Moray: Spectrally Parameterized Neural Inverse Reconstruction

Harshvardhan C. Takawale
Department of Computer Science
University of Maryland, College Park
htakawal@umd.edu

Aritrik Ghosh
Department of Computer Science
University of Maryland, College Park
aghosh98@umd.edu

Nirupam Roy
Department of Computer Science
University of Maryland, College Park
niruroy@umd.edu

Abstract

In neural inverse reconstruction, the forward model is typically treated as a fixed simulator that maps a neural scene representation to measurements. This work studies how the parameterization of the forward operator itself shapes the optimization landscape of coherent inverse problems. We introduce *Moray*, a reconstruction framework built on spectrally parameterized forward operators that bypass conventional time-domain synthesis and instead evaluate measurements directly through closed-form spectral kernels. This parameterization induces a substantially shorter and better-conditioned differentiable graph compared to the time-domain approaches. We further introduce *Phase-Coherent Manifold Parameterization*, a scene representation that jointly learns scene reflectivity and a deformable surface manifold, allowing the reconstruction to adapt to geometric deviations from the assumed imaging plane and thereby reducing common artifacts in coherent reconstruction. We instantiate these ideas for 77 GHz radar imaging using real measurements from a synthetic-aperture mmWave system. Across multiple challenging scenes, Moray consistently improves reconstruction fidelity, suppresses background artifacts, and degrades more gracefully under data limitations. More broadly, the results suggest that, in neural inverse problems, exposing appropriate analytical structure of the sensing physics to the optimizer can be as important as the choice of neural representation itself.

1 Introduction

Neural Inverse Reconstruction and Forward Models. Many recent reconstruction methods formulate imaging as an analysis-by-synthesis problem: a neural scene representation is optimized so that measurements synthesized through a differentiable forward model match the observations. This paradigm underlies neural rendering and surface reconstruction methods and has increasingly been extended to physical sensing modalities including transient imaging, ultrasound, RF propagation, and coherent acoustics. In these systems, the forward model encodes the sensing physics that maps the scene field to measurements through propagation, scattering, sampling, and signal processing operations. As a result, optimization behavior depends not only on the neural representation itself, but also on how this physical operator is represented inside the differentiable pipeline. This becomes particularly important in coherent sensing systems such as radar, sonar, and optical coherence tomography, where measurements retain carrier-phase information and the forward model operates on highly oscillatory complex-valued signals.

40th Conference on Neural Information Processing Systems (NeurIPS 2026).

Limits of Conventional Parameterizations. Most existing differentiable sensing pipelines implement the forward model by directly mirroring the physical acquisition process. In coherent sensing systems, this typically involves synthesizing a time-domain signal by integrating scene responses against the transmitted excitation, sampling the resulting waveform, and finally applying spectral transforms such as the discrete Fourier transform before evaluating the reconstruction loss. While this formulation is physically faithful and straightforward to implement, it also fixes a particular computational structure for the differentiable operator. Gradients must propagate through the entire synthesis pipeline, including integration, sampling, and spectral transformation stages, regardless of whether all intermediate computations contribute useful information for reconstruction. In coherent sensing regimes operating at multi-GHz carrier frequencies, where measurements contain rapidly varying phase terms and scene information is often concentrated in a small subset of spectral components, this parameterization can lead to unstable optimization, unnecessary computation, and characteristic reconstruction artifacts. Importantly, these effects arise not from the sensing physics itself, but from how the forward operator is represented within the differentiable optimization pipeline.

Spectral Parameterization: The Key Insights. Rather than synthesizing the full time-domain signal and applying a spectral transform afterward, we parameterize the forward model directly in the spectral domain. For a broad class of coherent sensing systems, the output at each spectral bin can be expressed in closed form as an integral against an analytically derived complex kernel. This produces the same physical input-output mapping, but through a substantially shorter differentiable graph in which individual spectral components can be evaluated independently. In coherent ranging systems, spectral bins correspond directly to geometric quantities such as propagation delay and scene depth, and only a subset of bins typically contains scene-supported information for a bounded imaging volume. The spectral parameterization therefore enables supervision to be restricted to geometrically meaningful bins without requiring synthesis of the full signal, while preserving the carrier-phase structure of the measurement explicitly throughout optimization. We further pair this forward-operator formulation with a complementary scene parameterization that allows reconstruction geometry to adapt to deviations from the assumed imaging plane, reducing characteristic artifacts that arise under coherent measurements. Together, these ideas form the basis of *Moray*, our framework for coherent neural reconstruction.

Contributions and Impacts. Advances in coherent neural reconstruction benefit sensing and perception applications including robotic vision, autonomous operation, industrial inspection, and low-visibility safety systems, while also enabling lower-cost sensing from sparse measurements. At the same time, high-resolution hidden-object sensing may raise privacy and surveillance concerns, motivating responsible deployment. Fundamentally, this work studies how the parameterization of sensing physics within differentiable forward models influences optimization and reconstruction behavior in neural inverse problems. We study this through three contributions:

- We derive a closed-form spectral forward formulation for coherent sensing systems that enables direct per-bin evaluation and selective supervision over scene-supported spectral components.
- We propose Phase-Coherent Manifold Parameterization (PCMP), which jointly learns scene reflectivity and reconstruction geometry to mitigate fringe artifacts arising from geometric mismatch under coherent measurements.
- We validate these ideas on real 77 GHz synthetic-aperture radar measurements across multiple challenging scenes, where *Moray* improves reconstruction quality and robustness to data limitation compared to classical and neural reconstruction baselines.

2 Related Work

Inverse Problems and Differentiable Rendering. The analysis-by-synthesis paradigm originates with NeRF Mildenhall et al. [2021] and has since been extended to surface reconstruction Park et al. [2019], Mescheder et al. [2019], Wang et al. [2021], Yariv et al. [2021] and to physical modalities – transient light Malik et al. [2023], ultrasoundWysocki et al. [2024], RF channels Lu et al. [2024], and coherent acoustics Reed et al. [2023] (see Tewari et al. [2022] for a survey). Prior work largely takes the forward operator's parameterization as given; Moray instead studies how time-versus spectral-domain formulations shape optimization in coherent radar.

Signal and Spectral Representations. A separate line of work studies how signals themselves should be represented inside neural pipelines. Positional encodings Tancik et al. [2020], Mildenhall et al. [2021], sinusoidal activations Sitzmann et al. [2020], multi-resolution hash grids Müller et al.

[2022], band-limited coordinate networks Lindell et al. [2022], and wavelet activations Saragadam et al. [2023] all attack spectral bias in MLPs to recover high-frequency content. At the operator level, Fourier neural operators Li et al. [2020] parameterize a learned PDE kernel directly in Fourier space, showing that spectral parameterization can be substantially better-conditioned. *Moray* is closer to this second perspective, but rather than introduce a new input encoding for the INR, we parameterize the *forward operator* of a known physical measurement model in the spectral domain.

Physical Forward Models. Embedding a known physical model in a learned pipeline is a recurring strategy. Physics-informed neural networks Raissi et al. [2019] enforce PDE residuals at training time; differentiable signal processing Engel et al. [2020] treats classical DSP blocks as layers; and computational imaging routinely combines learned priors with model-based reconstruction Ulyanov et al. [2018], Venkatakrishnan et al. [2013]. These works use physical models as inductive bias for reconstruction, but typically adopt a fixed time- or space-domain formulation of the forward model. In contrast, we study spectral-domain parameterizations for coherent FMCW imaging and show that they enable shorter differentiable graphs, selective spectral supervision, and more stable optimization. Seismic full-waveform inversion likewise restricts the inversion to selected frequencies, either in the frequency domain Pratt [1999], Sirgue and Pratt [2004] or through a multiscale decomposition that starts from low frequencies Bunks et al. [1995], and its time-domain adjoint computations rely on checkpointing to reduce memory Griewank and Walther [2000], Symes [2007].

Radar Reconstruction. Classical radar imaging is dominated by backprojection, range-Doppler processing, and synthetic aperture pipelines that assume dense sampling and far-field geometry. Recent neural approaches like Radar Fields Borts et al. [2024], DART Huang et al. [2024], NeuRadar Rafidashti et al. [2025], and RadarHD Prabhakara et al. [2023] operate on post-processed 2D range–azimuth or range–Doppler tensors from planar MIMO arrays under far-field assumptions, targeting large-scale automotive or indoor scenes where coarse spatial resolution and magnitude-only representations suffice. The closest analog in coherent imaging is NVR for synthetic aperture sonar Reed et al. [2023], which uses a time-domain pulse-echo forward model. *Moray* instead operates directly on raw complex FMCW chirps and supervises reconstruction through analytically synthesized spectral bins. We focus on the coherent imaging regime where phase behavior and spectral parameterization significantly affect reconstruction quality.

3 Problem Formulation

This section formalizes the class of neural inverse problems we consider, in a form agnostic to whether the forward operator is expressed in the time or spectral domain.

Neural Reconstruction with Forward Models. Let $\Omega \subset \mathbb{R}^d$ be a bounded region of interest and $\sigma_\theta : \Omega \to \mathbb{R}$ a continuous map of the *scene field*, parameterized by an implicit neural representation (INR) with weights θ. A measurement is acquired under a configuration $p \in \mathcal{P}$, which collects all observation parameters that distinguish one measurement from another (e.g., sensor poses, channel index, or transmit signal parameters). For each p, the sensor produces a complex-valued observation $\mathbf{y}_p \in \mathbb{C}^N$ obtained from a known physical *forward operator*:

$$\Phi_p : \mathcal{V} \to \mathbb{C}^N, \qquad \mathbf{y}_p \;=\; \Phi_p[\sigma], \tag{1}$$

where $\mathcal{V} \subset \Omega$. Φ_p encodes the propagation, scattering, and sensing physics. The neural inverse problem is to recover σ from a finite set of observations $\{\mathbf{y}_p\}_{p\in\mathcal{P}}$ by optimizing θ so that the synthesized measurements $\hat{\mathbf{y}}_p = \Phi_p[\sigma_\theta]$ explain the observed ones.

Forward-Operator Parameterization. Although Φ_p in (1) is fixed by physics, the way it is *computed* in a differentiable pipeline is not. The same input–output mapping can be realized by different chains of intermediate variables, each producing the same $\hat{\mathbf{y}}_p$ but routing gradients to θ through different paths; we refer to such a choice as a *parameterization* of Φ_p. This is distinct from learning the operator itself, as in neural-operator approaches Li et al. [2020]: here the operator is fixed, and the design choice is how it is represented for differentiable evaluation. The choice of operator matters because optimization behavior depends on the structure of the differentiable graph, not just on the input–output map: two parameterizations that agree as functions of σ_θ can produce different graph depths, different gradient stability, and different opportunities for evaluating individual output components in isolation Engel et al. [2020].

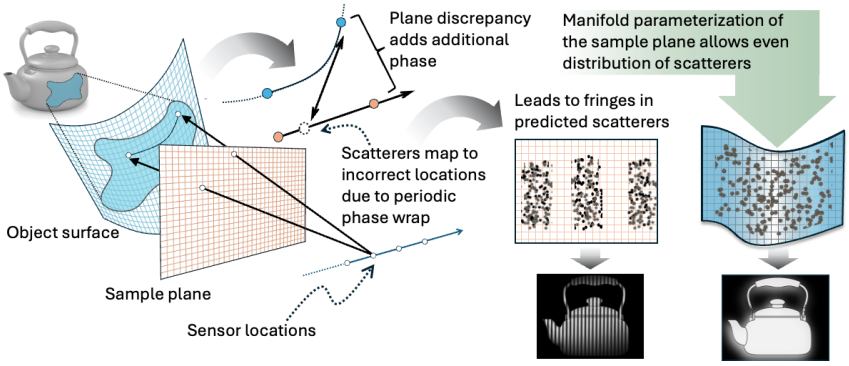


Figure 1: Phase-Coherent Manifold Parameterization mitigates the fringe effect due to discrepancy between the object surface and the sample plane.

Learning Objective and Supervision. Given a parameterization of Φ_p, training optimizes θ so that synthesized measurements explain observed ones. The objective takes the standard form

$$\mathcal{L}(\theta) \;=\; \sum_{p\in\mathcal{P}} \ell\big(\Phi_p[\sigma_\theta],\, \mathbf{y}_p\big), \tag{2}$$

where ℓ is a per-measurement discrepancy. Two design choices couple back to the parameterization. First, the form of ℓ is not pinned down by Φ_p: with complex-valued measurements, defining the discrepancy on the complex plane versus on magnitudes yields qualitatively different optimization behavior. Second, the loss need not be applied uniformly across all components of $\Phi_p[\sigma_\theta]$. When measurement components carry unequal information about σ (due to sensor noise or out-of-volume effects), restricting supervision to the informative ones can stabilize training and cut per-step cost, provided the parameterization exposes them individually.

4 Spectrally Parameterized Forward Models

This section develops the spectral parameterization in three steps.

(1) Closed-Form Spectral Forward Model. For a configuration $p \in \mathcal{P}$, the time-domain forward model maps a scene to a continuous-time signal $b_p(t)$ via integration against a known per-scatterer contribution; uniform sampling at rate $1/\Delta t$ producing a length-N discrete signal; a length-N DFT produces the per-bin measurement. Symbolically,

$$\Phi_p \;=\; \mathcal{F}_N \circ \Pi_{\Delta t} \circ T_p, \tag{3}$$

where $T_p : V(\Omega) \to L^2(\mathbb{R})$ is the modality-specific time-domain forward, $\Pi_{\Delta t}$ is the uniform sampling operator, and $\mathcal{F}_N : \mathbb{C}^N \to \mathbb{C}^N$ is the length-N DFT. The conventional differentiable parameterization implements Φ_p in the order written: synthesize $b_p(t)$ at N samples, then take the FFT, then index into the result.

A different parameterization is available when each scatterer's contribution to the time-domain signal has a tractable analytic form. Suppose that for a scatterer at $\mathbf{x} \in \Omega$, the per-scatterer time-domain contribution can be written as $g_p(\mathbf{x}, t)$, with g_p a known function of $\mathbf{x}$ and t encoding propagation, modulation, and sensing physics. Then the time-domain forward is the linear map $[T_p\sigma](t) = \int_\Omega \sigma(\mathbf{x})\, g_p(\mathbf{x}, t)\, d\mathbf{x}$. Composing with sampling and the DFT and exchanging the order of integration yields a per-bin closed form,

$$[\Phi_p\sigma](k) \;=\; \int_\Omega \sigma(\mathbf{x})\, h_p(\mathbf{x}, k)\, d\mathbf{x}, \qquad h_p(\mathbf{x}, k) \;=\; \frac{1}{N}\sum_{n=0}^{N-1} g_p(\mathbf{x}, n\Delta t)\, e^{-\mathrm{i}\,\beta_k n}, \tag{4}$$

where $\beta_k = 2\pi k/N$ is the angular frequency of the k-th DFT bin. The kernel $h_p(\mathbf{x}, k)$ is the N-point DFT of the per-scatterer contribution evaluated at the configuration p. We refer to Eq. (4) as the *spectral parameterization* of Φ_p: the same input–output mapping as Eq. (3), computed through a different chain of intermediate variables. This parameterization has two structural properties the time-domain composition does not. First, the bin index k enters h_p as an explicit parameter, so $h_p(\cdot, k)$ can be evaluated for any single k in isolation; the time-domain composition produces all N bins jointly through the FFT. Second, when $g_p(\mathbf{x}, n\Delta t)$ is a complex exponential in n, which

holds for the chirp modalities mentioned above, the geometric-series sum collapses to a closed-form Dirichlet-kernel factor that captures inter-bin spectral leakage exactly. We work out this collapse for FMCW radar in Section 6.

(2) Selective Spectral Supervision. Section 3 introduced a per-configuration discrepancy ℓ over the bins of $\Phi_p[\sigma_\theta]$ and noted that supervision need not be applied uniformly across all N bins. The spectral parameterization makes this concrete: with $h_p(\mathbf{x}, k)$ evaluable per bin, restricting the loss to a mask $M_p \subseteq \{0, \ldots, N-1\}$ requires evaluating only the kernel slices $\{h_p(\cdot, k) : k \in M_p\}$, with per-step compute scaling as $|M_p|$ rather than N. The time-domain composition of Eq. (3) cannot exploit this: the FFT's gradient is dense with respect to its input, so masking output bins after the FFT does not reduce the work of the backward pass through the time-domain forward pass and the FFT itself. Per-component supervision translates into per-component compute only when the parameterization exposes per-component access. The choice of M_p is a physical prior. Bounded scene support imposes an upper bound on the round-trip delay $\tau_p(\mathbf{x})$ over $\mathbf{x} \in \Omega$, which in turn restricts the bins where scene-supported energy can appear. Bins outside this support carry mostly sensor noise, out-of-volume scattering, and DC artifacts, and supervising on them contaminates the gradient. M_p should retain only the bins whose spectral support overlaps the scene under the modality's range-frequency mapping.

(3) Phase-Coherent Manifold Parameterization. A neural forward-model pipeline only sees measurements and cannot detect if the assumed scene region Ω is misaligned with the actual scatterers. When Ω is a fixed imaging plane and the object surface is tilted relative to it, the optimizer still deposits albedo somewhere in Ω to fit the measurements. Under a coherent forward model, depth mismatch induces a linear trend in round-trip delay that wraps into a periodic phase pattern, which the optimizer reproduces as fringes whose period is set by the carrier wavelength (Figure 1). At 77 GHz ($\lambda \approx 3.9\,\text{mm}$), even sub-centimeter mismatches produce visible fringes (Figure 6). We provide a detailed discussion on the wrapped-phase-depth ambiguity in Appendix E.

We solve this issue by reducing the constraints on scene representation by jointly learning a manifold $\mathcal{M}_\phi$ (which captures the object surface) and the intensity of the scatterer. The manifold is the representation of a learned depth-offset field over the image-plane region of interest: $\mathcal{M}_\phi = \{(x, y, z_0 + \delta_\phi(x, y)) : (x, y) \in \Omega_{xy}\} \subset \mathbb{R}^3$, where $\Omega_{xy} \subset \mathbb{R}^2$ is the ROI, z_0 is the nominal imaging plane, and $\delta_\phi : \Omega_{xy} \to [-\delta_{\max}, \delta_{\max}]$ is a learned depth-offset field with parameters ϕ. Reflectivity is the composition $\sigma_\theta(x, y)$ evaluated at the surface point. The integral becomes a surface integral over $\mathcal{M}_\phi$. Our forward model remains unchanged, only the domain onto which we perform integration changes. We refer to this as *Phase-Coherent Manifold Parameterization* (PCMP).

5 Optimization and Learning Dynamics

The spectral parameterization of Section 4 and the time-domain composition of Eq. (3) compute the same input–output mapping but route gradients through different intermediate operations. We isolate two consequences: per-step computational cost and empirical optimization behavior during training. Both are properties of the differentiable graph, not of the function being represented. In this view the forward model is part of the network for the purposes of optimization, and its analytic structure determines what the optimizer sees.

Gradient Sparsity and Per-Step Cost. Selective spectral supervision restricts the loss to a bin set M_p with $|M_p| \ll N$. The two parameterizations exploit this restriction differently. The Jacobian of the spectral forward operator is $J_p^{\text{spec}}[k, \theta] = \int_{\mathcal{M}_\phi} \nabla_\theta \sigma_\theta(\mathbf{x})\, h_p(\mathbf{x}, k)\, dA$, $k \in M_p$, which is computed by exactly $|M_p|$ matrix–vector products against bin-specific kernel slices, since $h_p(\cdot, k)$ is evaluable per bin (Eq. (4)). Its rank is bounded by $|M_p|$ by construction. The Jacobian of the time-domain operator with bin-selective masking, $J_p^{\text{time}} = \mathbf{P}_{M_p}\, \mathcal{F}_N\, J_{T_p}$ has the same rank $|M_p|$ but is computed through the full length-N forward and DFT regardless of the mask: the projection $\mathbf{P}_{M_p}$ can only act *after* both J_{T_p} and $\mathcal{F}_N$ have been evaluated. The backward pass therefore does $\mathcal{O}(N)$ work for an $\mathcal{O}(|M_p|)$-rank result. With $|M_p| = 7$ and $N = 256$ in our setup, the time-domain pipeline does $N/|M_p| \approx 37\times$ more per-step computation. Per-component supervision translates into per-component compute only when the parameterization exposes per-component evaluability of the operator. Section 7 reports the measured cost (Table 4).

Empirical Gradient Stability and Loss. Beyond per-step cost, the two parameterizations yield qualitatively different optimization behavior under matched compute. Figure 2 reports two com-

plementary measurements on a controlled comparison where the network architecture, optimizer, schedule, and active-bin set are held fixed and only the forward operator differs. Panels (a, b) show that the time-domain parameterization's gradient variance grows over training rather than stabilizing, while the spectral parameterization plateaus and stays tight throughout. Panels (c, d) show that this gradient behavior is consistent with the local geometry of the loss surface around the converged solution: the spectral parameterization sits in a single smooth, well-conditioned basin, while the time-domain parameterization's landscape contains multiple nearby local minima along the dominant gradient directions. An optimizer running on (d) sees a notably more rugged surface than one running on (c) for the same scene and training budget.

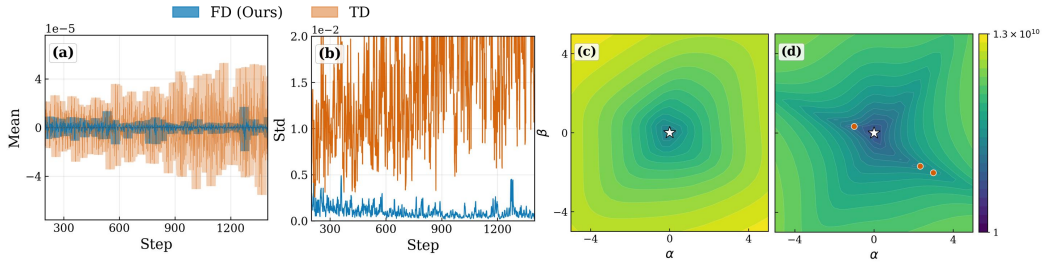


Figure 2: Optimization behavior of spectral (blue) and time-domain (orange) parameterizations. **(a, b)** Mean and standard deviation of first-layer MLP gradients over the first 1500 training steps. **(c, d)** Loss landscapes near the converged solution, marked by a white star, visualized along the top two PCA directions of parameter gradients (α, β). Orange dots in (d) are local minima.

Both observations trace back to where the bin-selection mask sits in the differentiable graph. In the time-domain parameterization, the projection $\mathbf{P}_{M_p}$ acts only on the output, after the time-domain integral, sampling, and DFT have already run. Because the DFT couples every time sample to every output bin, the backward pass routes minibatch noise from the $N - |M_p|$ unsupervised bins back into the gradient on θ, on top of the signal from the bins we actually supervise. The spectral parameterization avoids this by evaluating $h_p(\cdot, k)$ only for $k \in M_p$: the mask is enforced at the forward, not just the loss, and no gradient flows through bins that contribute nothing to the objective. Together with the $\approx 37\times$ per-step FLOP overhead, this puts the time-domain parameterization at a structural disadvantage that we expect contributes to the reconstruction-quality gap in Section 7, alongside the gradient stability shown above. More broadly, the spectral parameterization is an instance of a recurring pattern: changing how a fixed mapping is computed, without changing what it computes, can substantially reshape the optimization.

6 Instantiation: Radar Image Reconstruction

We instantiate the *Moray* framework on FMCW radar imaging application.

Signal model. An FMCW radar transmits a chirp $m(t) = \exp(\mathrm{i}\, 2\pi(f_0 t + \frac{1}{2} S t^2))$ for $0 \le t < T_c$, where f_0 is the start (carrier) frequency and $S = B/T_c$ is the chirp slope. A point scatterer at position $\mathbf{x} \in \Omega$ with reflectivity $\sigma(\mathbf{x})$ returns a delayed copy of $m(t)$, attenuated by $a_p(\mathbf{x})$ and offset by the round-trip delay $\tau_p(\mathbf{x}) = (\|\mathbf{o}_T(p) - \mathbf{x}\| + \|\mathbf{o}_R(p) - \mathbf{x}\|)/c$, where $\mathbf{o}_T(p)$ and $\mathbf{o}_R(p)$ are the transmitter and receiver positions for configuration p. The received signal is dechirped by mixing with $m^*(t)$, sampled at N uniformly spaced points $t_n = n\,\Delta t$ with $\Delta t = T_c/N$, and passed through a length-N DFT to produce the measurement $\mathbf{y}_p \in \mathbb{C}^N$. We index DFT bins by k as $\beta_k = 2\pi k/N$.

Spectral forward operator. Substituting the per-scatterer dechirped contribution into the kernel of Eq. (4) reduces the geometric-series sum to closed form. With normalized angular beat frequency $\alpha_p(\mathbf{x}) = 2\pi S\, \tau_p(\mathbf{x})\, \Delta t$, the per-bin operator becomes

$$[\Phi_p \sigma](k) \;=\; \int_\Omega \sigma(\mathbf{x})\, a_p(\mathbf{x})\, \underbrace{e^{\mathrm{i}\, 2\pi f_0\, \tau_p(\mathbf{x})}}_{\text{carrier phase}}\, \underbrace{\frac{1}{N}\, \frac{1 - e^{\mathrm{i}\, \alpha_p(\mathbf{x})\, N}}{1 - e^{\mathrm{i}(\alpha_p(\mathbf{x}) - \beta_k)}}}_{\text{Dirichlet kernel}}\, \mathrm{d}\mathbf{x}. \tag{5}$$

A complete derivation is given in Appendix D. The kernel exhibits the per-bin evaluability of Section 4: k enters as an explicit parameter, so the operator can be evaluated at any subset of bins independently. The carrier phase is preserved exactly, and the Dirichlet kernel captures inter-bin leakage in closed form as the N-point DFT of a windowed complex tone.

Selective bin supervision and loss. A scene bounded to maximum round-trip delay $\tau_{\max}$ produces beat frequencies in $[0, S\tau_{\max}]$, so the supervision mask M_p contains only bins $k \le \lceil S\tau_{\max}/\Delta f \rceil$ where $\Delta f = 1/T_c$. Equivalently, a scatterer at range R falls in bin $b(R) = 2BR/c = R/\Delta r$ with $\Delta r = c/2B = 4.165\,\text{cm}$, so only the bandwidth enters. In our geometry the nearest scene point ($R = z_0 = 0.25\,\text{m}$) falls in bin 6.0 and the farthest (aperture corner to opposite region corner, $R \approx 0.417\,\text{m}$) in bin 10.0. A margin of two bins below the near edge covers the Dirichlet main lobe (about one bin) and the learned depth offset ($|\delta_\phi| \le 20\,\text{mm}$, or 0.48 bin), which gives $M_p = \{4, \dots, 10\}$, 7 of $N = 256$ bins; Appendix A.2 reports the sensitivity to this choice. We apply the complex-residual discrepancy on the masked bins:

$$\ell(\hat{\mathbf{y}}, \mathbf{y}) \;=\; \sum_{k \in M_p} \Big[\big(\Re(\hat{y}_k - y_k)\big)^2 + \big(\Im(\hat{y}_k - y_k)\big)^2 \Big]. \tag{6}$$

A smoothness prior $\lambda_\nabla \, \mathbb{E}_{x,y} \|\nabla \delta_\phi(x, y)\|_2^2$ on the depth-offset field of Section 4 is added to the training objective; the weight and the schedule are reported in Appendix A.2.

7 Experiments

We evaluate *Moray* on real radar data acquired from a synthetic-aperture radar module, against learned and classical baselines.

Physical Data, Models, and Training. We use a TI AWR1843-class single-chip 77 GHz FMCW radar mounted on a two-axis motorized rail (Figure 3). We use the bandwidth of $B = 3.599\,\text{GHz}$ and start frequency $f_0 = 77.328\,\text{GHz}$, with $N = 256$ samples per chirp. Full hardware specifications, the multi-channel tiling protocol, and acquisition timing are given in Appendix C. The albedo field σ_θ is a multi-resolution hash-grid encoder Müller et al. [2022] followed by a 3-layer MLP of width 128, and the depth-offset field δ_ϕ is a 240×240 bilinear grid bounded to $|\delta_\phi| \le 20\,\text{mm}$. Encoder, MLP, and surface contribute $\approx 1.4 \times 10^8$, $\approx 21\,\text{K}$, and $\approx 58\,\text{K}$ parameters respectively. Encoder hyperparameters are in Appendix A. We optimize all parameters jointly with Adam at peak learning rate 10^{-3}, decayed via cosine annealing, for 10,000 steps with a batch of 64 TX configurations per step. The spectral forward model (5) is evaluated on the 7 active bins over the 240×192 scene grid; the loss is the complex residual (6) plus a smoothness regularizer $\lambda_\nabla \|\nabla \delta_\phi\|_2^2$ with $\lambda_\nabla = 10^{-3}$. Two training-time choices (1) a one-shot bias correction at step 0 and (2) a delayed unfreeze of δ_ϕ at step 5,000, improve stability and are detailed in Appendix A. Each run completes in ≈ 1.5 hours on a single RTX A5000.

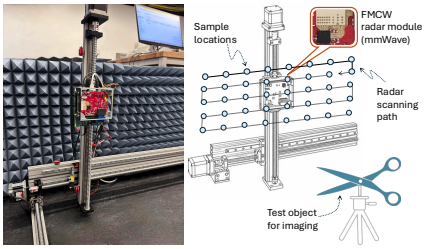


Figure 3: Physical radar setup.

Baselines and Evaluation Protocol. DART Huang et al. [2024], and Radar Fields Borts et al. [2024] address large-scale automotive and indoor scenes using planar MIMO arrays under far-field assumptions, operating on coarse range–azimuth or range–Doppler tensors where magnitude-only representations are sufficient. *Moray* instead targets a complementary approach through raw 1D complex FMCW chirps from a near-field synthetic aperture, where phase and finite-window spectral leakage dominate sub-wavelength reconstruction quality. Although not designed for this regime, these methods are the closest prior work in neural radar reconstruction; therefore these comparisons should be read as a positioning of design spaces rather than a head-to-head benchmark. DART expects range–Doppler images and Radar Fields range–azimuth scans, so we convert our complex measurements into each representation; the conversion runs one way and discards phase. Conversely, our operator uses the exact round-trip delay with no far-field approximation, but cannot be evaluated on their datasets, which contain no complex measurements (Appendix I). *Moray* is also compared against a non-learning baseline - Backprojection (Backproj.) Duersch [2013].

Metrics. Seven metrics are reported here and they are organized by reference signal (full definitions in Appendix B). *Photo-aligned image-domain metrics*, computed against a homography-warped optical photograph of the imaged object, score how faithfully the reconstruction reproduces the visible shape of the target: edge correlation (EdgeCorr ↑), LPIPS Zhang et al. [2018] (↓), PSNR (↑), and pixel Pearson correlation (↑). *No-reference metric*: target-to-background ratio (TBR ↑).*Held-out measurement metrics*, computed by re-rendering each method's reconstructed scene at TXs withheld from training (shared held-out split across methods): SSIM on the predicted complex beat signal magnitude (SSIM-magnitude ↑) and the equivalent Gaussian-noise SNR matching the residual MSE

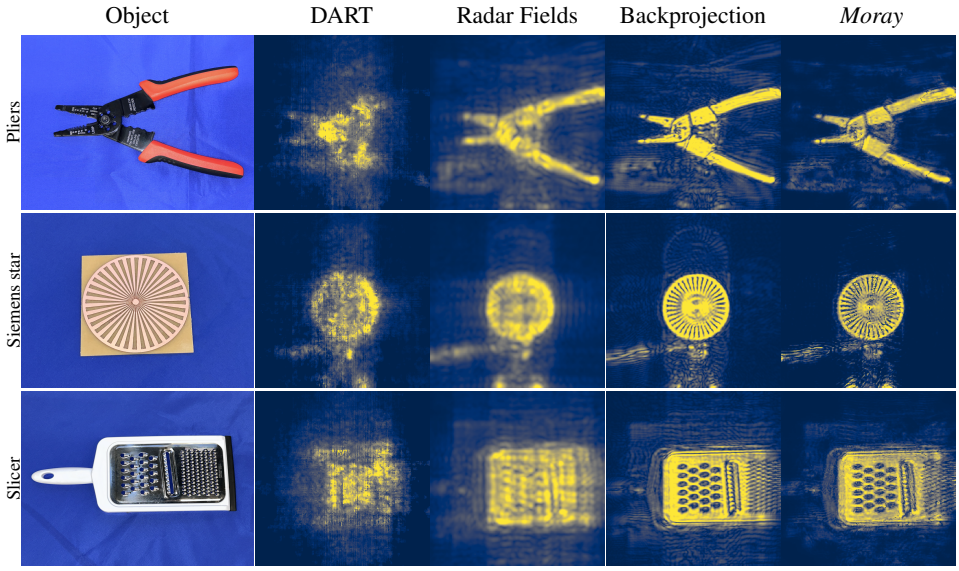


Figure 4: Qualitative comparison of representative objects. *Moray* gives the best reconstructions among the neural methods and the cleanest background; backprojection remains the visual reference at full aperture. Partial tripod structures are visible because objects were tripod-mounted.

Table 1: Full-aperture reconstruction quality. Four methods are shown across 5 datasets (scissor, pliers, pattern, slicer, siemens star) ↑/↓ denote higher/lower. Best values indicated in bold.

	Photo-aligned				No-ref	Held-out	
Method	EdgeCorr↑	LPIPS↓	PSNR↑	Pearson↑	TBR↑	SSIM-mag↑	dB↓
Moray	**0.159**	**0.446**	**9.98**	**0.34**	**287**	0.795	**-3.44**
Backproj.	0.141	0.49	9.64	**0.34**	65.4	**0.84**	8.12
Radar Fields	0.0652	0.783	8.8	0.319	23.6	1.01e-02	12.36
DART	0.0178	0.709	9.26	0.235	18.4	1.11e-02	11.03

(dB-noise ↓). SSIM-magnitude compares magnitude envelopes only, while dB-noise scores the complex residual and is the only metric that sees phase, which carries the geometry; 0 dB is the score of predicting an empty scene, and negative values predict held-out measurements better than that. The aperture-subsampling experiment reports binarised support metrics (SSIM-binary, F1@5%) alongside TBR and dB-noise.

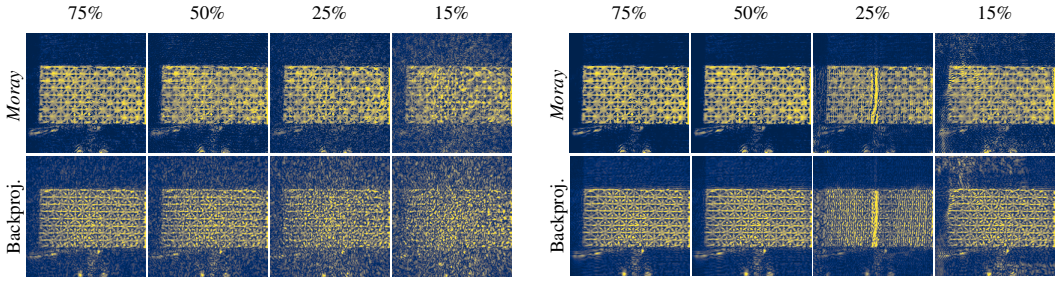
Table 1 reports quantitative results across five scenes; Figure 4 shows qualitative reconstructions. Due to the more stable spectral formulation, *Moray* produces superior reconstruction compared to neural baselines, and its background is far cleaner than that of every other method. At full aperture, backprojection remains the visual reference. Backprojection performs poorly in quantitative evalaution because backprojection puts 26.5% of a scatterer's energy outside its main lobe (integrated sidelobe ratio -4.43 dB). With many scatterers this energy fills the gaps between them. As a result objects look solid. To check this we ran a controlled experiment in simulation and found 5.9% of its energy lands in true gaps against 0.5% for *Moray*, lowering its TBR and edge correlation (Appendix J).

Impacts of Data Limitation. We probe behavior under reduced radar sampling by retaining a fraction $r \in \{75\%, 50\%, 25\%, 15\%\}$ of the original 240×240 sample positions and reconstructing with each method under the same training budget. Evaluation uses the binarised support metrics defined in Appendix B. We consider two subsampling strategies: (1) random subsampling of TXs across the aperture, and (2) uniform subsampling, which enlarges the inter-element spacing. The overall quantitative results can be found in Table 2 and qualitative results for each subsampling strategy can be found in Figures 5(a) and 5(b) respectively. The two subsampling strategies highlight different failure modes - a non-uniform aperture can lead to ghosting artifacts, and higher inter-element distance leads to aliasing in the image plane . *Moray* shows a graceful degradation across both subsampling strategies compared to backprojection. From 75% to 15% of the positions kept

Table 2: Effect of sub-sampling on object reconstruction

	TX (%)	SSIM-binary ↑		F1@5% ↑		TBR ↑		dB noise ↓	
		Moray	Backproj.	*Moray*	Backproj.	*Moray*	Backproj.	*Moray*	Backproj.
Random	75	**0.193**	0.170	**0.106**	0.103	**45.5**	15.8	**-1.97**	18.03
	50	**0.181**	0.161	**0.108**	0.102	**37.3**	11.1	**-1.98**	19.28
	25	**0.185**	0.119	**0.108**	0.099	**17.0**	7.0	**-1.45**	17.12
	15	**0.162**	0.079	**0.105**	0.099	**13.3**	7.0	**-0.16**	19.03
Uniform	75	**0.198**	0.175	**0.113**	0.103	**65.6**	29.8	**-2.22**	17.23
	50	**0.200**	0.175	**0.109**	0.103	**73.4**	29.6	**-2.22**	18.12
	25	**0.170**	0.110	0.104	**0.106**	**68.1**	36.1	**-0.14**	19.16
	15	**0.185**	0.160	**0.109**	0.102	**10.8**	6.3	**-2.01**	17.74

at random, the binarised SSIM of backprojection falls from 0.170 to 0.079, and that of *Moray* only from 0.193 to 0.162 as seen in Table 2.



(a) (b)

Figure 5: Effect of (a) random and (b) uniform subsampling. *Moray* is more robust than classical backprojection because its analysis-by-synthesis pipeline can synthesize novel signal views.

7.1 Ablation Study

To compare the effects of various components of *Moray*, we run ablations covering the relevant axes of the design space. To isolate forward-modeling and supervision effects, we define two variants: *TF-TS*, a time-domain forward model with time-domain ℓ_2 supervision adapted to FMCW; and *TF-SS*, the same forward model with spectral supervision (FFT before the loss), a hybrid that keeps the time-domain graph but supervises in frequency. We additionally ablate the PCMP and calibration-head components.

Calibration Head. Real-radar acquisition introduces per-channel residuals that survive standard hardware pre-processing: path-length mismatch between transmitters, phase-center calibration error, and sub-bin range offsets in FMCW chirp. We absorb these residuals through a *calibration head* $\mathcal{C}_\phi$ applied inside the differentiable forward model, parameterized by three real-valued scalars per-physical-channel: log-gain $a_c \in \mathbb{R}$, absorbing path-length and channel-gain mismatch; phase offset $\delta\varphi_c \in [-\pi, \pi]$, absorbing phase-center calibration residuals; range bias $f_c \in \mathbb{R}$ in DFT-bin units, implementing an $\exp(\mathrm{i}\, 2\pi f_c\, t)$ frequency shift inside the forward model that absorbs sub-bin range offsets in the chirp parameters [Yanik, 2020, Eq. 12].

With 8 monostatic channels, $\mathcal{C}_\phi$ adds only $3 \times 8 = 24$ parameters. The calibration head improves *Moray* across all metrics and objects, both quantitatively and qualitatively (Tables 3, 6; Figure 7). To confirm that it captures hardware-side residuals, we train *Moray* independently on five scenes and the learnt $(a, \delta\varphi, f)$ closely match across runs, indicating $\mathcal{C}_\phi$ models hardware rather than scene content. The evaluation for this experiment can be found in Appendix F.

PCMP Ablation. We isolate the contribution of phase-coherent manifold parameterization by holding everything else fixed—spectral forward operator, INR backbone, optimizer, training data, total compute—and varying only the scene parameterization. Figure 6 shows the effect of learning albedo only on the z_0 imaging plane such that δ_ϕ is not learnt and fixed at 0. Even a small mismatch between object plane and imaging plane leads to fringe patterns. PCMP alleviates this issue as it learns the contour of the object and learns δ_ϕ accordingly.

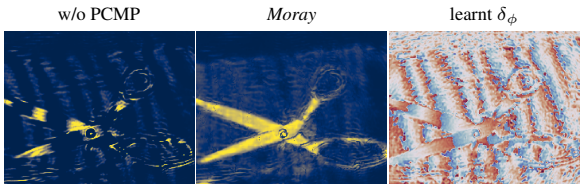


Figure 6: Scissors-scene reconstructions with different scene parameterizations. *Left:* volumetric. *Center: Moray* with PCMP. *Right:* learned δ_ϕ.

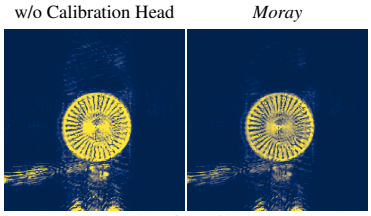


Figure 7: Lack of calibration leads to unclear reconstruction and ghost images (*left*).

Table 3: Reconstruction quality across radar datasets (scissor, pattern, pliers, siemens star, slicer).

	Photo-aligned				No-ref	Held-out	
Method	EdgeCorr↑	LPIPS↓	PSNR↑	Pearson↑	TBR↑	SSIM-mag↑	dB↓
Moray	**0.159**	**0.446**	**9.98**	**0.34**	**287**	**0.795**	**-3.44**
Moray w/o cal	0.151	**0.446**	9.67	0.337	192	0.776	-2.3
TF-SS	0.00641	0.663	9.79	0.0118	4.15	0.123	2.27
TF-TS	-0.0211	0.686	7.34	0.0711	69.2	0.00435	5.71e-06

Computational Cost. Table 4 compares the per-step cost of the spectral forward operator with that of the time-domain composition of Eq. (3), whose analytic ratio is $N/|M_p| \approx 37\times$ (Section 5). In complex64, the precision we train in, the spectral graph takes $19\times$ less time and $32\times$ less peak memory per step, and the gap grows linearly with N (Appendix H). Under a fixed memory budget the smaller graph affords a larger batch B, and since minibatch gradient variance falls as $1/B$ McCandlish et al. [2018], the spectral form needs 2.2–$4.1\times$ fewer optimizer steps to the same held-out target ($N = 256$ to $4{,}096$; Appendix H).

Table 4: Per-step cost of the forward operator (forward and backward pass), measured on an RTX A6000 at the training configuration ($46{,}080$ pixels, 64 TX configurations, $N = 256$, $|M_p| = 7$). Memory is the operator's peak, not the whole trainer's.

	complex64 (as trained)		complex128	
Forward operator	ms/step	Peak MB	ms/step	Peak MB
Spectral (*Moray*)	6.1	208	15.5	410
Time-domain	115.8	6,691	344.3	13,377
Ratio	19×	32×	22×	33×

8 Limitations and Conclusion

Limitations. Our claims are bounded by two modeling choices and by the scope of our data. First, the forward operator sums single-bounce returns from a view-independent reflectivity field, so it does not capture multipath or view-dependent scattering, which matter in cluttered or strongly curved scenes. Second, PCMP suppresses fringes by learning one depth per (x, y). This suits the near-planar objects imaged here but not volumetric or layered scenes. The spectral forward operator does not share this restriction and applies unchanged to volumetric fields. Third, all real-data results come from one 77 GHz radar and five tabletop scenes. Because true radar reflectivity cannot be observed, the photo-aligned metrics measure agreement with an optical proxy. Carrier generality rests on simulation based results, and transfer to FMCW LiDAR, OCT, ultrasonic FMCW or SAR is left for future work.

Conclusion. *Moray* introduces a spectrally parameterized forward model and a manifold scene parameterization (PCMP) for neural inverse reconstruction in coherent active sensing. Applied to 77 GHz FMCW radar imaging, *Moray* improves reconstruction quality over neural baselines and degrades more gracefully than backprojection under data limitation.

Acknowledgments and Disclosure of Funding

The authors would like to thank the reviewers for their helpful comments. This work was partially supported by NSF CAREER Award 2238433 and the Artificial Intelligence Interdisciplinary Institute at Maryland (AIM) seed grant. The authors declare no competing interests.

References

David Borts, Erich Liang, Tim Broedermann, Andrea Ramazzina, Stefanie Walz, Edoardo Palladin, Jipeng Sun, David Brueggemann, Christos Sakaridis, Luc Van Gool, et al. Radar fields: Frequency-space neural scene representations for fmcw radar. In *ACM SIGGRAPH 2024 Conference Papers*, pages 1–10, 2024.

Carey Bunks, Fatimetou M. Saleck, Stéphane Zaleski, and Guy Chavent. Multiscale seismic waveform inversion. *Geophysics*, 60(5):1457–1473, 1995.

Michael Israel Duersch. *Backprojection for synthetic aperture radar*. Brigham Young University, 2013.

Jesse Engel, Lamtharn Hantrakul, Chenjie Gu, and Adam Roberts. Ddsp: Differentiable digital signal processing. *arXiv preprint arXiv:2001.04643*, 2020.

Andreas Griewank and Andrea Walther. Algorithm 799: revolve: an implementation of checkpointing for the reverse or adjoint mode of computational differentiation. *ACM Transactions on Mathematical Software*, 26(1):19–45, 2000.

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 *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 24118–24129, 2024.

Zongyi Li, Nikola Kovachki, Kamyar Azizzadenesheli, Burigede Liu, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Fourier neural operator for parametric partial differential equations. *arXiv preprint arXiv:2010.08895*, 2020.

David B Lindell, Dave Van Veen, Jeong Joon Park, and Gordon Wetzstein. Bacon: Band-limited coordinate networks for multiscale scene representation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 16252–16262, 2022.

Haofan Lu, Christopher Vattheuer, Baharan Mirzasoleiman, and Omid Abari. Newrf: A deep learning framework for wireless radiation field reconstruction and channel prediction. *arXiv preprint arXiv:2403.03241*, 2024.

Anagh Malik, Parsa Mirdehghan, Sotiris Nousias, Kyros Kutulakos, and David Lindell. Transient neural radiance fields for lidar view synthesis and 3d reconstruction. *Advances in neural information processing systems*, 36:71569–71581, 2023.

Sam McCandlish, Jared Kaplan, Dario Amodei, and OpenAI Dota Team. An empirical model of large-batch training. *arXiv preprint arXiv:1812.06162*, 2018.

Lars Mescheder, Michael Oechsle, Michael Niemeyer, Sebastian Nowozin, and Andreas Geiger. Occupancy networks: Learning 3d reconstruction in function space. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4460–4470, 2019.

Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. *Communications of the ACM*, 65(1):99–106, 2021.

Thomas Müller, Alex Evans, Christoph Schied, and Alexander Keller. Instant neural graphics primitives with a multiresolution hash encoding. *ACM transactions on graphics (TOG)*, 41(4): 1–15, 2022.

Jeong Joon Park, Peter Florence, Julian Straub, Richard Newcombe, and Steven Lovegrove. Deepsdf: Learning continuous signed distance functions for shape representation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 165–174, 2019.

Akarsh Prabhakara, Tao Jin, Arnav Das, Gantavya Bhatt, Lilly Kumari, Elahe Soltanaghai, Jeff Bilmes, Swarun Kumar, and Anthony Rowe. High resolution point clouds from mmwave radar. In *2023 IEEE International Conference on Robotics and Automation (ICRA)*, pages 4135–4142. IEEE, 2023.

R. Gerhard Pratt. Seismic waveform inversion in the frequency domain; Part 1, theory and verification in a physical scale model. *Geophysics*, 64(3):888–901, 1999.

Mahan Rafidashti, Ji Lan, Maryam Fatemi, Junsheng Fu, Lars Hammarstrand, and Lennart Svensson. Neuradar: Neural radiance fields for automotive radar point clouds. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 2488–2498, 2025.

Maziar Raissi, Paris Perdikaris, and George E Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational physics*, 378:686–707, 2019.

Albert Reed, Juhyeon Kim, Thomas Blanford, Adithya Pediredla, Daniel Brown, and Suren Jayasuriya. Neural volumetric reconstruction for coherent synthetic aperture sonar. *ACM Transactions on Graphics (TOG)*, 42(4):1–20, 2023.

Vishwanath Saragadam, Daniel LeJeune, Jasper Tan, Guha Balakrishnan, Ashok Veeraraghavan, and Richard G Baraniuk. Wire: Wavelet implicit neural representations. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 18507–18516, 2023.

Laurent Sirgue and R. Gerhard Pratt. Efficient waveform inversion and imaging: A strategy for selecting temporal frequencies. *Geophysics*, 69(1):231–248, 2004.

Vincent Sitzmann, Julien Martel, Alexander Bergman, David Lindell, and Gordon Wetzstein. Implicit neural representations with periodic activation functions. *Advances in neural information processing systems*, 33:7462–7473, 2020.

William W. Symes. Reverse time migration with optimal checkpointing. *Geophysics*, 72(5):SM213–SM221, 2007.

Matthew Tancik, Pratul Srinivasan, Ben Mildenhall, Sara Fridovich-Keil, Nithin Raghavan, Utkarsh Singhal, Ravi Ramamoorthi, Jonathan Barron, and Ren Ng. Fourier features let networks learn high frequency functions in low dimensional domains. *Advances in neural information processing systems*, 33:7537–7547, 2020.

Ayush Tewari, Justus Thies, Ben Mildenhall, Pratul Srinivasan, Edgar Tretschk, Wang Yifan, Christoph Lassner, Vincent Sitzmann, Ricardo Martin-Brualla, Stephen Lombardi, et al. Advances in neural rendering. In *Computer Graphics Forum*, volume 41, pages 703–735. Wiley Online Library, 2022.

Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. Deep image prior. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 9446–9454, 2018.

Singanallur V Venkatakrishnan, Charles A Bouman, and Brendt Wohlberg. Plug-and-play priors for model based reconstruction. In *2013 IEEE global conference on signal and information processing*, pages 945–948. IEEE, 2013.

Peng Wang, Lingjie Liu, Yuan Liu, Christian Theobalt, Taku Komura, and Wenping Wang. Neus: Learning neural implicit surfaces by volume rendering for multi-view reconstruction. *arXiv preprint arXiv:2106.10689*, 2021.

Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004.

Magdalena Wysocki, Mohammad Farid Azampour, Christine Eilers, Benjamin Busam, Mehrdad Salehi, and Nassir Navab. Ultra-nerf: Neural radiance fields for ultrasound imaging. In *Medical Imaging with Deep Learning*, pages 382–401. PMLR, 2024.

Muhammet Emin Yanik. *Millimeter-wave Imaging using MIMO-SAR techniques*. The University of Texas at Dallas, 2020.

Muhammet Emin Yanik, Dan Wang, and Murat Torlak. Development and demonstration of mimo-sar mmwave imaging testbeds. *IEEE Access*, 8:126019–126038, 2020.

Lior Yariv, Jiatao Gu, Yoni Kasten, and Yaron Lipman. Volume rendering of neural implicit surfaces. *Advances in neural information processing systems*, 34:4805–4815, 2021.

Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 586–595, 2018.

Appendix

A Implementation Details

This appendix expands on the implementation summary in Section 7. Code and datasets are available on our webpage (`https://harshvardhan-takawale.github.io/Moray/`). All details about the hyperparameters, environment specifications, and real-world experiment setup are provided in the appendix or the website.

A.1 Network and surface parameterization

Hash encoder. We use the `tinycudann` HashGrid encoder configured with $L = 16$ levels, $F = 2$ features per level, $T = 2^{23}$ entries per hash table, base resolution $N_{\min} = 16$, and geometric per-level scale $b = 1.35$. The finest-level resolution is $N_{\max} = N_{\min} \cdot b^{L-1} \approx 1655$ voxels per side. Table entries total $\approx 1.4 \times 10^8$; we report the nominal table size, noting that the effective number of learned slots is bounded by the number of unique queried positions per level. Encoder weights are stored in fp16; the MLP and forward model run in fp32.

MLP. A 3-layer MLP with widths $32 \to 128 \to 128 \to 1$, SiLU activations between hidden layers, and Kaiming-uniform initialization. The first two linear layers have no bias; the final layer carries a single scalar bias, initialized to 10^{-3} and shifted at step 0 by the bias-match procedure described below. The output is passed through a softplus to enforce non-negative reflectivity without an upper bound. Total MLP parameter count: 20,609.

Surface module. The depth-offset field $\delta_\phi(x, y)$ is parameterized as a 240×240 grid of scalar values queried bilinearly at arbitrary (x, y) and then squashed via $\delta_{\max} \cdot \tanh$ with $\delta_{\max} = 20\,\text{mm}$. The grid is initialized to zero. This module contributes 57,600 trainable parameters and is registered as a separate parameter group in the optimizer so that its updates can be gated independently (delayed-δ schedule, below).

Coordinate normalization. The lateral region of interest is $[-12, 12]\,\text{cm} \times [-12, 12]\,\text{cm}$. World coordinates are linearly mapped to $[0, 1]^2$ before being passed to the encoder, with the depth coordinate $z = z_0 + \delta_\phi(x, y)$ rescaled to the same range using the same lateral half-width. The nominal imaging plane is at $z_0 = 0.25\,\text{m}$, and δ_ϕ provides up to $\pm 20\,\text{mm}$ of learned deformation about this plane.

A.2 Training protocol

Optimizer. Adam with $\beta_1 = 0.9$, $\beta_2 = 0.999$, $\varepsilon = 10^{-8}$, no weight decay. The encoder + MLP and the surface grid receive the same peak learning rate of 10^{-3} and share a single cosine schedule with $T_{\max} = 36{,}000$ and $\eta_{\min} = 10^{-6}$. Training is terminated at 10,000 steps—before the schedule reaches $\eta_{\min}$—and gradient norms are clipped to 1.0 on each parameter group separately.

Per-step compute. Each step samples a batch of 64 transmitter configurations from the 240×30 acquisition grid. For each configuration, the forward operator of Eq. (5) is evaluated on the 240×192 pixels of the imaging region and projected onto the 7 active spectral bins $M_p = \{4, \ldots, 10\}$, giving $\approx 2.95 \times 10^6$ scatterer-bin evaluations per step.

Sensitivity to the supervision band. Narrowing the window $M_p = \{4, \ldots, 10\}$ derived in Section 6, or shifting it by two bins, changes held-out fidelity by less than 2%. Seven bins drawn at random outside the scene band reach a held-out SSIM-magnitude of only 0.030, against 0.771 for the derived window on the same scene, and supervising all 256 bins lowers held-out fidelity by 2.5% while needing $8\times$ the peak training memory (48.4 vs. 6.1 GB).

Loss assembly.

$$\mathcal{L}(\theta, \phi) \;=\; \frac{1}{|B|} \sum_{p \in B} \frac{1}{|M_p|} \sum_{k \in M_p} \left[(\Re \hat{y}_{p,k} - \Re y_{p,k})^2 + (\Im \hat{y}_{p,k} - \Im y_{p,k})^2 \right] \;+\; \lambda_\nabla \, \mathbb{E}_{x,y} \|\nabla \delta_\phi(x, y)\|_2^2,$$

with $\lambda_\nabla = 10^{-3}$, $|M_p| = 7$, and $|B| = 64$ throughout training. Eq. (6) uses no separate magnitude term: we found no benefit from one in this setting, as the real and imaginary residuals already constrain the magnitude implicitly. The smoothness regularizer is gated to zero during the surface-frozen phase (steps 0–4,999) and active thereafter.

Initial bias matching. Before the first optimization step, we run one forward pass at $p = 0$ and compute the ratio $\rho = \log\big(\overline{|y_p|}/\overline{|\hat{y}_p|}\big)$, where the averages are taken over the active-bin set. We then add ρ to the bias of the final MLP layer. This eliminates a roughly 10^4 initial magnitude mismatch that would otherwise dominate the gradient at step 0; the procedure has no effect once gradient descent begins.

Delayed-δ schedule. The surface module δ_ϕ is held at zero for the first 5,000 steps by zeroing its gradients before the optimizer step. From step 5,000 onward, δ_ϕ is updated normally; the transition is hard, with no warmup ramp. The motivation is the PCMP decoupling rationale of Section 6: with $\delta_\phi \equiv 0$, the albedo field σ_θ converges against a flat reference plane at $z = z_0$, and geometric deformation enters only after the albedo has stabilized. Joint optimization from step 0 produced consistently worse reconstructions in our preliminary runs, with the surface module absorbing residuals that should have been encoded in σ_θ.

Schedule and command line. Table 5 lists when each component trains. The schedule-relevant flags of the training command are, verbatim:

```
--max_steps 10000 --delta_warmup_steps 5000 --plane_warmup_steps 5000 \
--cal_warmup_steps 3500 --init_bias_match_data --use_calibration \
--active_bins_lo 4 --active_bins_hi 10 --batch_size 64 \
--learning_rate 1e-3 --seed 42
```

`--active_bins_hi` is inclusive, so the active set $\{4, \ldots, 10\}$ has seven bins.

Table 5: Training schedule, in optimization steps.

Component	Schedule
Total optimization steps	10,000
Depth-offset field δ_ϕ	frozen 0–4,999; updated from 5,000
Smoothness regularizer on $\nabla\delta_\phi$	inactive 0–4,999; active from 5,000
Calibration head	frozen 0–3,499; updated from 3,500
One-shot bias match	single forward pass before step 0

Reproducibility. All RNG sources (Python `random`, NumPy, PyTorch CPU and CUDA) are seeded with 42 before model construction. `torch.backends.cudnn.deterministic` is not set, so kernel-level nondeterminism in `tinycudann` and PyTorch CUDA may produce small run-to-run variation; this is below the level of any reported metric difference in our experience. Repeating the full model with seeds 42, 1337 and 2024 gives a spread of ± 0.005 in held-out SSIM-magnitude and ± 0.007 in EdgeCorr. Software versions: Python 3.8, PyTorch 2.0.1+cu118, CUDA 11.8, `tinycudann` 1.7, NumPy 1.24.

Wallclock. A complete training run takes 1 hour 27 minutes on a single RTX A5000.

B Evaluation Metrics

This appendix gives the precise definition of each reconstruction-quality metric reported in the paper, the reference signal it compares against, and the direction of improvement. All metrics take a non-negative reflectivity field as input; the photo-aligned metrics are computed inside a per-scene evaluation polygon rather than over the full imaging region (see "Evaluation polygons" below).

Reference signals. Three reference signals are used: (i) a per-scene homography-warped grayscale photograph G of the imaged object, used for photo-aligned image-domain metrics; (ii) no reference, for TBR; (iii) the measured complex beat signal $\mathbf{y}_p$ at a held-out subset $\mathcal{P}_{\text{test}}$ of TX configurations,

used to assess whether each method's reconstructed scene reproduces measurements it never saw during training. The same held-out split is used across all methods.

Common preprocessing. For every photo-aligned metric, both the reconstruction R and the warped grayscale G are mapped to $[0, 1]$ by clipping at the per-image 99th percentile and dividing by it. Photo polarity is normalised so that bright = object across all scenes.

Evaluation polygons. The photo-aligned metrics are computed inside one polygon per scene, annotated manually once on the homography-warped photograph, stored with the dataset, and applied unchanged to every method and every reference. The object occupies 48–69% of the polygon (mean 59%), so background agreement accounts for roughly 41% of the scored pixels. To check that the ranking does not depend on where the polygon sits, we re-scored all methods under 180 boundary perturbations within the annotation uncertainty: the EdgeCorr ordering between *Moray* and backprojection flips in only 1.1% of them.

B.1 Photo-aligned image-domain metrics (continuous)

EdgeCorr (↑). Pearson correlation between Sobel-gradient magnitudes of R and G,

$$\mathrm{EdgeCorr} \;=\; \mathrm{corr}\big(\,\|\nabla R\|_2,\, \|\nabla G\|_2\,\big), \qquad \|\nabla I\|_2 = \sqrt{(S_x{*}I)^2 + (S_y{*}I)^2},$$

with S_x, S_y the standard $3{\times}3$ Sobel kernels.

Pearson (↑). Pixelwise Pearson correlation $\mathrm{corr}(R, G)$.

PSNR (↑, dB). Peak signal-to-noise ratio on $[0, 1]$-normalised images, $\mathrm{PSNR} = -10\log_{10}\big(\mathrm{MSE}(R, G)\big)$ with peak value $\mathrm{MAX}_I = 1$.

LPIPS (↓). Learned Perceptual Image Patch Similarity Zhang et al. [2018] using the pretrained AlexNet backbone (`lpips` v0.1.4 with `net='alex'`). Inputs are the $[0, 1]$-normalised images linearly rescaled to $[-1, 1]$ and replicated to three channels.

B.2 Photo-aligned image-domain metrics (binarised)

The reconstruction is binarised by keeping the top $\tau\%$ brightest pixels: $R_\tau = 1\{R \geq \mathrm{percentile}(R, 100 - \tau)\}$. These metrics are used in the aperture-subsampling experiment, where the recovered support is the quantity of interest. The reference is the binary fill mask M derived from the photo annotation.

SSIM-binary (↑). Wang et al. [2004] SSIM (`skimage.metrics.structural_similarity`, $\mathrm{data_range} = 1$) between R_τ (cast to $\{0, 1\}$ floats) and M, at $\tau = 5\%$.

F1@5% (↑). Harmonic mean of precision $|R_5 \cap M|/|R_5|$ and recall $|R_5 \cap M|/|M|$ at $\tau = 5\%$.

B.3 No-reference metric

TBR (↑). Target-to-background ratio, $\mathrm{TBR} = \max_{i\in\mathcal{T}} R_i \,/\, \mathrm{mean}_{i\in\mathcal{B}} R_i$, where $\mathcal{T}$ and $\mathcal{B}$ are per-scene target and background regions. Aggregated across scenes by the geometric mean.

B.4 Held-out measurement metrics

For every method, the trained scene representation is used to predict the complex beat signal $\hat{\mathbf{y}}_p$ at each held-out $p \in \mathcal{P}_{\mathrm{test}}$, restricted to the active spectral bins $M_p = \{4, \ldots, 10\}$. Predictions and measurements are jointly normalised by the global mean magnitude $g = \mathrm{mean}_{p,k\in M_p} |y_{p,k}|$.

SSIM-magnitude (↑). 1-D SSIM (Wang et al. formula, $C_1{=}(0.01L)^2$, $C_2{=}(0.03L)^2$) between the flattened magnitude vectors $|\hat{\mathbf{y}}_p|$ and $|\mathbf{y}_p|$ over $\mathcal{P}_{\mathrm{test}} \times M_p$. This is the metric reported in the "SSIM-mag" column of Table 1; it does *not* measure image-domain similarity.

Equivalent dB-noise (↓). The reference SNR (in dB) at which i.i.d. Gaussian noise on the measurements would produce the same complex-residual MSE as this method. Lower is better. Predicting an empty scene scores exactly 0 dB, so a negative value means a reconstruction predicts held-out measurements better than an empty scene would. SSIM-magnitude compares magnitude envelopes only: two reconstructions can match them equally well while differing in whether they reproduce the phase that carries the geometry, and only the dB column separates them.

Aggregation. Per-scene values are aggregated to the single numbers reported in Table 1 by the arithmetic mean across the five scenes, except for TBR (geometric mean). Cells where a metric is undefined for a method are reported as "—" and excluded from aggregation. Bold marks the column-best with absolute tolerance 0.01 on dimensionless metrics, and 0.5 dB on PSNR / dB-noise.

C Hardware Setup Details

The TI AWR1843-class radar Yanik et al. [2020] provides 3 TX and 4 RX antennas; we activate only the 2 TX whose phase centers are co-located with the RX array along the vertical axis, yielding $2\times4 = 8$ virtual monostatic channels uniformly spaced at $\lambda_w/4$ along y, where $\lambda_w = c/(79\,\mathrm{GHz}) \approx 3.795\,\mathrm{mm}$ is the wavelength near the center of the chirp band. The synthetic aperture is acquired by a two-axis raster scan: 240 horizontal positions at $\Delta x \approx 1\,\mathrm{mm}$ spacing ($\approx \lambda_w/4$), each repeated at 30 vertical positions on the motorized rail with vertical step $\Delta y = 8\lambda_w/4 \approx 7.59\,\mathrm{mm}$ so that the 8 virtual channels of consecutive vertical scans tile uniformly along y. This produces an effective aperture of $\approx 240 \times 230\,\mathrm{mm}$ with $240 \times 240 = 57{,}600$ measurement points at uniform $\lambda_w/4$ density on both axes.

Chirp parameters (full). Slope $S = 70.295\,\mathrm{MHz}/\mu\mathrm{s}$, sample rate $f_s = 5\,\mathrm{MHz}$, $N = 256$ samples per chirp; bandwidth $B = S \cdot N/f_s = 3.599\,\mathrm{GHz}$. Start frequency $f_0 = 77.328\,\mathrm{GHz}$. Range-bin width $c/(2B) \approx 4.165\,\mathrm{cm}$; carrier wavelength $\lambda_c = c/f_0 \approx 3.877\,\mathrm{mm}$. The ratio $f_0/B \approx 21.5$ governs the wrapped-phase severity of Section 4.

D Derivation of Spectral Transformation for Forward Model

To clarify the mathematical foundation of our *spectral forward model* in Moray (Ours), we revisit the discrete Fourier analysis of a single complex tone. This explicit derivation highlights two properties that are crucial for radar reconstruction: (i) how a finite-length DFT distributes energy into neighbouring bins when a scatterer's beat frequency is *off-bin*, and (ii) why modelling both magnitude *and* phase is essential for physically faithful supervision.

Signal model. We consider the sampled beat signal from one ideal point scatterer,

$$S_t \;=\; M\,e^{\mathrm{i}(\alpha t+\phi)}, \qquad \alpha \;=\; \tfrac{2\pi f}{N}, \quad t = 0,\dots,N-1, \tag{7}$$

where M is the amplitude, ϕ the initial phase, f the (normalised) frequency and N the analysis window length.

DFT definition. The unit-normalised DFT coefficient at bin index k is

$$Z_k \;=\; \frac{1}{N}\sum_{t=0}^{N-1} S_t\,e^{-\mathrm{i}\beta_k t}, \qquad \beta_k \;=\; \frac{2\pi k}{N}. \tag{8}$$

Forming a geometric series. Substituting S_t and collecting constants gives

$$Z_k \;=\; \frac{M}{N}\,e^{\mathrm{i}\phi}\sum_{t=0}^{N-1}\big(e^{\mathrm{i}(\alpha-\beta_k)}\big)^t, \tag{9}$$

which is a finite geometric series with common ratio $r = e^{\mathrm{i}(\alpha-\beta_k)}$.

Closed-form sum. Using the standard sum formula $\sum_{t=0}^{N-1} r^{\,t} = (1-r^N)/(1-r)$ we obtain

$$Z_k \;=\; \frac{M}{N}\,e^{\mathrm{i}\phi}\,\frac{1-e^{\mathrm{i}(\alpha-\beta_k)N}}{1-e^{\mathrm{i}(\alpha-\beta_k)}}. \tag{10}$$

Eliminating the bin phase. Because $\beta_k N = 2\pi k$ is an integer multiple of 2π, $e^{-\mathrm{i}\beta_k N} = 1$ and (10) simplifies to the closed form used in the main paper:

$$Z_k \;=\; \boxed{\frac{M}{N}\, e^{\mathrm{i}\phi}\, \frac{1 - e^{\mathrm{i}\alpha N}}{1 - e^{\mathrm{i}(\alpha - \beta_k)}}}. \qquad \text{(5 revisited)}$$

Interpretation. If the scatterer frequency *exactly* matches a bin centre ($\alpha = \beta_k$), both numerator and denominator of (5 revisited) vanish; evaluating the limit (or returning to the original sum) yields $Z_k = M\, e^{\mathrm{i}\phi}$ for that bin and $Z_{k'} = 0$ elsewhere. For all *off-bin* frequencies ($\alpha \neq \beta_k$), (5 revisited) predicts a sinc-shaped spectral envelope—the familiar *spectral leakage* modelled. Capturing this leakage, including complex phase rotations, is what allows our frequency-domain supervision to remain physically consistent even when a scatterer's true delay does not align with a DFT bin centre.

E Derivation of the Wrapped-Phase Depth Ambiguity

This appendix establishes the wrapped-phase property stated in Section 4. The argument proceeds in three steps: (i) the carrier-phase factor in the spectral forward operator is exactly invariant under per-pixel depth shifts of $\lambda_c/2$; (ii) the Dirichlet-kernel factor acquires a small phase rotation under the same shift; (iii) the residual rotation scales as B/f_0 and is small in our regime.

Setup. Recall the spectral forward operator from Eq. (5):

$$[\Phi_p \sigma](k) \;=\; \int_{\mathcal{M}_\phi} \sigma_\theta(\mathbf{x})\, a_p(\mathbf{x})\, e^{\mathrm{i}\, 2\pi f_0\, \tau_p(\mathbf{x})}\, \frac{1}{N}\, \frac{1 - e^{\mathrm{i}\, \alpha_p(\mathbf{x})\, N}}{1 - e^{\mathrm{i}(\alpha_p(\mathbf{x}) - \beta_k)}}\, dA,$$

where $\tau_p(\mathbf{x}) = (\|\mathbf{o}_T(p) - \mathbf{x}\| + \|\mathbf{o}_R(p) - \mathbf{x}\|)/c$ is the round-trip delay and $\alpha_p(\mathbf{x}) = 2\pi S\, \tau_p(\mathbf{x})\, \Delta t$ is the normalized angular beat frequency. The integration is over the manifold $\mathcal{M}_\phi$ defined by $\delta_\phi(x, y)$.

Consider a per-pixel shift of the depth-offset field,

$$\delta_\phi(x, y) \;\to\; \delta'_\phi(x, y) \;\; \delta_\phi(x, y) + n(x, y)\, \tfrac{\lambda_c}{2}, \qquad n : \Omega_{xy} \to \mathbb{Z},$$

where $\lambda_c = c/f_0$ is the carrier wavelength. The corresponding round-trip delay shifts to

$$\tau'_p(\mathbf{x}) \;=\; \tau_p(\mathbf{x}) + \frac{2}{c} \cdot \frac{\partial \tau_p}{\partial \delta} \cdot n(x, y)\, \tfrac{\lambda_c}{2} \;=\; \tau_p(\mathbf{x}) + \frac{n(x, y)}{f_0} + O\big(\tfrac{n\, \lambda_c}{c}\big),$$

where the leading term assumes near-normal incidence on the imaging plane (so $\partial \tau_p / \partial \delta \approx 2/c$); deviations from normal incidence introduce a per-pixel cosine factor that we absorb into the small residual addressed below.

Step (i): Carrier-phase invariance. The carrier-phase factor under the shifted delay is

$$e^{\mathrm{i}\, 2\pi f_0\, \tau'_p(\mathbf{x})} \;=\; e^{\mathrm{i}\, 2\pi f_0\, \tau_p(\mathbf{x})} \cdot e^{\mathrm{i}\, 2\pi\, n(x, y)} \;=\; e^{\mathrm{i}\, 2\pi f_0\, \tau_p(\mathbf{x})},$$

where the second equality follows from $n(x, y) \in \mathbb{Z}$. The carrier-phase factor is therefore pointwise invariant under per-pixel shifts of $\lambda_c/2$, regardless of the integer field n.

Step (ii): Dirichlet-kernel rotation. The beat-frequency variable α_p shifts to

$$\alpha'_p(\mathbf{x}) \;=\; \alpha_p(\mathbf{x}) + 2\pi S\, \frac{n(x, y)}{f_0}\, \Delta t \;=\; \alpha_p(\mathbf{x}) + 2\pi\, n(x, y) \cdot \frac{S\, \Delta t}{f_0} \;=\; \alpha_p(\mathbf{x}) + 2\pi\, n(x, y) \cdot \frac{B/N}{f_0}.$$

The shift in the Dirichlet argument is therefore $\Delta\alpha = 2\pi\, n\, (B/N)/f_0$, much smaller than the bin spacing $\beta_k - \beta_{k-1} = 2\pi/N$ by the factor B/f_0. The Dirichlet kernel

$$D_N(\alpha) \;=\; \frac{1 - e^{\mathrm{i}\, \alpha N}}{1 - e^{\mathrm{i}(\alpha - \beta_k)}}$$

is smooth in α, so to first order

$$D_N(\alpha'_p(\mathbf{x})) \;=\; D_N(\alpha_p(\mathbf{x}))\, \big(1 + \mathcal{O}(B/f_0)\big),$$

acquiring a small multiplicative correction.

Step (iii): Magnitude of the residual. With the chirp parameters of our setup ($f_0 = 77.328\,\text{GHz}$, $B = 3.599\,\text{GHz}$), the ratio $B/f_0 \approx 1/21.5$. Per-bin Dirichlet-kernel rotation under a single $\lambda_c/2$ wrap is therefore bounded by approximately 4.7% in argument, and proportionally small in magnitude. Across the active-bin set M_p, the per-bin corrections do not coherently sum into a distinguishable signature: the gradient of the loss with respect to a uniform $n(x, y) \equiv n_0$ wrap is bounded by the same $\mathcal{O}(B/f_0)$ factor, far below the gradient noise floor under stochastic minibatching.

F Additional Evaluation

In this section we evaluate the performance of *Moray* and compare it across all datasets. Table 6 show that *Moray* consistently outperforms other ablation baselines across all dataset.

To confirm that it captures hardware-side residuals, we train *Moray* independently on five scenes and the learnt $(a, \delta\varphi, f)$ closely match across runs is consistent as shown in Figure 8. This shows that calibration head does not learn any random quantity but rather it learns the actual residual due to actual hardware imperfections.

Table 6: *Moray* reconstructs measurements at TX locations held out from training, with *Moray* achieving SSIM-magnitude of 0.72–0.913 across five datasets.

Method	Pattern	Pliers	Scissor	Siemens star	Slicer	Mean
Moray	**0.756**	**0.813**	**0.771**	**0.72**	**0.913**	**0.795**
Moray w/o cal	0.715	0.781	0.76	0.717	0.905	0.776
TF-SS	0.0878	0.114	0.142	0.125	0.145	0.123
TF-TS	0.0017	0.00321	0.00687	0.00386	0.00613	0.00435

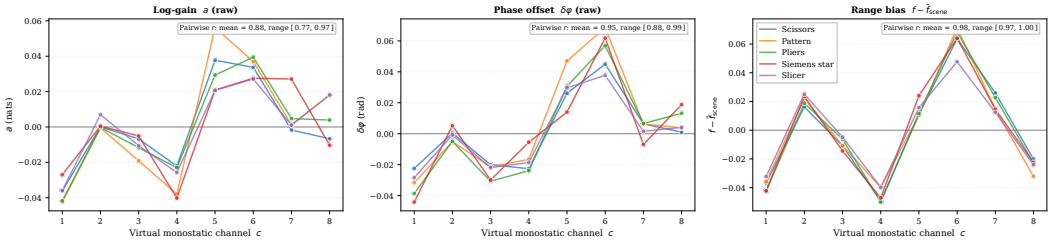


Figure 8: Cross-scene calibration consistency. Independently learned calibration parameters converge across scenes, suggesting that *Moray* learns hardware-specific imperfections.

G Effect of Tilt on Coherent Image Reconstruction

Forward-model based image reconstruction rests on two coupled estimates: the spatial location of each scatterer and its reflectivity at that location. Both are recovered by coherently combining the receive-side echoes against an assumed propagation geometry, so the fidelity of the assumed geometry directly bounds the fidelity of the reconstructed image.

Suppose the true surface is not flat but tilted slightly relative to the assumed plane. Each reflector then occupies a small nonzero displacement from where the imaging model places it, and the two-way path to that reflector differs from the path used during reconstruction by an amount ΔR. This mismatch enters the coherent sum as a residual round-trip phase $\Delta\phi = (4\pi/\lambda)\,\Delta R$. Because the displacement varies approximately linearly with position along the slope direction, the residual phase also varies linearly across the reflector array. A linear phase ramp summed coherently is precisely the operation that produces a periodic stripe pattern superimposed on the focused line.

G.1 Geometry and forward signal model

Let's assume a two-dimensional FMCW transceiver array with $N \times N$ elements at uniform spacing d lies in the plane $z = 0$, centred at the origin. The n-th element occupies the position $(x_n, y_n, 0)$.

A flat rectangular reflector patch of width W_x along $\hat{x}$ and W_y along $\hat{y}$ lies nominally at $z = z_0$. Each reflector at lateral coordinate $\mathbf{s}_m = (s_{x,m}, s_{y,m})$, with $|s_{x,m}| \le W_x/2$ and $|s_{y,m}| \le W_y/2$, contributes to a unit-amplitude monostatic round-trip return at carrier wavelength $\lambda = c/f_c$.

The patch is rotated by an angle θ about an in-plane axis $\hat{a}$ passing through the origin; lateral positions are held fixed and the height becomes

$$z(\mathbf{s}) \;=\; z_0 + (\hat{p}\cdot\mathbf{s})\sin\theta, \qquad \hat{p} \;=\; \hat{z}\times\hat{a}, \tag{11}$$

so $\hat{p}$ is the unit slope direction in the plane. After de-chirping, the narrow-band phasor at element n is the coherent sum

$$b_n \;=\; \sum_m \exp\!\left(j\frac{4\pi}{\lambda}R^{\mathrm{act}}_{n,m}\right), \qquad R^{\mathrm{act}}_{n,m} = \sqrt{(x_n - s_{x,m})^2 + (y_n - s_{y,m})^2 + z_m^2}, \tag{12}$$

with $z_m = z(\mathbf{s}_m)$. And reconstruction on the assumed flat plane evaluates the matched coherent sum as

$$I(x_p, y_p) \;=\; \sum_n b_n \exp\!\left(-j\frac{4\pi}{\lambda}R^{\mathrm{ass}}_{n,p}\right), \qquad R^{\mathrm{ass}}_{n,p} = \sqrt{(x_n - x_p)^2 + (y_n - y_p)^2 + z_0^2}, \tag{13}$$

which collapses to a focused image when the model and the truth agree.

G.2 Residual round-trip phase

Define the path mismatch between the actual and assumed geometries as

$$\Delta R_{n,m} \;\equiv\; R^{\mathrm{act}}_{n,m} - R^{\mathrm{ass}}_{n,m}, \tag{14}$$

where $R^{\mathrm{ass}}_{n,m}$ uses the un-tilted reflector position $(s_{x,m}, s_{y,m}, z_0)$. Expanding $R^{\mathrm{act}}_{n,m}$ to first order in $\sin\theta$ and using $z_m = z_0 + (\hat{p}\cdot\mathbf{s}_m)\sin\theta$ yields

$$\Delta R_{n,m} \;\approx\; \frac{z_0}{R^{\mathrm{ass}}_{n,m}}\cdot(\hat{p}\cdot\mathbf{s}_m)\sin\theta \;=\; (\hat{p}\cdot\mathbf{s}_m)\sin\theta\cdot\cos\beta_{n,m}, \tag{15}$$

with $\cos\beta_{n,m} = z_0/R^{\mathrm{ass}}_{n,m}$ the angle from broadside to the line of sight from element n to reflector m. The corresponding residual round-trip phase is

$$\Delta\phi_{n,m} \;=\; \frac{4\pi}{\lambda}\,\Delta R_{n,m} \;\approx\; \frac{4\pi\sin\theta}{\lambda}\,(\hat{p}\cdot\mathbf{s}_m)\cos\beta_{n,m}. \tag{16}$$

Two structural features of 16 drive everything that follows. The residual is linear in the slope-direction coordinate $\hat{p}\cdot\mathbf{s}$, and—in the paraxial regime where both the aperture and the rectangle are small compared with z_0, so $\cos\beta_{n,m}\to 1$—it is independent of the observing element n. A linear ramp in $\mathbf{s}$, sampled coherently by the imaging operator, is exactly the configuration that produces a periodic pattern in image space.

G.3 Image-domain fringe pattern and period

Substituting 16 into 13 and evaluating near a ground-truth reflector at $(x_p, y_p) = (s_x, s_y)$, the coherent sum reduces, in the paraxial limit, to

$$I(x_p, y_p;\theta) \;\approx\; I(x_p, y_p; 0)\cdot\exp\!\left(j\frac{4\pi\sin\theta}{\lambda}\,(\hat{p}\cdot\mathbf{r}_p)\cos\beta\right), \tag{17}$$

with $\mathbf{r}_p = (x_p, y_p)$ and $\cos\beta$ a representative broadside factor absorbed into a single multiplicative term. The argument is linear in $\hat{p}\cdot\mathbf{r}_p$, so $|I|$ is modulated by stripes oriented orthogonal to $\hat{p}$ with spatial period

$$\boxed{\Lambda_{\mathrm{fringe}} \;=\; \frac{\lambda}{2\sin\theta\cdot\cos\beta} \;\approx\; \frac{\lambda}{2\sin\theta} \quad \text{(paraxial)}.} \tag{18}$$

Table 7: Predicted ($N = 2W_{\hat{p}} \sin\theta/\lambda$, paraxial) versus observed fringe counts for three rotation axes.

Axis	θ	$N_{\text{predicted}}$	N_{observed}
$\hat{y}$ ($W_{\hat{p}} = 80$ mm)	5°	3.58	3.5
$\hat{y}$	10°	7.13	7.0
$\hat{x}$ ($W_{\hat{p}} = 40$ mm)	5°	1.79	2.0
$\hat{x}$	10°	3.57	3.5

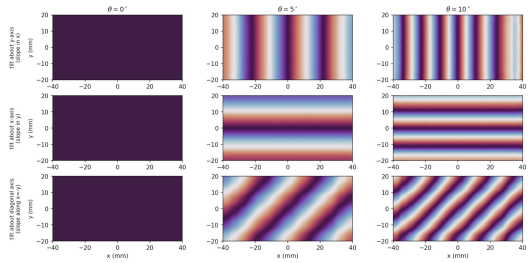


Figure 9: Fringe pattern of a 80 ×40 mm rectangle, with three rotation axes and three tilt angles.

G.4 Fringe-count formula

Dividing the rectangle's slope-direction extent $W_{\hat{p}}$ by Λ_{fringe} gives the number of full fringes across the image of the patch:

$$N_{\text{fringes}} = \frac{W_{\hat{p}}}{\Lambda_{\text{fringe}}} = \frac{2W_{\hat{p}} \sin\theta}{\lambda} \cdot \cos\beta \approx \frac{2W_{\hat{p}} \sin\theta}{\lambda}. \tag{19}$$

Three rotation axes specialize 19:

- $\hat{a} = \hat{y}$: $W_{\hat{p}} = W_x$; stripes orthogonal to $\hat{x}$ (vertical).
- $\hat{a} = \hat{x}$: $W_{\hat{p}} = W_y$; stripes orthogonal to $\hat{y}$ (horizontal).
- $\hat{a} = (\hat{x} + \hat{y})/\sqrt{2}$: $W_{\hat{p}} = (W_x + W_y)/\sqrt{2}$ measured corner-to-corner along $\hat{p}$; stripes oriented at 45°.

What appears in 19 is $W_{\hat{p}}$, θ, λ, and the broadside factor $\cos\beta$. What does *not* appear is the stand-off range z_0, the array aperture D, or the antenna count N: these set imaging resolution but not fringe periodicity. The fringe count is a property of the geometric mismatch alone—the rectangle's slope-direction extent, the tilt angle, and the wavelength.

G.5 Numerical verification

A controlled simulation at $f_c = 77$ GHz ($\lambda = 3.893$ mm) was used to validate 19. The setup places a 120×120 element array at 1 mm spacing and an 80×40 mm rectangular patch at $z_0 = 250$ mm. Tilt angles $\theta \in \{0^\circ, 5^\circ, 10^\circ, 15^\circ\}$ were applied about three axes ($\hat{y}$, $\hat{x}$, and the 45° diagonal), and the resulting images $|I(x_p, y_p)|$ were inspected for periodic structure.

Counting full periods directly from the simulated images and comparing with the paraxial prediction $2W_{\hat{p}} \sin\theta/\lambda$ gives the agreement summarized in Table 7 and visualized in Figure 9. Predictions match observations within $\pm 12\%$ across all rotation axes and angles, consistent with the residual paraxial error $\cos\beta \geq 0.985$ for the geometry above.

H Per-step cost and memory scaling

Table 4 compares the two forward operators at the training configuration. The two measurements below support Section 5. Both are operator-level: they measure the forward operator's differentiable graph, not the whole trainer, whose additional state lowers the time-domain batch ceiling further.

Scaling with record length. Holding the physical scene band fixed (fixed range resolution, longer unambiguous range), $|M_p|$ stays constant while N grows. The spectral cost is therefore flat in N and the time-domain cost grows linearly (Table 8). At $N = 16{,}384$ the time-domain graph needs 24.6 GB, while the spectral graph is unchanged at 13 MB.

Table 8: Per-step cost against record length N with the scene band fixed ($|M_p|$ constant), at a batch of 16 TX configurations, a smaller configuration than Table 4. Ratios are computed from the rounded values shown.

	Spectral		Time-domain		Ratio	
N	ms/step	Peak MB	ms/step	Peak MB	time	memory
256	1.8	13	5.6	482	3.1×	37×
1,024	1.8	13	20.8	1,538	11.6×	118×
4,096	1.8	13	82.1	6,147	45.6×	473×
16,384	1.8	13	326.3	24,582	181×	1,891×

Steps to a target at fixed memory. Minibatch gradient variance falls as $1/B$ McCandlish et al. [2018], so the batch a parameterization can afford under a fixed memory budget sets both the gradient noise the optimizer sees and the learning rate it can use. The spectral form's batch ceiling does not depend on N, while the time-domain ceiling falls as N grows: 45, 24 and 6 TX configurations at $N = 256$, $1{,}024$ and $4{,}096$. Table 9 counts optimizer steps to a fixed held-out target at matched objective and precision, so that only the affordable batch differs. These runs use their own batch sizes and learning rates, not the training setup of Appendix A.2 ($B = 64$, peak learning rate 10^{-3}). Across this batch range the usable learning rate rises $8\times$, from 2.5×10^{-4} at $B = 6$ to 2×10^{-3} at $B = 432$, and the standard deviation of the gradient noise falls by $\sqrt{432/6} \approx 8.5\times$.

Table 9: Optimizer steps to a fixed held-out target under a fixed operator-memory budget, at matched objective and precision. The time-domain form runs at the largest batch that fits; the spectral form at $B = 432$ for every N.

N	Time-domain batch	Time-domain steps	Spectral steps	Fewer steps
256	45	8,625	3,988	2.2×
1,024	24	9,875	3,988	2.5×
4,096	6	16,300	3,988	4.1×

I Baselines and carrier-frequency studies

DART and Radar Fields. Both methods were designed for planar MIMO arrays and consume post-processed tensors: DART Huang et al. [2024] takes range–Doppler images and Radar Fields Borts et al. [2024] range–azimuth scans. We converted our raw complex chirps into the representation each expects. The conversion runs one way, from complex measurements to power, so neither method sees the phase that our forward model uses. Conversely, our operator uses the exact round-trip delay with no far-field approximation, but running it on their captures would need complex measurements that those datasets do not contain. The comparisons that share our input, raw complex chirps, are the time-domain variants in the ablation of Section 7 and the NVR port below.

NVR port. NVR Reed et al. [2023] is the closest coherent analysis-by-synthesis method; it uses a time-domain pulse-echo forward model for synthetic-aperture sonar. We ported it to our geometry and verified the port in two ways. NVR's own backprojector, fed our adapted measurements, reconstructs the scissor scene at a normalized cross-correlation of 0.967 against our independent beamformed reference, and the port reproduces NVR's published in-air result end to end. We then simulated data with the bandwidth and range resolution fixed and only the carrier varied. NVR retains 79–92% of its resolution-limited reconstruction ceiling from 1 to 60 GHz, but only 45% at 77 GHz, where the coherent phase wraps about $77\times$ faster than at 1 GHz. Single-variable controls rule out range sampling, shell clipping and undertraining as causes. This demonstrates a mechanism on idealized simulation; it is not a claim about hardware performance.

Carrier-range simulation. The spectral kernel is derived for general chirp parameters (f_0, B, S, N). To test this beyond our hardware, we reconstruct simulated data on the same fixed 0.24 m rig at chirp start frequencies from 0.01 to 81 GHz with the bandwidth fixed. The range spans baseband, the legacy 24 GHz automotive band, the unlicensed 60 GHz band, our 77 GHz operating point and the 81 GHz band edge. Realistic per-channel hardware non-idealities (gain, phase, sub-bin range bias and noise) are injected, and each configuration recomputes its own supervision band (Section 6). Image SSIM against the known ground-truth albedo stays between 0.79 and 0.93 across the range, and the number of resolvable scene features scales as $1/\delta_{\mathrm{cr}} \propto f_0$, from one at baseband to about $1{,}250$ at 81 GHz. On hardware, we validate only the 77 GHz system.

J Full-aperture comparison with backprojection

At full aperture, backprojection Duersch [2013] is the visual reference: in Figure 4, *Moray* comes close to it but does not exceed it, while Table 1 favors *Moray* on most columns. This appendix explains the difference.

Point response. Simulating a single point scatterer on our acquisition geometry and reconstructing it with the same beamformer gives an integrated sidelobe ratio of -4.43 dB: 26.5% of a scatterer's energy lands outside its own main lobe. In a scene with many scatterers, that energy superposes into the gaps between them, filling interiors and producing a solid appearance; the streaks above and below each object in the backprojection column of Figure 4 are such sidelobe energy, not scene structure. On simulated data with exactly known reflectivity, backprojection places 5.9% of its energy in true gaps (pixels inside the object's envelope where there is no object), $12\times$ more than *Moray* (0.5%). *Moray* leaves those gaps empty, which looks less solid, but it agrees better with the photograph on where the edges are: EdgeCorr is 0.159 against 0.141, and 0.121 against 0.106 inside the object alone, so the difference is not a background effect.

Resolution. The spokes of a Siemens star rise in spatial frequency toward its centre, so one image gives a contrast-transfer curve. Measuring the modulation depth at the spoke harmonic against radius, *Moray* resolves spokes down to a radius of 23.5 mm, inside the 22.7–26.3 mm band set by the aperture's diffraction limit (cross-range resolution $\delta \approx 1.98$–2.29 mm). Radar Fields and DART show no resolvable spoke structure at any radius.

What each metric family measures. EdgeCorr correlates Sobel-gradient magnitudes with the warped photograph, so it responds to where edges sit. LPIPS is a learned perceptual distance over patches (0.446 for *Moray* against 0.49 for backprojection, and 0.476 against 0.519 on an object bounding-box crop). PSNR and Pearson measure pointwise agreement; they are dominated by large uniform regions and insensitive to edge placement, so their margins are small (9.98 against 9.64, and 0.34 for both). TBR divides the peak inside the target by the mean of the background (287 against 65.4). SSIM-magnitude and dB-noise are not image metrics: they re-render each reconstruction at withheld transmitters and compare it with the measured radar signal (Appendix B.4). The photo-aligned metrics are computed inside per-scene evaluation polygons in which the object occupies 48–69% of the pixels (Appendix B), so background agreement does not dominate them. The pointwise columns look alike because both methods place the object in the same place; the columns that separate the two methods are those tied to the task, edge structure and background suppression. Under aperture subsampling the gap widens: with 15% of the positions kept at random, binarised support SSIM is 0.162 against 0.079 (Table 2).

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The claims made in the abstract and introduction accurately reflect the paper's contributions and scope

   Guidelines:

   - The answer [N/A] means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A [No] or [N/A] answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: The limitations can be found in section 8.

   Guidelines:

   - The answer [N/A] means that the paper has no limitation while the answer [No] means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [Yes]

Justification: The paper provides detailed proofs in Appendix sections D, E and G

Guidelines:

- The answer [N/A] means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: The details are mentioned in Appendix section A.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - If the paper includes experiments, a [No] answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
   - While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
     (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
     (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
     (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
     (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: [Yes]

Justification: The details are mentioned in Appendix section A.

Guidelines:

- The answer [N/A] means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so [No] is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer) necessary to understand the results?

   Answer: [Yes]

   Justification: The details are mentioned in Appendix section A.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [Yes]

   Justification: Appendix A.2 reports the spread of held-out SSIM-magnitude and EdgeCorr over three random seeds.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The authors should answer [Yes] if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
   - The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
   - The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
   - The assumptions made should be given (e.g., Normally distributed errors).
   - It should be clear whether the error bar is the standard deviation or the standard error of the mean.

- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g., negative error rates).
- If error bars are reported in tables or plots, the authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: The details are mentioned in Appendix section A.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: The paper conforms, in every respect, with the NeurIPS Code of Ethics.

   Guidelines:

   - The answer [N/A] means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer [No], they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [Yes]

    Justification:The paper discusses potential societal impacts in the Section 1 of the paper.

    Guidelines:

    - The answer [N/A] means that there is no societal impact of the work performed.
    - If the authors answer [N/A] or [No], they should explain why their work has no societal impact or why the paper does not address societal impact.
    - Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.

- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate Deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pre-trained language models, image generators, or scraped datasets)?

    Answer: [N/A]

    Justification: The paper poses no such risks.

    Guidelines:

    - The answer [N/A] means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [N/A]

    Justification: The paper does not use existing assets.

    Guidelines:

    - The answer [N/A] means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.
    - For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
    - If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
    - For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.

- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [Yes]

    Justification: The details are mentioned in Appendix section A.

    Guidelines:

    - The answer [N/A] means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [N/A]

    Justification: The paper does not involve crowdsourcing nor research with human subjects.

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [N/A]

    Justification: the paper does not involve crowdsourcing nor research with human subjects

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
    - We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
    - For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does *not* impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

Answer: [N/A]

Justification: The core method development in this research does not involve LLMs as any important, original, or non-standard components.

Guidelines:

- The answer [N/A] means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
- Please refer to our LLM policy in the NeurIPS handbook for what should or should not be described.