Human-Centric RF Sensing

Pose Estimation, Cardiac Monitoring, and Self-Supervised Learning

Dr.Yan Chen**, Dr. Dongheng Zhang, **Dr. Zhi Lu

`https://ustc-ip-lab.github.io`

Intelligent Perception Lab
School of Cyber Science and Technology
University of Science and Technology of China

3, Dec, 2024, APSIPA Tutorial

Contents

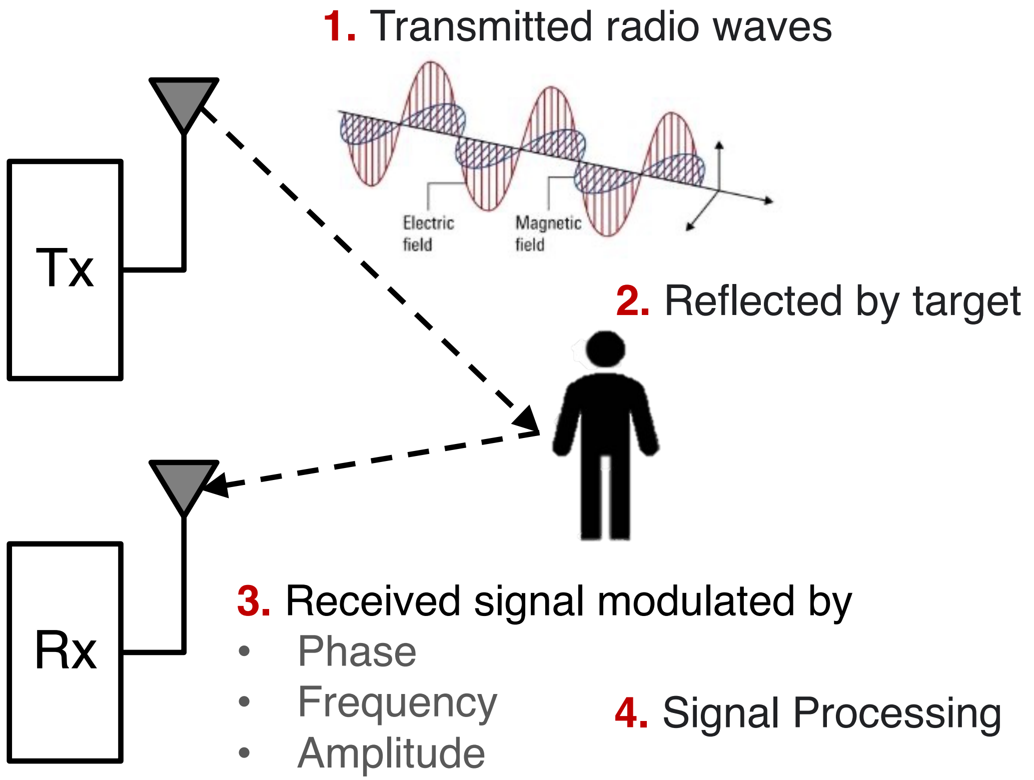
中国科学技术大学
University of Science and Technology of China

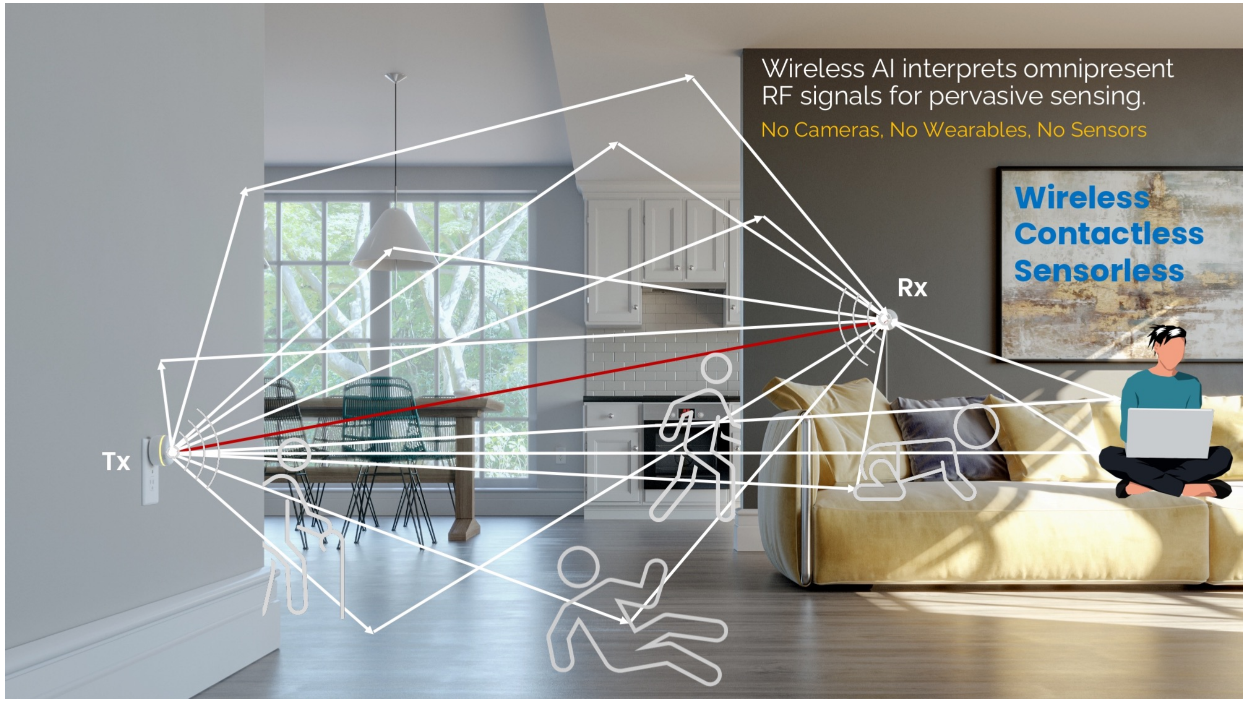
Introduction

Principle, Advantages and Challenges

RF-based human sensing uses RF signals' interaction with the human body—via reflection, absorption, and scattering—for presence, motion, physiological monitoring, etc.

Sensing Principle





https://innowings.engg.hku.hk/wireless-ai/

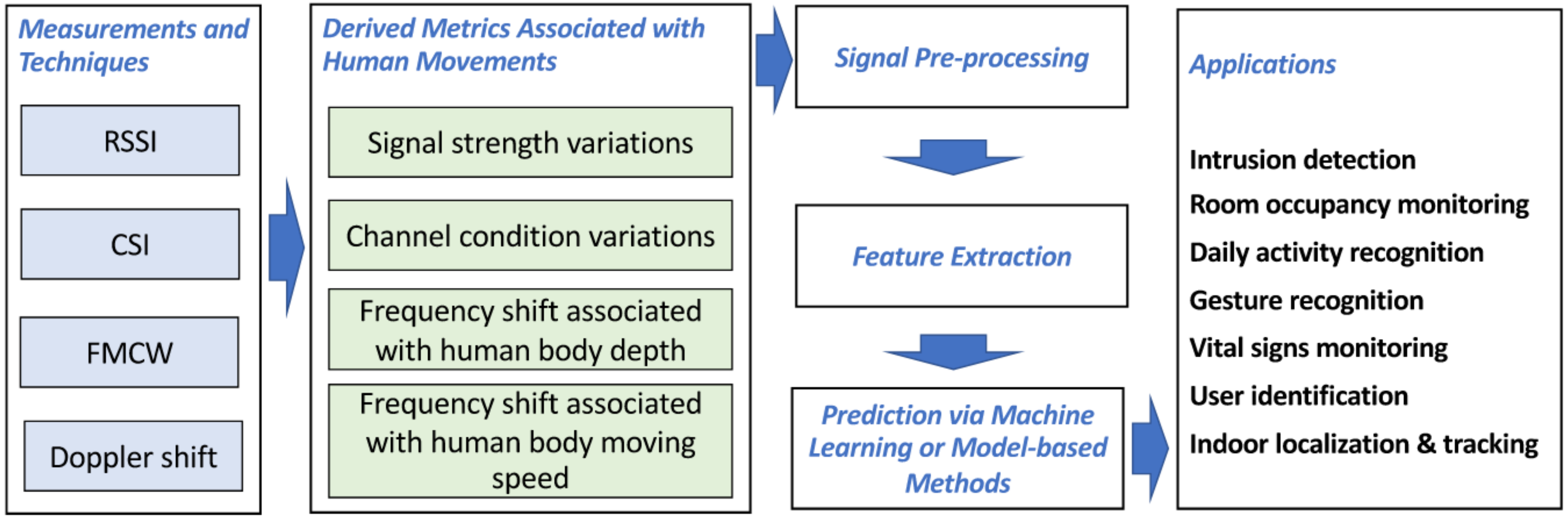
Introduction

Principle, Advantages and Challenges

RF-based human sensing uses RF signals' interaction with the human body—via reflection, absorption, and scattering—for presence, motion, physiological monitoring, etc.

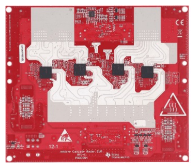
Workflow

Sensors → Information and Algorithm → Applications



Wireless Sensing for Human Activity (Liu et. al 2020)

Diverse types of RF signals vary significantly in terms of ***processing methods***, derived ***information***, and ***performance*** characteristics.



Radar

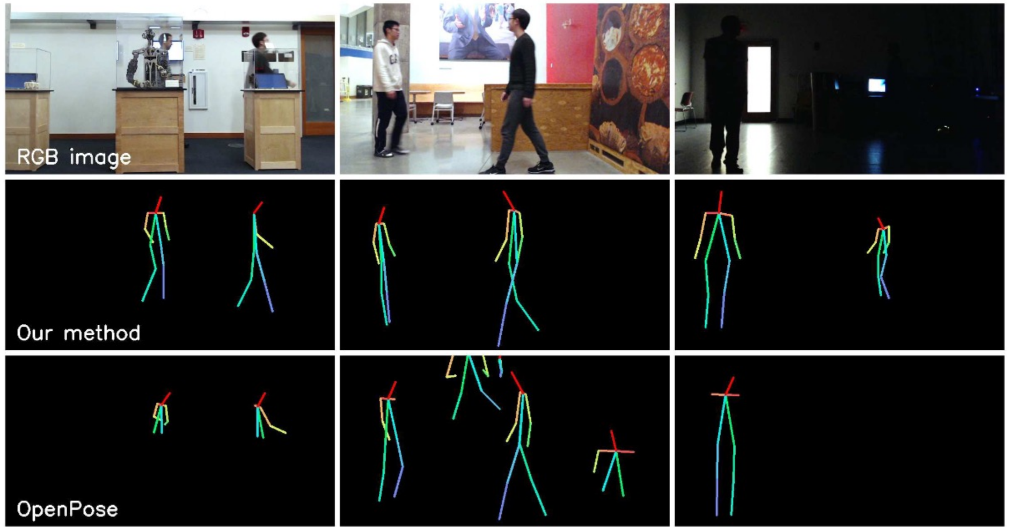
Router

Introduction

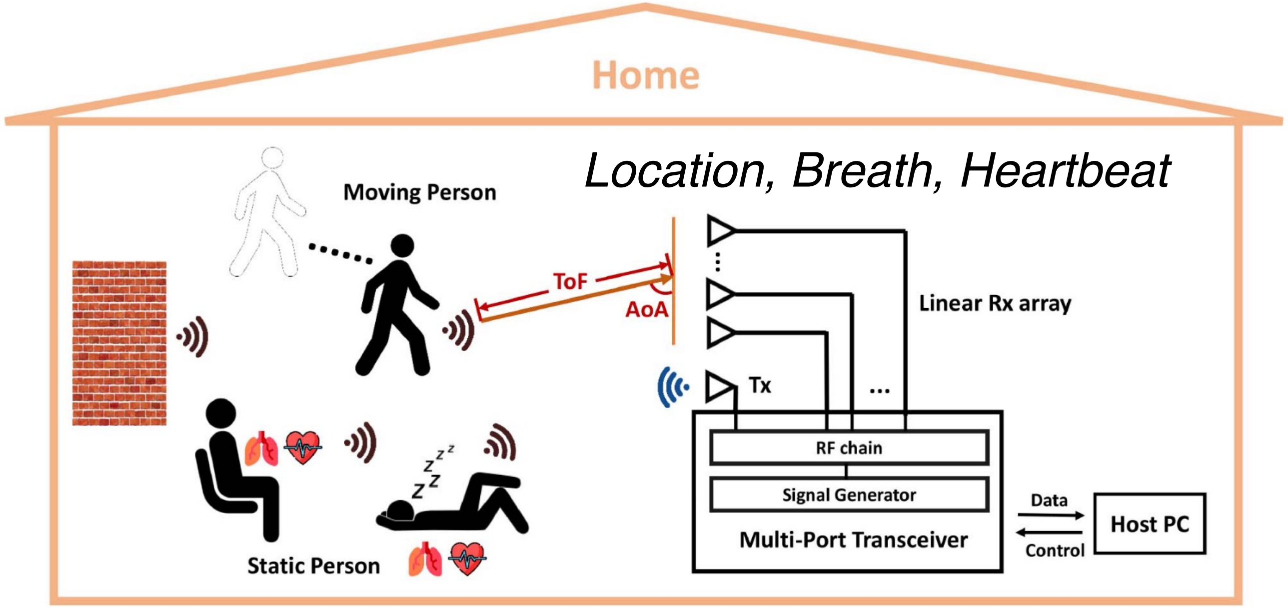
Principle, Advantages and Challenges

RF-based human sensing uses RF signals' interaction with the human body—via reflection, absorption, and scattering—for presence, motion, physiological monitoring, etc.

Applications



Pose estimation (Zhao et. al 2018)



Tracking and vital signs (Zhang et. al 2021)

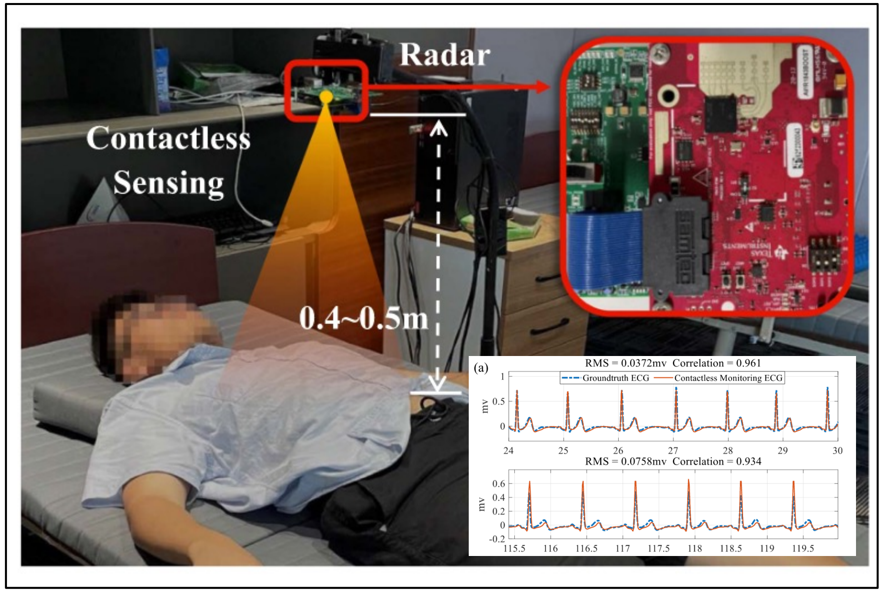
Non- line-of-sight, low-light conditions, privacy-preserving

Introduction

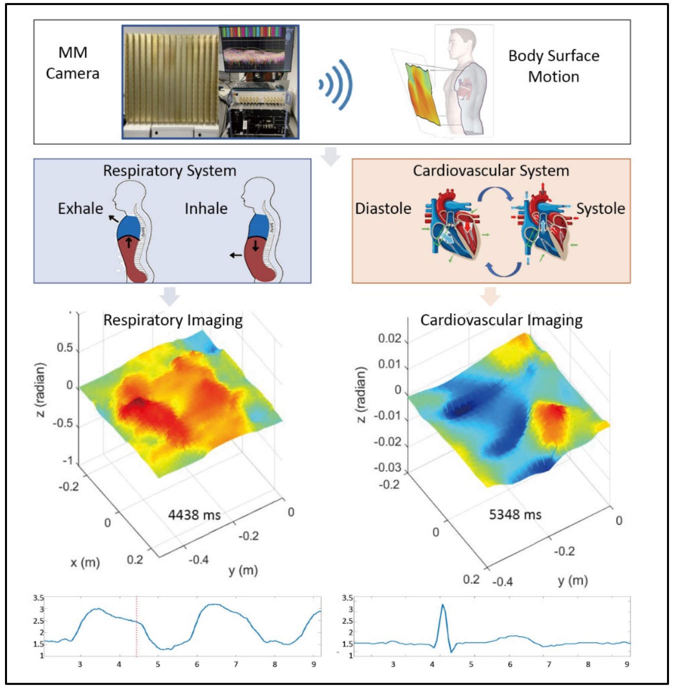
Principle, Advantages and Challenges

RF-based human sensing uses RF signals' interaction with the human body—via reflection, absorption, and scattering—for presence, motion, physiological monitoring, etc.

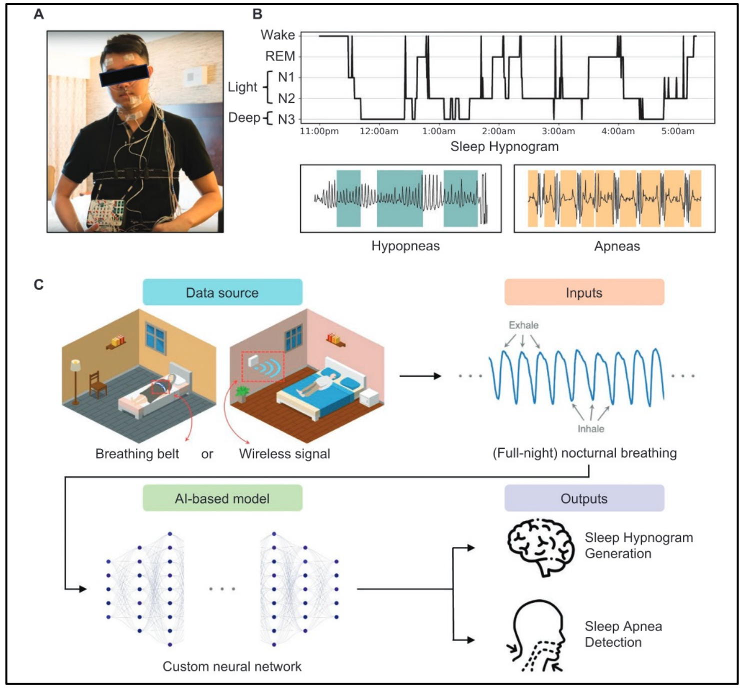
Applications



ECG Monitoring (Chen et. al 2022)



Imaging (Chen et. al 2022)



Sleep Monitoring (He et. al 2024)

Contactless sensing, user-friendly, cost-effective

Introduction

Principle, Advantages and Challenges

Challenges

- Low spatial resolution: challenges in precisely locating or differentiating small details.
- Environmental interference: susceptibility to noise from surrounding objects and signal obstructions.
- Hardware Limitations: dependence on specialized equipment, which may increase cost and complexity.
- Non-Intuitive Operation: difficulty in interpreting results compared to traditional, visually-driven methods.

Introduction

Selected Publications

Pose Estimation

1) Wu et al. RFMask: A Simple Baseline for Human Silhouette Segmentation with Radio Signals. ***TMM*** 2022.
2) Yu et al. RFPose-OT: RF-Based 3D Human Pose Estimation via Optimal Transport Theory. ***FITEE*** 2023.
3) Xie et al. RPM: RF-Based Pose Machines. ***TMM*** 2023.
4) Xie et al. RPM 2.0: RF-Based Pose Machines for Multi-Person 3D Pose Estimation. ***TCSVT*** 2023.
5) Yu et al. MobiRFPose: Portable RF-Based 3D Human Pose Camera. ***TMM**,* 2024.
6) Yu et al. RFGAN: RF-Based Human Synthesis. ***TMM*** 2023.

ECG Monitoring

1) Chen et al. MMCamera: An Imaging Modality for Future RF-Based Physiological Sensing. *In* ***MobiCom*** *(Poster)* 2022.
2) Chen et al. Contactless Electrocardiogram Monitoring with Millimeter Wave Radar. ***TMC*** 2024.
3) Zhang et al. Monitoring Long-Term Cardiac Activity with Contactless Radio Frequency Signals. ***Nature Communications*** 2024. (To appear)

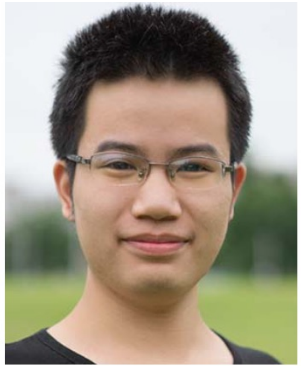
Self-supervised Learning

1) Song et al. RF-URL: Unsupervised Representation Learning for RF Sensing. *In* ***MobiCom*** 2022.
2) Fang et al. PRISM: Pre-Training RF Signals in Sparsity-Aware Masked Autoencoders. In ***INFOCOM*** 2024.
3) Song et al. Unleashing the Potential of Self-Supervised RF Learning With Group Shuffle. ***TMC*** 2024.

Introduction

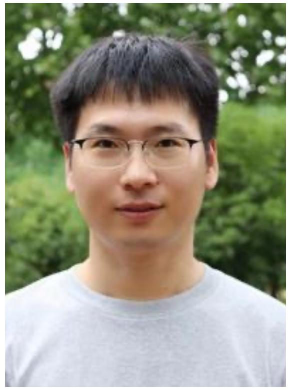
Acknowledgement

Dr. **Chunyang Xie**
Class of 2023
Pose Estimation



Dr. **Cong Yu**
Class of 2023
Pose Estimation

Zhi Wu,
PhD Student
Class of 2025
Pose Estimation



Dr. **Binbin Zhang**
Class of 2024
Vital Sign

Dr. **Jinbo Chen**
Class of 2024
Vital Sign

Ruiyuan Song
PhD Student
Class of 2025
SSL

Liang Fang
Master Student
Class of 2025
SSL

Intelligent Perception Lab

`IPLab: https://ustc-ip-lab.github.io`

Contents

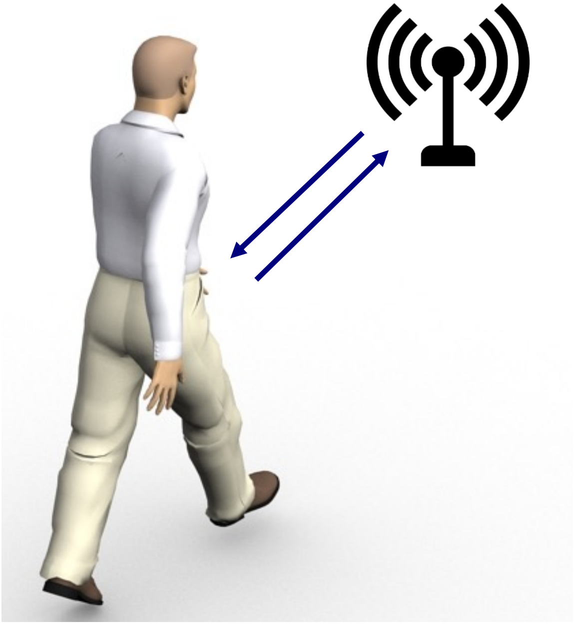
RF-Based Human Pose Sensing

Pose Estimation Methods

Camera & Wearable Devices

- occlusions
- lighting conditions
- active cooperation

RF Signals



RF-based: All-weather & Non-intrusive & Privacy-preserving

中国科学技术大学
University of Science and Technology of China

RF-Based Human Pose Sensing

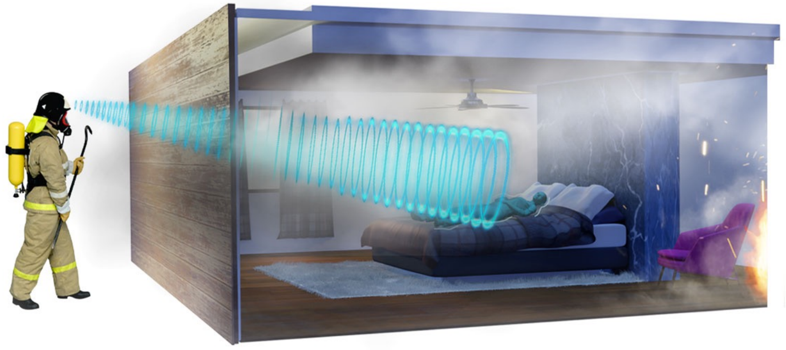
Goal

Obtain human behavior and posture information from RF signals, with all-weather, non-contact and non-line-of-sight characteristics.

Applications

Fall Detection

Anti-terrorism

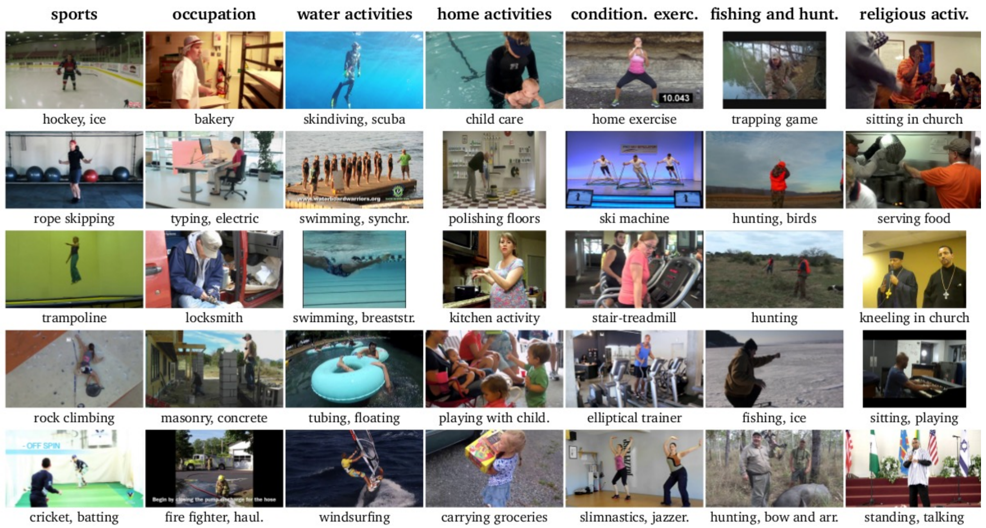


Disaster Rescue

Introduction

- ## Challenges

Publicly available human pose estimation datasets based on visual images



Publicly available human pose estimation datasets based on RF signals

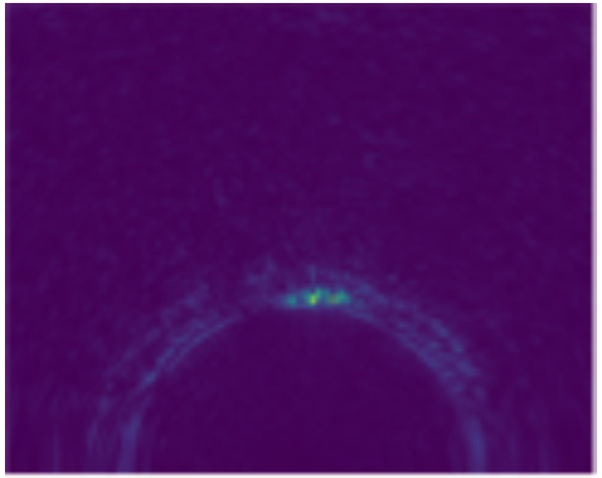
?

Lack of publicly available RF-based human pose sensing datasets

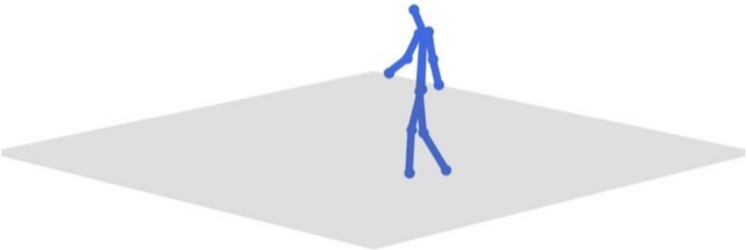
Introduction

- **Challenges**

Not intuitive
Only existence



RF Signals

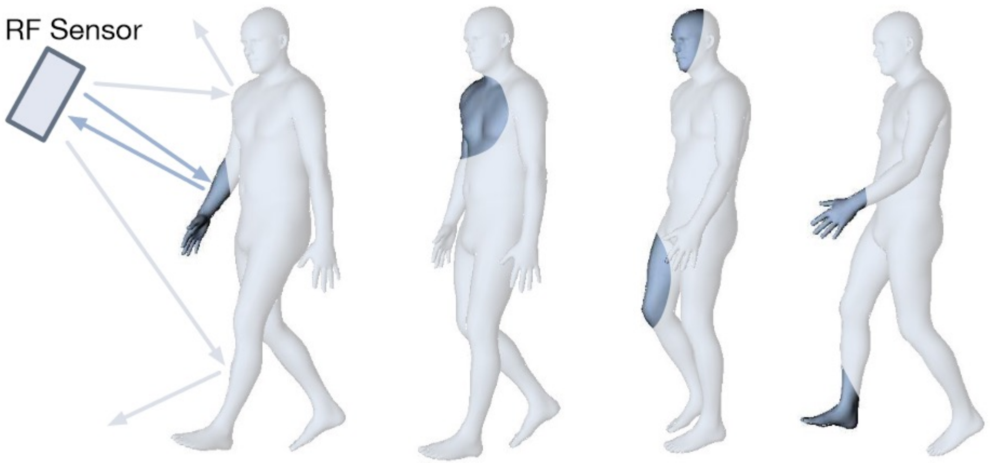


Human Pose

Significant structural difference between the RF signals and the human pose

Introduction

- **Challenges**

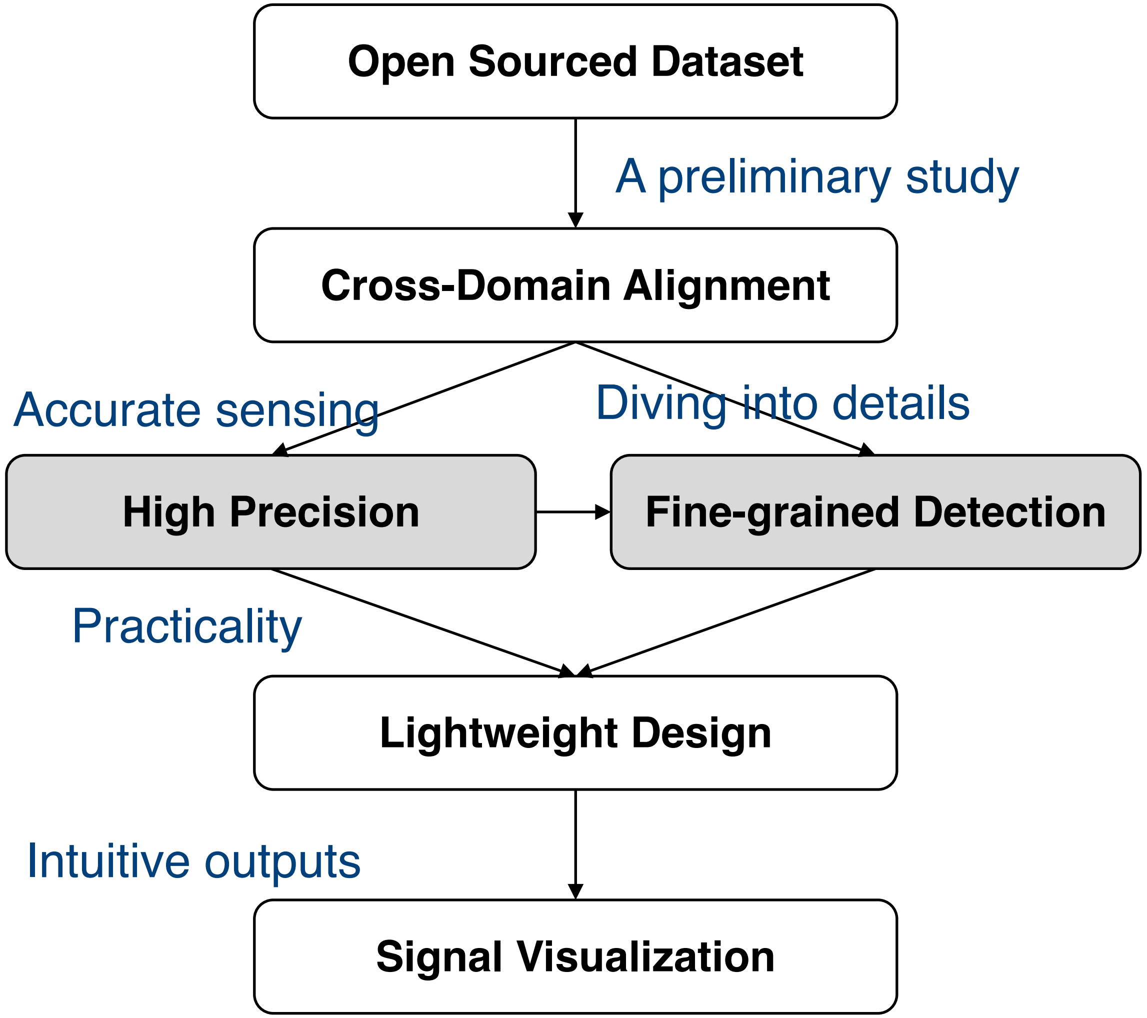


RF signal specular reflection characteristics

The specular reflection characteristics of the RF signals on the human body cause the signals to be sparse and incomplete

Introduction

Our works



- ❑ **RF-Based Human Pose Sensing Dataset**
- ❑ **Pose Estimation with Optimal Transport (FITEE 2023)**
- ❑ **RF-Based Human Pose Estimation with Spatio-Temporal Attention (TCSVT 2023, TMM 2023)**
- ❑ **RF-Based Human Pose Silhouette Segmentation (TMM 2022)**
- ❑ **Lightweight Pose Estimation for Mobile Devices (TMM 2024)**
- ❑ **Multimodal-Based Human Pose Visualization (TMM 2023)**

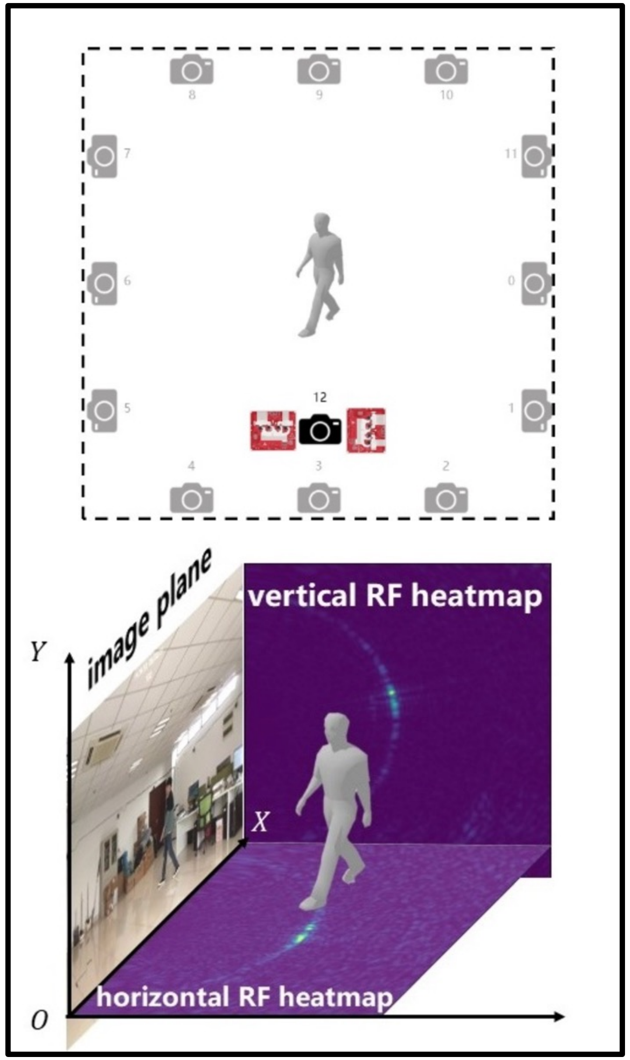
中国科学技术大学
University of Science and Technology of China

RF-Based Human Pose Sensing Dataset

中国科学技术大学
University of Science and Technology of China

RF-Based Human Pose Sensing Dataset

HIBER Dataset



https://github.com/Intelligent-Perception-Lab/HIBER

HIBER (Human Indoor Behavior Exclusive RF dataset)

HIBER(Human Indoor Behavior Exclusive RF dataset) is an open-source mmWave human behavior recognition dataset collected from multiple domains(i.e., various environments, users, occlusions, and actions). It can be used to study human position tracking, human behavior recognition, human pose estimation, and human silhouette generation tasks. The total size of the processed dataset is 400GB, including RF heatmaps, RGB images, 2D/3D human skeletons, bounding boxes, and human silhouette ground truth. Following, we introduce the composition and implementation details of this dataset.

How to access the HIBER dataset

To obtain the dataset, please sign the agreement by yourself. And **additionally**:

- If you are a researcher from China, please ensure that the agreement is stamped with the official seal of your institution.
- If you are not from China, please ask your director or team leader to sign the agreement.

Once stamped/signed, you can scan and send it to wzwyyx@mail.ustc.edu.cn. Then you will receive a notification email that includes the download links of the dataset within seven days. Thank you for your cooperation.

RF-Based Human Pose Sensing Dataset

University of Science and Technology of China

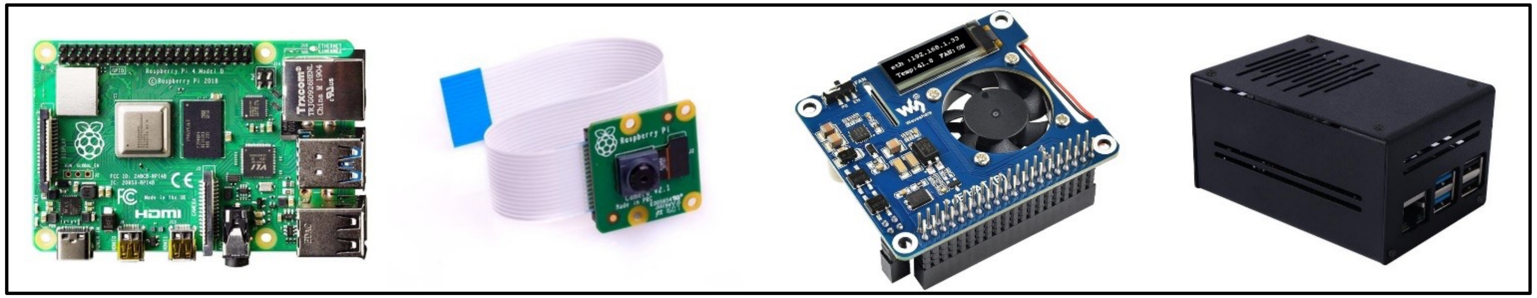
Problem: *To collect RF data and its corresponding human pose.*

Challenges

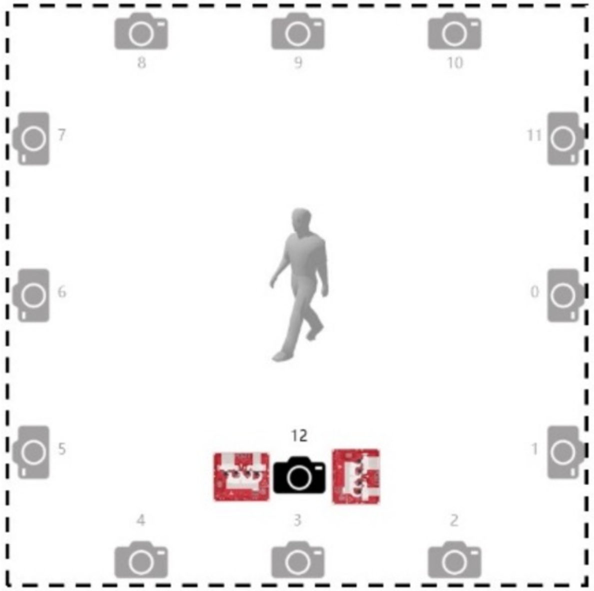
- Manually labeling ground-truth 3D pose for RF signal is infeasible
- Wearable devices, such as Vicon, are very expensive.

Solution

- Multi-camera system to obtain the 3D labels

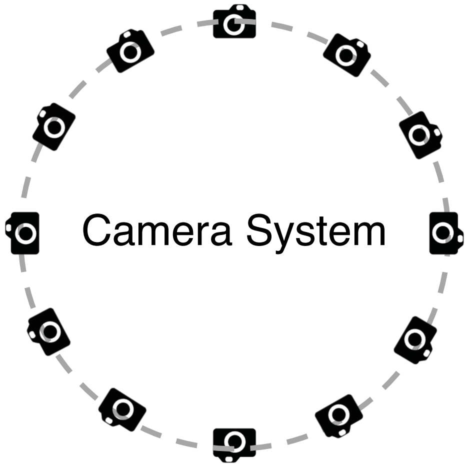


Raspberry Pi 4B with camera module (x12)

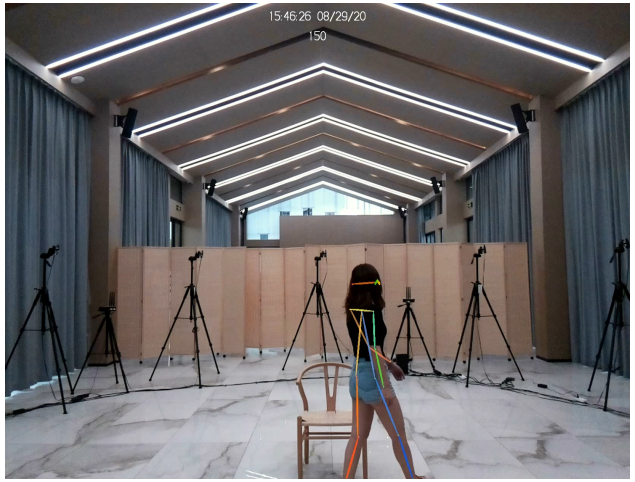


RF-Based Human Pose Sensing Dataset

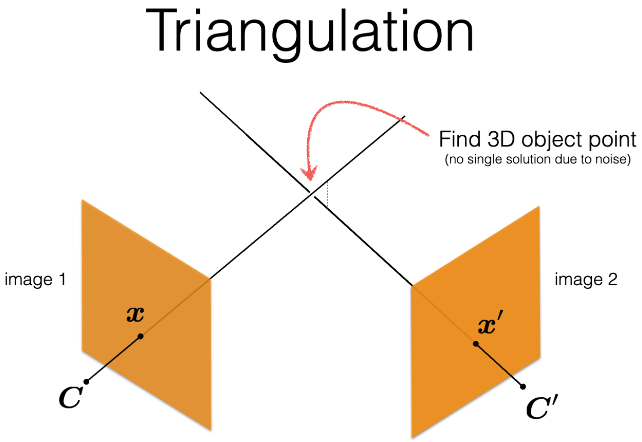
3D Pose



1. Calibrate the Camera System



2. Get 2D pose for each view



3. Reconstruct 3D pose with triangulation

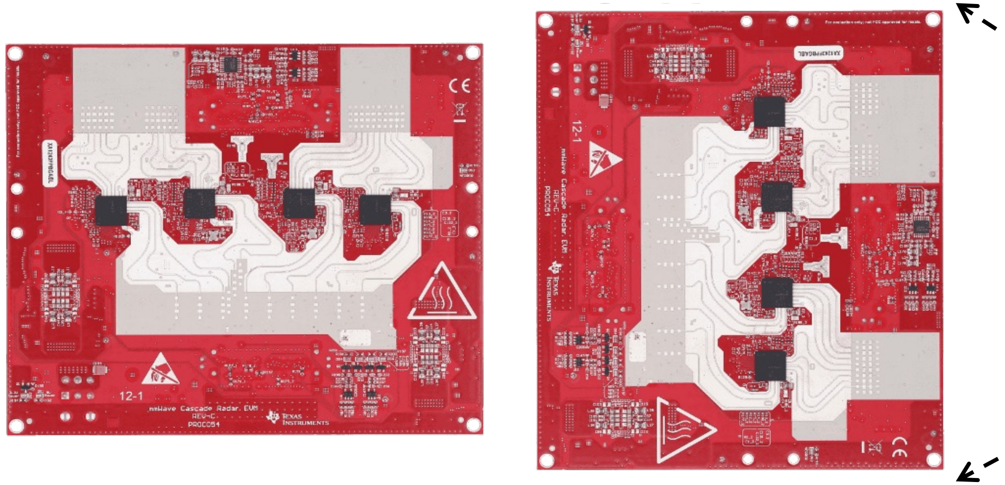
Note for replication

1. We use the well-known Zhang's method for camera calibration.
2. Synchronization between cameras is crucial.
3. We use AlphaPose (Fang et al. 2022) to create 2d Pose.

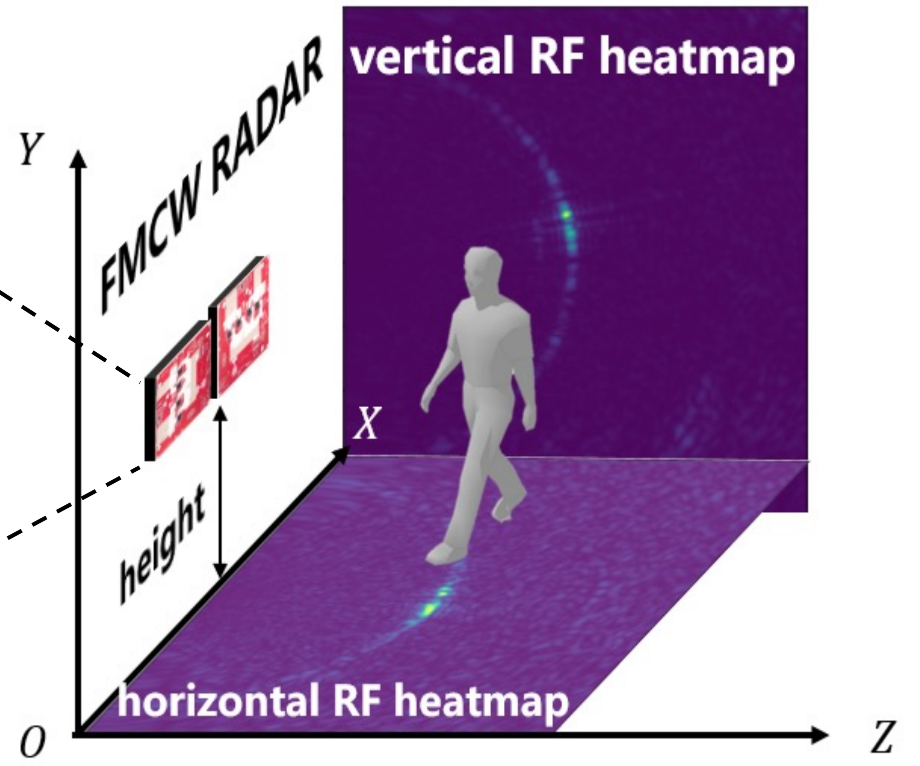
RF-Based Human Pose Sensing Dataset

Radar Device

Model: TI AWR2243



Single Chirp Configuration
77~79 GHz & 79～81 GHz



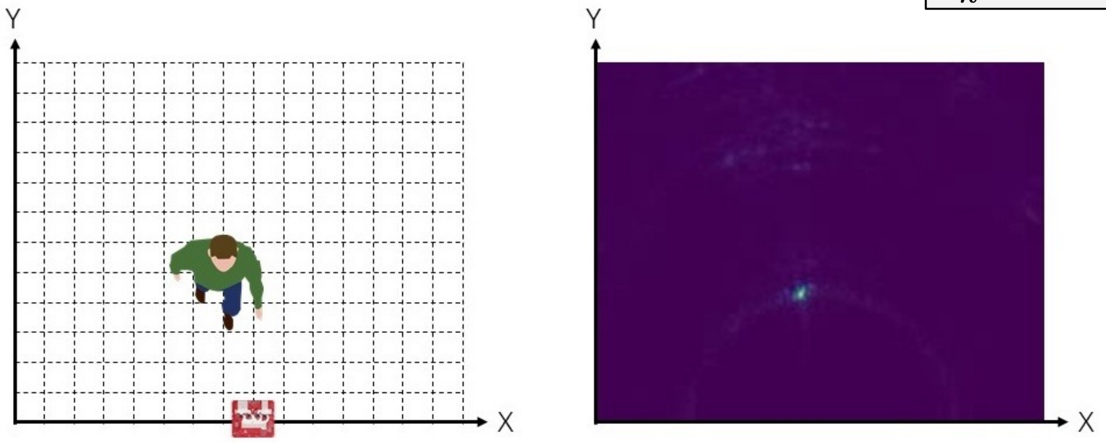
Two **perpendicular** radars

RF-Based Human Pose Sensing Dataset

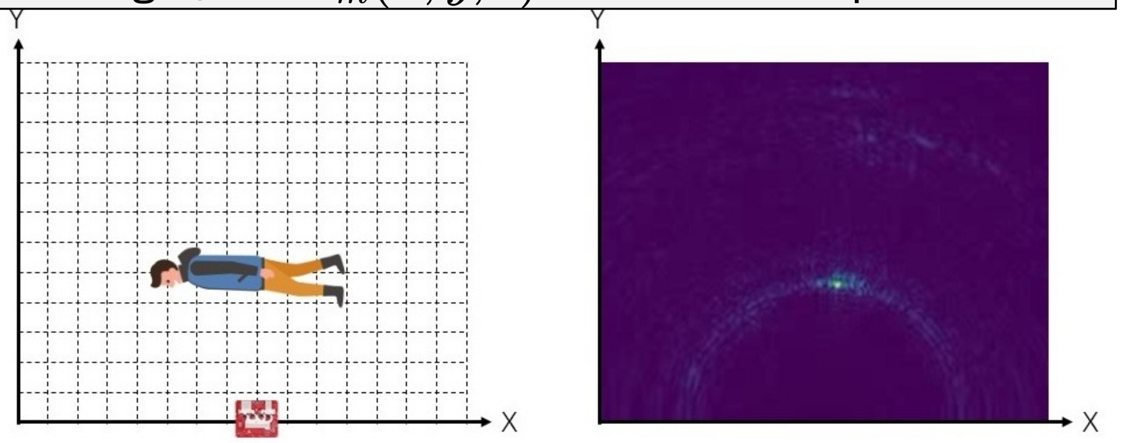
中国科学技术大学
University of Science and Technology of China

Signal Processing

$s_{k,m,t}$ is the signal from the k-th FMCW scan on the m-th antenna, λ_k is the wavelength, and $d_m(x, y, z)$ is the round-trip distance.



Horizontal grid and corresponding data



Vertical grid and corresponding data

Horizontal signal

$$y_{\text{hor}}(x, y, t) = \sum_{k=1}^{K} \sum_{m=1}^{M} s_{k,m,t} \cdot e^{j2\pi \frac{d_m(x,y)}{\lambda_k}}$$

Vertical signal

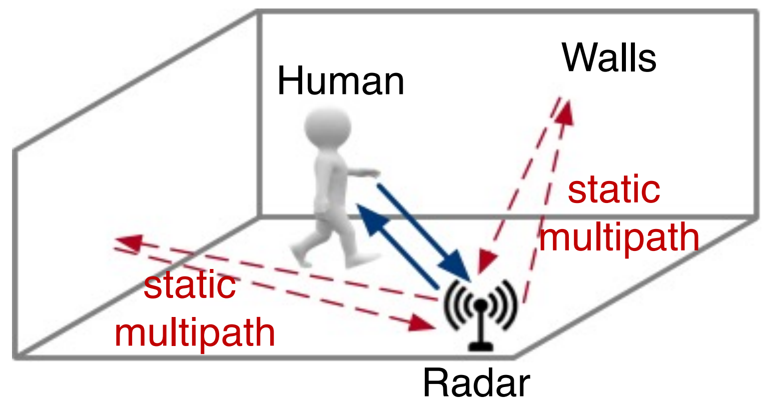
$$y_{\text{ver}}(y, z, t) = \sum_{k=1}^{K} \sum_{m=1}^{M} s_{k,m,t} \cdot e^{j2\pi \frac{d_m(y,z)}{\lambda_k}}$$

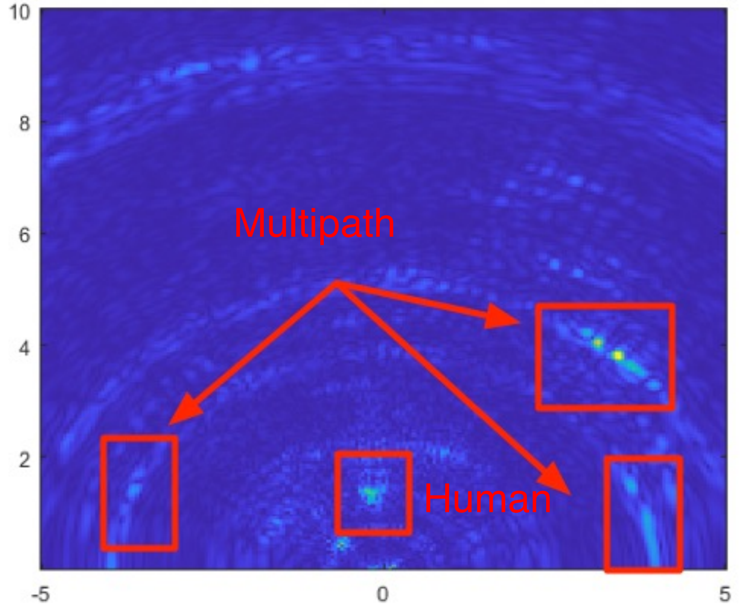
For simplicity, each pixel in RF image represents the signal reflected from this 2D grid.

RF-Based Human Pose Sensing Dataset

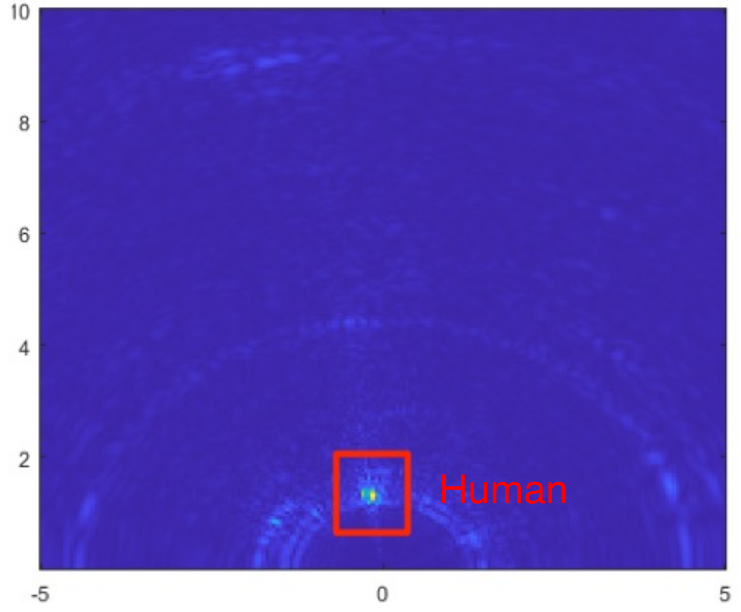
Signal Processing

multipath interface
static background





Before subtraction



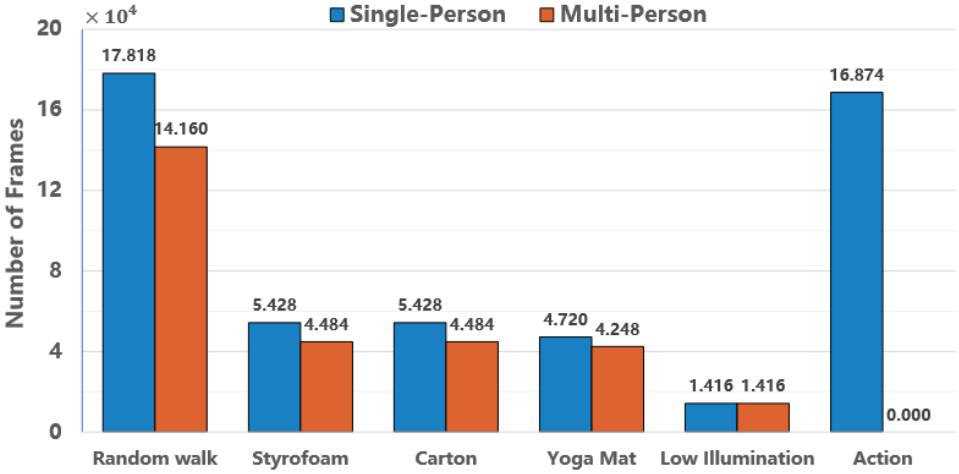
After subtraction

We utilize frame differencing to suppress the interface

RF-Based Human Pose Sensing Dataset

Data statistics

- Single or multiple persons
- Ten different environments
- Poses: standing, walking, squatting, sitting
- Challenge scenarios: obstruction and darkness



Statistics of the number of frames in each category

Ten Collection Environments

Four Occlusion Scenarios

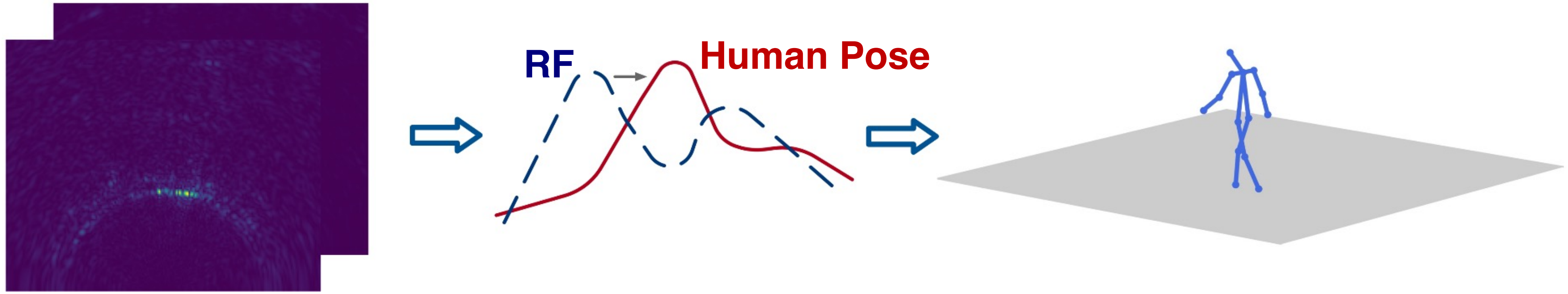
中国科学技术大学
University of Science and Technology of China

RF-Based Human Pose Estimation with Optimal Transport

Pose Estimation with Optimal Transport

Cross-domain Pose Alignment and Estimation

RF Signals | **Feature Space** | **3D Human Pose**



Challenges

- Measure the difference between RF and human pose domains
- Represent human poses in the feature space

Advantage

- Simplifying mapping complexity
- Improving interpretability
- Enabling cross-domain modeling

Yu et al. RFPose-OT. ***FITEE*** 2023.

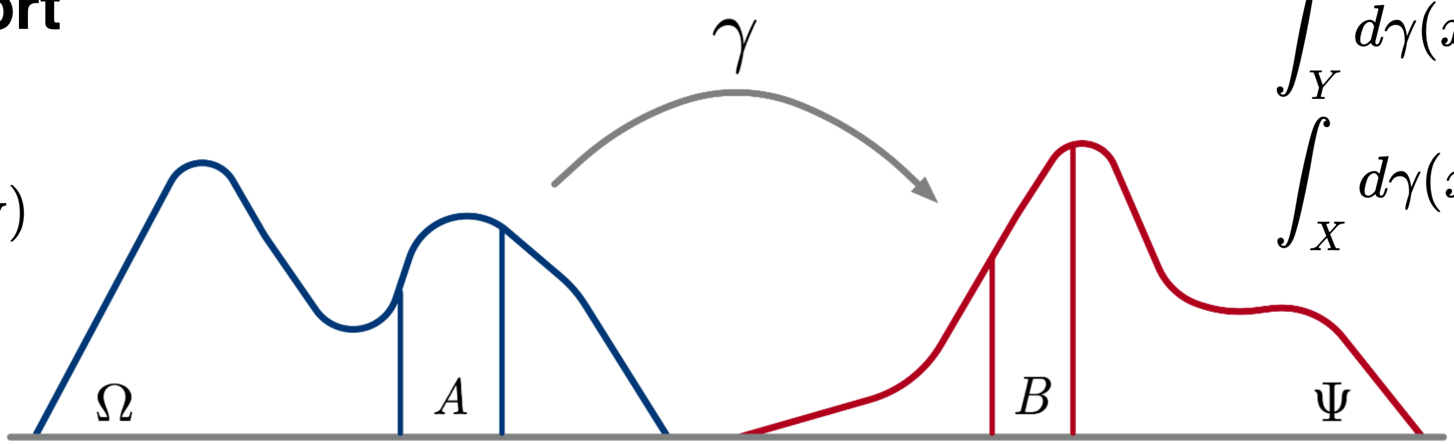
Pose Estimation with Optimal Transport

University of Science and Technology of China

Preliminary

- **Optimal Transport**

Distributions $\mathrm{A(x), B(y)}$
Transport plan $\gamma(x, y)$



$$\int_Y d\gamma(x, y) = A(x),$$
$$\int_X d\gamma(x, y) = B(y).$$

- **Minimum transport cost as metric**

$$\min_{\gamma \in \Pi(\mu, \nu)} \int_{X \times Y} C(x, y) d\gamma(x, y)$$

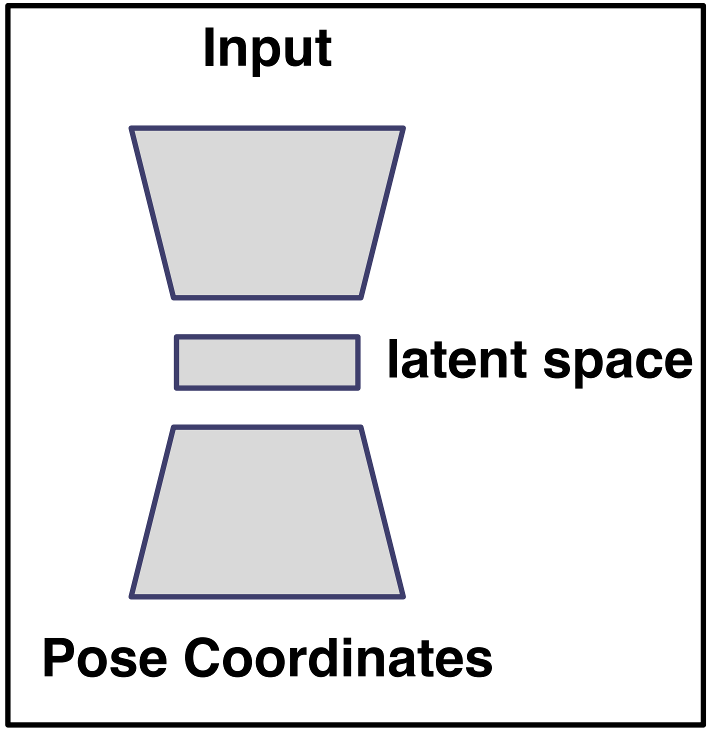
cost function   joint distribution over $X \times Y$

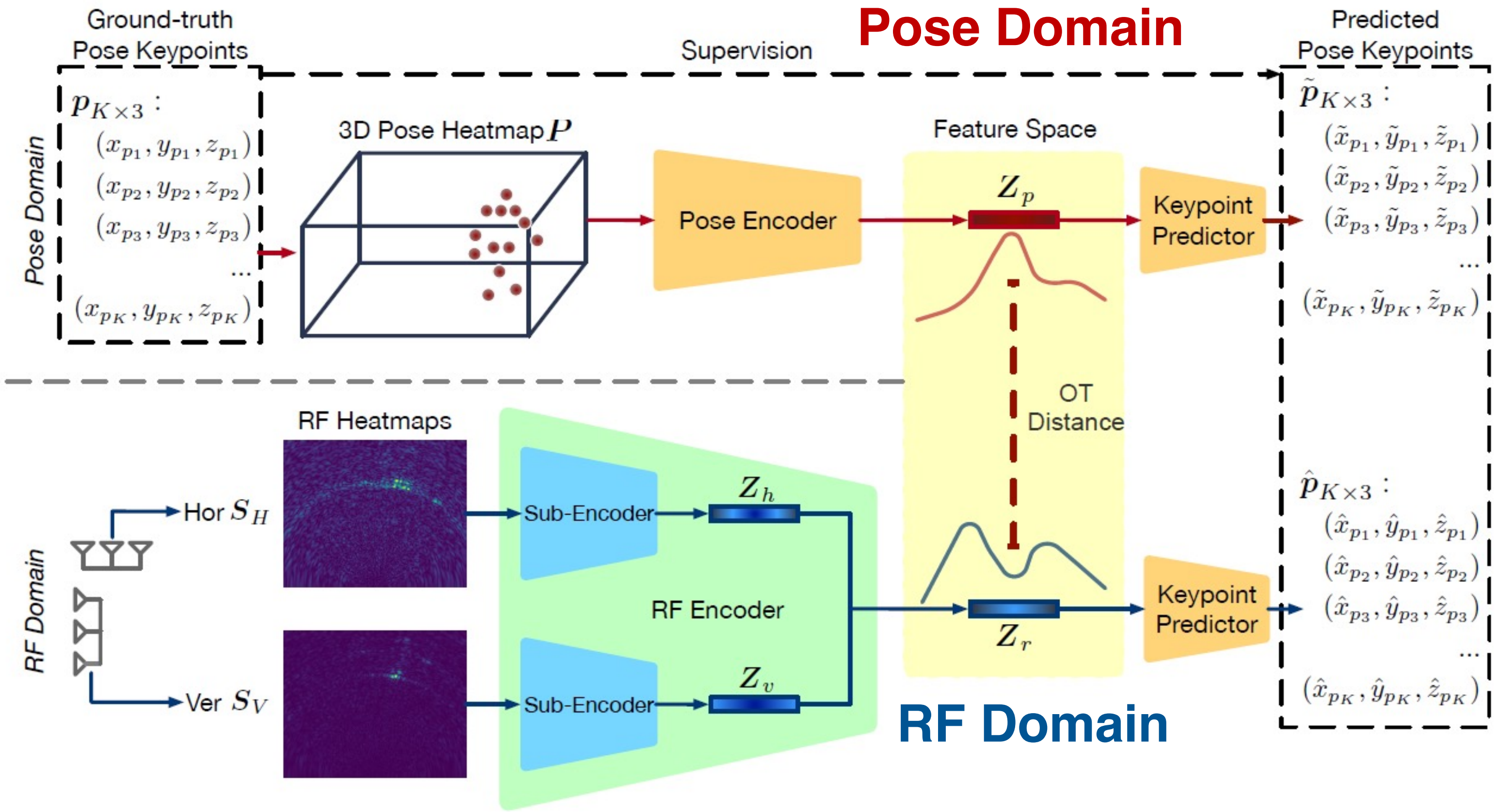
- The minimum transport cost is defined as the optimal transport distance
- An optimal transport distance equal to zero means that the two distributions are the same

Yu et al. RFPose-OT. ***FITEE*** 2023.

Pose Estimation with Optimal Transport

Model Architecture

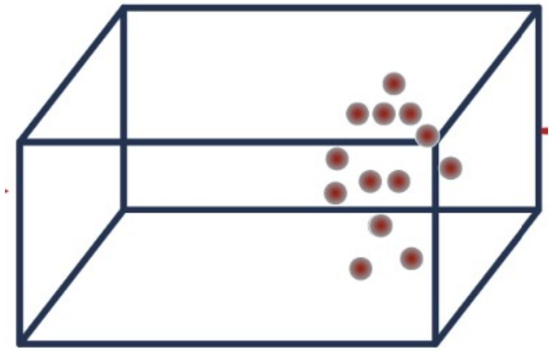




Yu et al. RFPose-OT. ***FITEE*** 2023.

Pose Estimation with Optimal Transport

Pose Domain



1. Construct the Pose Heatmap

Human Keypoint

$$\boldsymbol{P}_k(x,y,z) = \exp\left[-\frac{(x-x_{p_k})^2 + (y-y_{p_k})^2 + (z-z_{p_k})^2}{2\sigma^2}\right] \quad \boldsymbol{P}(x,y,z) = \sum_{k=1}^{K} \boldsymbol{P}_k(x,y,z)$$

2. Loss Functions

Absolute Error $\mathcal{L}_P = \left\|\tilde{\boldsymbol{p}}_{K\times 3} - \boldsymbol{p}_{K\times 3}\right\|_2$

Pose-only Error $\mathcal{L}_{PO} = \left\|\left(\tilde{\boldsymbol{p}}_{K\times 3} - \frac{1}{K}\sum_k^K \tilde{p}_k\right) - \left(\boldsymbol{p}_{K\times 3} - \frac{1}{K}\sum_k^K p_k\right)\right\|_2$

Yu et al. RFPose-OT. ***FITEE*** 2023.

Pose Estimation with Optimal Transport

RF Domain

1. Learning RF representation with minimal distance to pose space

$$\mathcal{L}_{\mathrm{OT}} = \int_{\boldsymbol{Z}_{\mathrm{r}} \times Z_{\mathrm{p}}} C\left(\boldsymbol{Z}_{\mathrm{r}}, \boldsymbol{Z}_{\mathrm{p}}\right) \mathrm{d}\gamma\left(\boldsymbol{Z}_{\mathrm{r}}, \boldsymbol{Z}_{\mathrm{p}}\right),$$

2. Fine-tuning pose estimator based on RF domain

Absolute Error

$$\mathcal{L}_P = \left\| \tilde{\boldsymbol{p}}_{K\times 3} - \boldsymbol{p}_{K\times 3} \right\|_2$$

Pose-only Error

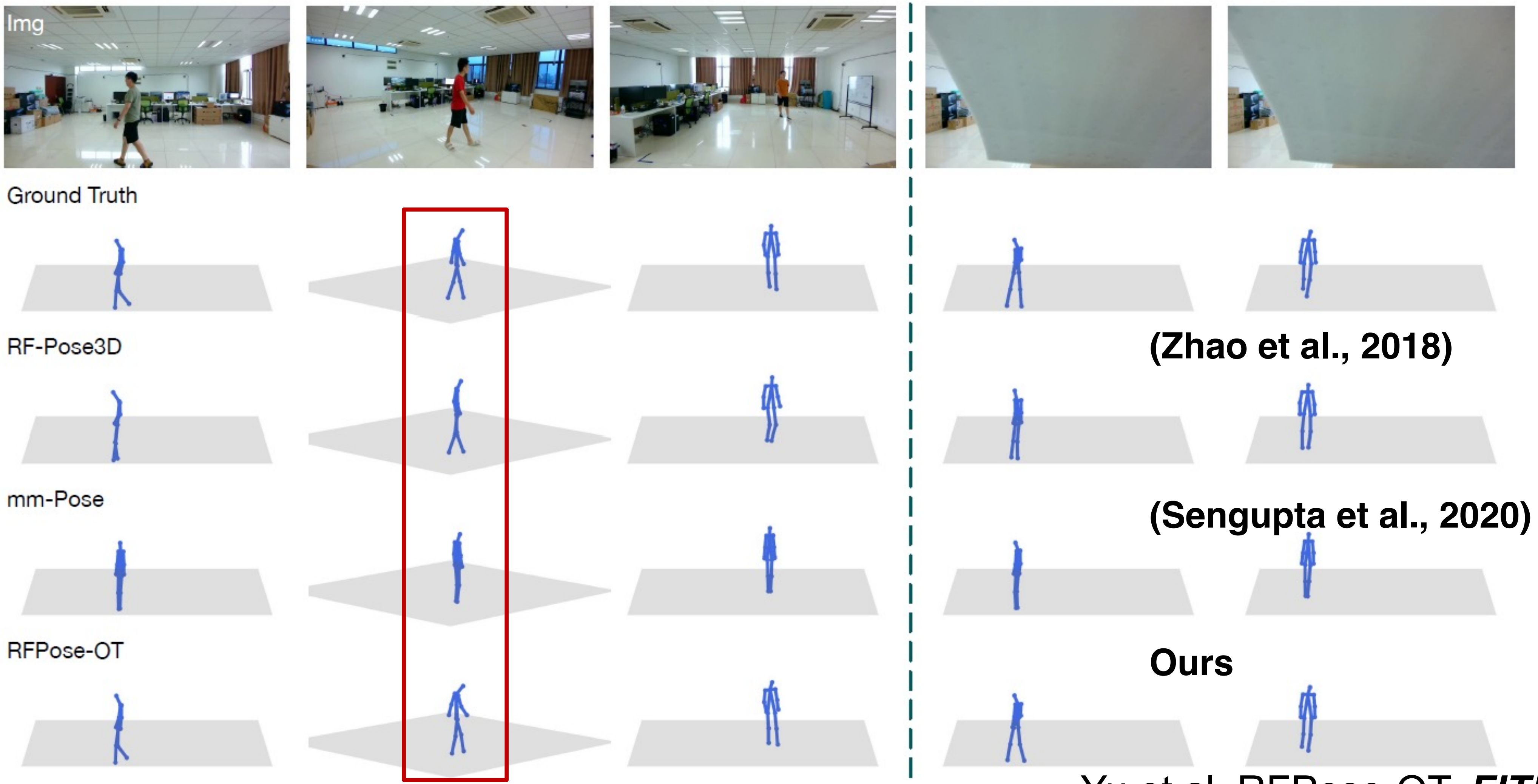
$$\mathcal{L}_{PO} = \left\| \left(\tilde{\boldsymbol{p}}_{K\times 3} - \frac{1}{K} \sum_{k}^{K} \tilde{p}_k \right) - \left(\boldsymbol{p}_{K\times 3} - \frac{1}{K} \sum_{k}^{K} p_k \right) \right\|_2$$

Yu et al. RFPose-OT. ***FITEE*** 2023.

Pose Estimation with Optimal Transport

University of Science and Technology of China

Experiments



Yu et al. RFPose-OT. ***FITEE*** 2023.

Pose Estimation with Optimal Transport

Experiments

Envs	Methods	Nose	Neck	Shoulders	Elbows	Wrists	Hips	Knees	Ankles	Overall
(a)	RF-Pose3D (Zhao et al., 2018b)	8.11	**5.21**	7.57	9.92	15.74	6.64	11.31	21.10	11.27
	mm-Pose (Sengupta et al., 2020)	8.19	5.30	7.23	9.67	15.29	**6.20**	10.83	19.04	10.72
	RFPose-OT	**7.90**	6.14	**6.76**	**7.99**	**11.67**	6.39	**8.34**	**12.60**	**8.68**
(b)	RF-Pose3D (Zhao et al., 2018b)	**6.53**	4.86	6.65	8.75	14.05	6.95	11.26	21.52	10.70
	mm-Pose (Sengupta et al., 2020)	6.64	**3.88**	**6.34**	9.16	14.84	6.98	11.28	19.28	10.45
	RFPose-OT	7.85	6.42	6.78	**7.90**	**11.41**	**6.82**	**9.35**	**14.05**	**9.07**

Quantitative comparison experiment (keypoint estimation error):
(a) basic scene & (b) occluded scene, Unit: cm

Yu et al. RFPose-OT. ***FITEE*** 2023.

Pose Estimation with Optimal Transport

Ablation Study

Envs	Models	Nose	Neck	Shoulders	Elbows	Wrists	Hips	Knees	Ankles	Overall
(a)	RFPose	8.69	6.82	7.58	8.97	12.93	7.19	9.31	14.11	9.69
	RFPose-L2	8.02	**5.93**	6.88	8.12	12.02	6.36	8.57	12.96	8.84
	RFPose-OT w/o $\mathcal{L}_{PO+RO}$	8.32	6.04	7.01	8.52	12.66	**6.30**	8.57	13.94	9.17
	RFPose-OT (full)	**7.90**	6.14	**6.76**	**7.99**	**11.67**	6.39	**8.34**	**12.60**	**8.68**
(b)	RFPose	7.87	6.64	7.35	8.69	12.10	7.57	10.11	15.27	9.76
	RFPose-L2	7.88	6.59	7.23	8.62	12.47	7.33	10.08	15.20	9.74
	RFPose-OT w/o $\mathcal{L}_{PO+RO}$	**7.83**	**6.11**	7.01	8.36	12.04	**6.61**	9.51	15.55	9.44
	RFPose-OT (full)	7.85	6.42	**6.78**	**7.90**	**11.41**	6.82	**9.35**	**14.05**	**9.07**

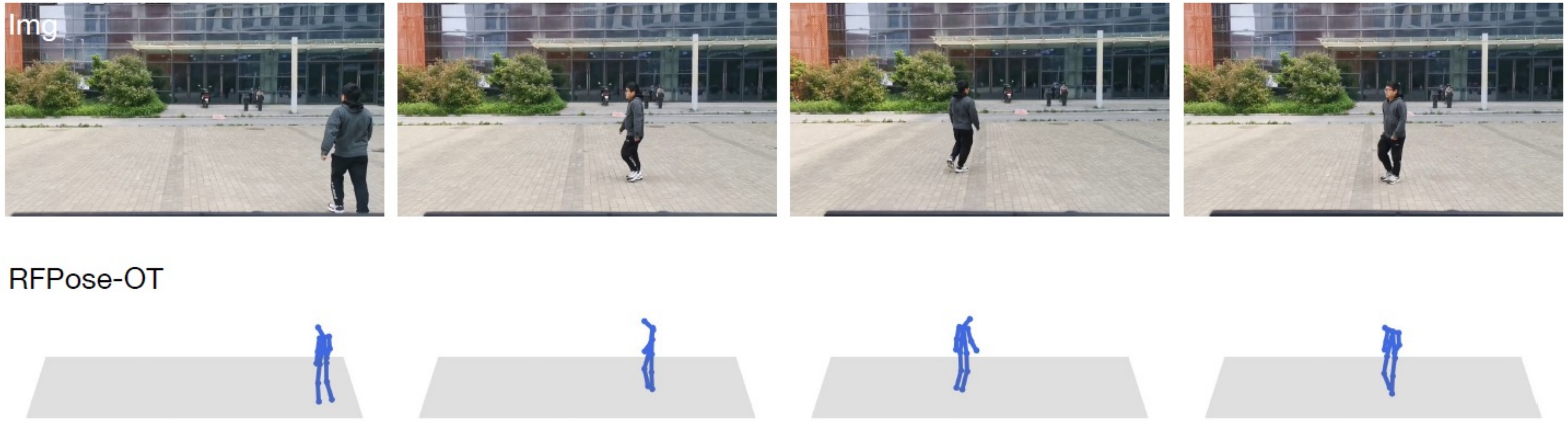
Ablation experiment (keypoint estimation error):
(a) basic scene & (b) occluded scene, Unit: cm

Yu et al. RFPose-OT. ***FITEE*** 2023.

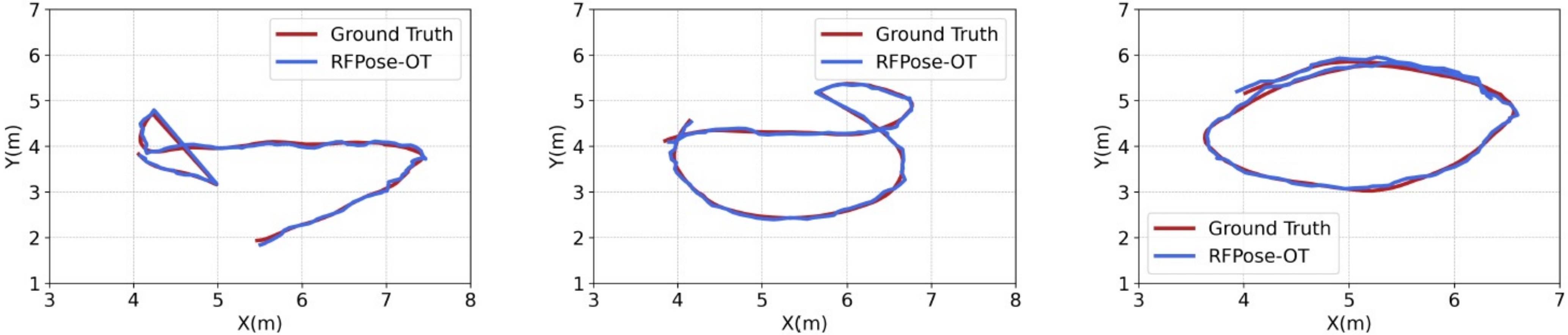
Pose Estimation with Optimal Transport

Unseen Scenario

Outdoor



Trajectory Tracking



Yu et al. RFPose-OT. ***FITEE*** 2023.

Pose Estimation with Optimal Transport

University of Science and Technology of China

Summary

1. Optimal transport theory to align RF and pose domains, ensuring accurate feature matching.
2. Demonstrated strong generalization across diverse scenarios:
   - Indoor basic and occlusion scenes
   - Outdoor environments

Outcome: Achieved accurate and versatile human pose estimation across challenging settings.

Yu et al. RFPose-OT. ***FITEE*** 2023.

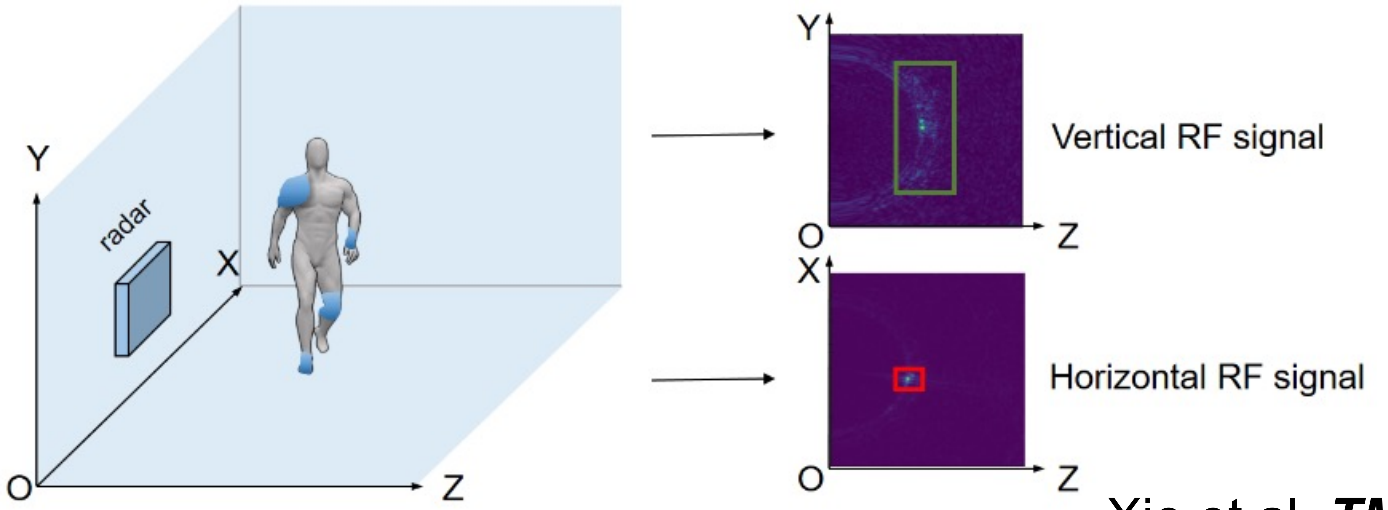
RF-Based Human Pose Estimation with Spatio-Temporal Attention

Single-Person Pose Estimation

中国科学技术大学
University of Science and Technology of China

Challenges

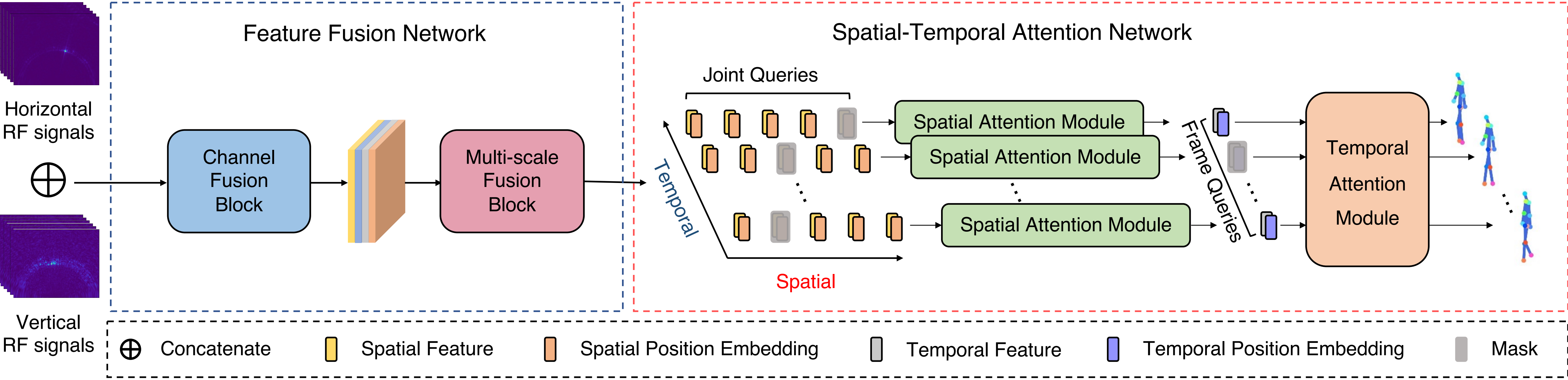
1. **Sparse and Incomplete RF Signals**: RF signals are often sparse and incomplete.
2. **Feature Fusion Across Dimensions**: RF signals from horizontal and vertical planes have distinct characteristics.



Xie et al. *TMM* 2023 & *TCSVT* 2023

Single-Person Pose Estimation

RPM Framework



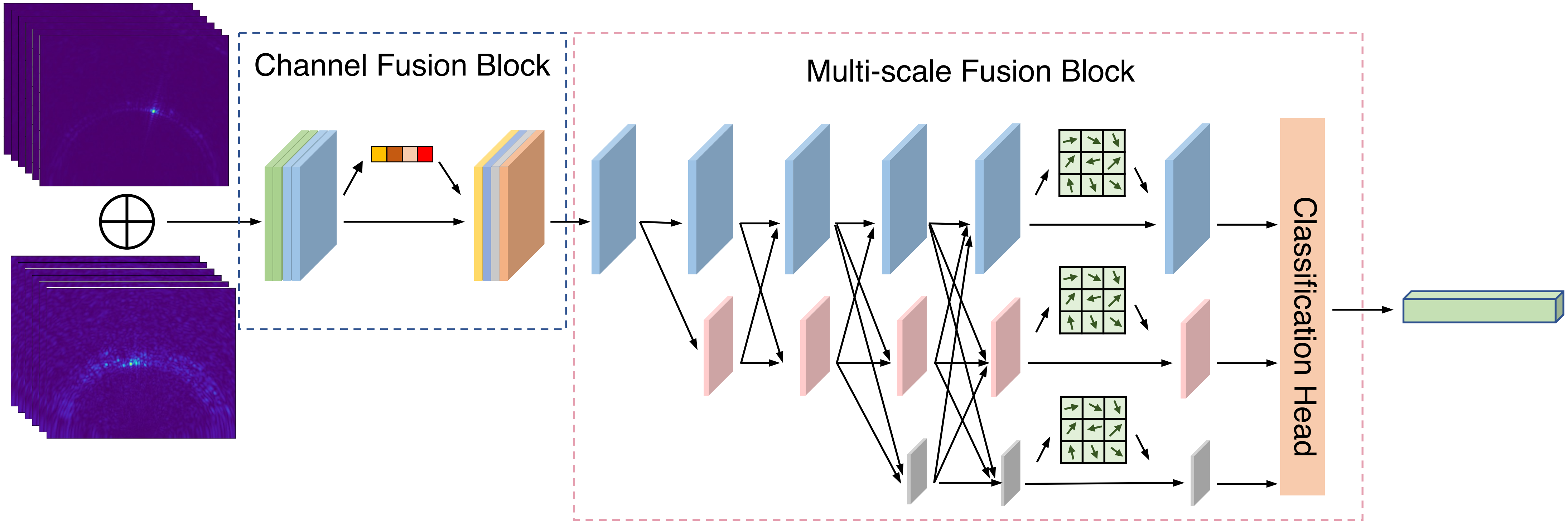
- Feature Fusion Network (FFN): Combines horizontal and vertical RF features.
- Spatial-Temporal Attention Network (STAN):
  - *Spatial Attention Module (SAM)*: Recovers missing body parts.
  - *Temporal Attention Module (TAM)*: Refines 3D skeleton sequences

Xie et al. ***TMM*** 2023 & ***TCSVT*** 2023

Single-Person Pose Estimation

中国科学技术大学 University of Science and Technology of China

Feature Fusion Network: *Combines horizontal and vertical RF features*



- Channel Fusion Block: Group convolution, bottleneck blocks.
- Multi-Scale Fusion Block: Deformable convolutions, scale/shape adaptation.
- Lightweight MLP: 2048-dimensional feature vector.

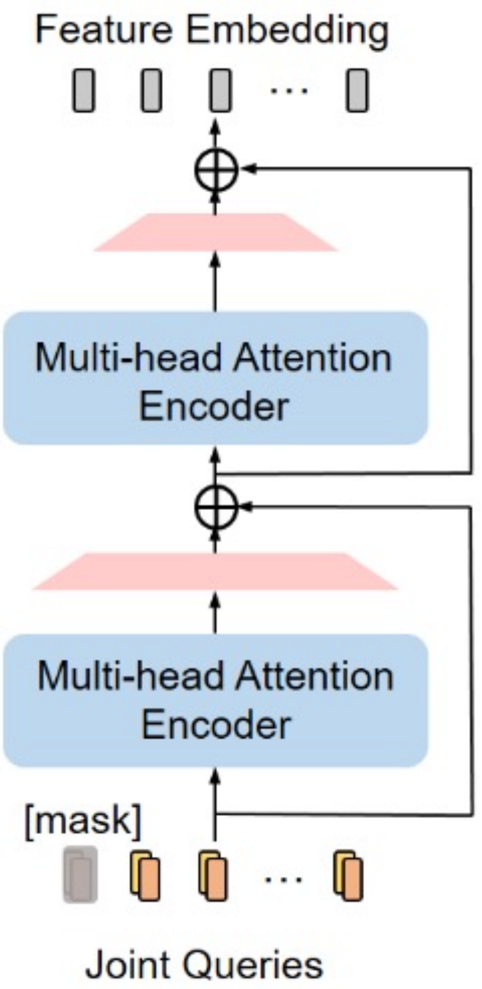
Xie et al. ***TMM*** 2023 & ***TCSVT*** 2023

Single-Person Pose Estimation

Spatial Attention Module: *Recovers missing body parts*

1. Masked Joint Modeling (MJM): Simulates missing body parts by masking random joint queries, encouraging the model to infer missing information.

2. Multi-Head Self-Attention: Captures non-local joint relationships by modeling dependencies across all joints.



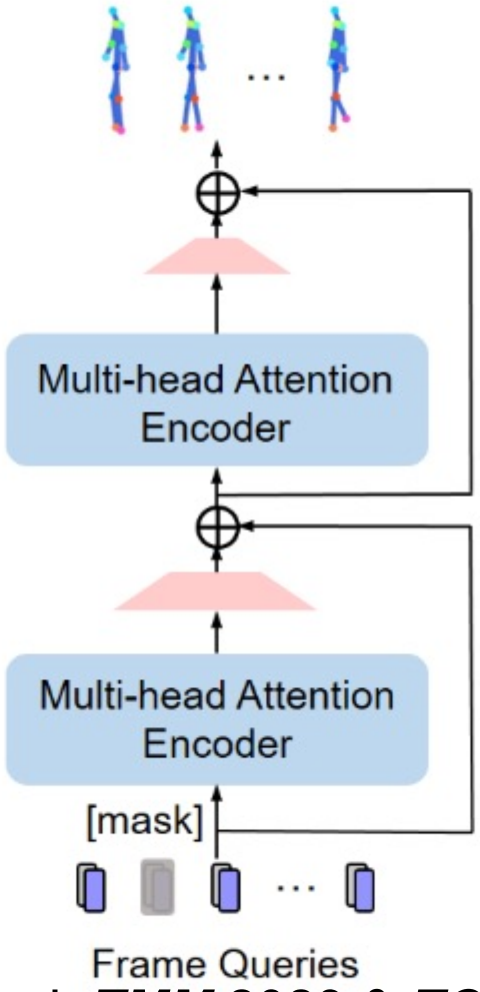
Xie et al. ***TMM*** 2023 & ***TCSVT*** 2023

Single-Person Pose Estimation

Temporal Attention Module (TAM): Refines 3D skeleton sequences

1. Masked Frame Modeling (MFM): Masks input queries from random frames to simulate missing temporal information.

2. Multi-Head Self-Attention: Captures temporal dependencies across frames.



Xie et al. ***TMM*** 2023 & ***TCSVT*** 2023

Single-Person Pose Estimation

Loss Function

- Location Loss

$$\mathcal{L}_{loc} = \frac{1}{\boxed{F}} \sum_{i=1}^{F} \left\| \boxed{P_i^{root}} - \boxed{\hat{P}_i^{root}} \right\|_2$$

sequence length — body center prediction — body center label

- Pose Loss

$$\mathcal{L}_{pose} = \frac{1}{FJ} \sum_{i=1}^{F} \sum_{k=1}^{J} \left\| (\boxed{P_i^k} - P_i^{root}) - (\boxed{\hat{P}_i^k} - \hat{P}_i^{root}) \right\|_2$$

3D pose prediction — 3D pose label

- Overall Loss

$$\mathcal{L} = \mathcal{L}_{loc} + \mathcal{L}_{pose}$$

Xie et al. ***TMM*** 2023 & ***TCSVT*** 2023

Single-Person Pose Estimation

- ## Comparison Test

Quantitative comparison experiments of different methods (unit: mm)

Method	Params (M)	MACs (G)	Nose	Neck	Sho	Elb	Wri	Hip	Knee	Ank	MPJPE (↓)
RFPose3D [103]	10.92	25.33	81.4	52.0	90.1	120.4	135.1	89.9	144.9	167.4	116.3
mm-Pose [107]	33.08	0.25	165.8	67.3	203.3	233.8	261.6	138.6	162.3	169.9	183.7
RPM	81.67	1375	**57.5**	**37.2**	**49.1**	**64.9**	**68.2**	**46.5**	**58.1**	**65.1**	**57.1**

MPJPE： measures the Euclidean distances between the ground truth joints and the predicted joints

Params： measures the number of all trainable parameters in the model

MACs： measures the amount of all multiply-accumulate operations in the model

RPM leads baseline methods significantly in pose estimation accuracy

Xie et al. ***TMM*** 2023 & ***TCSVT*** 2023

Single-Person Pose Estimation

- **Ablation Study**

Ablation experiments for Feature Fusion Network (unit: mm)

Backbone	Multi-scale Modeling	Deformable Conv	MPJPE (↓)
FFN	-	-	63.2
	✓	-	57.6
	✓	✓	**57.1**

Ablation experiments of Spatio-Temporal Attention Network (unit: mm)

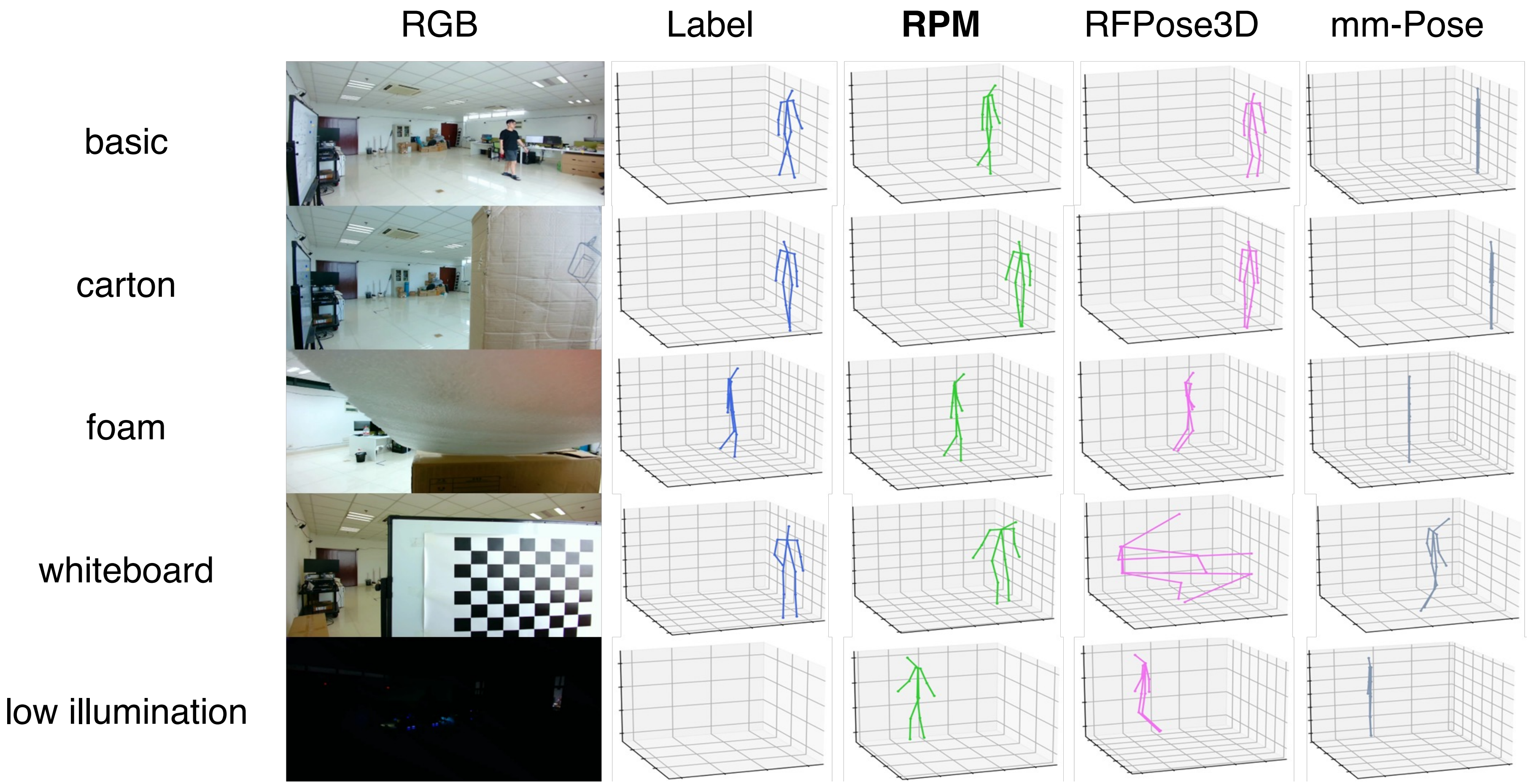
Method	SAM	TAM	MPJPE (↓)
RPM	-	-	99.2
	✓	-	75.5
	✓	✓	**57.1**

Xie et al. ***TMM*** 2023 & ***TCSVT*** 2023

Single-Person Pose Estimation

中国科学技术大学
University of Science and Technology of China

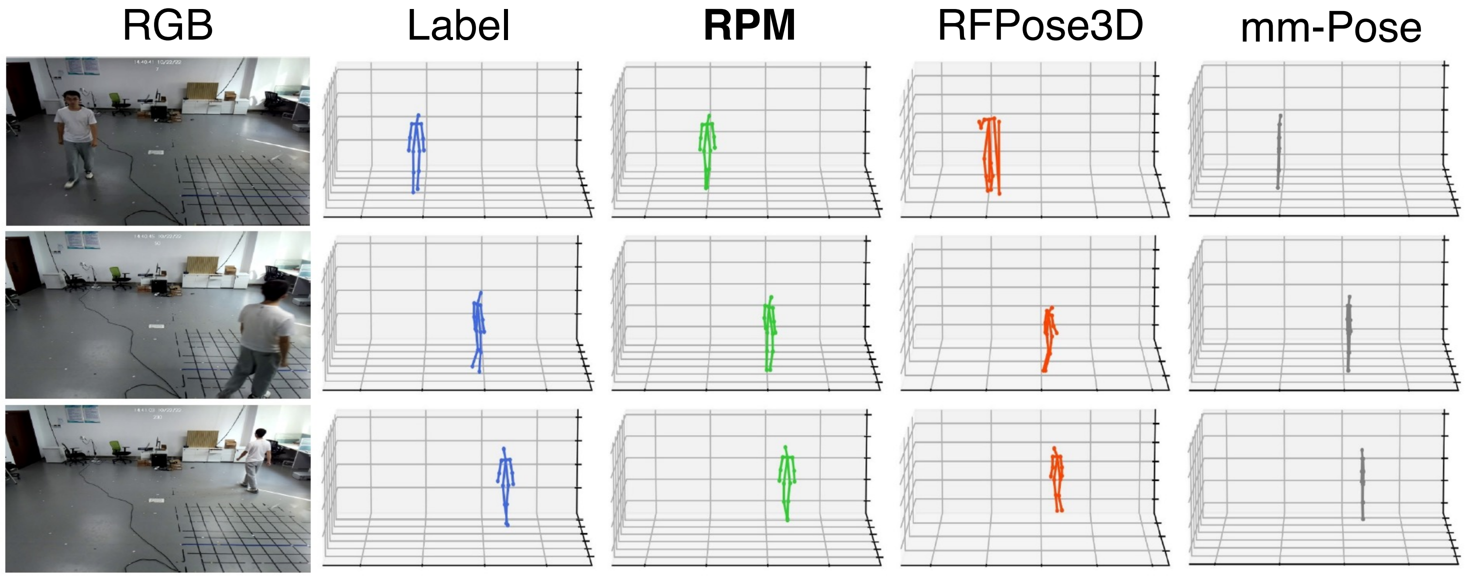
- **Visualization Results in Basic Scenarios**



Xie et al. ***TMM*** 2023 & ***TCSVT*** 2023

Single-Person Pose Estimation

- **Visualization Results in New Scenario**



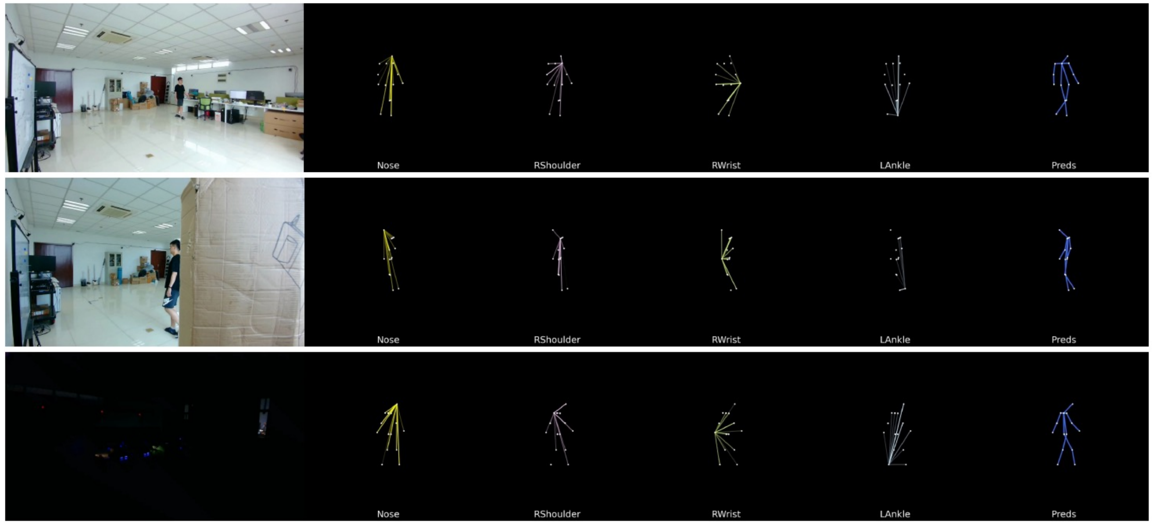
Comparison of pose estimation performance under new scenarios (unit: mm)

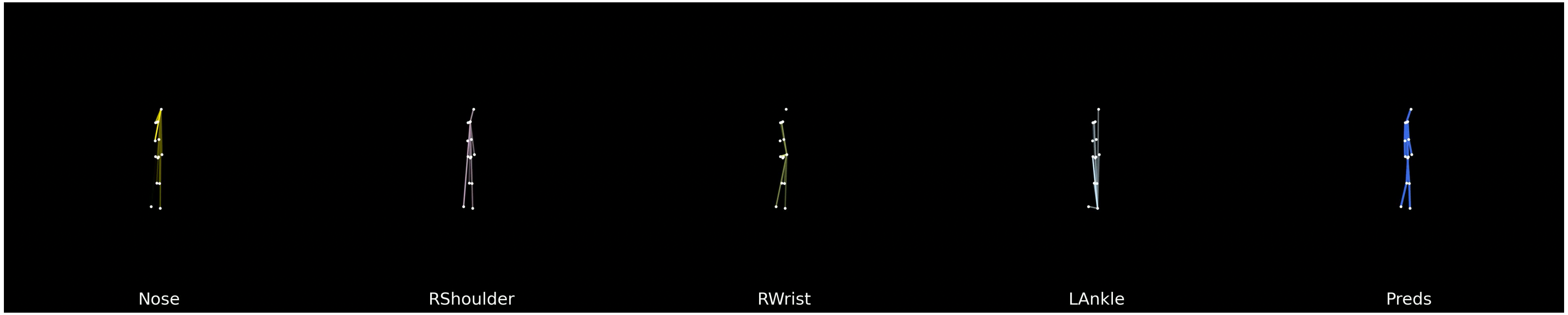
Method	Mean (↓)
RFPose3D [103]	146.5
mm-Pose [107]	205.2
RPM	**78.5**

Xie et al. ***TMM*** 2023 & ***TCSVT*** 2023

Single-Person Pose Estimation

- **Spatial Attention Visualization**

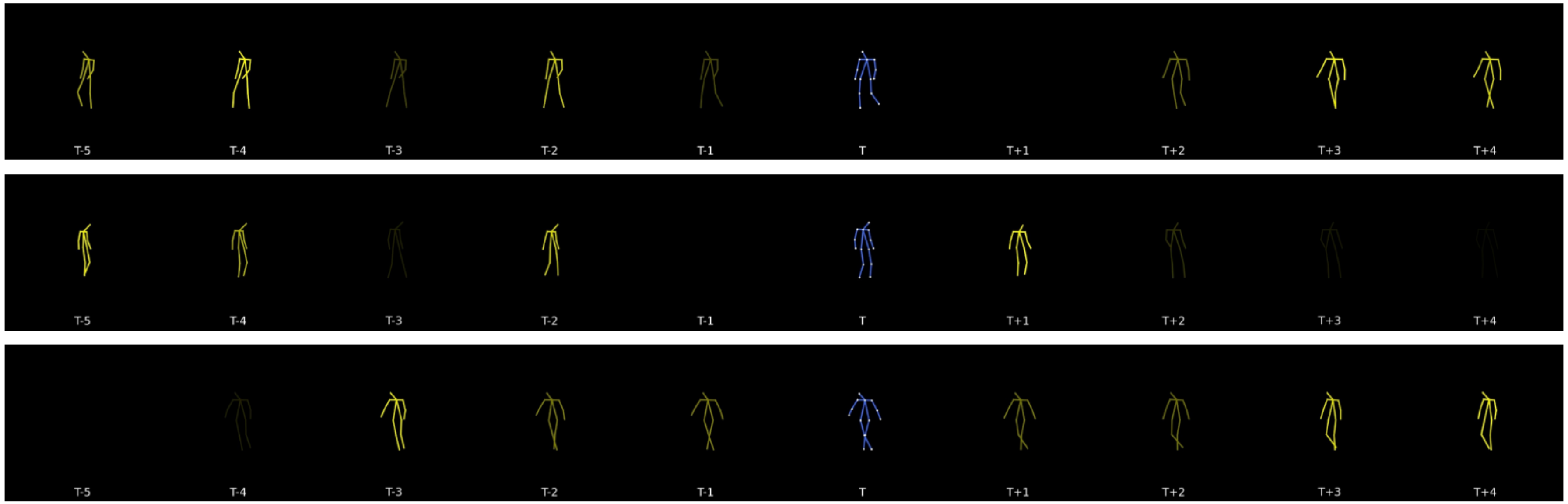


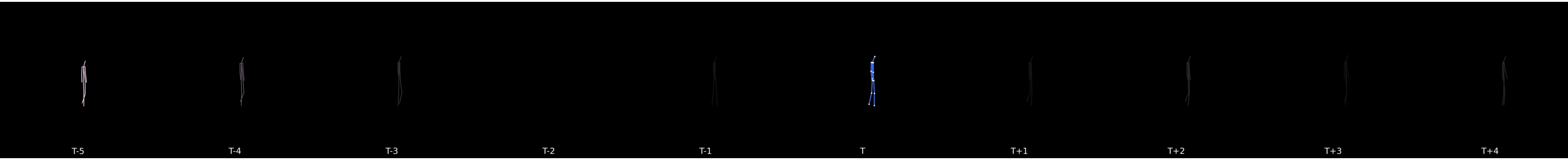


RPM can adaptively adjust the spatial attention of **keypoints**

Single-Person Pose Estimation

- **Temporal Attention Visualization**





RPM can refine human pose sequences based on **temporal attention**

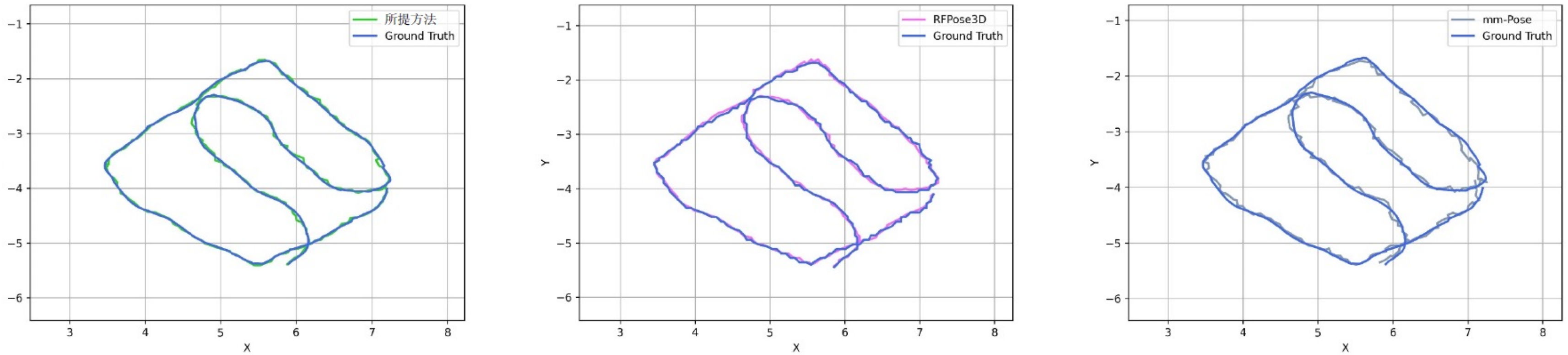
Single-Person Pose Estimation

- **Indoor Localization**

Indoor localization performance comparison (unit: cm)

Method	X	Y	Z	Mean (↓)
RFPose3D [103]	2.6	2.8	1.6	5.0
mm-Pose [107]	3.8	3.4	1.9	6.3
RPM	**2.3**	**1.8**	**1.1**	**3.6**

Visualization of indoor localization trajectories



RPM also performs well on indoor human localization tasks

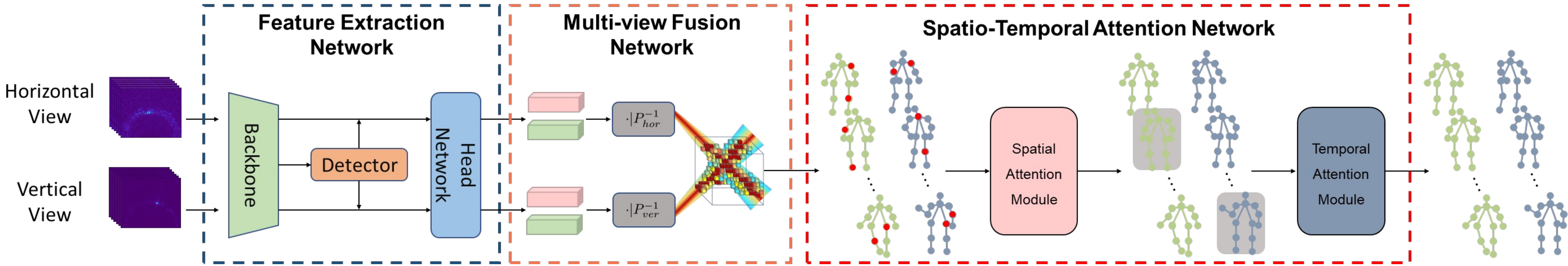
Single-Person Pose Estimation

- Attention mechanism for robust pose estimation.
- Multi-scale feature fusion using channel attention and deformable convolution.
- Fine-grained human perception in regular, occluded, and dark scenes.
- Accurate indoor human localization.

Xie et al. ***TMM*** 2023 & ***TCSVT*** 2023

Multiple-Person Pose Estimation

- **RPM2.0 Framework**

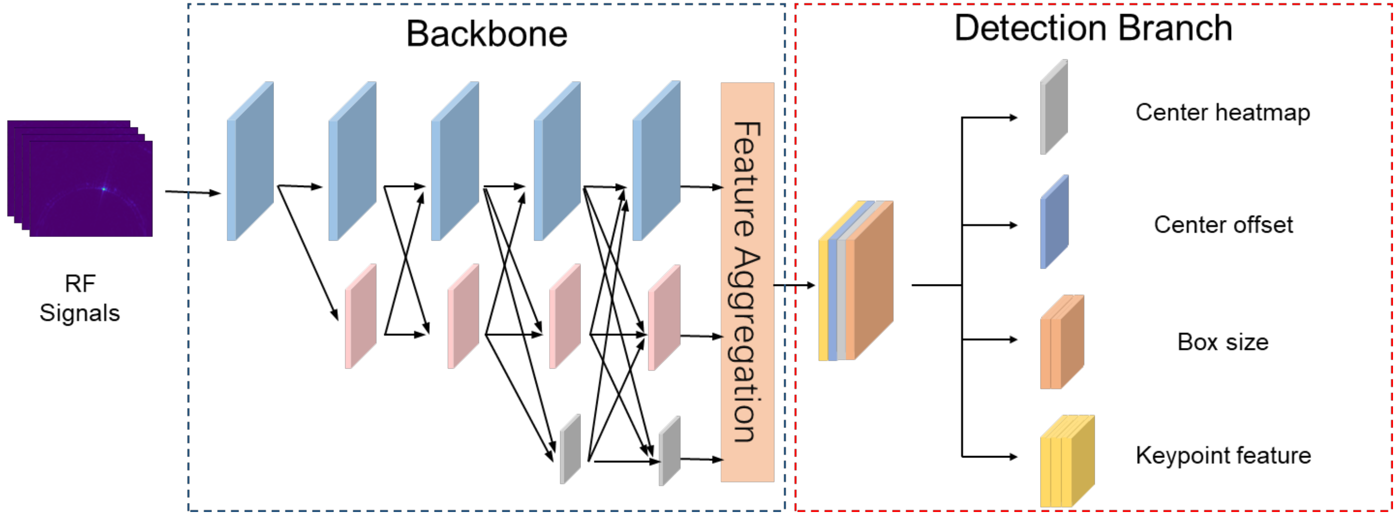


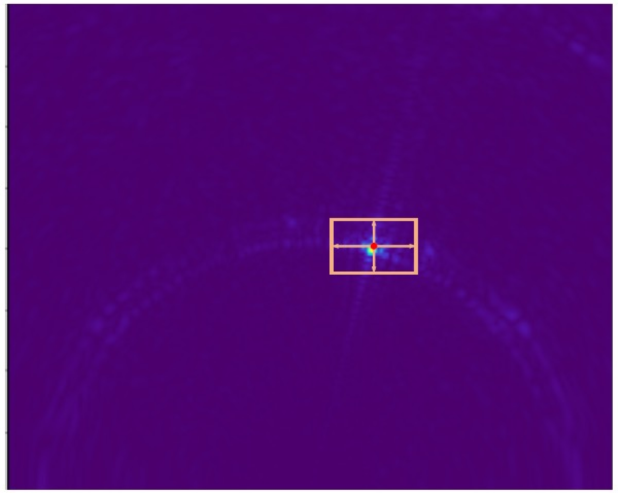
- Feature Extraction Network: extracts features of multi-person separately
- Multi-view Fusion Network: mapping multiple radar views to a uniform space
- Spatio-temporal Attention Network: modeling the correlation of multi-person

Xie et al. *TMM* 2023 & *TCSVT* 2023

Multiple-Person Pose Estimation

Feature Extraction Network: *extracts features of multi-person separately*





1. **HRNet-18 Backbone:** Ensures robust multi-scale feature extraction for small-scale targets.
2. **Anchor-Free Detection**: Simplifies person detection using center points and offsets.
3. **Heatmaps:** Generates center and keypoint heatmaps for precise 3D pose estimation.

Xie et al. ***TMM*** 2023 & ***TCSVT*** 2023

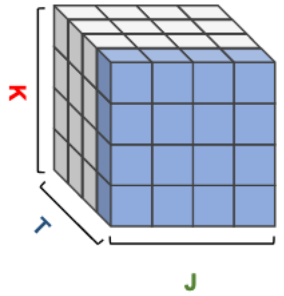
Multiple-Person Pose Estimation

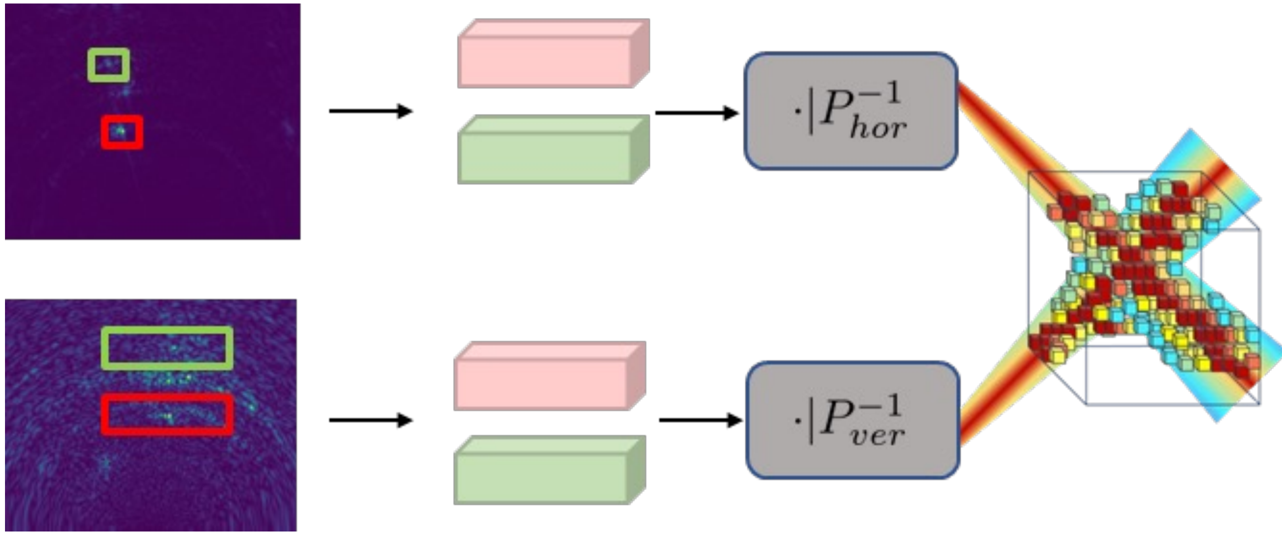
Multi-view Fusion Network: mapping multiple views to a uniform space

- Cropped Feature Extraction: bounding boxes are used to crop regions of interest to enhance each view

- Canonical 3D Space: The horizontal and vertical features (2D) are projected into a shared canonical space

- Fusion of Horizontal and Vertical Features

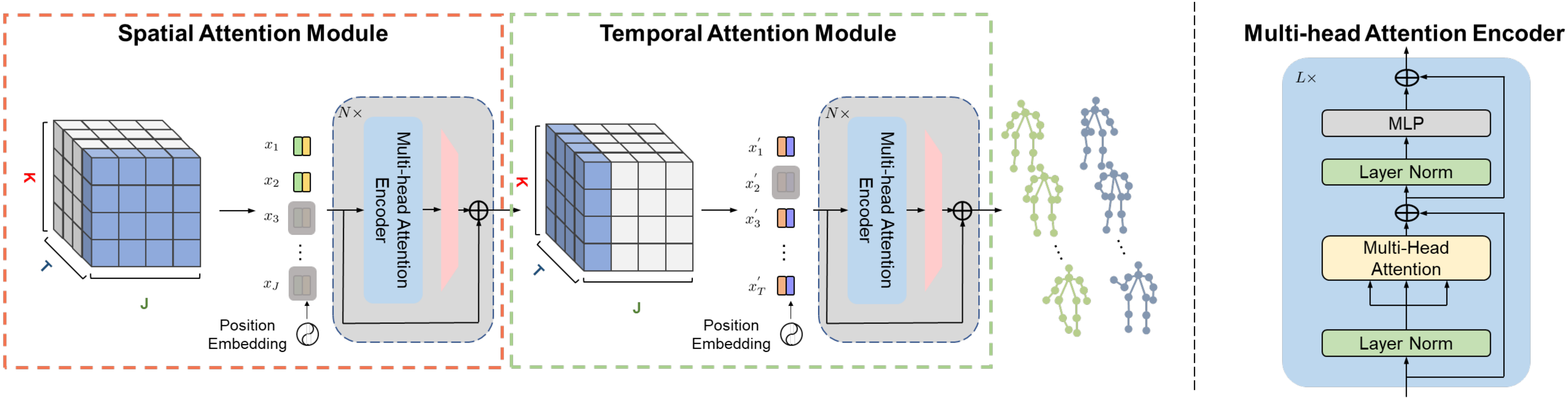




Xie et al. ***TMM*** 2023 & ***TCSVT*** 2023

Multiple-Person Pose Estimation

Spatio-temporal Attention Network: modeling the correlation of multi-person



Xie et al. ***TMM*** 2023 & ***TCSVT*** 2023

Multiple-Person Pose Estimation

- **Comparison Test**

Comparison of single-person pose estimation performance by different methods (unit: mm)

Method	Nose	Neck	Sho	Elb	Wri	Hip	Knee	Ank	Mean (↓)
Single-Person									
RFPose3D [103]	100.5	73.7	117.2	149.7	159.4	107.6	149.5	168.0	134.1
RFPose-OT [159]	86.9	69.8	84.7	108.5	111.0	84.6	110.4	122.4	100.0
RPM [158]	57.7	49.0	52.4	**64.6**	**65.8**	51.5	60.5	65.8	59.2
RPM2.0 (concat)	55.8	37.0	50.5	69.1	71.8	47.4	60.0	66.1	58.4
RPM2.0 (sum)	56.1	37.9	52.6	70.3	74.5	48.4	59.9	67.2	59.8
RPM2.0 (softsum)	**55.4**	**37.0**	**49.8**	67.2	69.4	**47.1**	**58.3**	**64.8**	**57.5**

better results than previous methods on single-person pose estimation task

Multiple-Person Pose Estimation

- **Comparison Test**

Comparison of the performance of different methods for multi-person pose estimation (unit: mm)

Method	Nose	Neck	Sho	Elb	Wri	Hip	Knee	Ank	Mean (↓)
Multi-Person									
RFPose3D [103]	114.0	84.3	133.0	162.7	175.5	117.7	155.7	172.5	145.2
RPM2.0 (concat)	71.9	53.3	**65.4**	**83.7**	86.3	62.0	74.0	81.6	74.6
RPM2.0 (sum)	73.1	53.3	69.6	88.5	90.5	63.8	76.6	83.4	76.5
RPM2.0 (softsum)	**69.2**	**50.3**	65.7	84.8	**85.7**	**61.0**	**73.5**	**80.4**	**73.0**

RPM2.0 achieves substantial performance gains over the baseline methods for multi-person pose estimation task

Xie et al. ***TMM*** 2023 & ***TCSVT*** 2023

Multiple-Person Pose Estimation

- **Comparison Test**

Comparison of Model Size and Computational Complexity

Method	Params (M)	MACs (G)	Inference Speed (ms)
RFPose-OT [159]	**7.02**	**1.30**	**29**
RFPose3D [103]	10.92	25.33	32
RPM [158]	81.67	1375	49
RPM2.0	38.41	184.93	43

RPM 2.0 is simpler and more efficient than previous RPM

Performance comparison of different detectors

Detector	Detection Speed (ms)	AP	MPJPE (mm)
RPN [160]	3.2	0.72	74.6
DETR [156] (2 stage)	15.6	**0.76**	75.2
Anchor-free detector (ours)	**0.4**	0.75	**73.0**

Anchor-free detector achieves a balance of speed and accuracy

Xie et al. ***TMM*** 2023 & ***TCSVT*** 2023

Multiple-Person Pose Estimation

- **Ablation Study**

Ablation experiments for RPM2.0 (unit: mm)

Backbone	MFN	SAM	TAM	MPJPE (↓)
HRNet	-	-	-	174.1
	-	✓	-	127.5
	-	-	✓	120.2
	-	✓	✓	95.7
	✓	-	-	155.0
	✓	✓	-	102.4
	✓	-	✓	98.6
	✓	✓	✓	**73.0**

Analysis on different signal input (unit: mm)

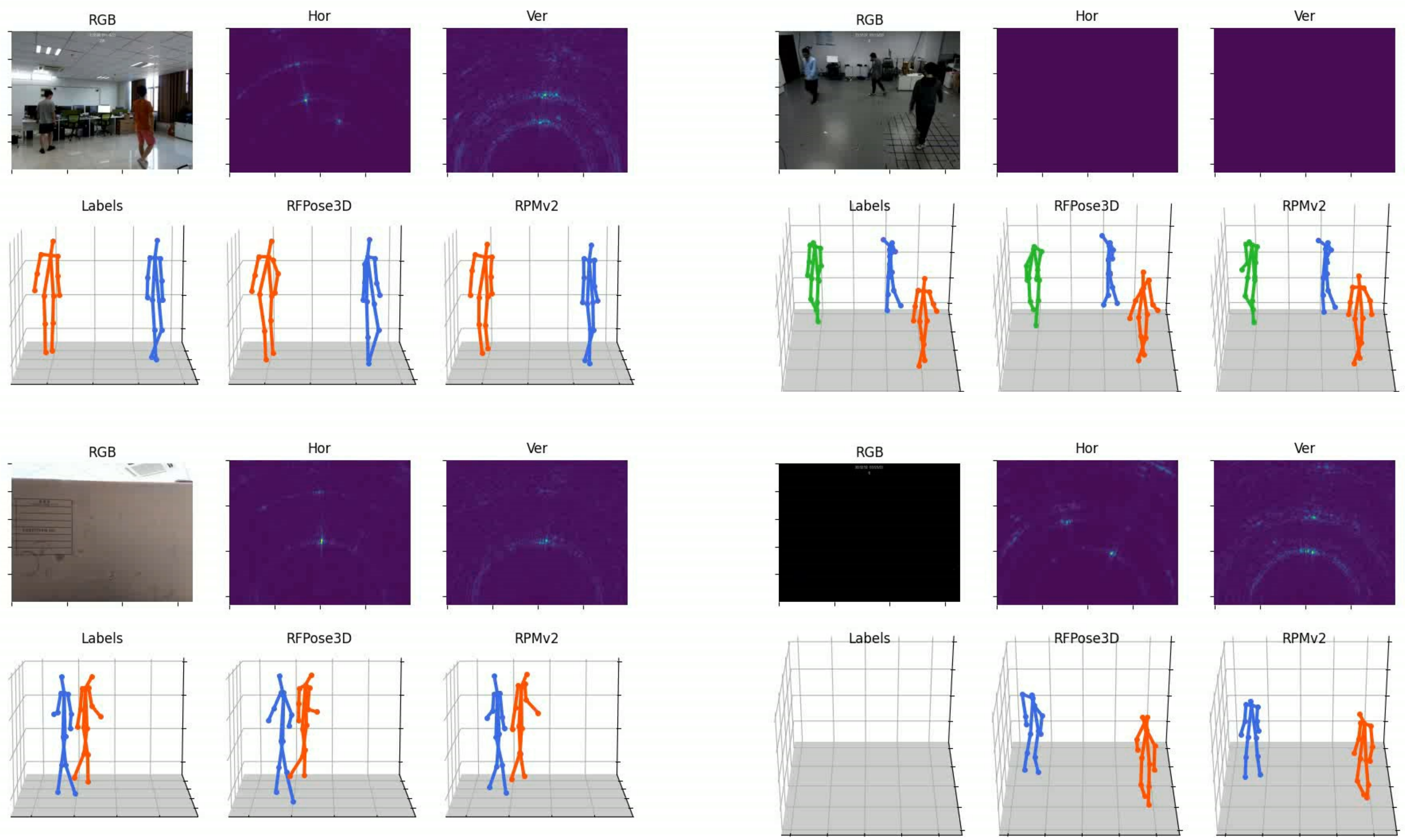
Method	Horizontal RF signal	Vertical RF signal	MPJPE (↓)
RPM2.0	✓	-	103.2
	-	✓	132.4
	✓	✓	**73.0**

RPM2.0 can fully fuse RF signals from multiple views for optimal performance

Xie et al. ***TMM*** 2023 & ***TCSVT*** 2023

Multiple-Person Pose Estimation

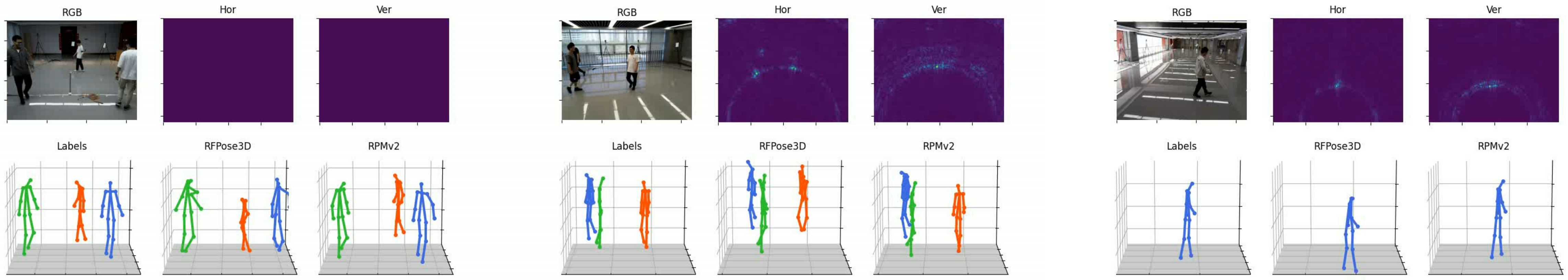
- **Visualization Results in Basic Scenarios**



Multiple-Person Pose Estimation

中国科学技术大学
University of Science and Technology of China

- **Visualization Results in New Scenarios**



Comparison of pose estimation performance in new scenarios (unit: mm)

Method	Mean (↓)
RFPose3D [50]	253.4
RPM 2.0	**116.2**

RPM 2.0 can accurately recover multi-person poses on new data

Multiple-Person Pose Estimation

Summary

1. **Lightweight Anchor-Free Detector**: It efficiently identifies multiple targets with reduced computational complexity.

2. **Multi-View Feature Fusion Network:** This network integrates radar views based on spatial relationships, ensuring robust and unified 3D pose representation.

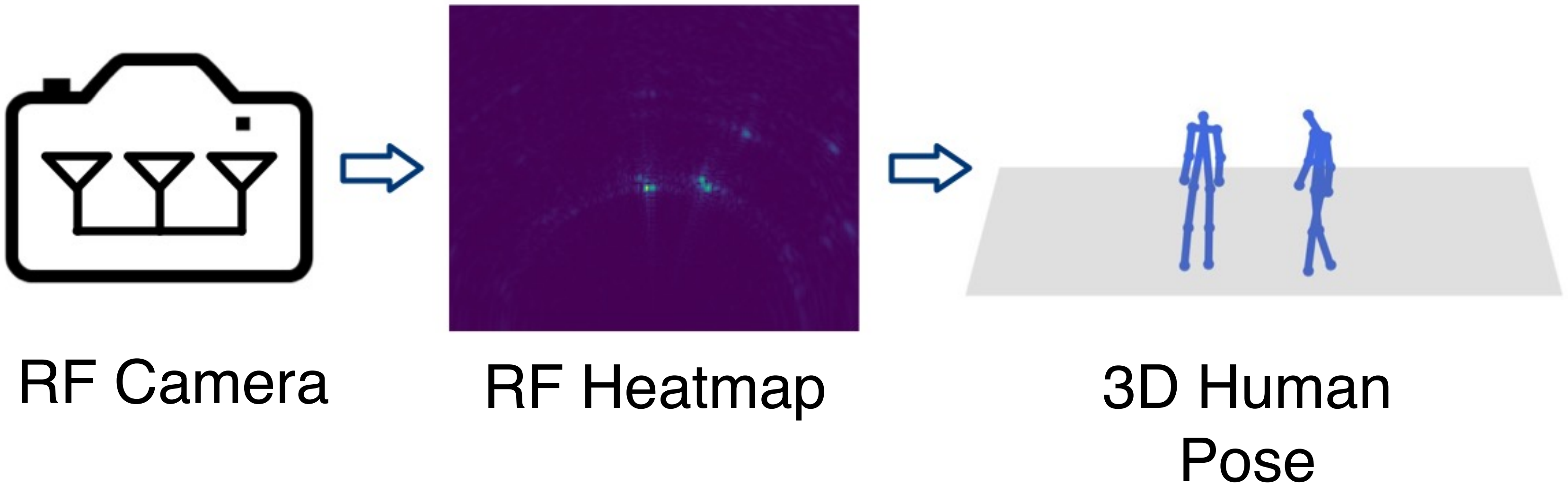
Xie et al. ***TMM*** 2023 & ***TCSVT*** 2023

中国科学技术大学
University of Science and Technology of China

Lightweight RF-Based Human Pose Estimation for Mobile Devices

Lightweight Pose Estimation for Mobile Devices

- **Problem**



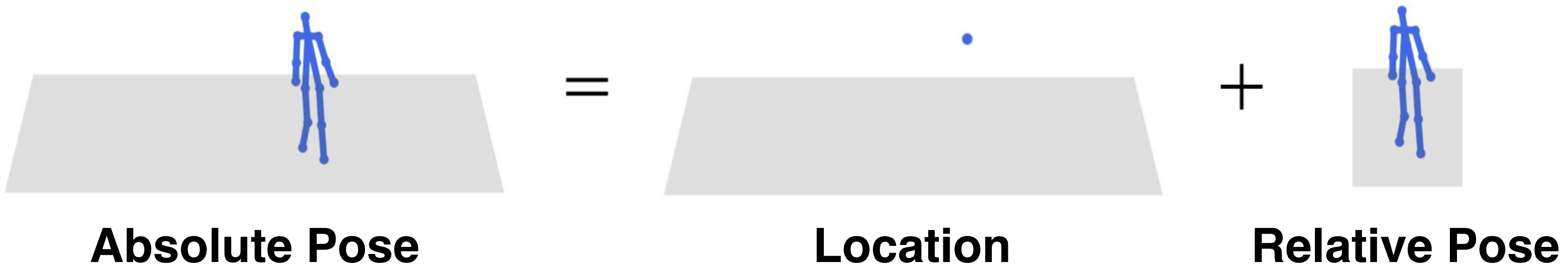
- **How to estimate human poses using RF signals in real time?**

  - Large amount of input signal data
  - Heavyweight model

Yu et al. MobiRFPose ***TMM***, 2024.

Lightweight Pose Estimation for Mobile Devices

- **Solution**



$$\hat{\boldsymbol{p}}_o = \mathrm{HLN}(\boldsymbol{S}) \qquad \hat{\boldsymbol{p}}_p = \mathrm{PEN}(\tilde{\boldsymbol{S}})$$

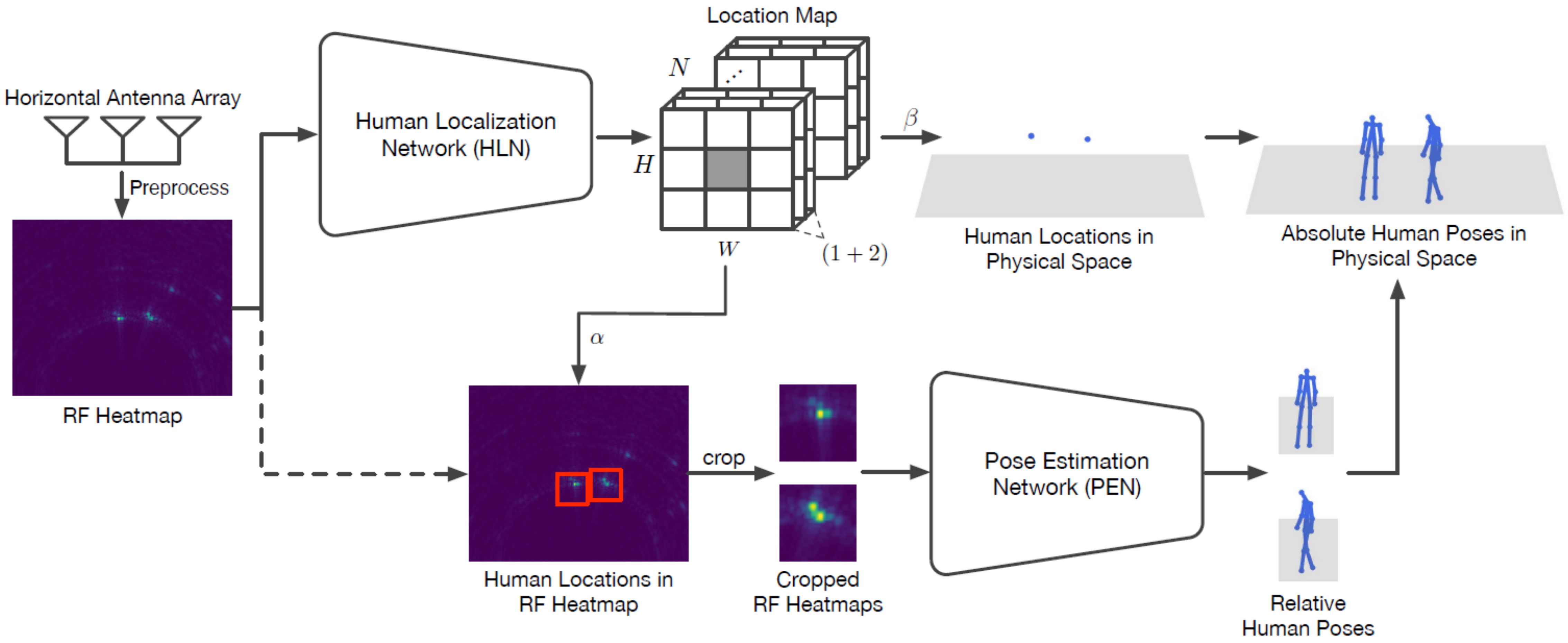
⇩

$$\hat{\boldsymbol{p}} = \hat{\boldsymbol{p}}_o + \hat{\boldsymbol{p}}_p$$

Yu et al. MobiRFPose ***TMM***, 2024.

Lightweight Pose Estimation for Mobile Devices

- **MobiRFPose**



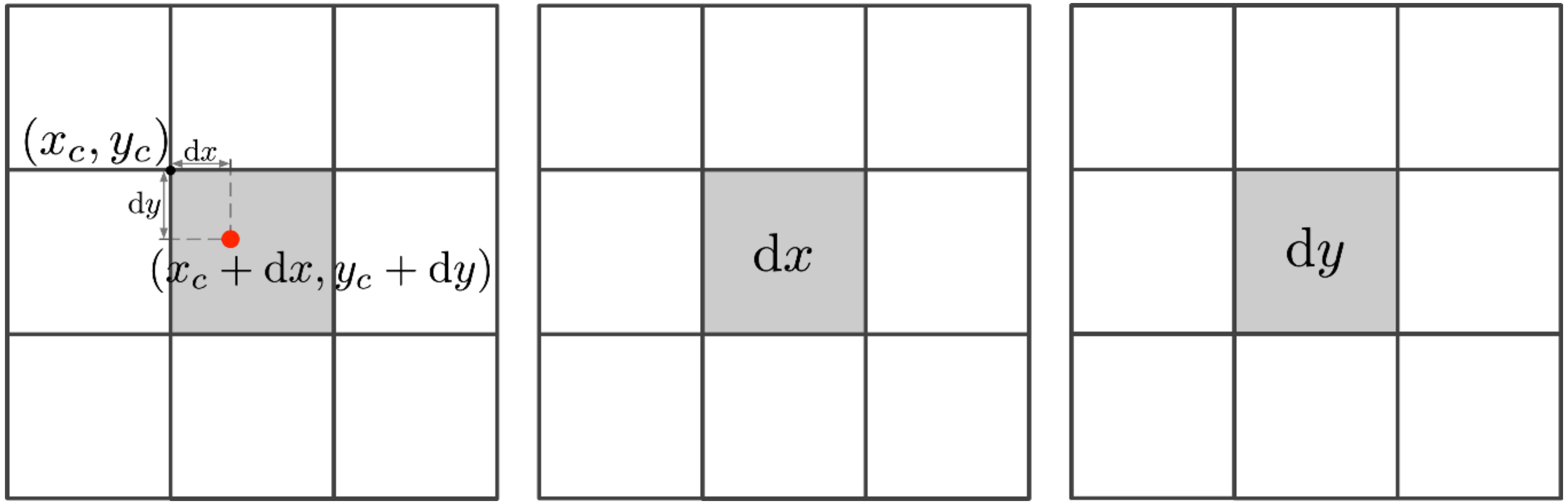
Yu et al. MobiRFPose ***TMM***, 2024.

Lightweight Pose Estimation for Mobile Devices

- **Human Localization:** Each cell produces (c, dx, dy)

Location Map $\hat{\boldsymbol{M}} = E_L(\boldsymbol{S})$. When $c > \delta$, human object exists Human Location in the Location Map

$$x_l = x_c + \mathrm{d}x$$
$$y_l = y_c + \mathrm{d}y$$



Human Location in the RF Heatmap

$$\hat{\boldsymbol{p}}_r = (\hat{x}_r, \hat{y}_r) = \alpha(x_l, y_l)$$

Human Location in the Physical Space

$$\hat{\boldsymbol{p}}_o = (\hat{x}_o, \hat{y}_o, 0) = \beta(x_l, y_l, 0)$$

Loss Function

$$\mathcal{L}_{HLN} = \|\hat{\boldsymbol{M}} - \boldsymbol{M}\|_2$$

Lightweight Pose Estimation for Mobile Devices

University of Science and Technology of China

Human Pose Estimation

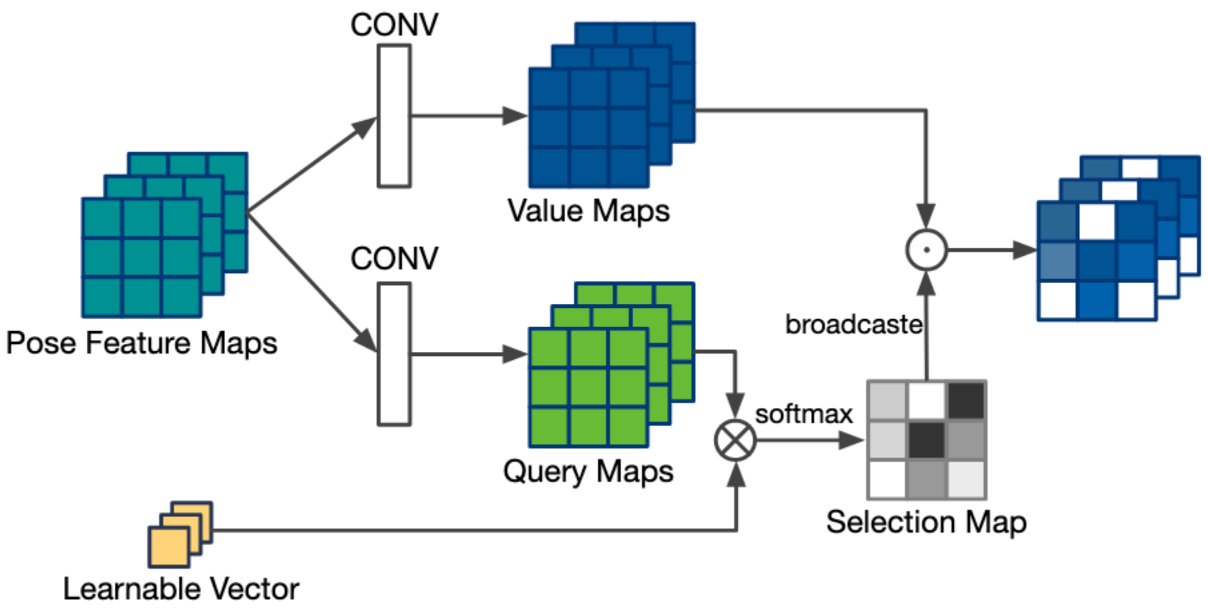
1. Given cropped pose feature Map

$$\boldsymbol{f}_{\tilde{\boldsymbol{S}}} = E_P(\tilde{\boldsymbol{S}})$$

2. Use attention to refine the feature map

$$f_V = \mathrm{CONV}_1\left(f_{\tilde{S}}\right), f_Q = \mathrm{CONV}_2\left(f_{\tilde{S}}\right),$$

$$\boldsymbol{f}'_{\tilde{S}} = \boldsymbol{f_V} \odot \mathrm{B}^{N_f}\left[\mathrm{softmax}\left(\boldsymbol{f_Q}\boldsymbol{w}^T\right)\right]$$



Pose Feature Selection (PFS)

3. Pose Keyoints Prediction

$$\hat{\boldsymbol{p}}_p = F(\boldsymbol{f}'_{\tilde{\boldsymbol{S}}})$$

Loss Function

$$\mathcal{L}_{PEN} = \|\hat{\boldsymbol{p}}_p - \boldsymbol{p}_p\|_2$$

Yu et al. MobiRFPose ***TMM***, 2024.

Lightweight Pose Estimation for Mobile Devices

- **Experiments**

Table 1. Quantitative evaluation results of different methods.

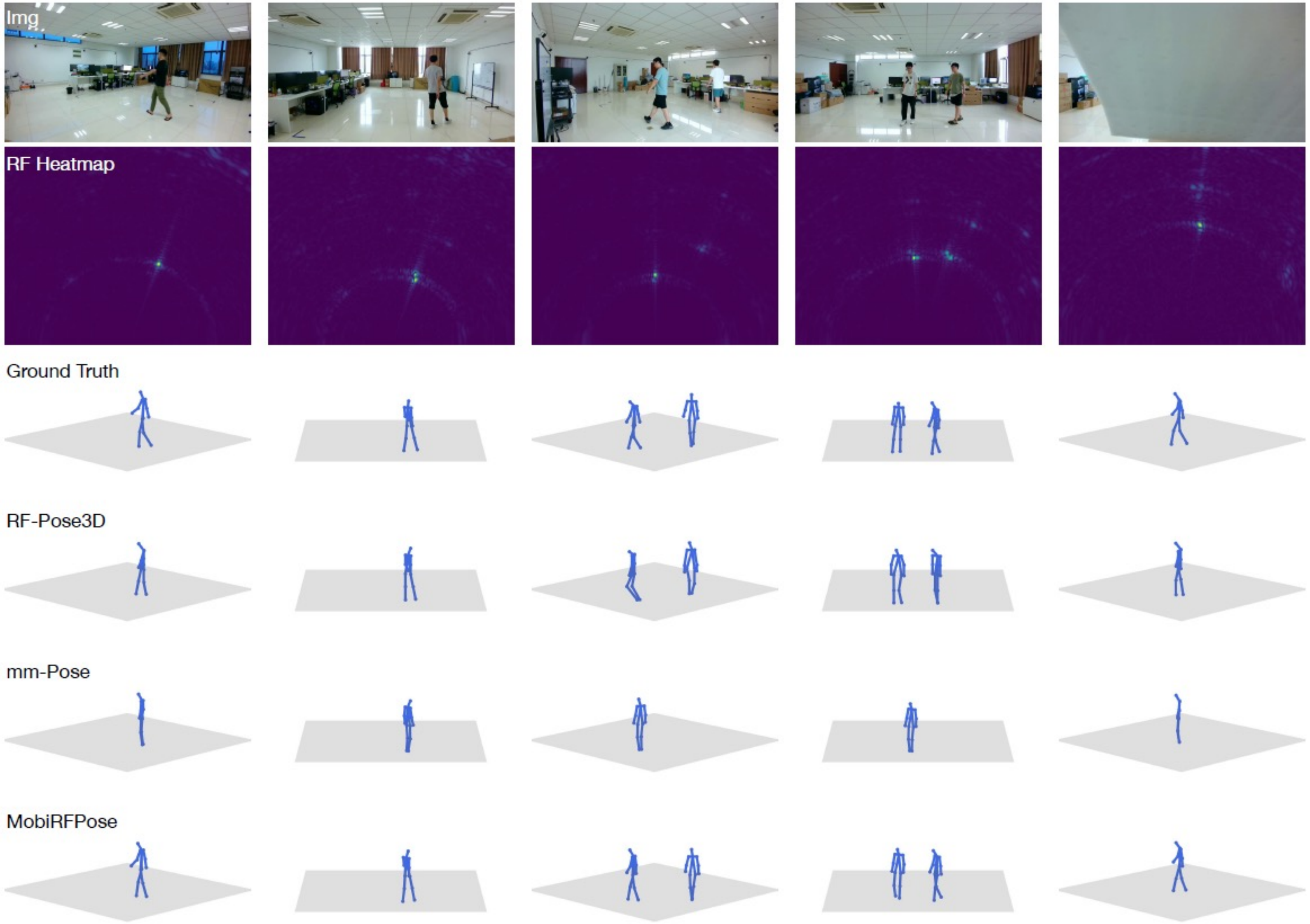
Methods	RF-HPED-A		RF-HPED-B		Params (M) ↓	MACs (G) ↓
	HDA (%) ↑	MPJPE (cm) ↓	HDA (%) ↑	MPJPE (cm) ↓		
RF-Pose3D [16]	97.76	13.68	94.32	14.04	9.492	25.33
mm-Pose [23]	**100.0**	**10.26**	–	–	33.08	0.246
Fast RFPose	98.54	11.05	**96.13**	**11.29**	**0.813**	**0.170**

RF-HPED-A: Single Person
RF-HPED-B: Multiple Persons
HDA: Human Detection Accuracy
MACs: Multiply-Accumulate Operations

Yu et al. MobiRFPose ***TMM***, 2024.

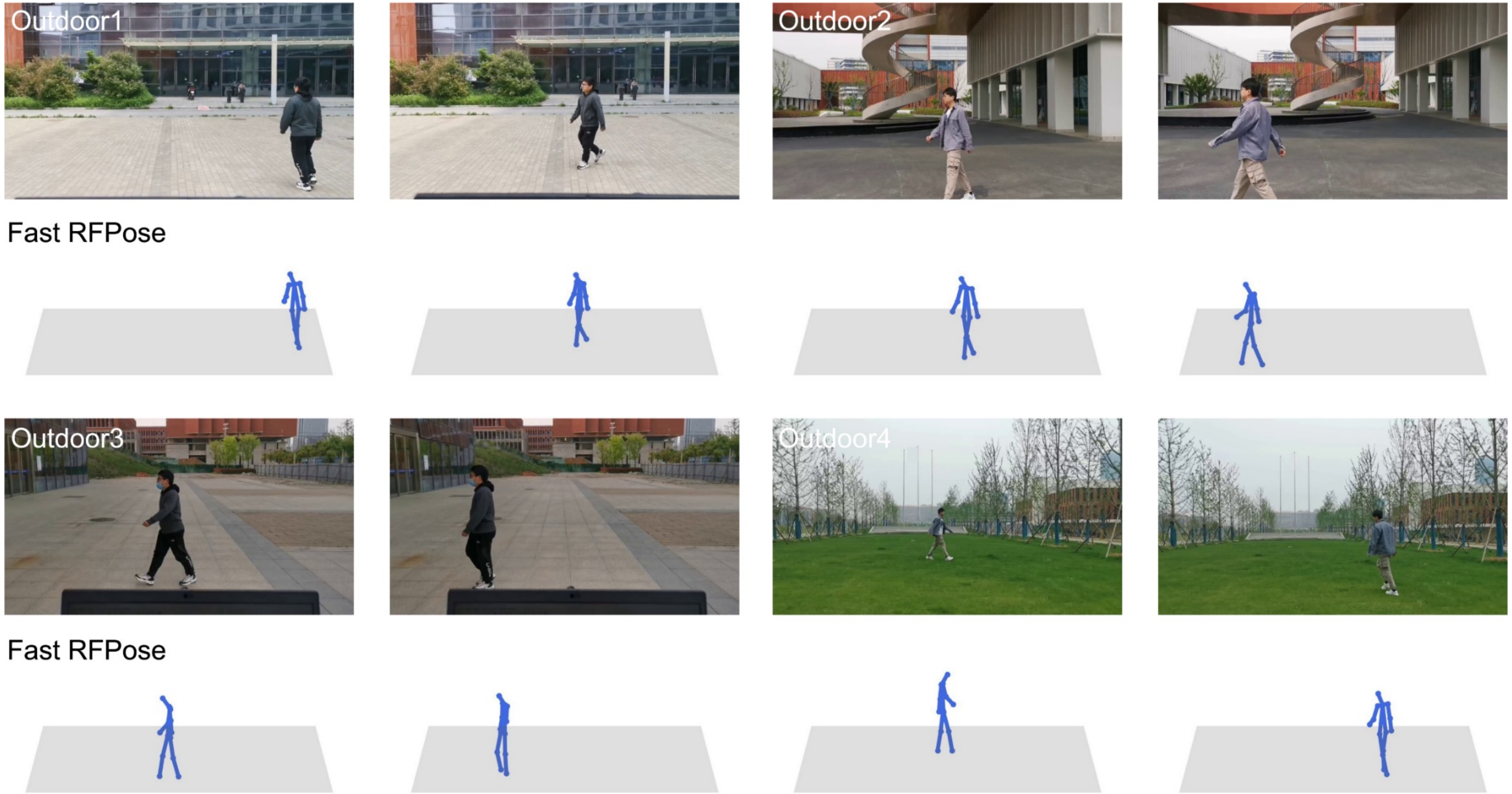
Lightweight Pose Estimation for Mobile Devices

- **Experiments**



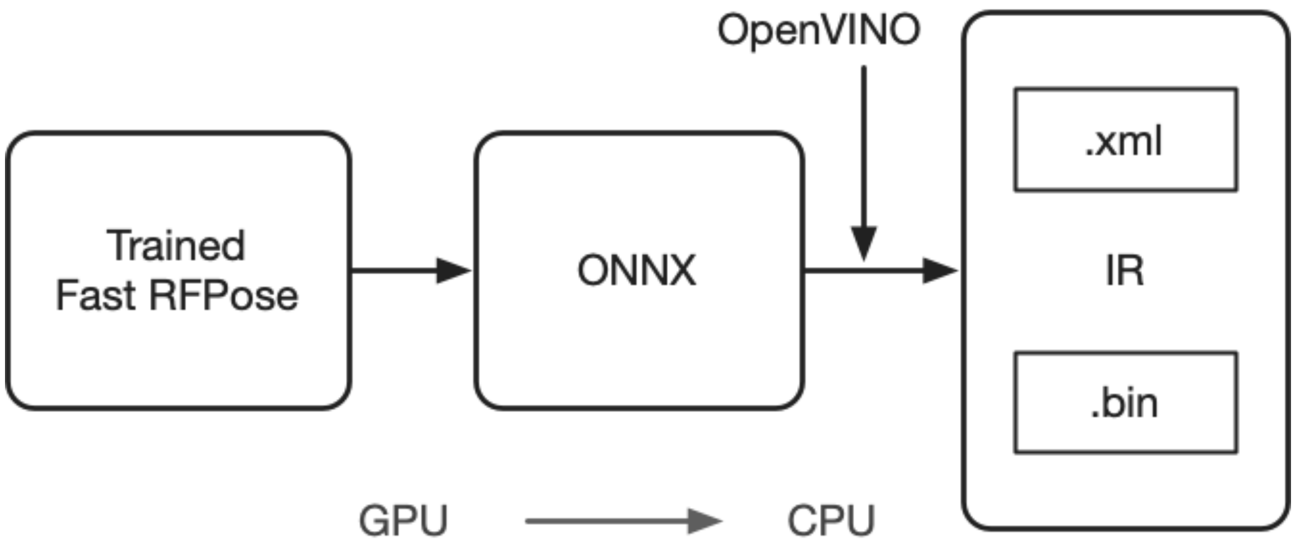
Lightweight Pose Estimation for Mobile Devices

- **Outdoor**



Lightweight Pose Estimation for Mobile Devices

- **Deployment on Mobile Devices**



CPU: 1 processor with 1.6 GHz and 2 cores

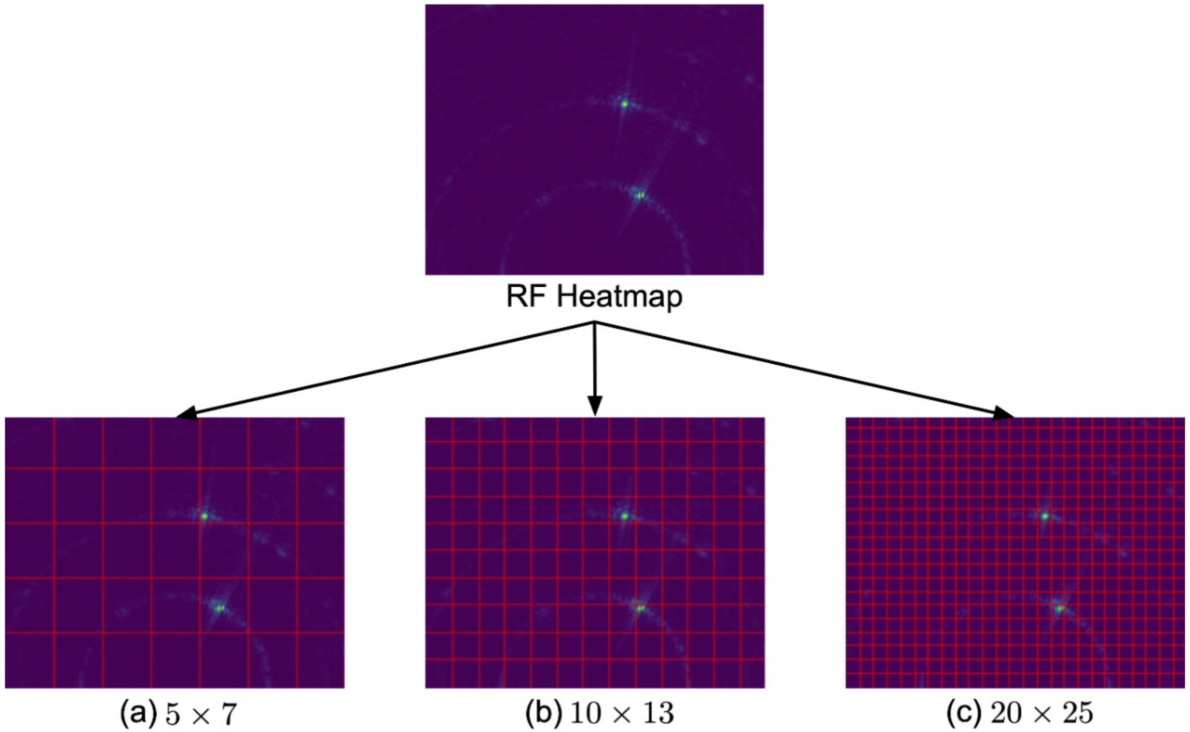
Speed: 66 FPS

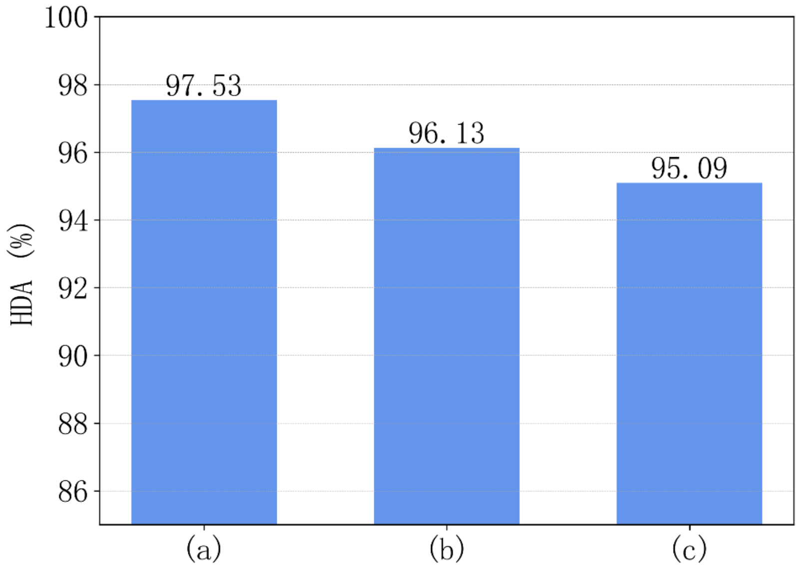
Yu et al. MobiRFPose ***TMM***, 2024.

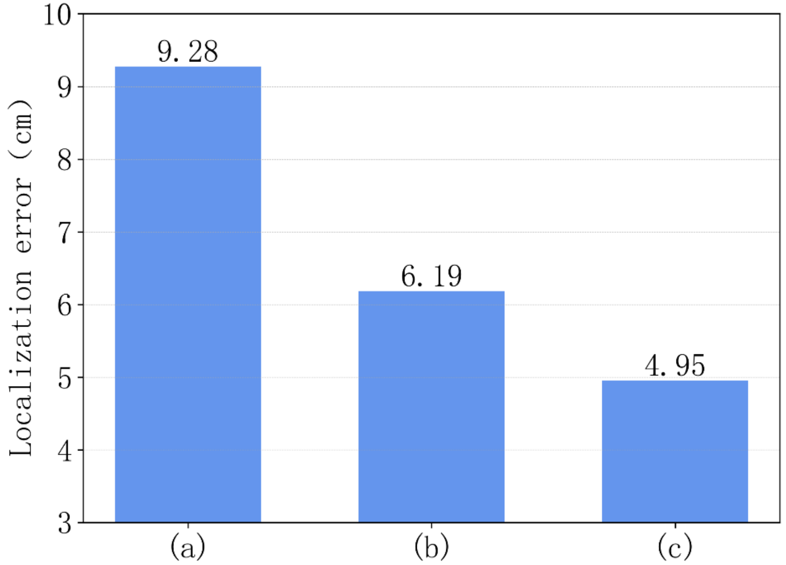
中国科学技术大学
University of Science and Technology of China

Lightweight Pose Estimation for Mobile Devices

- **Detection Grid**







Sparse detection grid: HDA ↑ Localization error ↑

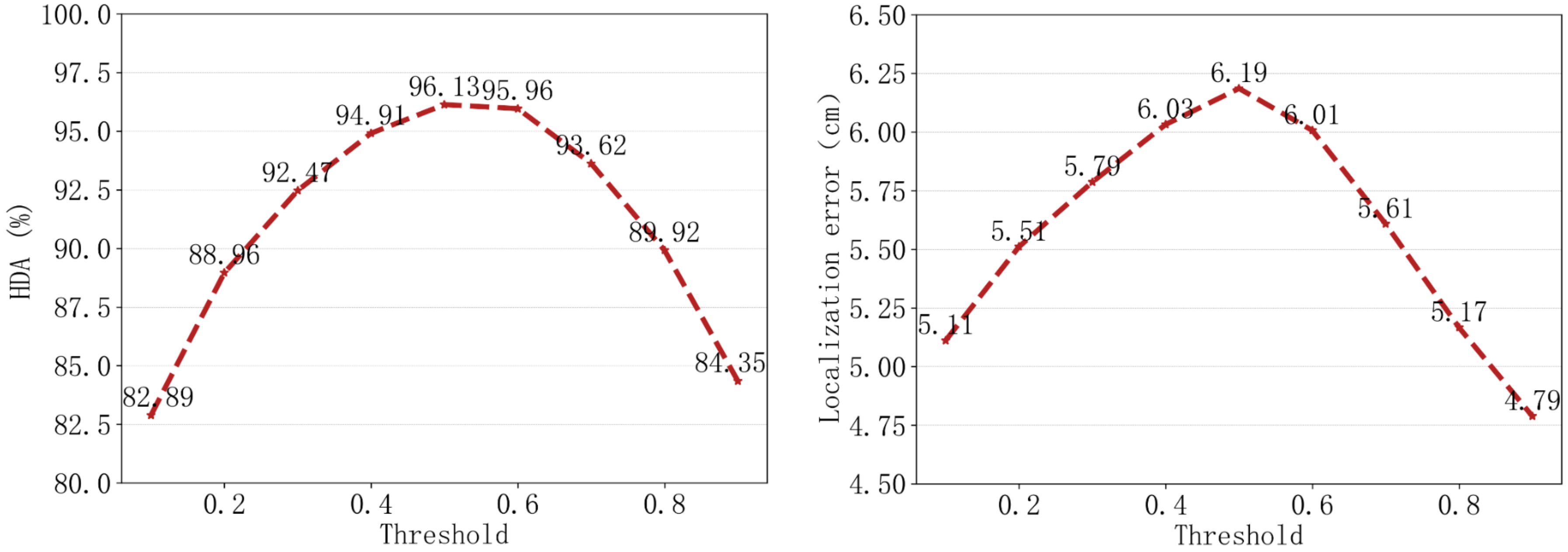
Dense detection grid: HDA ↓ Localization error ↓

HDA: Human Detection Accuracy

Yu et al. MobiRFPose ***TMM***, 2024.

Lightweight Pose Estimation for Mobile Devices

- **Detection Threshold**



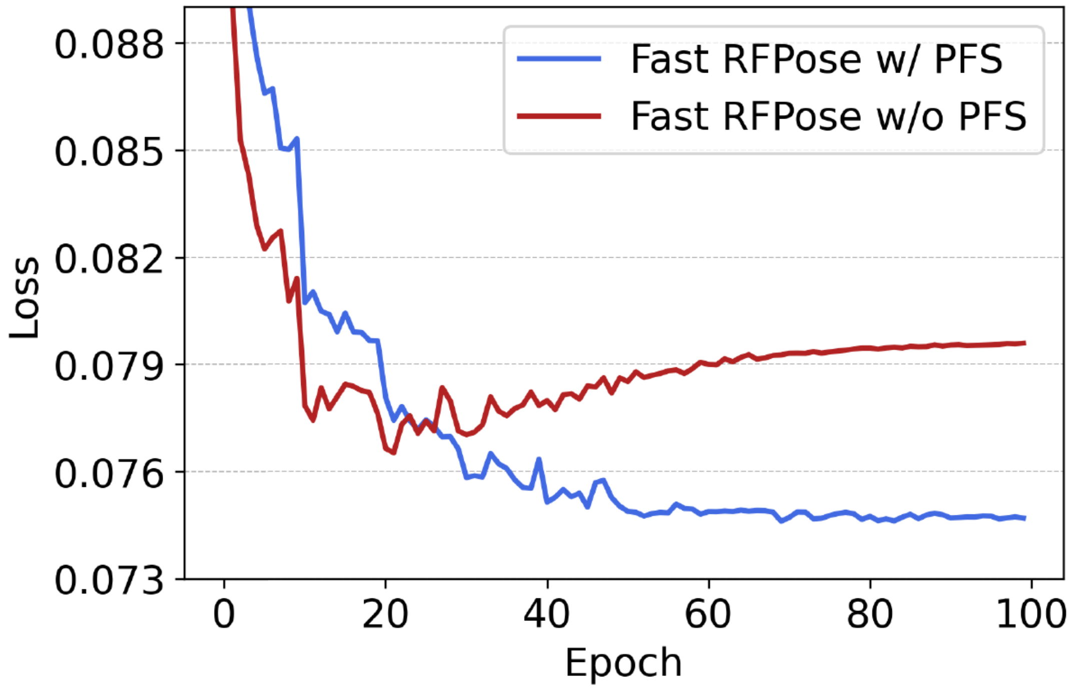
Small thresholds: Over-detection, Large thresholds: Miss-detection

Localization error ↓

Yu et al. MobiRFPose ***TMM***, 2024.

Lightweight Pose Estimation for Mobile Devices

- **Effect of the PFS Module**



The PFS module can prevent the model from overfitting, and it is extremely lightweight, with Params at 0.008M and MACs at 0.0005G

Yu et al. MobiRFPose ***TMM***, 2024.

Lightweight Pose Estimation for Mobile Devices

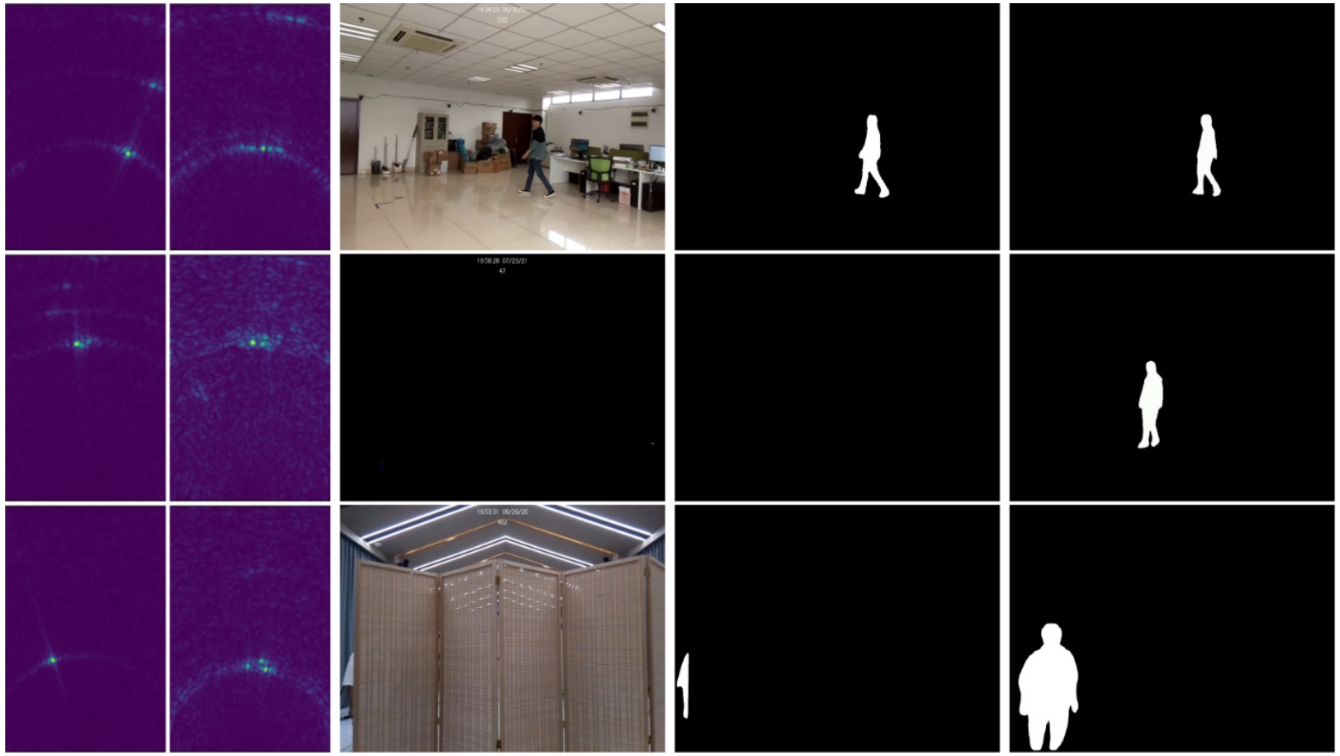
- Lightweight RF-based human pose estimation model
  - Transmitting and receiving signals using only horizontal antenna arrays
  - Model structure from whole to local
- Design Pose Feature Selection (PFS) to efficiently extract human posture information
  - Prevent the model from overfitting with very few parameters and computations
- Mobile deployment validates the lightweight and real-time characteristics
  - Runs at 66FPS

Yu et al. MobiRFPose ***TMM***, 2024.

RF-Based Human Pose Silhouette Segmentation

RF-Based Human Pose Silhouette Segmentation

- **Problem**



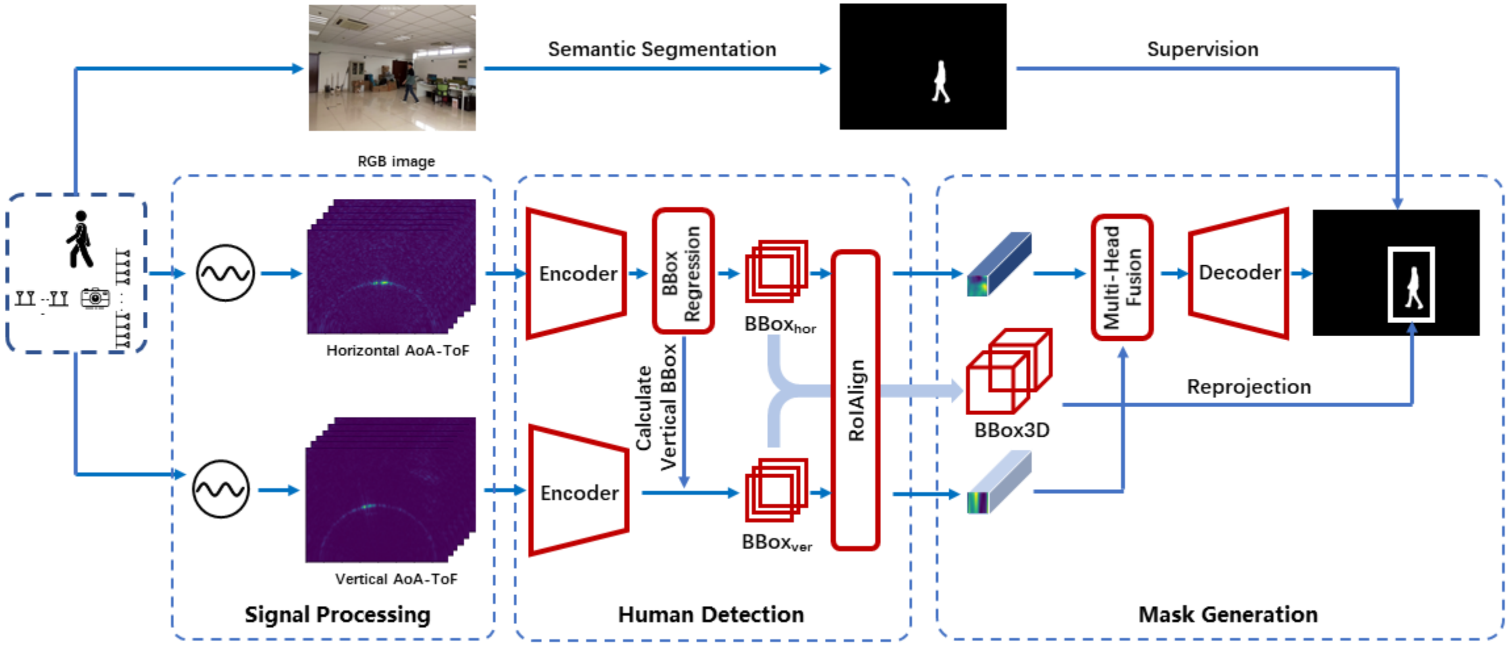
How to achieve more fine-grained RF-based human posture sensing?

Extracting fine-grained information for body contouring

Wu et al. RFMask **TMM** 2022

RF-Based Human Pose Silhouette Segmentation

- **RFMask System Overview**

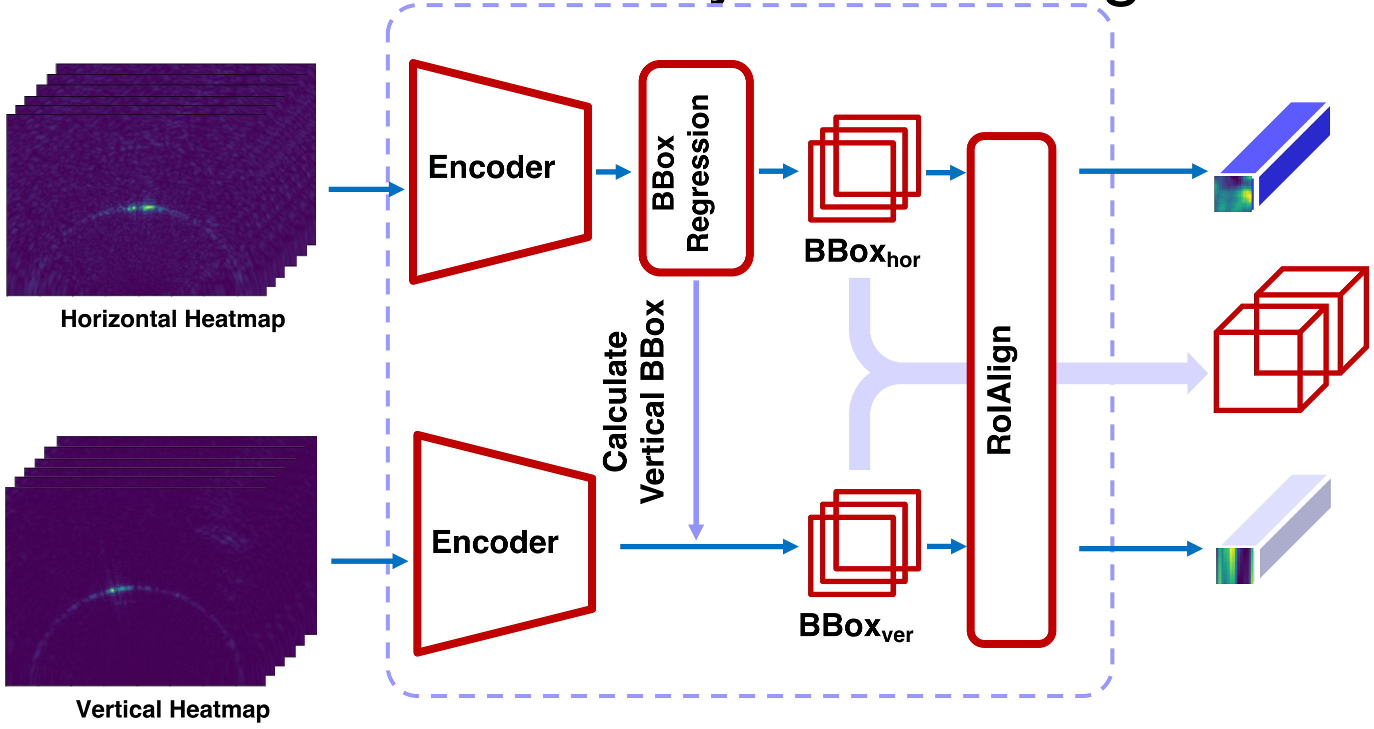


- **Signal Processing Module**: Generate horizontal and vertical heatmaps from raw wireless signals.
- **Human Detection Module**: Detect the target's 3D position using horizontal signals and geometric relationships with vertical signals.
- **Mask Generation Module**: Extract RoI features, decode the semantic segmentation map, and merge it with the original image for a complete segmentation result.

Wu et al. RFMask **TMM** 2022

RF-Based Human Pose Silhouette Segmentation

Human Detection: locate and identify human targets in 3D space

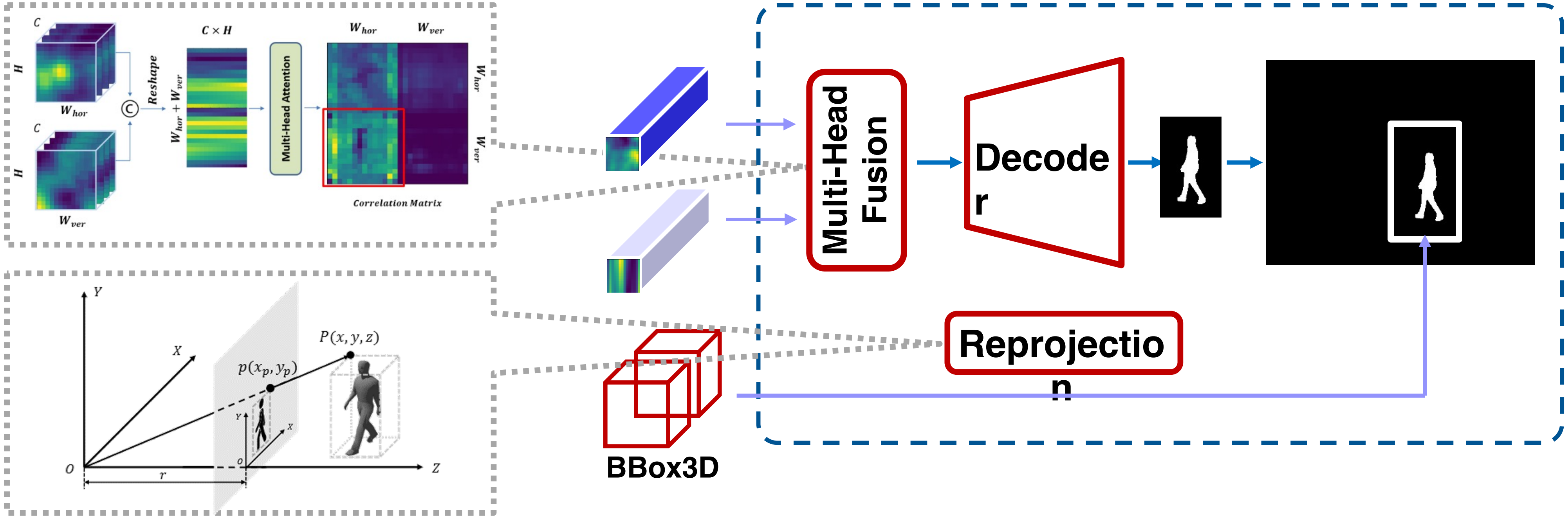


- **Bounding-box regression module:** Predicts target positions and bounding boxes.
- **Feature cropping module:** Extracts horizontal and vertical features based on target positions.
- **3D bounding box composition module:** Merges horizontal and vertical boxes into a 3D bounding box using geometric relationships.

Wu et al. RFMask **TMM** 2022

RF-Based Human Pose Silhouette Segmentation

- **Mask Generation Module:** generate a semantic segmentation map



- **Multi-Head Fusion Module:** The multi-head fusion module utilizes attention mechanisms to combine spatial information with features.
- **Decoder Module:** The decoder module decodes the fused features to generate human contour information.
- **Reprojection Module:** The reprojection module projects the 3D bounding boxes onto the imaging plane, complementing the missing position information in the RoI features, and generating a complete contour segmentation map.

RF-Based Human Pose Silhouette Segmentation

- **Loss Function**
  - Human detection loss

$$L_{detect}(p, p^u, v, t^u) = L_{cls}(\underbrace{\boxed{p}}_{\text{predicted probability}}, p^u) + \underbrace{\boxed{\lambda_{det}}}_{\text{balance weight}} [u \geq 1] L_{box}(\underbrace{\boxed{t^u}}_{\text{predicted coordinate}}, v),$$

  - Silhouette prediction loss

$$L_{mask} = \frac{1}{\underbrace{\boxed{N_{box}}}_{\text{detected number of targets}}} \sum_{i=1} L_m(\underbrace{\boxed{m_{i,k}}}_{\text{predicted silhouette}}, m_k^*),$$

  - Total loss

$$L = L_{detect} + L_{mask}.$$

Where L_{cls} and L_{m} represent binary cross-entropy loss, and L_{box} is the smooth L_1 loss.

Wu et al. RFMask **TMM** 2022

RF-Based Human Pose Silhouette Segmentation

- **Comparative Experiments**

TABLE I
COMPARISONS WITH RFPOSE

Model	Single-Person	Multi-Person	Action
RFPose(4)	0.664	0.626	0.616
RFPose(12)	0.675	0.631	0.614
RFPose(32)	0.661	0.617	0.598
RFPose(64)	0.641	0.589	0.604
RFMask(4)	0.681	0.682	0.681
RFMask(12)	**0.706**	**0.711**	**0.705**

Mask IoU: Measures the similarity between predicted and ground truth segmentation by calculating the ratio of their intersection to their union.

The proposed method significantly outperforms the baseline methods in the silhouette generation task

RF-Based Human Pose Silhouette Segmentation

- **Comparison Experiment: Impact of Sequence Length and Backbone Network Architecture on Detection Performance.**

TABLE II
LOCATION ACCURACY

Model	Backbone	$AP_{50:95}$	AP_{50}	AP_{75}	Recall
RFMask(4)	ResNet-18	0.586	0.966	0.678	0.662
RFMask(4)	ResNet-34	0.590	0.966	0.689	0.665
RFMask(4)	ResNet-50	0.581	0.966	0.671	0.656
RFMask(12)	ResNet-18	0.621	0.967	0.783	0.691
RFMask(12)	ResNet-34	0.631	0.967	0.817	0.699
RFMask(12)	ResNet-50	**0.632**	**0.967**	**0.824**	**0.701**

AP (Average Precision): Precision represents the proportion of correctly detected targets among the detected targets, while Recall represents the proportion of correctly detected targets among all the actual targets. AP value represents the average area under the Precision-Recall curve.

The proposed method performs well across different sequence lengths and when using different backbone networks.

Wu et al. RFMask **TMM** 2022

RF-Based Human Pose Silhouette Segmentation

- **Ablation Study:**

TABLE III
ABLATION STUDY

	H	H & V	H & V
Dual-Branch		✓	✓
Multi-Head Fusion			✓
Single-Person(4)	0.634	0.644	**0.681**
Multi-Person(4)	0.587	0.604	**0.682**
Action(4)	0.582	0.585	**0.681**
Single-Person(12)	0.655	0.670	**0.706**
Multi-Person(12)	0.638	0.642	**0.711**
Action(12)	0.603	0.603	**0.705**

Each module and its structure in the proposed method make significant contributions to the final performance.

Wu et al. RFMask **TMM** 2022

RF-Based Human Pose Silhouette Segmentation

- **Visualization Results: (Typical Scenarios)**

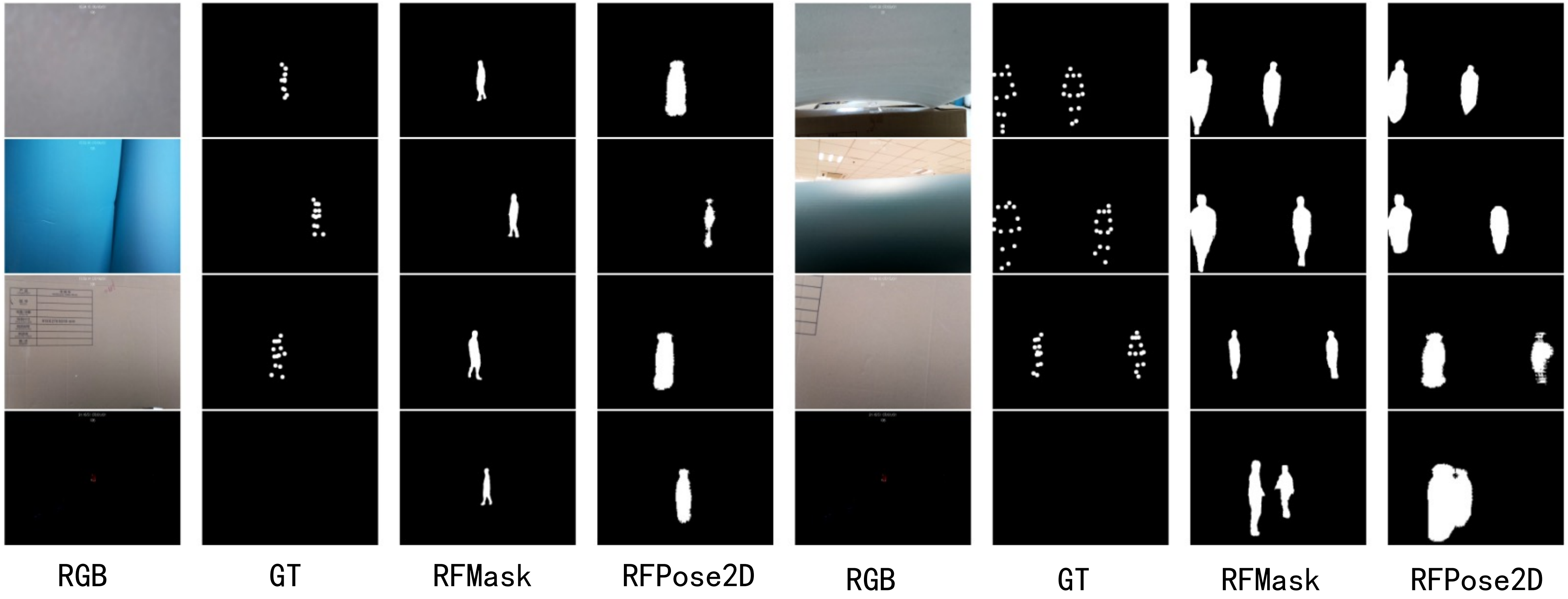
RGB GT RFMask RFPose2D RGB GT RFMask RFPose2D

The proposed method can effectively accomplish the contour generation task in typical scenarios.

Wu et al. RFMask **TMM** 2022

RF-Based Human Pose Silhouette Segmentation

- **Visual results display: (Special scenarios)**



The proposed method performs consistently well in special scenarios without significant performance degradation.

Wu et al. RFMask **TMM** 2022

RF-Based Human Pose Silhouette Segmentation

- Human silhouette extraction system from RF signals.
- Lightweight two-stage generation model.
- Spatial feature fusion based on geometric relationships.
- Handles single/multiple individuals, low-light, occluded, and dark environments.

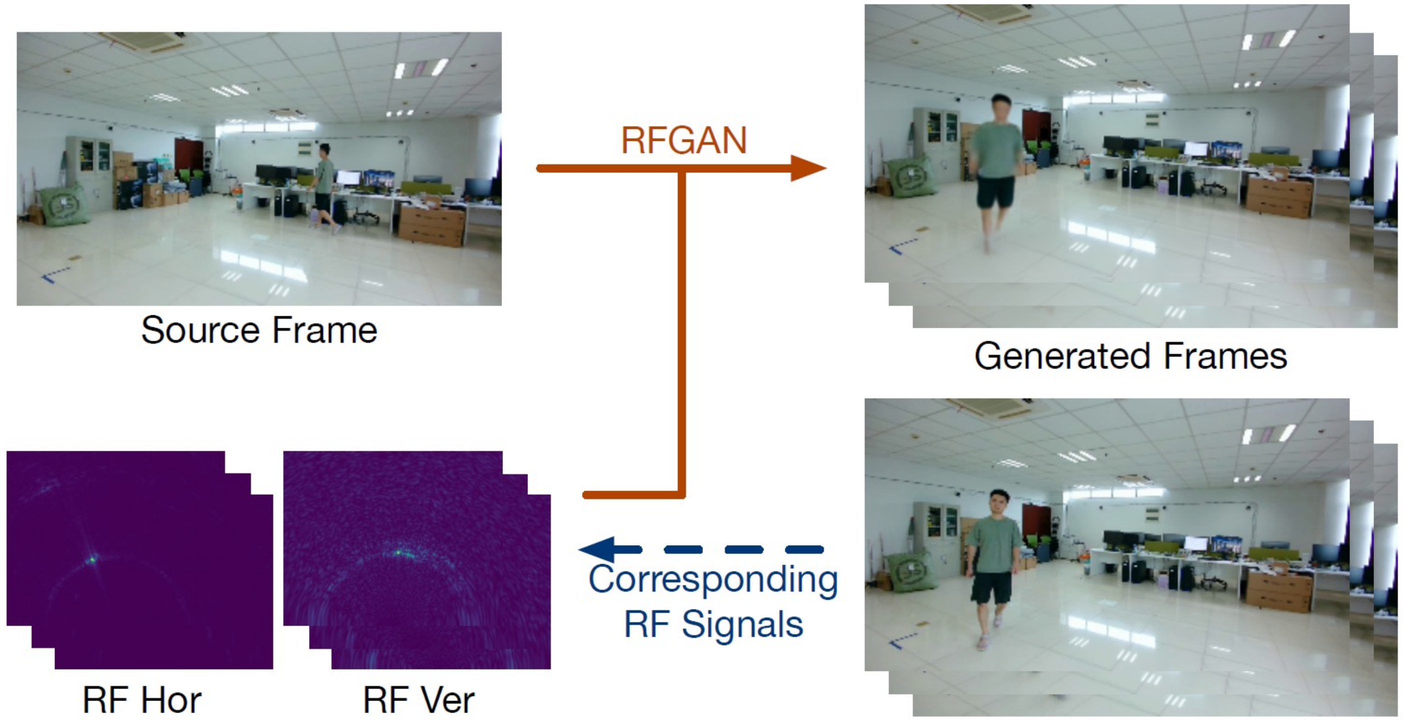
Wu et al. RFMask **TMM** 2022

Multimodal-Based Human Pose Visualization

Multimodal-Based Human Pose Visualization

Problem: RF-signal control Image to Video Synthesis

$$I_t = G(I_0, X_t^{\mathrm{RF}})$$



Challenges:

- Unsupervised RF Feature Extraction
- Fusion of Horizontal and Vertical RF Heatmaps
- Injecting RF Features into GAN

Yu et al. RFGAN **TMM** 2023

Multimodal-Based Human Pose Visualization

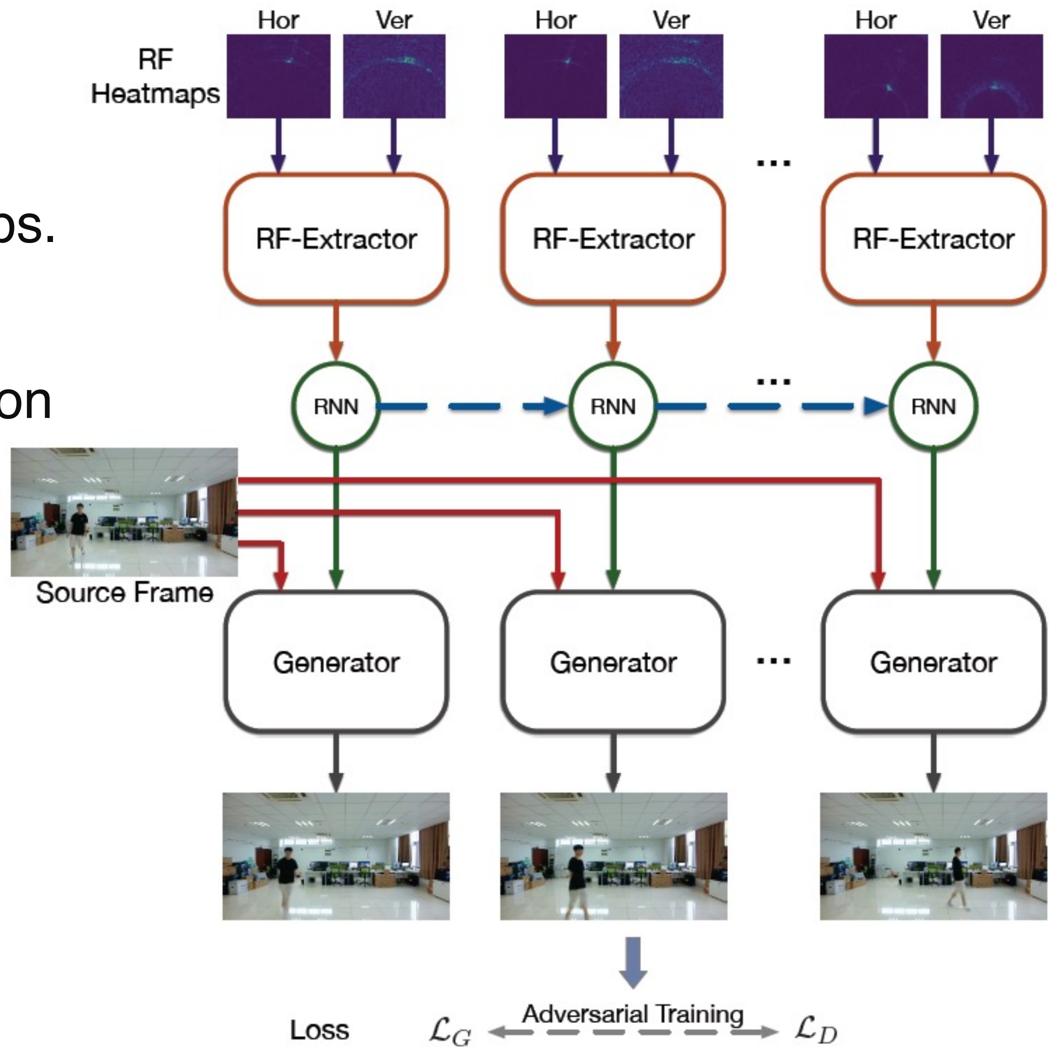
RFGAN

1. RF-Extractor with RNN:

- Two encoders for horizontal and vertical RF heatmaps.
- Feature maps are fused using RF-Fusion
- RNN capture temporal dependencies in human motion

2. RF-Based Generator

- Generate image given source image and features

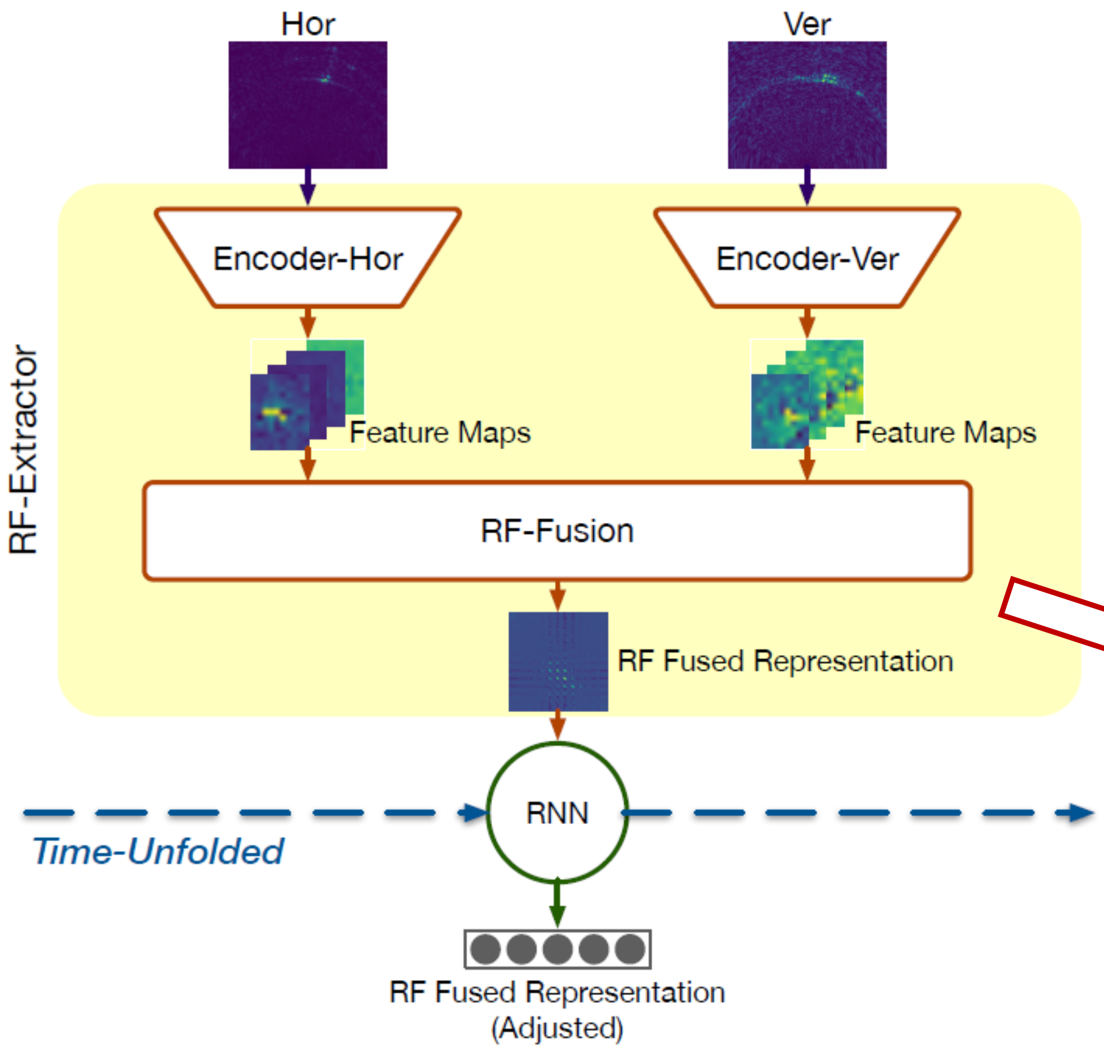


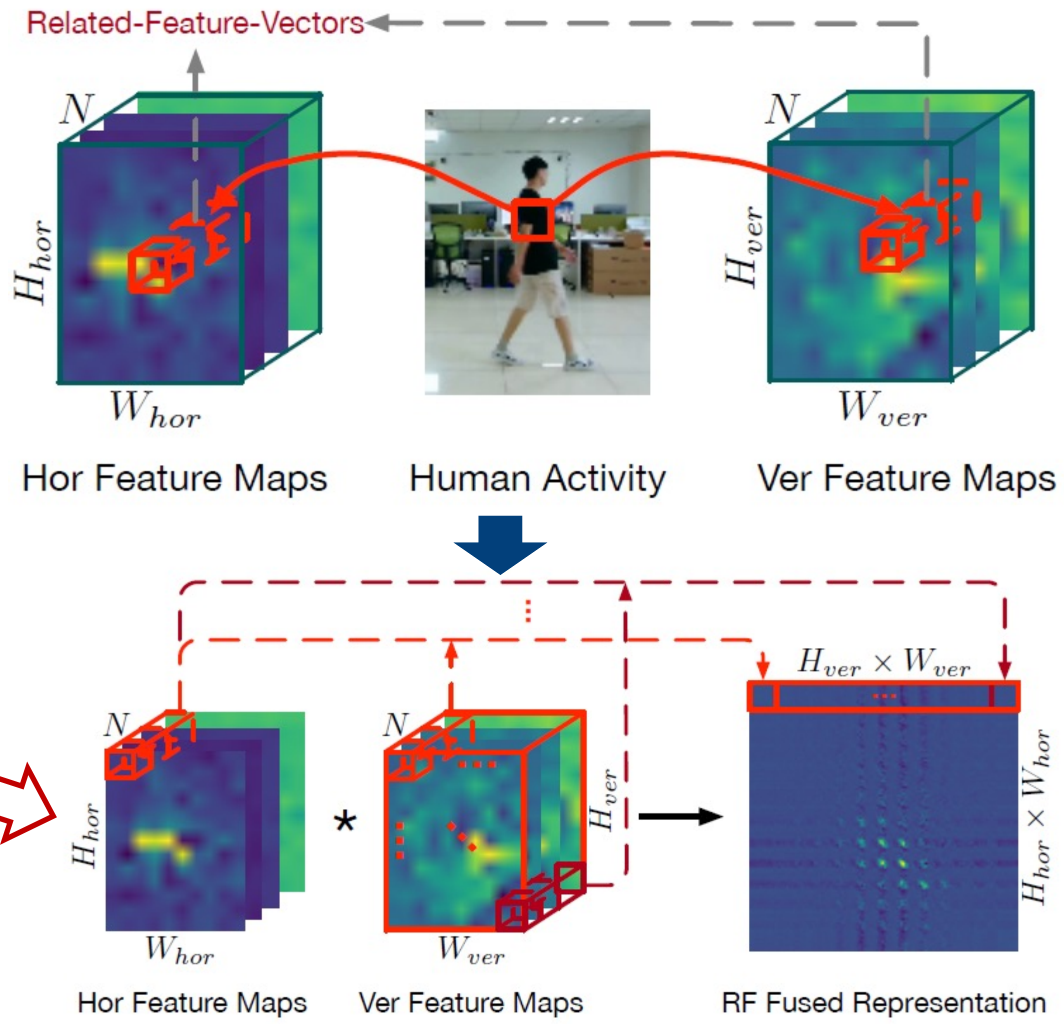
The first work to generate human images from the mmWave radar signals

Yu et al. RFGAN **TMM** 2023

Multimodal-Based Human Pose Visualization

- **RF-Extractor**



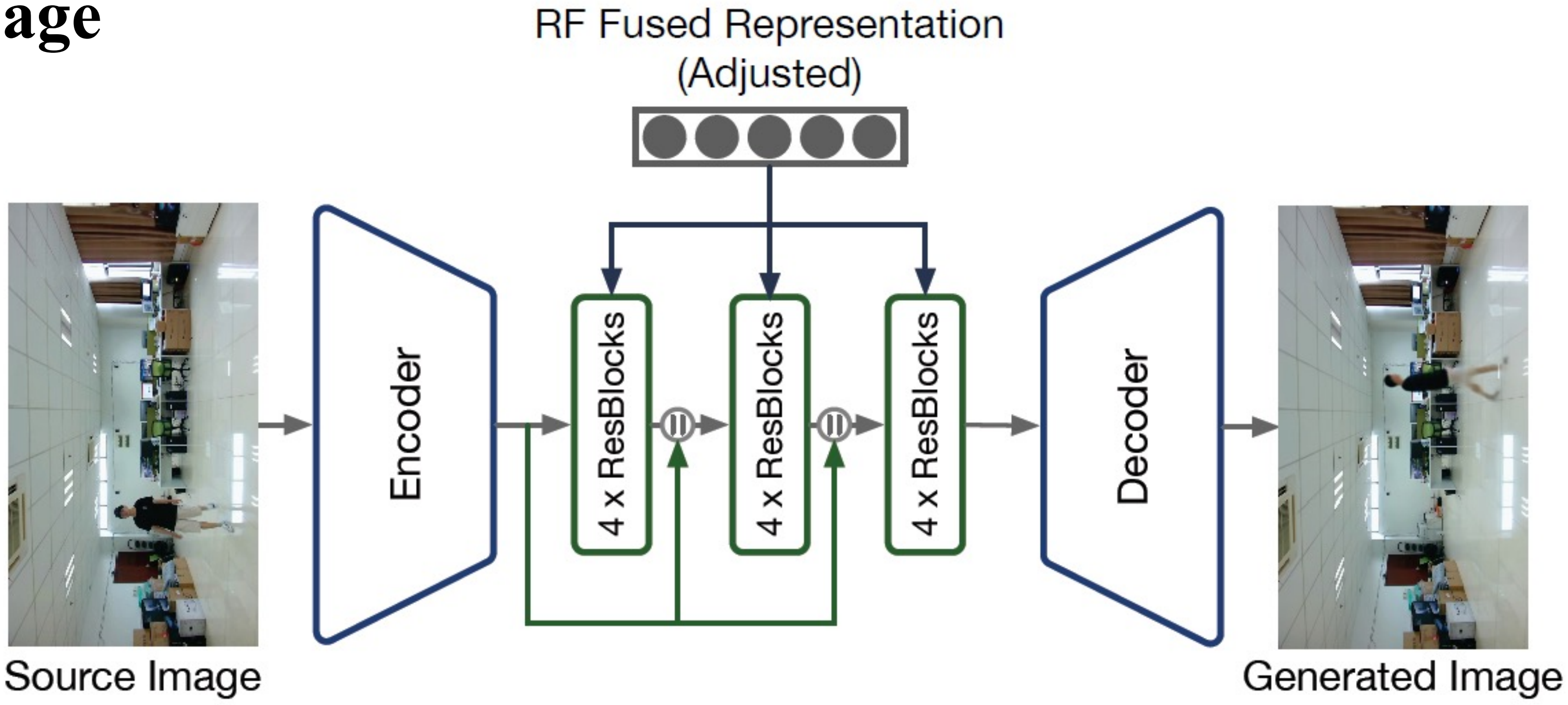


$$\boldsymbol{R}_{hv}(i,j) = \frac{\boldsymbol{R}_h(i)\boldsymbol{R}_v(j)^T}{\sqrt{N}}$$

中国科学技术大学
University of Science and Technology of China

Multimodal-Based Human Pose Visualization

- **Generator: generate new image with given RF signal and source image**

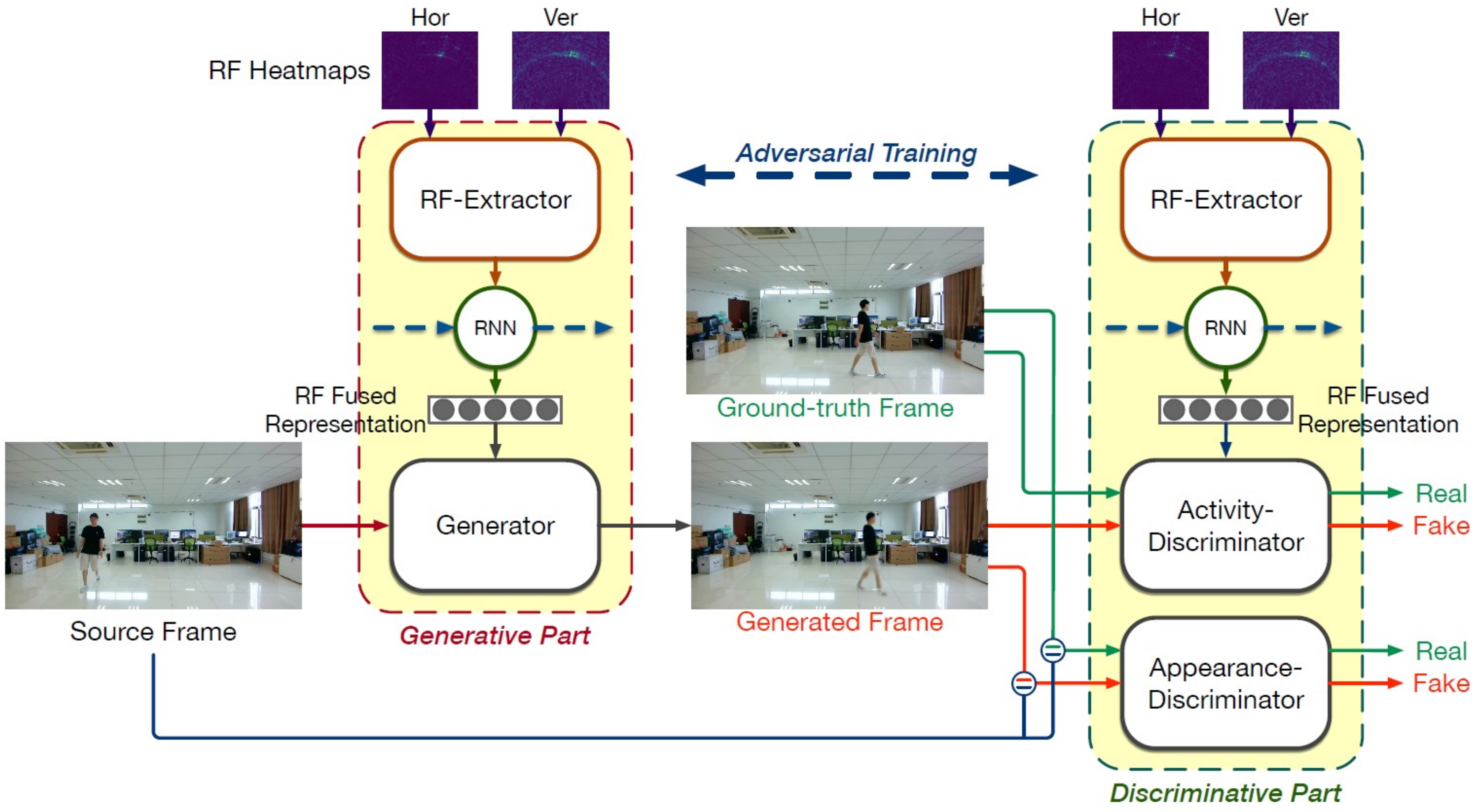


$$\text{RF-InNorm}(\boldsymbol{f}_{X_s,i}) = F_{\gamma,i}(\boldsymbol{h}) \cdot \frac{\boldsymbol{f}_{X_s,i} - \boldsymbol{\mu}_{X_s,i}}{\boldsymbol{\sigma}_{X_s,i}} + F_{\beta,i}(\boldsymbol{h})$$

Yu et al. RFGAN **TMM** 2023

Multimodal-Based Human Pose Visualization

Discriminative Part



1. Whether the image matches the pose?

D1(image, RF-features)

2. Whether the image matches the source image?

D2(image, source)

Yu et al. RFGAN **TMM** 2023

Multimodal-Based Human Pose Visualization

- **Loss Functions**

Activity-Discriminator

$$\mathcal{L}^{pos} = \mathcal{L}^{pos}_{LSD} + \lambda \mathcal{L}^{pos}_{GP}$$

$$\mathcal{L}^{pos}_{LSD} = \mathbb{E}_{\boldsymbol{X}_r \sim \mathbb{P}} \| D_{pos}(\boldsymbol{X}_r | E_{dis}(\boldsymbol{S}_h, \boldsymbol{S}_v)) - 1 \|_2^2 + \mathbb{E}_{\boldsymbol{X}_g \sim \mathbb{Q}} \| D_{pos}(\boldsymbol{X}_g | E_{dis}(\boldsymbol{S}_h, \boldsymbol{S}_v)) - 0 \|_2^2$$

$$\mathcal{L}^{pos}_{GP} = \mathbb{E}_{\boldsymbol{X}_r \sim \mathbb{P}} \| \nabla D_{pos}(\boldsymbol{X}_r | E_{dis}(\boldsymbol{S}_h, \boldsymbol{S}_v)) \|_2^2$$

Appearance-Discriminator

$$\mathcal{L}^{vis} = \mathcal{L}^{vis}_{LSD} + \lambda \mathcal{L}^{vis}_{GP}$$

$$\mathcal{L}^{vis}_{LSD} = \mathbb{E}_{\boldsymbol{X}_r \sim \mathbb{P}} \| (D_{vis}(\boldsymbol{X}_r | \boldsymbol{X}_s) - 1) \|_2^2 + \mathbb{E}_{\boldsymbol{X}_g \sim \mathbb{Q}} \| (D_{vis}(\boldsymbol{X}_g | \boldsymbol{X}_s) - 0) \|_2^2$$

$$\mathcal{L}^{vis}_{GP} = \mathbb{E}_{\boldsymbol{X}_r \sim \mathbb{P}} \| \nabla D_{vis}(\boldsymbol{X}_r | \boldsymbol{X}_s) \|_2^2$$

Generator

$$\mathcal{L}_G = \mathcal{L}_{LSG} + \alpha \mathcal{L}_{IMG} + \beta \mathcal{L}_{FEA}$$

$$\mathcal{L}_{LSG} = \mathbb{E}_{\boldsymbol{X}_g \sim \mathbb{Q}} \| (D_{pos}(\boldsymbol{X}_g | E_{dis}(\boldsymbol{S}_h, \boldsymbol{S}_v)) - 1 \|_2^2 + \mathbb{E}_{\boldsymbol{X}_g \sim \mathbb{Q}} \| D_{vis}(\boldsymbol{X}_g | \boldsymbol{X}_s) - 1) \|_2^2 .$$

$$\mathcal{L}_{IMG} = \mathbb{E}_{\boldsymbol{X}_g \sim \mathbb{Q}, \boldsymbol{X}_r \sim \mathbb{P}} \| \boldsymbol{X}_g - \boldsymbol{X}_r \|_1$$

$$\mathcal{L}_{FEA} = \sum_i^{N_c} \mathbb{E}_{\boldsymbol{X}_g \sim \mathbb{Q}, \boldsymbol{X}_r \sim \mathbb{P}} \left\| \boldsymbol{f}^{pos}_{\boldsymbol{X}_g, i} - \boldsymbol{f}^{pos}_{\boldsymbol{X}_r, i} \right\|_1 + \sum_i^{N_c} \mathbb{E}_{\boldsymbol{X}_g \sim \mathbb{Q}, \boldsymbol{X}_r \sim \mathbb{P}} \left\| \boldsymbol{f}^{vis}_{\boldsymbol{X}_g, i} - \boldsymbol{f}^{vis}_{\boldsymbol{X}_r, i} \right\|_1$$

Yu et al. RFGAN **TMM** 2023

Multimodal-Based Human Pose Visualization

- **Quantitative comparison experiments**
  - Walk

Methods	FID ↓	SSIM ↑	MSE ↓	FID (Crop) ↓	SSIM (Crop) ↑	MSE (Crop) ↓	AKD ↓
Img&RF	27.84	0.9622	14.923	142.77	0.6199	63.41	9.041
RF-Concat	21.08	0.9689	7.144	106.85	0.6548	54.67	7.947
RFGAN	**15.75**	**0.9695**	**6.691**	**78.68**	**0.6611**	**53.05**	**5.539**

  - Activity

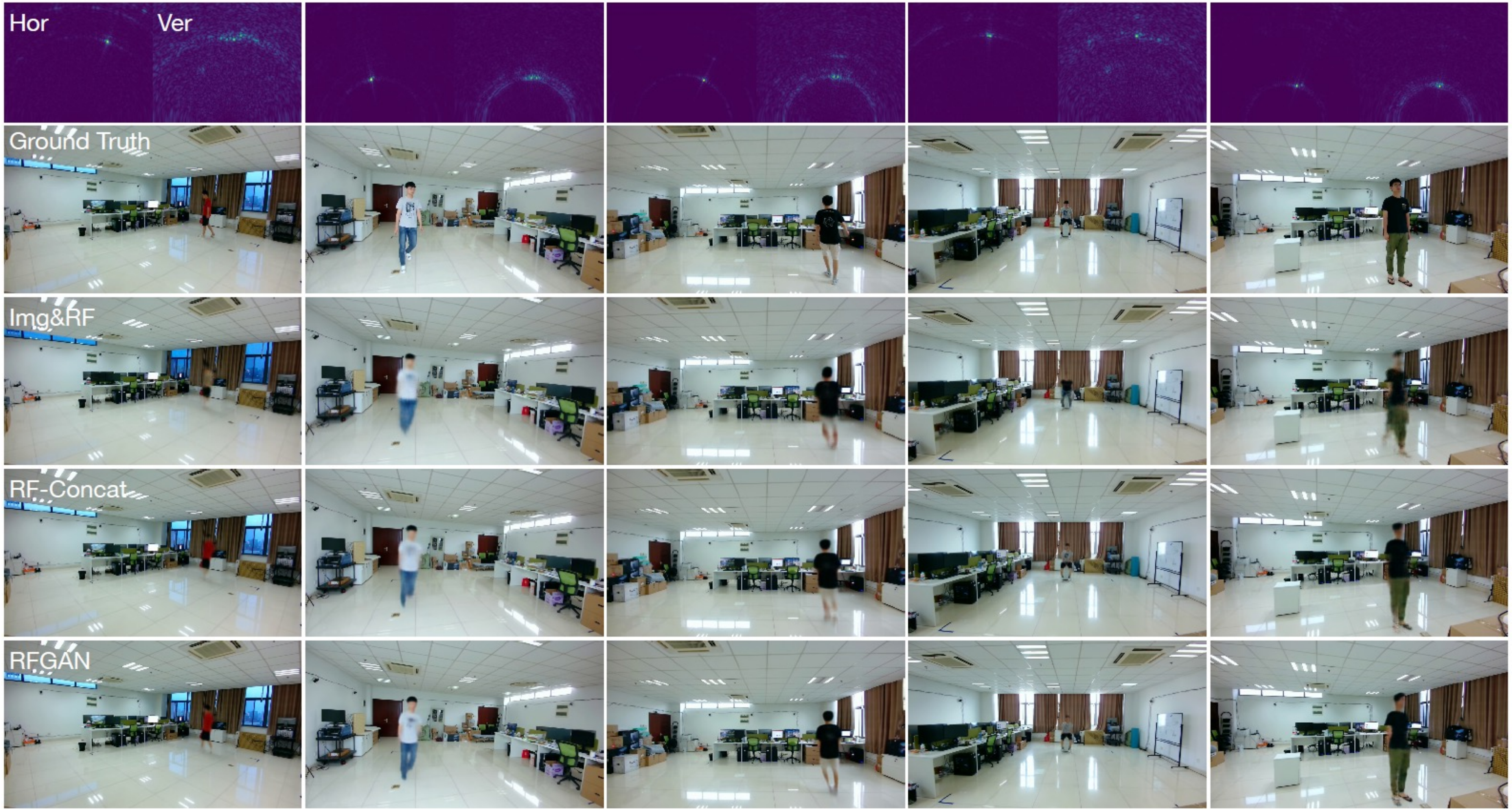
Methods	FID ↓	SSIM ↑	MSE ↓	FID (Crop) ↓	SSIM (Crop) ↑	MSE (Crop) ↓	AKD ↓
Img&RF	22.03	0.9643	12.862	133.20	0.6034	64.91	12.212
RF-Concat	19.19	0.9707	6.644	101.36	0.6501	55.49	8.996
RFGAN	**15.05**	**0.9708**	**6.572**	**76.16**	**0.6548**	**52.96**	**7.163**

Yu et al. RFGAN **TMM** 2023

Multimodal-Based Human Pose Visualization

中国科学技术大学
University of Science and Technology of China

- **Qualitative comparison experiments**



University of Science and Technology of China

Multimodal-Based Human Pose Visualization

- **Quantitative ablation experiments**
  - Walk

Methods	FID ↓	SSIM ↑	MSE ↓	FID (Crop) ↓	SSIM (Crop) ↑	MSE (Crop) ↓	AKD ↓
Img&RF	27.84	0.9622	14.923	142.77	0.6199	63.41	9.041
RF-Concat	21.08	0.9689	7.144	106.85	0.6548	54.67	7.947
RFGAN	**15.75**	**0.9695**	**6.691**	**78.68**	**0.6611**	**53.05**	**5.539**

  - Activity

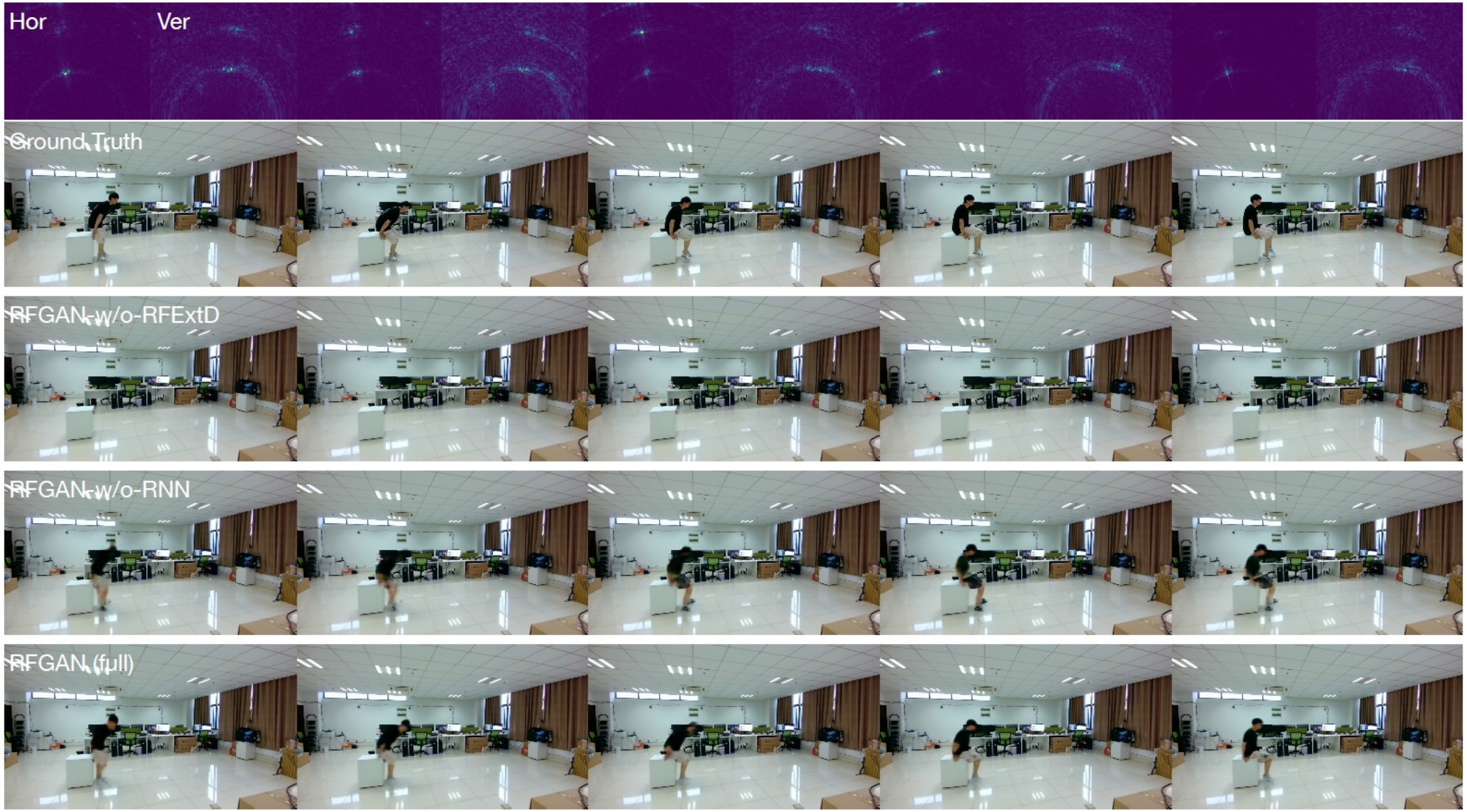
Methods	FID ↓	SSIM ↑	MSE ↓	FID (Crop) ↓	SSIM (Crop) ↑	MSE (Crop) ↓	AKD ↓
Img&RF	22.03	0.9643	12.862	133.20	0.6034	64.91	12.212
RF-Concat	19.19	0.9707	6.644	101.36	0.6501	55.49	8.996
RFGAN	**15.05**	**0.9708**	**6.572**	**76.16**	**0.6548**	**52.96**	**7.163**

Yu et al. RFGAN **TMM** 2023

Multimodal-Based Human Pose Visualization

中国科学技术大学
University of Science and Technology of China

- **Qualitative ablation experiments**



Multimodal-Based Human Pose Visualization

- **New Environments**

  - Quantitative

Methods	FID ↓	SSIM ↑	MSE ↓	FID (Crop) ↓	SSIM (Crop) ↑	MSE (Crop) ↓	AKD ↓
New Env 1	20.64	0.9739	5.735	184.98	0.6671	57.31	7.002
New Env 2	32.35	0.9192	25.76	122.40	0.5598	63.09	5.487

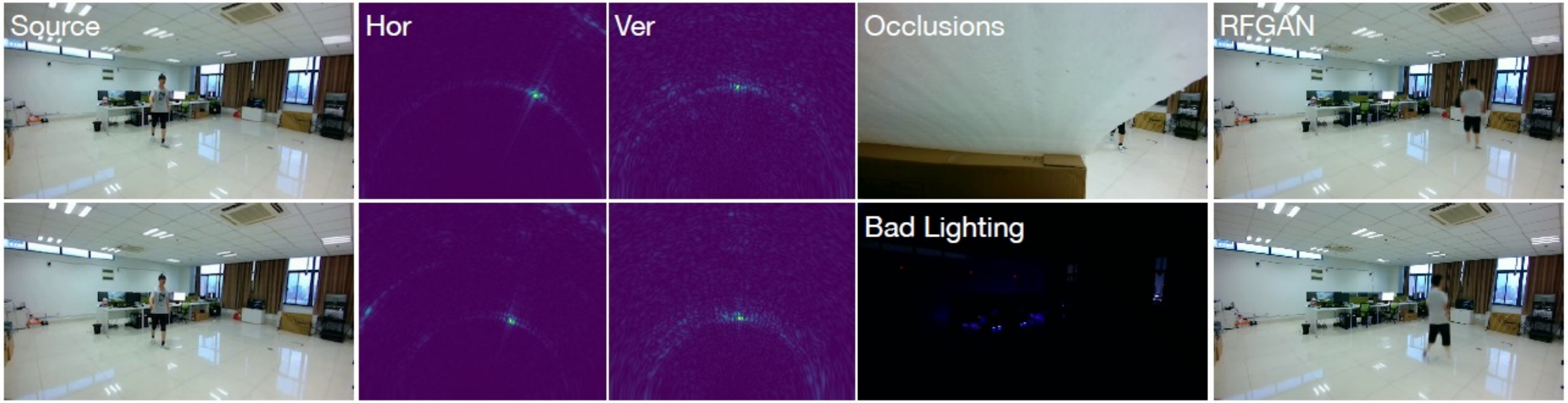
  - Qualitative

Multimodal-Based Human Pose Visualization

- **Occlusions and Bad Lighting**

  - When environmental conditions are good ➡ Source image
  - When environmental conditions deteriorate ➡ RF signals

  ⇨ All-Weather



Yu et al. RFGAN **TMM** 2023

Multimodal-Based Human Pose Visualization

Summary

- Multimodal Fusion for Human Pose Generation:

Combines RF signals and visual information, using adversarial learning and feature correlation matching to extract and fuse pose features seamlessly.

- Robust Performance in Diverse Conditions:

Achieves high-quality human pose generation with strong robustness in challenging scenarios, including dark, occluded, and unfamiliar environments.

Yu et al. RFGAN **TMM** 2023

Contents

RF-Based ECG Monitoring

Background

Cardiovascular disease (CVD) imposes a substantial burden on healthcare systems

Global Impact:
- CVD is a leading cause of death worldwide.
- Annual global deaths: 19 million.

Regional Impact:
- In China, CVD accounts for half of all deaths.
- In LMICs, 75% of global CVD deaths occur.

Economic Burden:
- Annual healthcare costs: $393 billion.
- Early medication use reduces individual CVD costs by 51%.

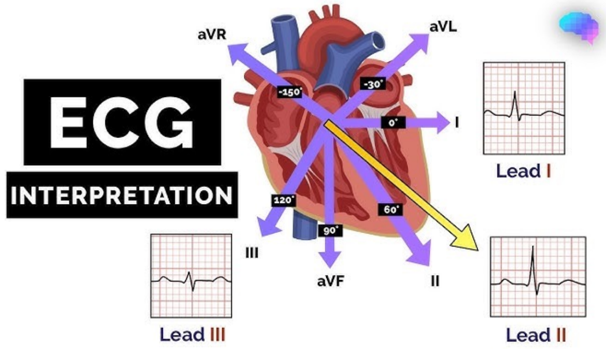
Good news: 80% of cardiovascular disease is preventable

RF-Based ECG Monitoring

Value of Electrocardiogram (ECG)

ECG shows significant ***prognosis value*** *and* ***diagnosis value*** *in potential.*

1. The **gold standard** for diagnosing various cardiac conditions.
   - Detects arrhythmias (e.g., atrial fibrillation).
   - Identifies myocardial infarctions (e.g., ST-segment changes).
   - Monitors bradycardia and heart block.
2. A critical tool for **postoperative** monitoring
   - Holter, External Loop, and Event Recorder
3. Significant **prognostic value**
   - HRV (beat-by-beat) shows significant prognostic value for cardiac events (HR = 1.47).
   - Baseline abnormalities link to overall mortality, CVD admissions, and major new abnormalities.



RF-Based ECG Monitoring

Dilemmas in Current ECG Workflow

1. **Economic burden**
   - The median cost of a routine ECG is $125 in the U.S.
   - In LMICs, ECG machines are often unaffordable for rural populations
2. Long-term monitoring is **inconvenient**
   - It relies on proper user operation
   - Wearable or adhesive-based devices can cause discomfort
3. Prognostic and screening value remains **underutilized**
   - Routine ECG struggles to capture intermittent or transient cardiac events
   - Implantable Cardiac Monitors (ICMs) are invasive and expensive

Our answer is a *cost-effective*, *contactless*, *continuous* cardiac monitoring device that is *connected-care* enabled and *convenient* for users.

RF-Based ECG Monitoring

Connected-care Cardiac Sensing System

Our goal is to achieve daily diagnosis and prevention, discover effective novel biomarkers for prognosis, and ultimately promote health equity.

- *Cost-effective:* Affordable and accessible for the majority.
- *Contactless:* Designed for ease of use without physical contact
- *Continuous:* Enables reliable, long-term monitoring
- *Convenient:* Easy to use and designed to fit effortlessly into daily life.

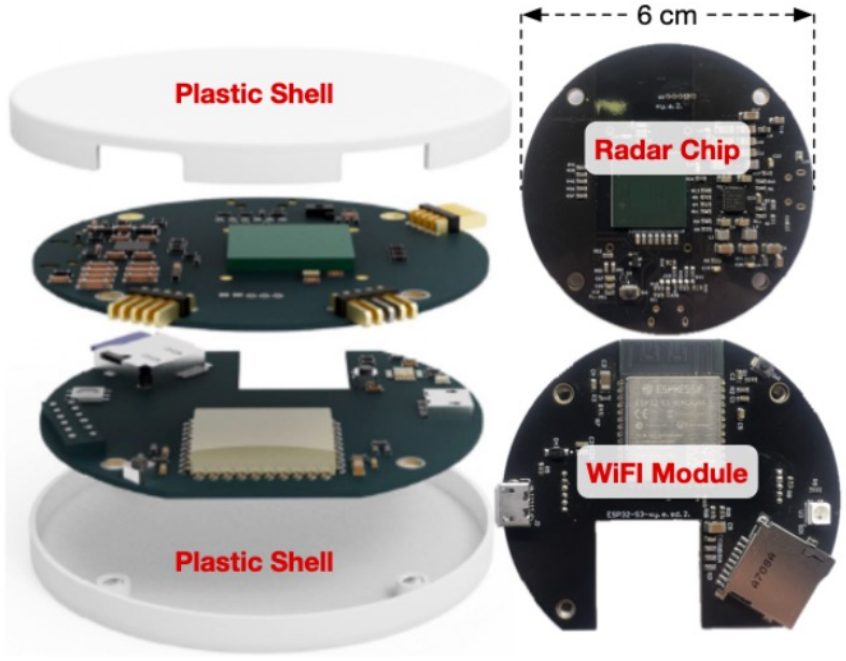
Our current efforts

- *Device:* Compact and efficient cardiac monitoring solution.
- *Feasibility:* Fine-grained cardiac imaging.
- *Preliminary:* Single-lead ECG monitoring to establish benchmarks.
- *Clinical validation:* Testing in real-world healthcare settings

RF-Based ECG Monitoring Device

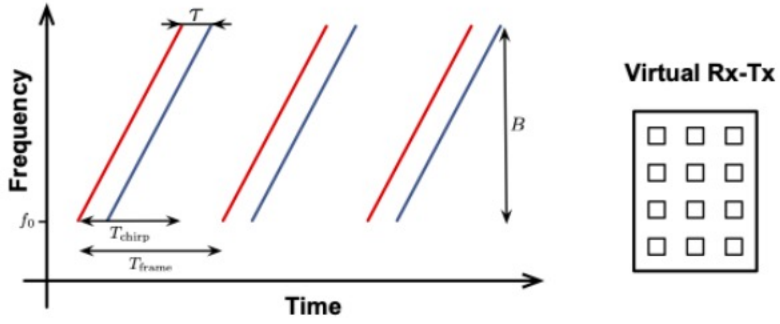
中国科学技术大学
University of Science and Technology of China

Our device



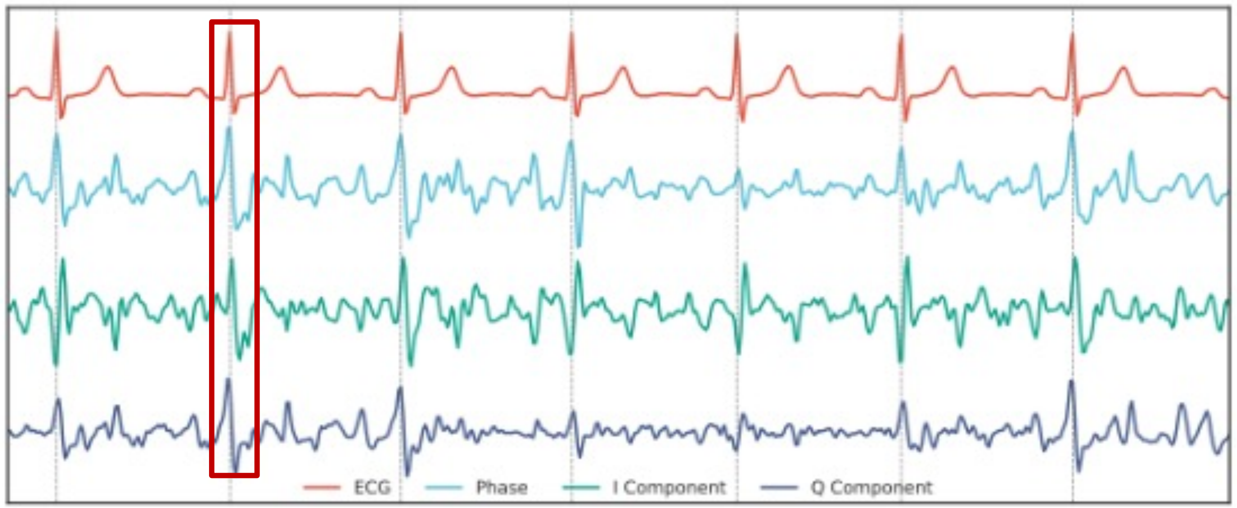
Costs $30

Single Chirp Configuration



frame rage: 100
data: 15 x 12 I/Q
rate: 576 Kbps

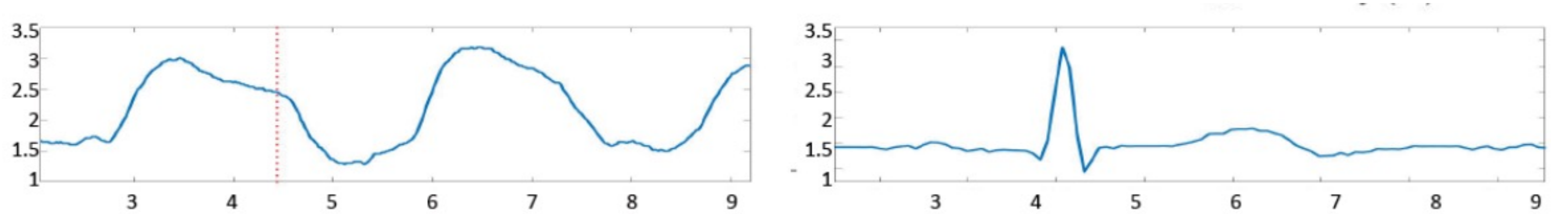
Example: R peak is aligned with signals



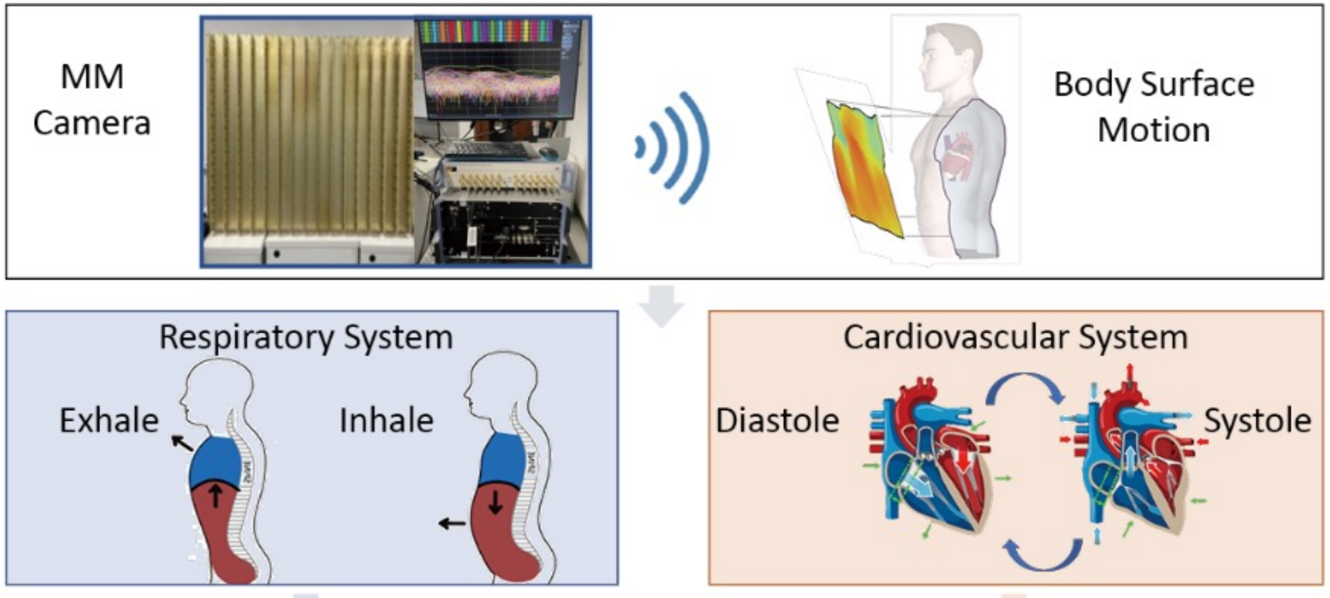
RF-Based Physiological Imaging

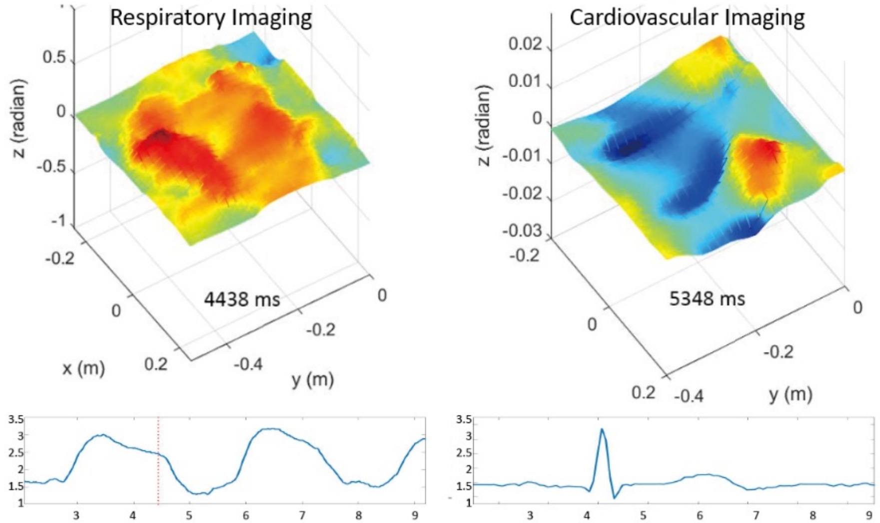
中国科学技术大学
University of Science and Technology of China

Problem: previous work usually estimates the 1d vital signal



Question: can we image the detailed body surface motion with RF signals





If true, we can analyze finer details to infer cardiac states

Chen et al. MMCamera. *In* ***MobiCom*** *(Poster)* 2022.

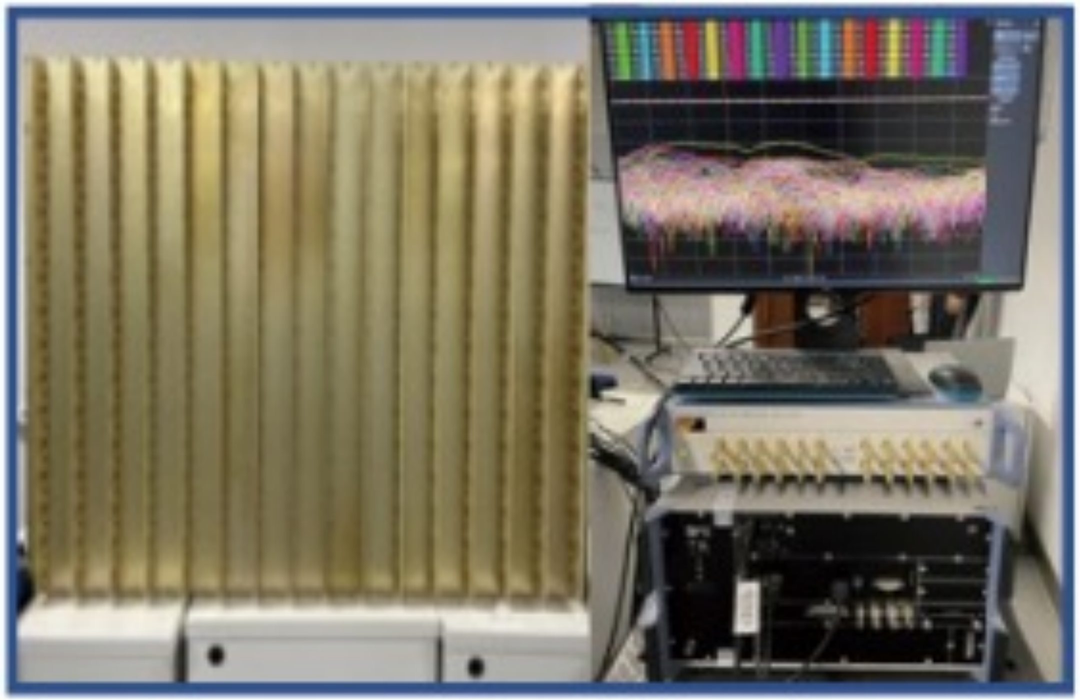
RF-Based Physiological Imaging

MMCamera Prototype Design

For concept validation purposes

Massive MIMO radio system

- 1.4Ghz bandwidth from 2.7 to 4.1Ghz
- 12 x 12 virtual planar array
- Aperture sizes: 75.72cm (h), 52.8cm (w)

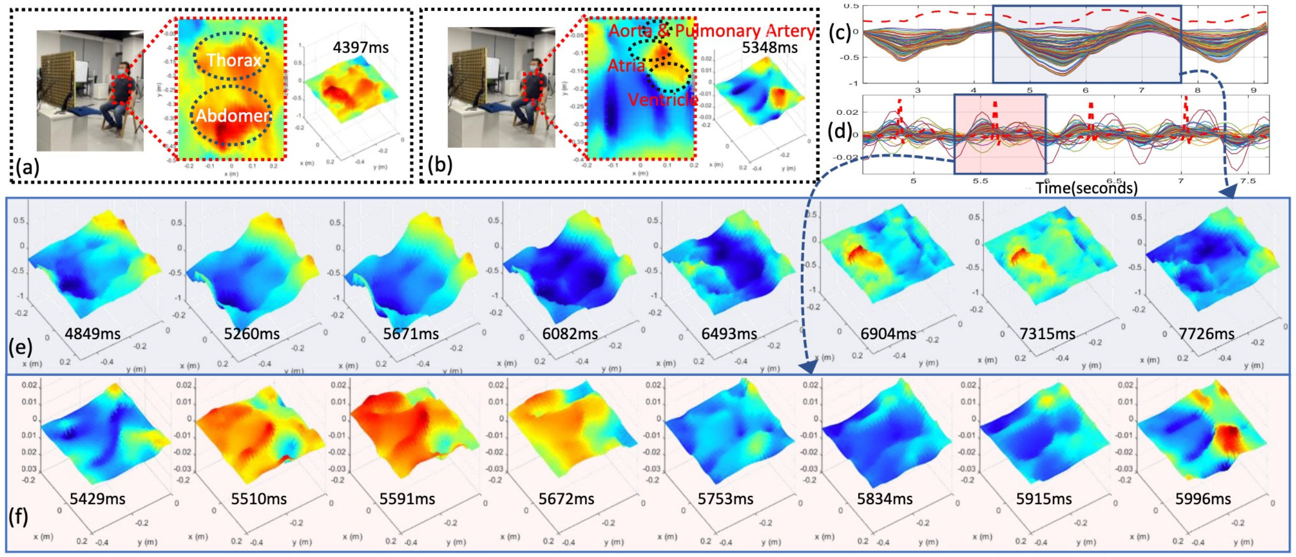


Chen et al. MMCamera. *In **MobiCom** (Poster)* 2022.

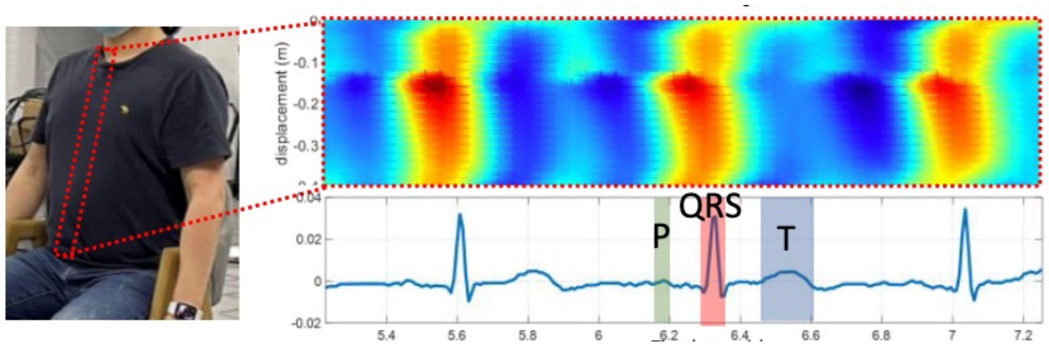
RF-Based Physiological Imaging

MMCamera Prototype Design

1. Back Projection: Reconstruct 3D voxel reflections using back projection
2. Multipath Elimination: 1D CFAR to detect direct reflections
3. Surface Projection: Project the closest reflections onto the imaging plane.
4. Noise Filtering: Use median filtering to smooth out noise and ensure continuity.
5. Motion Imaging: Extract phase variance to generate dynamic imaging.



Align with ECG



strong ventricular movements
signal with high amplitude

RF-Based Physiological Imaging

Conclusion

RF signals can capture the detailed movements associated with cardiac activity.

Questions

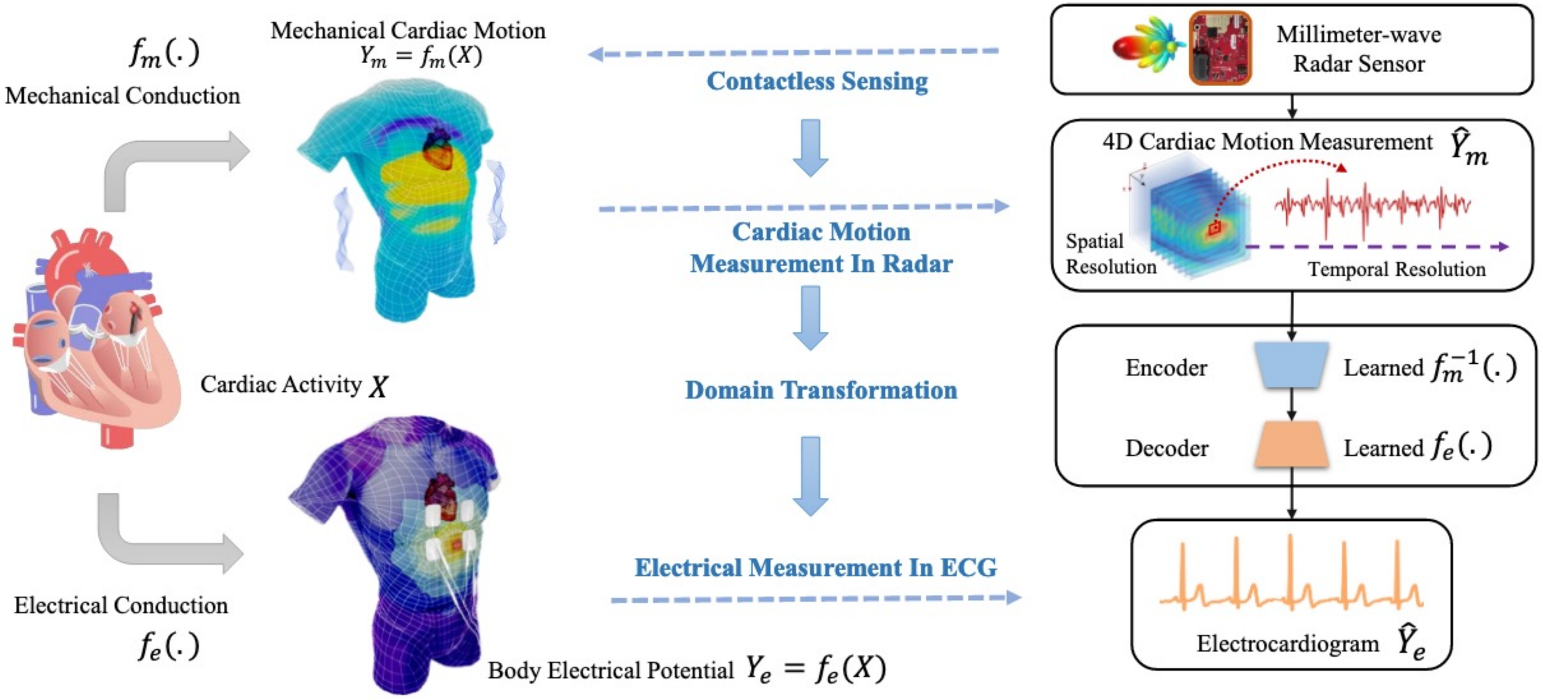
Is it possible to recover the ECG from surface motion?
How can ECG be monitored with an off-the-shelf radar chip?

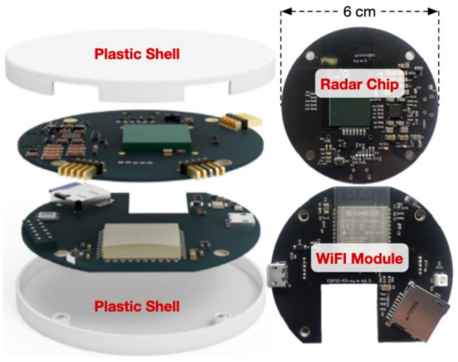
The data-driven approach using deep learning to address the limitations of radar performance

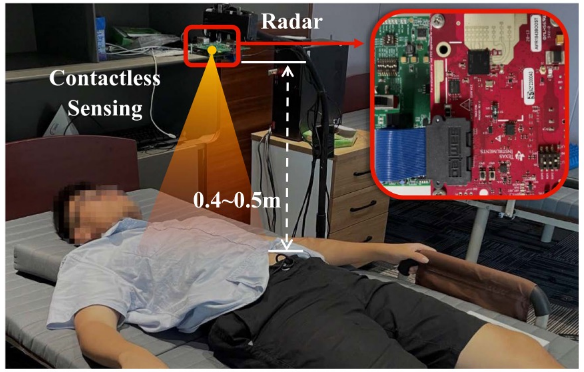
Chen et al. MMCamera. *In* ***MobiCom*** *(Poster)* 2022.

RF-Based Electrocardiogram Monitoring

Problem







TI evaluation board (same chip)

Can we monitor Electrocardiogram (ECG) with RF signal?

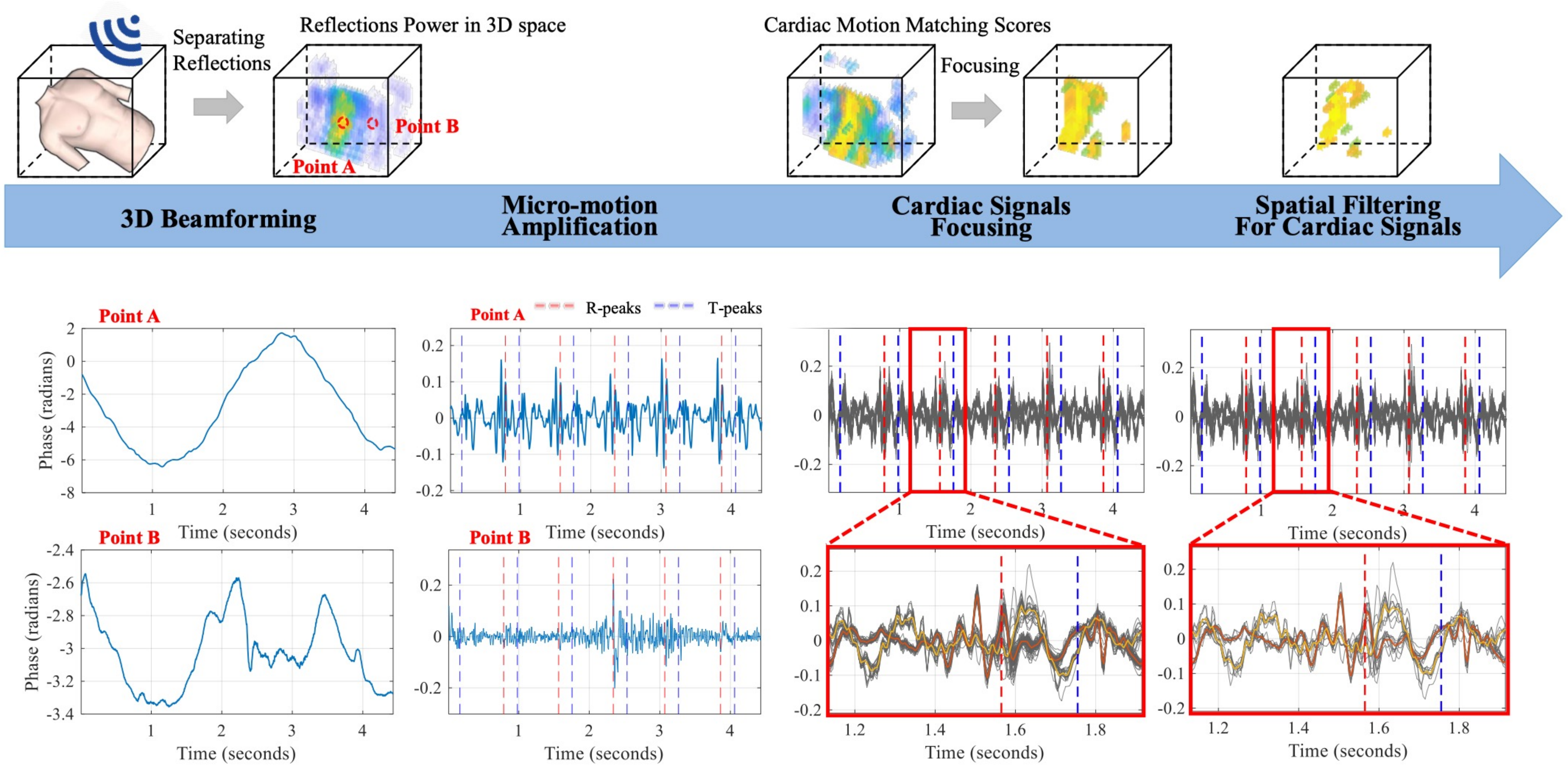
- Related works: breath and heartbeat (Adib et al. 2015), RR interval (Dong et al. 2020)
- Fundamental: Mechano-electrical Coupling (Bers 2002)

Chen et al. ***TMC*** 2022.

RF-Based Electrocardiogram Monitoring

Signal selection

Learning directly from raw data often fails to produce satisfactory results, even with larger datasets.



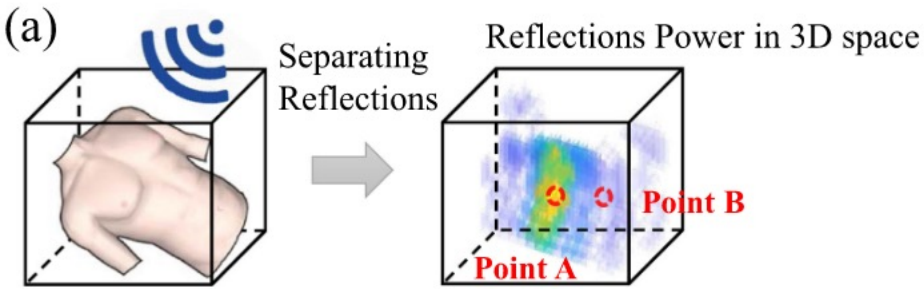
Chen et al. *TMC* 2022.

RF-Based Electrocardiogram Monitoring

Signal Pre-processing

1. Beamforming: reflections coming from different are separated

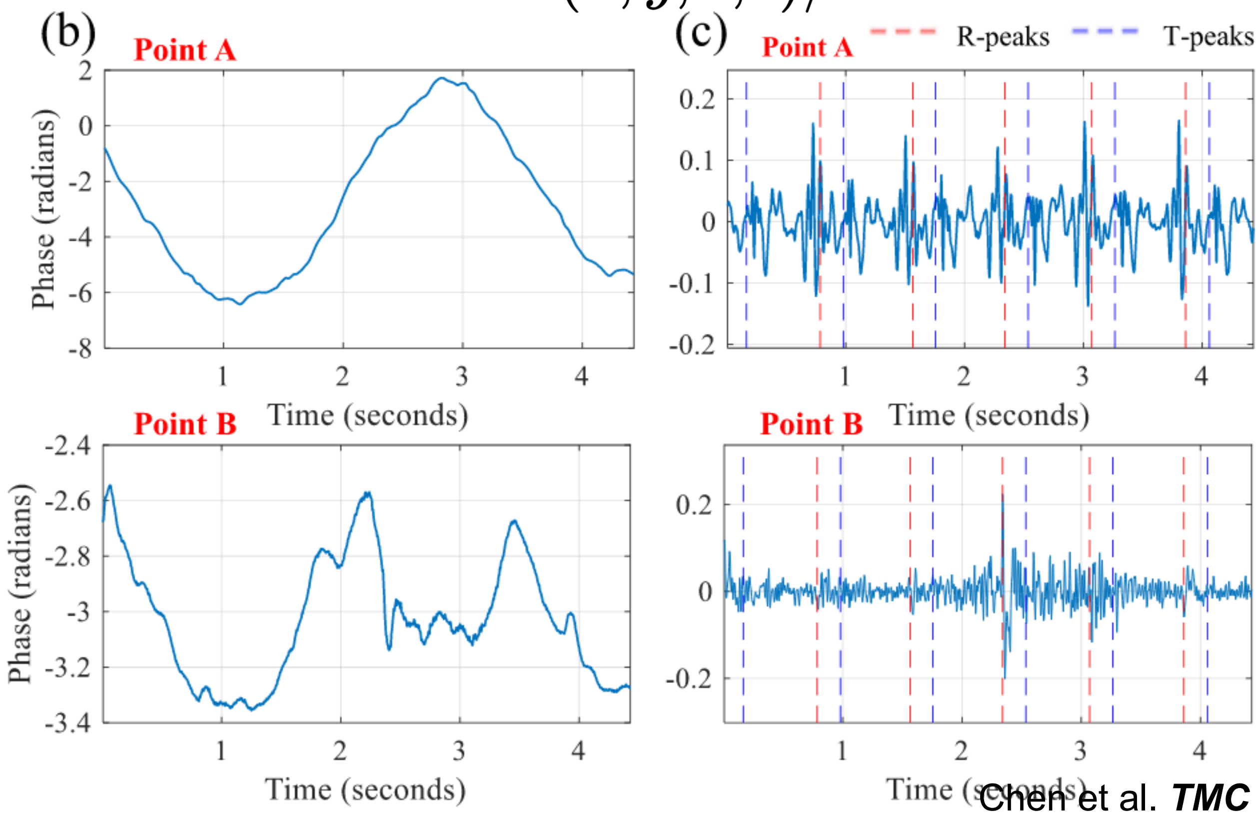
$$S(x, y, z, t) = \sum_{n=1}^{N} \sum_{t=1}^{T} y_{n,t} e^{j2\pi \frac{kr(x,y,z,n)}{c} t} e^{j2\pi \frac{r(x,y,z,n)}{\lambda}}$$



2. Micro-motion Amplification

RF signals are dominated by breathing

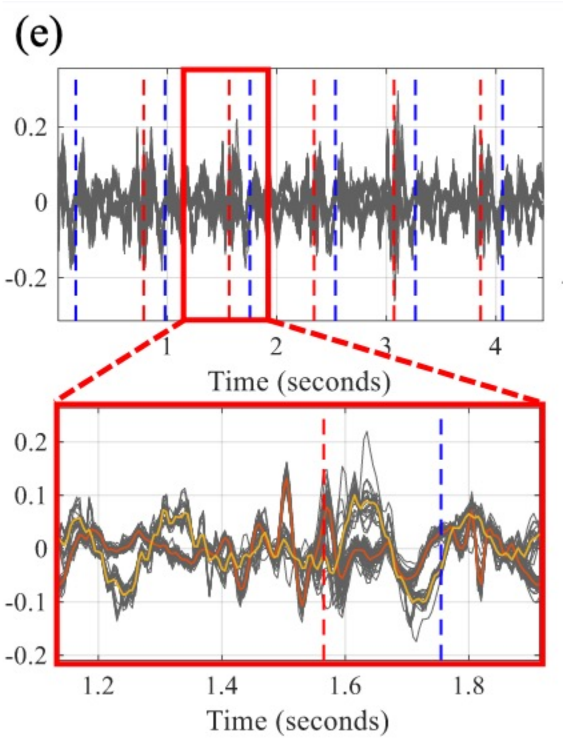
$$\mathrm{d}^2 S(x, y, z, t)/\mathrm{d}t^2$$



Chen et al. *TMC* 2022.

RF-Based Electrocardiogram Monitoring

Cardiac Signals Focusing



Motivation:

3D beamforming is redundant

Focus on the signals with cardiac cycle

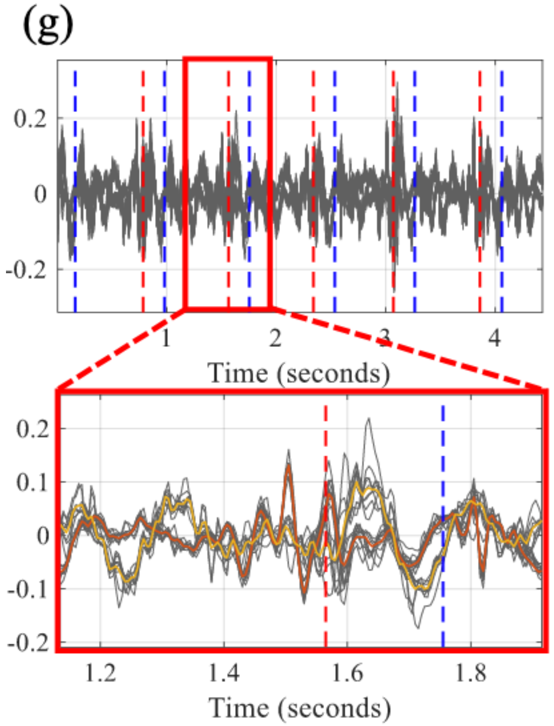
DTW thresholding

$$\mathrm{P}(S) = \frac{1}{l} \sum_{\bar{S}_i^{\dagger} \in \bar{S}^{\dagger}} \mathrm{DTW}\left(\bar{T}^{\dagger}, \bar{S}_i^{\dagger}\right) > \mathrm{Threshold}$$

Chen et al. ***TMC*** 2022.

RF-Based Electrocardiogram Monitoring

Spatial Filtering for Cardiac Signals



Motivation:

Cardiac signals are spread over the body surface

Similar motion trends with spatially nearby signals

K-means clustering

$$J=\sum_{i=1}^{m}\sum_{k=1}^{K}\left(w_{i,k}\rho_s\|s_i-\mu_k\|^2+w_{i,k}\rho_l\|l_i-l_{\mu_k}\|^2\right),$$

$$\mu_k=\frac{\sum_{i=1}^{m} w_{i,k}p_i s_i}{\sum_{i=1}^{m} w_{i,k}p_i},\quad l_{\mu_k}=\frac{\sum_{i=1}^{m} w_{i,k}p_i l_i}{\sum_{i=1}^{m} w_{i,k}p_i}$$

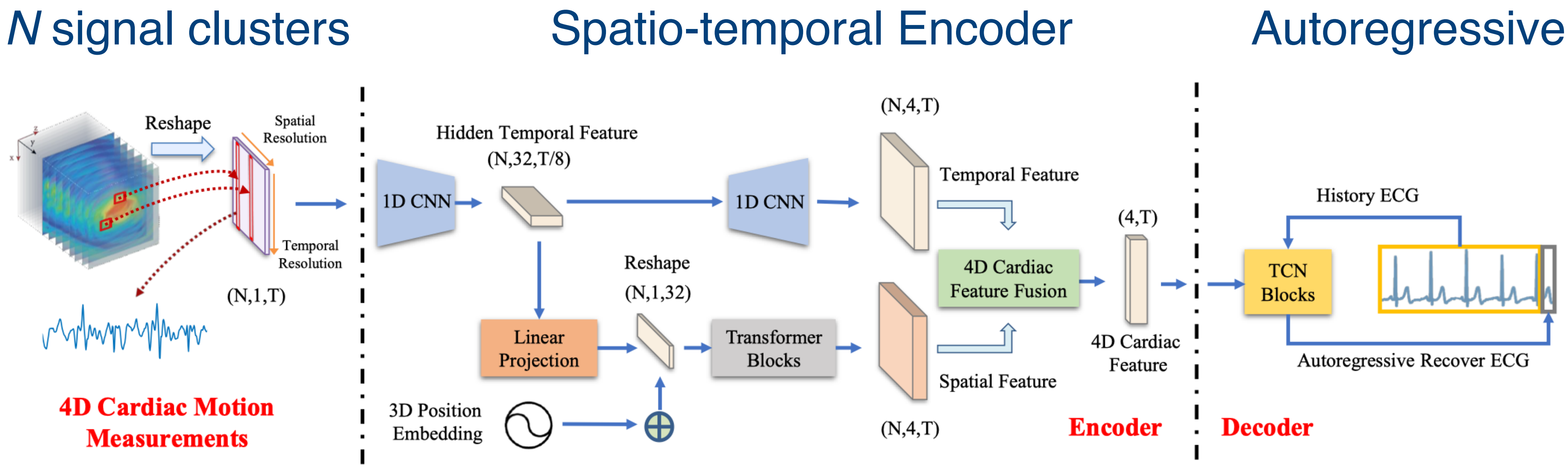
Chen et al. ***TMC*** 2022.

RF-Based Electrocardiogram Monitoring

中国科学技术大学
University of Science and Technology of China

Overall Framework

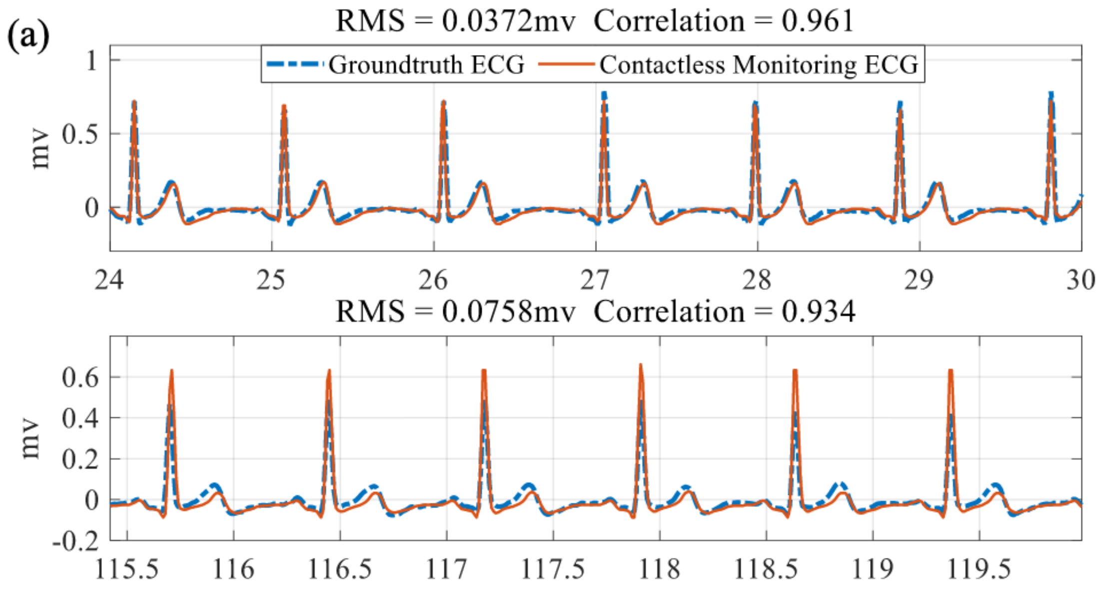
1. Signal selection
2. Spatio-temporal features
3. Autoregressive ECG restruction



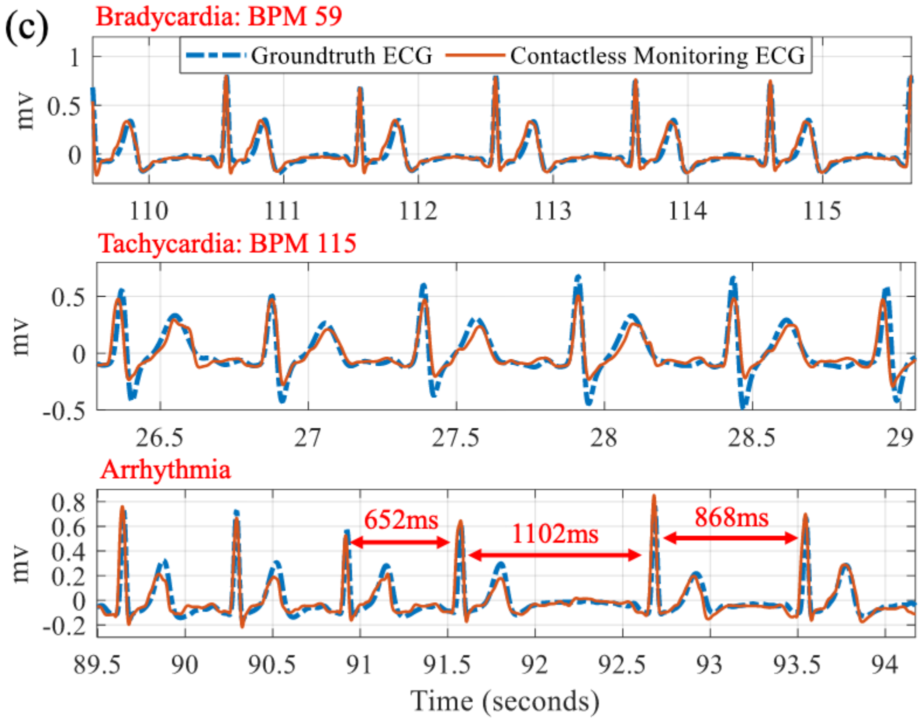
Chen et al. *TMC* 2022.

RF-Based Electrocardiogram Monitoring

ECG Reconstruction



Regular ECG



Irregular ECG

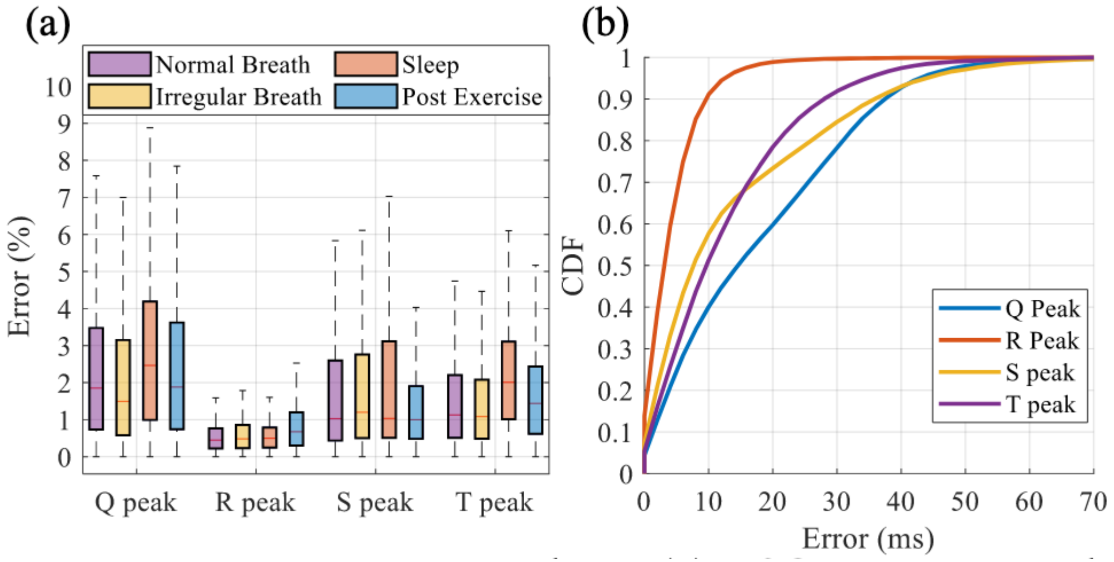
Accurate ECG reconstruction even in irregular case.

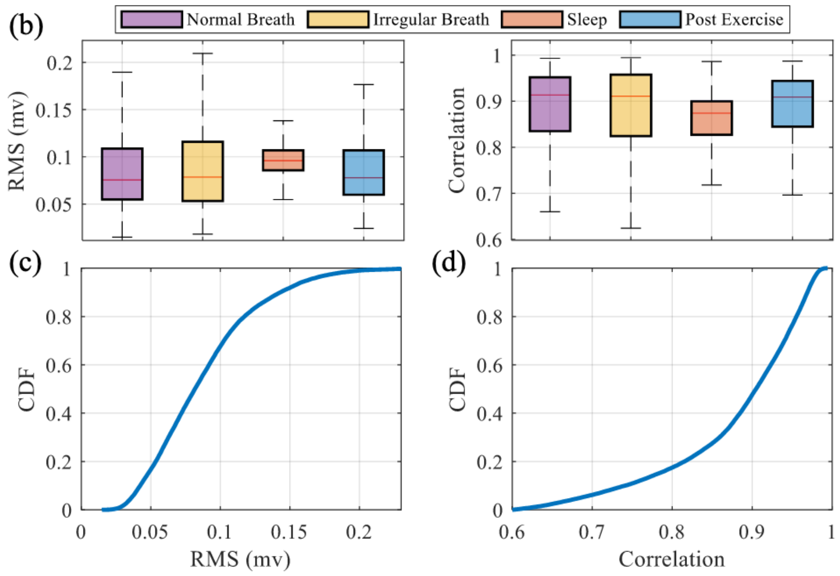
Chen et al. ***TMC*** 2022.

RF-Based Electrocardiogram Monitoring

Quantitative Results

Media: 14ms (Q), 3ms (R), 8ms (S), 10ms (T)





ECG Events Timing Accuracy

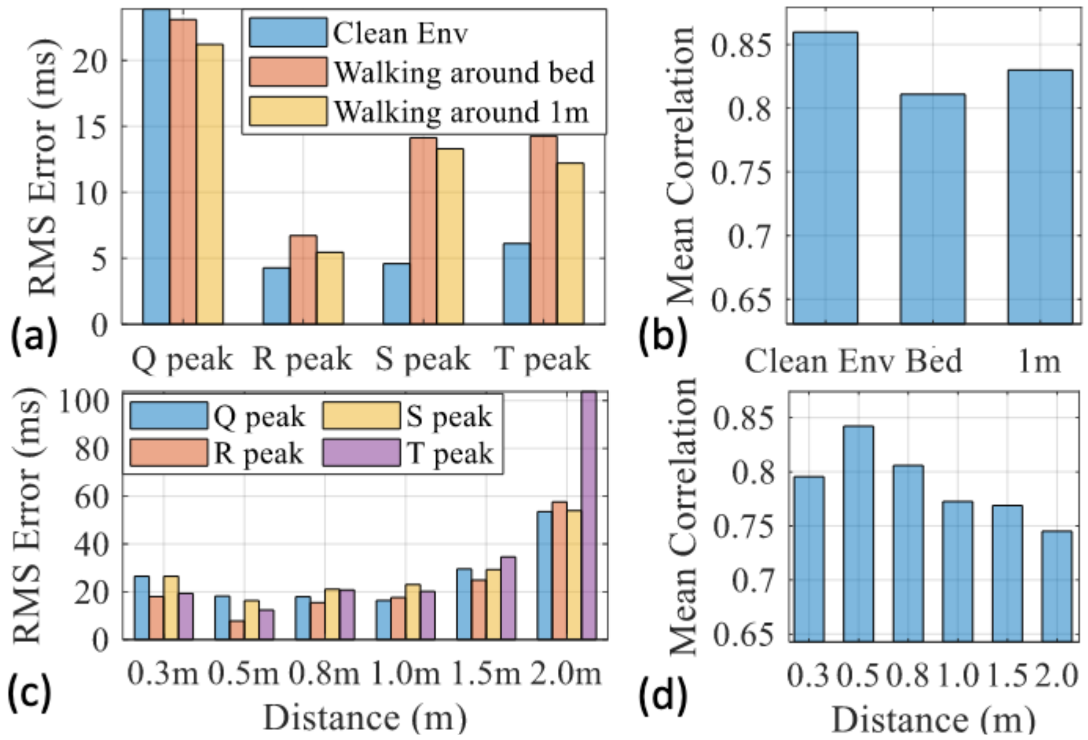
Correlation and MSE

Accurate Q, R, S, T peaks and morphology.

Chen et al. ***TMC*** 2022.

RF-Based Electrocardiogram Monitoring

Daily life usage



1. Environment robustness
2. Performance drop 10% at 2m

Different environments and different distance

RF-Based Electrocardiogram Monitoring

Conclusion

First RF-based system for contactless ECG monitoring

- High-accuracy cardiac mechanical activity sensing.
- Expands radar sensing capabilities.

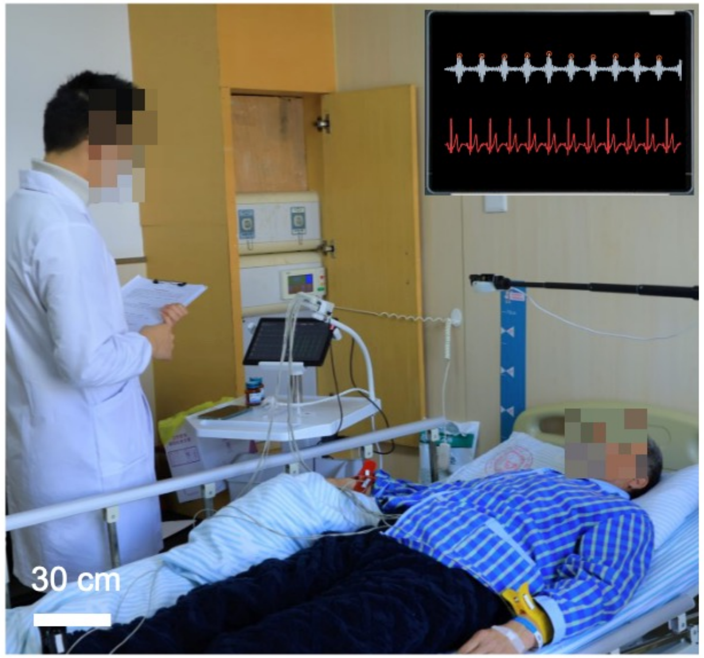
Further Challenge

Not validated on a large scale or in a clinical environment.

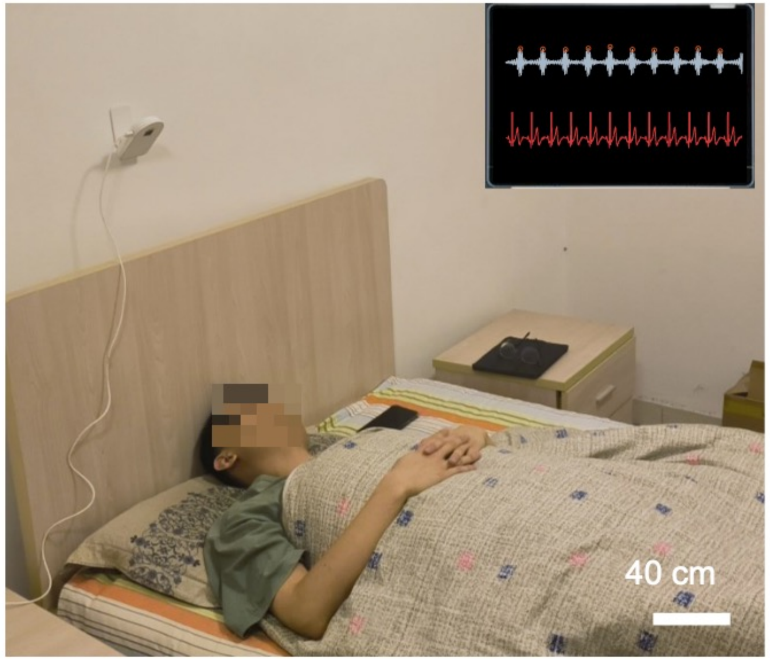
Chen et al. *TMC* 2022.

Long-term Cardiac Activity Monitoring

Long-term Cardiac Activity Monitoring



Clinical scenario



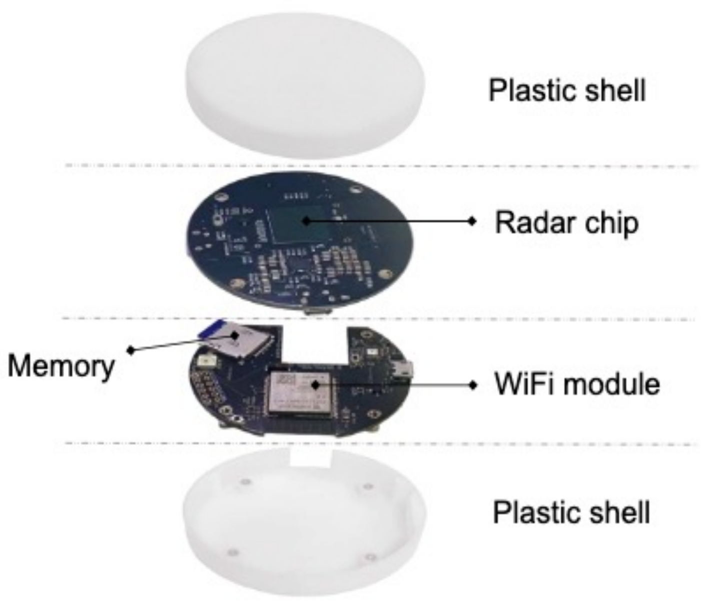
Daily life scenario

Goal: Long-term and continuous monitoring of cardiac activities

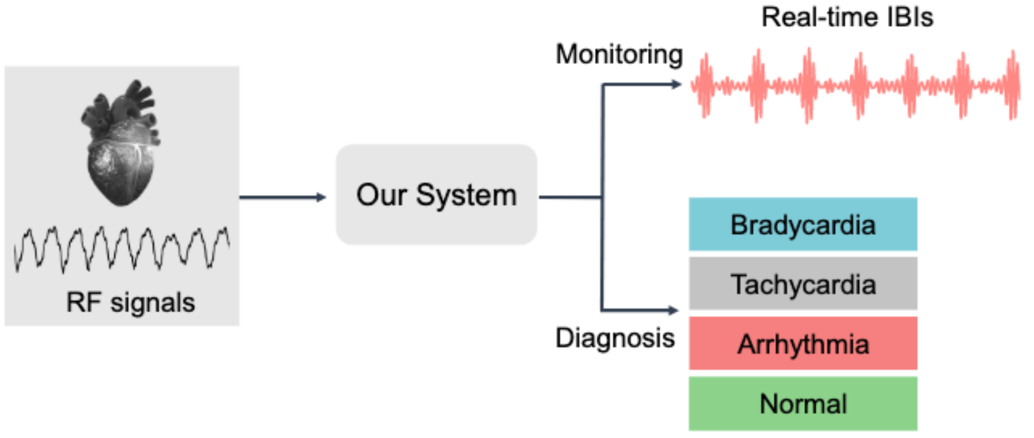
Long-term Cardiac Activity Monitoring

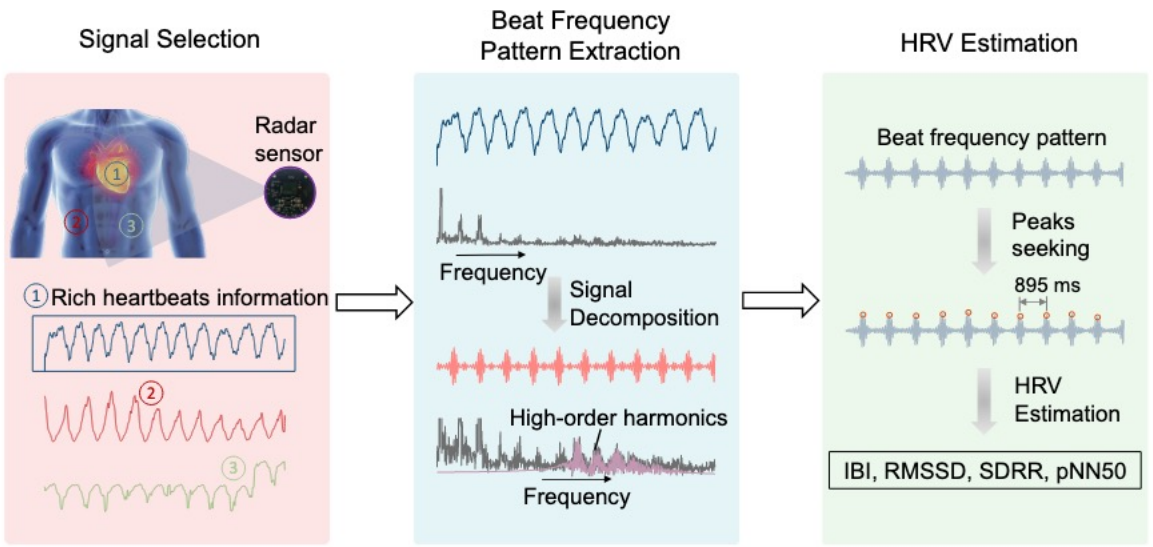
中国科学技术大学
University of Science and Technology of China

Proposed System



Our device





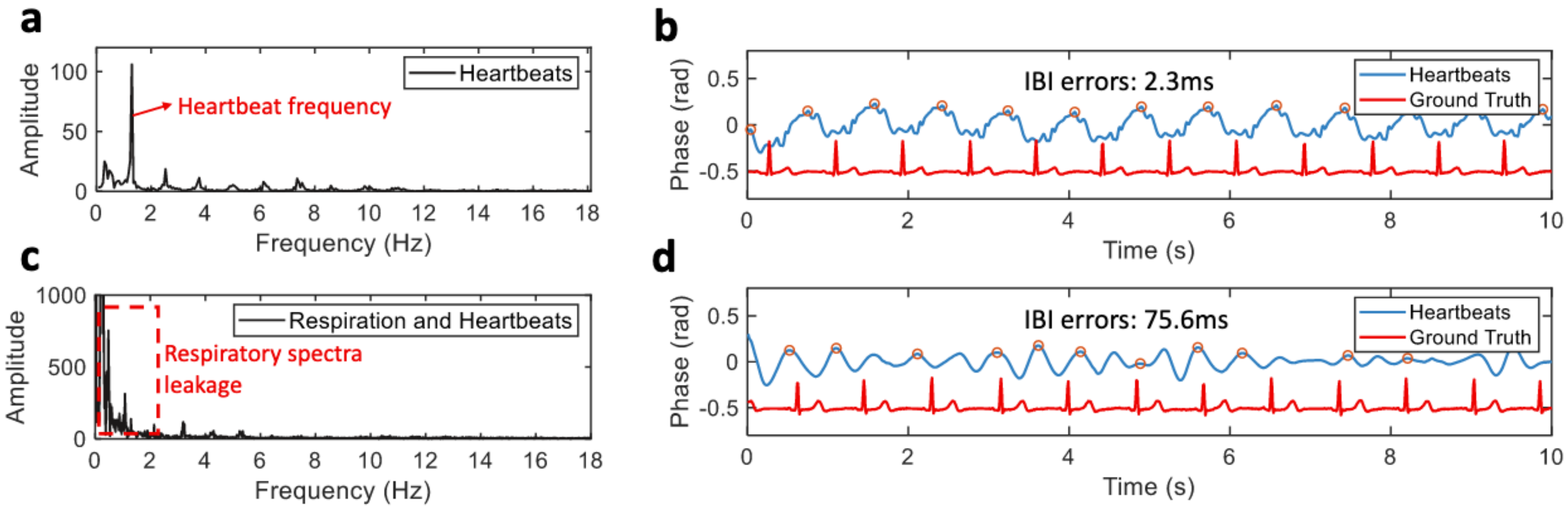
Heart Rate Variability / Inter-Beat Interval (IBI)

Long-term Cardiac Activity Monitoring

Heartbeat on Different conditions

Breath Holding

Normal Breath



Frequency

Phase

Zhang et al. ***Nature Communications*** 2024

RF-Based Cardiac Activity Monitoring

Periodic Signal

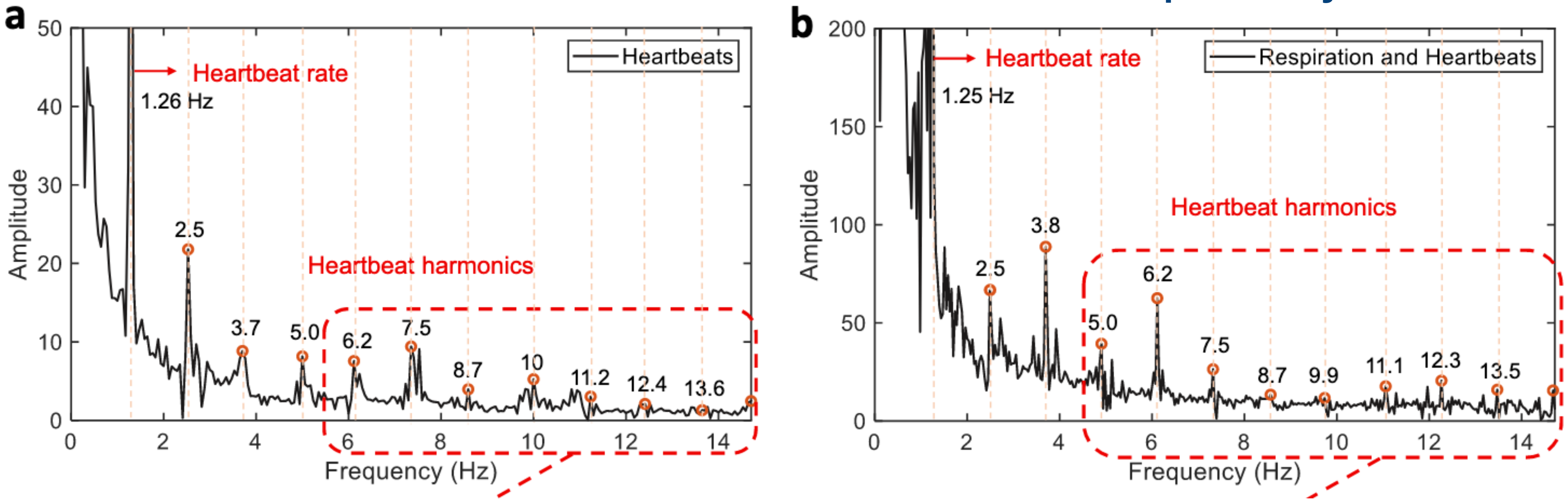
$$x(t) = \sum_{n=-\infty}^{\infty} c_n e^{jn\omega_0 t}$$

ω_0 : fundamental frequency

$n\omega_0$: harmonics

HR is clean

HR is corrupted by RR harmonics



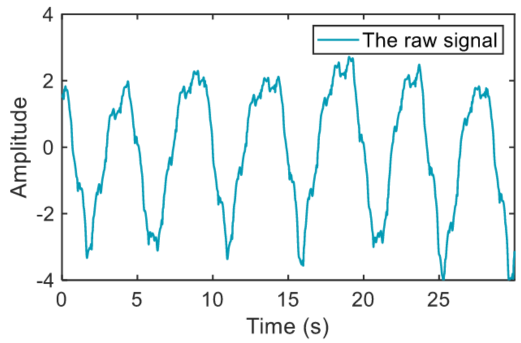
High order HR harmonics are much cleaner

Zhang et al. ***Nature Communications*** 2024

RF-Based Cardiac Activity Monitoring

- Method

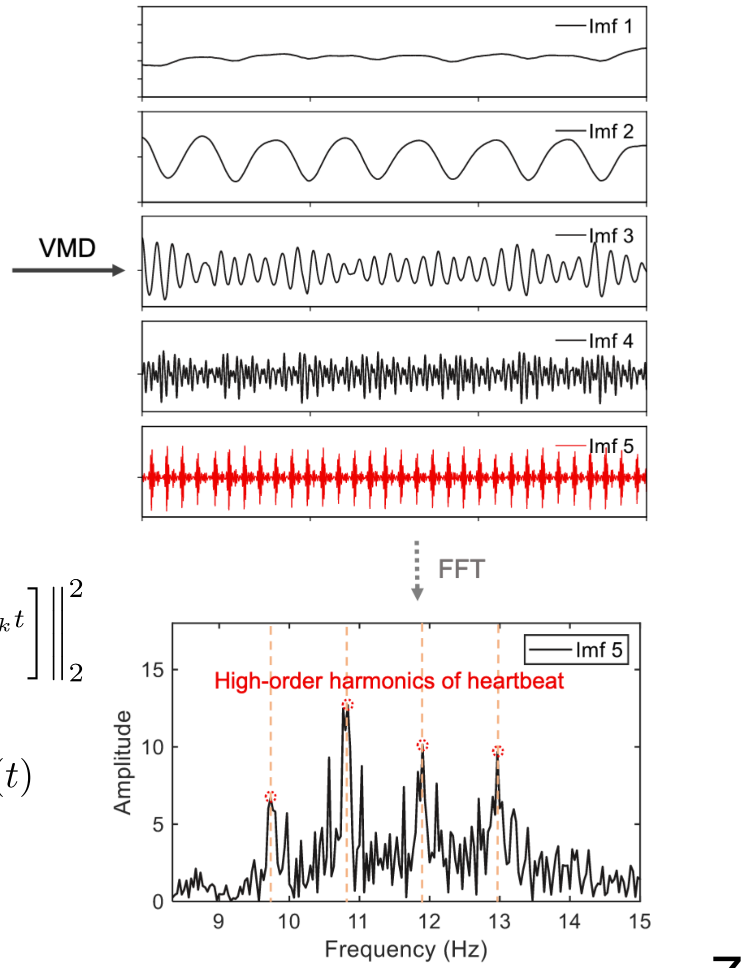
1. VMD Decomposition



$$\min_{\{u_k\},\{\omega_k\}} \sum_{k=1}^{K} \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_k(t) \cdot e^{-j\omega_k t} \right] \right\|_2^2$$

$$\sum_{k=1}^{K} u_k(t) = x(t)$$

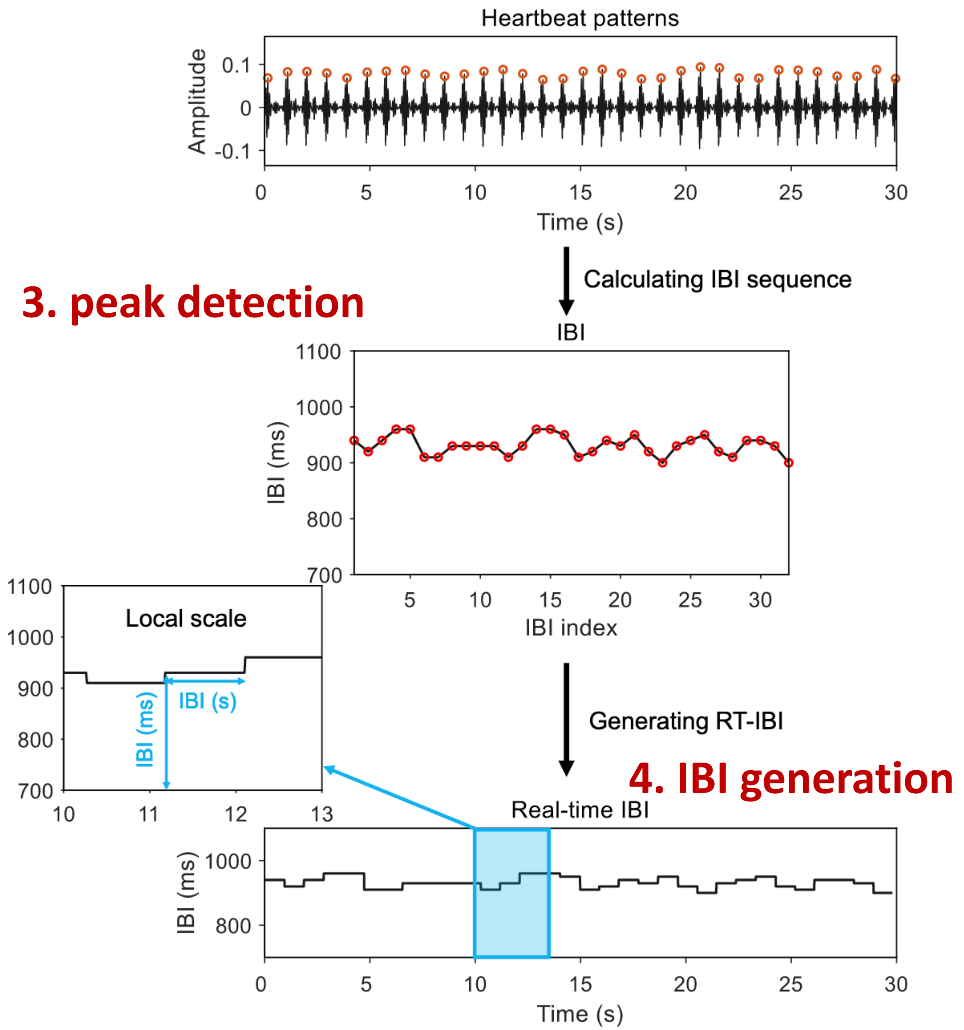
intrinsic mode functions (IMFs)



2. selected IMFs

3. peak detection

4. IBI generation

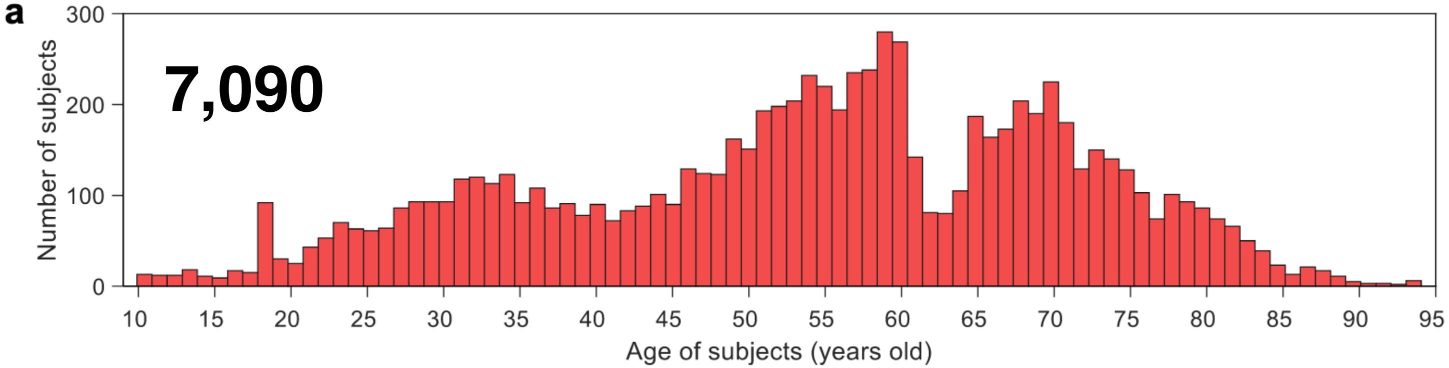


Zhang et al. ***Nature Communications*** 2024

RF-Based Cardiac Activity Monitoring

- Dataset

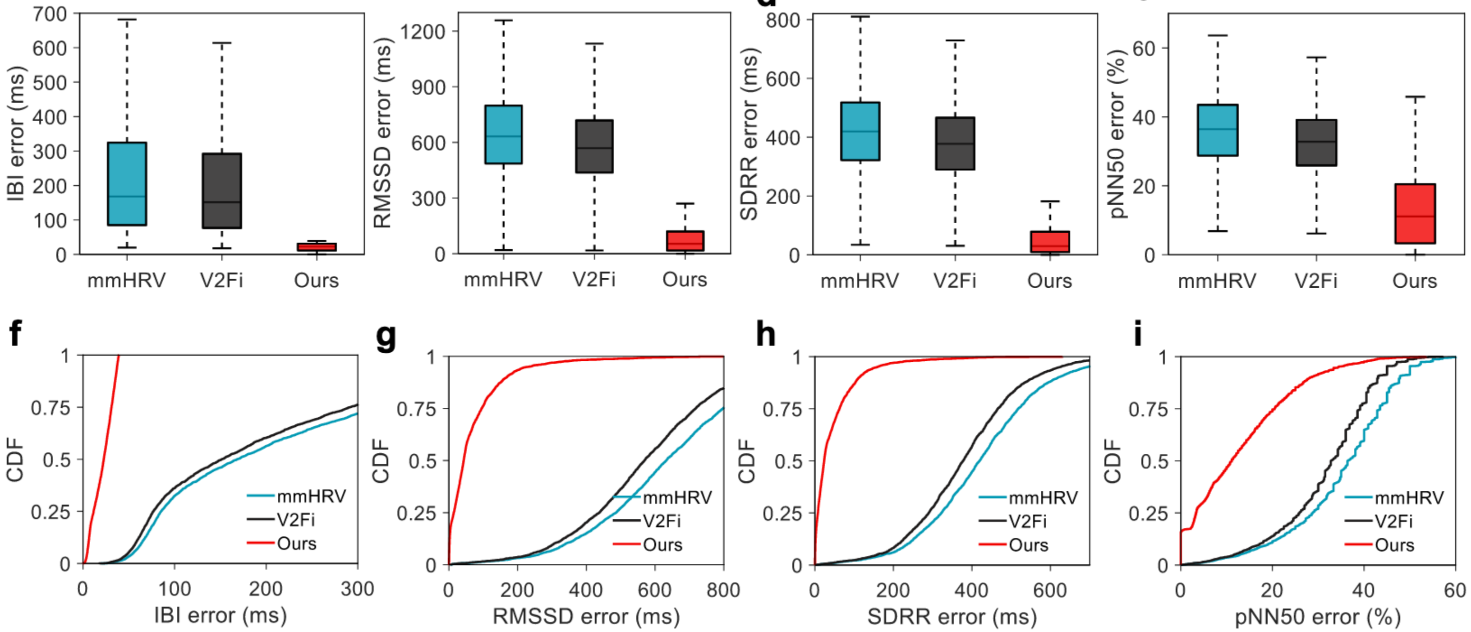
7,090



- Performance

26.1 ms 53.8 ms 29.4 ms 11.1%

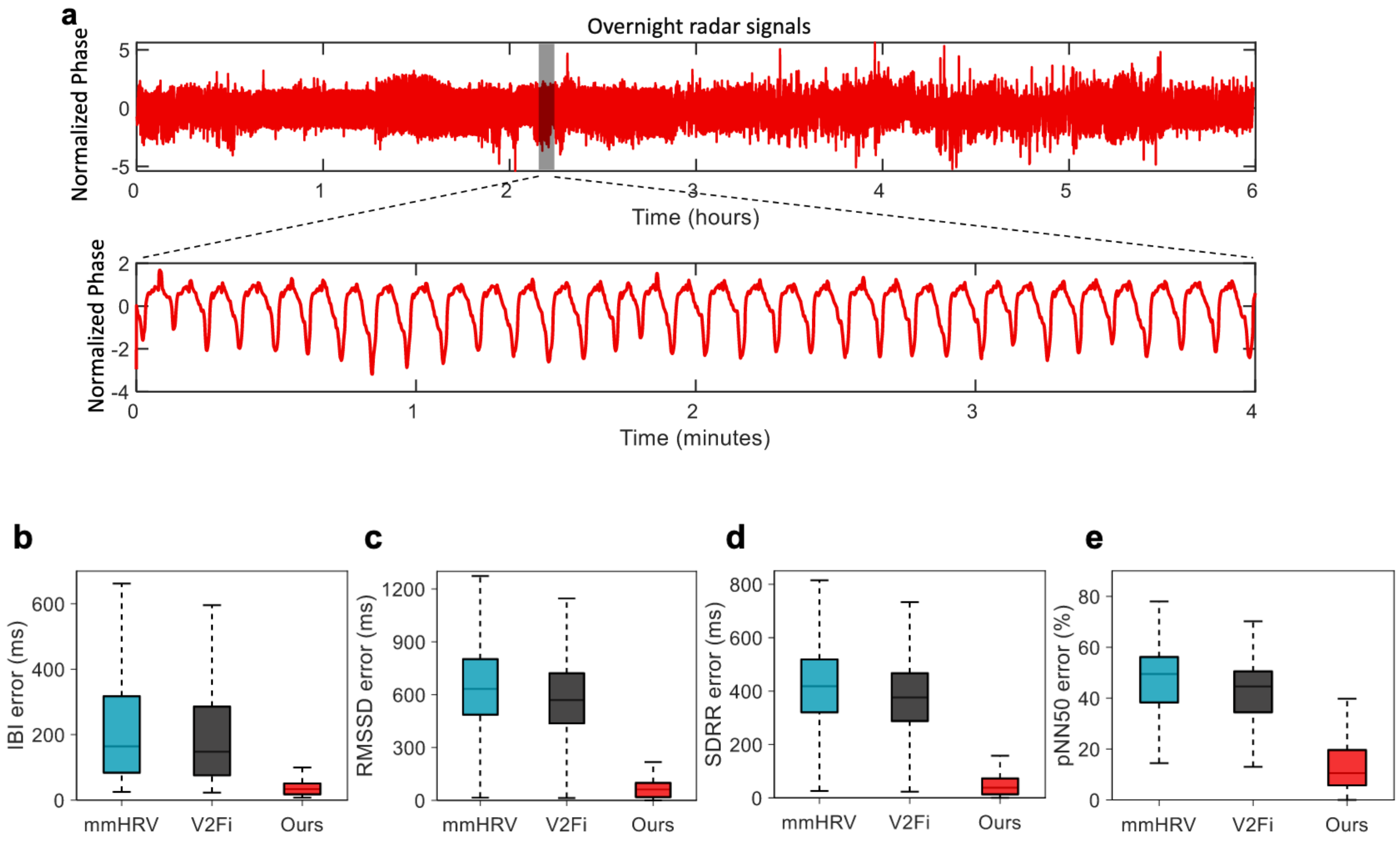
10x improvement



Zhang et al. ***Nature Communications*** 2024

RF-Based Cardiac Activity Monitoring

- Over-night Performance



Zhang et al. ***Nature Communications*** 2024

RF-Based Cardiac Activity Monitoring

Conclusion

We develop the first large-scale clinical-level cardiac monitoring system

- *High IBI accuracy*: Our system achieves superior accuracy in inter-beat interval (IBI) measurements, ensuring precise cardiac monitoring.
- *Works robustly in daily life*: It performs reliably across various real-world scenarios, including long-term and overnight monitoring.

Ongoing Research for Health Equity

- 12-lead ECG reconstruction.
- Disease Monitoring.
- ....

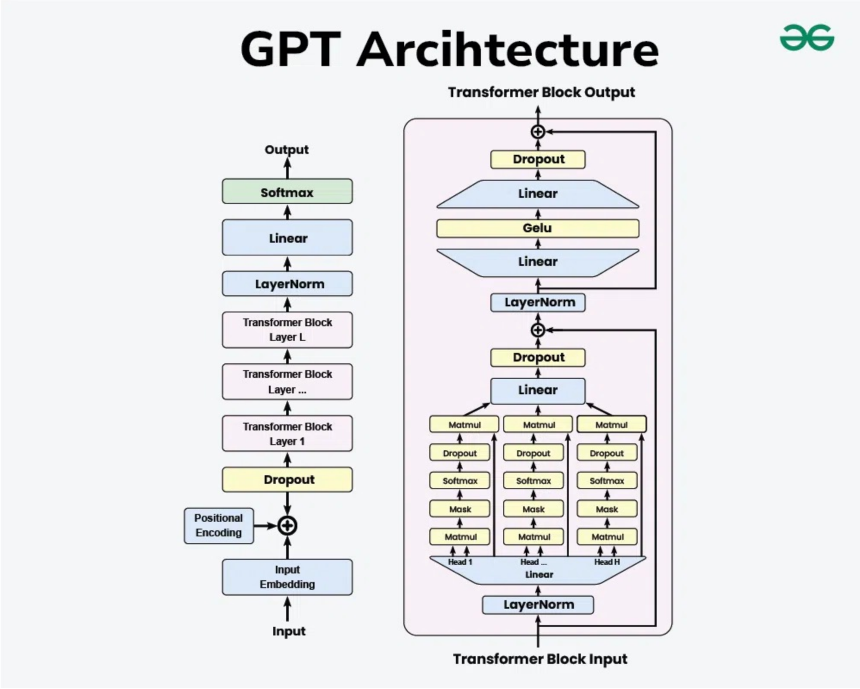
Contents

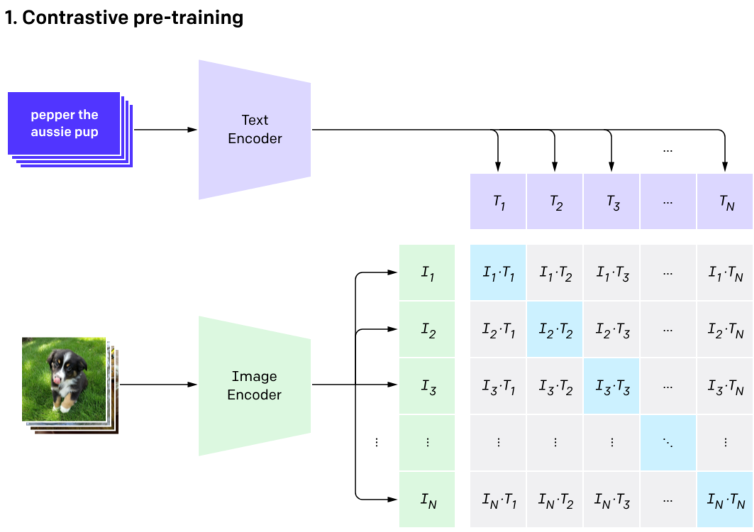
中国科学技术大学 University of Science and Technology of China

RF-Based Self-supervised Learning

Self-supervised learning leverages ***unlabeled data*** *to predict parts of the input, generating representations that can be* ***fine-tuned for downstream tasks*** *like classification, detection, and segmentation.*

Can we design SSL methods for the RF data



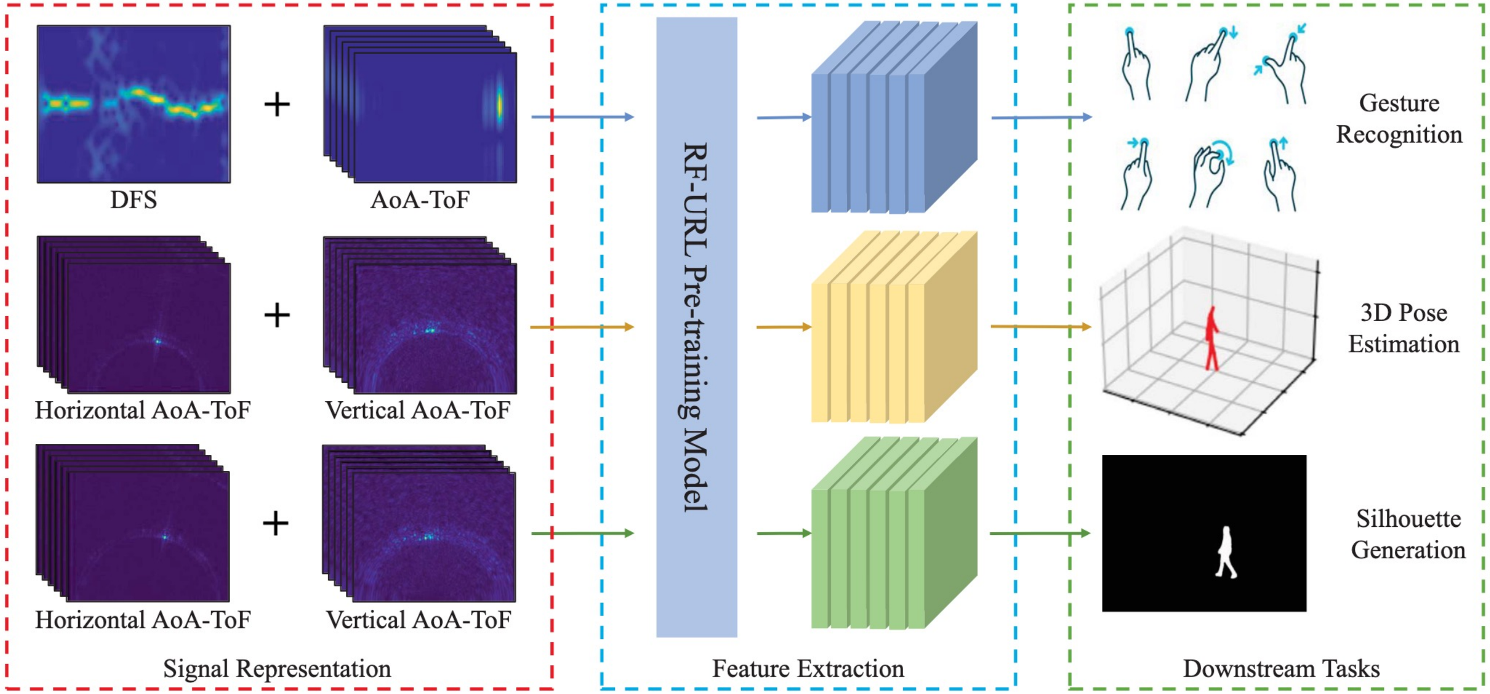


SimCLR **GPT** **CLIP**

RF-Based Self-supervised Learning

中国科学技术大学
University of Science and Technology of China

- Why SSL



Challenges
- non-intuitive
- hard to annotate
- much sparse

Easy to collect data

- Two devices (WiFi and Radar)
- Three tasks (gesture, pose and silhouette)

RF-Based Self-supervised Learning

中国科学技术大学
University of Science and Technology of China

- *Contrastive Learning*

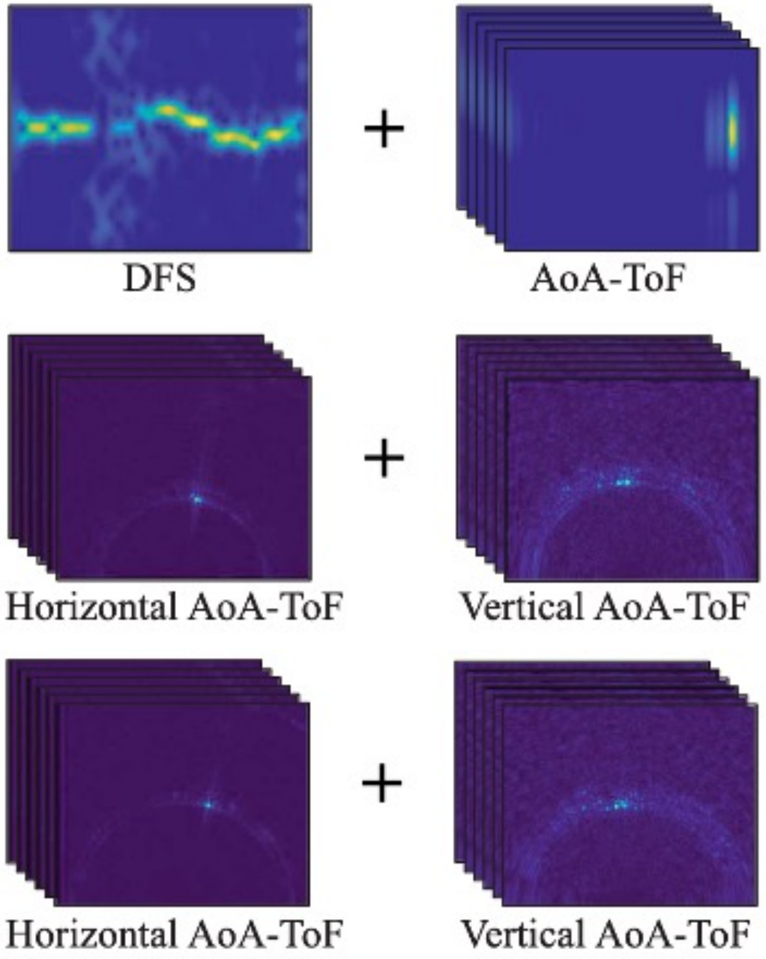
  Different signal representations as data augmentation

- *Masked Autoencoder*

  Eliminating the need for complex data augmentations

- *RF-aware Augmentation*

  Approach leveraging RF characteristics

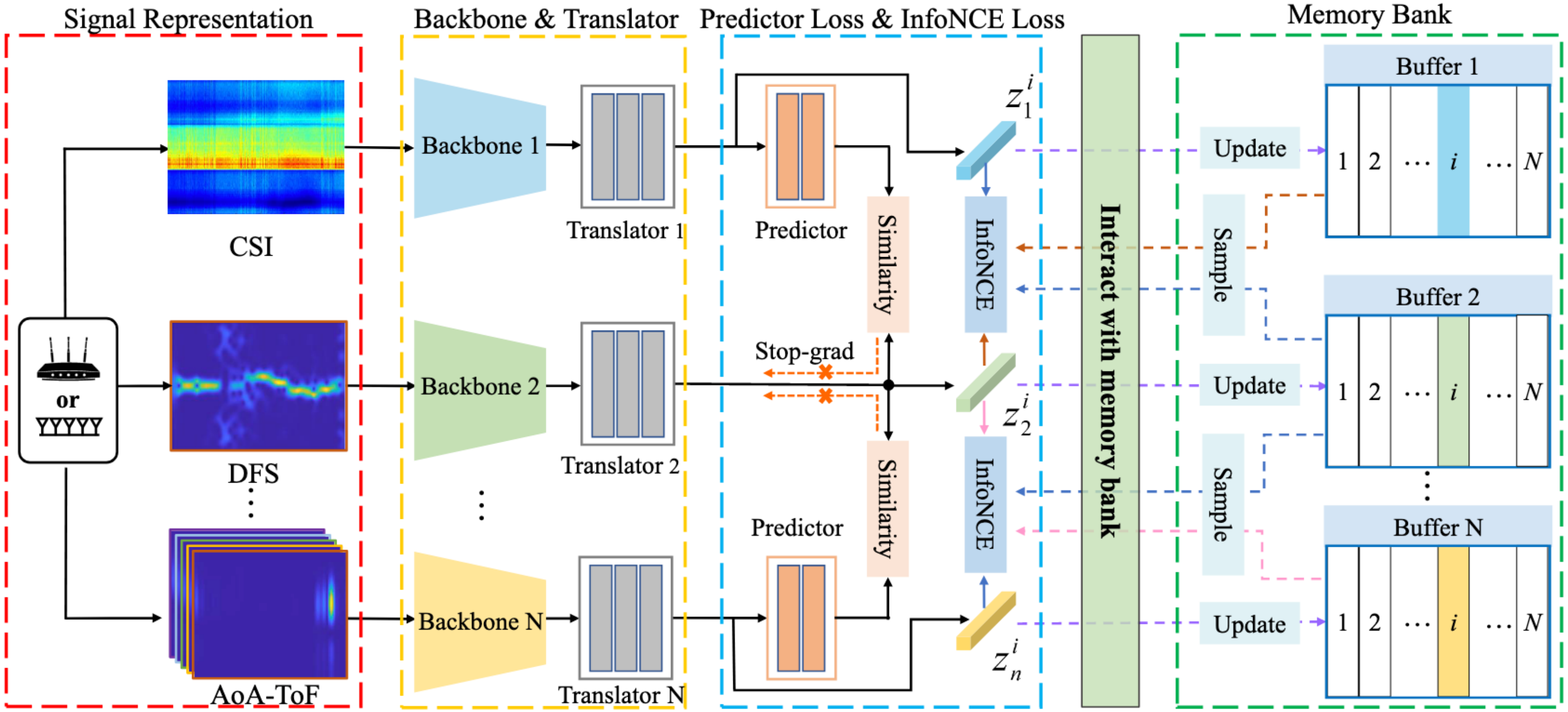


Signal Representation

Self-supervised Contrastive Learning

中国科学技术大学
University of Science and Technology of China

RF-URL (MobiCom’ 22)

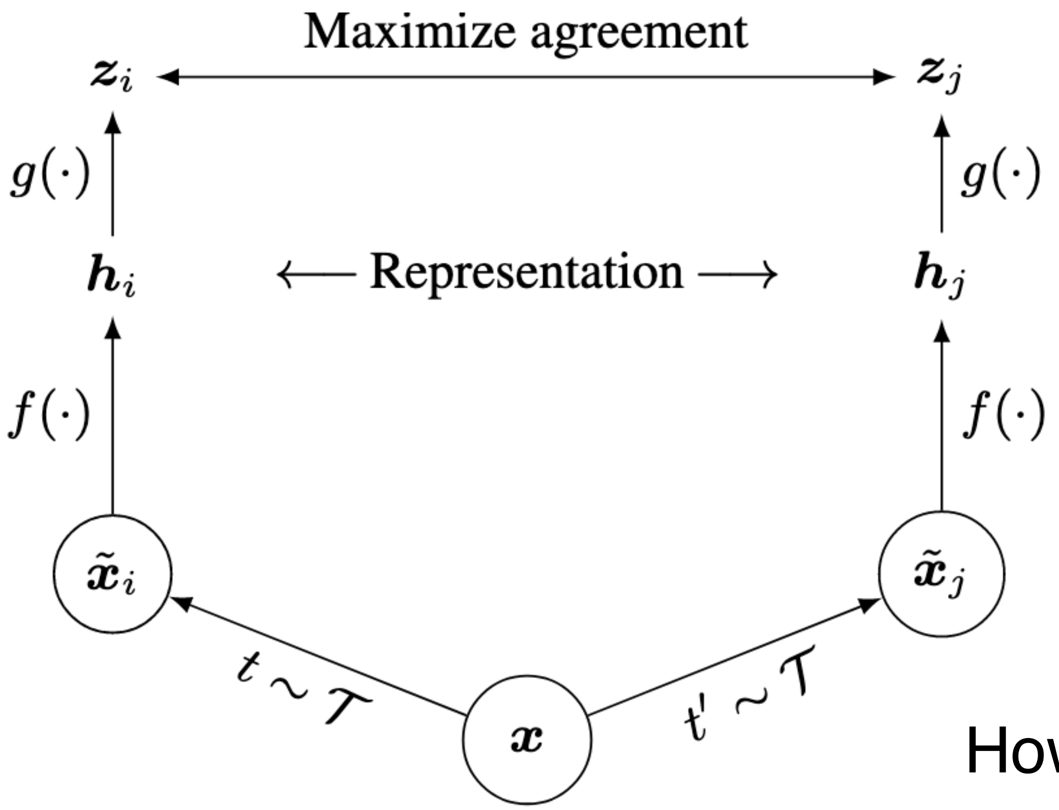


First work to do SSL for RF sensing tasks

Song et al. RF-URL In ***MobiCom*** 2022

Self-supervised Contrastive Learning

Contrastive learning



(Chen et al. 2020)

Contrastive learning aims to group similar samples (**positive**) closer and diverse samples (**negative**) far from each other.

How to create positive samples for RF signals?

RF-Based Self-supervised Learning

Data augmentations

Image augmentations:

- Crop and resize
- Color distort
- Gaussian noise
- Cutout
- ...

Augmentations for natural images may not suitable for RF signals

RF Signal representations:

- Doppler Frequency Shift (DFS)
- Angle of Arrival (AoA)
- Time of Flight (ToF)
- Channel state information (CSI)
- ...

Different representations should contains the semantics

Song et al. RF-URL In ***MobiCom*** 2022

RF-Based Self-supervised Learning

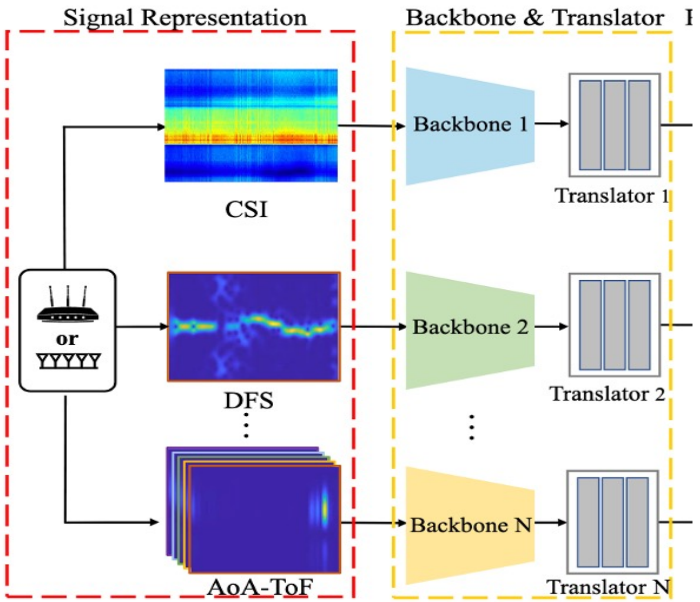
Model design

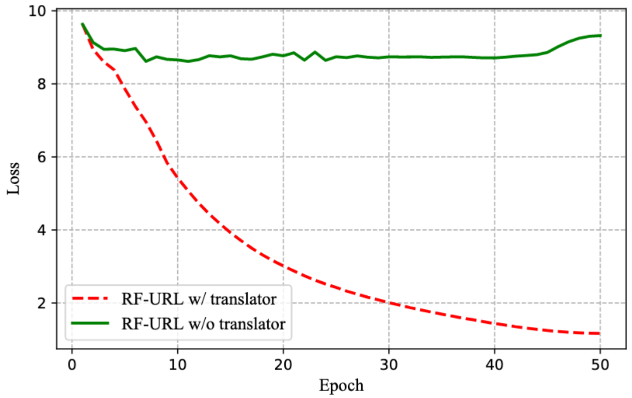
1. Backbone Encoder

Multi-branch backbone network for different signal representations

2. **Translator**

A mediator to transform the different RF signals into a unified latent space





Song et al. RF-URL In ***MobiCom*** 2022

RF-Based Self-supervised Learning

Model design

3. Predictor

A small shared-weight neural network *h* with stop gradient (*sg*) interacted with different branches

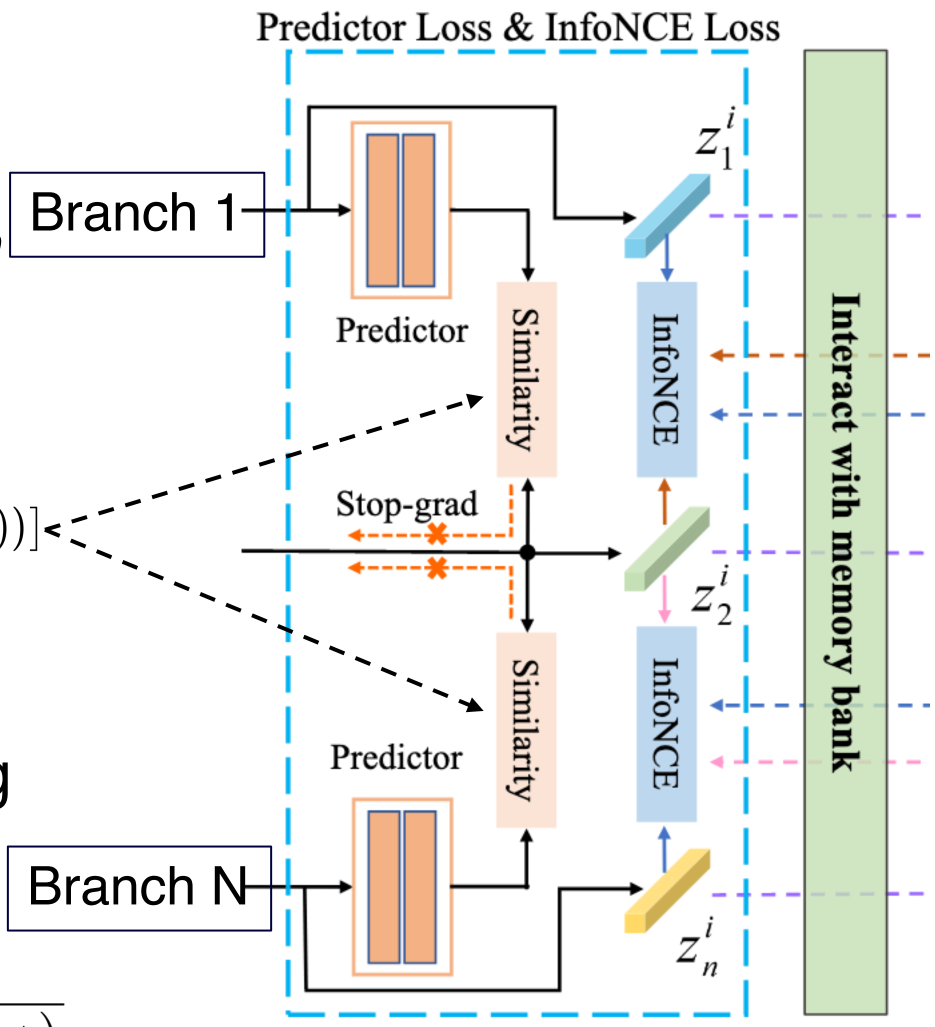
$$\mathcal{L}_P^i = \frac{1}{2(n-1)} \sum_{k=1}^{n-1} [\mathcal{D}(h(z_k^i), sg(z_{k+1}^i)) + \mathcal{D}(sg(z_k^i), h(z_{k+1}^i))]$$

4. InfoNCE

Preserve the shared information among different representations of the same signal

$$\mathcal{L}_c^i(z_k^i, z_{k+1}^i) = -\log \frac{\exp(s(z_k^i, z_{k+1}^i)/t)}{\exp(s(z_k^i, z_{k+1}^i)/t) + \sum_{j=1}^{K} \exp\left(s\left(z_k^i, z_{k+1}^j\right)/t\right)}$$

representations from other samples



Song et al. RF-URL In ***MobiCom*** 2022

RF-Based Self-supervised Learning

Model design

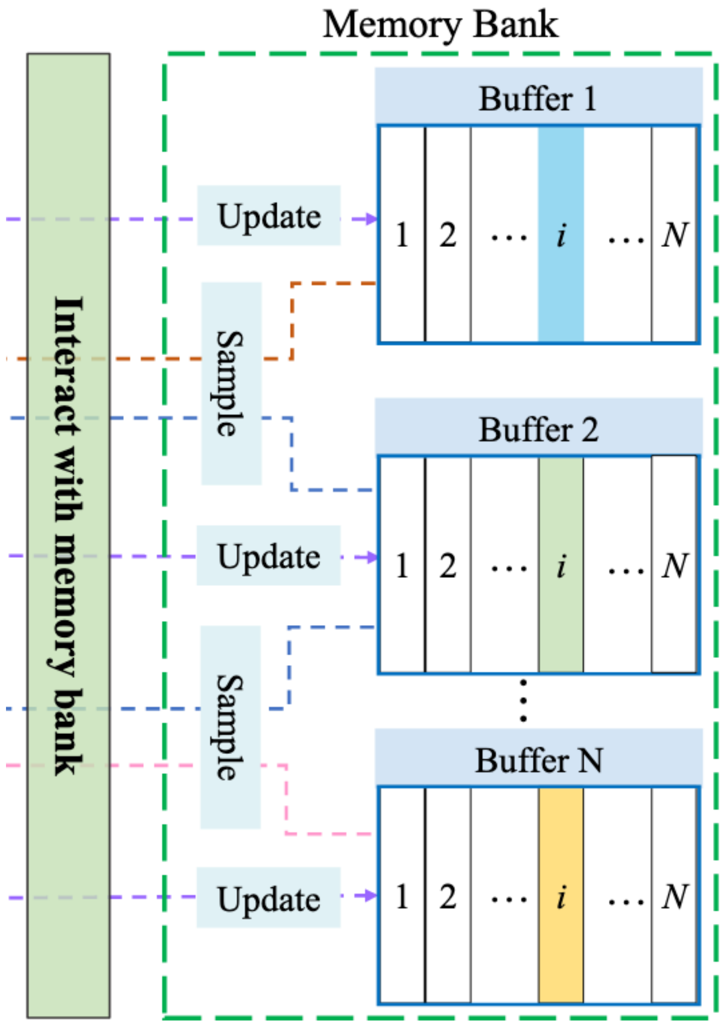
5. Memory bank

Stores representations of the training dataset used to negative samples in InfoNCE

update rule $z^{new} \leftarrow m \cdot z + (1-m) z^{\text{old}}$

6. Fine-tuning

Downstream tasks are fine-tuned with pre-trained models



Song et al. RF-URL In ***MobiCom*** 2022

RF-Based Self-supervised Learning

Representation quality

How effective are pre-trained features?

Gesture recognition

Model	Parameters	Random init	RF-URL (Frozen)
ResNet-17	11.18M	**46.539**	91.514
ResNet-35	21.85M	34.971	91.603
ResNet-50	25.55M	28.093	**92.407**
ResNet-101	44.54M	21.617	91.961
ResNet-152	60.19M	22.778	92.095

Silhouette Segmentation

Model	Parameters	Random init	RF-URL (Frozen)
RFSG-T	0.39M	0.225	0.552
RFSG-B	0.76M	0.239	0.556
RFSG-L	2.09M	0.248	0.536

Pose Estimation

Model	Parameters	Random init	RF-URL (Frozen)
RFP-T	2.66M	**198**	**68**
RFP-B	3.87M	206	71
RFP-L	7.09M	200	77

Pre-trained features extracted meaningful representations for RF signals.

Song et al. RF-URL In ***MobiCom*** 2022

RF-Based Self-supervised Learning

中国科学技术大学 University of Science and Technology of China

Detailed results

Gesture recognition

Model	Pre-training	100%labels	50%labels	10%labels	0%labels
ResNet-17	-	86.780	82.269	65.699	10.540
	Frozen	**91.514**	**89.549**	**82.314**	10.808
	Fine-tune	91.201	84.591	63.510	-
ResNet-50	-	89.013	84.815	64.448	11.121
	Frozen	92.407	**90.621**	**83.519**	10.630
	Fine-tune	**92.631**	90.174	71.103	-
ResNet-152	-	89.326	84.323	61.411	10.585
	Frozen	92.095	90.889	**84.323**	9.558
	Fine-tune	**94.060**	**91.157**	72.086	-

Size	100%	80%	60%	40%	20%	0%
Frozen	92.407	89.192	82.448	76.061	65.386	28.093
Fine-tune	92.631	92.586	89.951	84.949	84.055	84.011

Pre-training	Method	Accuracy
-	EI[28]	80.0
	Widar3.0[51]	**92.9**
-	ResNet-17	86.780
	ResNet-35	88.656
	ResNet-50	89.013
	ResNet-101	89.058
	ResNet-152	**89.326**
RF-URL (Fine-tune)	ResNet-17	91.201 (+4.421)
	ResNet-35	92.363 (+3.707)
	ResNet-50	92.631 (+3.618)
	ResNet-101	93.301 (+4.243)
	ResNet-152	**94.060** (+4.734)
RF-URL (Details)	ResNet-50 (baseline)	89.013
	+ RF-URL(frozen)	92.229 (+3.216)
	+ Predictor	92.407 (+0.178)
	+ Fine-tune	92.631 (+0.224)
	+ 3D CNN	84.323 (-8.308)
	+ feature in translator	96.784 (+12.461)
	+ Shuffle BN	**97.008** (+0.224)

Song et al. RF-URL In ***MobiCom*** 2022

RF-Based Self-supervised Learning

Experimental results

Pose Estimation

Model	Pre-training	100% labels	50% labels	10% labels
RFP-T	-	102	304	305
	Frozen	68	72	104
	Fine-tune	**63**	**68**	**103**
RFP-B	-	114	288	305
	Frozen	71	76	111
	Fine-tune	**64**	**70**	**109**
RFP-L	-	262	303	305
	Frozen	77	82	**119**
	Fine-tune	**64**	**71**	122

Size	100%	80%	60%	40%	20%	0%
Frozen	68	79	88	101	109	198
Fine-tune	63	67	72	78	83	86

Pre-training	Method	Pose Err.(mm)
-	RF-Pose3D[54]	112.7
-	RFP-T	**102**
	RFP-B	114
	RFP-L	262
RF-URL (Fine-tune)	RFP-T	**63** (-39)
	RFP-B	64 (-50)
	RFP-L	64 (-198)
RF-URL (Details)	Baseline: RFP-T(w/o IAM)	97
	+ RF-URL(frozen)	79 (-18)
	+ CSA	70 (-9)
	+ Predictor	68 (-2)
	+ Fine-tune	63 (-5)
	+ Shuffle BN	62 (-1)

Song et al. RF-URL In ***MobiCom*** 2022

RF-Based Self-supervised Learning

Experimental results

Silhouette Segmentation

Model	Pre-training	100% labels	50% labels	10% labels
RFSG-T	-	0.539	0.539	0.457
	Frozen	0.552	0.553	0.532
	Fine-tune	**0.610**	**0.611**	**0.581**
RFSG-B	-	0.556	0.550	0.481
	Frozen	0.557	0.552	0.537
	Fine-tune	**0.619**	**0.614**	**0.586**
RFSG-L	-	0.571	0.591	0.502
	Frozen	0.536	0.529	0.506
	Fine-tune	**0.613**	**0.612**	**0.565**

Size	100%	80%	60%	40%	20%	0%
Frozen	0.557	0.531	0.529	0.489	0.426	0.239
Fine-tune	0.619	0.602	0.585	0.573	0.562	0.556

Pre-training	Method	IoU
-	RF-Pose[52]	**0.583**
-	RFSG-T	0.539
	RFSG-B	0.556
	RFSG-L	**0.571**
RF-URL (Fine-tune)	RFSG-T	0.610 (+0.071)
	RFSG-B	**0.619** (+0.063)
	RFSG-L	0.613 (+0.042)
RF-URL (Details)	RFSG-B (baseline)	0.556
	+ RF-URL(frozen)	0.557 (+0.001)
	+ Fine-tune	0.611 (+0.054)
	+ Predictor	0.619 (+0.008)
	+ Shuffle BN	0.614 (-0.005)

Song et al. RF-URL In ***MobiCom*** 2022

RF-Based Self-supervised Learning

Conclusion

- A general self-supervised learning framework for RF sensing tasks
  - Propose data augmentations for RF signals
  - Enhance multiple sensing tasks in an unsupervised manner
- Conducting extensive experiments to demonstrate its effectiveness
  - Tested on multiple sensing tasks

Limitations

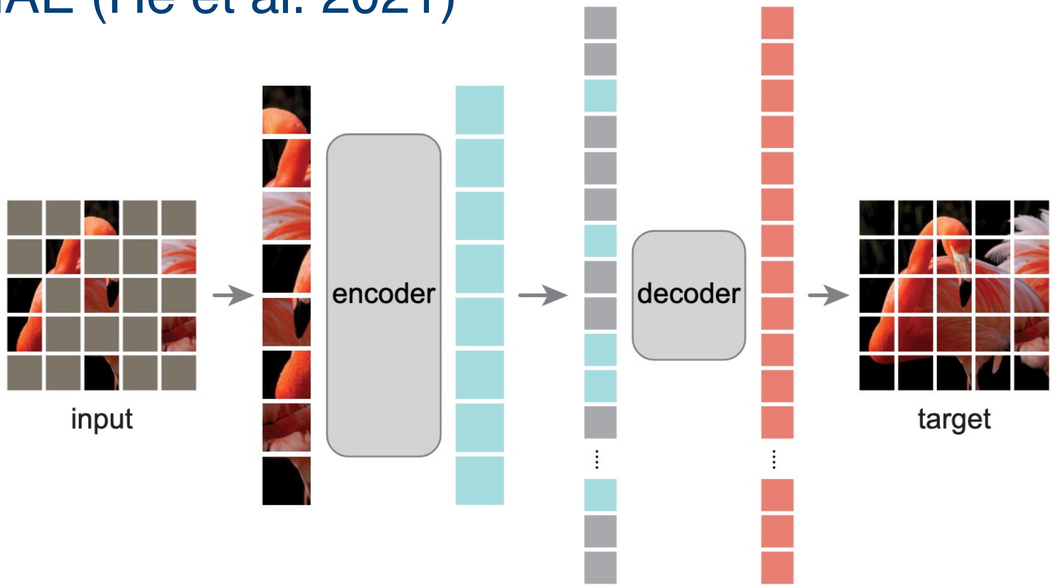
- The unique characteristics of RF signals are not considered.
- Number of signal representations can be limited

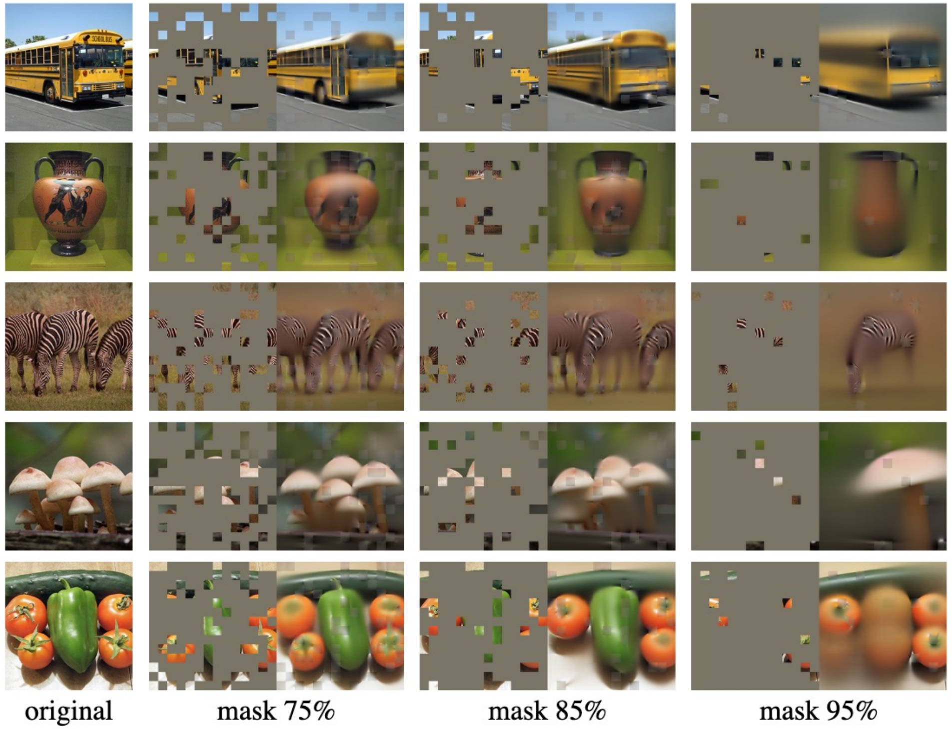
Song et al. RF-URL In ***MobiCom*** 2022

Self-Supervised Learning with MAE

Background

MAE (He et al. 2021)





Semantic information can be recovered from as few as 5% of the patches.

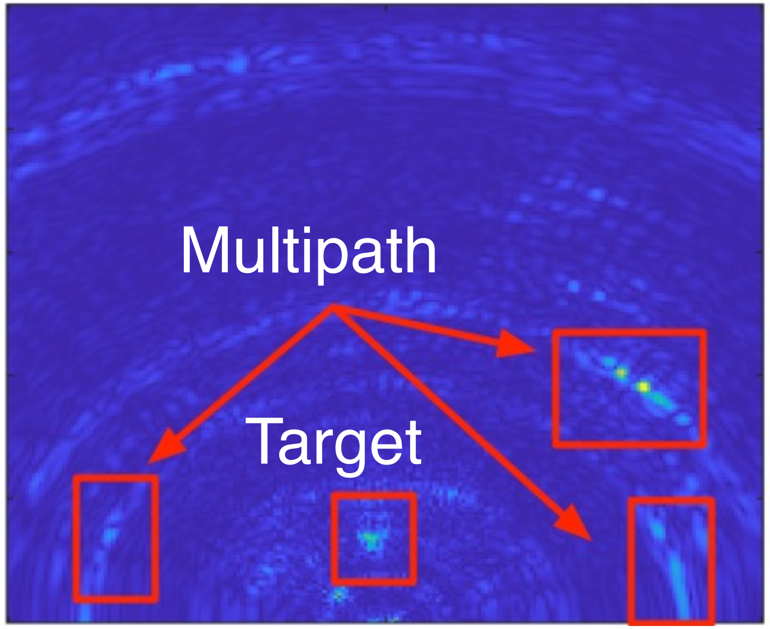
Can this level of recovery be achieved in the RF domain? **Very difficult**

Fang et al. PRISM In ***INFOCOM*** 2024

Self-Supervised Learning with MAE

Challenges

RF data is **sparse** and **noisy**



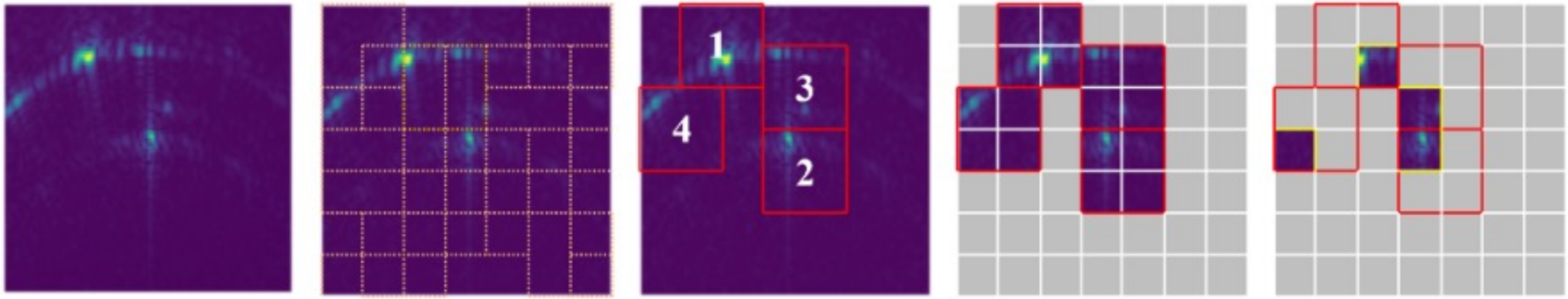
Random masking often results in losing the target.

Overfitting noise leads to learning non-informative features.

Fang et al. PRISM In ***INFOCOM*** 2024

Self-Supervised Learning with MAE

Solution: Sparsity-aware masking



1. Generate dense region proposals.
2. Rank regions by energy and select the top-k.
3. Mask the selected regions.
4. Reconstruct only the missing parts within these regions.

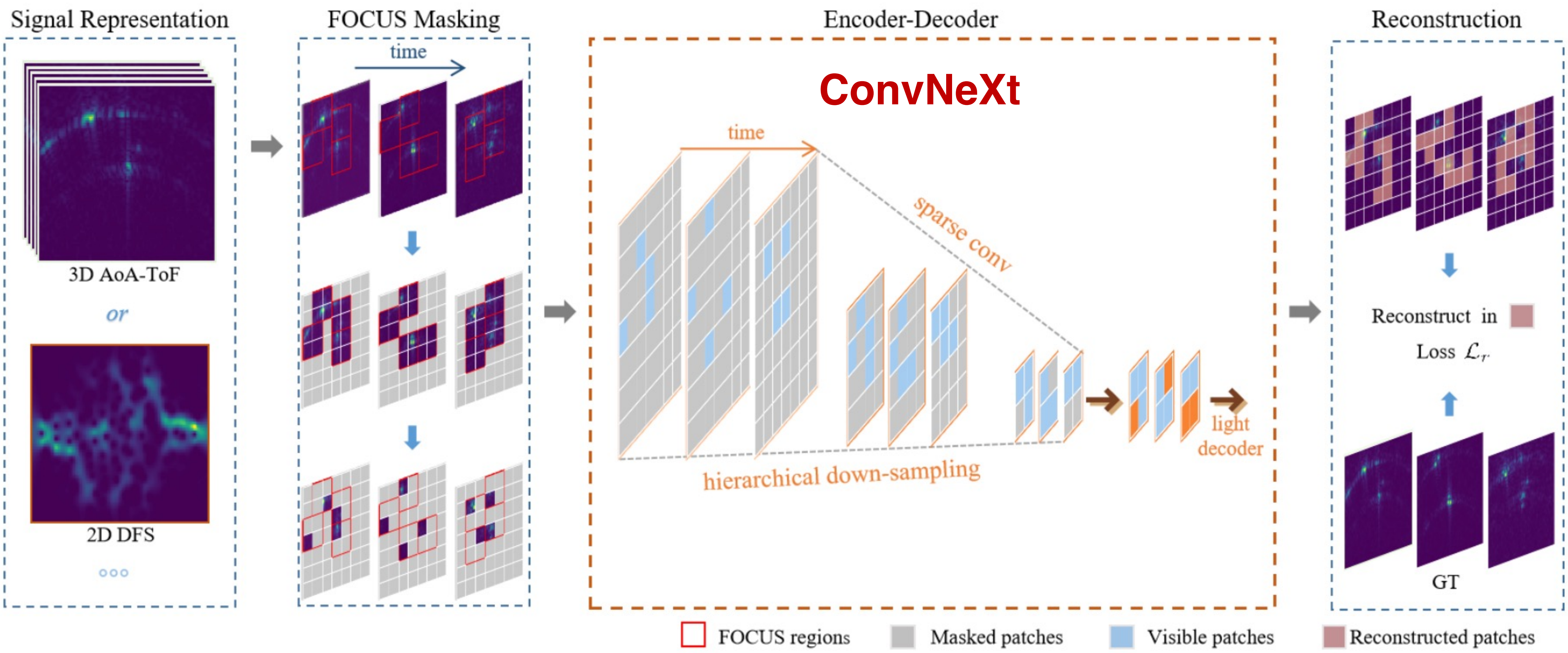
Strategy	Mask ratio	Silhouette (IoU↑)	Pose (MPJPE↓)
FOCUS	$\gamma = 92\%$	**0.7642**	**73.10**
Random	$\gamma = 92\%$	0.7593	79.11

Notable performance boost at the same mask ratio

Fang et al. PRISM In ***INFOCOM*** 2024

Self-Supervised Learning with MAE

Framework

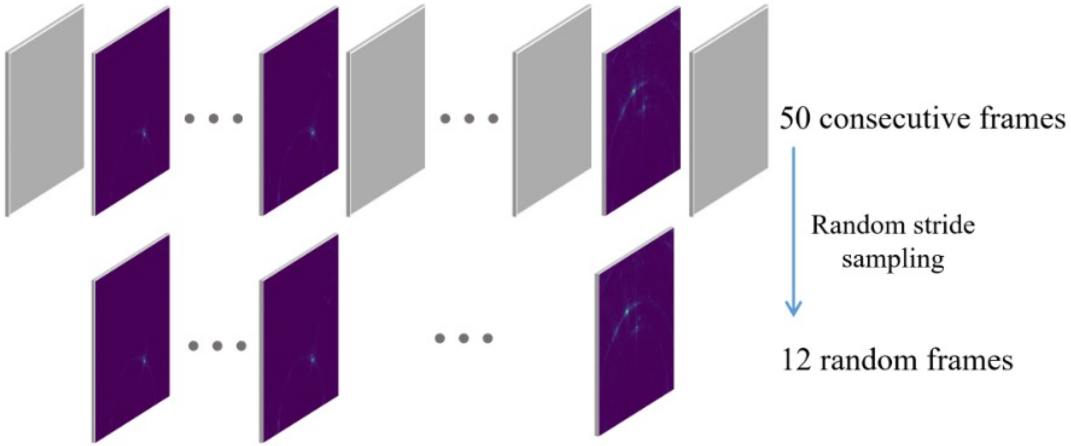


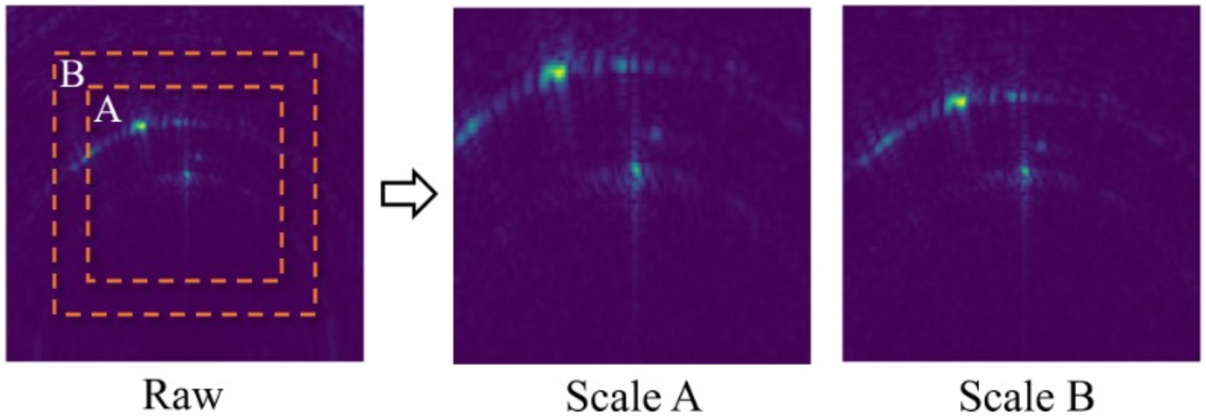
Fang et al. PRISM In ***INFOCOM*** 2024

Self-Supervised Learning with MAE

中国科学技术大学
University of Science and Technology of China

Implementation





Random stride sampling: to avoid information leak from consecutive frames

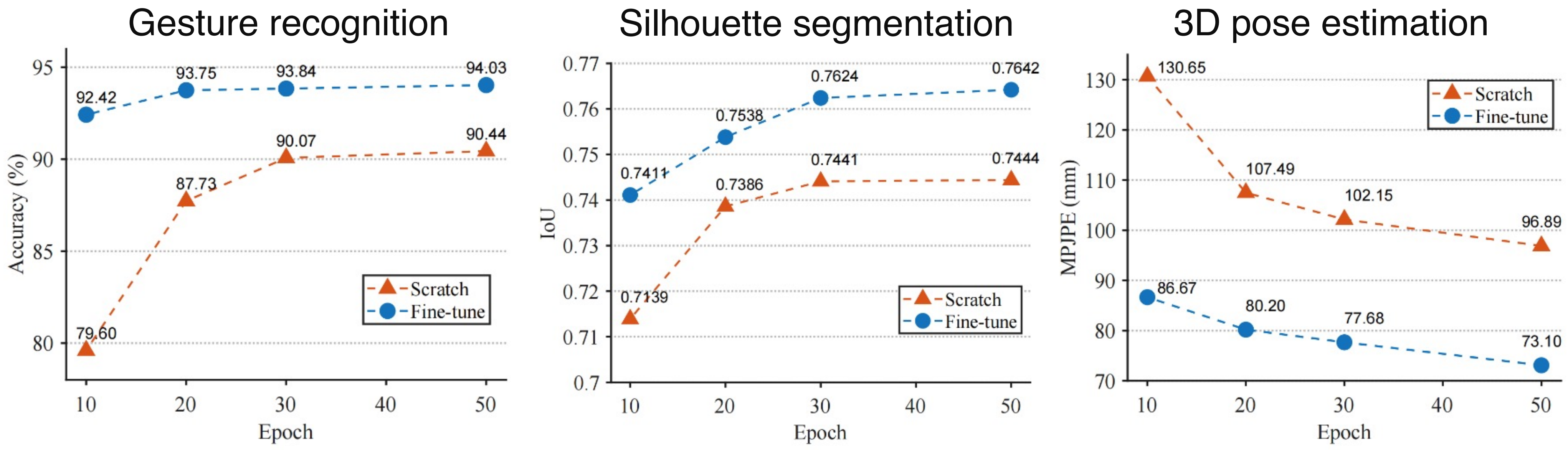
Multi-scale central crop: to amplifies the information-dense region

Component	Silhouette		Pose	
	IoU↑	Improv.	MPJPE↓	Improv.
Full (ConvX3D-S)	**0.7642**	-	**73.10**	-
- stride sample	0.7597	(-0.0045)	73.58	(+0.48)
- central crop	0.7575	(-0.0022)	76.21	(+2.63)

Fang et al. PRISM In ***INFOCOM*** 2024

Self-Supervised Learning with MAE

Performance



Converges faster & improved performance

Fang et al. PRISM In ***INFOCOM*** 2024

中国科学技术大学
University of Science and Technology of China

Self-Supervised Learning with MAE

Performance

TABLE III
THE RESULTS OF SILHOUETTE SEGMENTATION AND POSE ESTIMATION UNDER RADAR-BASED DATASET

Method	Bacbone	Silhouette Segmentation			Pose Estimation		
		IoU↑	w/ Fine-tune	Improv.	MPJPE↓	w/ Fine-tune	Improv.
Supervised	RF-Pose2D/3D	0.7060	-	-	110.56	-	-
RF-URL [7]	ResNet3D-50	0.7278	0.7338	(+0.0060)	**90.74**	87.54	(-3.20)
PRISM	ResNet3D-50	0.7278	0.7352	(+0.0074)	**90.74**	88.06	(-2.68)
PRISM	ConvX3D-S	0.7444	**0.7642**	(+0.0198)	96.89	73.10	(-23.79)
	ConvX3D-B	**0.7485**	0.7635	(+0.0150)	98.11	**72.67**	(-25.44)
	ConvX3D-L	0.7473	0.7630	(+0.0157)	99.83	80.51	(-19.32)

Better performance than RF-URL

Much higher performance with new backbone

TABLE IV
THE RESULTS OF GESTURE RECOGNITION UNDER WIFI-BASED DATASET

Method	Backbone	Accuracy↑	w/ Fine-tune	Improv.
Supervised	EI [11]	80.0	-	-
	Widar3.0 [8]	92.9	-	-
RF-URL [7]	ResNet2D-17	86.78	91.20	(+4.42)
	ResNet2D-50	89.01	92.63	(+3.62)
	ResNet2D-152	89.32	94.06	(+4.74)
PRISM	ConvX2D-S	**90.44**	94.03	(+3.59)
	ConvX2D-B	88.69	**94.72**	(+6.03)
	ConvX2D-L	87.59	94.49	(+6.90)

Much lower memory consumption & faster

TABLE V
THE COMPARISONS OF MEMORY AND SPEED

Method	Backbone	Enc #Para.(M)	Memory(G)	Speedup
RF-URL [7]	ResNet3D-50	4.29 × 2	75	1×
PRISM	ResNet3D-50	4.29	28	1.33×
	ConvX3D-S	**3.20**	**27**	**1.42**×
	ConvX3D-B	4.59	29	1.25×
	ConvX3D-L	8.11	32	1.11×

Fang et al. PRISM In *INFOCOM* 2024

Self-Supervised Learning with MAE

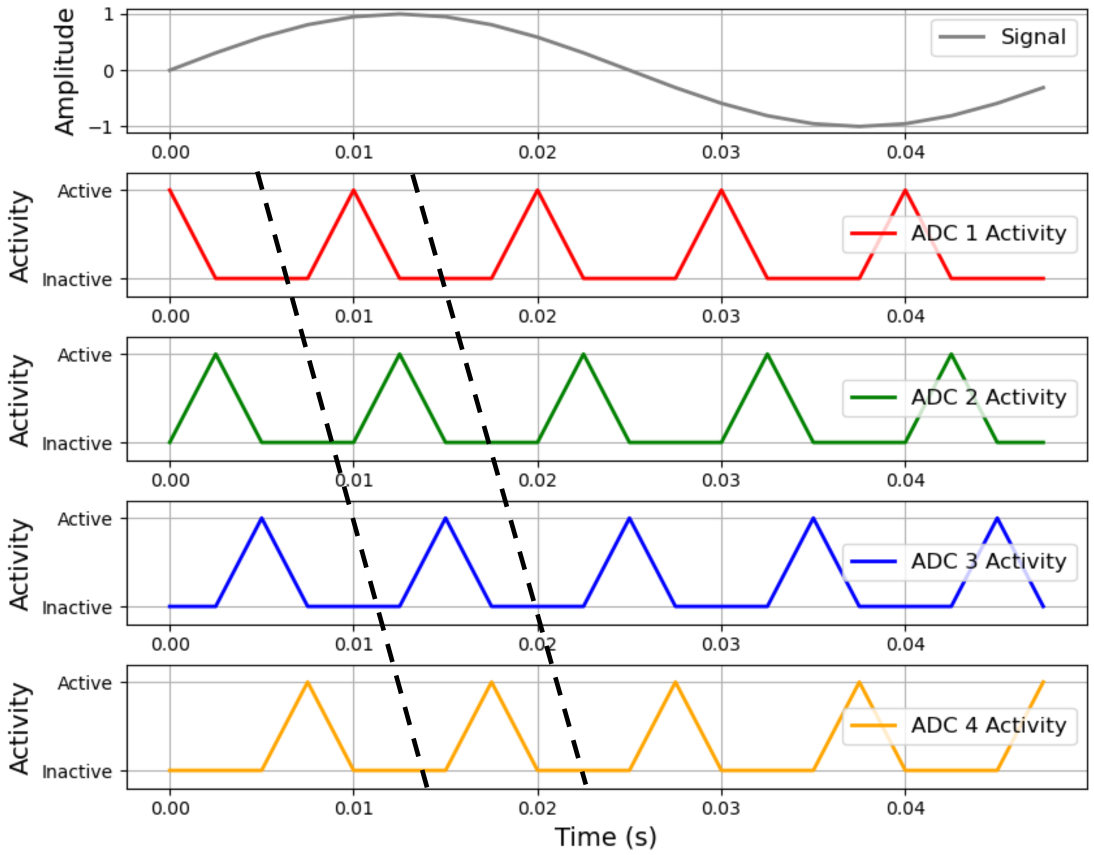
Conclusion

- An easy-to-follow approach with a sparsity-aware design.
- Does not rely on signal representations.
- Only considering the sparsity can largely simplify the framework

Self-Supervised Learning with Group Shuffle

Domain-aware design

Time-Interleaved Analog-to-Digital Converter (TI-ADC)



n ADCs can be arranged to achieve a nx sampling rate

Channel Swapping Error: the output from multiple ADC channels is mis-ordered

ADC1, ADC2, ADC3, ADC4, ..., ADC1, ADC2, ADC3, ADC4

↓

ADC2, ADC1, ADC3, ADC4, ..., ADC2, ADC1, ADC3, ADC4

randomly shuffling each group

Song et al. **TMM** 2024

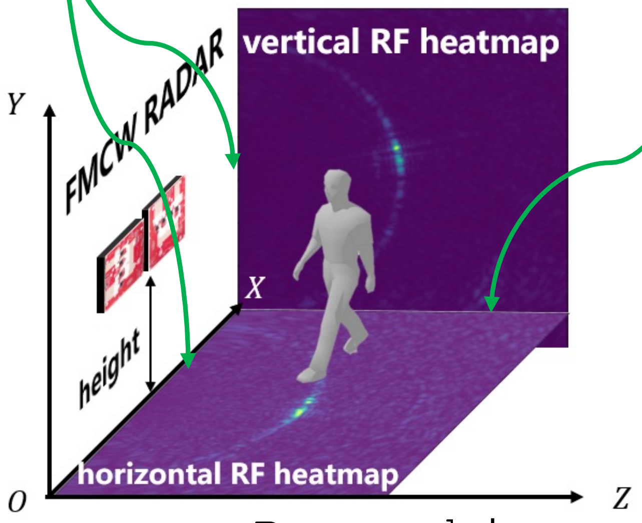
Self-Supervised Learning with Group Shuffle

Group Shuffle in Image Patch

Temporal shuffling → FFT → Range bin shuffling

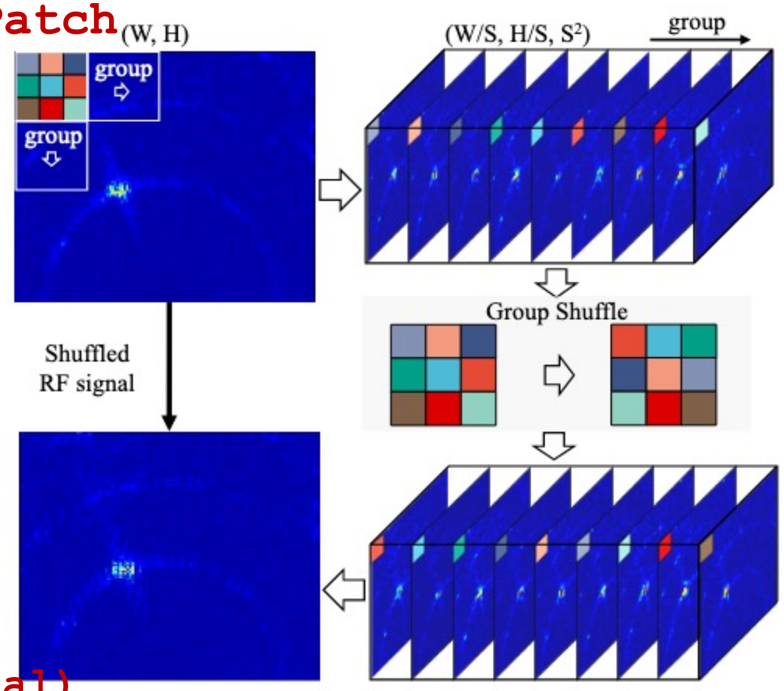
Antenna shuffling → Azimuth or Elevation shuffling

⟹ Radar **Image Patch Shuffling**



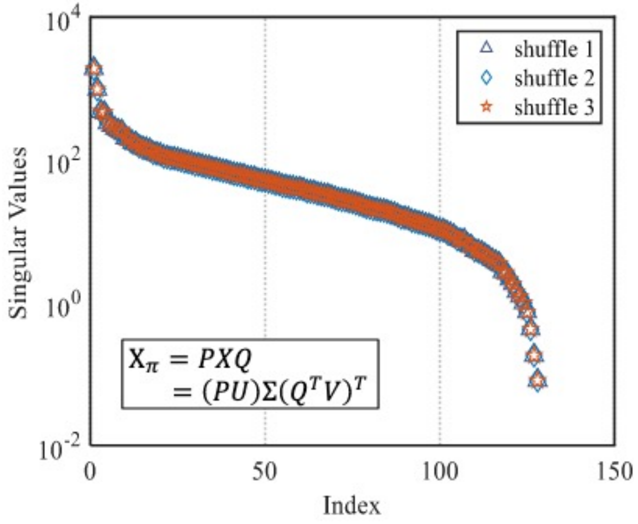
Range bin

Patch



(a)

1. **All Patches follow the same shuffling**



(b)

2. **shuffling matrix (orthogonal)**

$$\mathbf{P} = \begin{bmatrix} \Pi_r & & \\ & \ldots, & \\ & & \Pi_r \end{bmatrix}, \mathbf{Q} = \begin{bmatrix} \Pi_c & & \\ & \ldots, & \\ & & \Pi_c \end{bmatrix}$$

$$\mathbf{X}_\pi = \mathbf{P}\mathbf{X}\mathbf{Q} = (\mathbf{P}\mathbf{U})\mathbf{\Sigma}(\mathbf{Q}^T\mathbf{V})^T$$

3. **singular values stay unchanged**

Song et al. **TMM** 2024

中国科学技术大学 University of Science and Technology of China

Self-Supervised Learning with Group Shuffle

Group Shuffle in Raw Signal

Raw Signal Shuffle

$$\mathbf{R}_\pi(m,k) = \pi(\mathcal{R}(m,k)) = [\mathbf{P}\mathcal{R}\mathbf{Q}]_{m,k}$$

antenna frequency

$$\mathcal{R} \in \mathbb{C}^{\frac{M}{S} \times \frac{K}{S} \times S^2}$$

raw signal

$$\Pi_r = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \Pi_r = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

permutation

Beamforming

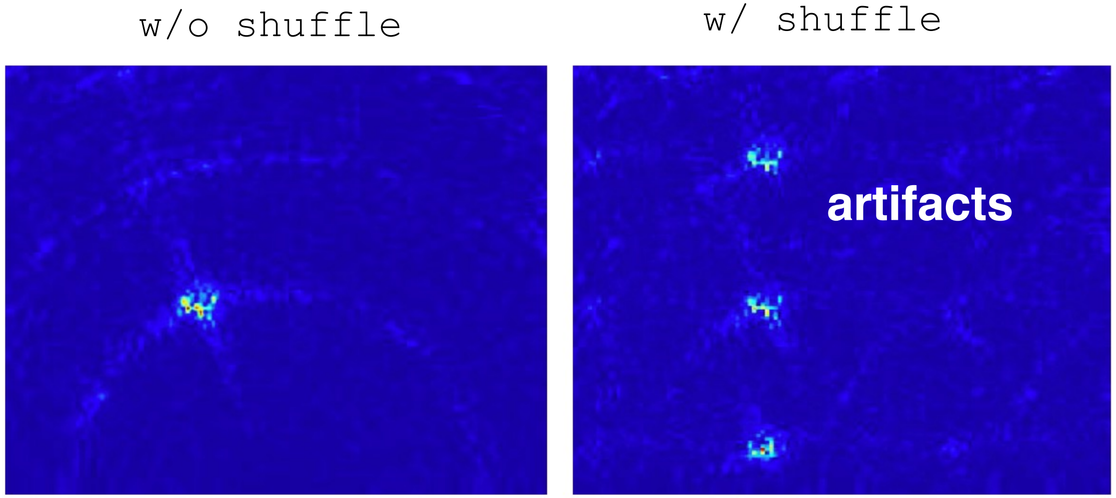
$$\mathbf{X}_\pi(x,y) = \sum_{m=0}^{\frac{M}{2}} \sum_{k=0}^{\frac{K}{2}} \mathbf{R}_{2m,2k} e^{j2m\left(\frac{2m+1}{2m}\Phi_\theta\right)} e^{j\left(\frac{2m+1}{2m}\Phi_\tau\right)}$$

$$+ \sum_{m=0}^{\frac{M}{2}} \sum_{k=0}^{\frac{K}{2}} \mathbf{R}_{2m+1,2k+1} e^{j(2m+1)\left(\frac{2m}{2m+1}\Phi_\theta\right)} e^{j\left(\frac{2m}{2m+1}\Phi_\tau\right)}$$

energy re-distribution

$$(\Phi_\theta, \Phi_\tau) \nearrow \left(\frac{2m}{2m+1}\Phi_\theta, \frac{2m}{2m+1}\Phi_\tau\right)$$

$$(\Phi_\theta, \Phi_\tau) \searrow \left(\frac{2m+1}{2m}\Phi_\theta, \frac{2m+1}{2m}\Phi_\tau\right)$$

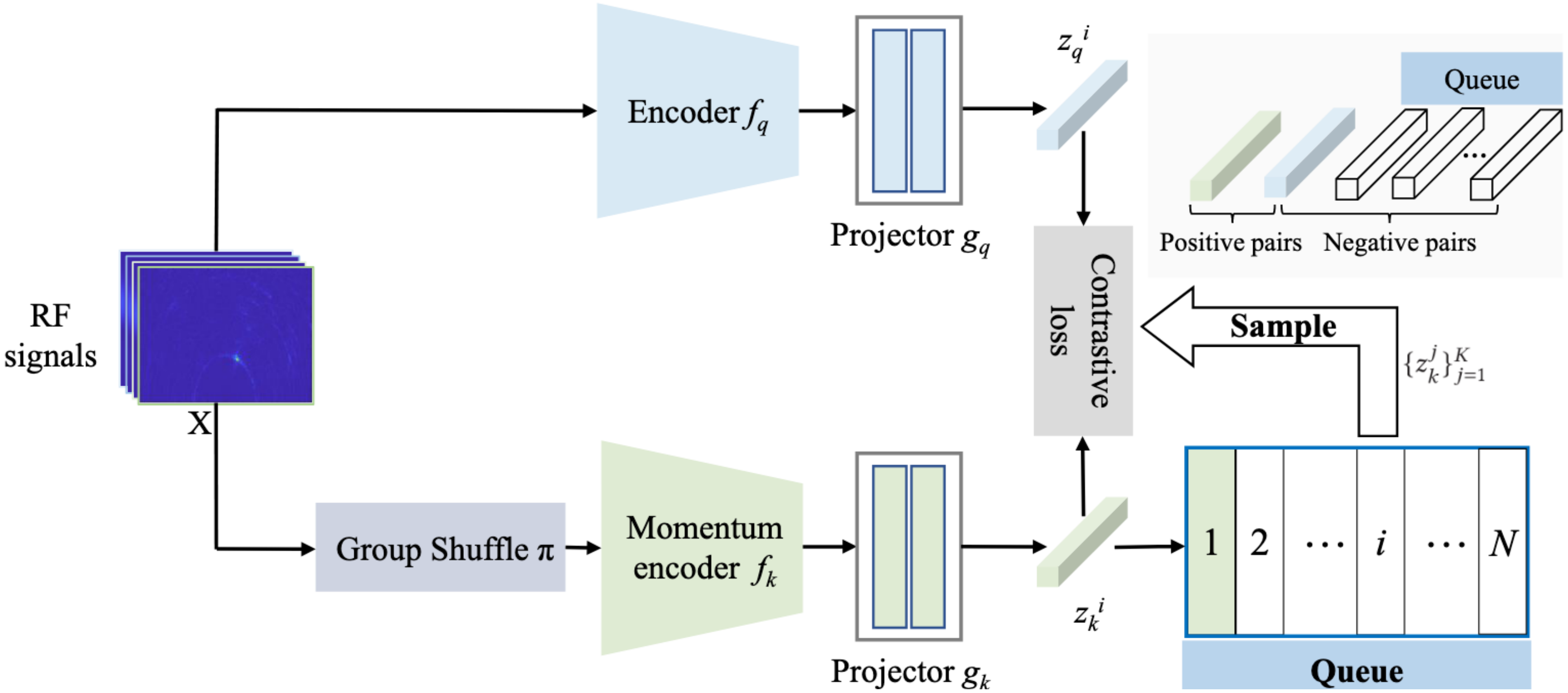


Song et al. **TMM** 2024

Self-Supervised Learning with Group Shuffle

Framework

No. of Shuffles: $A(m, m)^2 = m!^2 \quad A(4, 4)^2 = 576$



Asymmetry augmentation

	3D Pose	Silhouette	Action
w/o Asym.	128.96	0.709	91.483
w Asym.	121.13	0.727	92.298

Song et al. **TMM** 2024

Self-Supervised Learning with Group Shuffle

Performance

TABLE 2
Evaluation of different models on 3D pose estimation, silhouette generation, and action recognition under fine-tuning setting, the relative improvements over supervised training from scratch, and comparison with the SOTA model.

Method	backbone	3D pose estimation (MPJPE mm)			Silhouette generation (IoU)			Action recognition (Acc. %)		
		Scratch	Fine-tune	Improve	Scratch	Fine-tune	Improve	Scratch	Fine-tune	Improve
Supervised	RF-Pose [23]	-	-	-	0.697	-	-	-	-	-
	RF-Pose3D [30]	165.00	-	-	-	-	-	-	-	-
TGUL [8]	CNX3D-B	134.90	128.56	+6.34	0.699	0.698	-0.001	85.211	89.018	+3.807
RF-URL [7]	CNX3D-B	134.90	133.91	+0.99	0.699	0.707	+0.008	85.211	88.930	+3.719
GSAA	CNX3D-S	**133.84**	123.56	+10.28	0.696	0.724	+0.028	**85.938**	92.077	+6.139
	CNX3D-B	134.90	121.13	+13.77	**0.699**	0.727	+0.028	85.211	92.298	+7.087
	CNX3D-L	135.64	**120.66**	+14.98	**0.699**	**0.728**	+0.029	85.185	**92.380**	+7.195

	3D pose estimation (MPJPE mm)				Silhouette generation (IoU)				Action recognition (Acc. %)			
	p=0.5	0.7	0.9	1.0	p=0.5	0.7	0.9	1.0	p=0.5	0.7	0.9	1.0
group shuffle	124.97	**124.23**	124.46	125.17	0.721	0.722	**0.724**	0.722	90.097	**90.911**	90.889	90.674
rows or columns	126.04	125.27	126.09	125.65	0.719	0.720	0.722	0.723	90.449	90.471	90.735	90.405
shuffle all	125.70	124.24	124.66	126.16	0.721	0.721	0.721	0.722	90.537	90.801	89.679	90.889

Song et al. **TMM** 2024

Contents

中国科学技术大学
University of Science and Technology of China

University of Science and Technology of China

Conclusion

- **Unique Capabilities**: *It offers advantages in privacy-preserving sensing, through-wall detection, and operation in low-light or occluded environments that vision and audio cannot achieve*.

- **Current Limitations**: *lacks the spatial and contextual richness of vision or the sequential detail of audio/text, making it less comprehensive as a standalone modality in general*.

- **Potential for Growth**: *Advances in advanced devices and deep learning approaches could enable RF to bridge its limitations, making it more comparable to traditional modalities.*

Key Takeaways

- **Pose Estimation**
  - Irreplaceable in specific scenarios
  - A promising area of research
  - Publicly available datasets
- **Fine-grained Vital Signs**:
  - More fine-grained physiological signals can be sensed.
  - Limited existing research
- **Self-supervised Learning:**
  - An important and relatively new topic
  - Generalization issue
  - Unlabeled data are easy to collect

Thanks！

Q&A

中国科学技术大学
University of Science and Technology of China