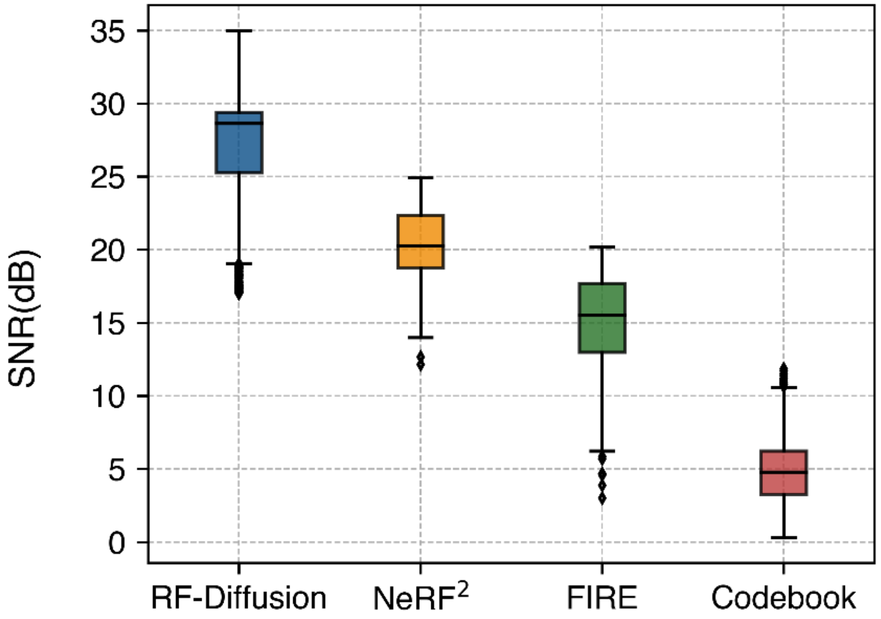
Case Study: Channel Prediction

Taking uplink channel CSI as a condition, RF-Diffusion can generate the downlink CSI:

- Compared to the SOTAs, RF-Diffusion achieved over 5.9 dB performance gain;
- The MSE of channel estimation has been reduced by about 70%;



Predicted CSI amplitude and phase



SNR Performance

Data Synthesis/Generation

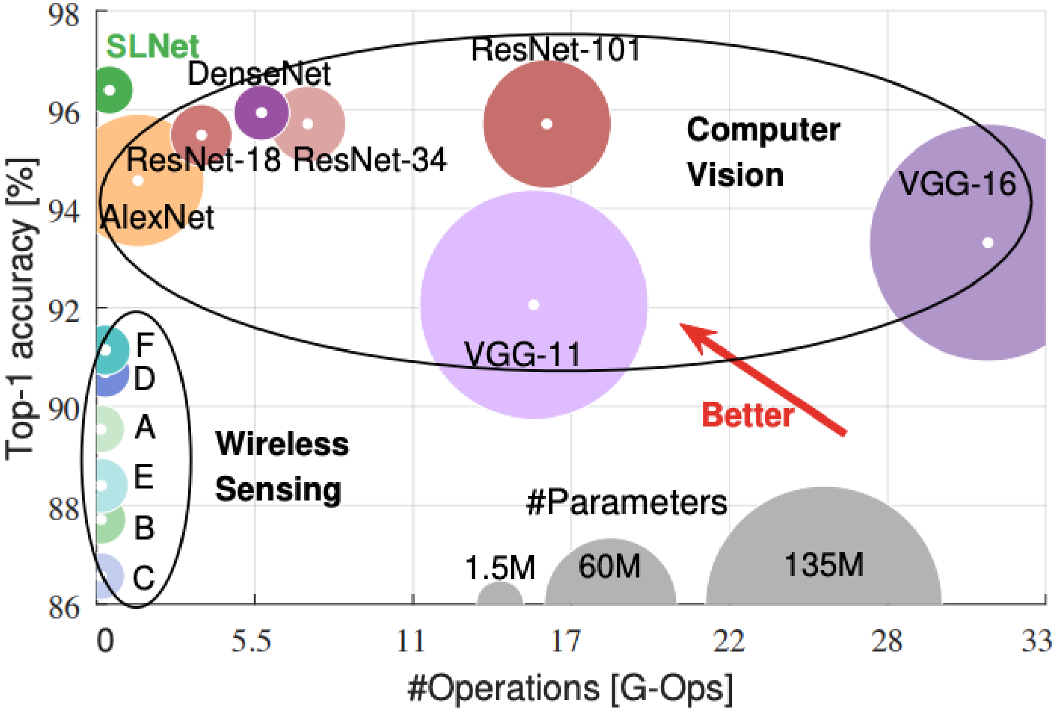
- GenAI for wireless data generation is a promising direction
- NeRF (Neural Radiance Field) for Radio Frequency

- Open question:
  - How to generate large-scale data?
  - How to generate large-scale, high-quality data?

SLNet: Spectrogram Learning Neural Network

- Do we need separate design for DWS or not? And how?

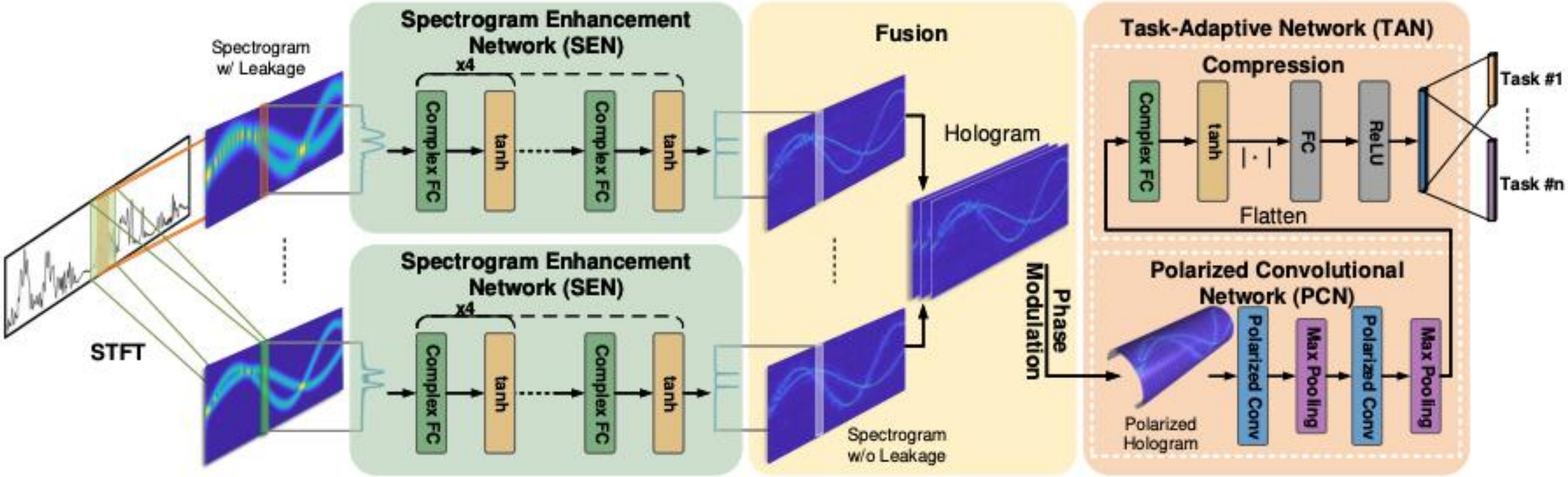
- Many are based on mature CV models, accurate but quite big
- Existing DWS models are relatively small but less accurate



SLNet: A Spectrogram Learning Neural Network for Deep Wireless Sensing, NSDI’23

SLNet

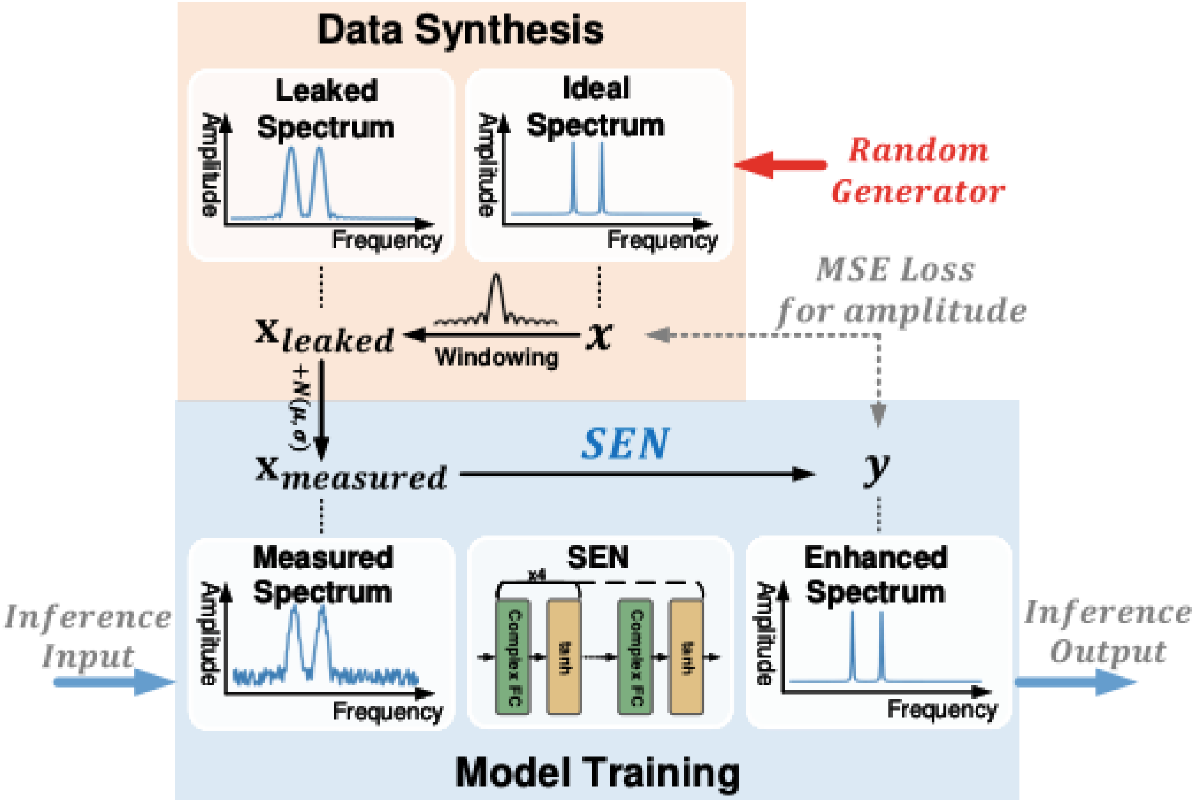
- An attempt for DWS models
    - Based on spectrogram learning



SLNet

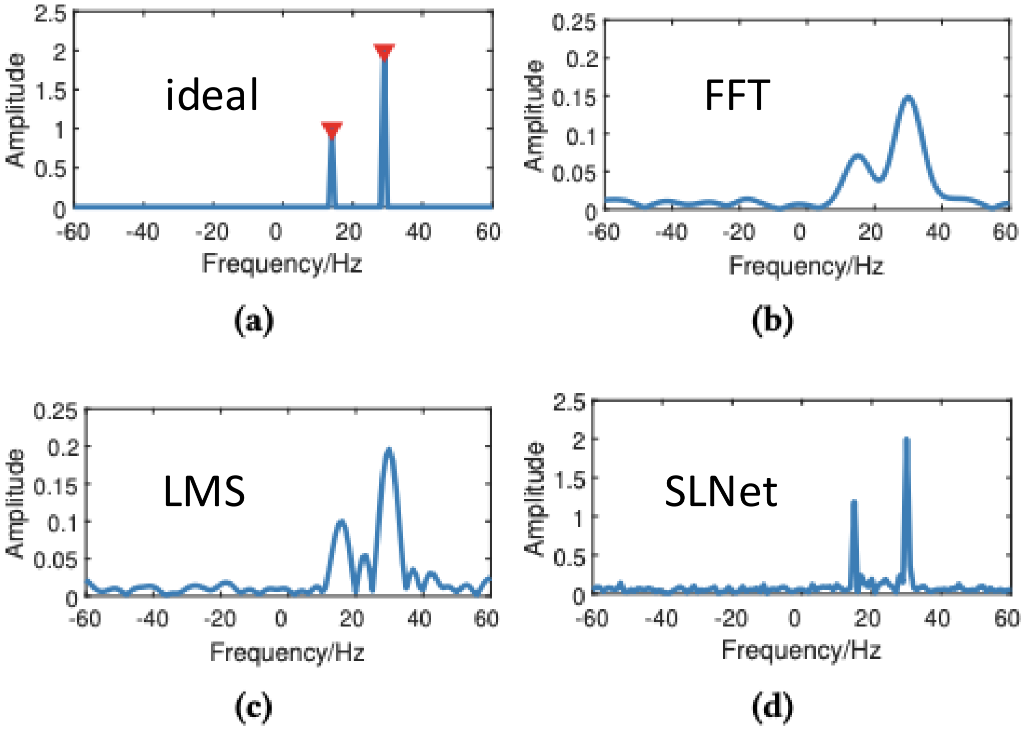
- How to generate high-fidelity spectrograms?

- Spectrograms generated by STFT suffer from spectral leakage, an inherent issue of FFT.
- SEN: Spectrogram Enhancement Network to learn the best function to minimize or nearly eliminate the leakage.



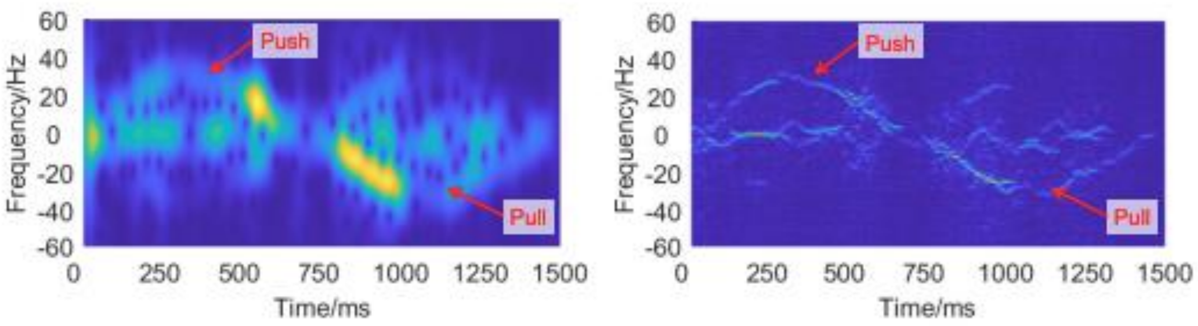
SLNet

- How to generate high-fidelity spectrograms?



(a) (b)

(c) (d)



STFT spectrogram

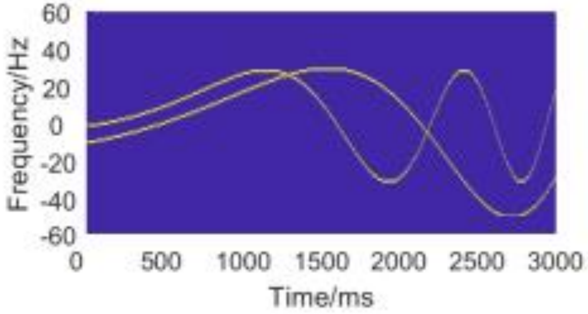
SEN spectrogram

The spectrogram of a pushing and pulling gesture.

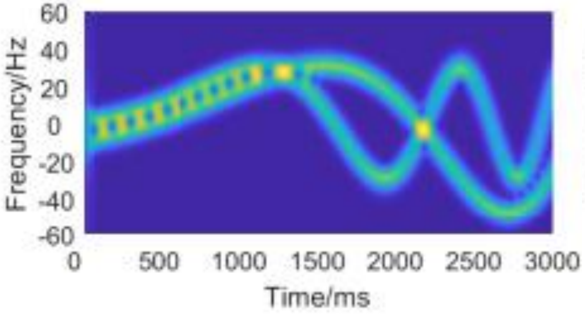
SLNet

- How to trade-off time-frequency resolution?
  - Achieve high frequency resolution by using long windows
  - Capture fast-changing frequencies by using short windows

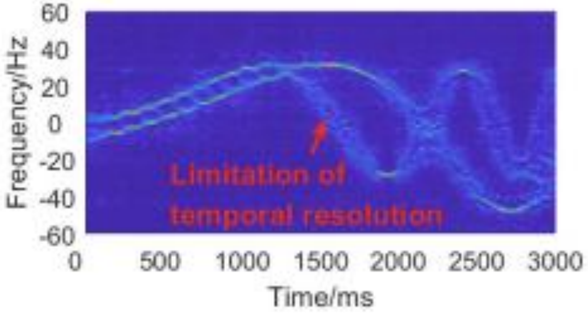
- Use a bank of sliding windows with different lengths
  - Concatenated as multiple channels, forming a multi-channel “hologram”



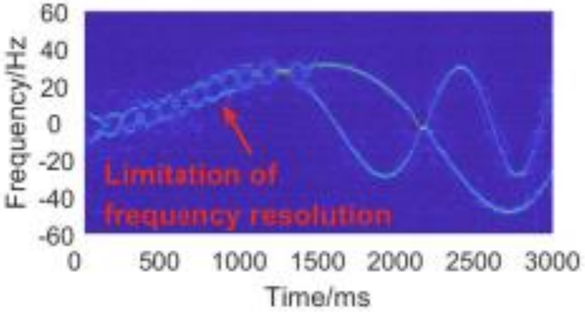
Ideal spectrogram



STFT spectrogram (w=251 ms)



SEN spectrogram (w=251 ms)



SEN spectrogram (w=125 ms)

SLNet

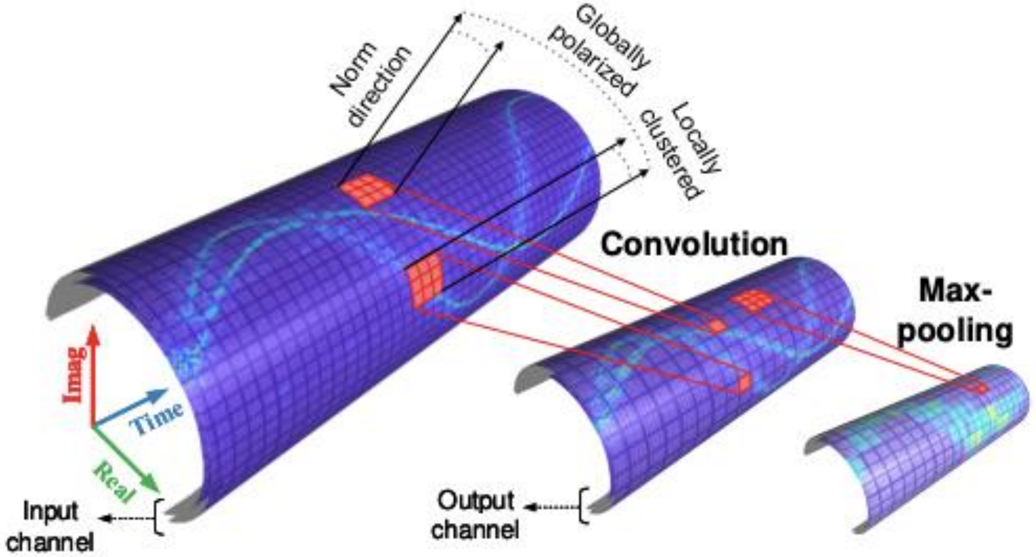
- How does a model simultaneously preserve local dependency and global discrimination?

- Problem
  - CNN mainly learns local features irrespective of global locations of objects in an image
  - Not ideal for spectrogram learning, as the global locations, i.e., frequencies, are correlated with the physical properties of a person’s activities, which is not shift-invariant.

SLNet: Spectrogram Learning NN

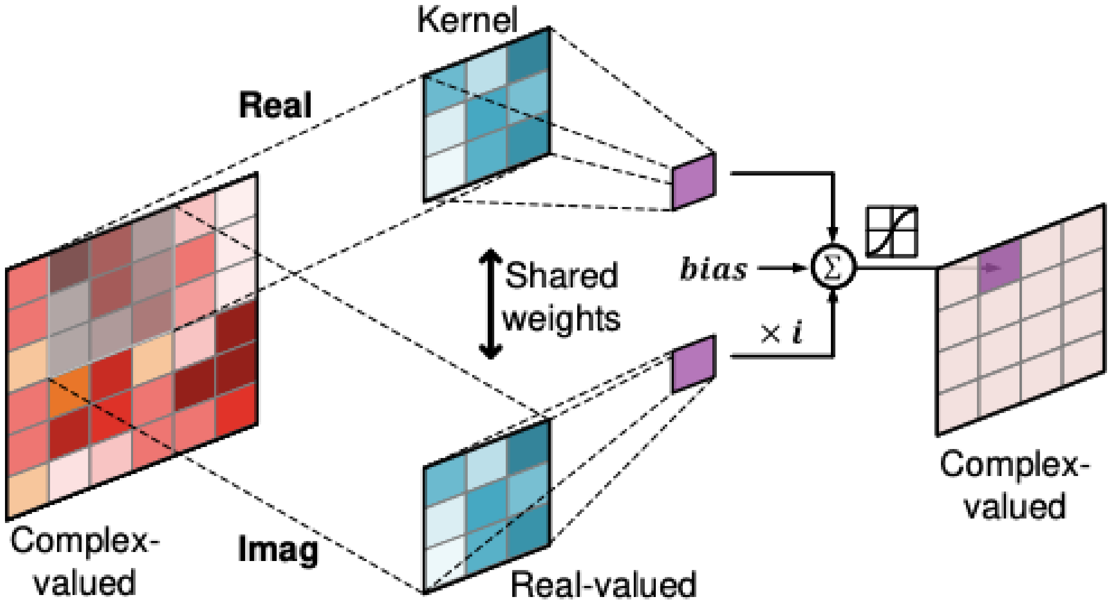
- How does a model simultaneously preserve local dependency and global discrimination?

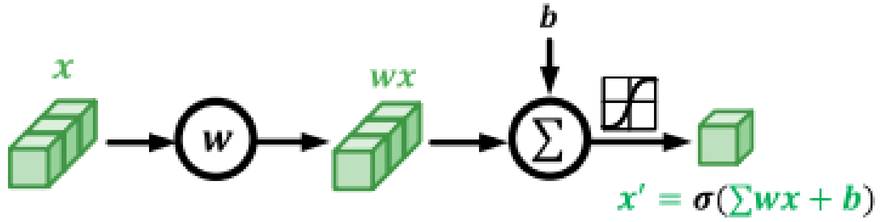
- Solution: Polarized Convolutional Network
  - Polarize the spectrograms via linearly modulated phase information
  - Adjacent frequency components have similar phases while the distant ones have discriminative phases
  - Use phases that vary linearly along the frequency dimension, making them locally unaltered while globally differentiated



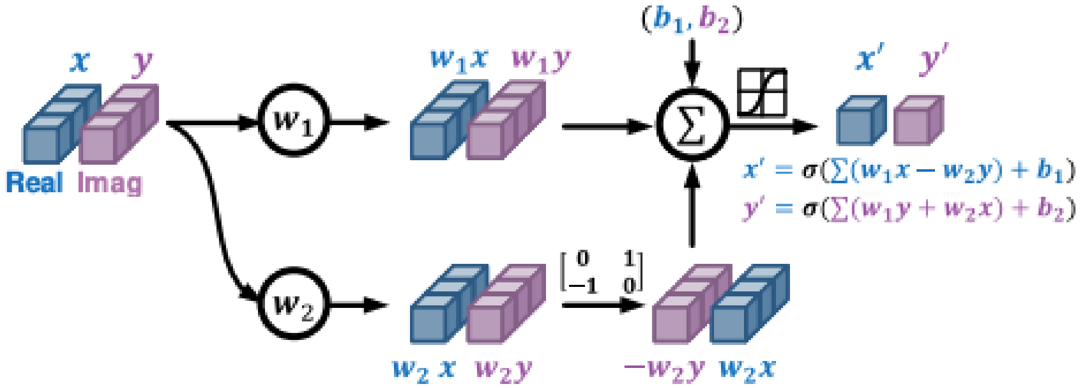
SLNet

- How to learn from the polarized complex-valued spectrograms?





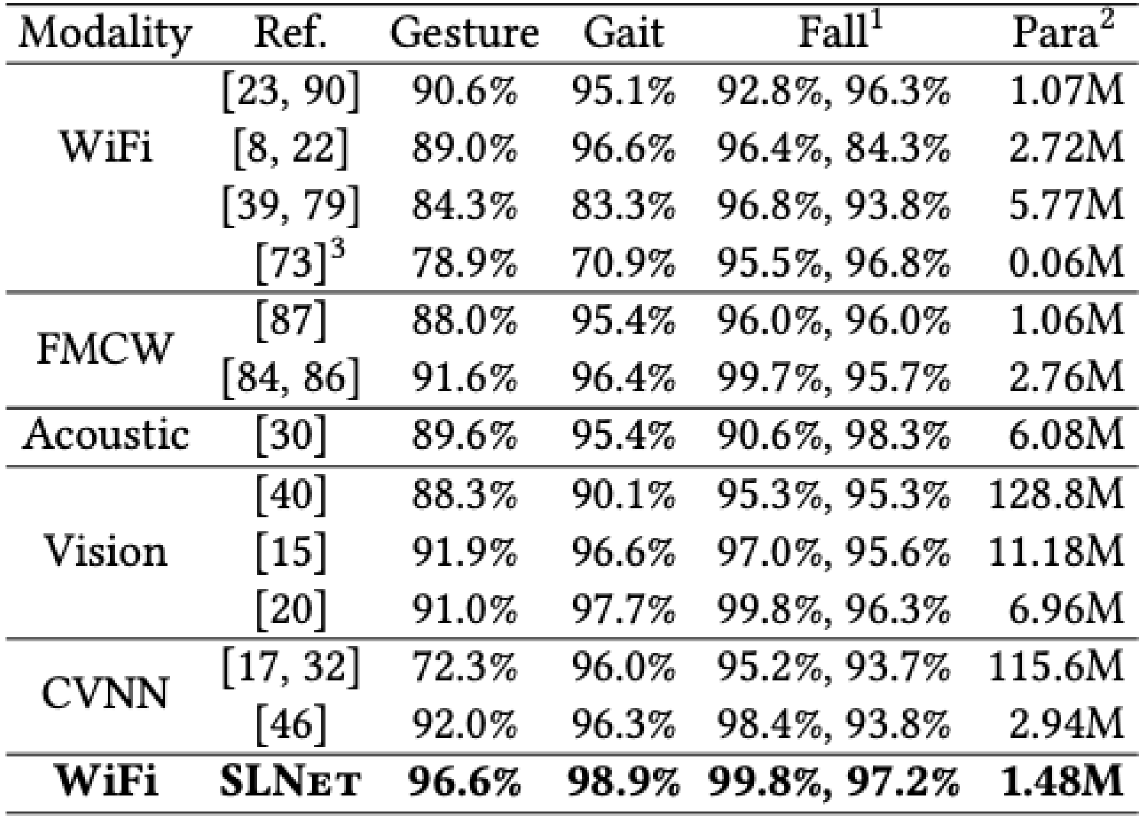
real-valued vs. complex-valued neurons



x y
Real Imag
w_1
w_2
$w_1 x$ $w_1 y$
(b_1, b_2)
Σ
x' y'
$x' = \sigma(\sum(w_1 x - w_2 y) + b_1)$
$y' = \sigma(\sum(w_1 y + w_2 x) + b_2)$
$w_2\, x$ $w_2 y$
$\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$
$-w_2 y$ $w_2 x$

SLNet

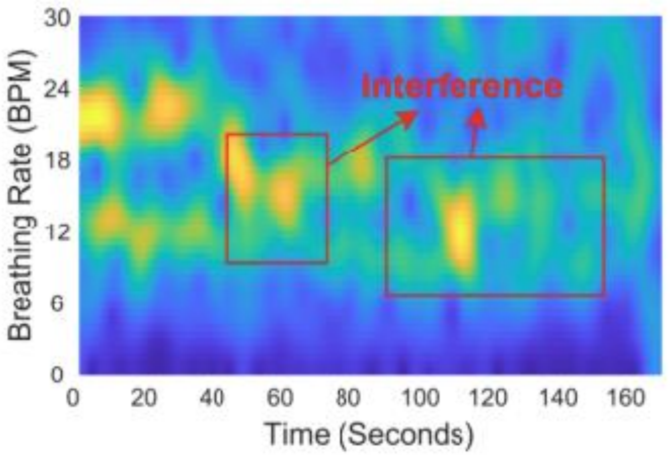
- How well does it work?



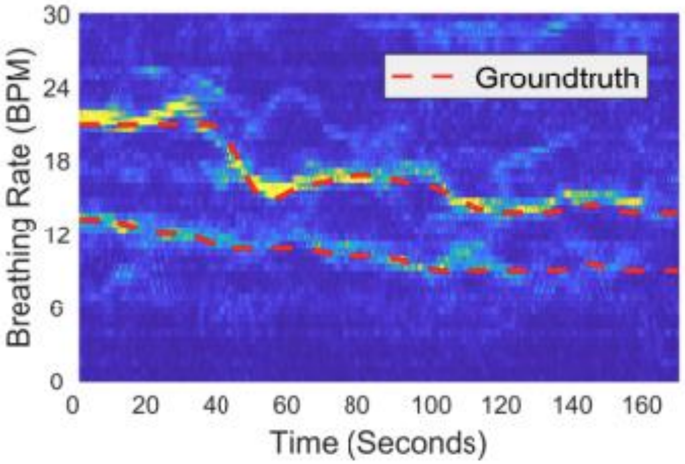
Modality	Ref.	Gesture	Gait	Fall[1]	Para[2]
WiFi	[23, 90]	90.6%	95.1%	92.8%, 96.3%	1.07M
	[8, 22]	89.0%	96.6%	96.4%, 84.3%	2.72M
	[39, 79]	84.3%	83.3%	96.8%, 93.8%	5.77M
	[73][3]	78.9%	70.9%	95.5%, 96.8%	0.06M
FMCW	[87]	88.0%	95.4%	96.0%, 96.0%	1.06M
	[84, 86]	91.6%	96.4%	99.7%, 95.7%	2.76M
Acoustic	[30]	89.6%	95.4%	90.6%, 98.3%	6.08M
Vision	[40]	88.3%	90.1%	95.3%, 95.3%	128.8M
	[15]	91.9%	96.6%	97.0%, 95.6%	11.18M
	[20]	91.0%	97.7%	99.8%, 96.3%	6.96M
CVNN	[17, 32]	72.3%	96.0%	95.2%, 93.7%	115.6M
	[46]	92.0%	96.3%	98.4%, 93.8%	2.94M
WiFi	**SLNet**	**96.6%**	**98.9%**	**99.8%, 97.2%**	**1.48M**

SLNet

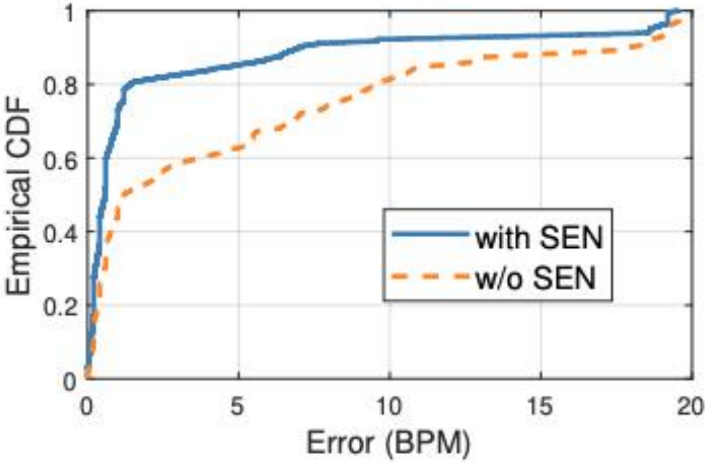
- How well does it work?
  - Multi-user breath estimation



(a) The raw spectrogram from traditional FFT. Spectral leakage causes severe interference for the close frequency components, making it hard to detect the breath rates of different people.



(b) The enhanced spectrogram with SEN. Frequency components can be clearly discriminated.



(c) The accuracy of breath rate estimation with raw and enhanced spectrograms.

DWS Models

- There are too many to list…

- Cross-domain generalizability is still a big issue
  - which is difficult to verity without a comprehensive dataset.
- Target to solve more challenging problems using DL, because well-addressed applications do not necessitate DL.

- A step further, Foundation Model for wireless?

Questions?

- Thank you!