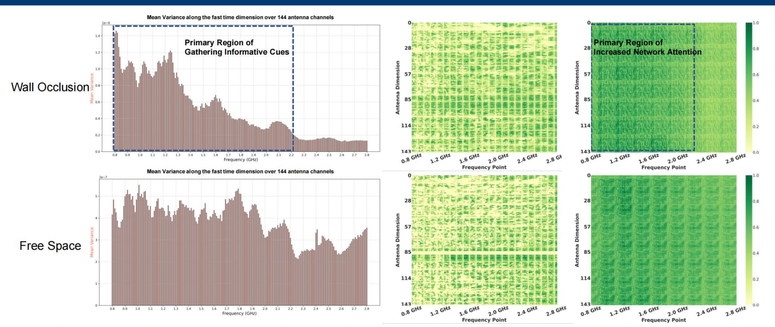
RFMamba: Frequency-Aware State Space Model for RF-Based Human-Centric Perception

Rui Zhang[1], *Ruixu Geng*[1], *Yadong Li*[2], *Ruiyuan Song*[1], *Hanqin Gong*[1], *Dongheng Zhang*[1], *Yang Hu*[1], *Yan Chen*[1*]

[1] *University of Science and Technology of China* [2] *University of Washington*

*ruizh@mail.ustc.edu.cn, eecyan@ustc.edu.cn**

Motivation



(a) Variance Distribution Across Different Frequency Ranges (b) Transformer-based Method (c) Our Mamba-based Method

A natural question raised:

Can a more efficient yet effective solution be developed to capture the long-range dependencies across large-sequence RF?

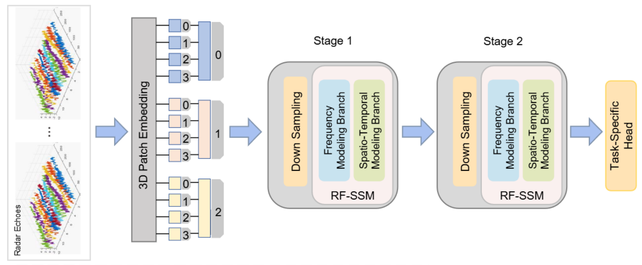
Contributions

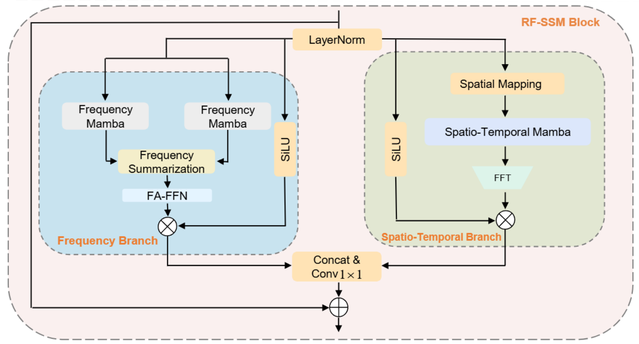
- We **pioneer the first state space model for RF-based human perception**, demonstrating the potential of Mamba for efficient yet effective global modeling in long-sequence RF signals.
- We introduce a novel RF-SSM block which integrates **both frequency domain and spatio-temporal domain modeling** to effectively capture critical characteristics of RF signals for human perception.
- We propose a **six-way scanning strategy** in the frequency modeling branch, which ensures comprehensive interaction of amplitude and phase information across all dimensions and is capable of adaptively selecting the most informative frequency cues.

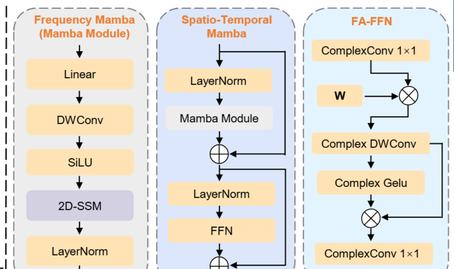
RFMamba Framework

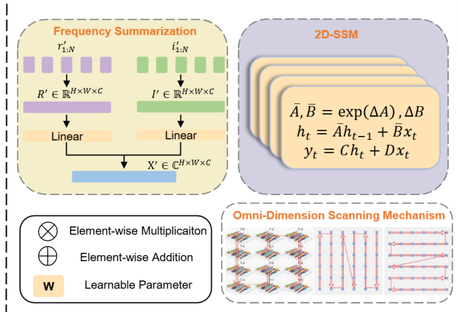
Three Key Insights:

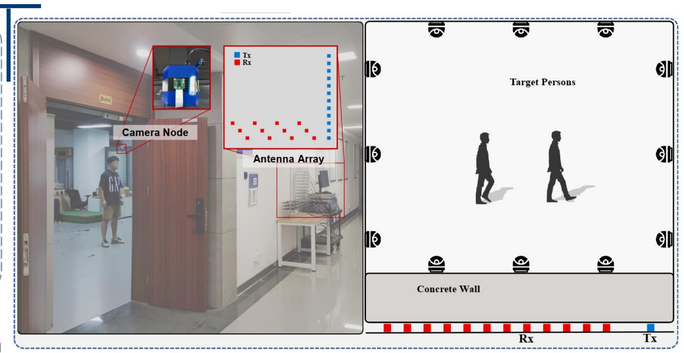
1. **Amplitude-Phase Joint Modeling** effectively captures the unique characteristics of radio signals.
2. **Spatio-Temporal Joint Modeling** better facilitates the learning process from the spatial domain perspective.
3. **Omni-Dimensional Scanning Mechanism** ensures comprehensive interaction across all dimensions.

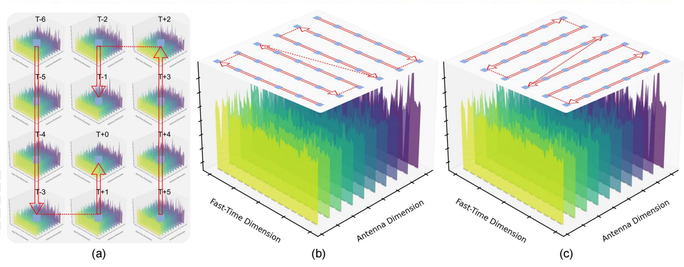




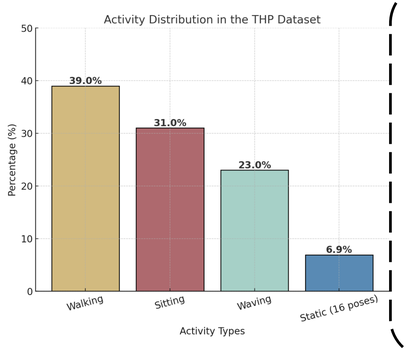








Setting and Baseline

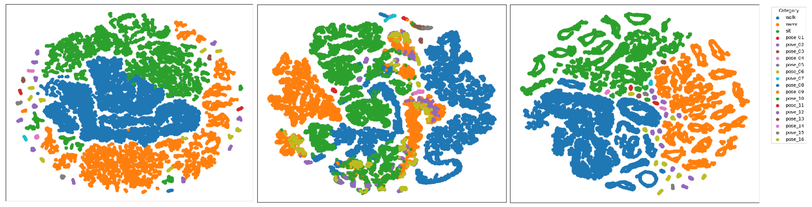


Overall, our THP dataset contains the largest number of paired optical images and radar echoes among existing datasets, and a high-precision multi-camera system is employed to provide high-quality labels.

COMPARISONS OF THE THP DATASET AND RELATED THROUGH-WALL DATASETS.

Dataset	Dataset Size	Person	Environment	Devices for Collecting Labels	Label Type
RPSNet [54]	9,301	5	Wall-occlusive Free-space	RGB Camera with 3 FPS	2D Joint Coordinate Human Shape
MIMDSN [8]	181,780	10	Wall-occlusive Free-space Low-visibility	RGB-D Camera with 3 FPS	3D Joint Coordinate
RadarFormer [9]	194,760	10	Wall-occlusive (without labels, 32,480) Free-space (162,280)	RGB-D Camera with 3 FPS	3D Joint Coordinate Action Category Person Category
Our THP	491,000	10	Wall-occlusive (347,000) Wooden-occlusive (24,000) Free-space (120,000)	Multi-Camera System with 12 FPS (12 camera nodes)	3D Joint Coordinate Action Category Person Category

Experimental Results



(a) CNN-based ResNet3D-50 (b) Transformer-based RadarFormer (c) Mamba-based RFMamba

Quantitative and Qualitative Results on Diverse Downstream Perception Tasks.

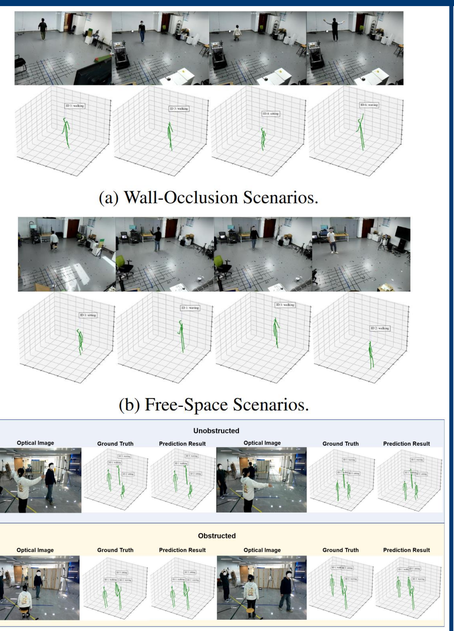
Method	Action Recognition		Person Re-ID	
	Accuracy	F1 Score	mAP	CMC-1
ResNet3D-50	0.9871	0.9820	0.8575	0.8979
RadarFormer	0.8918	0.7020	0.4817	0.8986
RFMamba	**0.9997**	**0.9994**	**0.9991**	**0.9967**

Table 2: Evaluation of Action Recognition and Person Re-ID.

Method	Nose	Neck	Shoulder	Elbow	Wrist	Hip	Knee	Ankle	Eye	Ear	Mean (mm) ↓	Params (M)
RF-Pose3D	81.30	62.98	78.95	96.93	122.06	75.84	80.47	83.44	78.63	79.67	85.35	10.91
ResNet3D-50	105.26	86.10	98.38	114.51	161.95	88.62	87.35	97.82	106.13	97.60	105.34	352.30
mm-Pose	160.92	147.69	159.82	189.97	247.54	145.57	138.19	140.37	161.65	156.41	165.98	22.07
RadarFormer	281.26	239.97	248.36	277.24	344.03	231.88	221.23	208.42	280.02	258.29	258.90	12.88
RFMamba	**51.85**	**41.31**	**44.03**	**52.73**	**68.89**	**41.53**	**46.32**	**55.18**	**53.30**	**47.24**	**50.64**	1.94

Table 1: Quantitative Evaluation Results for Pose Estimation Task. The notation '↓': lower is better.

Visualization



(a) Wall-Occlusion Scenarios.

(b) Free-Space Scenarios.

Conclusions

1. **Framework**: We propose a novel RFMamba architecture, integrating omni-dimensional frequency scanning technology to achieve breakthrough capabilities in through-wall human sensing using radio frequency signals. This significantly enhances the accuracy of pose estimation, activity recognition, and person re-identification tasks.
2. **Generation:** The system maintains high robustness in complex conditions (e.g., free space and multi-person environments), outperforming existing methods and providing a reliable real-time monitoring solution.
3. **Application:** This technology enables new applications in medical monitoring, smart homes, and security systems, balancing non-intrusive sensing with privacy protection.