Prompt-to-Radar: A Generative Radar World Model for Real2Sim2Real Robot Navigation

Anuvab Sen **Saibal Mukhopadhyay**
School of Electrical and Computer Engineering, Georgia Institute of Technology
asen74@gatech.edu anuvabsen.com/demos/prompt-to-radar

Abstract: Radar perception requires labeled data for difficult scenes that are costly to collect on physical robots. We present Prompt-to-Radar, a generative radar world model that reconstructs real scenes, varies obstacles and trajectories, and renders labeled radar through a configurable physical sensor model. The generated data improves real RADIal detection by 15–16 percentage points in AP and F1 across three detector architectures. We also complete a Real2Sim2Real deployment on a LoCoBot: with an identical navigation policy, combining generated and real radar raises success in unseen corridors from 87/100 to 95/100 runs and reduces runs with a collision from eight to two. These results demonstrate transfer of learned radar perception to physical navigation and motivate controllable hard-scene generation for robot learning.

Keywords: Real2Sim2Real, radar perception, world models

1 Real2Sim2Real Radar Learning

Radar measurements depend on scene geometry, materials, multipath, and sensor configuration. A useful training simulator must therefore connect editable scenes to the observations of a particular radar. Prompt-to-Radar combines real-scene reconstruction with generative scene variation and radar synthesis. Building on Real2Sim2Real learning [1], the transferred component is learned perception: we train SSMRadNet [2] on generated radar and deploy it on a LoCoBot [3] while retaining the same navigation policy (Fig. 1). The new evidence is improved real-data detection across architectures and successful navigation transfer to unseen corridors.

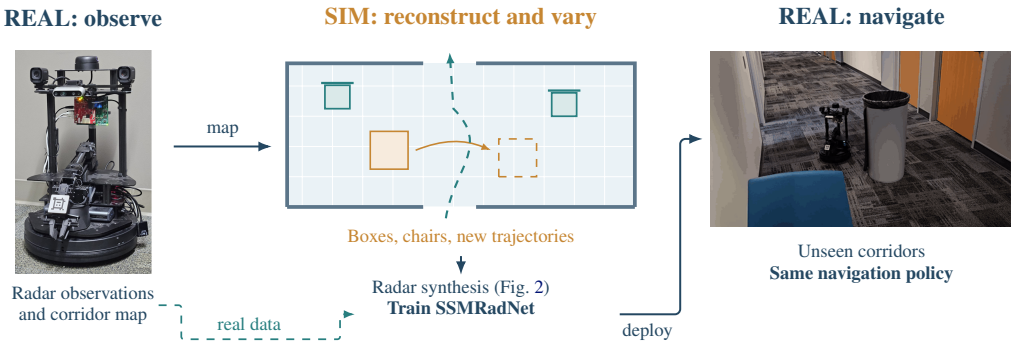


Figure 1: **Real2Sim2Real with a fixed navigation policy.** Real observations anchor an editable scene. Radar synthesis supplies labeled data for perception training, and the trained detector returns to the physical robot. The scene drawing is schematic; photographs show the LoCoBot and a corridor run.

10th Conference on Robot Learning (CoRL 2026), Austin TX, USA.

2 Radar Simulator and Evaluation Protocol

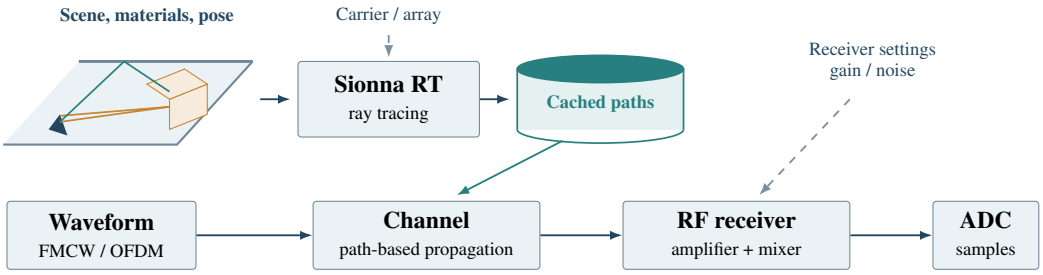


Figure 2: **Physics-based radar simulator.** Scene geometry, materials, pose, and array configuration determine ray-traced paths. Cached paths drive the propagation channel; waveform and receiver models produce ADC samples. Labels come from the same scene.

Scene generation and radar synthesis: Inputs are real video, a robot map, or video from a prompt-conditioned diffusion transformer [4, 5] trained on nuScenes, Waymo, Argoverse 2, and BDD100K [6, 7, 8, 9]. We lift observations into a coarse 3D layout and vary obstacles and trajectories to generate hard scenes with matched labels. The *world model* couples this scene generator to Sionna RT [10], cached propagation paths, and a configurable waveform/receiver/ADC chain. Re-running a frame takes 0.7–2.2 s versus 74–96 s for re-tracing, enabling offline data generation.

Physical-robot protocol: The LoCoBot uses a TI IWR1443BOOST radar [11]: 76–81 GHz, three transmitters and four receivers in TDM-MIMO, yielding 12 virtual channels. We evaluate SSMRadNet trained with real radar only, generated radar only, or both. Each condition has 100 navigation trials in **unseen corridors containing boxes and chairs**. The navigation policy is identical across conditions. We report successful runs and runs with a collision as separate outcomes. We also evaluate FFT-RadNet [12], T-FFTRadNet [13], and SSMRadNet on real RADIal test data using AP and F1.

3 Results

Table 1: **RADIal test detection.** AP/F1 (%). Both training conditions are from this study.

	Real only		+ World model	
Detector	AP	F1	AP	F1
FFT-RadNet	75	77	91	93
T-FFTRadNet	77	79	92	**95**
SSMRadNet	78	80	**94**	**95**

Table 2: **LoCoBot navigation.** SSMRadNet, 100 trials per condition. Collisions count affected runs.

Training data	Success	Collisions
Real only	87/100 (87%)	8
World model only	83/100 (83%)	6
World model + real	**95/100 (95%)**	**2**

World-model-assisted training improves AP and F1 by 15–16 points across detectors (Table 1). On LoCoBot, combined training raises success from 87% to 95% and reduces collision runs from eight to two; generated-only training reaches 83% success (Table 2). This demonstrates perception transfer under a fixed navigation policy.

Scope: Results cover one robot in indoor corridors. Coarse reconstruction can affect radar fidelity, and the gains do not isolate hard-scene selection from additional training data. Broader environments and controlled ablations remain future work.

References

[1] M. Torne, A. Simeonov, Z. Li, A. Chan, T. Chen, A. Gupta, and P. Agrawal. Reconciling reality through simulation: A real-to-sim-to-real approach for robust manipulation. In *Robotics: Science and Systems (RSS)*, 2024.

[2] A. Sen, M. S. Mohammad, and S. Mukhopadhyay. SSMRadNet: A sample-wise state-space framework for efficient and ultra-light radar segmentation and object detection. In *IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 2026.

[3] A. Murali, T. Chen, K. V. Alwala, D. Gandhi, L. Pinto, S. Gupta, and A. Gupta. PyRobot: An open-source robotics framework for research and benchmarking. *arXiv preprint arXiv:1906.08236*, 2019.

[4] W. Peebles and S. Xie. Scalable diffusion models with transformers. In *IEEE/CVF International Conference on Computer Vision (ICCV)*, 2023.

[5] Wan Team. Wan: Open and advanced large-scale video generative models. *arXiv preprint arXiv:2503.20314*, 2025.

[6] H. Caesar, V. Bankiti, A. H. Lang, S. Vora, V. E. Liong, Q. Xu, A. Krishnan, Y. Pan, G. Baldan, and O. Beijbom. nuScenes: A multimodal dataset for autonomous driving. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020.

[7] P. Sun, H. Kretzschmar, X. Dotiwalla, A. Chouard, V. Patnaik, P. Tsui, J. Guo, Y. Zhou, Y. Chai, B. Caine, et al. Scalability in perception for autonomous driving: Waymo open dataset. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020.

[8] B. Wilson, W. Qi, T. Agarwal, J. Lambert, J. Singh, S. Khandelwal, B. Pan, R. Kumar, A. Hartnett, J. K. Pontes, D. Ramanan, P. Carr, and J. Hays. Argoverse 2: Next generation datasets for self-driving perception and forecasting. In *NeurIPS Datasets and Benchmarks Track*, 2021.

[9] F. Yu, H. Chen, X. Wang, W. Xian, Y. Chen, F. Liu, V. Madhavan, and T. Darrell. BDD100K: A diverse driving dataset for heterogeneous multitask learning. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020.

[10] J. Hoydis, F. Aït Aoudia, S. Cammerer, M. Nimier-David, N. Binder, G. Marcus, and A. Keller. Sionna RT: Differentiable ray tracing for radio propagation modeling. In *IEEE Globecom Workshops*, 2023.

[11] Texas Instruments. IWR1443: Single-chip 76- to 81-GHz mmwave sensor. `https://www.ti.com/product/IWR1443`. Product specifications, accessed October 2026.

[12] J. Rebut, A. Ouaknine, W. Malik, and P. Pérez. Raw high-definition radar for multi-task learning. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022.

[13] J. Giroux, M. Bouchard, and R. Laganière. T-FFTRadNet: Object detection with Swin vision transformers from raw ADC radar signals. In *IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, 2023.