RADARSIMRAD: FROM CLEAN SUPPORT TO MEASURED RESPONSE: SENSOR-CONDITIONED PAIRED SCHRÖDINGER BRIDGES

Anonymous authors
Paper under double-blind review

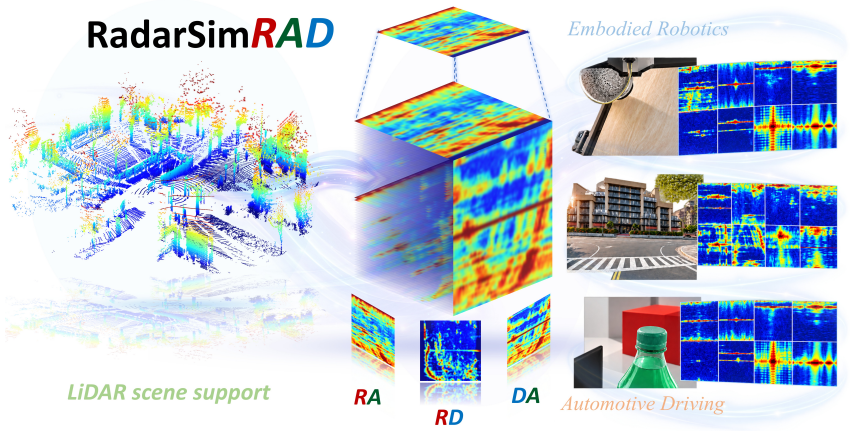


Figure 1: RadarSimRAD generates realistic radar observations from geometric scene support and powers seamless integration with CARLA, RoboTwin, and LIBERO, enabling scalable downstream research in autonomous driving and embodied intelligence.

ABSTRACT

Doppler radar provides complementary geometric and motion cues for autonomous perception, yet large-scale radar data remain costly to acquire at scale. LiDAR offers an attractive source for radar synthesis through accurate geometric support and mature acquisition and annotation pipelines; however, existing simulators either rely on expensive physics-based modeling or use generic learned generation that underexploits source-determined structure and leaves radar-specific measurement formation largely implicit. This limitation is particularly severe for range–azimuth–Doppler (RAD) data, where range, azimuth, and Doppler exhibit distinct response statistics and Doppler ambiguity, often leading generic generators to over-smooth weak responses or hallucinate unsupported structures. We introduce **RadarSimRAD**, a sensor-conditioned paired Schrödinger bridge that learns stochastic transport from an alias-aware clean geometric support endpoint to measured radar response. To model the heterogeneous formation process, we further propose a Radar Spread Adapter to explicitly capture sensor-dependent, non-exchangeable response formation across range, azimuth, and Doppler. Extensive experiments demonstrate that RadarSimRAD substantially improves radar-response fidelity over baselines, reducing axis-gradient error by 55.8% and spectral distance by 38.1% over baselines, while exhibiting zero-shot cross-sensor transfer on independently collected real-world measurements. Generated RAD further improves multiple downstream tasks under limited measured-data availability, demonstrating its practical value for radar learning in data-scarce regimes.

1 INTRODUCTION

Millimeter-wave radar provides Doppler measurements and robust sensing under poor illumination and adverse weather (Paek et al., 2022; Kramer et al., 2022; Sheeny et al., 2021; Hong et al., 2020).

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

Yet learning from range–azimuth–Doppler (RAD) measurements requires specialized hardware, synchronized acquisition, and sensor-specific calibration (Yin et al., 2019; Wise et al., 2023). Coverage across environments, weather, and devices is therefore expensive, while perception models need both geometrically faithful returns and realistic measurement statistics. LiDAR-conditioned RAD synthesis offers a route to scalable, label-aligned augmentation, provided that it reproduces what a radar measures rather than merely rendering scene geometry (Wang et al., 2020; Shen et al., 2026).

This objective exposes two coupled challenges. First, the modalities do not differ only in appearance: LiDAR specifies scene support, while reflectivity, multipath, object motion, sidelobes, and clutter remain underdetermined (Zheng et al., 2024; Wang et al., 2025). Deterministic translation therefore averages plausible measurements, whereas Gaussian-start generation expends capacity rediscovering geometry already present in the condition. Second, RAD is not an isotropic image volume. Range, azimuth, and periodic Doppler encode different physics and resolutions, and their response statistics change with the sensor (Wang et al., 2026a; Xu et al., 2025). Consequently, preserving geometry and forming radar-specific response require different inductive biases.

As shown in Figure 2, our core insight is to make this distinction explicit: *start from what the scene determines, and learn only how the sensor turns that support into a response*. Calibrated LiDAR occupancy and ego motion define an alias-aware clean RAD endpoint with exact RA/DA marginals, but no handcrafted spreading, target power, or background. A stochastic path from this endpoint to measured radar can then focus on the unresolved response. We call this formulation **support-to-response generation**.

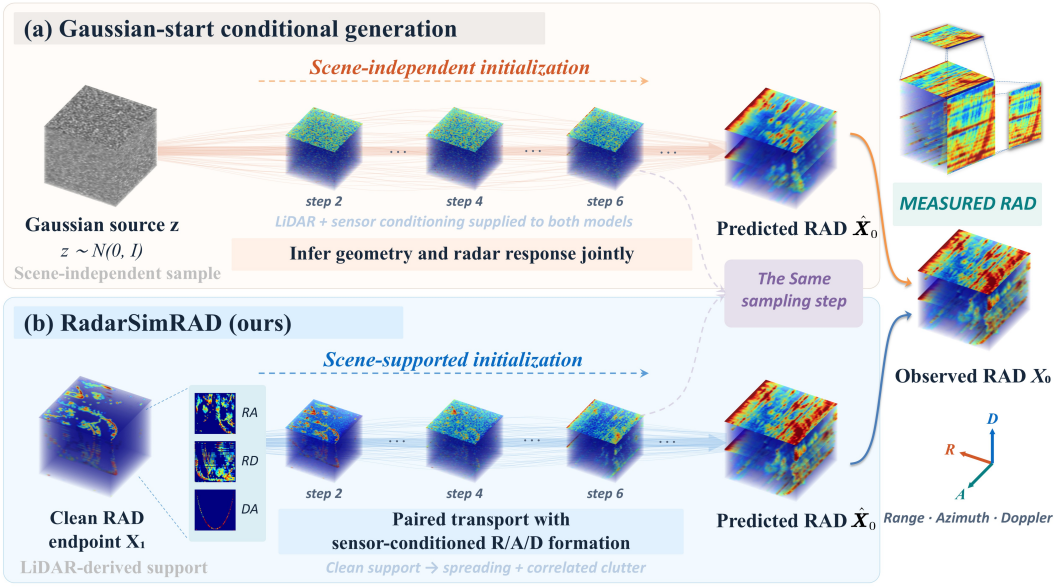


Figure 2: Motivation of RadarSimRAD. Traditional methods initialize from scene-independent noise and jointly infers support and response; RadarSimRAD starts from a physically structured, alias-aware RAD support endpoint derived from LiDAR and learns the residual measurement response.

We instantiate this principle in **RadarSimRAD**, a sensor-conditioned paired Schrödinger bridge. Our contributions are fourfold. 1) To our knowledge, we are the first to cast dense LiDAR-to-Radar RAD synthesis as paired stochastic transport from an alias-aware physical endpoint, preserving spatial–Doppler priors while substantially shortening transport paths. 2) We introduce the Radar Spread Adapter for heterogeneous radar anisotropic response to capture non-exchangeable R/A/D response statistics without eroding source-determined geometry. 3) Extensive experiments across four heterogeneous radar domains demonstrate superior response-formation fidelity over generative baselines. Synthetic RAD improves multiple downstream tasks under scarce measured data. We deploy a TI AWR1843–LiDAR sensing platform, validating zero-shot cross-sensor transfer on independently collected real-world measurements. 4) We integrate RadarSimRAD as a plug-and-play radar synthesis engine into CARLA, LIBERO, and RoboTwin, enabling scalable radar augmentation for embodied simulation and radar foundation-model development. We will release the code, checkpoints, integration interfaces, and evaluation assets.

2 Related Work

Radar simulation. Radar simulation spans physics-based, human-centric, and learned generative paradigms. Classical approaches reproduce propagation and sensor formation through ray tracing, electromagnetic or signal-level modeling (Schüßler et al., 2021; Mock et al., 2025; Fidelis et al., 2023), while RadSimReal incorporates measured radar responses for realistic synthetic-to-real automotive perception (Bialer & Haitman, 2024). For human sensing, Vid2Doppler synthesizes Doppler signatures from video (Ahuja et al., 2021), and SimHumalator generates human micro-Doppler measurements from motion-capture-driven physical simulation (Vishwakarma et al., 2021). RF Genesis combines ray tracing with generative diffusion for scalable cross-modal RF synthesis (Chen & Zhang, 2023), while RF-Diffusion develops time–frequency diffusion for general RF and FMCW signal generation (Chi et al., 2024).

Scene-level radar generation. Recent work targets scene-level automotive radar. L2R GAN translates LiDAR into radar spectra, while diffusion-based methods extend generation to range–azimuth maps and sparse 4D radar point clouds (Wang et al., 2020; 2026b; Kwok et al., 2025). RadarGen generates automotive radar point clouds from multi-view imagery with explicit RCS and Doppler attributes (Borreda et al., 2026). Beyond sparse outputs, L2RLDB translates LiDAR into full 4D radar tensors through latent alignment and a diffusion bridge (Shen et al., 2026), while SA-Radar generates RAD cubes conditioned on waveform-derived sensor attributes (Xiao et al., 2026). Complementary neural-rendering approaches, including DART, RadarSplat, and RF4D, synthesize radar observations from learned scene representations (Huang et al., 2024; Kung et al., 2025; Zhang et al., 2026a).

Despite this progress, existing methods either target partial or sparse radar representations, require scene-specific reconstruction or explicit sensor-response parameterization, or leave dense RAD formation largely implicit. More fundamentally, treating Doppler as an ordinary Euclidean axis obscures its periodic quotient structure: alias-equivalent velocities collapse to the same observation, making the conditional mapping intrinsically many-to-one. Without explicitly resolving this ambiguity and the non-exchangeable geometry of R/A/D, generic generators must absorb motion uncertainty and response formation jointly, often encouraging conditional averaging or structurally inconsistent echoes. RadarSimRAD instead preserves source-derived spatial–Doppler support under sensor-specific Doppler folding and learns the unresolved stochastic response directly in dense RAD space, explicitly modeling the non-exchangeable formation of range, azimuth, and Doppler.

3 Method

As shown in Fig 3, RadarSimRAD separates source-determined support from stochastic radar-response formation. It first constructs an *alias-aware RAD endpoint* from calibrated LiDAR geometry and ego motion in Sec 3.1, and then transports this clean support toward measured RAD through a paired Schrödinger bridge in Sec 3.2. A *Radar Spread Adapter* models sensor-dependent, non-exchangeable R/A/D formation in Sec 3.3, while complementary objectives supervise source-local response spreading and off-support texture in Sec 3.4.

3.1 Alias-Aware RAD Endpoint

We model the conditional law of a normalized measured RAD cube $X_0 \in [-1, 1]^{R\times A\times D}$ given radar-aligned LiDAR conditions $L_{\mathrm{RA}} \in \mathbb{R}^{R\times A}$, $L_{\mathrm{DA}} \in \mathbb{R}^{D\times A}$, and sensor identity s. The paired target is acquired at the same time as the conditions. We lift the deterministic geometry into a clean endpoint X_1. Everything not specified by that endpoint is learned as conditional measurement response as shown in Appendix A.1. Let $O(r, a)$ denote calibrated LiDAR occupancy, $\nu(a)$ the radial velocity induced by ego-motion and object motion, and $m_A(a) = \sum_r O(r, a)$. The occupancy $O(r, a)$ can be readily constructed by frustum-based voxelization of radar-aligned LiDAR points (Zhang et al., 2026b; Wang et al., 2026a), while $\nu(a)$ can be obtained from available flow or motion estimator (Lin & Caesar, 2024). For sensor velocity interval $[-v_{\max}^{(s)}, v_{\max}^{(s)})$, we wrap $\nu(a)$ before nearest-bin

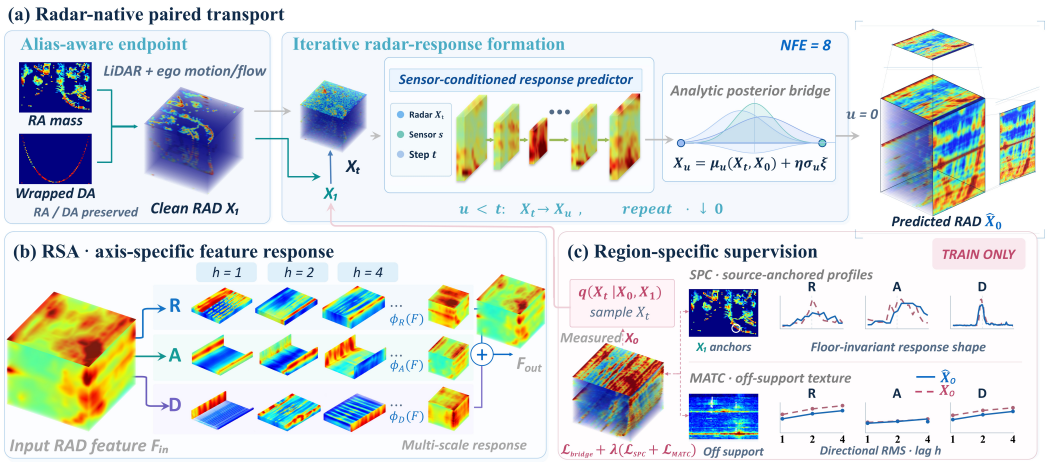


Figure 3: Overview of **RadarSimRAD. (a)** Alias-aware RA/DA lifting supplies deterministic scene support; a paired bridge and sensor-conditioned predictor form the stochastic measured response. **(b)** The scale-16 view shows an actual EMA decoder RSA execution at $t = 0.5$: colored slabs are diagnostic channel-RMS projections of the gated $h \in \{1, 2, 4\}$ branches under one branch scale; dark rails denote the underlying signed tensors on which the dilation mean and residual operate. **(c)** SPC and MATC supervise local spread and off-support texture during training.

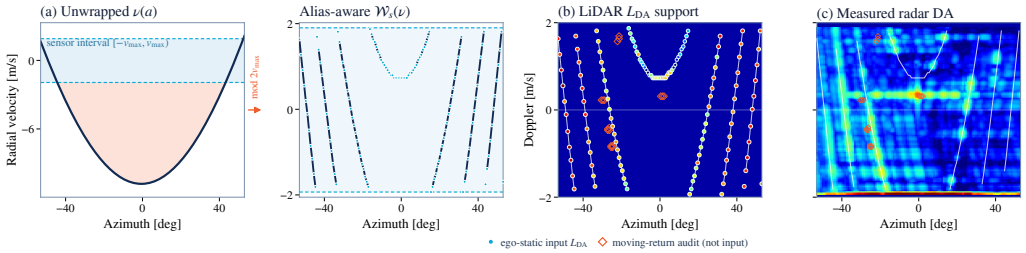


Figure 4: Alias-aware Doppler support construction.

quantization $\mathcal{Q}_s$:

$$\mathcal{W}_s(\nu) = \nu - 2v_{\max}^{(s)} \left\lfloor \frac{\nu + v_{\max}^{(s)}}{2v_{\max}^{(s)}} \right\rfloor, \qquad d_s(a) = \mathcal{Q}_s(\mathcal{W}_s(\nu(a))) \,. \tag{1}$$

Equation 1 maps motion-derived radial velocity into the sensor-specific Doppler quotient space, such that velocities differing by integer multiples of the unambiguous interval correspond to the same observable Doppler state. As illustrated in Figure 4, this construction explicitly resolves the periodic folding inherent to FMCW Doppler measurements, avoiding artificial discontinuities or inconsistent supervision near ambiguity boundaries. The resulting DA support therefore preserves the physically observable motion structure, including both ego- and object-induced components before it is lifted into the clean RAD endpoint.

The resulting source conditions are $L_{\mathrm{RA}}(r, a) = O(r, a)$ and $L_{\mathrm{DA}}(d, a) = m_A(a)\mathbf{1}[d = d_s(a)]$, encoding geometric occupancy and alias-aware motion support, respectively. The raw endpoint is

$$X_1^{\mathrm{raw}}(r, a, d) = \begin{cases} L_{\mathrm{RA}}(r, a)L_{\mathrm{DA}}(d, a)/M(a), & M(a) > 0, \\ 0, & \text{otherwise}, \end{cases} \qquad M(a) = \sum_d L_{\mathrm{DA}}(d, a). \tag{2}$$

We obtain the normalized bridge endpoint as $X_1 = \mathcal{N}_{s,\mathrm{src}}(X_1^{\mathrm{raw}})$, using the sensor-specific source calibration defined in Appendix F.1. Summing over Doppler or range exactly recovers the respective input marginal. Invalid-range voxels are masked throughout training and evaluation.

3.2 Support-Anchored Stochastic Transport

Once the source-determined support is fixed in X_1, radar synthesis reduces to modeling the unresolved stochastic response around that support. Rather than restarting from scene-independent noise, we

therefore learn a paired transport from X_1 to the measured RAD X_0, concentrating generative capacity on unresolved measurement formation rather than rediscovering source-provided structure. Following I^{2}SB (Liu et al., 2023), we condition a Brownian reference with variance rate $\beta(t)$ on the observed pair (X_0, X_1). Writing $B_t = \int_0^t \beta(\tau)d\tau$ and $\bar{B}_t = B_1 - B_t$ gives

$$q(X_t \mid X_0, X_1) = \mathcal{N}\left(\frac{\bar{B}_t}{B_1}X_0 + \frac{B_t}{B_1}X_1,\ \frac{B_t\bar{B}_t}{B_1}I\right). \tag{3}$$

Since X_1 already captures source-determined spatial–Doppler structure, the bridge only needs to model the unresolved measurement response. Its conditional mean is the minimum-action path, $\mathcal{A}(X_0, X_1) = \|X_1 - X_0\|_2^2/(2B_1)$, making a physically informed endpoint substantially closer to the target than a scene-independent Gaussian source in our normalized state space. This converts RAD generation from unconstrained synthesis into a shorter, support-anchored stochastic transport problem. Derivations and empirical validation are given in Appendices F and B.1.

As in diffusion-based generative models (Ho et al., 2020; Song et al., 2021; Lipman et al., 2022), support-anchored stochastic transport consists of two complementary stages: learning endpoint prediction from intermediate bridge states and progressively transporting the constructed support toward the radar measurement distribution at inference.

Training. Training samples one of 1,000 nonzero bridge indices and draws X_t directly from equation 3. With $C = (X_1, L_{\mathrm{RA}}, L_{\mathrm{DA}}, s)$, the predictor uses

$$\epsilon_t^\star = \frac{X_t - X_0}{\sqrt{B_t}}, \qquad \widehat{X}_0 = X_t - \sqrt{B_t}\,\epsilon_\theta(X_t, C, t). \tag{4}$$

Sampling. For reverse times $0 \le u < t \le 1$, Gaussian conditioning gives the update

$$X_u = \frac{B_t - B_u}{B_t}\widehat{X}_0 + \frac{B_u}{B_t}X_t + \eta\sqrt{\frac{B_u(B_t - B_u)}{B_t}}\,\xi, \qquad \xi \sim \mathcal{N}(0, I). \tag{5}$$

Sampling begins at the observed X_1 and ends at $t = 0$. We use $\eta = 1$ and eight number-of-function evaluations (NFE). $\eta = 0$ is a deterministic diagnostic. Condition encodings are cached, predicted endpoints are clipped to $[-1, 1]$. More details are provided in Appendix F.

3.3 Stochastic Response: Sensor-Conditioned Anisotropy

Range, azimuth, and Doppler are physically non-exchangeable, motivating a predictor that preserves their distinct response structure rather than treating RAD as an isotropic volume. We adopt a four-stage 3D U-Net with widths $48/96/192/384$, where separate RA and DA encoders construct multi-scale condition pyramids and inject them into the volumetric stream through their shared azimuth axis. Sensor identity and bridge time modulate residual blocks via FiLM, decoupling hardware-dependent response statistics from transport-time dynamics while retaining a single predictor across radar domains. A low-rank sensor-style residual captures global floor and contrast variation, while the Radar Spread Adapter (RSA) introduces axis-specific response capacity at 32^3, 16^3, and 8^3 resolutions. For feature F, physical axis $q \in \{R, A, D\}$, and dilation $h \in \{1, 2, 4\}$, RSA applies learned depth-wise length-five kernels $K_{q,h}$:

$$\mathrm{RSA}(F; s, t) = F + \frac{1}{3}\sum_{q,h} g_{q,h}(s, t)\,\mathcal{S}(K_{q,h} *_q F), \qquad \mathcal{S}(U) = \frac{U}{\sqrt{1 + \mathrm{RMS}(U)^2}}. \tag{6}$$

Centered sensor–time gates $g_{q,h}$ and zero-initialized kernels make RSA a bounded identity-initialized residual. The separable branches expose axis-dependent response scales without prescribing a fixed radar point-spread function. Appendix A.4 gives its axis-wise frequency interpretation and bounds.

3.4 Formation-Aware Training Objective

Voxel regression can preserve scene layout while smoothing unresolved radar background. We therefore supervise local response shape around source support and directional texture outside it with complementary statistics.

Source-anchored spread profiles. Spread Profile Consistency (SPC) selects RA anchors $\mathcal{A}(X_1^{\text{raw}})$ only from source-supported local maxima. At each anchor α, range and azimuth profiles $\ell_{q,\alpha}$ are radius-K cuts through the radar maximum-over-Doppler projection. The Doppler profile instead spans the full circular axis, max-pooled over a 3×3 RA neighborhood and shifted to source-relative Doppler coordinates. This construction retains off-center moving returns even when the source specifies only ego-static velocity (Appendix F.4). We subtract each line's lower quartile, apply softplus, and normalize to a probability vector $p(\ell)$, and the exact transform is in equation 46. The loss is

$$\mathcal{L}_{\text{profile}} = \mathbb{E}_{\alpha \in \mathcal{A}(X_1^{\text{raw}})} \frac{1}{3} \sum_q \left\| p(\ell_{q,\alpha}^{\text{pred}}) - p(\ell_{q,\alpha}^{\text{gt}}) \right\|_1 . \tag{7}$$

All axes use source-selected anchors, without target-peak selection. Consequently, SPC supervises how response forms around known support rather than discovering target peaks. Its floor invariance and moment bounds are derived in Appendix A.5.

Off-support anisotropic texture. Masked Anisotropic Texture Consistency (MATC) uses valid voxels M_{bg} with $X_1^{\text{raw}} = 0$. For each axis q and lag $h \in \{1, 2, 4\}$, define

$$E_{q,h}(\ell) = \left[\frac{\sum_{i,\, i+he_q \in M_{\text{bg}}} (\ell(i + he_q) - \ell(i))^2}{\max(1, N_{q,h})} + \varepsilon^2 \right]^{1/2}, \tag{8}$$

where $N_{q,h}$ counts valid voxel pairs. MATC matches log energies:

$$\mathcal{L}_{\text{texture}} = \frac{1}{9} \sum_{q,h} \text{SmoothL1}_{0.25} \big(\log E_{q,h}(\ell^{\text{pred}}) - \log E_{q,h}(\ell^{\text{gt}}) \big) . \tag{9}$$

The logarithm compares multiplicative excess and deficit symmetrically, while separate axes prevent cancellation and multiple lags constrain different spectral bands. MATC therefore teaches off-support response organization complementary to SPC's local profiles. Appendix A.6 gives its covariance and spectral interpretation.

With valid-volume mask M, let $\mathcal{L}_{\text{bridge}} = \|M \odot (\epsilon_\theta - \epsilon_t^\star)\|_2^2 / \|M\|_1$. The complete objective is

$$\mathcal{L} = \mathcal{L}_{\text{bridge}} + \lambda(k)(\mathcal{L}_{\text{profile}} + \mathcal{L}_{\text{texture}}), \qquad \lambda(k) = 0.05 \min(1, k/500). \tag{10}$$

The auxiliary weight warms up over 500 optimizer steps. SPC and MATC operate on the predicted endpoint during training and add no inference passes.

4 Experiments

4.1 Experimental Setup

We train RadarSimRAD with a four-stage 3D predictor of widths $48/96/192/384$. We use EMA weights, stochastic sampling with NFE$= 8$. Radar measurements and LiDAR-derived support are co-registered onto a unified 64^3 RAD lattice. All experiments were conducted on a Linux server running Ubuntu 22.04.5 LTS with kernel 6.8.0. The server is equipped with two AMD EPYC 9555 64-core CPUs, providing 256 CPU threads in total, 755 GB of system memory, and eight NVIDIA L40S GPUs. Training the full model for 14.5k optimizer steps took approximately 1.5 days. All calibration statistics are estimated from training sequences only. Architecture, optimization, sampling, and split details are provided in Appendix G.

Datasets. We evaluate four heterogeneous radar configurations: K-Radar (Paek et al., 2022), RaDelft (Roldan et al., 2024), and the cascade and single-chip configurations of ColoRadar (Kramer et al., 2022). Across the four heterogeneous radar configurations, raw ADC data or radar cubes are accompanied by aligned LiDAR and motion annotations (e.g., 3D boxes or odometry), allowing us to derive both the geometric support and paired RAD targets used for RadarSimRAD training. For zero-shot transfer, we additionally use an independently acquired indoor sequence from a TI AWR1843 radar and a Hesai JT128 LiDAR. This sensor and scene are absent from training.

Table 1: Quantitative comparison on RAD generation quality over the same held-out test set. All methods are evaluated using identical frame identities and evaluation protocols. Complete metrics and per-domain results are reported in the appendix.

Method	Corr ↑	Peak-F1 ↑	Spread ↓	RA Corr ↑	AGLRE ↓
L2RLDB 2026	0.2104 ↓70.3%	0.1135 ↓63.4%	0.7089 ↑30.7%	0.4571 ↓33.9%	1.0974 ↑301.4%
Conditional DDPM	0.6568 ↓7.2%	0.2349 ↓24.3%	0.5898 ↑8.8%	0.5966 ↓13.7%	0.6216 ↑127.4%
Flow matching	0.7020 ↓0.8%	0.2884 ↓7.1%	0.5606 ↑3.4%	0.6683 ↓3.4%	0.5311 ↑94.3%
RadarSimRAD	**0.7077**	**0.3104**	**0.5423**	**0.6915**	**0.2734**

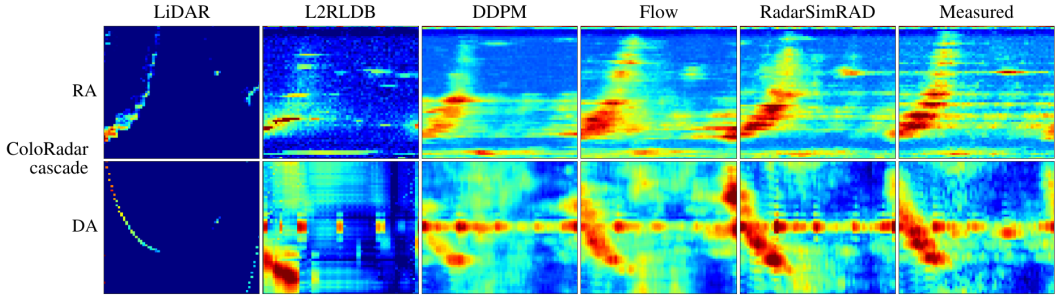


Figure 5: Qualitative comparison of RA and DA synthesis on the ColoRadar cascade dataset. Additional visualizations and details are provided in Figs. 8 and 9.

Baselines. We conduct Conditional DDPM (Ho et al., 2020) and flow matching (Lipman et al., 2022) using the same 3D predictor, RA/DA encoders, sensor conditioning, preprocessing, and data splits as RadarSimRAD. Strict ablations remove MATC or RSA while retaining the paired bridge and all other settings. We additionally compare against L2RLDB (Shen et al., 2026) on dense-RAD validation and qualitative results. Due to its distinct representation, training pipeline, and 50-step inference, it is evaluated as an external reference rather than a compute-matched baseline. Details are given in Appendix E.1.

Metrics. Corr (ρ) and RA Corr quantify volumetric and range–azimuth agreement. Peak-F1 measures exact overlap among strong-response voxels, and Spread compares source-anchored R/A/D response profiles. AGLRE (G) measures off-support directional-gradient mismatch, while PSD-W (W) compares normalized non-DC background spectra. Figure 8 abbreviates these cube-level scores as $\rho/G/W$. Corr, RA Corr, and Peak-F1 are higher-is-better. All remaining metrics are lower-is-better. Exact definitions, MAE and supplementary statistics, and sequence-level significance tests appear in Appendix G.2.

4.2 RAD Generation Quality

Quantitative results. Table 1 shows that RadarSimRAD provides its largest and most consistent gains in radar-response formation, while maintaining competitive voxel-level fidelity. Relative to flow matching, Corr and Peak-F1 rise to 0.7077 and 0.3104, while AGLRE falls from 0.5311 to 0.2734. Against L2RLDB, RadarSimRAD improves Corr by 0.4973 and reduces AGLRE by 75.1%, demonstrating substantially stronger volumetric agreement and directional response fidelity. These joint gains indicate that stronger support agreement does not come at the expense of radar-specific anisotropy or background structure.

Qualitative results. Figure 5 shows that RadarSimRAD preserves LiDAR-supported scene layout while recovering domain-specific DA branches, localized spreading, and off-support texture. Unlike L2RLDB, the generated responses remain structured in both RA and DA, and unlike DDPM or flow matching, they more closely reproduce the measured directional organization. The consistency across four heterogeneous domains supports the proposed separation between deterministic support and sensor-dependent response formation. Additional examples and details appear in Figure 8, 9.

Table 2: Downstream utility across measured-RAD availability. Percentage suffixes report relative performance changes with respect to the corresponding $\text{Real}_{p=100}$ reference.

	Translation ↓			Ego-speed ↓			Occupancy ↑			Localization ↑		
p (%)	Real	Mix	Gen	Real	Mix	Gen	Real	Mix	Gen	Real	Mix	Gen
10	0.188↓51.6	0.101↑18.5	0.205↓65.3	0.193↓1.6	**0.148**↑22.1	0.240↓26.3	0.211↓35.1	0.286↓12.0	0.177↓45.5	0.769↓0.5	0.773	0.759↓1.8
20	0.180↓45.2	0.099↑20.2	0.180↓45.2	0.179↑5.8	0.173↑8.9	0.191↓0.5	0.213↓34.5	0.298↓8.3	0.181↓44.3	0.758↓1.9	0.770↓0.4	0.778↑0.6
30	**0.088**↑29.0	**0.091**↑26.6	0.209↓68.5	0.212↓11.6	0.186↑2.1	0.189↑0.5	0.286↓12.0	0.309↓4.9	0.235↓27.7	0.786↑1.7	0.772↓0.1	0.764↓1.2
40	0.204↓64.5	0.197↓58.9	0.188↓51.6	0.180↑5.3	0.168↑11.6	0.161↑15.3	0.267↓17.8	0.284↓12.6	0.263↓19.1	**0.790**↑2.2	**0.784**↑1.4	0.760↓1.7
50	0.194↓56.5	0.177↓42.7	0.205↓65.3	**0.175**↑7.9	0.171↑10.0	0.146↑23.2	0.293↓9.8	0.305↓6.2	0.261↓19.7	0.762↓1.4	0.769↓0.5	0.770↓0.4
60	0.190↓53.2	0.109↑12.1	**0.084**↑32.3	0.191↓0.5	0.165↑13.2	**0.124**↑34.7	0.285↓12.3	0.322↓0.9	**0.271**↓16.6	0.762↓1.4	0.778↑0.6	0.761↓1.6
70	0.155↓25.0	0.202↓62.9	0.177↓42.7	0.177↑6.8	0.165↑13.2	0.193↓1.6	0.310↓4.6	0.312↓4.0	0.249↓23.4	0.770↓0.4	0.779↑0.8	0.781↑1.0
80	0.180↓45.2	0.198↓59.7	0.182↓46.8	0.191↓0.5	0.213↓12.1	0.186↑2.1	0.287↓11.7	0.303↓6.8	0.243↓25.2	0.766↓0.9	0.779↑0.8	0.769↓0.5
90	0.201↓62.1	0.190↓53.2	0.150↓21.0	0.197↓3.7	0.186↑2.1	0.163↑14.2	0.306↓5.8	0.310↓4.6	0.261↓19.7	0.768↓0.6	0.782↑1.2	**0.789**↑2.1
100	0.124	0.124	0.106↑14.5	0.190	0.190	0.149↑21.6	**0.325**	**0.325**	0.240↓26.2	0.773	0.773	0.771↓0.3

Ablation studies. We ablate MATC and RSA within the same paired-transport framework to isolate their contributions to radar-specific response formation. Removing MATC increases AGLRE from 0.2750 to 0.5238, showing that off-support texture constraints are essential for recovering directional response organization. Removing RSA raises AGLRE to 0.2948 and degrades axis-dependent agreement, confirming that multi-scale, axis-specific modulation complements the distributional constraint imposed by MATC. Together, these results establish that MATC governs response statistics outside LiDAR support, whereas RSA provides the directional capacity required to represent radar anisotropy, with complete metrics and per-domain ablations reported in Appendix G.

Computational efficiency. At NFE$=2$, RadarSimRAD requires only 54.6 ms/frame yet outperforms flow matching at NFE$=8$ in Corr (0.7455 vs. 0.7013), Peak-F1 (0.3193 vs. 0.2877), and AGLRE (0.3289 vs. 0.5326). This operating point uses one quarter of the function evaluations and reduces latency by 62.9%, demonstrating that the proposed bridge achieves stronger response formation under a substantially smaller sampling budget. Increasing RadarSimRAD to NFE$=8$ further improves directional agreement, providing a flexible quality–efficiency trade-off without retraining. More details are in Appendix G.4, E.4.

4.3 Downstream Utility of Generated RAD

We evaluate generated RAD on odometry, ego-speed estimation, RA occupancy, and foreground localization. Task predictors are trained on K-Radar sequences with richer task-level annotations and evaluated exclusively on measured RAD. Complete secondary task metrics appear in Appendix I.3.

Table 2 shows that the benefit of generated RAD is task-dependent. For ego-speed estimation, increasing real-data availability does not monotonically reduce error, suggesting rapid saturation of this low-dimensional motion cue and sensitivity to realization-specific radar variability. Synthetic augmentation alleviates this effect: Mix10 reduces MAE from 0.193 to 0.148, while Gen100 achieves 0.149 versus 0.190 for Real_{100}. This suggests that response diversity helps suppress finite-sample bias toward incidental clutter and spectral patterns, promoting motion-stable representations.

The gains are more systematic for spatial tasks. For occupancy, Mix consistently outperforms Real across 10–90% availability, improving F1 from 0.211/0.213 to 0.286/0.298 at 10%/20%. Translation shows even larger gains, with Mix reducing MAE by 46.3%, 45.0%, and 42.6% at 10%, 20%, and 60%, respectively; Gen60 further reaches 0.084. Localization follows the same trend: Mix surpasses Real in seven of nine scarce-data settings, while Gen90 reaches 0.789, exceeding the Real_{100} reference of 0.773. These results suggest that RadarSimRAD acts as a support-aligned augmentation mechanism: it expands geometry–motion coverage while preserving realistic measurement variability, thereby improving downstream learning under limited real RAD supervision.

4.4 Zero-Shot Cross-Domain Transfer

We evaluate the frozen model under simultaneous sensor and scene shift using an independently collected TI AWR1843–Hesai JT128 sequence. Inference uses the existing sensor condition without target-scene adaptation. Acquisition and synchronization details appear in Appendix H.1.

Zero-shot transfer to an unseen radar–LiDAR platform in Fig. 6 further demonstrates that the learned response model generalizes beyond the sensors and environments seen during training. Without any adaptation, the frozen model raises RA Corr from 0.3074 to 0.5623 while reducing RA-JSD and RA-TV by 30.0% and 14.2%, respectively; RA-JSD improves on 97.3% of frame pairs, indicating consistent radar-response formation under simultaneous sensor and scene shifts.

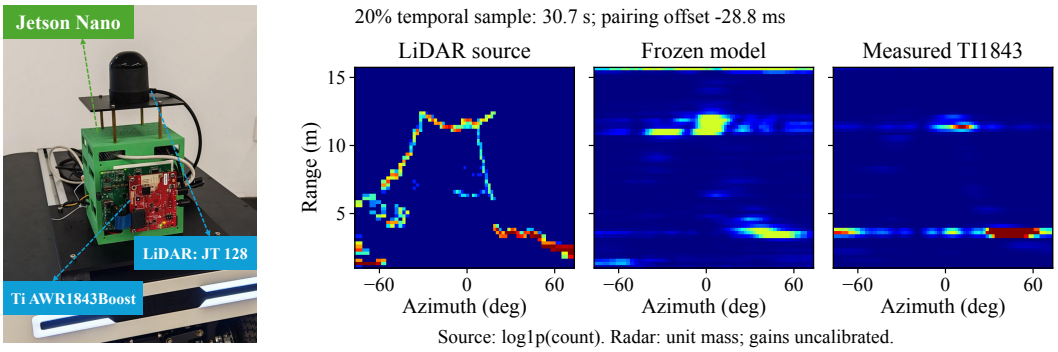


Figure 6: **Zero-shot transfer to an unseen radar–LiDAR platform.** The physical rig and source/generated/measured RA projections are shown at a fixed sequence position. Additional cases are provided in Figures 7 and 10.

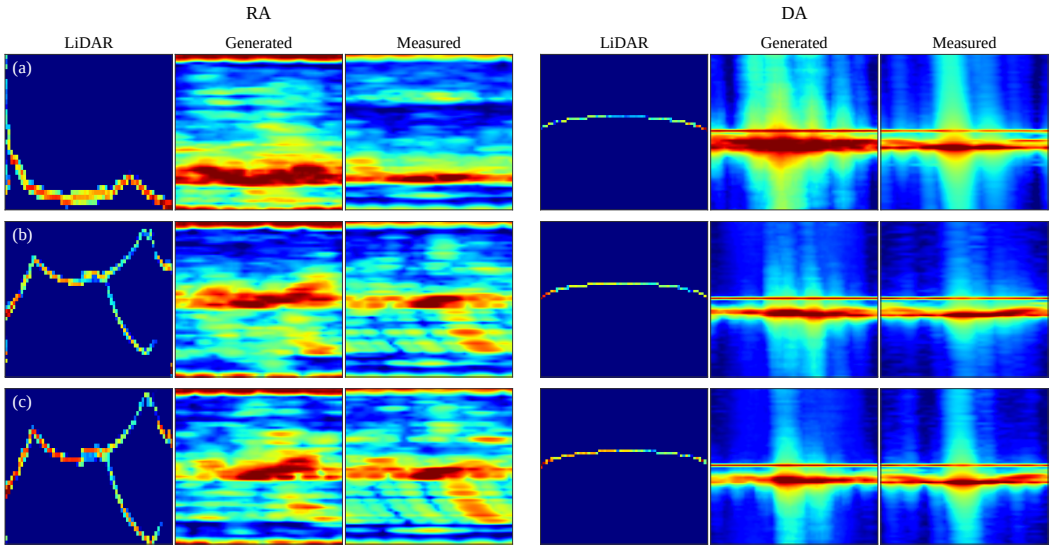


Figure 7: **Additional zero-shot cases from the independent acquisition.** Each row shows RA and DA projections of the LiDAR-derived source, the frozen model output, and the paired measured TI AWR1843 RAD. Generated results use the frozen model without target-scene adaptation, while measured panels show the synchronized radar acquisition. Additional visualization is in H.3

5 Conclusion

In this paper, we introduce RadarSimRAD which firstly formulates dense RAD synthesis as explicit support-to-response generation: preserve what LiDAR determines, then learn the stochastic, sensor-dependent measurement formation. RadarSimRAD realizes this principle with an alias-aware paired bridge and anisotropic response modeling. Experiments demonstrate strong response fidelity, downstream utility, and zero-shot cross-sensor transfer, highlighting RadarSimRAD's potential for scalable radar foundation models and embodied perception.

AI USE STATEMENT

In this work, we used generative AI tools for language polishing, figure refinement, and limited assistance with code debugging and LaTeX formatting. We did not use generative AI to generate experimental data or results. All AI-assisted content was reviewed, revised, and verified by the authors, who take full responsibility for the final content of this work.

REFERENCES

Karan Ahuja, Yue Jiang, Mayank Goel, and Chris Harrison. Vid2doppler: Synthesizing doppler radar data from videos for training privacy-preserving activity recognition. In *Proceedings of the 2021 CHI conference on human factors in computing systems*, pp. 1–10, 2021.

Oded Bialer and Yuval Haitman. Radsimreal: Bridging the gap between synthetic and real data in radar object detection with simulation. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 15407–15416. IEEE, 2024.

Tomer Borreda, Fangqiang Ding, Sanja Fidler, Shengyu Huang, and Or Litany. Radargen: Automotive radar point cloud generation from cameras. In *European Conference on Computer Vision*, pp. 628–646. Springer, 2026.

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*, pp. 28–42, 2023.

Guoxuan Chi, Zheng Yang, Chenshu Wu, Jingao Xu, Yuchong Gao, Yunhao Liu, and Tony Xiao Han. Rf-diffusion: Radio signal generation via time-frequency diffusion. In *Proceedings of the 30th Annual International Conference on Mobile Computing and Networking*, pp. 77–92, 2024.

Eduardo C Fidelis, Fabio Reway, Herick Ribeiro, Pietro L Campos, Werner Huber, Christian Icking, Lester A Faria, and Torsten Schön. Generation of realistic synthetic raw radar data for automated driving applications using generative adversarial networks. *arXiv preprint arXiv:2308.02632*, 2023.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020.

Ziyang Hong, Yvan Petillot, and Sen Wang. Radarslam: Radar based large-scale slam in all weathers. In *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 5164–5170. IEEE, 2020.

Tianshu Huang, John Miller, Akarsh Prabhakara, Tao Jin, Tarana Laroia, Zico Kolter, and Anthony Rowe. Dart: Implicit doppler tomography for radar novel view synthesis. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 24118–24129. IEEE, 2024.

Andrew Kramer, Kyle Harlow, Christopher Williams, and Christoffer Heckman. Coloradar: The direct 3d millimeter wave radar dataset. *The International Journal of Robotics Research*, 41(4): 351–360, 2022.

Pou-Chun Kung, Skanda Harisha, Ram Vasudevan, Aline Eid, and Katherine A Skinner. Radarsplat: Radar gaussian splatting for high-fidelity data synthesis and 3d reconstruction of autonomous driving scenes. In *2025 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 27596–27606. IEEE, 2025.

Jimmie Kwok, Holger Caesar, and Andras Palffy. 4d-radiff: Latent diffusion for 4d radar point cloud generation. *arXiv preprint arXiv:2512.14235*, 2025.

Yancong Lin and Holger Caesar. Icp-flow: Lidar scene flow estimation with icp. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 15501–15511. IEEE, 2024.

Yaron Lipman, Ricky TQ Chen, Heli Ben-Hamu, Maximilian Nickel, and Matt Le. Flow matching for generative modeling. *arXiv preprint arXiv:2210.02747*, 2022.

Guan-Horng Liu, Arash Vahdat, De-An Huang, Evangelos Theodorou, Weili Nie, and Anima Anandkumar. I^2SB: Image-to-image schrödinger bridge. In Andreas Krause, Emma Brunskill, Kyunghyun Cho, Barbara Engelhardt, Sivan Sabato, and Jonathan Scarlett (eds.), *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 22042–22062. PMLR, 23–29 Jul 2023. URL https://proceedings.mlr.press/v202/liu23ai.html.

Alexander Mock, Martin Magnusson, and Joachim Hertzberg. Radarays: Real-time simulation of rotating fmcw radar for mobile robotics via hardware-accelerated ray tracing. *IEEE Robotics and Automation Letters*, 10(3):2470–2477, 2025.

Dong-Hee Paek, Seung-Hyun Kong, and Kevin Tirta Wijaya. K-radar: 4d radar object detection for autonomous driving in various weather conditions. *Advances in Neural Information Processing Systems*, 35:3819–3829, 2022.

Ignacio Roldan, Andras Palffy, Julian FP Kooij, Dariu M Gavrila, Francesco Fioranelli, and Alexander Yarovoy. A deep automotive radar detector using the radelft dataset. *IEEE Transactions on Radar Systems*, 2:1062–1075, 2024.

Christian Schüßler, Marcel Hoffmann, Johanna Bräunig, Ingrid Ullmann, Randolf Ebelt, and Martin Vossiek. A realistic radar ray tracing simulator for large mimo-arrays in automotive environments. *IEEE Journal of Microwaves*, 1(4):962–974, 2021.

Marcel Sheeny, Emanuele De Pellegrin, Saptarshi Mukherjee, Alireza Ahrabian, Sen Wang, and Andrew Wallace. Radiate: A radar dataset for automotive perception in bad weather. In *2021 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 1–7. IEEE, 2021.

Dazhong Shen, Jingjing Gu, Qiang Zhou, Meng Zhao, and Ying Sun. Lidar-to-4dradar diffusion bridge via cross-modal alignment and translation in latent space. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 17110–17120, 2026.

Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models. In *International Conference on Learning Representations*, 2021. URL https://openreview.net/forum?id=St1giarCHLP.

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.

Leichen Wang, Bastian Goldluecke, and Carsten Anklam. L2r gan: Lidar-to-radar translation. In *Asian Conference on Computer Vision*, pp. 704–720. Springer, 2020.

Shengpeng Wang, Kuangyu Wang, and Wei Wang. Rauf: Learning the spatial uncertainty field of radar. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 42474–42483, June 2026a.

Wei Wang, Chunshen Li, Bixin Zeng, Lieke Chen, Liang Sun, Kai Luo, and Da Chen. Mitigating interference for automotive millimeter-wave radar perception in dense traffic scenarios. *IEEE Transactions on Mobile Computing*, 2025.

Zhaoze Wang, Changxu Zhang, Tai Fei, Christopher Grimm, Yi Jin, Claas Tebruegge, Ernst Warsitz, and Markus Gardill. Synthetic fmcw radar range azimuth maps augmentation with generative diffusion model. *arXiv preprint arXiv:2601.06228*, 2026b.

Emmett Wise, Qilong Cheng, and Jonathan Kelly. Spatiotemporal calibration of 3-d millimetre-wavelength radar-camera pairs. *IEEE Transactions on Robotics*, 39(6):4552–4566, 2023.

Weiqing Xiao, Hao Huang, Chonghao Zhong, Yujie Lin, Nan Wang, Xiaoxue Chen, Zhaoxi Chen, Saining Zhang, Shuocheng Yang, Pierre Merriaux, et al. Controllable radar simulation with waveform parameter embedding. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 6424–6434, 2026.

Yang Xu, Qiucan Huang, Shaojie Shen, and Huan Yin. Incorporating point uncertainty in radar slam. *IEEE Robotics and Automation Letters*, 10(3):2168–2175, 2025.

Jiapeng Yin, Peter Hoogeboom, Christine Unal, Herman Russchenberg, Fred Van Der Zwan, and Erik Oudejans. Uav-aided weather radar calibration. *IEEE Transactions on Geoscience and Remote Sensing*, 57(12):10362–10375, 2019.

Jiarui Zhang, Zhihao Li, Chong Wang, and Bihan Wen. Rf4d: Neural radar fields for novel view synthesis in outdoor dynamic scenes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 15387–15397, 2026a.

Ruijie Zhang, Bixin Zeng, Shengpeng Wang, Fuhui Zhou, and Wei Wang. Rald: Generating high-resolution 3d radar point clouds with latent diffusion. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 18773–18781, 2026b.

Le Zheng, Jiamin Long, Marco Lops, Fan Liu, Xueyao Hu, and Chuanhao Zhao. Detection of ghost targets for automotive radar in the presence of multipath. *IEEE Transactions on Signal Processing*, 72:2204–2220, 2024.

APPENDIX CONTENTS

A Mathematical and Radar-Physical Foundations . . . 13
Radar Observability from LiDAR Support . . . 13
Clean Endpoint as a Least-Committed Coupling . . . 14
LiDAR-Supported RAD Transport . . . 15
Radar Spread Adapter as an Anisotropic Response Operator . . . 15
Why SPC Measures Response Shape . . . 15
MATC as a Directional Variogram and Spectral Constraint . . . 16
B Empirical Evidence for Support-Anchored Transport . . . 16
LiDAR Support Reduces Path Length and Source Variance . . . 16
Repeated-Sample Conditional Validation . . . 17
C Implementation Details . . . 18
D Per-Domain Results . . . 19
Additional Curated Visualizations . . . 19
E Matched L2RLDB Evaluation . . . 19
Matched-Frame Protocol and Representation Conversion . . . 19
Aggregate Validation Results . . . 20
Formation Metrics Independent of MATC . . . 21
Inference Cost . . . 22
F Endpoint and Predictor Implementation Details . . . 22
G Extended Experimental Protocol and Results . . . 26
Metric Conventions and Statistical Families . . . 27
H Frozen-Model Transfer to an Independently Acquired Scene . . . 29
Acquisition, Frozen Inference, and Pairing . . . 29
Transfer Metrics and Interpretation . . . 29
Additional Visualizations of Zero-Shot Transfer . . . 30
I Downstream Augmentation under Restricted Measured-RAD Availability . . . 31
Data, Mixing Protocol, and Generator Exposure . . . 31
Tasks and Metrics . . . 31
Downstream augmentation across measured-RAD availability . . . 32

A MATHEMATICAL AND RADAR-PHYSICAL FOUNDATIONS

These results connect the method to the sensing structure of RAD data: what LiDAR determines, how its RA/DA marginals lift into a periodic Doppler volume, why this support shortens transport to measured radar, and how RSA, SPC, and MATC represent sensor-dependent response formation.

A.1 RADAR OBSERVABILITY FROM LIDAR SUPPORT

The central modeling distinction is between geometric support that is identifiable from LiDAR and measured response that is not. A local idealization of an FMCW RAD cell before log compression is

$$Y_s(r,a,d) = \left| \sum_{j=1}^{J} \alpha_j h_R^{(s)}(r-r_j) h_A^{(s)}(a-a_j) h_D^{(s)}(d-d_j) + n_s(r,a,d) \right|^2 , \tag{11}$$

where (r_j, a_j, d_j) is the discretized delay–angle–Doppler location of a scatterer, $\alpha_j \in \mathbb{C}$ contains reflectivity and propagation phase, $h_q^{(s)}$ is the sensor-s response along physical axis q, and n_s collects receiver noise and unresolved interference. Equation 11 separates the information supplied by geometry from the response learned from radar. Calibrated LiDAR constrains (r_j, a_j) and ego motion constrains the static-scene Doppler class, but neither determines complex reflectivity, independently moving velocity, multipath, coherent cancellation, or the realized background. The normalized target $X_0 = \mathcal{N}_{s,\text{target}}(Y_s)$ must therefore remain stochastic even when its geometric support is known.

The three axes are also not metrically exchangeable. To first order, their native resolutions scale as

$$\Delta r_s \simeq \frac{c}{2\mathcal{B}_s}, \qquad \Delta v_s \simeq \frac{\lambda_s}{2T_{\mathrm{CPI},s}}, \qquad \Delta\phi_s \propto \frac{\lambda_s}{D_{\mathrm{ap},s}}, \tag{12}$$

with sweep bandwidth $\mathcal{B}_s$, wavelength λ_s, coherent processing interval $T_{\mathrm{CPI},s}$, and effective aperture $D_{\mathrm{ap},s}$. An azimuth offset $h\Delta\phi_s$ further corresponds to the range-dependent transverse distance $rh\Delta\phi_s$, whereas Doppler lives on a periodic velocity interval. Consequently, a voxel displacement of h bins has a different physical meaning along range, azimuth, and Doppler and across sensors. This observation motivates both the sensor-conditioned gates and the explicitly directional operators used by RSA.

A.2 Clean Endpoint as a Least-Committed Coupling

For one nonempty azimuth column, let $M(a) = \sum_r L_{\mathrm{RA}}(r,a) = \sum_d L_{\mathrm{DA}}(d,a)$ and define normalized marginals

$$p_R(r \mid a) = \frac{L_{\mathrm{RA}}(r,a)}{M(a)}, \qquad p_D(d \mid a) = \frac{L_{\mathrm{DA}}(d,a)}{M(a)}. \tag{13}$$

Equation 2 is equivalently $X_1^{\mathrm{raw}}(r,a,d) = M(a)p_R(r \mid a)p_D(d \mid a)$. This factorization has two properties that are useful beyond implementation convenience.

Lemma A.1 (Exact marginal preservation). For every a with $M(a) > 0$,

$$\sum_d X_1^{\mathrm{raw}}(r,a,d) = M(a)p_R(r \mid a)\sum_d p_D(d \mid a) = L_{\mathrm{RA}}(r,a), \tag{14}$$

$$\sum_r X_1^{\mathrm{raw}}(r,a,d) = M(a)p_D(d \mid a)\sum_r p_R(r \mid a) = L_{\mathrm{DA}}(d,a). \tag{15}$$

Empty columns are zero in both input marginals and in the joint tensor. Thus the lift preserves every available LiDAR RA mass and ego-static DA mass without copying any measured radar value.

Proposition A.1 (Maximum-entropy endpoint). Among all nonnegative couplings with the normalized marginals in equation 13, the product coupling $p_1(r,d \mid a) = p_R(r \mid a)p_D(d \mid a)$ maximizes Shannon entropy.

Proof. For any feasible coupling $\pi(r,d \mid a)$,

$$D_{\mathrm{KL}}(\pi \,\|\, p_R p_D) = \sum_{r,d} \pi(r,d \mid a)\log\frac{\pi(r,d \mid a)}{p_R(r \mid a)p_D(d \mid a)} \tag{16}$$

$$= H(p_R) + H(p_D) - H(\pi) \geq 0. \tag{17}$$

Hence $H(\pi) \leq H(p_R) + H(p_D)$, with equality exactly for the product coupling on its support. Multiplication by $M(a)$ recovers X_1^{raw}. □

This result gives the endpoint a precise interpretation: subject to the two source-derived marginals, it introduces no additional range–Doppler association. Independent-object velocity, target amplitude, and point-spread structure are therefore learned from the conditional radar response rather than inserted into the source.

Doppler quotient geometry. For sensor s, Doppler velocity belongs to the quotient $\mathbb{R}/(2v_{\max}^{(s)}\mathbb{Z})$. The wrapping map satisfies

$$\mathcal{W}_s(\nu + 2kv_{\max}^{(s)}) = \mathcal{W}_s(\nu), \qquad k \in \mathbb{Z}, \tag{18}$$

by direct substitution into equation 1. If $\mathcal{Q}_s$ is nearest-bin quantization with velocity spacing Δv_s, then the circular quantization error obeys

$$\mathrm{dist}_{\mathbb{S}_s^1}\big(\mathcal{W}_s(\nu), v_{d_s(a)}\big) \leq \frac{\Delta v_s}{2}. \tag{19}$$

The DA condition and therefore X_1 are invariant to the choice of equivalent unwrapped velocity representative. The endpoint respects sensor ambiguity instead of presenting an aliased velocity as an absolute physical label.

A.3 LiDAR-Supported RAD Transport

The I^{2}SB solver supplies the standard conditional Gaussian path. The radar-specific question is whether a LiDAR-derived source makes that path easier than a scene-independent source. Let $C = (L_{\mathrm{RA}}, L_{\mathrm{DA}}, s)$ and $X_1 = g(C)$. Under the same Brownian reference, the mean-path actions from the paired support and from independent $Z \sim \mathcal{N}(0, I_n)$ are proportional to $\|X_0 - g(C)\|_2^2$ and $n + \|X_0\|_2^2$, respectively. This motivates the *RAD support transport ratio*

$$\kappa_{\mathrm{RAD}} = \frac{\mathbb{E}\|X_0 - g(C)\|_2^2}{\mathbb{E}[n + \|X_0\|_2^2]}. \tag{20}$$

$\kappa_{\mathrm{RAD}} < 1$ means that calibrated LiDAR support removes model-space distance that a Gaussian-start generator must learn from conditioning alone. The same pairing removes the independent-source variance term $\lambda_t^2 I$ at intermediate states while retaining the bridge innovation $v_t I$. Appendix B.1 measures both quantities on held-out RAD cubes and obtains $\kappa_{\mathrm{RAD}} = 0.260$. This connects the transport construction directly to the information supplied by LiDAR rather than re-deriving the general bridge solver.

A.4 Radar Spread Adapter as an Anisotropic Response Operator

Under the local response model in equation 11, different sensors induce different correlation lengths along the three radar coordinates. A full dense 3D kernel can represent this behavior but does not expose which physical axis produces a response and costs $125C^2$ spatial weights for a 5^3 C-channel convolution. RSA uses nine length-five depth-wise kernels, requiring $9 \times 5C = 45C$ spatial weights, and assigns their relative contributions through sensor–time gates.

Ignoring the RMS stabilizer only for a local small-signal analysis, the Fourier response of equation 6 is

$$\widehat{\mathrm{RSA}(F)}(\boldsymbol{\omega}) = \left[1 + \frac{1}{3} \sum_{q \in \{R,A,D\}} \sum_{h \in \{1,2,4\}} g_{q,h}(s,t) \widehat{K}_{q,h}(\omega_q)\right] \widehat{F}(\boldsymbol{\omega}). \tag{21}$$

Each branch depends on exactly one component of $\boldsymbol{\omega} = (\omega_R, \omega_A, \omega_D)$. It therefore changes directional frequency content without asserting an isotropic 3D point-spread function. At native resolution, dilation h corresponds to $h\Delta r_s$, $h\Delta\phi_s$ (or approximately $rh\Delta\phi_s$ transversely), and circular $h\Delta v_s$ offsets on the respective axes. The sensor embedding is necessary because these physical scales change across the four hardware domains even after all tensors are resampled to 64^3.

Centering the gate logits makes their relative axis/scale allocation invariant to a common sensor-conditioned offset. The tanh gates and RMS stabilizer keep each branch bounded, while zero initialization begins from the shared backbone. These are implementation stabilizers. The radar-specific content is the mapping from branch dilation to sensor-dependent range, angle, and circular-Doppler scales above.

A.5 Why SPC Measures Response Shape

SPC converts each source-anchored projection-aware R/A profile or full circular Doppler profile (Appendix F.4) into the distribution $p(\ell)$ in equation 46. Quantile translation equivariance gives $Q_{0.25}(\ell + c\mathbf{1}) = Q_{0.25}(\ell) + c$, and hence

$$e_i(\ell + c\mathbf{1}) = e_i(\ell), \qquad p_i(\ell + c\mathbf{1}) = p_i(\ell). \tag{22}$$

SPC is therefore exactly invariant to an additive local log-power floor. Softplus keeps every bin positive and supplies a recovery gradient even for a collapsed profile. The normalization makes p a probability vector, so its ℓ_1 discrepancy lies in $[0, 2]$.

The distributional form also controls interpretable spread statistics. For any bounded axis statistic $f(i)$ and two profiles p and q, total variation implies

$$\left|\sum_i f(i)[p_i - q_i]\right| \le \frac{\max_i f(i) - \min_i f(i)}{2} \|p - q\|_1. \tag{23}$$

For R/A, choosing $f(i) = i$ or $f(i) = i^2$ bounds errors in profile centroid and second moment. for circular Doppler, sine/cosine statistics bound circular-moment errors. An off-center-bin indicator bounds response-mass error in either setting. Thus minimizing SPC does more than compare a visual line shape: it controls a family of physically meaningful range, beam, and Doppler-spread summaries around source-selected reflectors. Source-only anchor selection makes these response statistics computable from LiDAR support without target-selected measured-radar peaks.

A.6 MATC as a Directional Variogram and Spectral Constraint

Let $Z(i)$ denote an approximately second-order stationary off-support log-power field. Its directional semivariogram is

$$\gamma_q(h) = \frac{1}{2}\mathbb{E}\left[(Z(i + he_q) - Z(i))^2\right] = C(0) - C(he_q), \tag{24}$$

where C is the autocovariance. Ignoring only finite-mask boundary effects, equation 8 estimates $E^2_{q,h} \simeq 2\gamma_q(h) + \varepsilon^2$. It therefore measures how quickly radar background decorrelates along each physical coordinate rather than how closely a generated speckle phase matches one target realization.

The same quantity has a frequency interpretation. If $S(\boldsymbol{\omega})$ is the background power spectral density, Parseval's identity gives

$$2\gamma_q(h) = 2\int [1 - \cos(h\omega_q)]\, S(\boldsymbol{\omega}) d\boldsymbol{\omega} \tag{25}$$

$$= 4\int \sin^2\left(\frac{h\omega_q}{2}\right) S(\boldsymbol{\omega}) d\boldsymbol{\omega}. \tag{26}$$

A single lag could be matched by concentrating error in an unobserved frequency band. The lags $\{1, 2, 4\}$ apply three different high-pass transfer functions and reduce this degeneracy, while separate $q \in \{R, A, D\}$ terms prevent one over-textured axis from cancelling an under-textured one. The logarithm in equation 9 compares multiplicative energy ratios symmetrically, and SmoothL1 limits the leverage of isolated high-energy cells. Masking source support turns the statistic into a background estimator. because a finite irregular mask breaks exact stationarity, equation 26 is the population interpretation rather than an equality claimed for every masked sample.

Connection to support-to-response generation. The analysis identifies non-overlapping responsibilities. The maximum-entropy endpoint preserves predictable support. The paired bridge learns the residual response from that scene-specific source. RSA supplies sensor-dependent directional capacity. SPC constrains local reflector response modulo the floor, and MATC constrains off-support directional covariance modulo phase. None of SPC, MATC, or RSA changes the bridge law, and none of the bridge equations prescribes a radar point-spread function. This separation is what makes the components a single support-to-response construction rather than a collection of losses.

B Empirical Evidence for Support-Anchored Transport

B.1 LiDAR Support Reduces Path Length and Source Variance

We evaluate the RAD support transport ratio from Appendix A.3 in the normalized state space used to train all generators. Let $C = (L_{\mathrm{RA}}, L_{\mathrm{DA}}, s)$ and let the paired source be the deterministic $X_1 = g(C)$. As an analytic Gaussian-source control, let $Z \sim \mathcal{N}(0, I_n)$ be independent of (X_0, C). Under the same Brownian reference, conditioning on X_0 and C gives

$$\mathcal{A}(X_0, X_1) = \frac{\|X_1 - X_0\|_2^2}{2B_1}, \qquad \mathbb{E}_Z\mathcal{A}(X_0, Z) = \frac{n + \|X_0\|_2^2}{2B_1}. \tag{27}$$

The numerator is the energy of response formation left after LiDAR support is fixed. the denominator is the distance traversed from scene-independent noise. Their ratio is κ_{RAD} in Equation 20.

The same control isolates source-induced nuisance variance. Write $\lambda_t = B_t/B_1$ and $v_t = B_t\bar{B}_t/B_1$. Conditioned on a fixed measured endpoint X_0 and condition C, the paired marginal and the Gaussian-

source mixture satisfy

$$\mathrm{Var}(X_t \mid X_0, C, X_1) = v_t I, \tag{28}$$

$$\mathrm{Var}_{Z,W}(X_t \mid X_0, C) = (v_t + \lambda_t^2) I. \tag{29}$$

The additional $\lambda_t^2 I$ comes from drawing a scene-independent source. For the symmetric schedule used here, $B_1 = 0.30005$, $\lambda_{0.5} = 0.5$, and $v_{0.5} = 0.07501$. A fixed paired endpoint therefore reduces the common-reference intermediate variance from 0.32501 to 0.07501 per voxel at $t = 0.5$, removing 76.9% of reference-source variance while retaining Brownian innovation for stochastic response formation.

Empirical path verification. We compute equation 27 on the same 1,920 held-out frames as the main comparison, excluding invalid range voxels. For Gaussian endpoints, the expectation $n + \|X_0\|^2$ is evaluated analytically rather than from a favorable random draw. Metrics are averaged within sequence and domain before the four-domain macro average. The paired endpoint is shorter on all 15 sequences (two-sided exact sign test, $p = 6.10 \times 10^{-5}$), including every sensor domain. Thus LiDAR geometry and ego-motion remove 73.98% of the model-space path that a Gaussian-start generator must otherwise infer through conditioning.

Table 3: Model-space path comparison over 15 held-out sequences. The confidence interval is a domain-stratified sequence bootstrap. The ratio is computed per sequence before domain macro-averaging, so it need not equal the ratio of the two separately averaged squared distances.

Statistic	Domain macro
Paired endpoint squared distance per voxel	0.3292
Expected Gaussian endpoint squared distance per voxel	1.3098
RAD support transport ratio κ_{RAD}	0.2602
Relative path-action reduction	73.98%
95% CI for reduction	[70.45%, 77.63%]

B.2 Repeated-Sample Conditional Validation

Path shortening removes scene-independent source variation, but a useful simulator must still produce non-degenerate measured responses. We therefore draw $K = 8$ samples for each of 120 conditions: eight temporally uniform frames from every held-out sequence. Conditions and random seeds are shared across methods. In addition to conditional sample variance and pairwise RMSE, we use two proper ensemble scores. For generated samples $\{\widehat{X}^{(k)}\}_{k=1}^{K}$ and target Y, voxelwise CRPS and the multivariate energy score have the empirical forms

$$\widehat{\mathrm{CRPS}} = \frac{1}{K}\sum_k \|\widehat{X}^{(k)} - Y\|_1 - \frac{1}{K(K-1)}\sum_{k<j} \|\widehat{X}^{(k)} - \widehat{X}^{(j)}\|_1, \tag{30}$$

$$\widehat{\mathrm{ES}} = \frac{1}{K}\sum_k \|\widehat{X}^{(k)} - Y\|_2 - \frac{1}{K(K-1)}\sum_{k<j} \|\widehat{X}^{(k)} - \widehat{X}^{(j)}\|_2, \tag{31}$$

where norms are averaged over valid voxels before sequence and domain aggregation. These scores penalize both inaccurate samples and unsupported dispersion. We also measure the fraction of target voxels lying between the minimum and maximum generated values. Under exchangeability, its nominal coverage for K samples is $(K-1)/(K+1) = 0.7778$. The table reports absolute calibration error from this value.

The frequency-domain metrics do not appear in any training loss. Source-supported voxels are filled with the background mean, each valid line is mean-centered, and complete non-DC power spectra are normalized along range, azimuth, and Doppler. We then average the one-dimensional Wasserstein distance (PSD-W) and Jensen–Shannon divergence (PSD-JSD) over the three axes. Unlike MATC, these metrics compare the complete frequency distribution rather than finite-difference energy at lags $\{1, 2, 4\}$.

Table 4: Eight-sample conditional evaluation on 120 held-out conditions. Variance and pairwise RMSE quantify dispersion and have no universal optimum. CRPS, energy score, coverage error, PSD-W, and PSD-JSD are lower-is-better.

Method	Cond. var.	Pair RMSE	CRPS ↓	Energy ↓	Coverage err. ↓	PSD-W ↓	PSD-JSD ↓
Conditional DDPM	0.2261	0.6345	0.4414	0.6631	0.3280	0.05495	0.02105
Flow matching	0.2167	0.6210	**0.3803**	**0.5704**	0.2902	0.04973	0.01719
RadarSimRAD	**0.1707**	**0.5706**	0.4178	0.6395	**0.2482**	**0.03372**	**0.01077**

The paired bridge reduces conditional variance by 24.5% relative to DDPM and 21.2% relative to flow. Pairwise RMSE is $10.1\%/8.1\%$ lower than DDPM/flow but remains nonzero at 0.5706. These dispersion reductions retain Holm-corrected $p = 0.04742/0.02112$. Its full-spectrum distance is also lower: PSD-W improves by $38.6\%/32.2\%$ and PSD-JSD by $48.8\%/37.3\%$ relative to DDPM/flow. PSD-JSD improves on all 15 sequences against both controls, and PSD-W does so on 14/15 and 15/15 sequences, respectively. All four spectral comparisons retain Holm-corrected $p \leq 0.00085$. The conditional-variance reductions also remain significant after correction ($p = 0.02881$ and 0.01074). This evidence is independent of the finite-lag MATC metric family.

Disabling posterior noise in RadarSimRAD provides a deterministic comparison. The deterministic output has MAE 0.7126 and PSD-W 0.06579, whereas stochastic draws average 0.6041 and 0.03372. Stochasticity is therefore active and beneficial rather than an unused formal component. Flow attains lower generic CRPS and energy scores, whereas RadarSimRAD achieves substantially lower coverage error and full-spectrum distances, indicating a different advantage in radar-specific response structure.

C Implementation Details

Table 5: Sequence-disjoint dataset split used throughout the paper. Nineteen K-Radar sequence-tail frames with missing static curves are excluded before manifest construction.

Domain	Train	Validation	Test	Test sequences
K-Radar	7,348	2,982	2,666	4
RaDelft	11,608	2,380	2,987	1
ColoRadar cascade	21,307	4,365	3,706	5
ColoRadar single-chip	40,503	8,721	7,407	5
Total	80,766	18,448	16,766	15

Preprocessing. Disk tensors use axes $[R, D, A]$ and are transposed to $[R, A, D]$ before non-overlapping reduction. Radar power is max-pooled to preserve peaks, whereas occupancy and DA mass are sum-pooled. Invalid near-range bins are masked in training and metrics. Robust normalization uses the 0.001 and 0.999 log-power quantiles from fixed, uniformly sampled training frames. Its manifest digest is stored with the statistics.

Model and training. The final model uses batch size 24 on one NVIDIA L40S, bf16 mixed precision, gradient checkpointing, and deterministic data ordering with seed 42. The triangular bridge rate has $\beta_{\min} = 10^{-4}$ and $\beta_{\max} = 0.6$. RSA uses length-five depth-wise kernels at dilations 1, 2, and 4, centered sensor–time gates with $g_{\max} = 0.35$, and insertion scales 32^3, 16^3, and 8^3. SPC uses at most 32 source anchors, R/A profile radius 8, and full 64-bin source-relative Doppler spectra. MATC uses lags 1, 2, and 4 on each physical axis. The auxiliary weight warms linearly to 0.05 over 500 optimizer steps.

Evaluation. Controlled test generators use the same 1,920 frame identities and per-frame sample seed. The separate L2RLDB validation benchmark shares frame identities and evaluation settings, with sampling provenance recorded separately for its supplied exports. Ratios whose optimum is one are scored by distance to one rather than by treating larger as better. The original test family—correlation, MAE, Peak-F1, spread, RA correlation, off-center Doppler error, lag-1 error,

Table 6: Per-domain results for the final RadarSimRAD checkpoint and matched generative controls. Each cell reports Corr / MAE / Peak-F1 / Spread / AGLRE.

Domain	RadarSimRAD	Conditional DDPM	Flow matching
Cascade	0.7288 / 0.9247 / 0.2581 / 0.5532 / 0.3162	0.6519 / 1.0779 / 0.1199 / 0.6136 / 0.6493	0.7157 / 0.9341 / 0.2268 / 0.5732 / 0.5409
Single-chip	0.7374 / 0.7866 / 0.3963 / 0.5301 / 0.1799	0.6744 / 0.8331 / 0.3224 / 0.6108 / 0.4108	0.7307 / 0.7737 / 0.3701 / 0.5622 / 0.3857
K-Radar	0.6944 / 0.4715 / 0.2702 / 0.6696 / 0.2665	0.6912 / 0.4573 / 0.2851 / 0.6803 / 0.6766	0.7026 / 0.4668 / 0.2742 / 0.6805 / 0.5598
RaDelft	0.6700 / 0.2212 / 0.3171 / 0.4162 / 0.3313	0.6097 / 0.2236 / 0.2121 / 0.4547 / 0.7498	0.6591 / 0.2174 / 0.2825 / 0.4265 / 0.6379

AGLRE, worst-axis error, texture ratio, and background-standard- deviation ratio—was fixed before complete sequence evaluation. PSD-W, PSD-JSD, and repeated-sample proper scores are post-training diagnostics defined before evaluating all 1,920 frames and 120 conditional ensembles. ASCE, ASAE, PSE-E, and MS-ACF were added after training and before evaluating the matched validation benchmark in Appendix E. All are evaluation-only diagnostics.

D Per-Domain Results

D.1 Additional Curated Visualizations

The eight curated conditions in Figures 8 and 9 illustrate geometry and response formation across scene types and sensor power scales. Their per-frame scores distinguish conditional agreement from background-gradient and spectral fidelity. Selection is illustrative rather than statistically representative. aggregate conclusions use the fixed test and matched-validation frame sets.

E Matched L2RLDB Evaluation

E.1 Matched-Frame Protocol and Representation Conversion

The matched-validation benchmark in Table 1 contains 15 sequences (four K-Radar, five ColoRadar cascade, five ColoRadar single-chip, and one RaDelft) and 128 temporally uniform frame identities per sequence, yielding 1,920 predictions per method. Frame identities are resolved from exporter manifests. Metrics are averaged from frames to sequences, sequences to domains, and equally across the four domains.

L2RLDB is trained separately in each domain with its three-stage radar-VAE, LiDAR-VAE, and latent-bridge procedure. The delivered predictions are physical $\log(1+Y)$ RAD arrays in $[R, A, D]$ order. For ColoRadar and RaDelft, its training representation places RAD in a dummy eight-bin elevation dimension. Export takes the elevation maximum and retains the native 64^3 RAD grid. For K-Radar, the generated crop has $(D, R, A, E) = (63, 192, 96, 32)$ and is embedded into the official DREA canvas before elevation maximization, range pooling, azimuth resampling, and log compression. This conversion has support only in output range bins $0{:}48$. We consequently apply the identical $R < 48$ validity mask to predictions and ground truth for *all* methods in K-Radar, rather than reward or penalize the baseline for unsupported bins. The other domains use their full 64^3 support. For Spread, the mask restricts anchor centers but not every voxel in the extracted profile window. K-Radar profiles near the crop boundary can therefore extend beyond $R < 48$.

Supervised representation versus dense projections. L2RLDB (Shen et al., 2026) represents radar power using its Doppler mean, a CFAR key-cell mask, and full spectra at the selected cells. Non-key spectra are completed with a stochastic prior scaled by the predicted mean power. For a nonnegative N_D-bin spectrum with mean μ, the peak can lie anywhere in $[\mu, N_D\mu]$. Thus, even exact mean-power recovery does not identify RA peak contrast outside the key set, while non-key completion can influence DA maxima and dense spectral statistics. The non-key spectral shape is prescribed rather than directly supervised. This representation therefore emphasizes a different statistic from the dense maximum-projection and background-formation metrics evaluated here.

Checkpoint and export provenance. The evaluation uses the delivered stochastic L2RLDB exports after elevation maximization and native-grid conversion. DDPM, flow matching, and RadarSimRAD use the same evaluator, frame identities, and physical RAD support.

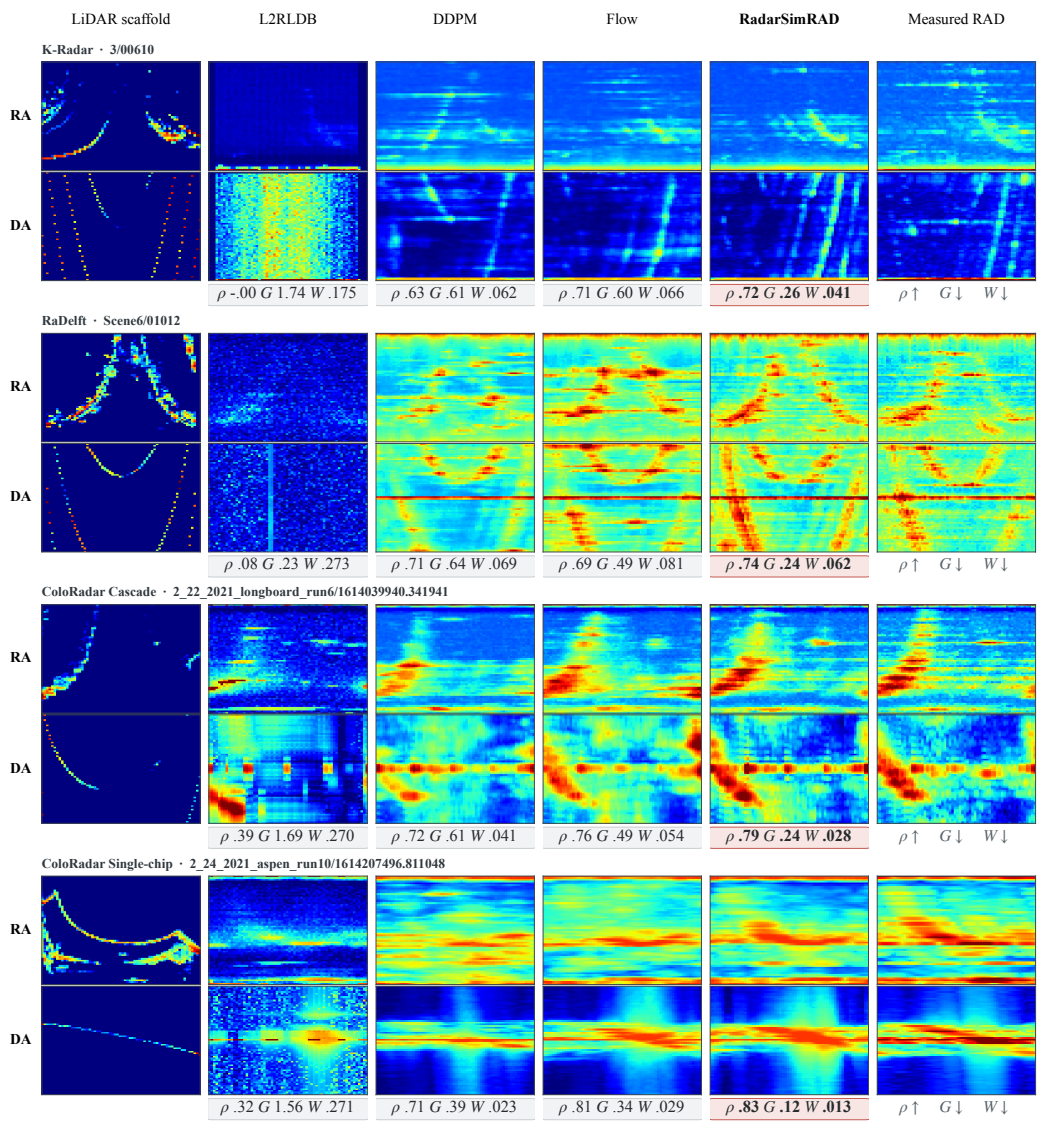


Figure 8: **Conditional RAD synthesis across four radar domains.** Columns show the LiDAR endpoint, L2RLDB reproduction, DDPM, flow matching, RadarSimRAD, and measured RAD; each domain contains RA and DA maximum projections. Radar panels within a frame share log-power limits, while the LiDAR endpoint uses a separate scale. Insets report cube-level Corr $\rho \uparrow$, AGLRE $G \downarrow$, and PSD-W $W \downarrow$.

E.2 Aggregate Validation Results

Aggregate comparison. Table 7 shows that the final RadarSimRAD model improves correlation from 0.1793 to 0.6532 and Peak-F1 from 0.0928 to 0.2809 relative to the L2RLDB reproduction, while reducing MAE by 84.3% and spread distance by 17.4%. Its AGLRE, PSD-W, and MS-ACF errors fall by 78.8%, 86.6%, and 86.0%, respectively. Corr, MAE, Peak-F1, and all spectral and autocorrelation diagnostics improve on every sequence. Relative to flow, the final model also improves Spread, RA Corr, and all seven formation metrics. Complete per-domain, per-sequence, and significance results follow below.

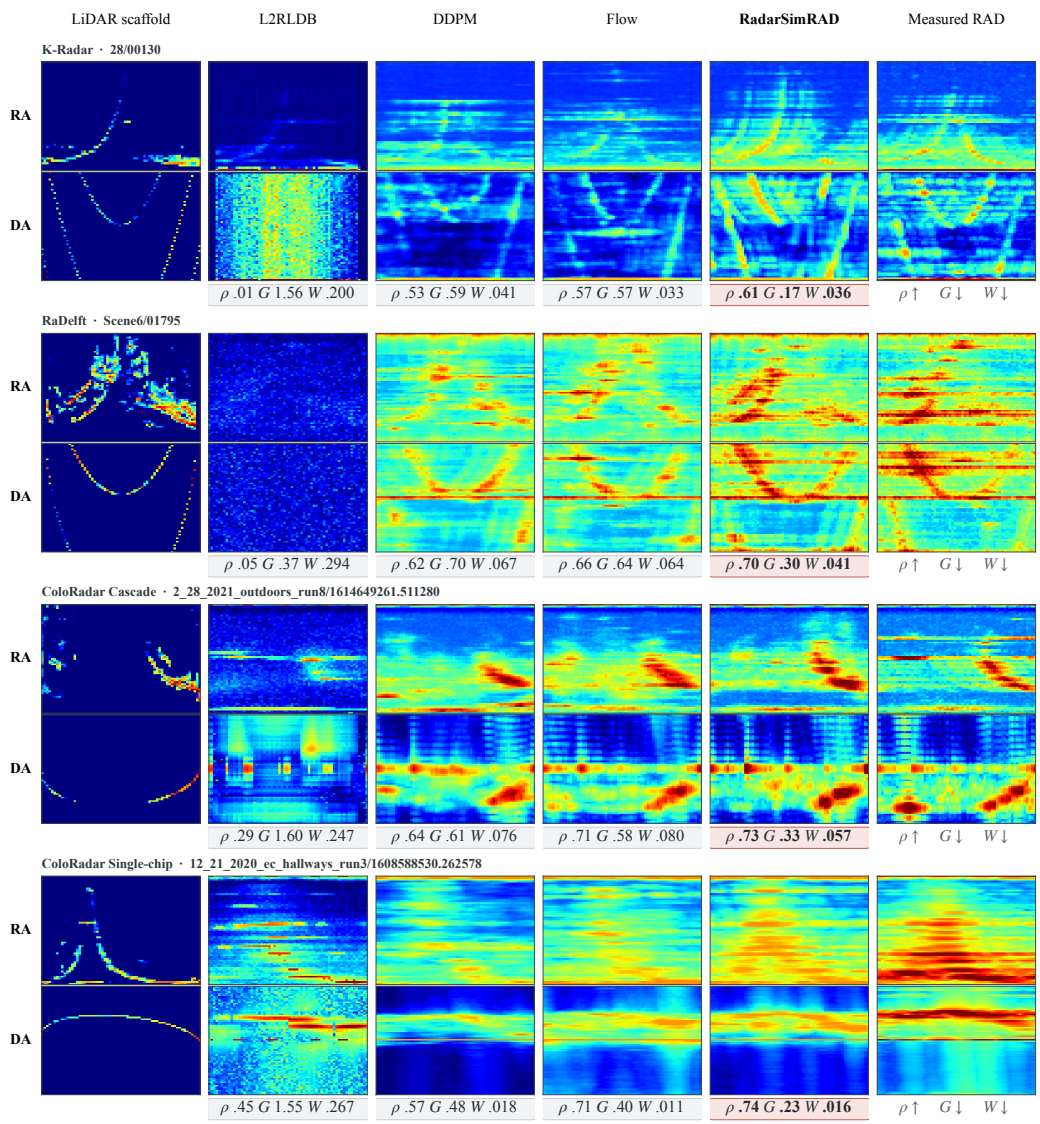


Figure 9: **Additional curated validation examples.** Each domain contributes a different scene or acquisition time from Figure 8, using the same six columns, RA/DA projections, shared radar scales, and validity masks. Insets report Corr $\rho\uparrow$, AGLRE $G\downarrow$, and PSD-W $W\downarrow$, computed on each displayed cube.

E.3 Formation Metrics Independent of MATC

Let $\mathcal{B}$ be the common valid off-support mask. Along physical axis $q \in \{R, A, D\}$, masked values are filled with the sample background mean, each line is mean-centered, and the line spectra are averaged. After removing DC, let $P_q(k) \geq 0$ denote the resulting normalized power spectrum and $f_k \in [0, 1]$ its normalized frequency. Its centroid and entropy are

$$c_q(P) = \sum_{k=1}^{K_q} f_k P_q(k), \qquad h_q(P) = -\frac{1}{\log K_q} \sum_{k=1}^{K_q} P_q(k) \log(P_q(k) + \varepsilon). \tag{32}$$

We define axis spectral-centroid error and power-spectral-entropy error as

$$\text{ASCE} = \frac{1}{3} \sum_q |c_q(\widehat{P}) - c_q(P)|, \qquad \text{PSE-E} = \frac{1}{3} \sum_q |h_q(\widehat{P}) - h_q(P)|. \tag{33}$$

Table 7: Matched validation comparison on 1,920 identical frames. K-Radar uses the common $R < 48$ voxel mask. Spread windows have the boundary qualification in Appendix E.1. Other domains use their full grids. L2RLDB uses one 50-step reproduction per domain. The other methods use a cross-domain model at NFE$= 8$. Green/blue indicate best/second best, with our rows tinted.

Method	Corr ↑	MAE ↓	Peak-F1 ↑	Spread ↓	RA Corr ↑	AGLRE ↓	PSD-W ↓	MS-ACF ↓
L2RLDB (Shen et al., 2026)	0.1793	4.0704	0.0928	0.6515	0.4762	1.2503	0.2317	0.4012
Conditional DDPM	0.5898	0.6859	0.2247	0.5885	0.5761	0.6541	0.0483	0.0884
Flow matching	0.6457	**0.6340**	0.2639	0.5564	0.6471	0.5612	0.0483	0.0844
RadarSimRAD w/o MATC	0.5821	0.7480	0.2053	**0.5351**	0.6222	0.6370	0.0420	0.1031
RadarSimRAD w/o RSA	0.5992	0.7300	0.2260	0.5452	0.6303	0.4416	0.0390	0.0741
RadarSimRAD (matched, 10k)	0.5778	0.7766	0.2499	0.5494	0.6266	0.4315	0.0404	0.0868
RadarSimRAD (final, 14.5k)	**0.6532**	0.6373	**0.2809**	0.5380	**0.6742**	**0.2649**	**0.0311**	**0.0561**

The centroid triplet also isolates anisotropy. With $a_q(P) = c_q(P)/(\sum_j c_j(P) + \varepsilon)$, axis spectral-anisotropy error is

$$\mathrm{ASAE} = \|a(\widehat{P}) - a(P)\|_1. \tag{34}$$

ASCE measures the overall spatial-frequency scale, whereas ASAE is invariant to a common centroid rescaling and tests how that scale is apportioned across the non-exchangeable radar axes. PSD-W and PSD-JSD retain the complete spectrum and therefore need not agree with either moment summary.

For a second, non-Fourier check, define the masked normalized autocorrelation

$$\rho_q(h, X) = \frac{\sum_{i:(i,i+h)\in\mathcal{B}} (X_i - \bar{X}_h)(X_{i+h} - \bar{X}'_h)}{\sqrt{\sum_i (X_i - \bar{X}_h)^2 \sum_i (X_{i+h} - \bar{X}'_h)^2} + \varepsilon}. \tag{35}$$

The multi-scale error is

$$\text{MS-ACF} = \frac{1}{12} \sum_{q\in\{R,A,D\}} \sum_{h\in\{1,2,4,8\}} |\rho_q(h, \widehat{X}) - \rho_q(h, X)|. \tag{36}$$

MATC instead matches unnormalized finite-difference RMS at lags $\{1, 2, 4\}$ during training. It contains neither a Fourier transform, a spectral centroid or entropy, nor normalized covariance, and it never observes lag eight. Equations equation 33– equation 36 are thus evaluation-only diagnostics rather than restatements of the loss. They are post-training diagnostics applied unchanged to every method and reported alongside conventional fidelity scores rather than collapsed into a tuned composite.

E.4 Inference Cost

The supplied L2RLDB timing on an NVIDIA L40S is 125.55 ± 10.72 ms per K-Radar frame (100 timed frames after 20 warm-up frames, fp32, batch size one, 50 steps). It times only GPU LiDAR encoding, bridge sampling, radar decoding, and 4D reconstruction for the $(63, 192, 96, 32)$ crop. It excludes I/O, checkpoint loading, RAD export, and metrics. Table 14, in contrast, reports end-to-end evaluator time for 64^3 RAD at NFE$= 8$. The timings are listed separately because they include different pipeline components.

F Endpoint and Predictor Implementation Details

This section retains the coordinate, normalization, and predictor definitions underlying Section 3. Their radar-specific consequences are developed in Appendix A.

F.1 Source Coordinates and Calibration

The conditional measurement model is

$$X_0 \sim p_\theta(X_0 \mid X_1, L_{\mathrm{RA}}, L_{\mathrm{DA}}, s), \tag{37}$$

Table 8: Matched-validation per-domain radar-response formation. Green/blue denote best/second best within a domain.

Domain	Method	AGLRE ↓	PSD-W ↓	PSD-JSD ↓	ASCE ↓	ASAE ↓	PSE-E ↓	MS-ACF ↓
ColoRadar cascade	L2RLDB	1.4778	.2497	.1500	.2575	.3865	.3129	.3383
	DDPM	.6345	.0457	.0217	.0459	.1993	.0932	.0748
	Flow	.5376	.0485	.0213	.0492	.1898	.1078	.0712
	w/o MATC	.5073	.0378	.0159	.0381	.1363	.0807	.0680
	w/o RSA	.3143	.0343	.0146	.0339	.1279	.0640	.0529
	Ours-10k	.3140	**.0330**	.0149	**.0322**	.1271	**.0621**	.0532
	Ours-14.5k	**.3046**	.0335	**.0142**	.0329	**.1250**	.0628	**.0525**
ColoRadar single-chip	L2RLDB	1.6447	.2321	.1336	.2385	.3947	.2523	.3059
	DDPM	.4072	.0396	.0153	.0373	.1183	.0654	.0757
	Flow	.3847	.0355	.0111	.0342	.0993	.0647	.0716
	w/o MATC	.3864	.0322	.0095	.0311	.1182	.0578	.0679
	w/o RSA	.1883	**.0269**	.0078	**.0257**	**.0853**	**.0438**	**.0510**
	Ours-10k	.1804	.0273	.0079	.0265	.0924	.0445	.0524
	Ours-14.5k	**.1748**	.0281	**.0077**	.0274	.0923	.0442	.0542
K-Radar	L2RLDB	1.5863	.1924	.1428	.1931	.8820	.2254	.4737
	DDPM	.8065	.0504	.0227	.0493	.2366	.1000	.1214
	Flow	.6954	.0497	.0203	.0487	.2282	.1001	.1149
	w/o MATC	1.0548	.0512	.0256	.0510	.2399	.0726	.1980
	w/o RSA	.9179	.0550	.0245	.0528	.2462	.0698	.1404
	Ours-10k	.9139	.0618	.0311	.0612	.3029	.0842	.1891
	Ours-14.5k	**.2448**	**.0258**	**.0078**	**.0230**	**.0624**	**.0332**	**.0665**
RaDelft	L2RLDB	**.2925**	.2526	.1254	.2606	.1269	.2190	.4869
	DDPM	.7680	.0573	.0174	.0575	.2712	.0991	.0817
	Flow	.6274	.0594	.0164	.0603	.2309	.1070	.0799
	w/o MATC	.5996	.0468	.0112	.0474	.1510	.0789	.0784
	w/o RSA	.3459	.0400	.0090	.0395	.1390	.0586	.0519
	Ours-10k	.3178	.0394	.0090	.0387	.1251	.0556	.0525
	Ours-14.5k	.3355	**.0370**	**.0081**	**.0362**	**.1163**	**.0517**	**.0512**

Table 9: Complete matched-validation per-sequence Peak-F1 ↑ results. Green/blue denote the best/second-best method on each sequence.

Domain	Sequence	L2RLDB	DDPM	Flow	w/o MATC	w/o RSA	Ours-10k	Ours-14.5k
ColoRadar cascade	12_21_2020_arpg_lab_run3	.0637	.0205	.0934	.0736	.1271	**.1333**	.1306
	12_21_2020_ec_hallways_run3	.1203	.0122	.0908	.0737	.1280	.1259	**.1369**
	2_22_2021_longboard_run6	.1449	.2861	.2922	.2830	.3069	.2974	**.3106**
	2_24_2021_aspen_run10	.2959	.1532	.3612	.3198	.3717	.3833	**.3999**
	2_28_2021_outdoors_run8	.2486	.2331	.3458	.3382	.3714	.3511	**.3790**
ColoRadar single-chip	12_21_2020_arpg_lab_run3	.0735	.2240	.2496	.2622	.3006	.2900	**.3039**
	12_21_2020_ec_hallways_run3	.0995	.1837	.2271	.2431	.2748	.2768	**.2863**
	2_22_2021_longboard_run6	.0834	.4624	.4782	**.4943**	.4487	.4504	.4495
	2_24_2021_aspen_run10	.2057	.3513	.4346	.4283	.4710	.4599	**.4861**
	2_28_2021_outdoors_run8	.1193	.3844	.4505	**.4779**	.4384	.4408	.4378
K-Radar	3	.0782	.3087	**.3171**	.0824	.1042	.2089	.2879
	11	.0788	**.2805**	.2772	.0445	.0565	.1726	.2480
	22	.0819	.2112	.2125	.0015	.0049	.0720	**.2549**
	28	.0807	**.3334**	.3186	.0884	.1036	.2296	.2880
RaDelft	Scene6	.0004	.1530	.1694	.1681	.1890	.1871	**.1897**

Table 10: Complete matched-validation per-sequence AGLRE ↓ results. Green/blue denote the best/second-best method on each sequence.

Domain	Sequence	L2RLDB	DDPM	Flow	w/o MATC	w/o RSA	Ours-10k	Ours-14.5k
ColoRadar cascade	12_21_2020_arpg_lab_run3	1.3315	.6653	.5626	.5396	.3849	.3661	**.3650**
	12_21_2020_ec_hallways_run3	1.3848	.7019	.5036	.5198	.3579	.3514	**.3508**
	2_22_2021_longboard_run6	1.6348	.6703	.6746	.5691	.3106	.3380	**.3099**
	2_24_2021_aspen_run10	1.4342	.5423	.4295	.4074	.2456	**.2246**	.2376
	2_28_2021_outdoors_run8	1.6036	.5926	.5177	.5005	.2727	.2900	**.2595**
ColoRadar single-chip	12_21_2020_arpg_lab_run3	1.5983	.4010	.3847	.4088	.2054	**.1898**	.1908
	12_21_2020_ec_hallways_run3	1.5442	.4575	.3949	.4229	.2165	.2181	**.2067**
	2_22_2021_longboard_run6	1.7198	.4191	.3912	.3714	.1688	.1658	**.1615**
	2_24_2021_aspen_run10	1.6444	.3863	.3733	.3704	.1756	.1592	**.1573**
	2_28_2021_outdoors_run8	1.7167	.3722	.3792	.3584	.1749	.1693	**.1576**
K-Radar	3	1.6332	.7918	.7542	1.0041	.8795	.8721	**.2709**
	11	1.6003	.8305	.7622	1.0487	.9157	.9134	**.2738**
	22	1.4767	.8650	.6810	1.1515	.9987	1.0160	**.2446**
	28	1.6348	.7388	.5840	1.0147	.8778	.8540	**.1900**
RaDelft	Scene6	**.2925**	.7680	.6274	.5996	.3459	.3178	.3355

Table 11: Complete matched-validation per-sequence PSD-W ↓ results. Green/blue denote the best/second-best method on each sequence.

Domain	Sequence	L2RLDB	DDPM	Flow	w/o MATC	w/o RSA	Ours-10k	Ours-14.5k
ColoRadar cascade	12_21_2020_arpg_lab_run3	.2491	.0354	.0402	.0334	.0306	.0286	**.0270**
	12_21_2020_ec_hallways_run3	.2458	.0463	.0484	.0460	.0426	.0432	**.0422**
	2_22_2021_longboard_run6	.2612	.0518	.0543	.0360	.0320	.0317	**.0311**
	2_24_2021_aspen_run10	.2522	.0330	.0368	.0297	.0255	.0251	**.0243**
	2_28_2021_outdoors_run8	.2403	.0621	.0628	.0439	.0406	**.0367**	.0429
ColoRadar single-chip	12_21_2020_arpg_lab_run3	.2570	.0285	.0302	.0308	.0226	.0238	**.0219**
	12_21_2020_ec_hallways_run3	.2574	.0278	.0306	.0320	**.0231**	.0249	.0235
	2_22_2021_longboard_run6	.2077	.0448	.0363	.0277	**.0269**	.0277	.0290
	2_24_2021_aspen_run10	.2366	.0425	.0362	.0369	.0295	.0289	**.0287**
	2_28_2021_outdoors_run8	.2016	.0544	.0442	.0334	.0321	**.0310**	.0372
K-Radar	3	.1864	.0574	.0573	.0494	.0560	.0604	**.0289**
	11	.1917	.0528	.0541	.0533	.0571	.0646	**.0263**
	22	.1972	.0501	.0423	.0513	.0588	.0632	**.0219**
	28	.1943	.0414	.0451	.0507	.0479	.0590	**.0261**
RaDelft	Scene6	.2526	.0573	.0594	.0468	.0400	.0394	**.0370**

Table 12: Complete matched-validation per-sequence MS-ACF ↓ results. Green/blue denote the best/second-best method on each sequence.

Domain	Sequence	L2RLDB	DDPM	Flow	w/o MATC	w/o RSA	Ours-10k	Ours-14.5k
ColoRadar cascade	12_21_2020_arpg_lab_run3	.3239	.0521	.0518	.0579	.0451	.0437	**.0400**
	12_21_2020_ec_hallways_run3	.3288	.0701	.0599	.0689	**.0572**	.0623	.0576
	2_22_2021_longboard_run6	.3743	.0901	.0950	.0772	.0609	.0603	**.0591**
	2_24_2021_aspen_run10	.3356	.0532	.0487	.0521	.0357	.0370	**.0330**
	2_28_2021_outdoors_run8	.3291	.1083	.1004	.0839	.0653	**.0626**	.0730
ColoRadar single-chip	12_21_2020_arpg_lab_run3	.3394	.0464	.0480	.0494	**.0337**	.0363	.0347
	12_21_2020_ec_hallways_run3	.3219	.0551	.0534	.0590	**.0398**	.0452	.0428
	2_22_2021_longboard_run6	.2384	.0907	.0887	.0762	**.0635**	.0644	.0663
	2_24_2021_aspen_run10	.3627	.0740	.0683	.0707	.0490	.0480	**.0476**
	2_28_2021_outdoors_run8	.2670	.1121	.0997	.0843	.0691	**.0680**	.0798
K-Radar	3	.4580	.1440	.1500	.1981	.1511	.1938	**.0781**
	11	.4757	.1300	.1297	.2018	.1500	.1966	**.0790**
	22	.4700	.0902	.0772	.1683	.1036	.1608	**.0496**
	28	.4912	.1213	.1026	.2237	.1571	.2053	**.0594**
RaDelft	Scene6	.4869	.0817	.0799	.0784	.0519	.0525	**.0512**

where the target is measured at the same acquisition time as the LiDAR conditions. For calibrated occupancy $O(r, a)$ and radial velocity $\nu(a)$ induced by ego-motion and object motion, wrapping and nearest-bin quantization follow equation 1. Writing $m_A(a) = \sum_r O(r, a)$, the conditions are

$$L_{\mathrm{RA}}(r, a) = O(r, a), \qquad L_{\mathrm{DA}}(d, a) = m_A(a)\,\mathbf{1}[d = d_s(a)]. \tag{38}$$

Measured dynamic-object labels are excluded because their Doppler indices depend on target radar peaks. Disk radar tensors use native axes $[R, D, A]$ at 128^3. we transpose to $[R, A, D]$ and max-pool to 64^3 to preserve localized peaks. LiDAR occupancy and DA mass are sum-pooled to preserve point count.

For sensor s and tensor role k, let $q^-_{s,k}$ and $q^+_{s,k}$ be robust training-only log-power quantiles, and set $c_{s,k} = (q^-_{s,k} + q^+_{s,k})/2$ and $h_{s,k} = (q^+_{s,k} - q^-_{s,k})/2$. For nonnegative input u, normalization is

$$\mathcal{N}_{s,k}(u) = \mathrm{clip}_{[-1,1]}\left(\frac{\log(1+u) - c_{s,k}}{h_{s,k}}\right). \tag{39}$$

Source and target roles use separate calibration. Invalid-range voxels are set to zero after normalization and excluded from every objective and metric. Fixed training frames determine all robust bounds.

F.2 Reference Schedule and Sampling

The time-inhomogeneous Brownian reference is

$$dX_t = \sqrt{\beta(t)}\,dW_t, \qquad B_t = \int_0^t \beta(\tau)d\tau, \qquad \bar{B}_t = B_1 - B_t. \tag{40}$$

Its variance rate is symmetric and triangular:

$$\beta(t) = \beta_{\min} + 2(\beta_{\max} - \beta_{\min})\min(t, 1-t), \qquad \beta_{\min} = 10^{-4}, \quad \beta_{\max} = 0.6. \tag{41}$$

Training samples an integer index uniformly from $\{1, \ldots, 1000\}$. Index zero is excluded because its forward standard deviation vanishes. Inference uses a decreasing time grid from $t = 1$ to $t = 0$ and caches the condition encodings across all evaluations. The bridge marginal and posterior follow the standard I^{2}SB construction in Equations 3 and 5. Our method-specific analysis instead asks how much LiDAR support reduces RAD transport relative to a Gaussian source (Appendix A.3).

F.3 Directional Feature Modulation

The U-Net uses separate RA/DA feature pyramids, sensor and sinusoidal time embeddings, and FiLM modulation. The condition pyramids and sensor code are computed once per generated cube. A zero-initialized low-rank sensor-style residual is applied at 64^3 and 32^3. RSA operates at 32^3, 16^3, and 8^3. For physical axis q and dilation h, its stabilized branch is

$$H_{q,h} = \mathcal{S}(K_{q,h} *_q F), \qquad \mathcal{S}(U) = \frac{U}{\sqrt{1 + \mathrm{RMS}(U)^2}}, \tag{42}$$

where the depth-wise length-five kernels use $h \in \{1, 2, 4\}$. Sensor–time logits are centered before bounded gating:

$$\widetilde{z}_{q,h} = z_{q,h}(s,t) - \frac{1}{9}\sum_{q',h'} z_{q',h'}(s,t), \qquad g_{q,h} = g_0 + (g_{\max} - g_0)\tanh(\widetilde{z}_{q,h}). \tag{43}$$

Directional kernels and gate-output layers are zero-initialized, so the initial adapter is the identity. Appendix A.4 connects these branches to sensor-dependent physical scales and directional RAD frequency response.

F.4 Projection-Aware SPC Profiles

Anchors are selected from the source RA maximum projection: positive, valid 5×5 local maxima are ranked by source strength, restricted to an interior margin $K = 8$, and greedily retained at pairwise Chebyshev distance greater than K, up to 32 per sample. Both prediction and target use

these same anchors and source-derived Doppler shifts. Predicted normalized values are mapped back to log-power units without clamping before profile extraction.

For a log-power cube V, define $P_V(r,a) = \max_d V(r,a,d)$. At anchor $\alpha = (r_\alpha, a_\alpha)$, the spatial profiles are

$$\ell_{R,\alpha}(j) = P_V(r_\alpha + j, a_\alpha), \qquad \ell_{A,\alpha}(j) = P_V(r_\alpha, a_\alpha + j), \quad -K \le j \le K. \tag{44}$$

For Doppler, let $d_\alpha = \arg\max_d X_1^{\text{raw}}(r_\alpha, a_\alpha, d)$ and $c = \lfloor D/2 \rfloor$. The full source-relative spectrum is

$$\ell_{D,\alpha}(j) = \max_{u,v\in\{-1,0,1\}} V(r_\alpha + u, a_\alpha + v, (j + d_\alpha - c) \bmod D), \quad 0 \le j < D. \tag{45}$$

Thus the R/A profiles each have 17 bins, while the Doppler profile has all $D = 64$ bins. The spatial maximum projections preserve moving echoes independently of the source Doppler bin. The small RA neighborhood tolerates registration and binning offsets, and the circular shift respects Doppler periodicity.

For each resulting log-power profile $\ell \in \mathbb{R}^{n_q}$, where $n_R = n_A = 2K + 1$ and $n_D = D$, SPC uses

$$e_i(\ell) = \tau \,\text{softplus}\left(\frac{\ell_i - Q_{0.25}(\ell)}{\tau}\right), \qquad p_i(\ell) = \frac{e_i(\ell) + \varepsilon}{\sum_j (e_j(\ell) + \varepsilon)}. \tag{46}$$

Subtracting the lower quartile removes a constant local floor before normalization. softplus keeps the response differentiable. We use $\tau = 0.5$ and $\varepsilon = 10^{-6}$. Each axis contributes an equally weighted normalized-profile ℓ_1 error, independently of its profile length. The implementation averages over selected anchors in the batch and returns zero when none is available. The invariance and associated bounds on response moments are derived in Appendix A.5.

G Extended Experimental Protocol and Results

G.1 Experimental Setup

We evaluate four radar domains from K-Radar (Paek et al., 2022), RaDelft (Roldan et al., 2024), and the cascade and single-chip configurations of ColoRadar (Kramer et al., 2022). Before constructing the splits, we remove sequences in which unreliable odometry supervision prevents consistent alignment between radar DA observations and motion flow, including tunnel scenes and extremely adverse-weather conditions. The reported sample counts therefore correspond to the filtered, motion-aligned subsets used in our experiments rather than the original dataset totals.

The sequence-disjoint train/validation/test splits contain respectively 7,348/2,982/2,666 K-Radar frames, 11,608/2,380/2,987 RaDelft frames, 21,307/4,365/3,706 ColoRadar-cascade frames, and 40,503/8,721/7,407 ColoRadar single-chip frames. Test data comprise 15 sequences (4/1/5/5 by domain). For balanced evaluation, we sample 128 temporally uniform frames from every sequence, yielding 1,920 generated cubes per method. Metrics are averaged within sequence, then within domain, and finally equally across the four domains, so the largest dataset cannot dominate the headline number. No adjacent frames cross a split boundary, and every normalization statistic is estimated from training sequences only.

Implementation details. Native 128^3 targets are max-pooled to 64^3 to retain localized peaks. LiDAR occupancy is sum-pooled. The U-Net uses channels $48/96/192/384$, 50.31M parameters, and a 192-dimensional condition code. RSA contributes 130,806 parameters (0.26%). We train with AdamW at 2×10^{-4} learning rate and 10^{-4} weight decay, 1,000-step warm-up followed by cosine decay, bf16 precision, gradient clipping at 1.0, and EMA decay 0.99. Training uses an NVIDIA L40S and effective batch size 24 for the matched bridge experiments. The main bridge uses $\beta_{\max} = 0.6$, 1,000 training intervals, at most 32 anchors with R/A radius eight and full circular Doppler profiles, MATC lags $\{1, 2, 4\}$, and a 500-step auxiliary warm-up. The matched-exposure RadarSimRAD and ablation checkpoints use 10,000 optimizer steps and 240k examples. The DDPM and flow controls use 232k examples. We additionally report the predetermined final RadarSimRAD checkpoint at 14,500 steps (348k examples) to separate exposure from mechanism. Unless noted, the controlled generators use EMA weights, one stochastic draw with a fixed per-frame seed, and NFE$= 8$. Their sample identities and sampling seeds are shared.

Baselines. The conditional DDPM uses the same four-stage U-Net, RA/DA encoders, sensor/style conditioning, splits, and normalization, with a cosine schedule, v-prediction, Min-SNR weighting, and DPM-Solver++ sampling. The flow-matching control uses the same predictor and conditioning with rectified-flow vector-field regression and Euler sampling. Neither matched baseline uses RSA, SPC, or MATC. We also ablate MATC and RSA individually from the paired bridge while retaining all other hyperparameters. For the closest external baseline, L2RLDB, we use a domain-adapted reproduction trained in three stages separately on each domain using our sequence split, and evaluate its supplied validation exports on exactly the same frame identities as all other methods. Its native range–azimuth–elevation–Doppler prediction is reduced to RAD by elevation maximization and the documented native-grid conversion. Because the K-Radar model covers the first 48 of 64 output range bins, the matched comparison uses a common K-Radar voxel mask. The other domains retain all 64^3 voxels (Appendix E.1).

Metrics and statistics. Conditional fidelity is measured by voxel correlation, de-normalized log-power MAE, strong-response Peak-F1, and RA-projection correlation. Correlation removes additive calibration offsets, whereas MAE retains absolute log-power sensitivity. Reporting both prevents either pattern agreement or amplitude agreement from standing alone. Peak-F1 thresholds each ground-truth cube at its 0.995 valid-voxel quantile and applies the same absolute threshold to the prediction, penalizing missing strong responses and indiscriminate bright clutter. Spread is the source-anchored R/A/D profile distance from equation 7. These metrics can still favor an over-smoothed conditional mean, so we additionally measure off-support background formation. For off-support gradient RMS G_q along axis q, define $r_q = \max\{G_q^{\mathrm{pred}} / \max(G_q^{\mathrm{gt}}, \varepsilon), \varepsilon\}$ with numerical floor $\varepsilon = 10^{-6}$. The implemented errors are

$$\mathrm{AGLRE} = \frac{1}{3} \sum_{q \in \{R,A,D\}} |\log r_q|, \quad \mathrm{Worst} = \max_q |\log r_q|. \tag{47}$$

Background lag-1 correlation error evaluates local spatial dependence rather than the marginal intensity histogram. We also record the absolute error in the fraction of softplus floor-subtracted log-response mass outside the central five Doppler bins (Off-D), and diagnostic texture and background-standard-deviation ratios whose ideal value is one. Off-D is not a physical linear-power energy fraction. These ratio diagnostics are ranked by distance to one rather than treated as higher-is-better scores. To decouple evaluation from MATC, we additionally compare the complete normalized non-DC background power spectrum along each physical axis using one-dimensional Wasserstein distance (PSD-W) and Jensen–Shannon divergence (PSD-JSD). We further record axis spectral-centroid error (ASCE), spectral-anisotropy error (ASAE), normalized spectral-entropy error (PSE-E), and the R/A/D autocorrelation error over lags $\{1, 2, 4, 8\}$ (MS-ACF). None of these metrics, nor a Fourier transform or normalized autocorrelation, appears in the training objective. Appendix E gives their definitions and relation to the original metric family. The independent statistical unit is a held-out sequence. We use 100,000 domain-stratified paired bootstrap replicates for 95% intervals and an exact one-sided paired sign-flip test over all 2^{15} assignments, followed by Holm correction across the original 11-metric family.

G.2 Metric Conventions and Statistical Families

All volume scores use the same $[R, A, D]$ order. Corr, MAE, Peak-F1, RA Corr, AGLRE, and spectral/ACF diagnostics operate on de-normalized $\log(1 + \mathrm{power})$, with predicted log power clamped to be nonnegative. Peak-F1 requires positive values at or above the target's 0.995 valid-voxel quantile, uses exact voxel overlap without a spatial tolerance, and is averaged per frame. Spread is the SPC profile distance: it inverse-normalizes without the nonnegative clamp before quartile subtraction, softplus, and normalization. Its R/A windows have radius eight. Its Doppler profile spans the full circular axis. The off-support mask is valid volume intersected with zero source occupancy, not a semantic annotation of clutter.

For normalized non-DC spectra $P_q, \widehat{P}_q$ with K_q retained bins, let $F_q(k) = \sum_{j \le k} P_q(j)$ and define

$$\text{PSD-W} = \frac{1}{3} \sum_{q \in \{R,A,D\}} \frac{1}{K_q} \sum_{k=1}^{K_q} |\widehat{F}_q(k) - F_q(k)|. \tag{48}$$

Table 13: Equal-domain macro results on the held-out test split, using stochastic EMA sampling at NFE$= 8$. Bold indicates the best method.

Method	Corr ↑	MAE ↓	Peak-F1 ↑	Spread ↓	RA Corr ↑	AGLRE ↓	Worst ↓	Lag-1 ↓
Conditional DDPM	0.6568	0.6480	0.2349	0.5898	0.5966	0.6216	0.9445	0.0585
Flow matching	0.7020	**0.5980**	0.2884	0.5606	0.6683	0.5311	0.8176	0.0583
RadarSimRAD (10k)	0.6990	0.6170	0.3007	0.5489	0.6810	0.2750	0.3780	**0.0403**
RadarSimRAD (14.5k)	**0.7077**	0.6010	**0.3104**	**0.5423**	**0.6915**	**0.2734**	**0.3759**	0.0405

Table 14: Quality and end-to-end evaluation time on 32 frames per sequence. Time includes loading and metrics. Peak memory is 0.74 GiB for every row.

Method	NFE	Corr ↑	Peak-F1 ↑	AGLRE ↓	ms/frame ↓
RadarSimRAD	2	0.7455	0.3193	0.3289	54.64
	8	0.7081	0.3100	0.2755	168.80
	32	0.6628	0.2775	0.1172	644.72
DDPM	2	0.6820	0.1841	0.9640	48.57
	8	0.6565	0.2357	0.6285	144.76
	32	0.6291	0.2400	0.2809	533.20
Flow	2	0.7556	0.2799	0.6045	49.01
	8	0.7013	0.2877	0.5326	147.19
	32	0.6583	0.2674	0.2121	538.38

This is the discrete mean cumulative-spectrum distance used by the evaluator. PSD-JSD uses natural logarithms. The displayed $\rho/G/W$ abbreviate cube-level Corr/AGLRE/PSD-W, not scores computed from the plotted projection images. Spread and AGLRE are related to SPC and MATC, respectively. The spectral and normalized ACF diagnostics are evaluation-only, not statistically independent of finite differences.

All point estimates aggregate frame to sequence to domain, then average domains equally. The original test family contains 11 metrics. Later spectral, paired-core, and matched-validation families contain 2, 8, and 12, respectively, with one-sided tests. The eight-metric ensemble evaluation uses two-sided tests. Holm correction is performed within each family and method comparison, irrespective of how many columns are displayed. Final-checkpoint significance is not assigned to the 10k exposure-matched comparison. Intervals resample sequences within domain and are conditional on this sampling design. The single RaDelft sequence contributes no within-domain bootstrap variation.

G.3 Main Comparison

Table 13 supplements the main comparison with worst-axis and lag-1 errors. For the final checkpoint versus DDPM, all 11 original metrics remain significant after Holm correction ($p_{\mathrm{H}} \leq 0.0109$). Against flow, Peak-F1, Spread, and RA Corr have $p_{\mathrm{H}} = 0.0284/0.00073/0.00320$. Lag-1, AGLRE, Worst, texture-ratio closeness, and background-standard-deviation closeness improve on all 15 sequences ($p_{\mathrm{H}} = 0.00034$). Corr and MAE are not significant against flow ($p_{\mathrm{H}} = 0.2017/0.7630$). The final AGLRE advantages are 0.3482 over DDPM and 0.2577 over flow, with paired 95% intervals $[0.3400, 0.3563]$ and $[0.2472, 0.2704]$. Table 6 provides the corresponding per-domain values.

G.4 NFE–Quality–Cost Trade-off

Table 14 shows the fidelity–formation trade-off with sampling depth. The full five-point sweep is slightly non-monotonic at NFE$= 4$. From 4 to 32, AGLRE decreases from 0.3676 to 0.1172 while Corr decreases from 0.7314 to 0.6628. At NFE$= 8$, RadarSimRAD reduces flow's AGLRE by 48.3% with 14.7% higher end-to-end evaluation time. At NFE$= 2$, it obtains Corr 0.7455 in 54.64 ms/frame. All methods use 0.74 GiB peak evaluation memory.

H FROZEN-MODEL TRANSFER TO AN INDEPENDENTLY ACQUIRED SCENE

This appendix specifies the zero-shot cross-scene evaluation in Section 4.4: the training-domain model is applied to a separate indoor acquisition without parameter updates, a new learned sensor embedding, or target-fitted normalization. The evaluation tests transfer of the learned support-to-response mapping across the scene, acquisition pipeline, and radar hardware.

H.1 ACQUISITION, FROZEN INFERENCE, AND PAIRING

Acquisition. The independently recorded indoor sequence contains 1,523 LiDAR frames and 1,532 radar ADC frames over approximately 154 seconds. The rig combines a Hesai JT128 LiDAR and a TI AWR1843 radar with three transmitters and four receivers (Figure 6). Each radar frame comprises 128 TDM loops, three transmitters, four receivers, and 128 ADC samples. The supplied DSP forms range and Doppler FFTs of length 128, max-pools adjacent bins on those axes, and uses eight virtual azimuth channels with a 64-point zero-padded FFT. Its exported linear-power array has order $[R, D, A]$, which is transposed to the manuscript's $[R, A, D]$ order.

Frozen conditions and checkpoint. Inference uses the same final 14,500-step checkpoint as the test evaluation, EMA weights, NFE$= 8$, and the existing ColoRadar single-chip condition code, selected by the hardware class before scoring. Zero-shot transfer to an unseen radar platform without target-specific parameter adaptation or a newly learned sensor embedding. Training uses only the four benchmark domains in Appendix C. All source and target normalization constants are frozen training statistics. The recorded radar is used for evaluation, not as a generator input or fine-tuning target. LiDAR point counts provide the scaffold. Point-intensity renders are not substituted for occupancy. Preprocessing removes ceiling points while retaining the ground, applies a 0.4 m near-range filter, shifts height by -0.13 m, and applies the fixed rig yaw of -45° to points and velocities using the independently validated radar–LiDAR extrinsic. KISS-ICP ego motion supplies static-world relative velocity $-(v + \omega \times p)$, projected onto the line of sight before the same alias-aware lifting as in Section 3.1. No independently moving-object velocity is supplied.

Time association. For each generated frame, we select the measured radar timestamp with the smallest absolute difference, without fitting a clock offset or optimizing association by image similarity. A fixed 100 ms gate, one nominal 10 Hz frame period, retains 1,518 of 1,523 generated frames. Five are excluded solely for temporal mismatch. The accepted pairs cover 1,504 distinct measured frames. The complete nearest-pair distribution has median 22.41 ms, 95th percentile 30.29 ms, and maximum 218.68 ms. Association need not be one-to-one, and no scene-motion warping is applied. Scores are means across the accepted, temporally dependent frames of this sequence.

H.2 TRANSFER METRICS AND INTERPRETATION

RA Corr retains the main paper's Pearson correlation of de-normalized $\log(1 + \text{power})$ maximum projections, evaluated after physical RA alignment. For the LiDAR input control, the corresponding transform is $\log(1 + \text{count})$. Because raw DSP gain is not radiometrically matched to the training radar, we add explicitly shape-based metrics rather than reinterpret absolute MAE or threshold-based Peak-F1 as calibrated cross-sensor accuracy. Let $Y_i \geq 0$ be the aligned linear-power RA maximum projection over the common pixel set Ω, and define

$$p_i = \frac{\widehat{Y}_i}{\sum_{j\in\Omega} \widehat{Y}_j}, \qquad q_i = \frac{Y_i}{\sum_{j\in\Omega} Y_j}, \qquad m_i = \tfrac{1}{2}(p_i + q_i). \tag{49}$$

We report

$$\text{RA-JSD} = \tfrac{1}{2} \sum_{i\in\Omega} \left[p_i \log \frac{p_i}{m_i} + q_i \log \frac{q_i}{m_i} \right], \qquad \text{RA-TV} = \tfrac{1}{2} \sum_{i\in\Omega} |p_i - q_i|. \tag{50}$$

Natural logarithms give ranges $[0, \log 2]$ and $[0, 1]$, respectively. These metrics remove a global multiplicative gain without fitting a gain to the measurement. They compare discrete mass on the shared bins, not a density per square meter or degree. RA-JSD is a spatial statistic, distinct from

the background spectral PSD-JSD in Appendix G.2. The source control replaces $\widehat{Y}$ with the count scaffold's RA projection. It is an input-to-output formation diagnostic, not a trained competing generator. The results shows all three metrics over every accepted frame: RA Corr improves from 0.3074 to 0.5623, RA-JSD falls from 0.4427 to 0.3098, and RA-TV falls from 0.7597 to 0.6522. The corresponding median scores are 0.2894/0.5836, 0.4570/0.2848, and 0.7825/0.6322 (source/generated). RA-JSD decreases on 1,477 of 1,518 paired frames. Its median paired change is -0.1377. These descriptive counts do not treat neighboring frames as independent statistical units.

With the single-chip style fixed before scoring, the improvement over the LiDAR source shows that the frozen mapping adds transferable radar-response structure on this acquisition. Evaluation across more hardware, scenes, and weather remains a natural extension.

H.3 Additional Visualizations of Zero-Shot Transfer

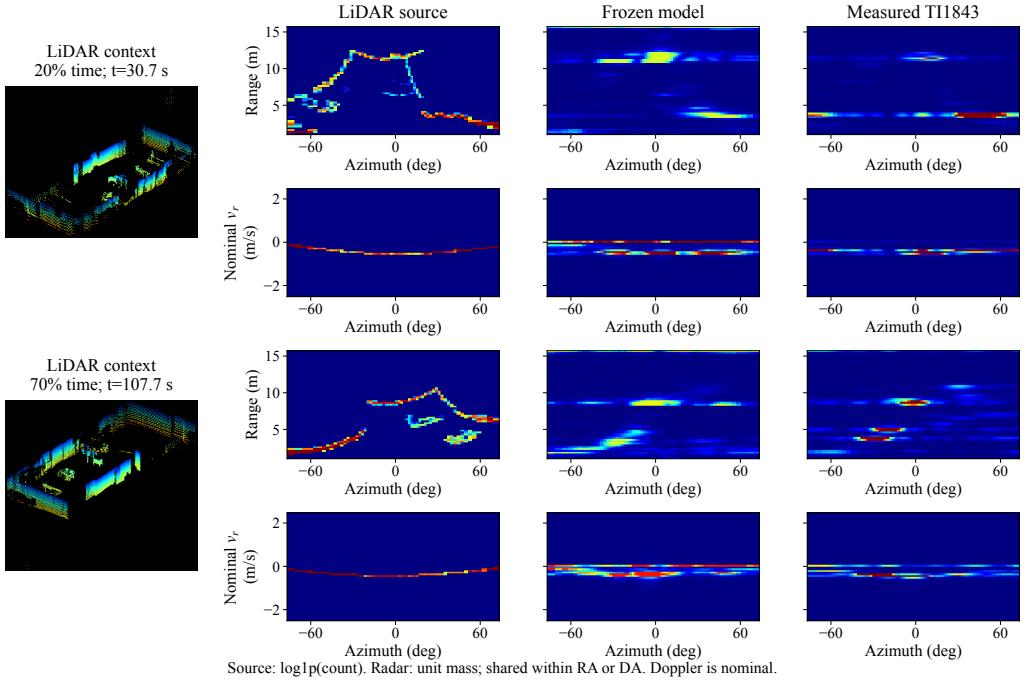


Figure 10: **Fixed-time examples from the independent acquisition.** LiDAR context and source, frozen-model output, and measured radar at the 20% and 70% frame positions. RA uses common physical coordinates. DA uses the nominal configuration-derived velocity axis. Radar colors represent unit-mass projection shape with shared limits, not calibrated absolute power.

I Downstream Augmentation under Restricted Measured-RAD Availability

This appendix gives the complete protocol and availability sweep for the downstream study in Section 4.3. It tests whether generated RAD can complement measured inputs for task-specific learners, alongside the sequence-disjoint generative benchmark and independent TI AWR1843 acquisition.

I.1 Data, Mixing Protocol, and Generator Exposure

Populations and splits. The study uses K-Radar sequences where eligible frames have both a measured RAD cube and aligned annotations. They are divided 80/10/10 before constructing temporal windows, and no temporally adjacent frames from the same sequence appear across downstream train/val/test. Windows cannot cross boundaries. Validation and test inputs are always measured RAD, regardless of the training regime.

What the availability fraction controls. For each task, a fixed permutation with subset seed 43 orders its N training windows. The first $\mathrm{round}(pN/100)$ indices define the same subset for Real_p, Mix_p, and Gen_p. Real retains only these measured windows. Gen replaces their inputs with generated RAD. Mix retains their measured inputs and fills the remaining indices with generated RAD. Every window is entirely real or entirely generated, and generated coverage is complete for all three window lengths. Thus, Real and Gen have the same selected-window count, whereas Mix has N windows. Mix_{100} is identical to Real_{100}, not a repeated experiment. Because windows can share constituent frames, p is a window-level availability fraction, not an exact percentage of unique radar frames or acquisition time.

Supervision is retained. All regimes inherit labels from the same H5 source files, including those supplying the synthetic complement. Frame-to-frame pose matrices provide odometry and ego-motion targets. LiDAR occupancy provides RA spatial targets. This setting measures the value of cross-modal label reuse when measured radar inputs are withheld: p controls measured-RAD availability, while supervision remains available. Generator pretraining is not restricted by p.

I.2 Tasks and Metrics

Odometry. A two-frame 3D CNN predicts six pose-increment coordinates. Translation MAE is $\frac{1}{3}\sum_{q\in\{x,y,z\}} |\widehat{\Delta q} - \Delta q|$, averaged over test windows, in meters. Forward-axis MAE is also reported. Yaw MAE wraps the angular difference before conversion to degrees. These are frame-to-frame errors, not trajectory ATE, RPE, or drift percentages.

Ego-speed proxy. A two-frame 3D CNN with intensity-weighted pooling predicts the magnitude of forward displacement, $|\Delta x|$, from the pose-increment label. Its MAE is in m/frame. Converting to m/s at nominal 10 Hz would multiply the error by ten. The archived `speed_mae_moving` field includes stationary samples: per-sample evaluation falls back to ordinary MAE when no moving sample is present. We therefore report all-frame ego-speed-proxy MAE, not moving-only MAE. The separate MAE@80+ column restricts targets to the 80 km/h-or-above bin using 10 Hz conversion, but its error unit remains m/frame. We use the unsmoothed CSV. The optional nine-frame median-filtered output is not substituted into these results.

RA occupancy and foreground localization. The occupancy learner uses a one-frame 3D stem, Doppler max reduction, and a 2D U-Net. H5 LiDAR occupancy is thresholded and elevation-reduced to a binary RA target. Gaussian targets are used for training. F1 and IoU are RA-pixel metrics, not full-volume voxel metrics. The foreground learner uses 16 frames and a 3D-to-2D encoder–decoder with occupancy-centroid targets. Its region F1 applies radius-five binary dilation before computing pixel precision/recall. Peak F1 matches local maxima to occupancy-component centroids within five bins. Foreground IoU uses undilated masks at threshold 0.30. None of these is category-aware box detection or mAP. The foreground peak score also differs from the exact-voxel RAD Peak-F1 in Table 1.

Bounded cross-task summary. To summarize heterogeneous task units without averaging raw F1 and MAE, we use the fixed monotone transforms

$$\tau_{\uparrow}(m, \alpha) = 100 \tanh(m/\alpha), \qquad \tau_{\downarrow}(e, \sigma) = 100 \exp[-e^2/(2\sigma^2)], \tag{51}$$

and average four components:

$$\tau\text{-RAD} = \tfrac{1}{4}\left[\tau_{\uparrow}(\mathrm{F1}_{\mathrm{occ}}, 0.22) + \tau_{\downarrow}(\mathrm{MAE}_{\mathrm{vel}}, 0.20) + \tau_{\downarrow}(\mathrm{MAE}_{\mathrm{trans}}, 0.22) + \tau_{\uparrow}(\mathrm{F1}_{\mathrm{region}}, 0.50)\right]. \tag{52}$$

The same scales are applied to every regime and availability fraction, preserving each component's ordering. τ-RAD is a bounded cross-task coordinate. All comparisons remain accompanied by the native metrics from which it is computed.

I.3 Downstream augmentation across measured-RAD availability

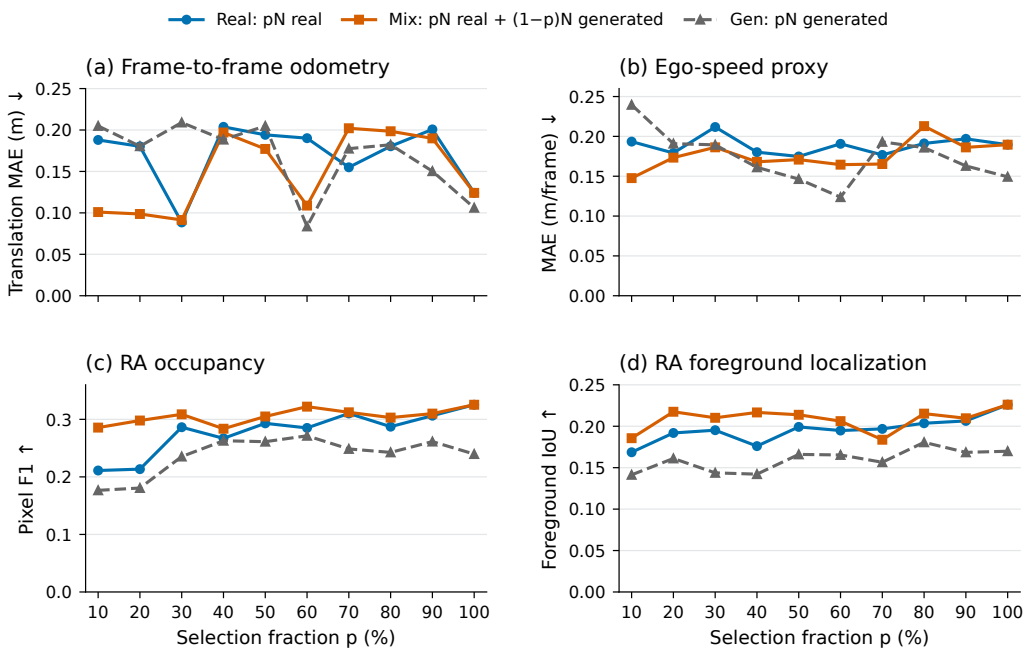


Figure 11: **Downstream augmentation across measured-RAD availability.** τ-RAD, its four native components, and its 10%-budget component scores, evaluated on measured task-test inputs across all availability settings. Real and Gen use the same selected fraction of training windows. Mix retains that real subset and fills the complement with generated inputs. Its x-axis is not a label or equal-update budget. Mix_{100} aliases Real_{100}.

At limited measured-RAD availability, synthetic augmentation provides substantial and consistent downstream gains. At 10% real-data availability, Mix reduces translation and ego-speed-proxy MAE by 46.29% and 23.68%, increases occupancy and localization F1 by 7.46 and 0.36 percentage points, and raises τ-RAD from 74.4 to 85.9. Similar gains persist at 20%, where translation and ego-speed-proxy MAE decrease by 45.27% and 3.25%, occupancy and localization F1 increase by 8.45 and 1.24 points, and τ-RAD improves from 76.0 to 84.5. Across the availability sweep, Mix improves τ-RAD in seven of nine settings, with consistent gains across translation, ego-speed estimation, occupancy, and localization. These results indicate that generated RAD is particularly effective at complementing limited measured observations, where additional response diversity translates directly into improved downstream supervision.

Synthetic RAD is most effective as a complement to limited measured RAD rather than a universal replacement for measured observations. At full-scale synthetic training, Gen_{100} reduces aggregate translation and ego-speed errors by 14.51% and 21.29% relative to Real_{100}, while closely matching localization region F1 (0.7710 versus 0.7727). Its overall τ-RAD further reaches 83.9 compared with

82.6 for Real_{100}, showing that the learned RAD distribution preserves task-relevant geometric and motion cues even without measured RAD as downstream training input. With the availability sweep in Figure 11, the result establishes generated RAD as an effective source of task-relevant training diversity rather than merely a visual surrogate for measured observations.

Implications for scarce paired RAD. RadarSimRAD converts readily available geometry and motion supervision into diverse, label-aligned radar observations, thereby extending the effective training distribution beyond the limited set of measured cubes. The pronounced gains in low-availability regimes suggest a practical role as a data engine for radar perception: measured RAD anchors the target domain, while generated responses expand scene and measurement coverage without requiring additional paired acquisition.