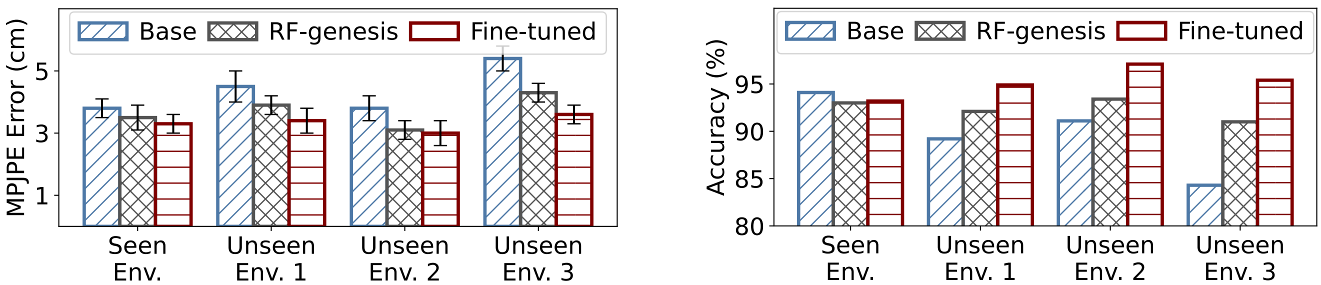
mmWEAVER: Supplementary Material

Mahathir Monjur
UNC Chapel Hill
mahathir@cs.unc.edu

Shahriar Nirjon
UNC Chapel Hill
nirjon@cs.unc.edu



(a) Pose estimation

(b) Gesture recognition

Figure 1. Performance comparison in different environments.

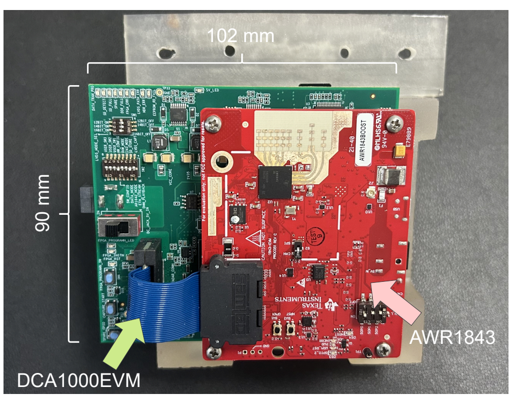


Figure 2. Radar system setup.

1. Background

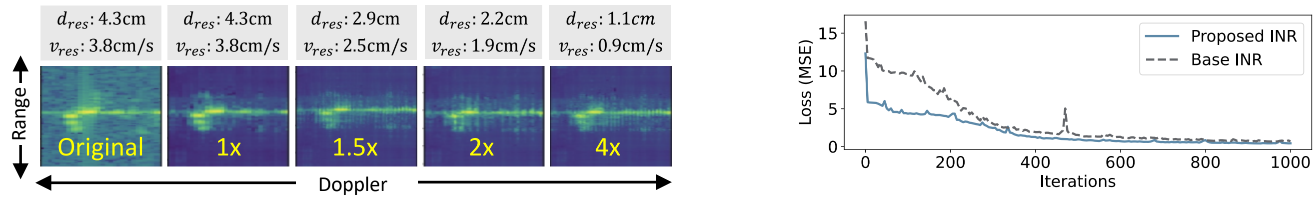
To acquire radar data for human pose estimation, we construct a hardware pipeline comprising a mmWave radar sensor and a high-speed data acquisition interface. Specifically, we utilize the TI AWR1843 [4], a compact single-chip radar system equipped with three transmit and four receive antennas. This sensor is integrated with the DCA1000EVM, which serves as the interface for streaming raw intermediate frequency (IF) signals to a host machine in real time. The radar transmits FMCW signals and receives their echoes to infer spatial and motion-related characteristics of targets. This setup forms the foundation of our self-collected dataset, described in Section 4 of the main paper. Figure 2 provides a visual overview of the sensing components, including their physical dimensions and connectivity.

2. Preliminary Study

The effectiveness of synthetic data in improving classifier performance relies on its ability to represent the target environment adequately. Techniques like data augmentation [6], electromagnetic simulations [5, 6], and model-based generative approaches [2, 8, 9] enhance dataset to a certain degree but often fail to capture critical environmental nuances in specific indoor settings. State-of-the-art methods like RF-Genesis [1], while capable of generating large synthetic datasets, often fall short in adequately representing the target environment, resulting in a performance gap. This study investigates that gap.

2.1. Experimental Setup

Data Collection: For this study, we collect mmWave data using a TI AWR1843 radar from six indoor environments, including 2 living rooms, 2 offices, 1 kitchen and a lab space. These rooms contain typical furniture such as chairs, tables and couches with walls, floors, and ceilings made of concrete and wood. Six participants perform six different gestures and activities with repetitions. This process results in a dataset comprising 672 samples. We use data from a living room and an office for testing and the remaining four environments for training.



(a) Signal Representation and Super-resolution Capability.

(b) Faster Convergence.

Figure 3. mmWeaver enables multi-resolution signal representation, 49x storage reduction, and faster convergence.

Data Augmentation: To augment the collected dataset, we implement the state-of-the-art mmWave generation technique [1], which employs diffusion models to synthesize arbitrary environments and motions and then simulates mmWave signals based on these scenarios.

Tasks and Classifiers: We implement two downstream tasks: *pose estimation*, which predicts location of 15 body joints, and *activity classification*, which recognizes six human gestures and activities–*clockwise turn, anti-clockwise turn, swipe left, swipe right, clapping and waving*. We use two state-of-the-art models: *mmMesh* [7] for pose estimation and *DI-Gesture* [3] for activity classification, respectively.

2.2. Experimental Observations

Three Model Variants: We compare three models for each task: the *Base* model, trained on data from four known environments and tested on three unseen environments; the *RF-Genesis* model, which extends the base model by further training on an augmented dataset and is also tested on the same unseen environments; and the *Fine-Tuned* model, which is fine-tuned using a portion of the test environment's data. Since the *Fine-Tuned* model is trained on data from the same environment where it is tested, it is expected to achieve the highest accuracy by directly adapting to the environment's specific characteristics.

☐ **Observation 1 – Performance Drops in Unseen Environments.** Figure 1 reports pose estimation errors and gesture recognition accuracies, respectively. When applied to unseen environments, the *Base* model's pose estimation error increases by 0.5–2.1 cm, and its gesture classification accuracy drops by 3%–10%.

☐ **Observation 2 – Generic Data Augmentation Provides Limited Gains.** Retraining with synthetic datasets generalized for different environments reduces the *Base* model's pose estimation error by 0.6–1.1 cm, as shown in Figure 1. Similarly, generalized data synthesis improves classification accuracy by 2%–7%. However, challenges persist in addressing discrepancies caused by environmental variability and complex motion dynamics.

☐ **Observation 3 – Environment-Specific Fine-Tuning Yields Superior Results.** As illustrated in Figure 1, the environment-specific *Fine-Tuned* model outperforms both the *Base* model and models trained with environment-agnostic synthetic data. This underscores the importance of learning mmWave signal propagation functions specific to unseen environments to generate tailored datasets. Such targeted datasets can significantly improve performance, effectively bridging the gap between generalized data augmentation techniques and optimal results.

3. INR Architecture Evaluation

To evaluate the effectiveness of Implicit Neural Representations (INRs) in modeling complex radar signals, we use a sequence of 10 radar frames capturing a person waving. Our neural network, with 10,018 trainable parameters, modeled data dimensions of frames (10) $\times$ range (64) $\times$ Doppler (32) $\times$ angular (12) $\times$ complex values (2), totaling 491,520 data points. This setup achieves a compression ratio (CR) of approximately 49.06:

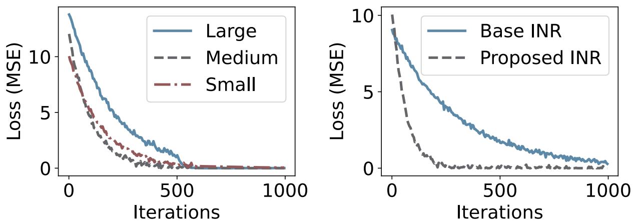
$$\text{CR} = \frac{\text{Total Data Points}}{\text{Number of Parameters}} = \frac{491,520}{10,018} \approx 49.06.$$

Figure 3b compares the average training loss over iterations between our INR with temporal modulation and a baseline INR that directly incorporates spatial and temporal coordinates as inputs. Our approach, which utilizes separate modules, demonstrates faster convergence, enabling improved representation of complex radar signals.

In Figure 3a, we compare range-Doppler spectrograms of the original and INR-reconstructed signals for frame 8. By modeling signals as continuous functions, our INR enables higher-resolution spectrograms. We showcase reconstructions at standard resolution (1x) and upsampled resolutions (1.5x, 2x, 4x), preserving Doppler characteristics of human waving while enhancing range (location) and Doppler (movement patterns) details, benefiting downstream analysis.

4. HyperNetwork Architecture Evaluation

To evaluate the effectiveness of our INR with a hypernetwork-based approach, we perform an ablation study in Figure 4b. The plot shows iteration-wise signal MSE loss over 1000 iterations, where our modulated INR converges faster and achieves 10× lower loss than the



(a) Effect of model size. (b) Faster convergence.

Figure 4. Evaluation of INR generalization.

baseline. This improvement demonstrates more accurate mmWave signal reconstruction from environmental images and pose sequences. By explicitly separating static and dynamic motion contexts, our INR and hypernetwork improve generalization across diverse signals.

Figure 4a further examines the impact of INR size (trainable parameters) on performance. We use a medium-sized INR (10K parameters) for all subsequent experiments. While larger models require more iterations to converge, they consistently achieve lower MSE loss at equivalent iteration counts.

References

[1] Xingyu Chen and Xinyu Zhang. Rf genesis: Zero-shot generalization of mmwave sensing through simulation-based data synthesis and generative diffusion models. In *Proceedings of the 21st ACM Conference on Embedded Networked Sensor Systems*, page 28–42, New York, NY, USA, 2024. Association for Computing Machinery. 1, 2

[2] Hao Huang, Guan Gui, H. Gaanin, C. Yuen, H. Sari, and F. Adachi. Deep regularized waveform learning for beam prediction with limited samples in non-cooperative mmwave systems. *IEEE Transactions on Vehicular Technology*, 2023. 1

[3] Yadong Li, Dongheng Zhang, Jinbo Chen, Jinwei Wan, Dong Zhang, Yang Hu, Qibin Sun, and Yan Chen. Towards domain-independent and real-time gesture recognition using mmwave signal. *IEEE Transactions on Mobile Computing*, 22(12): 7355–7369, 2022. 2

[4] Texas Instruments. Awr1843: 77-ghz radar-on-chip device, 2023. Accessed: 2023-12-05. 1

[5] Shelly Vishwakarma, Wenda Li, Chong Tang, Karl Woodbridge, Raviraj Adve, and Kevin Chetty. Simhumalator: An open-source end-to-end radar simulator for human activity recognition. *IEEE Aerospace and Electronic Systems Magazine*, 37(3):6–22, 2021. 1

[6] Zhiming Wang, Dechen Jiang, Bin Sun, and Y. Wang. A data augmentation method for human activity recognition based on mmwave radar point cloud. *IEEE Sensors Letters*, 2023. 1

[7] Hongfei Xue, Yan Ju, Chenglin Miao, Yijiang Wang, Shiyang Wang, Aidong Zhang, and Lu Su. mmmesh: towards 3d real-time dynamic human mesh construction using millimeter-wave. In *Proceedings of the 19th Annual International Conference on Mobile Systems, Applications, and Services*, page 269–282, New York, NY, USA, 2021. Association for Computing Machinery. 2

[8] Hongfei Xue, Qiming Cao, Chenglin Miao, Yan Ju, Haochen Hu, Aidong Zhang, and Lu Su. Towards generalized mmwave-based human pose estimation through signal augmentation. In *Proceedings of the ACM Conference*. ACM, 2023. 1

[9] Xiaotong Zhang, Zhenjiang Li, and Jin Zhang. Synthesized millimeter-waves for human motion sensing. In *Proceedings of the 20th ACM Conference on Embedded Networked Sensor Systems*, pages 377–390, 2022. 1