MORPHOLOGY-GUIDED SPARSE STATE-SPACE ROUTING FOR SAR IMAGE RESTORATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Radio-frequency interference corrupts synthetic aperture radar (SAR) echoes and produces structured artifacts during coherent focusing, making pre-focusing recovery important for SAR image restoration (SAR-IR). However, fixed state-space scan geometries limit adaptation to diverse interference morphologies, while dense multi-directional scanning incurs redundant computation. We propose MorphRoute-Mamba, a time-frequency SAR-IR framework that learns morphology-guided routing over predefined directional state-space experts. Guided by full-record spectral context and Fast-SCNN-predicted interference masks, the framework selects a single expert per block and adaptively fuses its output with context from reliability-ranked anchors. This combines directional state-space propagation with reliability-guided contextual access. For efficient deployment, validation-guided route compilation fixes block-wise scan choices while retaining input-dependent conditioning and context fusion. Simulated experiments demonstrate reconstruction gains across the time-frequency, echo, and focused-image domains, while measured SAR scenes provide complementary evidence of interference suppression and structural preservation. Component ablations support the proposed design, and deployment profiling demonstrates a favorable reconstruction-latency trade-off.

1 INTRODUCTION

Synthetic aperture radar (SAR) enables high-resolution Earth observation under diverse environmental conditions, but its coherent imaging mechanism makes it vulnerable to radio-frequency interference (RFI) (Cumming & Wong, 2005; Meyer et al., 2013). We study SAR image restoration (SAR-IR) under RFI, targeting focused-image quality through pre-focusing echo recovery. RFI contaminates echoes before image formation and is propagated through coherent focusing. Let $\mathbf{R} = \mathbf{S} + \mathbf{N} + \mathbf{I}$ denote a raw SAR block containing scene echo, receiver noise, and interference. For fixed acquisition parameters, the focusing operator $\mathcal{P}$ approximately satisfies

$$\mathbf{Z}_{\mathrm{foc}} = \mathcal{P}(\mathbf{R}) = \mathcal{P}(\mathbf{S}) + \mathcal{P}(\mathbf{N}) + \mathcal{P}(\mathbf{I}), \tag{1}$$

indicating that residual interference is superposed on scene structures after focusing. Pre-focusing time-frequency domain (TFD) representations expose interference ridges, stripes, and intermittent supports before this redistribution. As illustrated in Figure 1, we perform SAR-IR by restoring complex TFD coefficients before inverse transformation and focusing. Here, interference morphology refers to these time-frequency support patterns, while propagation geometry denotes the token order used by a state-space scan.

Exploiting interference morphology remains challenging for SAR-IR. Traditional filtering, sparse decomposition, and low-rank approaches rely on manually designed assumptions, while representative learning-based methods use fixed architectures without explicitly selecting propagation directions from interference cues (Tao et al., 2019; Su et al., 2017; Shen et al., 2022). We use two complementary conditions: the full-record Fourier-domain (FD) representation supplies global spectral context, while an automatically predicted interference mask localizes candidate interference support in the TFD. The challenge is to integrate these conditions with propagation suited to different interference morphologies.

Recent state-space models (SSMs), particularly Mamba (Gu & Dao, 2024), provide long-range dependency modeling through linear-complexity sequence propagation. Serialization determines

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

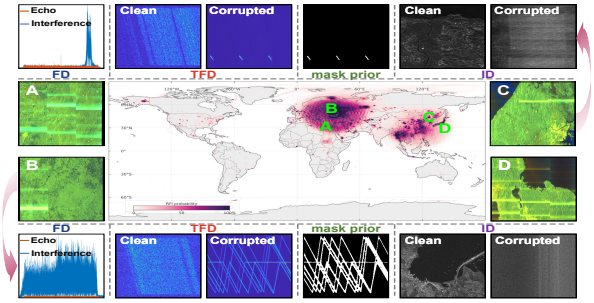


Figure 1: Domain motivation. FD supplies fine global band context; TFD and the predicted mask expose local morphology and candidate support; focusing redistributes residual RFI.

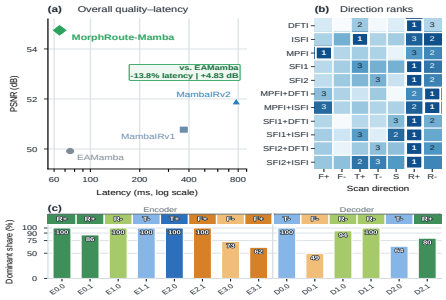


Figure 2: Routing motivation. Scan preference varies across interference types and network blocks, motivating sparse directional routing.

causal access and the sequence of state transitions, directly affecting information flow. For SAR-IR, the scan preferences in Figure 2 vary across interference types and network blocks, while dense multi-scan execution adds computation. These observations motivate learning morphology-conditioned selection within a predefined scan bank. The candidate permutations remain fixed; the learned routing policy chooses how features propagate through them.

Based on this insight, we propose **Morphology-Guided Sparse State-Space Routing**, implemented in *MorphRoute-Mamba* for SAR-IR. Frequency-Mask Compatibility Gating (FMCG) integrates spectral and mask conditions; the Sparse Directional Routing Module (SDRM) activates one expert from a predefined directional bank and combines its output with reliability-ranked anchor context. Transformer-to-Mamba distillation supports training, and validation-guided route compilation fixes one expert per block for deployment while retaining input-dependent conditioning, anchors, and fusion.

Our contributions are summarized as follows: ❶ We formulate morphology-guided propagation for SAR-IR and bound scan-dependent conditional perturbations from corrupted feature support. ❷ We develop MorphRoute-Mamba, coupling spectral-spatial conditioning, single-expert directional routing, and reliability-guided context fusion. ❸ We analyze validation-guided static route compilation and evaluate its reconstruction-latency trade-off. ❹ Simulated experiments show gains across transform, echo, and focused-image domains; measured scenes provide complementary evidence of clearer structures and lower relative background energy.

2 Related Work

2.1 SAR-IR and Pre-Focusing Interference Mitigation

Pre-focusing RFI mitigation supports SAR-IR through model-driven and learning-based approaches. Classical methods exploit spectral, time-frequency, sparse, low-rank, and tensor priors for interference separation (Tao et al., 2019; Zhang et al., 2011; Zhou & Tao, 2015; Tao et al., 2016; Su et al., 2017; Liu et al., 2018; Huang et al., 2019; Su et al., 2020; 2022). These approaches rely on signal assumptions that may be challenged by complex or overlapping interference. Learning-based methods exploit feature learning and prior guidance (Fan et al., 2019; Zhang et al., 2021; Xu et al., 2024; Shen et al., 2022; Fang et al., 2024). Generic image restoration (IR) methods based on CNNs, Transformers, and degradation conditioning (Chen et al., 2022; Liang et al., 2021; Wang et al., 2022; Zamir et al., 2022; Kong et al., 2023; Li et al., 2022; Potlapalli et al., 2023; Ren et al., 2026) provide useful restoration architectures. SAR-IR shares the goal of recovering image quality, but our setting restores complex echoes before coherent focusing and uses interference morphology in the TFD to guide propagation. MorphRoute-Mamba combines this routing with spectral context and predicted-mask support cues.

2.2 State-Space Restoration and Geometry-Aware Routing

SSMs support long-range dependency modeling through structured state-space layers, Mamba, and visual models operating on serialized representations (Gu et al., 2022; Smith et al., 2023; Gu &

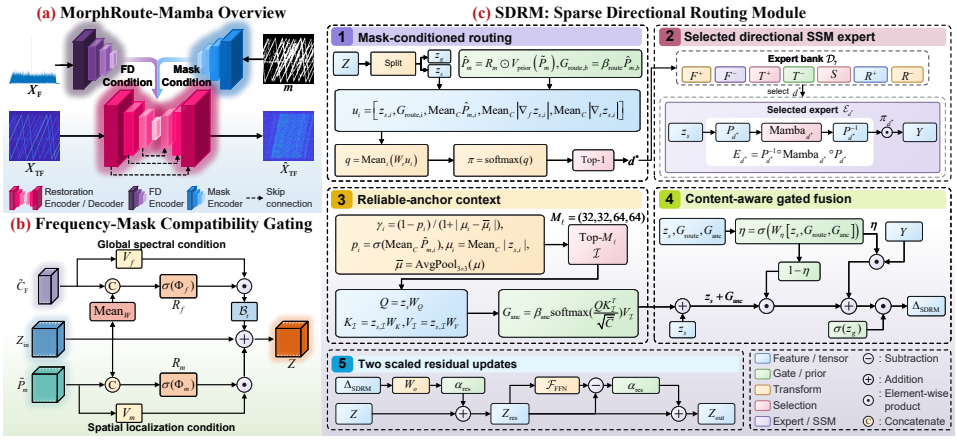


Figure 3: MorphRoute-Mamba: (a) conditional restoration network; (b) frequency-mask compatibility gating; (c) sparse directional routing with reliability-guided anchor retrieval and gated fusion.

Dao, 2024; Dao & Gu, 2024; Zhu et al., 2024; Liu et al., 2024). MambaIR, MambaIRv2, MaIR, and EAMamba investigate state-space modeling for IR (Guo et al., 2024; 2025; Li et al., 2025a; Lin et al., 2025). Adaptive scanning and routing approaches further modify traversals or select among experts (Zhou et al., 2025; Li et al., 2025b; Hu et al., 2025; Zhang et al., 2026; Wang et al., 2025; Lin et al., 2026). Our contribution is the SAR-IR-specific coupling of morphology-conditioned selection over axial, stripe, and ridge scan experts with reliability-ranked context retrieval. Spectral and predicted-mask conditions guide this computation, and validation-based compilation provides a fixed-route deployment option.

3 Problem Formulation

For SAR-IR under RFI, given a corrupted echo x_{td}, define $\chi(A) = [\mathrm{Re}\, A, \mathrm{Im}\, A]$ as stacking along the channel axis. The time-frequency and full-record Fourier views are

$$X_{\mathrm{TF}} = \chi(\mathrm{STFT}(x_{\mathrm{td}})), \qquad X_{\mathrm{F}} = \chi(\mathrm{FFT}(x_{\mathrm{td}})). \tag{2}$$

The mask $m \in [0,1]^{F_{\mathrm{TF}} \times T_{\mathrm{TF}}}$ is predicted from the corrupted TFD visualization by a frozen FastSCNN (Poudel et al., 2019) and used as a spatial condition for restoration during training and inference. Appendix B.2 details mask supervision and predictor selection. The RFI-free received echo $c_{\mathrm{td}} = s_{\mathrm{td}} + n_{\mathrm{td}}$ supplies the target $X^{\star}_{\mathrm{TF}} = \chi(\mathrm{STFT}(c_{\mathrm{td}}))$. We predict

$$\widehat{X}_{\mathrm{TF}} = f_\theta(X_{\mathrm{TF}}, X_{\mathrm{F}}, m). \tag{3}$$

The predicted coefficients are mapped to echo and focused-image domains by the fixed inverse transform and coherent focusing operator. Section 4.4 defines the supervised reconstruction objective and the training terms for distillation and sparse routing.

4 Method

4.1 Overview and Asymmetric Prior Conditioning

MorphRoute-Mamba uses an encoder–decoder backbone for SAR-IR, with four encoder stages (channel widths 32/64/128/256) and three skip-connected decoder stages. The deepest encoder stage forms the bottleneck. Each of the seven stages applies FMCG followed by two SDRM blocks, yielding 14 SDRM blocks in total (Figure 3). A convolutional head directly predicts $\widehat{X}_{\mathrm{TF}}$. FD and mask encoders supply aligned features $\widetilde{C}_{\mathrm{F}} \in \mathbb{R}^{C\times H\times 1}$ and $\widetilde{P}_m \in \mathbb{R}^{C\times H\times W}$. Suppressing batch/scale indices, FMCG maps the stage input Z_{in} to Z by

$$\begin{aligned} R_f &= \sigma\Big(\Phi_f[\mathrm{Mean}_W(Z_{\mathrm{in}}), \widetilde{C}_{\mathrm{F}}]\Big), \qquad & R_m &= \sigma\Big(\Phi_m[Z_{\mathrm{in}}, \widetilde{P}_m]\Big), \\ Z &= Z_{\mathrm{in}} + \mathcal{B}_t\Big(R_f \odot V_f(\widetilde{C}_{\mathrm{F}})\Big) + R_m \odot V_m(\widetilde{P}_m). \end{aligned} \tag{4}$$

Brackets concatenate channels; Φ, V are learned projections, σ is sigmoid, and $\odot$ is element-wise multiplication. The spectral gate $R_f \in \mathbb{R}^{C\times H\times 1}$ is broadcast in time by $\mathcal{B}_t$; the spatial gate $R_m \in \mathbb{R}^{1\times H\times W}$ broadcasts over channels. Both gates use the same Z_{in}. Each stage supplies R_m and $\widetilde{P}_m$ to its SDRM blocks; corresponding encoder/decoder stages share FMCG parameters, but recompute gates from their own features. Appendix C provides detailed network configurations.

4.2 Mask-Conditioned Sparse Directional Propagation

Experts and routing. On $\mathcal{G} = \{1,\dots,H\}\times\{1,\dots,W\}$, direction d defines a bijective scan-index map $\rho_d : \mathcal{G} \to \{1,\dots,HW\}$ and its feature permutation P_d. Each block has an independent bank

$$\mathcal{D}_7 = \{F^+, F^-, T^+, T^-, S, R^+, R^-\}, \qquad E_d = P_d^{-1} \circ \text{Mamba}_d \circ P_d. \tag{5}$$

Figure 4 illustrates the frequency-major, time-major, stripe-serpentine, and paired ridge orders. These permutations are predefined (Appendix A.1); expert parameters and the selection policy are learned.

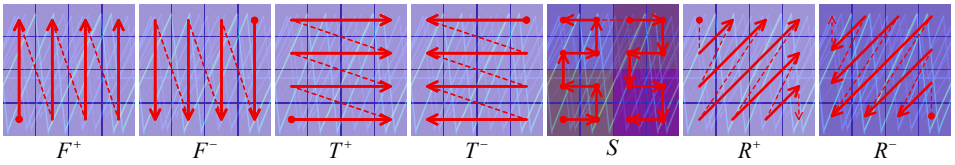


Figure 4: Seven predefined scan patterns in the directional expert bank.

Dynamic execution selects a direction per sample and block, while compiled execution uses the block-wise choices in Section 4.5. Normalization, local mixing, and projection split block input Z into $z_s, z_g \in \mathbb{R}^{\bar{C}\times H\times W}$, with $\bar{C} = 2C$. The block forms $\widehat{P}_m = R_m \odot V_{\text{prior}}(\widetilde{P}_m)$ and $G_{\text{route}} = \beta_{\text{route}}\widehat{P}_m$, using its own projection and learned sigmoid scalar. At position i, concatenate $u_i = [z_{s,i}, G_{\text{route},i}, \text{Mean}_C\, \widehat{P}_{m,i}, \text{Mean}_C\, |\nabla_f z_{s,i}|, \text{Mean}_C\, |\nabla_t z_{s,i}|]$. With logits $q = \text{Mean}_i(W_r u_i)$, each sample and block uses

$$\pi = \text{softmax}(q), \qquad d^\star = \arg\max_d \pi_d, \qquad Y = \pi_{d^\star} E_{d^\star}(z_s). \tag{6}$$

A single expert processes all spatial positions, weighted by its original softmax probability. For downstream loss $\ell(Y)$, holding expert outputs fixed and excluding selection ties gives

$$\frac{\partial \ell}{\partial q_r} = \pi_{d^\star}(\mathbf{1}[r = d^\star] - \pi_r)\,\langle \nabla_Y \ell, E_{d^\star}(z_s)\rangle\,. \tag{7}$$

The retained confidence passes task gradients to the router; Appendix A.3 details regularization.

Proposition 1 (scan-dependent conditional propagation). For serialized features x_n, fix the input-conditioned coefficients of $h_n = \overline{A}_n h_{n-1} + \overline{B}_n x_n$, $y_n = C_n h_n + D_n x_n$, with $h_0 = 0$. Assume $\|\overline{A}_k^{(d)}\|_2 \le e^{-\lambda_{d,k}}$, $\lambda_{d,k} \ge 0$, $\|\overline{B}_k^{(d)}\|_2 \le b_d$, and $\|C_k^{(d)}\|_2 \le c_d$. For $m = \rho_d(p) < n = \rho_d(q)$, let $\Lambda_d(p,q) = \sum_{k=m+1}^{n} \lambda_{d,k}$. For a feature perturbation e supported on $\mathcal{S}$, the conditional kernel and output change satisfy

$$K_d(p,q) = C_n^{(d)}\overline{A}_n^{(d)} \cdots \overline{A}_{m+1}^{(d)}\overline{B}_m^{(d)}, \qquad \|K_d(p,q)\|_2 \le c_d b_d e^{-\Lambda_d(p,q)},$$
$$\|\delta y_d(q)\|_2 \le c_d b_d \sum_{\substack{p\in\mathcal{S}\\ \rho_d(p)<\rho_d(q)}} e^{-\Lambda_d(p,q)}\|e_p\|_2, \qquad q \notin \mathcal{S}. \tag{8}$$

For $m > n$, the kernel is zero. For SAR-IR, the bound links corrupted support to scan-dependent access and contraction, motivating mask-conditioned routing complemented by reliability-ranked anchor context. Appendix A.2 provides the proof and feature-space interpretation.

4.3 Reliability-Guided Context Fusion

In SAR-IR, context used to reconstruct corrupted complex TFD coefficients may itself carry interference. Aggregating such context can bias coefficient estimates, leaving residual interference that

is redistributed during coherent focusing. Context selection therefore matters alongside propagation geometry.

With one directional expert per block, sparse scanning organizes interactions along a single serialization. Within its recurrent state path, a location receives information from preceding tokens through intervening state transitions (Proposition 1). Useful context may lie later in the sequence or be attenuated along the path, while earlier corrupted features can also contribute. Reliability-guided anchor retrieval complements this path by making selected context directly accessible across the feature grid, independently of scan order and without executing additional scan experts.

For each sample and block, define the reliability proxy $\gamma_i = (1 - p_i)/(1 + |\mu_i - \bar{\mu}_i|)$, where $p_i = \sigma(\mathrm{Mean}_C \widehat{P}_{m,i})$, $\mu_i = \mathrm{Mean}_C |z_{s,i}|$, and $\bar{\mu}$ is its local 3×3 average. The score favors lower mask-derived responses and smaller local feature-magnitude deviations. Queries attend to the shared top-M_ℓ anchor set $\mathcal{I}$ through learned Q/K/V projections:

$$\begin{aligned} G_{\mathrm{anc}} = \beta_{\mathrm{anc}} \,\mathrm{softmax}(QK_{\mathcal{I}}^{\top}/\sqrt{\bar{C}})V_{\mathcal{I}}, \qquad \eta = \sigma(W_\eta[z_s, G_{\mathrm{route}}, G_{\mathrm{anc}}]), \\ \Delta_{\mathrm{SDRM}} = \sigma(z_g) \odot [\eta \odot Y + (1 - \eta) \odot (z_s + G_{\mathrm{anc}})]. \end{aligned} \tag{9}$$

Gate η uses current features, routing conditions, and retrieved context to balance directional output Y and anchor-enriched features $z_s + G_{\mathrm{anc}}$ at each position and channel. Thus, sparse routing selects the recurrent propagation order, while anchor retrieval selects context sources for direct, query-dependent interaction. Their fusion supplements the single-scan representation while retaining the local features z_s. Attention costs $O(HWM_\ell\bar{C})$, excluding projections. Appendix A.4 provides implementation details.

4.4 Training Objective and Optimization

During distillation, MorphRoute-Mamba serves as the student, learning from RFI-free complex TFD targets and a frozen Transformer teacher under the same input conditions. The objective is

$$\mathcal{L}^{(s)} = \mathcal{L}_{\mathrm{rest}} + \lambda_{\mathrm{KD}}^{(s)}(\lambda_{\mathrm{out}}\mathcal{L}_{\mathrm{out}} + \lambda_{\mathrm{feat}}\mathcal{L}_{\mathrm{feat}}) + \lambda_{\mathrm{bal}}\mathcal{L}_{\mathrm{bal}} + \lambda_z\mathcal{L}_z. \tag{10}$$

Reconstruction and distillation. For a batch of B samples, let $\widehat{X}_{\mathrm{S},b}$, $\widehat{X}_{\mathrm{T},b}$, and $X_b^\star$ denote student predictions, teacher predictions, and targets. With Ω indexing the $2F_{\mathrm{TF}}T_{\mathrm{TF}}$ real/imaginary TFD entries per sample, we use

$$\mathcal{L}_{\mathrm{rest}} = \frac{1}{B|\Omega|}\sum_{b=1}^{B}\sum_{\nu\in\Omega}\sqrt{(\widehat{X}_{\mathrm{S},b,\nu} - X^\star_{b,\nu})^2 + \epsilon^2}, \qquad \epsilon = 10^{-3}. \tag{11}$$

This supervises both Cartesian components before coherent focusing. Output and feature transfer use elementwise ℓ_1 and squared Frobenius distances:

$$\begin{aligned} \mathcal{L}_{\mathrm{out}} &= \frac{1}{B|\Omega|}\sum_b \|\widehat{X}_{\mathrm{S},b} - \mathrm{sg}(\widehat{X}_{\mathrm{T},b})\|_1, \\ \mathcal{L}_{\mathrm{feat}} &= \sum_{\tau\in\mathcal{Q}}\frac{\omega_\tau}{BN_\tau}\sum_b \|\Pi_{\tau,\phi}(F^{(\tau)}_{\mathrm{S},b}) - \mathrm{sg}(F^{(\tau)}_{\mathrm{T},b})\|_{\mathrm{F}}^2. \end{aligned} \tag{12}$$

Here sg denotes stop-gradient; $\Pi_{\tau,\phi}$ aligns student feature $F_{\mathrm{S}}^{(\tau)}$ to teacher feature $F_{\mathrm{T}}^{(\tau)}$, and N_τ counts aligned elements per sample. The weighted keys $\mathcal{Q}$ cover multiscale encoder and decoder features, transferring intermediate representations alongside the output target.

Sparse-router optimization. For block j, batch-mean probabilities and Top-1 selection frequencies are

$$I_{j,d} = \frac{1}{B}\sum_b \pi^{(d)}_{j,b}, \qquad L_{j,d} = \frac{1}{B}\sum_b \mathbf{1}[d = d^\star_{j,b}]. \tag{13}$$

With $N_{\mathrm{blk}} = 14$ blocks and $D = 7$ experts, load balancing and router z-loss are

$$\mathcal{L}_{\mathrm{bal}} = \frac{D}{N_{\mathrm{blk}}}\sum_{j,d} I_{j,d}\,\mathrm{sg}(L_{j,d}), \qquad \mathcal{L}_z = \frac{1}{N_{\mathrm{blk}}B}\sum_{j,b}\left[\log\sum_d e^{q^{(d)}_{j,b}}\right]^2. \tag{14}$$

These terms regulate sample-wise expert utilization and the log partition, complementing the confidence-mediated task gradient in Eq. (7). A common logit shift leaves softmax unchanged but changes z-loss, constraining an otherwise undetermined offset; Appendix A.3 gives the auxiliary gradients. Appendix D.1 specifies feature alignment, loss weights, and the training schedule.

4.5 Static Route Compilation

Validation-guided static route compilation (VSRC) fixes block-wise expert indices and unit weights to bypass routing. SSM state updates, FMCG, anchor retrieval, and fusion remain input-dependent, preserving adaptive SAR-IR within a fixed scan schedule. Corrupted validation inputs determine:

$$s_j = \arg\max_d \sum_{b\in\mathcal{V}} \mathbf{1}[d = d^\star_{j,b}] = \arg\min_d \sum_{b\in\mathcal{V}} \mathbf{1}[d \neq d^\star_{j,b}], \qquad Y_j^{\text{stat}} = E_{s_j}(z_s). \tag{15}$$

Proposition 2 (conditional compilation sensitivity). At a shared block input and conditions, let $a = \pi_{d^\star}$, $g_j = \sigma(z_g) \odot \eta$, and $\varepsilon_j = \Pr(d^\star \neq s_j)$. If $\|E_{s_j}(z_s)\|_{\mathrm{F}} \leq B_j$, $\max_d \|E_d(z_s) - E_{s_j}(z_s)\|_{\mathrm{F}} \leq D_j$, and the Conv-FFN is L_j-Lipschitz on the comparison domain, then

$$\begin{aligned} \delta\Delta_j &= g_j \odot \big[a(E_{d^\star}(z_s) - E_{s_j}(z_s)) - (1-a)E_{s_j}(z_s)\big], \\ \mathbb{E}\|\delta Z_{\text{out},j}\|_{\mathrm{F}} &\leq \kappa_j D_j \varepsilon_j + \kappa_j B_j \mathbb{E}(1-a), \quad \kappa_j = \alpha\|W_{o,j}\|_2(1-\alpha+\alpha L_j), \quad \alpha = 0.1. \end{aligned} \tag{16}$$

Here δ denotes dynamic minus static execution. The shared anchor branch cancels; $g_j \in [0,1]$ gates route perturbations. Small expert disagreement, high routing confidence, and controlled κ_j favor low error. Appendix A.5 gives the tighter gated bound; Table 4 evaluates the network-level trade-off.

5 Experiments

5.1 Protocol

Training and validation use 11,000 simulated echo pairs from 73 SAR scenes. The simulated test set comprises 15,784 echoes from 11 scenes across 11 interference categories; four further scenes provide measured interference (MPFI, SFI2, SFI2+SFTI, UWI). The three scene groups are disjoint. Appendix B.1 details dataset construction and splits. Simulated metrics are complex TFD NMSE (cNMSE), time-domain NMSE, and focused-image RMSE/SSIM/PSNR; measured metrics are average gradient (AG) and non-scattering-to-reference ROI energy ratio (MNR), both computed directly on focused amplitude images. Same-scene revisits provide qualitative measured references. Validation selects checkpoints and VSRC routes; latency includes Fast-SCNN mask inference and the restoration-network forward pass (Appendix D).

5.2 Comparison with Existing Methods

Table 1: Simulated restoration across three domains. Pink and blue cells indicate the best and second-best results.

Method	TD NMSE (dB) ↓	TFD cNMSE (dB) ↓	RMSE ↓	SSIM ↑	PSNR (dB) ↑
ESP (Zhou & Tao, 2015)	1.64	1.64	2.9403	0.8237	41.39
RPCA (Su et al., 2017)	8.07	8.07	10.3880	0.3863	29.98
PISNet (Shen et al., 2022)	-3.55	-2.67	1.2822	0.9711	47.85
WBIM-GAN (Xu et al., 2024)	-6.19	-6.33	1.0877	0.9678	49.43
EAMamba (Lin et al., 2025)	-6.22	-6.10	1.0397	0.9811	49.92
MambaIRv1 (Guo et al., 2024)	-7.50	-7.56	0.9865	0.9839	50.77
MambaIRv2 (Guo et al., 2025)	-8.14	-8.14	0.8652	0.9891	51.90
MorphRoute-Mamba (static)	-10.79	-11.04	0.6019	0.9948	54.74

Consistency across reconstruction domains. The compiled model ranks first on all five metrics in Table 1. Relative to the strongest overall baseline, MambaIRv2, it reduces TFD cNMSE and TD NMSE by 2.90 and 2.65 dB, respectively, and raises focused-image PSNR by 2.84 dB. The matched inverse STFT links TFD and echo errors, so their joint reduction indicates that complex-coefficient corrections persist through echo reconstruction. After common focusing, lower RMSE

(0.6019 versus 0.8652) agrees with the PSNR gain, while higher SSIM (0.9948 versus 0.9891) adds evidence of structural agreement. The advantage holds across all 11 categories, including six compound types (Appendix Tables 13–14), supporting consistency across interference structures as well as reconstruction domains.

Table 2: Measured-scene AG and MNR, computed directly on focused amplitude images and averaged equally across scenes. Metric definitions appear in Appendix D.5.

Method	AG ↑					MNR (dB) ↓				
	MPFI	SFI2	SFI2+SFTI	UWI	Mean	MPFI	SFI2	SFI2+SFTI	UWI	Mean
ESP (Zhou & Tao, 2015)	18.66	16.23	12.37	18.65	16.48	5.45	1.54	7.31	-15.04	-0.18
RPCA (Su et al., 2017)	22.65	28.71	6.66	18.04	19.01	3.25	-2.97	16.10	-4.70	2.92
PISNet (Shen et al., 2022)	33.76	30.83	13.31	16.06	23.49	-2.50	-2.88	3.44	-15.43	-4.34
WBIM-GAN (Xu et al., 2024)	31.45	34.62	15.87	20.53	25.62	-7.40	-0.84	3.40	-8.13	-3.24
EAMamba (Lin et al., 2025)	35.64	33.08	15.75	18.44	25.73	-5.13	-3.04	-0.05	-13.22	-5.36
MambaIRv1 (Guo et al., 2024)	35.62	33.79	15.46	17.81	25.67	-4.25	-3.07	1.59	-15.02	-5.19
MambaIRv2 (Guo et al., 2025)	36.88	33.88	14.68	17.12	25.64	-3.70	-3.08	2.22	-15.33	-4.97
MorphRoute-Mamba (static)	37.50	34.96	21.55	21.31	28.83	-11.61	-4.03	-4.80	-16.83	-9.32

Transfer to unseen measured scenes. Trained and VSRC-calibrated on simulated interference, the static checkpoint achieves the highest AG and lowest MNR in four unseen measured scenes without restoration-model fine-tuning (Table 2). Mean AG reaches 28.83 versus 25.73 and mean MNR -9.32 versus -5.36 dB for the strongest baseline. Lower MNR indicates less non-scattering energy relative to reference ROIs; its agreement with higher AG supports suppression without uniform smoothing. SFI2+SFTI and UWI, absent from simulated training categories, retain the best ranks, supporting transfer across scenes and interference types, with a larger margin for SFI2+SFTI than UWI. Since routes are fixed, these results assess input-conditioned restoration under a common scan schedule.

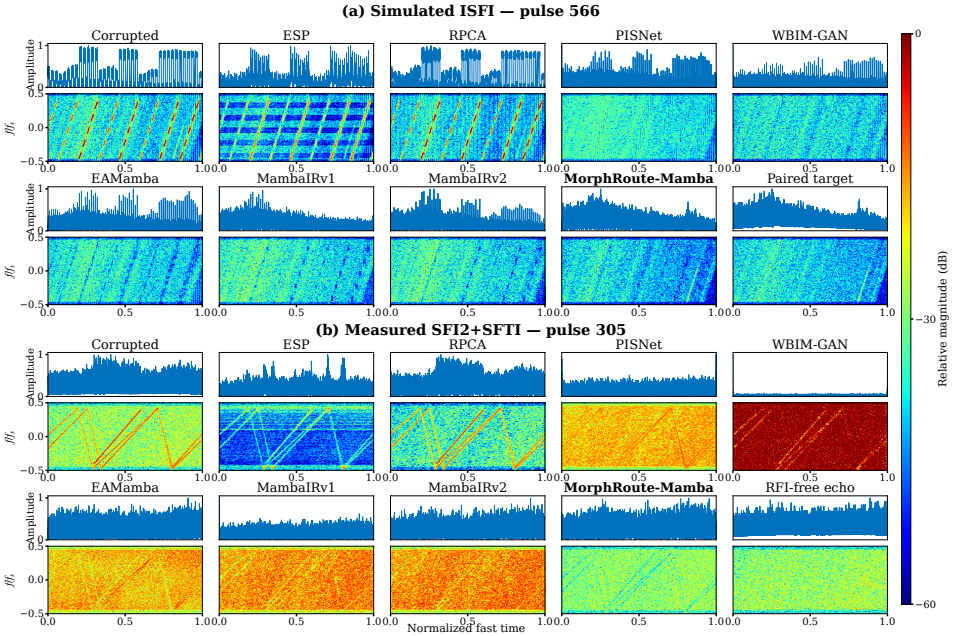
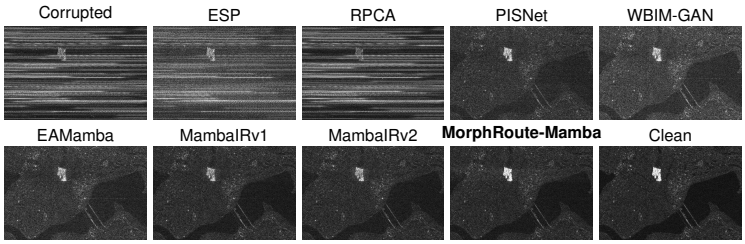
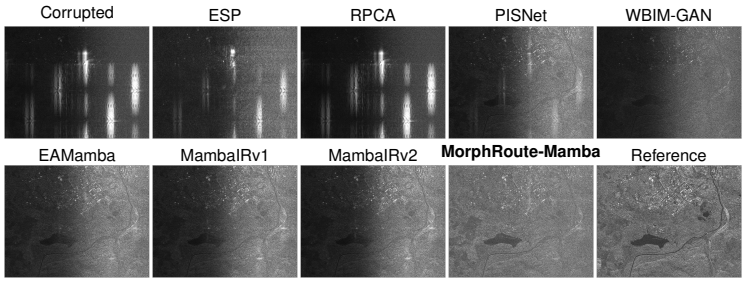


Figure 5: Waveform and spectrogram comparisons for (a) simulated ISFI with a paired target and (b) measured SFI2+SFTI with an RFI-free same-scene revisit as a qualitative reference.

Echo-to-image visual consistency. Figures 5–6 reveal two ISFI failure modes: ESP/RPCA retain oblique TFD ridges that focus into streaks; PISNet/WBIM-GAN suppress these ridges but weaken target-bearing TFD content and focused structures. MorphRoute-Mamba reduces streaks while preserving the paired target's waveform envelope and scatterers, consistent with lower TD/TFD errors and better image metrics in Table 1. In measured SFI2+SFTI, residual TFD traces yield bright focused bands, while stronger suppression weakens scene boundaries. MorphRoute-Mamba reduces both effects, consistent with higher AG and lower MNR for this scene in Table 2.



(a) Simulated ISFI



(b) Measured SFI2+SFTI

Figure 6: Focused-image comparisons: (a) simulated ISFI with a paired target; (b) measured SFI2+SFTI with a qualitative same-scene revisit. Display mapping is fixed within rows.

5.3 Cross-Method Feature Analysis

Figure 7 uses t-SNE (van der Maaten & Hinton, 2008) to visualize deep features from 100 shared, evenly spaced test pulses per category/scene. In simulated data, baselines show substantial category mixing, whereas MorphRoute-Mamba forms more label-consistent neighborhoods for single and compound interference. In measured data, its four scene groups are clearer than the overlapping baseline patterns. This organization suggests that its features distinguish heterogeneous input conditions, consistent with the reconstruction gains in Table 1 and transfer to unseen scenes in Table 2.

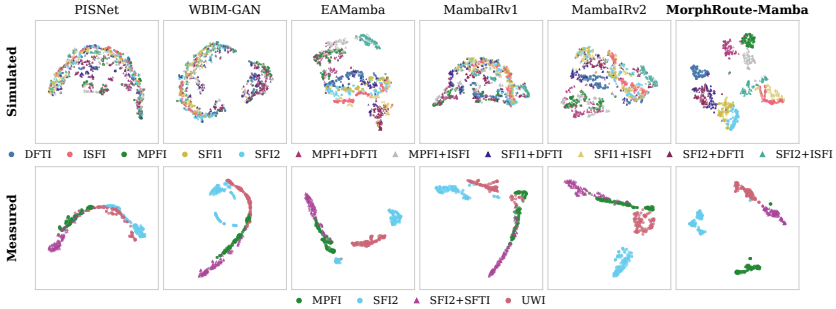


Figure 7: Deep-feature t-SNE: (a) simulated categories; (b) measured scenes.

5.4 Ablation Studies

Table 3: Condition-input ablation on simulated focused images.

Input	RMSE ↓	SSIM ↑	PSNR (dB) ↑
Full	0.6019	0.9948	54.74
w/o FD	0.6385	0.9879	53.82
w/o mask	0.6497	0.9837	53.53
TFD only	0.6579	0.9812	52.49

Table 4: Inference quality and cost. MorphRoute parameter counts and latency include Fast-SCNN.

Method	Params (M) ↓	FLOPs (G) ↓	Latency (ms) ↓	SSIM ↑	PSNR (dB) ↑
EAMamba	25.32	136.61	75.17	0.9811	49.92
MambaIRv1	15.77	1289.69	371.91	0.9839	50.77
MambaIRv2	22.66	1754.53	777.87	0.9891	51.90
MorphRoute dynamic	21.11	279.03	71.82	0.9950	54.95
MorphRoute static	19.49	277.55	64.79	0.9948	54.74

Condition sources. Table 3 shows PSNR rising from 52.49 dB with TFD alone to 53.53/53.82 dB with FD/mask and 54.74 dB with both; RMSE and SSIM agree. FD supplies fine full-record spectral

context, whereas the mask localizes candidate interference support for gating, routing, and anchors. Their complementary gains show that both global and local cues improve focused-image restoration.

Static-route deployment. Table 4 shows the static pipeline, including the frozen Fast-SCNN, at 19.49M parameters, 54.74 dB PSNR, and 64.79 ms, exceeding the MambaIRv2 checkpoint in this table by 2.84 dB while running about 12 times faster, and EAMamba, the fastest baseline, by 10.38 ms. Compared with the dynamic model, the tabulated static model has 1.62M fewer parameters; VSRC cuts latency by 9.79% at a cost of 0.21 dB PSNR and 0.0002 SSIM. Only expert selection and weighting are fixed; conditioning, anchors, and fusion stay adaptive. Among the checkpoints in Table 4, static routing has the lowest latency and a quality lead over the baselines.

Table 5: FMCG mechanism ablation (dB).

Variant	NMSE ↓	cNMSE ↓	PSNR ↑
Full model	-10.79	-11.04	54.74
w/o spectral compat.	-10.36	-10.61	54.27
w/o spatial compat.	-10.08	-10.39	53.98
Fixed learned gains	-9.63	-9.84	53.31
Naive concatenation	-9.89	-10.15	53.66

Table 6: SDRM component ablation (dB).

Variant	NMSE ↓	cNMSE ↓	PSNR ↑
Full model (7)	-10.79	-11.04	54.74
Axial-only bank (4)	-9.63	-9.89	53.56
w/o mask routing	-10.03	-10.12	53.79
w/o anchors	-10.22	-10.44	54.04

FMCG compatibility. With FD and mask retained, Table 5 isolates feature-dependent fusion (Appendix D.2). Replacing its spectral or spatial gate with a stagewise learned scalar costs 0.47 or 0.76 dB PSNR; replacing both costs 1.43 dB, and direct concatenation costs 1.08 dB. Time-domain NMSE and TFD cNMSE worsen for every variant, showing that adaptive gating improves complex echo recovery before focusing. The larger spatial penalty is consistent with R_m's roles in both mask injection and SDRM prior formation. FMCG thus matches global spectral and local interference cues to current features, beyond merely making these conditions available.

SDRM components. All three ablations worsen NMSE/cNMSE/PSNR (Table 6). The axial-only bank has the largest deficits (1.16/1.15/1.18 dB), although geometry and expert count change together. Removing mask cues only from routing yields 0.76/0.92/0.95 dB deficits despite retaining other mask pathways, supporting mask-informed expert selection. Removing anchors has the smallest deficits (0.57/0.60/0.70 dB), suggesting that reliability-guided retrieval complements directional propagation in echo and focused-image restoration.

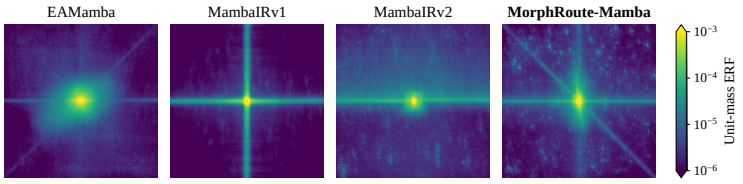


Figure 8: Mean conditional residual ERFs across restoration models.

Effective receptive fields. Shared-scale ERFs in Figure 8 show axial and oblique sensitivity for MorphRoute-Mamba, versus an axial cross for MambaIRv1, a horizontal band for MambaIRv2, and diffuse near-query sensitivity for EAMamba. Such orientation coverage matters when axial bands and sloped TFD ridges obscure different neighborhoods. These full-network ERFs are consistent with the scan-dependent paths in Proposition 1 and also reflect anchor-based context fusion.

6 Conclusion

MorphRoute-Mamba combines morphology-guided single-expert routing and reliability-ranked anchors for pre-focusing SAR-IR. Simulated results improve TFD, echo, and focused-image metrics; measured-scene AG/MNR gains suggest transfer without restoration-model fine-tuning. Static compilation reduces latency with a small quality loss while retaining adaptive conditioning and fusion. Future work will study joint interference-support estimation and cross-pulse context for scene-consistent restoration.

AI USE STATEMENT

The authors used a generative AI assistant to check grammar and polish selected passages of the manuscript. All AI-assisted edits were rigorously reviewed and verified by the authors, who take full responsibility for the final manuscript.

ETHICS STATEMENT

This work processes synthetic and measured SAR signals and does not involve human subjects or personally identifiable information. Interference mitigation can benefit scientific and environmental remote sensing, while restored sensing products may also have broader dual-use applications. Users of the method and data should follow applicable laws, data licenses, and responsible-use requirements.

REPRODUCIBILITY STATEMENT

Sections 3–5 and Appendices A–D document the restoration pipeline, routing, data and mask construction, architecture, training, and evaluation protocol.

REFERENCES

Liangyu Chen, Xiaojie Chu, Xiangyu Zhang, and Jian Sun. Simple baselines for image restoration. In *Computer Vision – ECCV 2022*, volume 13667 of *Lecture Notes in Computer Science*, pp. 17–33. Springer, 2022. doi: 10.1007/978-3-031-20071-7_2.

Ian G. Cumming and Frank H. Wong. *Digital Processing of Synthetic Aperture Radar Data: Algorithms and Implementation*. Artech House, 2005. ISBN 9781580530583. URL https://us.artechhouse.com/Digital-Processing-of-Synthetic-Aperture-Radar-Data-P1549.aspx.

Tri Dao and Albert Gu. Transformers are SSMs: Generalized models and efficient algorithms through structured state space duality. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 10041–10071. PMLR, 2024. URL https://proceedings.mlr.press/v235/dao24a.html.

Weiwei Fan, Feng Zhou, Mingliang Tao, Xueru Bai, Pengshuai Rong, Shuang Yang, and Tian Tian. Interference mitigation for synthetic aperture radar based on deep residual network. *Remote Sensing*, 11(14):1654, 2019. doi: 10.3390/rs11141654.

Fuping Fang, Yuanrong Tian, Dahai Dai, and Shiqi Xing. Synthetic aperture radar radio frequency interference suppression method based on fusing segmentation and inpainting networks. *Remote Sensing*, 16(6):1013, 2024. doi: 10.3390/rs16061013.

Albert Gu and Tri Dao. Mamba: Linear-time sequence modeling with selective state spaces. In *First Conference on Language Modeling (COLM)*, 2024. URL https://openreview.net/forum?id=tEYskw1VY2.

Albert Gu, Karan Goel, and Christopher Ré. Efficiently modeling long sequences with structured state spaces. In *International Conference on Learning Representations*, 2022. URL https://openreview.net/forum?id=uYLFoz1vlAC.

Hang Guo, Jinmin Li, Tao Dai, Zhihao Ouyang, Xudong Ren, and Shu-Tao Xia. Mambair: A simple baseline for image restoration with state-space model. In *European conference on computer vision*, pp. 222–241. Springer, 2024.

Hang Guo, Yong Guo, Yaohua Zha, Yulun Zhang, Wenbo Li, Tao Dai, Shu-Tao Xia, and Yawei Li. MambaIRv2: Attentive state space restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 28124–28133, 2025. URL https://openaccess.thecvf.com/content/CVPR2025/html/Guo_MambaIRv2_Attentive_State_Space_Restoration_CVPR_2025_paper.html.

Han Hu, Zhuoran Zheng, Liang Li, and Chen Lyu. VAMamba: An efficient visual adaptive mamba for image restoration, 2025.

Yan Huang, Lei Zhang, Jie Li, Wei Hong, and Arye Nehorai. A novel tensor technique for simultaneous narrowband and wideband interference suppression on single-channel SAR system. *IEEE Transactions on Geoscience and Remote Sensing*, 57(12):9575–9588, 2019. doi: 10.1109/TGRS.2019.2927764.

Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C. Berg, Wan-Yen Lo, Piotr Dollár, and Ross Girshick. Segment anything. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 4015–4026, 2023. URL https://openaccess.thecvf.com/content/ICCV2023/html/Kirillov_Segment_Anything_ICCV_2023_paper.html.

Lingshun Kong, Jiangxin Dong, Jianjun Ge, Mingqiang Li, and Jinshan Pan. Efficient frequency domain-based transformers for high-quality image deblurring. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5886–5895, 2023. URL https://openaccess.thecvf.com/content/CVPR2023/html/Kong_Efficient_Frequency_Domain-Based_Transformers_for_High-Quality_Image_Deblurring_CVPR_2023_paper.html.

Boyun Li, Xiao Liu, Peng Hu, Zhongqin Wu, Jiancheng Lv, and Xi Peng. All-in-one image restoration for unknown corruption. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 17452–17462, 2022. URL https://openaccess.thecvf.com/content/CVPR2022/html/Li_All-in-One_Image_Restoration_for_Unknown_Corruption_CVPR_2022_paper.html.

Boyun Li, Haiyu Zhao, Wenxin Wang, Peng Hu, Yuanbiao Gou, and Xi Peng. MaIR: A locality- and continuity-preserving mamba for image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 7491–7501, 2025a. URL https://openaccess.thecvf.com/content/CVPR2025/html/Li_MaIR_A_Locality-_and_Continuity-Preserving_Mamba_for_Image_Restoration_CVPR_2025_paper.html.

Tanzhe Li, Caoshuo Li, Jiayi Lyu, Hongjuan Pei, Baochang Zhang, Taisong Jin, and Rongrong Ji. DAMamba: Vision state space model with dynamic adaptive scan. In *Advances in Neural Information Processing Systems*, volume 38, 2025b. doi: 10.52202/085713-1326. URL https://proceedings.neurips.cc/paper_files/paper/2025/hash/38b23e2328096520e9c889ae03e372c9-Abstract-Conference.html.

Jingyun Liang, Jiezhang Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte. SwinIR: Image restoration using swin transformer. In *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, pp. 1833–1844, 2021. doi: 10.1109/ICCVW54120.2021.00210.

Jingbo Lin, Zhilu Zhang, Wenbo Li, Renjing Pei, Hang Xu, Hongzhi Zhang, and Wangmeng Zuo. UniRestorer: Universal image restoration via adaptively estimating image degradation at proper granularity. In *The Fourteenth International Conference on Learning Representations*, 2026. URL https://proceedings.iclr.cc/paper_files/paper/2026/hash/04fb0ea2010694af12735724a601a3b4-Abstract-Conference.html.

Yu-Cheng Lin, Yu-Syuan Xu, Hao-Wei Chen, Hsien-Kai Kuo, and Chun-Yi Lee. EAMamba: Efficient all-around vision state space model for image restoration. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 11708–11719, 2025. URL https://openaccess.thecvf.com/content/ICCV2025/html/Lin_EAMamba_Efficient_All-Around_Vision_State_Space_Model_for_Image_Restoration_ICCV_2025_paper.html.

Hongqing Liu, Dong Li, Yi Zhou, and Trieu-Kien Truong. Simultaneous radio frequency and wideband interference suppression in SAR signals via sparsity exploitation in time-frequency domain. *IEEE Transactions on Geoscience and Remote Sensing*, 56(10):5780–5793, 2018. doi: 10.1109/TGRS.2018.2825459.

Yue Liu, Yunjie Tian, Yuzhong Zhao, Hongtian Yu, Lingxi Xie, Yaowei Wang, Qixiang Ye, Jianbin Jiao, and Yunfan Liu. Vmamba: Visual state space model. In *Advances in neural information processing systems*, volume 37, pp. 103031–103063, 2024.

Franz J. Meyer, Jeremy B. Nicoll, and Anthony P. Doulgeris. Correction and characterization of radio frequency interference signatures in L-band synthetic aperture radar data. *IEEE Transactions on Geoscience and Remote Sensing*, 51(10):4961–4972, 2013. doi: 10.1109/TGRS.2013.2252469.

Vaishnav Potlapalli, Syed Waqas Zamir, Salman H. Khan, and Fahad Shahbaz Khan. PromptIR: Prompting for all-in-one image restoration. In *Advances in Neural Information Processing Systems*, volume 36, pp. 71275–71293, 2023. doi: 10.52202/075280-3121. URL https://proceedings.neurips.cc/paper_files/paper/2023/hash/e187897ed7780a579a0d76fd4a35d107-Abstract-Conference.html.

Rudra P. K. Poudel, Stephan Liwicki, and Roberto Cipolla. Fast-SCNN: Fast semantic segmentation network. In *Proceedings of the British Machine Vision Conference (BMVC)*, 2019. URL https://arxiv.org/abs/1902.04502.

Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. Learning transferable visual models from natural language supervision. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 8748–8763. PMLR, 2021. URL https://proceedings.mlr.press/v139/radford21a.html.

Bin Ren, Yawei Li, Xu Zheng, Yuqian Fu, Danda Pani Paudel, Hong Liu, Ming-Hsuan Yang, Luc Van Gool, and Nicu Sebe. Efficient degradation-agnostic image restoration via channel-wise functional decomposition and manifold regularization. In *International Conference on Learning Representations*, volume 2026, pp. 126404–126430, 2026.

Jiayuan Shen, Bing Han, Zongxu Pan, Guangzuo Li, Yuxin Hu, and Chibiao Ding. Learning time-frequency information with prior for SAR radio frequency interference suppression. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–16, 2022. doi: 10.1109/TGRS.2022.3225499. Art. no. 5239716.

Jimmy T. H. Smith, Andrew Warrington, and Scott W. Linderman. Simplified state space layers for sequence modeling. In *International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=Ai8Hw3AXqks.

Jia Su, Haihong Tao, Mingliang Tao, Ling Wang, and Jian Xie. Narrow-band interference suppression via RPCA-based signal separation in time–frequency domain. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 10(11):5016–5025, 2017. doi: 10.1109/JSTARS.2017.2727520.

Jia Su, Haojiang Li, Mingliang Tao, Yifei Fan, Ling Wang, and Haihong Tao. Wideband interference suppression for SAR by time–frequency–pulse joint domain processing. In *IGARSS 2020–2020 IEEE International Geoscience and Remote Sensing Symposium*, pp. 3774–3777, 2020. doi: 10.1109/IGARSS39084.2020.9323259.

Jia Su, Mengru Xi, Yanyun Gong, Mingliang Tao, Yifei Fan, and Ling Wang. Time-varying wideband interference mitigation for SAR via time–frequency–pulse joint decomposition algorithm. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–16, 2022. doi: 10.1109/TGRS.2022.3213318. Art. no. 5118816.

Mingliang Tao, Feng Zhou, and Zijing Zhang. Wideband interference mitigation in high-resolution airborne synthetic aperture radar data. *IEEE Transactions on Geoscience and Remote Sensing*, 54 (1):74–87, 2016. doi: 10.1109/TGRS.2015.2450754.

Mingliang Tao, Jia Su, Yan Huang, and Ling Wang. Mitigation of radio frequency interference in synthetic aperture radar data: Current status and future trends. *Remote Sensing*, 11(20):2438, 2019. doi: 10.3390/rs11202438.

Laurens van der Maaten and Geoffrey Hinton. Visualizing data using t-SNE. *Journal of Machine Learning Research*, 9:2579–2605, 2008. URL https://jmlr.org/papers/v9/vandermaaten08a.html.

Yongzhen Wang, Yongjun Li, Zhuoran Zheng, Xiao-Ping Zhang, and Mingqiang Wei. M2Restore: Mixture-of-experts-based mamba-cnn fusion framework for all-in-one image restoration. *IEEE Transactions on Image Processing*, 34:8086–8100, 2025. doi: 10.1109/TIP.2025.3638662.

Zhendong Wang, Xiaodong Cun, Jianmin Bao, Wengang Zhou, Jianzhuang Liu, and Houqiang Li. Uformer: A general u-shaped transformer for image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 17662–17672, 2022. doi: 10.1109/CVPR52688.2022.01716.

Xiaoyu Xu, Weiwei Fan, Siyao Wang, and Feng Zhou. WBIM-GAN: A generative adversarial network based wideband interference mitigation model for synthetic aperture radar. *Remote Sensing*, 16(5):910, 2024. doi: 10.3390/rs16050910.

Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, and Ming-Hsuan Yang. Restormer: Efficient transformer for high-resolution image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5728–5739, 2022. URL https://openaccess.thecvf.com/content/CVPR2022/html/Zamir_Restormer_Efficient_Transformer_for_High-Resolution_Image_Restoration_CVPR_2022_paper.html.

Shuangxi Zhang, Mengdao Xing, Rui Guo, Lei Zhang, and Zheng Bao. Interference suppression algorithm for SAR based on time–frequency transform. *IEEE Transactions on Geoscience and Remote Sensing*, 49(10):3765–3779, 2011. doi: 10.1109/TGRS.2011.2164409.

Zhaodong Zhang, Yi Ding, Weiwei Fan, Feng Zhou, and Bingbing Lu. Wide-band interference mitigation for SAR based on generative adversarial network. In *2021 CIE International Conference on Radar (Radar 2021)*, pp. 1897–1900, 2021. URL https://ieeexplore.ieee.org/document/10028142.

Ziyao Zhang, Qiankun Ma, Tong Zhang, Jie Chen, Hairong Zheng, and Wen Gao. Switch-UMamba: Dynamic scanning vision mamba UNet for medical image segmentation. *Medical Image Analysis*, 107:103792, 2026. doi: 10.1016/j.media.2025.103792.

Feng Zhou and Mingliang Tao. Research on methods for narrow-band interference suppression in synthetic aperture radar data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 8(7):3476–3485, 2015. doi: 10.1109/JSTARS.2015.2431916.

Kun Zhou, Xinyu Lin, and Jiangbo Lu. TSP-Mamba: The travelling salesman problem meets mamba for image super-resolution and beyond. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 28134–28143, 2025. doi: 10.1109/CVPR52734.2025.02620.

Lianghui Zhu, Bencheng Liao, Qian Zhang, Xinlong Wang, Wenyu Liu, and Xinggang Wang. Vision mamba: Efficient visual representation learning with bidirectional state space model. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 62429–62442. PMLR, 2024. URL https://proceedings.mlr.press/v235/zhu24f.html.

The appendix first develops the routing analysis (Section A), then specifies datasets and representations (Section B), network architectures (Section C), and training and evaluation protocols (Section D). Additional results and their analysis are collected in Section E.

A Sparse Directional Routing and Theoretical Analysis

A.1 Exact Scan Definitions

For each direction $d \in \mathcal{D}_7$, let $\rho_d : \mathcal{G} \to \{1, \dots, HW\}$ map a grid coordinate to its sequence index, and let P_d permute feature tokens accordingly. In Figure 4, frequency f increases upward and time t increases rightward. With the first lexicographic key denoting the outer traversal, the axial orders F^+, F^-, T^+, T^- are, respectively, $\mathrm{lex}(t\uparrow, f\uparrow)$, $\mathrm{lex}(t\downarrow, f\downarrow)$, $\mathrm{lex}(f\uparrow, t\uparrow)$, and $\mathrm{lex}(f\downarrow, t\downarrow)$. Thus, F^+ scans each column upward from left to right, while T^+ scans each row rightward from bottom to top; each minus order is the complete reverse of its plus order. The ridge order sorts coordinates lexicographically by $(t - f, f, t)$, and its paired order is the complete reversed list. The stripe-serpentine order partitions time into non-circular stripes of width $\max(2, \lfloor W_\ell/8 \rfloor)$, alternates frequency direction between stripes and local time direction between rows, and alternates zero and half-stripe shifts between consecutive blocks. Figure 4 illustrates these scan orders; its stripe-serpentine panel shows the unshifted phase.

A.2 Proof and Scope of Proposition 1

Exact conditional kernel. For one direction, let x_n denote the serialized feature entering the selective recurrence, and write

$$h_n = \overline{A}_n h_{n-1} + \overline{B}_n x_n, \qquad y_n = C_n h_n + D_n x_n, \qquad h_0 = 0. \tag{17}$$

Let $\Theta(x) = \{\overline{A}_k, \overline{B}_k, C_k, D_k\}_k$ collect the input-conditioned coefficients. Holding Θ fixed and unrolling the recurrence gives $y_n = \sum_m \mathcal{K}_{n,m} x_m$, where

$$\mathcal{K}_{n,m} = \left.\frac{\partial y_n}{\partial x_m}\right|_{\Theta \text{ fixed}} = \begin{cases} C_n \overline{A}_n \overline{A}_{n-1} \cdots \overline{A}_{m+1} \overline{B}_m, & m < n, \\ C_n \overline{B}_n + D_n, & m = n, \\ 0, & m > n. \end{cases} \tag{18}$$

The ordered matrix product is essential: transitions need not commute. Returning to grid coordinates sets $m = \rho_d(p)$, $n = \rho_d(q)$ and $K_d(p, q) = \mathcal{K}^{(d)}_{n,m}$. For $m < n$, submultiplicativity yields

$$\|K_d(p,q)\|_2 \le \|C_n^{(d)}\|_2 \|\overline{B}_m^{(d)}\|_2 \prod_{k=m+1}^{n} \|\overline{A}_k^{(d)}\|_2 \le c_d b_d \exp\left(- \sum_{k=m+1}^{n} \lambda_{d,k} \right). \tag{19}$$

This proves the kernel bound in Eq. (8); Eq. (18) covers the remaining cases. A uniform rate $\lambda_{d,k} \ge \lambda_d > 0$ gives $c_d b_d \alpha_d^{n-m}$, with $\alpha_d = e^{-\lambda_d} < 1$. Uniformity across inputs additionally requires uniform coefficient bounds.

Selective transitions and accumulated contraction. The implemented selective scan uses a negative diagonal continuous-time transition and positive input-dependent steps. Flattening channel/state indices into r, its discretized entries have the form $\overline{A}^{(d)}_{k,rr} = \exp(-a_{d,r}\tau_{d,k,r})$, with $a_{d,r} > 0$ and $\tau_{d,k,r} > 0$. Consequently,

$$\begin{aligned} \left\| \overline{A}_n^{(d)} \cdots \overline{A}_{m+1}^{(d)} \right\|_2 &= \exp\left[- \min_r \sum_{k=m+1}^{n} a_{d,r} \tau_{d,k,r} \right] \\ &\le \exp\left[- \sum_{k=m+1}^{n} \min_r a_{d,r} \tau_{d,k,r} \right]. \end{aligned} \tag{20}$$

Thus $\lambda_{d,k} = \min_r a_{d,r}\tau_{d,k,r}$ is an admissible choice, and the first line gives a potentially tighter path-specific bound. Different serializations change both causal access to a location and the intervening sequence of selective transitions. A uniform contraction rate additionally requires a positive lower bound on the steps across the input domain.

Perturbation propagation from corrupted feature support. Consider a feature perturbation e_p supported on grid locations $\mathcal{S}$ at the recurrence input. Evaluate the coefficients at a reference input and hold them fixed when comparing that input with its perturbed counterpart. The resulting output difference is $\delta y_d(q) = \sum_{p\in\mathcal{S}} K_d(p,q)e_p$. For a target $q \notin \mathcal{S}$, define $\mathcal{S}_d^-(q) = \{p \in \mathcal{S} : \rho_d(p) < \rho_d(q)\}$ and $\Lambda_d(p,q) = \sum_{k=\rho_d(p)+1}^{\rho_d(q)} \lambda_{d,k}$. Linearity of the conditional map, the triangle inequality, and Cauchy-Schwarz give

$$\begin{aligned}\|\delta y_d(q)\|_2 &\le c_d b_d \sum_{p\in\mathcal{S}_d^-(q)} e^{-\Lambda_d(p,q)}\|e_p\|_2 \\ &\le c_d b_d \left(\sum_{p\in\mathcal{S}_d^-(q)} e^{-2\Lambda_d(p,q)}\right)^{1/2} \left(\sum_{p\in\mathcal{S}_d^-(q)} \|e_p\|_2^2\right)^{1/2}.\end{aligned} \tag{21}$$

The first inequality proves the support-dependent bound in Eq. (8); the second expresses it in terms of perturbation energy. If every contributing source is at least $r_0 \ge 1$ transitions away and $\|\overline{A}_k\|_2 \le \alpha_d < 1$, bijectivity allows at most one source at each distance. Summing the geometric series yields

$$\|\delta y_d(q)\|_2 \le \frac{c_d b_d \alpha_d^{r_0}}{\sqrt{1-\alpha_d^2}} \|e_{\mathcal{S}_d^-(q)}\|_2. \tag{22}$$

For fixed nonnegative routing weights g_d, the mixed perturbation is bounded by the g_d-weighted sum of Eq. (21).

Implications for SAR-IR. The upstream support $\mathcal{S}_d^-(q)$ and path contraction $\Lambda_d(p,q)$ jointly control the bound on interference-induced feature perturbations reaching q. Thus, scan selection changes both which corrupted locations have a causal path to the target and how the state contracts along that path. The predicted TFD mask describes candidate interference support at the input representation; feature mixing can expand perturbations beyond this support, so $\mathcal{S}$ is defined separately at the recurrence input. The router combines mask cues with current features and directional variation to select among these propagation geometries. Reliability-ranked anchor attention complements the selected scan by accessing context independently of scan order. This interpretation concerns conditional influence bounds: a larger bound alone does not imply stronger actual influence or better restoration.

Relation to the full selective-map Jacobian. For the actual selective map $F_d(x;\Theta(x))$, the chain rule gives

$$D_{x_m} F_{d,n}(x;\Theta(x)) = \mathcal{K}_{n,m}^{(d)} + D_\Theta F_{d,n}(x;\Theta(x))[D_{x_m}\Theta(x)]. \tag{23}$$

Proposition 1 bounds the fixed-coefficient term and its response to support-restricted feature perturbations. The empirical conditional ERFs differentiate the complete network, including coefficient dependence, projections, routing, and anchors, while holding FD and mask inputs fixed. The proposition isolates propagation through the selective recurrence; the ERFs characterize the combined sensitivity of the restoration network.

A.3 Routing Descriptor and Optimization

For current block input Z, form $Z_{\mathrm{loc}} = \mathrm{Norm}(Z) + \mathrm{Local}(\mathrm{Norm}(Z))$ and split the learned input projection into $(z_s, z_g) = W_{\mathrm{in}} Z_{\mathrm{loc}}$. Both have $\bar{C}_\ell = 2C_\ell$ channels. For each block, let $\widehat{P}_m = R_m \odot V_{\mathrm{prior}}(\widetilde{P}_m)$ and $G_{\mathrm{route}} = \beta_{\mathrm{route}}\widehat{P}_m$. At position i, the router uses mask occupancy and directional feature variation:

$$\begin{aligned} m_{b,i} &= \mathrm{Mean}_C(\widehat{P}_{m,b,i}), \\ \delta_{a,b,i} &= \mathrm{Mean}_C\,|\nabla_a z_{s,b,i}|, \qquad a \in \{f,t\}, \\ q_b &= \mathrm{Mean}_i\, W_r[\, z_{s,b,i}, G_{\mathrm{route},b,i}, m_{b,i}, \delta_{f,b,i}, \delta_{t,b,i}\,]. \end{aligned} \tag{24}$$

All experiments use $K = 1$: $\pi_b = \mathrm{softmax}(q_b)$, $d_b^\star = \arg\max_d q_b^{(d)}$, and $Y_b = \pi_b^{(d_b^\star)} E_{d_b^\star}(z_{s,b})$. Retaining the selected softmax confidence yields the task-gradient path in Eq. (7); normalizing that weight to one would remove it. Hard selection is locally constant away from ties, and unselected experts receive no direct task gradient for that sample. Importance, load, and auxiliary losses follow Eqs. (13) and (14), with $D = 7$ for the full bank and $D = 4$ for the axial-only ablation.

Auxiliary gradients and common-shift decomposition. Let $\zeta_{j,b} = \log \sum_d \exp(q_{j,b}^{(d)})$ and $\mathcal{L}_{\mathrm{MoE}} = \lambda_{\mathrm{bal}}\mathcal{L}_{\mathrm{bal}} + \lambda_z\mathcal{L}_z$. Holding the discrete choices fixed, applying the softmax Jacobian to Eq. (14) gives

$$\frac{\partial \mathcal{L}_{\mathrm{MoE}}}{\partial q_{j,b}^{(r)}} = \frac{\pi_{j,b}^{(r)}}{N_{\mathrm{blk}}B}\left[\lambda_{\mathrm{bal}}D\left(L_{j,r} - \sum_d \pi_{j,b}^{(d)} L_{j,d}\right) + 2\lambda_z\zeta_{j,b}\right]. \tag{25}$$

The balancing gradient sums to zero over experts. The z-loss contribution sums to $2\lambda_z\zeta_{j,b}/(N_{\mathrm{blk}}B)$: adding a common offset c leaves softmax unchanged but changes its penalty to $(\zeta_{j,b}+c)^2$. The two terms therefore regulate expert utilization and logit offset, respectively, alongside the confidence-mediated task gradient.

A.4 Anchor Fusion and Residual Updates

Equation (9) uses $Q = z_s W_Q$, $K_{\mathcal{I}} = z_{s,\mathcal{I}} W_K$, and $V_{\mathcal{I}} = z_{s,\mathcal{I}} W_V$, with biases implicit and features reshaped into token rows. All queries in a sample/block share the selected anchor set, with query-dependent attention weights. The score γ_i provides a relative reliability ranking from mask-derived responses and local feature consistency. Ranking inputs are detached; task gradients pass through the selected features and attention projections. All spatial positions remain queries, and the fusion retains the local feature branch z_s. Anchor budgets and scalar-gate initialization are in Table 12.

The fusion tensors $Y, G_{\mathrm{route}}, G_{\mathrm{anc}}, \eta, \Delta_{\mathrm{SDRM}}$ have shape $B \times \bar{C}_\ell \times H_\ell \times W_\ell$, where $\bar{C}_\ell = 2C_\ell$; W_o projects to C_ℓ. The learned sigmoid scalar β_{anc} scales retrieved context, and $\sigma(z_g)$ modulates the fused update. The block applies the residual updates

$$\begin{aligned} Z_{\mathrm{res}} &= Z + \alpha W_o \Delta_{\mathrm{SDRM}}, \\ Z_{\mathrm{out}} &= Z_{\mathrm{res}} + \alpha[\mathcal{F}_{\mathrm{FFN}}(Z_{\mathrm{res}}) - Z_{\mathrm{res}}], \qquad \alpha = \alpha_{\mathrm{res}} = 0.1. \end{aligned} \tag{26}$$

The SiLU Conv-FFN includes an internal residual, so subtracting its input isolates the refinement update. At a shared block input, static compilation changes Y while preserving anchor ranking, context, and fusion gates.

A.5 Static Compilation: Conditional Adaptation and Error Analysis

Fixed expert identities with adaptive computation. VSRC replaces per-sample expert selection with a block-specific index s_j; different blocks can retain different directions. The selected expert still computes its selective state-space coefficients from the current features. Likewise, FMCG recomputes spectral and mask compatibility, anchor ranking and attention depend on the current sample, and the fusion gate retains G_{route}. These mechanisms preserve responses to spectral occupancy, corrupted support, and available context in SAR-IR. Training uses dynamic routing to learn the expert bank and its selection policy; compilation approximates that policy for deployment without retraining the experts. Both modes execute one expert per block, so the computational saving comes from bypassing routing computations.

The exact empirical objective. For the $N = |\mathcal{V}| = 1{,}100$ validation inputs, let $v_{j,d} = \sum_b \mathbf{1}[d = d_{j,b}^\star]$. The empirical disagreement of a fixed expert s is

$$\widehat{\varepsilon}_j(s) = \frac{1}{N}\sum_b \mathbf{1}[s \neq d_{j,b}^\star] = 1 - \frac{v_{j,s}}{N}. \tag{27}$$

Minimizing this expression proves Eq. (15). Count ties use accumulated probability, then canonical expert order. The calibration record stores counts, probabilities, selected directions, block names, direction schema, and checkpoint/configuration identifiers. This objective uses corrupted validation inputs and their conditions; clean targets separately select checkpoints. It is an empirical surrogate for preserving the learned routing policy. Its relevance to restoration depends on expert-output differences and fusion sensitivity, as quantified below; the complete compiled network is evaluated separately.

Feature deviation at a shared input. Let $d = d_j^\star(z_s)$, $a = \pi_d(z_s)$, and $s = s_j$. Adding and subtracting $aE_s(z_s)$ gives the exact identity

$$Y^{\mathrm{dyn}} - Y^{\mathrm{stat}} = a[E_d(z_s) - E_s(z_s)] - (1-a)E_s(z_s). \tag{28}$$

With the uniform-on-support constants D_j, B_j specified in the main paper, its pointwise consequence is

$$\|Y^{\mathrm{dyn}} - Y^{\mathrm{stat}}\| \le aD_j\mathbf{1}[d \neq s] + (1-a)B_j. \tag{29}$$

Taking expectations and using $0 \le a \le 1$ gives the expert-output bound $\mathbb{E}\|\delta Y\|_{\mathrm{F}} \le D_j\varepsilon_j + B_j\mathbb{E}(1-a)$, without an independence assumption. All probabilities and expectations refer to the same input distribution; on validation data they are empirical averages. Even when $d = s$, unit static weighting leaves discrepancy $(1-a)\|E_s(z_s)\|$. The comparison holds at a shared block input; upstream compilation can change downstream features.

Extension through fusion and both residual updates. At a shared input Z and stage conditions, $z_s, z_g, G_{\mathrm{anc}}, G_{\mathrm{route}}, \eta$ coincide between dynamic and static execution. In particular, compilation retains the router prior in the fusion gate while bypassing logits and selection.

Let $\delta Y = Y^{\mathrm{dyn}} - Y^{\mathrm{stat}}$, $g_j = \sigma(z_g) \odot \eta$, and $D_g = \mathrm{Diag}(\mathrm{vec}(g_j))$. Since both factors lie in $[0,1]$, $\|D_g\|_2 \le 1$. With Frobenius feature norms and the corresponding spectral norm of the pointwise projection W_o, Eq. (9) gives

$$\begin{aligned} \mathrm{vec}(\delta\Delta_{\mathrm{SDRM}}) &= D_g\,\mathrm{vec}(\delta Y), \\ \|\delta Z_{\mathrm{res}}\|_{\mathrm{F}} &\le \alpha_{\mathrm{res}}\|W_o\|_2\|g_j \odot \delta Y\|_{\mathrm{F}}. \end{aligned} \tag{30}$$

Suppose the internally residual map $\mathcal{F}_{\mathrm{FFN}}$ is $L_{\mathrm{FFN},j} = L_j$-Lipschitz on a domain containing the compared intermediate features. Its externally scaled update is $T_j(U) = (1-\alpha_{\mathrm{res}})U + \alpha_{\mathrm{res}}\mathcal{F}_{\mathrm{FFN}}(U)$. Because $0 < \alpha_{\mathrm{res}} = 0.1 < 1$, the triangle inequality yields

$$\begin{aligned} \|\delta Z_{\mathrm{out}}\|_{\mathrm{F}} &\le \kappa_j\|g_j \odot \delta Y\|_{\mathrm{F}}, \\ \kappa_j &= \alpha_{\mathrm{res}}\|W_o\|_2\big[(1-\alpha_{\mathrm{res}}) + \alpha_{\mathrm{res}}L_{\mathrm{FFN},j}\big], \\ \mathcal{R}_j &= \mathbb{E}\big[a\|g_j \odot (E_d - E_s)\|_{\mathrm{F}} + (1-a)\|g_j \odot E_s\|_{\mathrm{F}}\big], \\ \mathbb{E}\|\delta Z_{\mathrm{out}}\|_{\mathrm{F}} &\le \kappa_j\mathcal{R}_j \le \kappa_j D_j\varepsilon_j + \kappa_j B_j\mathbb{E}(1-a). \end{aligned} \tag{31}$$

Here E_d, E_s are evaluated at the same z_s, and $L_{\mathrm{FFN},j}$ is uniform on the comparison domain. Substituting Eq. (28) into the gated update proves both claims of Eq. (16). Both contributions inherit the gain κ_j; the derivation does not assume $\kappa_j \le 1$.

Interpretation for SAR-IR deployment. The retained anchor branch provides query-dependent context independently of scan order. At a shared input it contributes identically in both modes, so compilation perturbs only the directional contribution, weighted elementwise by g_j. The tighter quantity $\mathcal{R}_j$ preserves the dependence between routing confidence, expert differences, and fusion weights. Even frequent route changes can have limited effect when expert outputs are similar in the coordinates emphasized by fusion. Conversely, matching the selected expert still leaves the confidence-removal term. The bound therefore connects compilation sensitivity to feature differences and retained adaptive fusion, beyond route agreement alone.

Residual scaling explicitly weights this perturbation, while projection and FFN gains enter κ_j. The result is a conditional block bound; κ_j can exceed one, and preceding compiled blocks can change later features, anchor selections, and gates. End-to-end restoration is therefore assessed with the fully compiled checkpoint (Table 15); applying the analysis to a new input distribution requires its corresponding expectations and bounds.

B Datasets and Signal Representations

B.1 Scene Sources and Dataset Construction

RFI-free background echoes for model development are sampled from 73 distinct SAR acquisition scenes. Eleven separate scenes provide simulated testing, and four further scenes contain measured RFI. The three scene groups are disjoint; Figure 9 shows their geographic distribution.

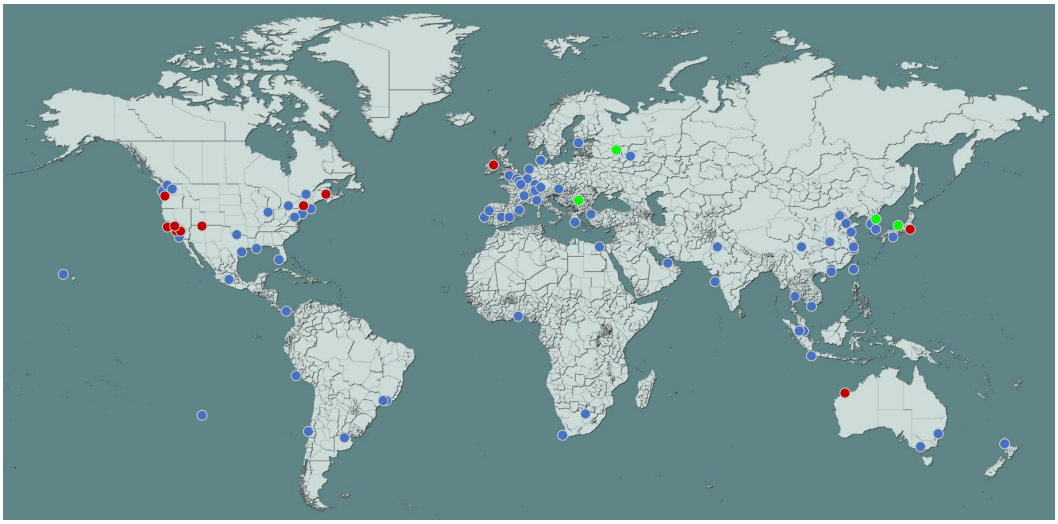


Figure 9: Geographic distribution of SAR scenes: blue, background-echo sources for simulated training and validation; red, simulated-test scenes; green, measured-test scenes.

For each candidate RFI-free time-domain echo from the development scenes, we form an STFT time-frequency image. We compare image-feature embeddings from a fine-tuned CLIP model (Radford et al., 2021) across candidate images and retain 11,000 with low mutual similarity. Their corresponding complex time-domain echoes serve as RFI-free backgrounds for simulation. We assign 1,000 echoes to each of the 11 interference categories, synthesize one corrupted echo per background using Section B.3, and retain the background as its paired restoration target. A category-stratified 9:1 split yields 900 training and 100 validation pairs per category. Thus, training and validation are separated at the sample level, whereas both test sets use scene-disjoint acquisitions.

The simulated test set covers the same 11 interference categories; the measured scenes contain MPFI, SFI2, SFI2 with single false-target interference (SFI2+SFTI), and UWI. Table 7 summarizes the splits. Simulated inputs have pulse-aligned RFI-free targets; measured RFI-free revisits provide same-scene qualitative references.

Table 7: Scene sources and echo counts by split. The 11 simulated categories comprise five single and six compound interference types.

Split	Source scenes	RFI types	Echoes
Simulated train	73-scene pool	11 simulated	9,900 (900/type)
Validation	Same pool	11 simulated	1,100 (100/type)
Simulated test	11 test scenes	11 simulated	15,784 (1,409–1,548/type)
Measured test	4 test scenes	4 measured	5,775 (UWI: 1,548; others: 1,409)

B.2 Signal Representations and Mask Construction

The full-record FFT and STFT operate on the same deterministic echo crop (Table 8). The FFT provides fine spectral resolution, while the STFT preserves temporal localization on the restoration grid. For mask supervision, we manually annotate 50 corrupted TFD images per interference category from the development data and use these annotations to fine-tune the Segment Anything Model (SAM) (Kirillov et al., 2023). We use SAM to generate masks for the training and validation splits and manually check them. The training masks supervise Fast-SCNN (Poudel et al., 2019), while the validation masks select its checkpoint. The selected Fast-SCNN is then frozen and predicts a mask from each corrupted TFD visualization supplied to the restoration network during training, validation, and inference.

B.3 Simulated Interference Construction

Each contaminated pulse combines a measured interference-free echo with simulated interference:

$$x(t) = c_{\text{meas}}(t) + i(t) = s_{\text{scene}}(t) + n(t) + i(t). \tag{32}$$

Table 8: Signal transforms. FFT and STFT use unnormalized, two-sided complex spectra in zero-centered frequency order.

Representation	Transform setting and coefficient grid
Echo crop	19,381 complex samples, shared by both transforms.
FD	Full-record FFT; 19,381 frequency bins.
TFD	STFT with a periodic Hamming window and FFT length 256, hop 75, and no temporal padding; 256×256 grid.

Since the scene echo and receiver noise are inseparable in c_{meas}, we use the interference-to-signal-plus-noise ratio (ISNR):

$$\mathrm{ISNR} = 10 \log_{10} \frac{E(i)}{E(c_{\mathrm{meas}})}, \qquad E(u) = \sum_{m=0}^{N-1} |u[m]|^2. \tag{33}$$

An unscaled realization i_0 is normalized to a target $\gamma \in [0, 30]$ dB by

$$i = \sqrt{\frac{E(c_{\mathrm{meas}}) 10^{\gamma/10}}{E(i_0)}}\, i_0. \tag{34}$$

Table 9 lists the simulation settings; f_c, B_s, and T_p denote carrier frequency, transmitted-signal bandwidth, and pulse width. Figure 10 illustrates the five single-interference types.

Table 9: Parameters for 11 simulated interference categories. Compound components are combined before scaling to the target ISNR.

Type	Parameter	Setting
All	Background echo	Measured interference-free SAR echo
	ISNR	$[0, 30]$ dB; Eqs. (33)-(34)
MPFI	Components N_f	10-200
	Bandwidth B_i	$[B_s, 2B_s]$
	Phase modulation	BPSK or QPSK
SFI1	Center frequency f_i	$f_c + [-B_s/2, B_s/2]$
	Bandwidth / sweep	$[B_s, 2B_s]$ / sinusoidal
SFI2	Center frequency f_i	$f_c + [-B_s/2, B_s/2]$
	Bandwidth / sweep	$[B_s, 2B_s]$ / sawtooth
DFTI	False targets / delay	5-15 / $[0.1T_p, T_p]$
	Retransmissions	3-10
ISFI	Sampling width T_w	$[0.1T_p, 0.5T_p]$
	Forwarding count	5-15
Compound	Composition	One of MPFI/SFI1/SFI2 plus one of DFTI/ISFI

Multi-Point Frequency Interference (MPFI). MPFI superposes N_f phase-modulated narrowband components:

$$i_{\mathrm{MPFI}}(t) = \sum_{k=1}^{N_f} A_k \exp(j[2\pi f_k t + \varphi_k(t)])\,, \tag{35}$$

where A_k, f_k, and $\varphi_k(t)$ are component amplitude, frequency, and phase law.

Sweep-Frequency Interference (SFI1 and SFI2). Both types share the carrier and integrated sweep model

$$\begin{aligned} i_{\mathrm{SFI}k}(t) &= A_i \exp(j2\pi[f_i t + \beta_i \phi_k(t)])\,, \\ \phi_k(t) &= \int_0^t g_k(\xi)\, d\xi, \qquad f_{\mathrm{inst},k}(t) = f_i + \beta_i g_k(t), \quad k \in \{1, 2\}. \end{aligned} \tag{36}$$

Here A_i, f_i, and β_i are amplitude, center frequency, and modulation index. SFI1 uses a sinusoidal g_1, producing smooth curved ridges; SFI2 uses a sawtooth g_2, producing linear sweeps with abrupt resets (Figure 10(b,c)).

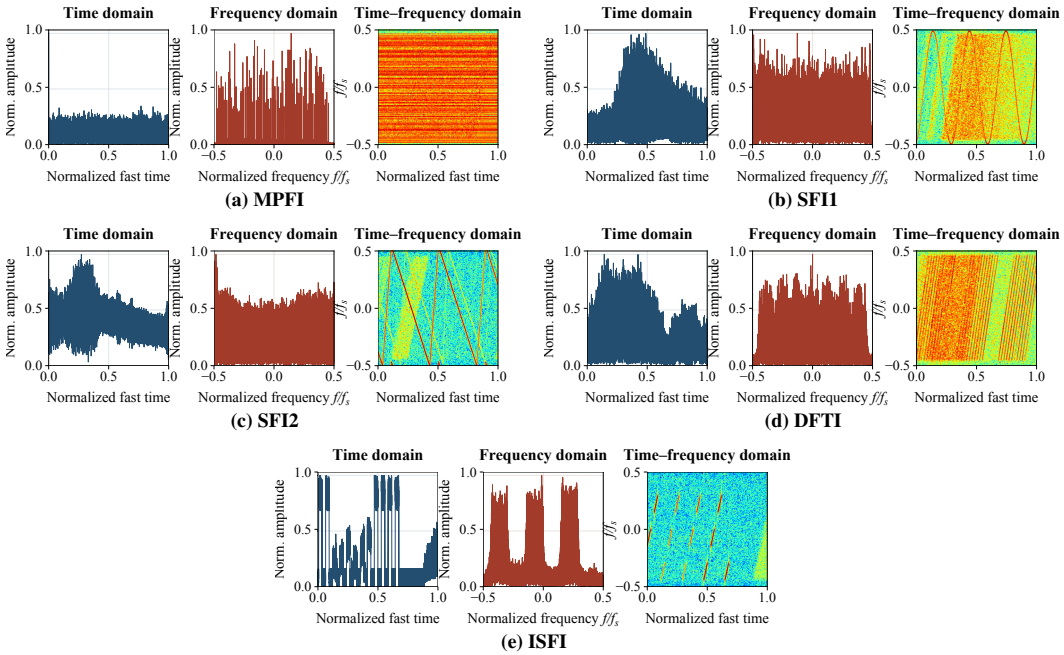


Figure 10: Five single simulated interference types, arranged in row-major order. Each panel shows time, frequency, and time-frequency views from left to right under the common transform and display settings.

Dense False-Target Interference (DFTI). The deception simulator coherently superposes delayed replicas of an intercepted waveform u_{int}, distinct from the unknown scene term in Eq. (32):

$$i_{\mathrm{DFTI}}(t) = \sum_{k=1}^{N_d} A_k u_{\mathrm{int}}(t - \tau_k). \tag{37}$$

Amplitudes A_k and delays τ_k follow the false-target and retransmission draws in Table 9.

Interrupted Sampling and Forwarding Interference (ISFI). ISFI gates and forwards fragments of the intercepted pulse:

$$i_{\mathrm{ISFI}}(t) = \sum_{r=1}^{N_{\mathrm{rep}}} \mathrm{rect}\left(\frac{t - (2r+1)T_w}{T_w}\right) u_{\mathrm{int}}(t - T_w), \tag{38}$$

where $\mathrm{rect}(u) = 1$ for $|u| \le 1/2$ and zero otherwise, T_w is the sampled-segment width, and N_{rep} is the forwarding count.

Compound interference. Each compound realization combines a suppressive component and a deception component before normalization:

$$i_{\mathrm{cmp}}(t) = \alpha_{\mathrm{sup}} i_{\mathrm{sup}}(t) + \alpha_{\mathrm{dec}} i_{\mathrm{dec}}(t). \tag{39}$$

Here i_{sup} is drawn from MPFI/SFI1/SFI2 and i_{dec} from DFTI/ISFI; the α coefficients set their relative powers. The combined waveform is then scaled by Eq. (34). The Cartesian product gives six compound types and, together with the five single types, 11 categories. Figure 11 shows how global spectral occupancy and echo-dependent support coexist in their TFDs.

B.4 Measured Data and Visualization

Measured data cover MPFI, SFI2, SFI2+SFTI, and unintentional wideband interference (UWI). Interference-free revisits cover the same nominal scenes. Figure 12 shows one deterministically selected corrupted pulse per setting under the transforms in Table 8.

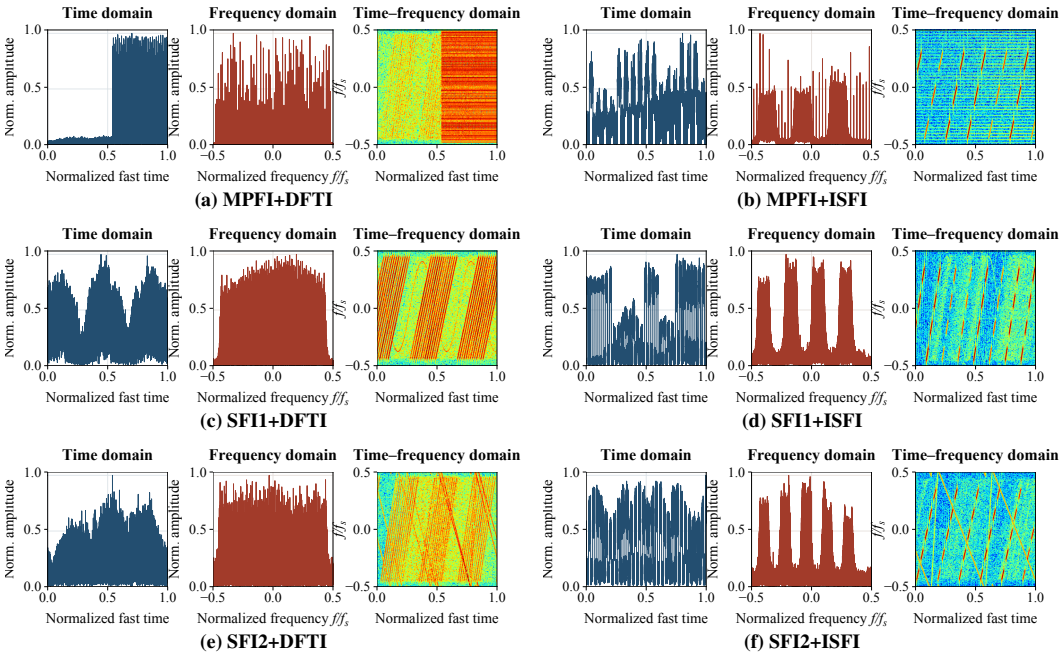


Figure 11: Six compound simulated interference types. Rows group the suppressive component (MPFI, SFI1, SFI2); columns group the deception component (DFTI, ISFI). Each panel shows time, frequency, and time-frequency views from left to right under the common display settings.

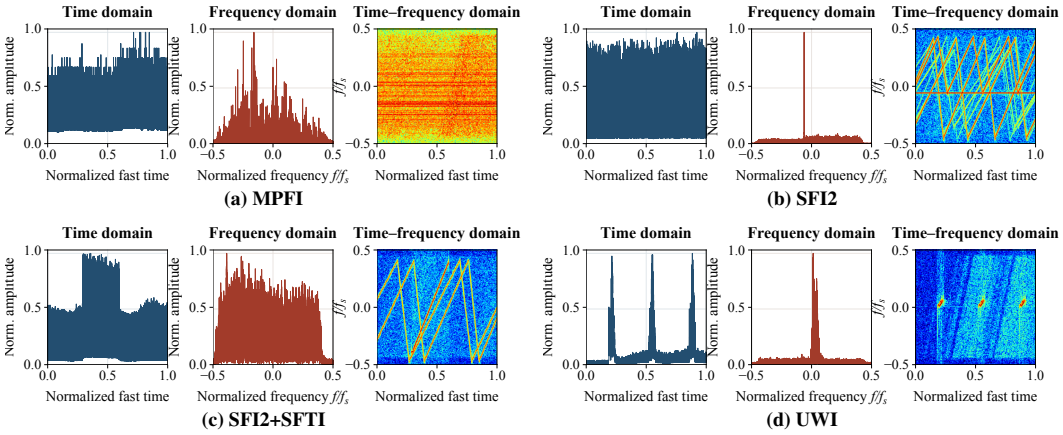


Figure 12: Measured interference examples: (a) MPFI, (b) SFI2, (c) SFI2+SFTI, and (d) UWI. Each panel shows time, frequency, and time-frequency views from left to right under the common transform and display settings.

C Student and Teacher Architectures

C.1 Backbones and Stage Organization

Both four-scale encoder-decoders receive complex TFD coefficients $X_{\mathrm{TF}} \in \mathbb{R}^{B\times 2\times 256\times 256}$, FD conditions $X_{\mathrm{F}} \in \mathbb{R}^{B\times 2\times 19,381}$, and masks $m \in [0,1]^{B\times 1\times 256\times 256}$. The two TFD channels encode real and imaginary parts. Table 10 specifies the SDRM student and a Restormer-style teacher (Zamir et al., 2022). The teacher uses multi-Dconv head transposed attention (MDTA) for channel attention with spatially enriched projections and a gated-Dconv feed-forward network (GDFN) for local feature transformation. Its wider, deeper stages supply reference features for multi-scale distillation.

Each decoder stage upsamples, concatenates the corresponding post-block encoder feature, and applies skip fusion, FMCG, and restoration blocks. FMCG executes once per stage occurrence in both networks, sharing parameters between encoder and decoder at the same resolution within a

Table 10: Student and teacher backbones. Shapes omit the batch axis; C is output width and n is block count. Decoder stages run from coarse to fine.

Stage	Grid	Student (SDRM)		Teacher (MDTA+GDFN)		
		C_{S}	n_{S}	C_{T}	n_{T}	Heads
Encoder 0	256×256	32	2	64	3	1
Encoder 1	128×128	64	2	128	4	2
Encoder 2	64×64	128	2	256	6	4
Encoder 3 (bottleneck)	32×32	256	2	512	6	8
Decoder 2	64×64	128	2	256	4	4
Decoder 1	128×128	64	2	128	3	2
Decoder 0	256×256	32	2	64	3	1
Total restoration blocks			14		29	

model. The student then executes two SDRM blocks per occurrence, yielding 14 SDRM block executions across seven stages; teacher block counts follow Table 10. Teacher and student have separate backbone and condition-encoder parameters. The head directly regresses restored complex TFD coefficients.

Table 11 specifies the TFD stem/head, scale changes, skip fusion, FD/mask encoders, and FMCG projections used by both networks. These operators establish their complex-TFD input/output interface; Table 12 details the distinct restoration blocks.

Table 11: Shared operators and conditioning. C_ℓ follows Table 10; GN uses eight groups and DWConv is depthwise convolution. Stride is one unless noted.

Component	Operator / configuration
TFD stem / head	3×3 Conv, $2 \to C_0$ / $C_0 \to 2$; padding one, with no output activation.
Downsampling	3×3 Conv, stride two, padding one, $C_\ell \to C_{\ell+1}$; GN and GELU.
Upsampling	2×2 transposed Conv, stride two, $C_{\ell+1} \to C_\ell$; 3×3 Conv, GN, GELU.
Skip fusion	Channel concatenation with encoder level ℓ, followed by 1×1 Conv, $2C_\ell \to C_\ell$.
FD stream	Seven-tap Conv1D stem, GN, GELU; residual frequency blocks per scale: student $[2, 2, 2, 2]$, teacher $[3, 4, 6, 6]$. Three-tap stride-two Conv1D, GN, GELU between scales.
Mask stream	3×3 Conv stem, GN, GELU; one residual mask block per scale; the same downsampling operator as above.
FMCG	Four scale-specific modules per model. Spectral and spatial compatibility/value projections are 1×1 convolutions; sigmoid compatibility gates initialize to 0.05.

C.2 Restoration Block Specifications

Block operations. For teacher head h, Q/K/V have shape $64 \times H_\ell W_\ell$. Spatially normalized Q/K produce channel attention $A_h = \mathrm{softmax}(\tau_h \widehat{Q}_h \widehat{K}_h^\top)$, applied as $A_h V_h$. GDFN expands to $2\lfloor 2.66C \rfloor$ channels before depthwise mixing and the split gated product in Table 12. Each student block has independent experts, router, and anchor projections. Step-projection ranks are $[4, 8, 16, 32]$; step biases use inverse-softplus values of log-uniform draws from $[10^{-3}, 10^{-1}]$. Exact scans and fusion appear in Section A.

C.3 Conditional Encoders and FMCG

Native FD features $C_{\mathrm{F}}^{(\ell)}$ have lengths $[19{,}381, 9{,}691, 4{,}846, 2{,}423]$. Each residual frequency block uses pre-normalized seven-tap depthwise Conv1D, GELU, and pointwise projection, followed by a pre-normalized pointwise feed-forward residual branch with expansion two and GELU. Frequency-only interpolation gives

$$\widetilde{C}_{\mathrm{F}}^{(\ell)} = \mathcal{A}_{H_\ell}^{(f)}(C_{\mathrm{F}}^{(\ell)}) \in \mathbb{R}^{B \times C_\ell \times H_\ell \times 1}. \tag{40}$$

The singleton time axis supplies frame-shared spectral context. Each mask block uses two 3×3 convolutions with intervening GELU, residual addition, and a 1×1 output projection. Its pyramid

Table 12: Student and teacher block configurations. Lists follow scales $\ell = 0, 1, 2, 3$; decoder settings match their resolution. C is the corresponding stage width.

Property	Student: SDRM	Teacher: MDTA+GDFN
Normalization	Eight-group GN before local mixing	Eight-group GN before attention and before GDFN
Local projection	1×1 Conv and 3×3 DWConv; input projection splits into two $2C$-channel tensors	1×1 QKV projection and 3×3 DWConv; 1×1 output projection
Global operator	Seven directional SSM experts; one executes per sample and block	Channel attention with heads $[1, 2, 4, 8]$, 64 channels per head
Internal sizes	Propagation width $2C$, state size 16, step-projection rank $\lceil 2C/16 \rceil$	GDFN hidden width $\lfloor 2.66C \rfloor$: $[170, 340, 680, 1361]$
Context / gating	Anchor budgets $[32, 32, 64, 64]$; sigmoid prior/anchor gates initialize to 0.05	Spatially normalized Q/K; learned per-head temperature initialized to one
Feed-forward	Two 3×3 convolutions, $C \to 2C \to C$, with GN and SiLU residual refinement	Gated depthwise feed-forward: $\mathrm{GELU}(u) \odot v$, then 1×1 projection to C
Residual scales	0.1 for propagation and Conv-FFN updates	Unit residual additions for attention and GDFN
Inference	Dynamic: non-renormalized Top-1; deployed static: one validation-selected expert per block	Frozen during distillation; absent from student inference

and alignment are

$$P_m^{(0)} = \phi_0(m), \qquad P_m^{(\ell+1)} = \phi_{\ell+1}(\mathrm{Down}(P_m^{(\ell)})),$$
$$\widetilde{P}_m^{(\ell)} = \mathcal{A}_{H_\ell, W_\ell}^{(f,t)}(P_m^{(\ell)}) \in \mathbb{R}^{B \times C_\ell \times H_\ell \times W_\ell}. \tag{41}$$

Alignment is identity when sizes match. In Eq. (4), $R_f \in \mathbb{R}^{B \times C_\ell \times H_\ell \times 1}$ broadcasts over time and $R_m \in \mathbb{R}^{B \times 1 \times H_\ell \times W_\ell}$ over channels. Compatibility and value projections are 1×1 convolutions; compatibility weights initialize to zero, with biases giving sigmoid gates of 0.05. $\mathcal{B}_t$ broadcasts the FD update without projection. Stage condition features and gates are cached for reuse by SDRM blocks; each block applies its own prior-value projection to the aligned mask.

D Training Configuration and Evaluation Protocol

D.1 Feature Alignment and Training Configuration

Training stages. The objective and batch averaging are defined in Eqs. (10)–(14). AdamW updates the student and alignment layers for 100 distillation epochs at 2×10^{-4}, with the teacher frozen. Validation restoration loss selects teacher and student checkpoints. We use $\epsilon = 10^{-3}$, $\lambda_{\mathrm{rest}} = \lambda_{\mathrm{KD}} = 1$ during distillation, $\lambda_{\mathrm{out}} = 0.5$, $\lambda_{\mathrm{feat}} = 0.1$, $\lambda_{\mathrm{bal}} = 10^{-2}$, and $\lambda_z = 10^{-3}$.

Feature alignment. $\Pi_{\tau,\phi}$ applies bilinear spatial resizing when needed, followed by learned 1×1 channel projections $32 \to 64, 64 \to 128, 128 \to 256, 256 \to 512$ at levels 0–3. Feature weights are $\omega_{\mathrm{restored}} = 1$, $\omega_{\mathrm{dec0}} = 0.75$, $\omega_{\mathrm{dec1}} = \omega_{\mathrm{dec2}} = 0.50$, and $\omega_{\mathrm{enc}\ell} = 0.25$ for $\ell = 0, \ldots, 3$. The restored-feature key aliases the final decoder feature before the head, but has a separate weighted loss and projection. Alignment layers are discarded at inference.

D.2 Controlled Ablation Settings

Condition sources. The full model receives TFD, FD, and mask inputs. The condition-source variants remove FD, the mask, or both while retaining the TFD prediction task.

FMCG interventions. All four variants retain TFD, FD, and mask inputs and modify FMCG's gates or fusion update. *Without spectral compatibility* replaces the input-dependent $R_f^{(\ell)}$ in Eq. (4) with a learned stage scalar $\sigma(a_f^{(\ell)})$, broadcast over its frequency–channel map; the spatial gate remains unchanged. *Without spatial compatibility* likewise replaces $R_m^{(\ell)}$ with $\sigma(a_m^{(\ell)})$, broadcast over

the spatial map, while retaining the spectral gate. *Fixed learned gains* applies both replacements. *Naive concatenation* broadcasts aligned FD features along time, concatenates them with the TFD and aligned mask features, and applies a 1×1 fusion projection in place of FMCG's gated update. The downstream SDRM router, reliability scoring, and anchor fusion retain their connections to the resulting stage features and spatial gate.

SDRM interventions. The *axial-only bank* removes stripe and ridge experts, retaining $\{F^+, F^-, T^+, T^-\}$; it changes both scan geometry and bank cardinality. *Without mask-aware routing* removes the mask-derived descriptor feature while preserving mask pathways in FMCG, anchor ranking, and fusion. *Without reliability anchors* removes anchor selection, attention, and context while retaining FMCG and routing. Conditioned stage features still enter the remaining experts.

D.3 Evaluation Data and Reconstruction

All methods share the test manifests in Section B.1. The 1,100-sample validation split selects checkpoints by restoration loss and VSRC experts by dynamic-route frequency on corrupted inputs (Eq. (15)). Paired clean targets define simulated-data references; measured revisits are qualitative.

For MorphRoute-Mamba, parameter counts include the frozen Fast-SCNN and the restoration network. Latency times Fast-SCNN mask prediction from the prepared TFD visualization followed by the restoration-network forward pass on TFD, FD, and the predicted mask, including learned conditioning and dynamic routing where applicable.

Inverse reconstruction matches the STFT window, hop, centering, normalization, sidedness, and frequency order in Table 8, yielding the natural 19,381-sample pulse. Restored pulses are reassembled at their original acquisition indices and focused once per complete echo block. The same transform and focusing pipeline applies to every method.

D.4 Simulated-Data Metrics and Aggregation

For a complex prediction $\widehat{X}$ and target $X^\star$, TFD cNMSE is

$$\mathrm{cNMSE}_{\mathrm{TF}} = 10 \log_{10} \frac{\|\widehat{X} - X^\star\|_{\mathrm{F}}^2}{\|X^\star\|_{\mathrm{F}}^2}. \tag{42}$$

Time-domain NMSE uses the same ratio after the matched inverse transform. Echo results aggregate all 15,784 test pulses. Focused-image RMSE, SSIM, and PSNR use paired images with the same deterministic intensity mapping across methods and are macro-averaged over 11 categories. Tables 13 and 14 provide all category values; their Overall rows follow these respective aggregation rules.

D.5 Measured-Scene Metric Implementation

AG and MNR are computed on the focused amplitude images, measuring spatial sharpness and relative background energy, respectively. All methods share the focusing pipeline and scene-specific ROIs. Revisit references guide structural inspection.

Let $I_s \in \mathbb{R}_{\geq 0}^{H_s \times W_s}$ denote the focused amplitude image for scene s. Define the forward differences $\delta_r I_s(r,c) = I_s(r+1,c) - I_s(r,c)$ and $\delta_c I_s(r,c) = I_s(r,c+1) - I_s(r,c)$. The implemented average gradient is

$$\mathrm{AG}(I_s) = \frac{1}{(H_s-1)(W_s-1)} \sum_{r=0}^{H_s-2} \sum_{c=0}^{W_s-2} \sqrt{\frac{\delta_r I_s(r,c)^2 + \delta_c I_s(r,c)^2}{2}}. \tag{43}$$

For fixed non-scattering and reference ROIs Ω_N and Ω_R, respectively, MNR is

$$\mathrm{MNR}(I_s) = 10 \log_{10} \frac{|\Omega_N|^{-1} \sum_{u \in \Omega_N} I_s(u)^2}{|\Omega_R|^{-1} \sum_{u \in \Omega_R} I_s(u)^2}. \tag{44}$$

E Additional Experimental Results and Analysis

E.1 Restoration across Interference Categories

Tables 13 and 14 report category-wise echo and focused-image results. MorphRoute-Mamba achieves the lowest time-domain NMSE, TFD cNMSE, and image RMSE, and the highest SSIM and PSNR across all 11 simulated categories, including the six compound interference categories. This consistency across interference types and reconstruction domains supports the method's broad effectiveness on the evaluated categories.

Table 13: Simulated echo errors by category (lower is better). Rankings use unrounded values; Overall aggregates all test pulses.

Category	Metric (dB)	ESP	RPCA	PISNet	WBIM-GAN	EAMamba	MambaIRv1	MambaIRv2	MorphRoute
DFTI	Time NMSE	0.99	4.81	-4.71	-9.11	-8.44	-9.56	-10.71	-13.18
	TFD cNMSE	0.98	4.81	-3.77	-9.40	-8.35	-9.70	-10.80	-13.65
ISFI	Time NMSE	1.82	5.26	-5.91	-5.95	-6.89	-6.16	-7.28	-12.58
	TFD cNMSE	1.81	5.24	-4.39	-6.16	-6.78	-6.17	-7.26	-12.77
MPFI	Time NMSE	-2.93	8.07	-2.01	-7.67	-5.58	-8.23	-8.64	-9.14
	TFD cNMSE	-2.94	8.06	-1.37	-7.62	-5.41	-8.31	-8.64	-9.45
SFI1	Time NMSE	0.20	3.54	-5.84	-8.15	-8.64	-10.68	-11.00	-14.00
	TFD cNMSE	0.20	3.52	-4.76	-8.37	-8.51	-10.73	-10.96	-14.15
SFI2	Time NMSE	1.03	3.70	-5.79	-7.81	-8.64	-9.50	-10.45	-13.76
	TFD cNMSE	1.02	3.68	-4.49	-8.04	-8.47	-9.55	-10.41	-13.96
MPFI+DFTI	Time NMSE	0.83	11.92	-0.05	-4.27	-2.63	-3.71	-4.38	-6.24
	TFD cNMSE	0.83	11.90	0.53	-4.24	-2.50	-3.72	-4.34	-6.37
MPFI+ISFI	Time NMSE	5.05	12.79	-1.35	-4.57	-4.06	-6.17	-6.58	-8.25
	TFD cNMSE	5.04	12.78	-0.75	-4.60	-3.94	-6.21	-6.55	-8.42
SFI1+DFTI	Time NMSE	1.68	9.95	-2.54	-6.23	-4.99	-6.71	-7.02	-9.64
	TFD cNMSE	1.67	9.93	-1.92	-6.36	-4.90	-6.74	-7.00	-9.76
SFI1+ISFI	Time NMSE	0.60	9.54	-4.32	-5.90	-6.48	-7.91	-8.05	-10.88
	TFD cNMSE	0.59	9.52	-3.52	-6.06	-6.39	-7.94	-8.03	-10.99
SFI2+DFTI	Time NMSE	1.64	9.56	-2.59	-4.30	-5.19	-6.33	-6.85	-9.74
	TFD cNMSE	1.64	9.56	-1.92	-4.41	-5.09	-6.39	-6.86	-10.00
SFI2+ISFI	Time NMSE	7.66	9.99	-3.97	-3.64	-6.71	-7.22	-8.24	-11.27
	TFD cNMSE	7.70	10.04	-3.06	-3.88	-6.62	-7.39	-8.33	-11.85
Overall	Time NMSE	1.64	8.07	-3.55	-6.19	-6.22	-7.50	-8.14	-10.79
	TFD cNMSE	1.64	8.07	-2.67	-6.33	-6.10	-7.56	-8.14	-11.04

E.2 Measured-Scene Restoration

Table 2 evaluates MorphRoute-Mamba, trained and VSRC-calibrated on simulated data, in four measured scenes disjoint from the training and validation scenes. Without restoration-model fine-tuning, it achieves the highest AG and lowest MNR in every scene: mean AG is 28.83 versus 25.73 for the strongest baseline, and mean MNR is -9.32 versus -5.36 dB. On the focused amplitude images, higher AG indicates sharper structures, while lower MNR indicates less non-scattering energy relative to reference scatterers; together, they support stronger target-detail retention and interference suppression. The advantage extends to SFI2+SFTI and UWI, interference types absent from the simulated training set. For SFI2+SFTI, the gains over the respective runners-up reach 5.68 AG and 4.75 dB MNR, supporting transfer to both unseen scenes and unseen interference types.

E.3 Directional Routing and Anchor Contributions

Figure 13 records the dynamic router's expert choices on measured inputs. Blocks E0.0 and E3.0 favor the same axial orders across all four scene/interference settings, whereas E2.0 favors time-major traversal for MPFI, stripe-serpentine traversal for SFI2 and UWI, and ridge traversal for SFI2+SFTI. Thus, the learned bank combines broadly used axial paths with non-axial options selected in particular settings.

The restoration effect is tested by the controlled SDRM ablations in Table 6. Relative to the full model, the axial-only bank increases time-domain NMSE and TFD cNMSE by 1.16 and 1.15 dB and lowers PSNR by 1.18 dB; this comparison changes both scan geometry and expert count. Re-

Table 14: Simulated image quality by category: RMSE ↓, SSIM ↑, PSNR ↑. Overall averages the 11 unrounded category values.

Category	Metric	ESP	RPCA	PISNet	WBIM-GAN	EAMamba	MambaIRv1	MambaIRv2	MorphRoute
DFTI	RMSE	1.1500	4.6038	0.6680	0.4140	0.4430	0.4176	0.3364	0.2584
	SSIM	0.8883	0.4736	0.9621	0.9842	0.9838	0.9826	0.9906	0.9945
	PSNR (dB)	40.33	28.29	45.05	49.21	48.62	49.13	51.01	53.30
ISFI	RMSE	7.4579	16.6139	1.8284	2.0394	1.8404	2.3897	2.0560	0.9043
	SSIM	0.8231	0.6326	0.9888	0.9804	0.9857	0.9742	0.9823	0.9979
	PSNR (dB)	41.28	34.33	53.49	52.55	53.44	51.17	52.47	59.61
MPFI	RMSE	0.9698	6.9641	0.8377	0.4569	0.6112	0.4566	0.4014	0.3790
	SSIM	0.9629	0.3352	0.9772	0.9925	0.9865	0.9914	0.9953	0.9957
	PSNR (dB)	47.48	30.36	48.75	54.02	51.49	54.03	55.15	55.64
SFI1	RMSE	2.0483	5.2290	0.9704	0.8230	0.7281	0.5813	0.5466	0.3848
	SSIM	0.7616	0.4765	0.9660	0.9700	0.9815	0.9885	0.9901	0.9953
	PSNR (dB)	39.50	31.36	45.99	47.42	48.48	50.44	50.97	54.02
SFI2	RMSE	2.2109	6.7678	1.1745	1.0969	0.8757	0.8902	0.7320	0.4780
	SSIM	0.9275	0.5240	0.9789	0.9732	0.9879	0.9860	0.9920	0.9971
	PSNR (dB)	43.32	33.60	48.82	49.41	51.37	51.22	52.92	56.62
MPFI+DFTI	RMSE	2.9433	22.8261	2.7258	1.8851	2.1603	2.1514	1.8168	1.4540
	SSIM	0.9736	0.3328	0.9793	0.9890	0.9872	0.9842	0.9915	0.9953
	PSNR (dB)	49.97	32.18	50.64	53.84	52.66	52.69	54.16	56.10
MPFI+ISFI	RMSE	5.5408	14.2795	1.3936	1.1399	1.1318	0.9142	0.7807	0.6526
	SSIM	0.6558	0.2479	0.9744	0.9727	0.9797	0.9873	0.9933	0.9956
	PSNR (dB)	37.30	29.08	49.29	51.04	51.10	52.95	54.32	55.88
SFI1+DFTI	RMSE	2.1484	11.3894	1.5125	1.1265	1.2408	0.9599	0.9257	0.7097
	SSIM	0.9192	0.3117	0.9601	0.9739	0.9706	0.9858	0.9867	0.9924
	PSNR (dB)	42.97	28.48	46.02	48.58	47.74	49.97	50.28	52.59
SFI1+ISFI	RMSE	2.0813	11.5543	1.2429	1.1494	1.0381	0.8172	0.8027	0.6064
	SSIM	0.8763	0.3059	0.9655	0.9560	0.9743	0.9862	0.9873	0.9927
	PSNR (dB)	41.65	26.76	46.12	46.80	47.69	49.77	49.92	52.36
SFI2+DFTI	RMSE	1.8631	8.5688	1.1726	1.1324	0.9165	0.7847	0.7271	0.5391
	SSIM	0.8565	0.2563	0.9544	0.9220	0.9711	0.9769	0.9824	0.9906
	PSNR (dB)	40.44	27.19	44.46	44.77	46.60	47.95	48.61	51.21
SFI2+ISFI	RMSE	3.9290	5.4714	0.5776	0.7007	0.4503	0.4891	0.3922	0.2549
	SSIM	0.4155	0.3523	0.9757	0.9322	0.9836	0.9799	0.9888	0.9956
	PSNR (dB)	31.09	28.21	47.74	46.06	49.90	49.19	51.10	54.85
Overall	RMSE	2.9403	10.3880	1.2822	1.0877	1.0397	0.9865	0.8652	0.6019
	SSIM	0.8237	0.3863	0.9711	0.9678	0.9811	0.9839	0.9891	0.9948
	PSNR (dB)	41.39	29.98	47.85	49.43	49.92	50.77	51.90	54.74

moving the mask descriptor only from routing increases time-domain NMSE and TFD cNMSE by 0.76 and 0.92 dB and lowers PSNR by 0.95 dB, showing that support-aware selection contributes beyond the mask pathways retained elsewhere. Removing reliability anchors raises NMSE/cNMSE by 0.57/0.60 dB and lowers PSNR by 0.70 dB, the smallest deficits among these ablations, consistent with direct context retrieval complementing scan-ordered propagation. The measured results in Table 2 use compiled block-wise routes; FMCG, anchors, and fusion remain input-dependent after compilation.

E.4 Training Progression

Table 15 separates training gains from deployment costs. With the dynamic architecture fixed at 21.11M parameters, 279.03G FLOPs, and 71.82 ms, KD raises focused-image PSNR/SSIM from 52.14 dB/0.9896 to 54.95 dB/0.9950 (+2.81 dB/+0.0054). Thus, the gain comes from the training stage without increasing the tabulated inference cost. Against MambaIRv2, the strongest distilled baseline in image quality, the dynamic model gains 0.58 dB/0.0016 with 6.3× fewer FLOPs and 10.8× lower latency.

Static compilation trades only 0.21 dB/0.0002 in PSNR/SSIM for a 7.03 ms (9.79%) latency reduction relative to the dynamic checkpoint; parameters and FLOPs also fall by 1.62M and 1.48G. Its 54.74 dB/0.9948 still exceeds every distilled baseline in both quality metrics. In particular, it leads MambaIRv2 by 0.37 dB/0.0014 while running 12.0× faster, and its 64.79 ms latency is the lowest in the table.

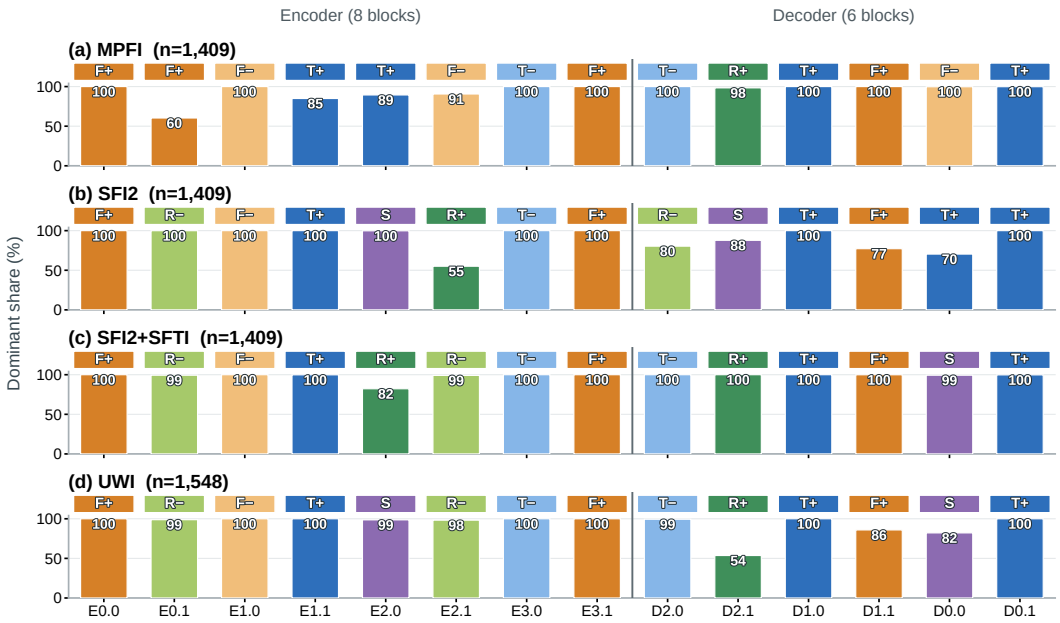


Figure 13: Dynamic-router dominant-direction shares across 14 SDRM blocks on measured scenes. Each row represents one scene; the colored header tile identifies the modal direction, and the bar height and in-bar label give the percentage of pulses selecting that direction. The vertical divider separates the eight encoder blocks from the six decoder blocks. F/T/S/R denote frequency-major, time-major, stripe-serpentine, and ridge traversals; $+/-$ refer to the paired orders in Section A.1.

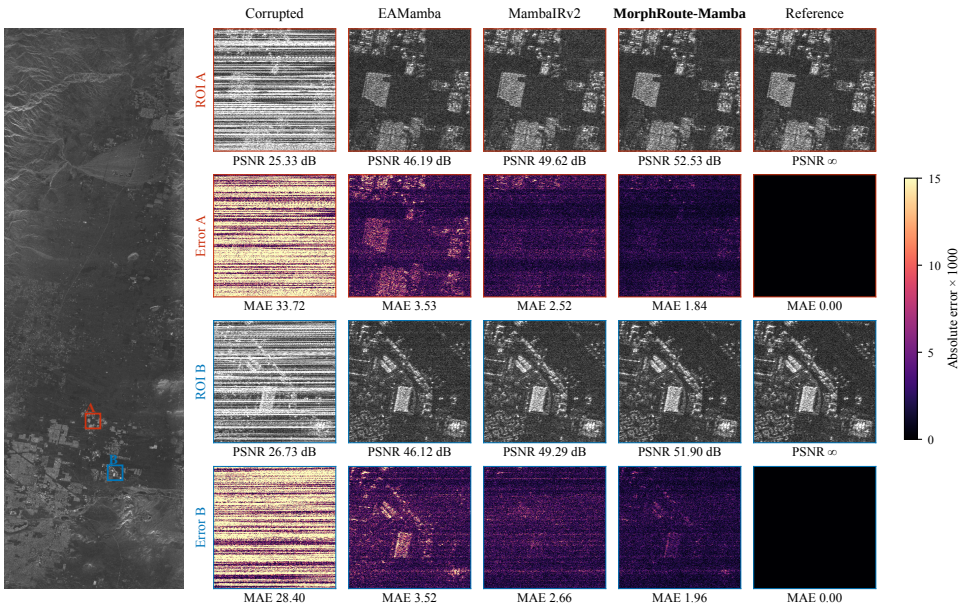
Table 15: Distilled baselines and MorphRoute-Mamba training-to-deployment progression on the simulated test set. SUP denotes supervised training and KD denotes knowledge distillation; dynamic/static use input-dependent/validation-compiled routes. MorphRoute parameter counts and latency include Fast-SCNN.

Model	Setting	Params (M) ↓	FLOPs (G) ↓	Latency (ms) ↓	SSIM ↑	PSNR (dB) ↑
EAMamba	KD	25.32	136.61	75.17	0.9893	52.46
MambaIRv1	KD	15.77	1289.69	371.91	0.9917	54.13
MambaIRv2	KD	22.66	1754.53	777.87	0.9934	54.37
MorphRoute-Mamba	SUP (dynamic)	21.11	279.03	71.82	0.9896	52.14
	+KD (dynamic)	21.11	279.03	71.82	0.9950	54.95
	+KD (static)	19.49	277.55	64.79	0.9948	54.74

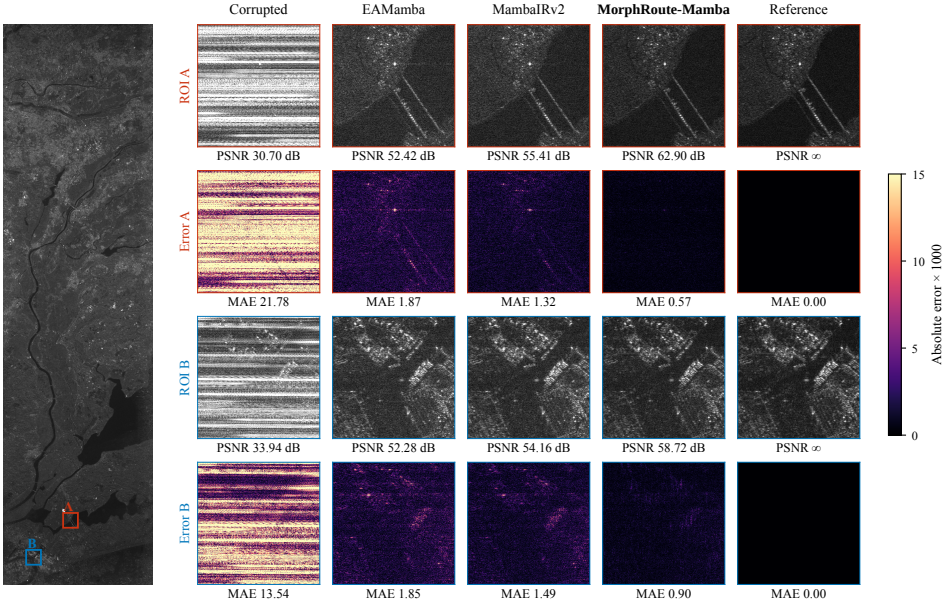
E.5 Image Comparisons and Local Analysis

Figures 14–19 compare local reconstructions with paired simulated targets. Across the displayed single and compound interference cases, MorphRoute-Mamba reduces residual bands while retaining scatterers and boundaries. Its error maps show weaker residuals around these structures; both ROIs in every case have higher local PSNR and lower MAE than EAMamba and MambaIRv2, consistent with the category-wise image results in Table 14.

Figure 20 extends the comparison to measured data. MPFI and UWI show less banding, SFI2 retains clearer local structure, and SFI2+SFTI preserves compact scatterers. These observations complement the AG/MNR gains in Table 2 on four unseen scenes, including two interference types absent from simulated training.

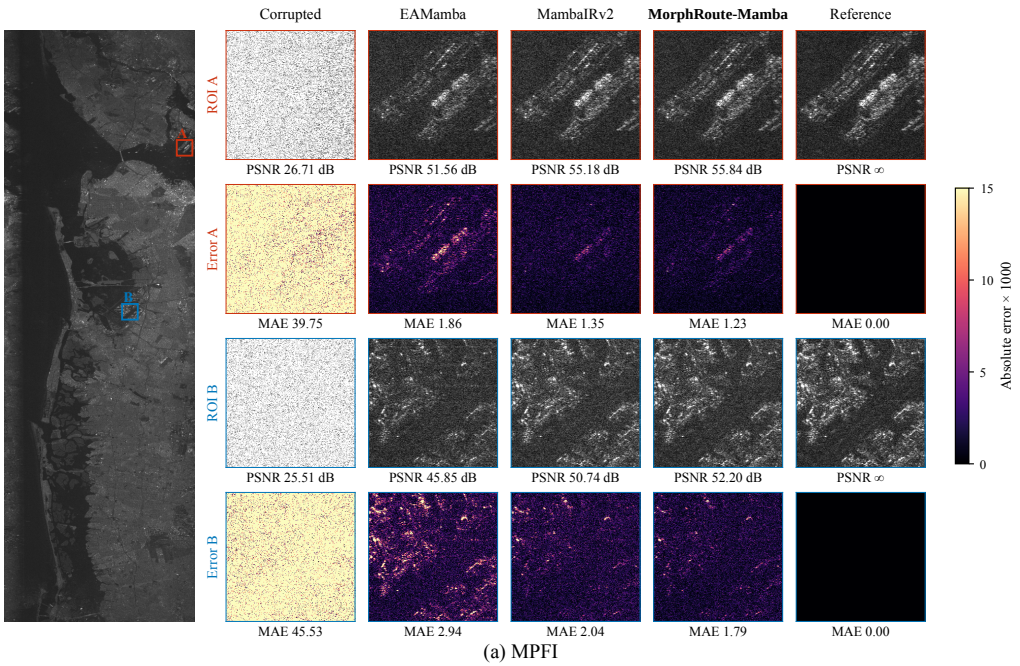


(a) DFTI

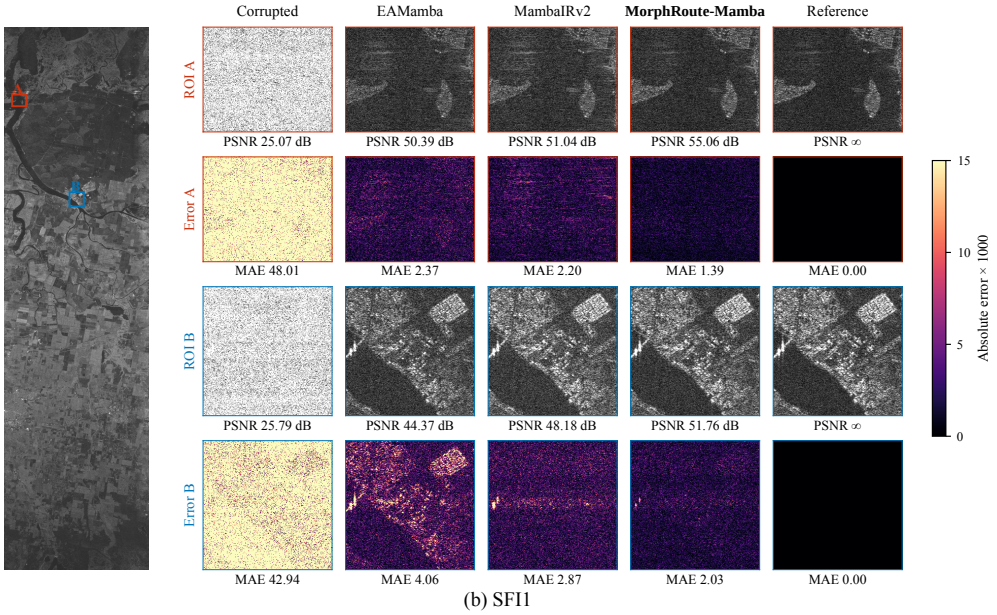


(b) ISFI

Figure 14: Simulated DFTI (a) and ISFI (b). Reference locators appear at left; each ROI row is followed by its error map.

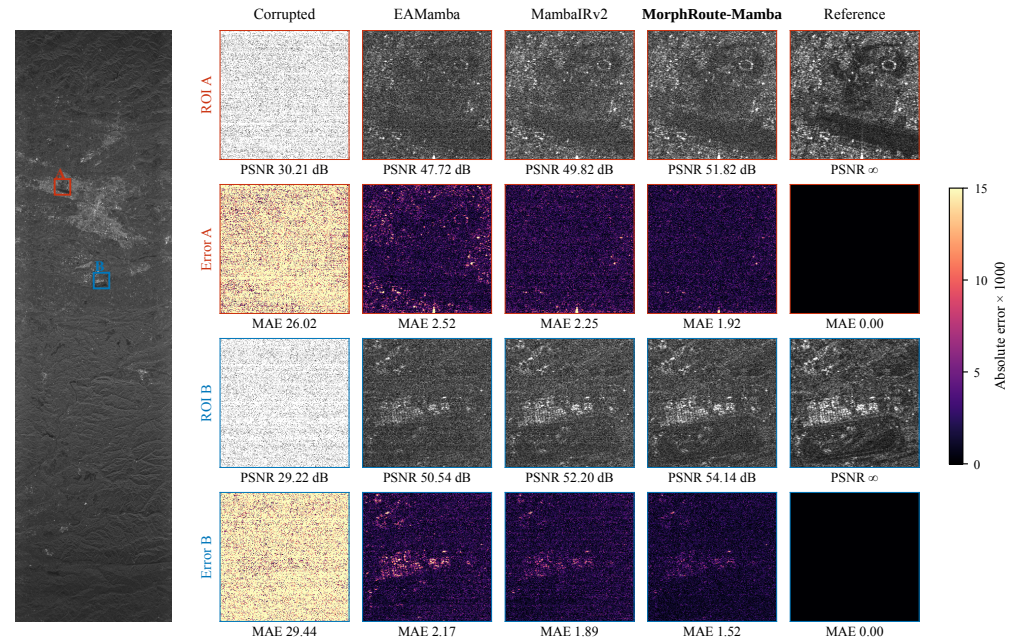


(a) MPFI

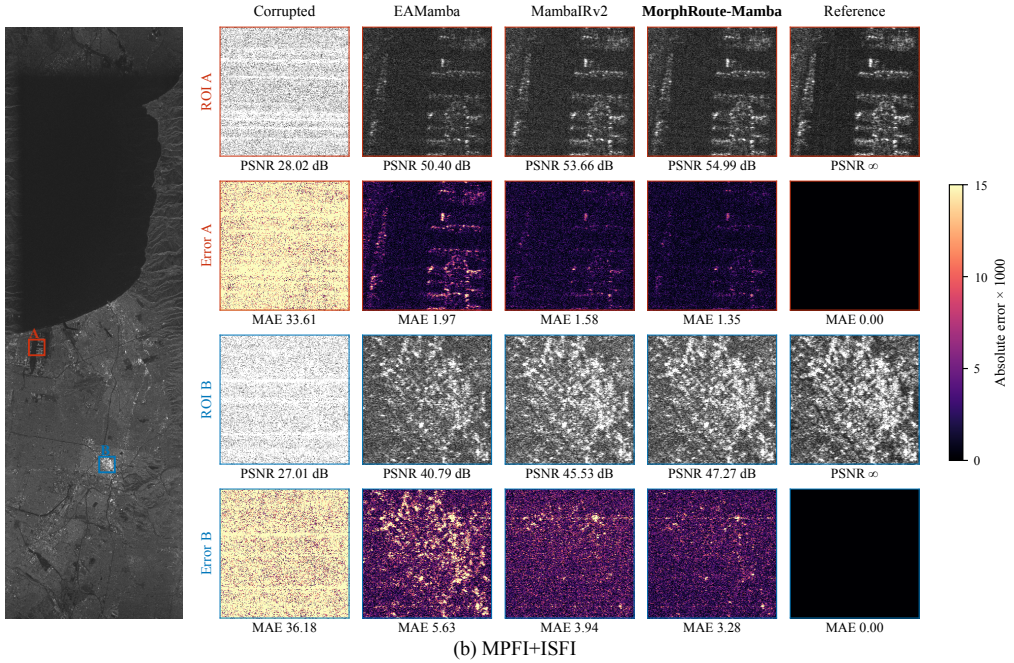


(b) SFI1

Figure 15: Simulated MPFI (a) and SFI1 (b), using the protocol of Figure 14.

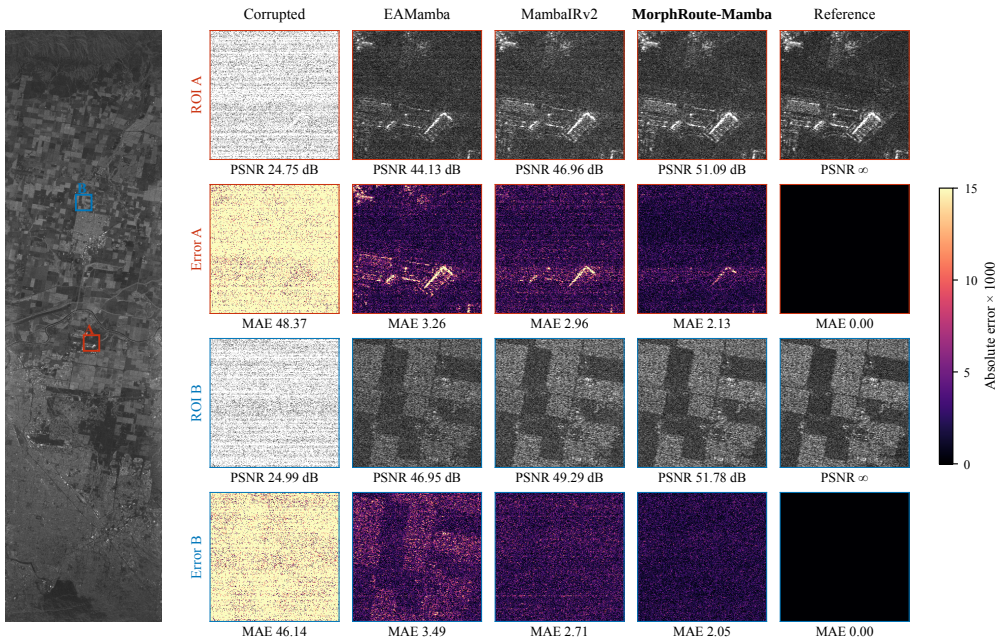


(a) MPFI+DFTI

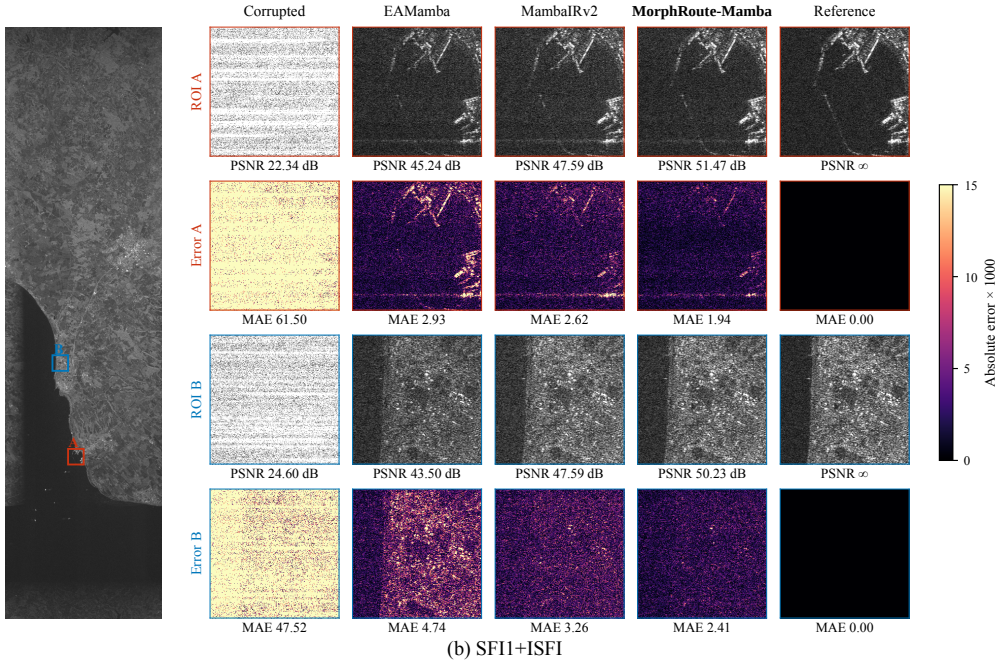


(b) MPFI+ISFI

Figure 16: Simulated MPFI+DFTI (a) and MPFI+ISFI (b), using the protocol of Figure 14.

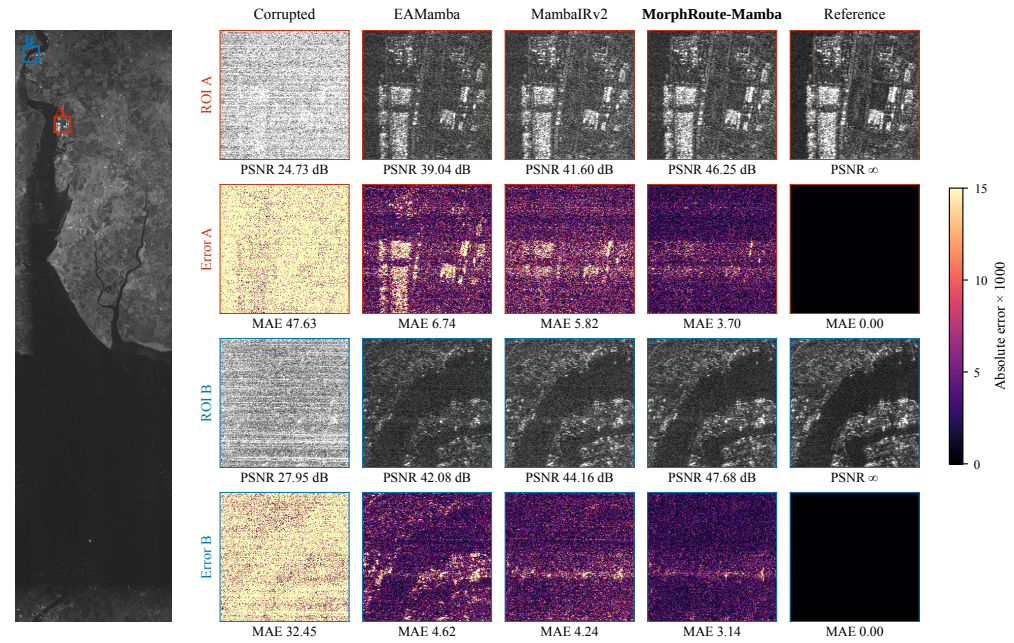


(a) SFI1+DFTI

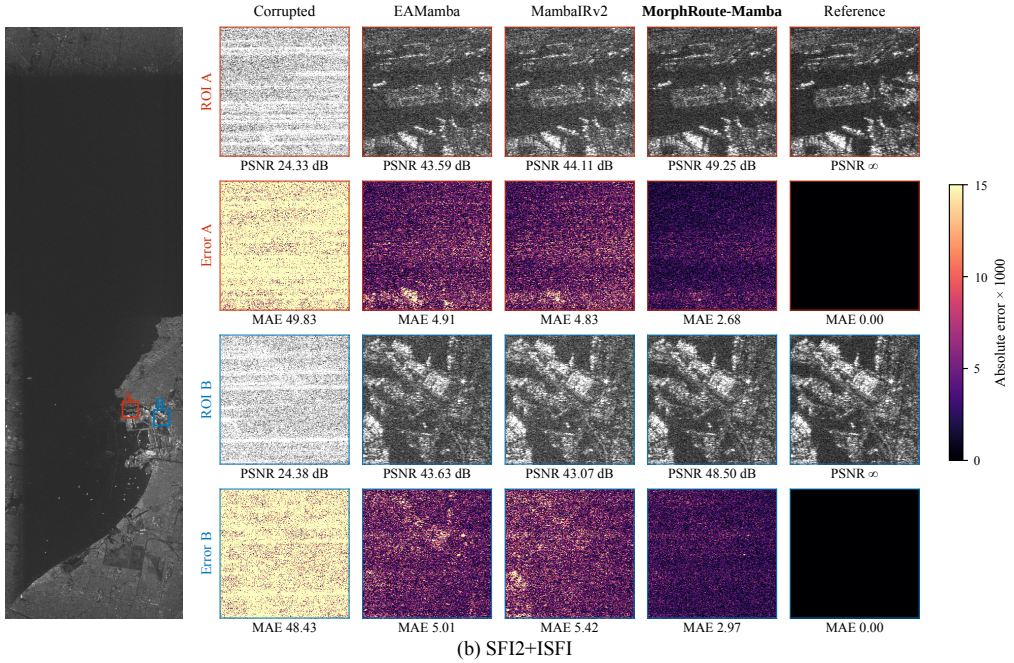


(b) SFI1+ISFI

Figure 17: Simulated SFI1+DFTI (a) and SFI1+ISFI (b), using the protocol of Figure 14.

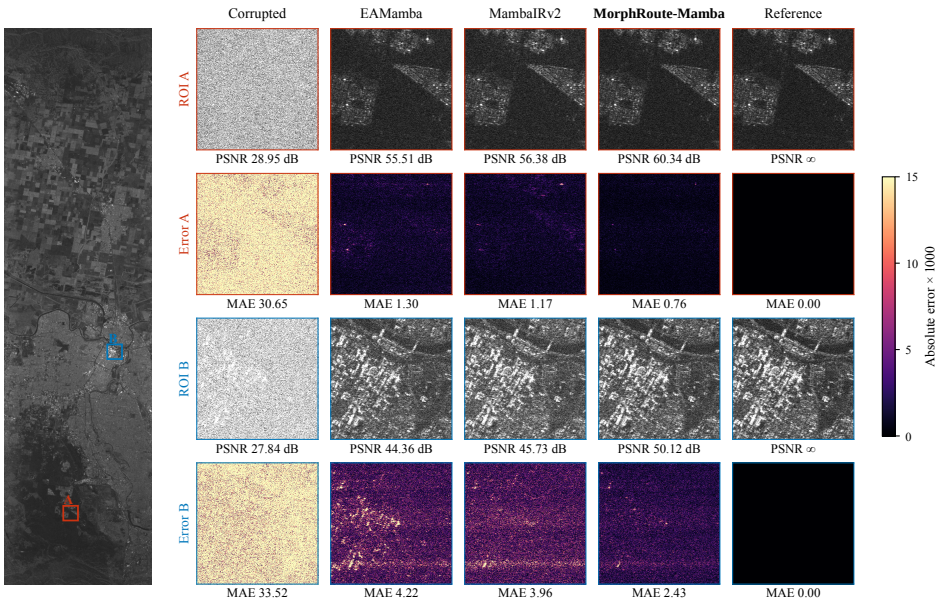


(a) SFI2+DFTI



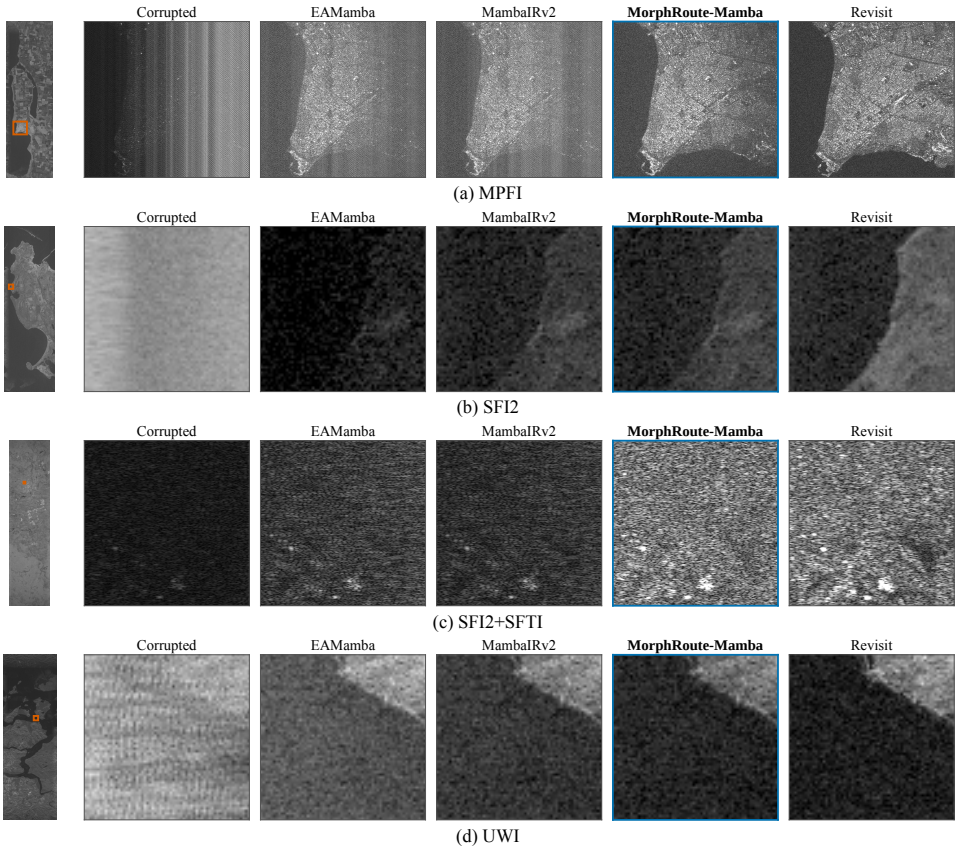
(b) SFI2+ISFI

Figure 18: Simulated SFI2+DFTI (a) and SFI2+ISFI (b), using the protocol of Figure 14.



(a) SFI2

Figure 19: Simulated SFI2, using the common ROI, metric, and display protocol of Figure 14.



(a) MPFI

(b) SFI2

(c) SFI2+SFTI

(d) UWI

Figure 20: Measured MPFI (a), SFI2 (b), SFI2+SFTI (c), and UWI (d). Each row shares coordinates and display processing; the translated revisit guides structural inspection.