GEOMETRY FIRST, SCATTERING WHEN RELIABLE: SPARSE-VIEW SAR TARGET GENERATION

Anonymous authors
Paper under double-blind review

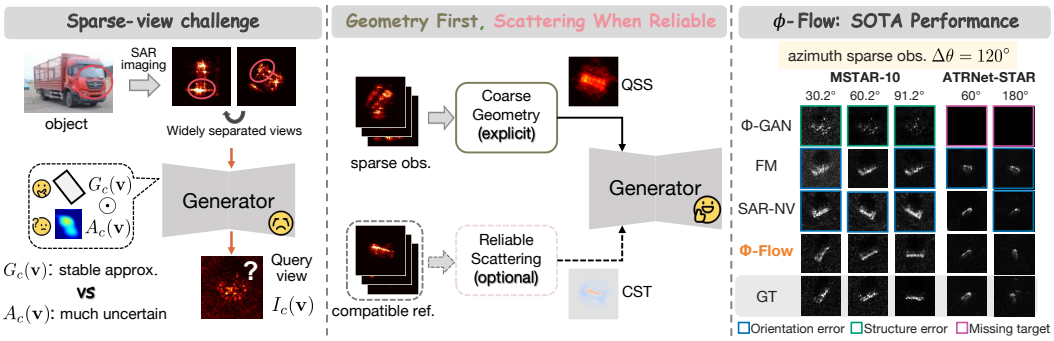


Figure 1: Sparse SAR observations leave unseen query views severely under-constrained, making it difficult for existing generators to infer both coarse geometry and view-dependent scattering. We therefore propose the "geometry first, scattering when reliable" principle and instantiate it with Φ-Flow, which explicitly constrains query-view geometry and incorporates scattering evidence only when reliable, improving generation under large angular gaps.

ABSTRACT

Sparse-view Synthetic Aperture Radar (SAR) undersamples the azimuth-depression space, making unseen-view generation severely under-constrained as view changes jointly alter coarse image-domain geometry and anisotropic electromagnetic scattering response. Yet coarse geometry admits a relatively stable cross-view approximation, whereas anisotropic scattering remains poorly determined across large angular gaps due to view-specific radar-target interactions. We therefore propose a **geometry-first, scattering-when-reliable principle**: coarse geometry is made explicit as a conditioning signal, while scattering information is introduced only when sufficient angular evidence supports its query-view variation. We instantiate this principle with Φ-Flow. For geometry, Φ-Flow represents coarse geometry as spatial support, constructing Canonical Spatial Support (CSS) from sparse observations and aligning it to the query view as Query-Aligned Spatial Support (QSS). For scattering, we first investigate which response information remains useful beyond QSS, and accordingly introduce Compatible Scattering Transfer (CST), which transfers relative, query-dependent scattering modulation from a compatible reference with denser angular coverage. QSS, optionally complemented by CST, conditions a flow-matching (FM) generator for unseen-view synthesis. Geometry ablations show that QSS provides query-aligned geometry that model scaling cannot recover. Scattering analyses reveal an inferability-utility trade-off: response information with greater complementary value beyond QSS is harder to infer from sparse observations. CST mitigates this limitation, consistently reducing azimuth error at unseen azimuths and even when the query depression is entirely unobserved for the target. Evaluated on MSTAR and ATRNet-STAR, Φ-Flow achieves state-of-the-art generation quality over all sparsity levels and the best recognition performance under the largest sparsity setting, with only 1.63% of the FM baseline's parameters.

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.

1 INTRODUCTION

Sparse-view Synthetic Aperture Radar (SAR) generation aims to synthesize unobserved views from only a few measured ones, reducing reliance on dense angular measurements and supporting recognition under unseen viewing conditions Huang et al. (2026; 2025); Zhang et al. (2023). This problem is particularly challenging in SAR because changes in viewing direction alter not only the target's coarse image-domain geometry, such as orientation, but also its anisotropic electromagnetic scattering response Datcu et al. (2023); Zhu et al. (2021); Guo et al. (2023). Local scattering responses may strengthen, weaken, appear, or disappear as the radar-target interaction changes. When observations are widely separated in the azimuth-depression space, large portions of this view-dependent variation remain unobserved, requiring the generator to infer both coarse geometry and local scattering behavior from limited angular evidence.

More broadly, novel-view synthesis from sparse observations in computer vision field faces similar ambiguity and has increasingly addressed it through generative priors Elata et al. (2025); Müller et al. (2024), structural guidance Müller et al. (2024), etc. SAR, however, introduces a distinct source of uncertainty because viewing changes affect not only image-domain geometry but also anisotropic electromagnetic scattering. We argue that coarse geometry and local scattering behavior require different degrees of generative inference under sparse observations. Coarse geometry is tied to the known viewing configuration and admits relatively stable cross-view approximations, allowing much of its query-view geometry to be specified explicitly rather than learned entirely from sparse observations. Anisotropic electromagnetic scattering, in contrast, depends on view-specific radar–target interactions and remains difficult to determine across large unobserved angular gaps. This suggests first reducing the avoidable geometric ambiguity, so that the generator can focus on the less constrained scattering variation and incorporate additional scattering information only when reliable angular evidence is available. **We summarize this principle as geometry first, scattering when reliable.**

Existing SAR generators Huang et al. (2026); Xiang et al. (2025); Sun et al. (2024); Han et al. (2026); Guan et al. (2026); Zhang et al. (2025) typically rely on angle codes, neighboring observations, or physical scattering constraints, which become increasingly insufficient as angular sampling gets sparser because query-view geometry remains implicit while reliable evidence for local scattering variation diminishes. To address this, we instantiate our principle with Φ-Flow through two mechanisms. Query-Aligned Spatial Support (QSS) realizes geometry first by explicitly representing coarse geometry as spatial support in the query frame, reducing the geometric burden on the generator. Compatible Scattering Transfer (CST) realizes scattering when reliable by transferring relative, query-dependent scattering modulation from a compatible reference, supplementing QSS with reliable scattering evidence. Accordingly, Φ-Flow jointly conditioned on QSS and CST, with CST included only when reliable scattering evidence is available.

We systematically validate the proposed principle through experiments. QSS provides query-view geometric information that cannot be recovered by simply scaling the generator, while scattering experiments reveal that additional response information is useful only when it is both reliable and complementary to QSS, with CST providing such evidence in controlled settings. Together, these findings translate into state-of-the-art sparse-view generation with a lightweight Φ-Flow, whose advantage becomes more pronounced under extreme view sparsity.

Contributions. 1) We propose a geometry-first, scattering-when-reliable principle for sparse-view SAR generation, treating coarse image-domain geometry and anisotropic electromagnetic scattering according to their different cross-view constraints. **2)** We instantiate this principle with Φ-Flow which represents geometry through Query-Aligned Spatial Support (QSS), and introduces Compatible Scattering Transfer (CST) to provide relative query-dependent scattering modulation from a compatible reference. **3)** Experiments reveal that QSS cannot be replaced by model scaling and demonstrate reliable scattering transfer via CST provides complementary evidence beyond QSS. These findings lead to SOTA sparse-view generation performance with a lightweight generator.

2 RELATED WORK

Novel-view synthesis from sparse observations. Novel-view synthesis from limited observations is inherently under-constrained, and prior work has addressed this ambiguity through generative

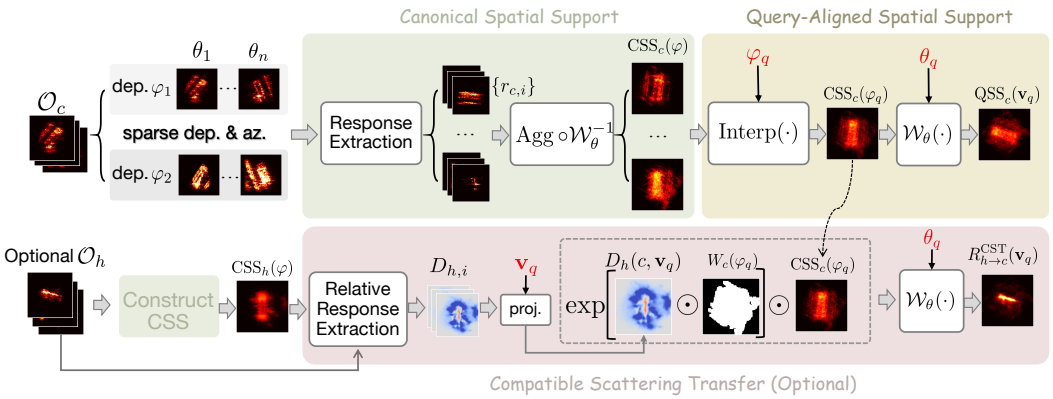


Figure 2: **Instantiation of the geometry-first, scattering-when-reliable principle in Φ-Flow.** Sparse observations are inverse-aligned and aggregated into Canonical Spatial Support (CSS), then aligned to the query view as Query-Aligned Spatial Support (QSS). When a compatible reference provides reliable angular evidence, CST adds relative query-dependent response modulation. QSS, optionally augmented by CST, conditions unseen-view generation.

priors Elata et al. (2025); Müller et al. (2024), structural guidance Müller et al. (2024), and reference-based conditioning Cao et al. (2024). These studies demonstrate the value of combining explicit geometric information with generative inference under sparse observations. SAR shares this sparse-view ambiguity but introduces additional uncertainty from anisotropic electromagnetic scattering. Existing SAR generators control target views through angle- or pose-based conditioning Oh & Kim (2021); Sun et al. (2023), neighboring-view or image-conditioned synthesis Wang et al. (2022), and SAR-specific physical, scattering, or limited-data constraints Han et al. (2026); Sun et al. (2023). Existing approaches improve controllable SAR synthesis and augmentation when neighboring views are informative, but under extreme sparsity, low-dimensional view conditioning leaves query-view geometry implicit and distant observations provide little evidence for local scattering variation. We address this regime by explicitly constraining reliably recoverable query-view geometry and adding scattering information only when supported by sufficient angular evidence.

Geometric representations. Recent work have shown that explicit spatial information can constrain SAR view synthesis, ranging from image-domain contours/support Xiang et al. (2025); Sun et al. (2024) to full 3D or differentiable representations Lei et al. (2024); Liu et al. (2026b). However, existing geometric representations often come with stronger data or modeling requirements Wu et al. (2026); Sun et al. (2024). It remains less explored what reliable geometry can be extracted from sparse SAR observations themselves, and how much explicitly providing such geometry can benefit unseen-view generation.

Scattering modeling. It aims to characterize how local electromagnetic responses vary with viewing direction, providing query-view response cues for physically plausible SAR synthesis. Existing methods model dominant scattering centers Han et al. (2026); Guan et al. (2026); Zhang et al. (2025), response amplitude Guan et al. (2026), visibility Zhang et al. (2026), or their view-dependent evolution Jiang et al. (2026) to improve local response fidelity at the requested view. Their effectiveness, however, depends on whether the available angular observations provide sufficient evidence. Across large angular gaps, scattering-center correspondence and local response evolution become increasingly uncertain.

3 Φ-Flow

3.1 Sparse-View Formulation

For the stationary ground-target SAR setting considered in this work, let $\mathbf{v} = (\theta, \varphi)$ denote a viewing direction, where θ and φ are the target aspect (azimuth) and depression angles, respectively. The sparse observations of target c are $\mathcal{O}_c = \{(I_{c,i}, \mathbf{v}_{c,i})\}_{i=1}^{N_c}$, where $I_{c,i}$ is the measured SAR image at view $\mathbf{v}_{c,i}$. Given an unobserved query view $\mathbf{v}_q = (\theta_q, \varphi_q)$, our goal is to synthesize $I_c(\mathbf{v}_q)$.

Sparse-view SAR generation is under-constrained because a change in viewing direction affects both coarse image-domain geometry and anisotropic electromagnetic scattering response. We conceptually write

$$I_c(\mathbf{v}) \approx G_c(\mathbf{v}) \odot A_c(\mathbf{v}) + C_c(\mathbf{v}), \tag{1}$$

where G_c describes coarse image-domain geometry, A_c the view-dependent scattering response, and C_c residual image-formation effects. Under sparse observations, $G_c(\mathbf{v}_q)$ remains relatively constrained by the known viewing configuration, whereas $A_c(\mathbf{v}_q)$ becomes increasingly uncertain across large angular gaps. This motivates our *geometry-first, scattering-when-reliable* principle: query-view geometry is made explicit, while scattering is primarily learned by the generator and supplemented only when reliable angular evidence is available.

3.2 Geometry First: Canonical and Query-Aligned Spatial Support

We represent the query-view coarse geometry $G_c(\mathbf{v}_q)$ through Query-Aligned Spatial Support (QSS), a soft spatial-support map in the query frame. QSS is constructed from sparse observations through Canonical Spatial Support (CSS), as demonstrated in Figure 2. Specifically, Let $r_{c,i}$ denote the response map extracted from the observed SAR image $I_{c,i}$, associated with observed view $\mathbf{v}_{c,i} = (\theta_{c,i}, \varphi_{c,i})$. Each $r_{c,i}$ is inverse-aligned in azimuth, and those on the same depression plane are aggregated into the CSS:

$$\bar{r}_{c,i} = \mathcal{W}_{\theta_{c,i}}^{-1}[r_{c,i}], \qquad \mathrm{CSS}_c(\varphi) = \mathrm{Agg}(\{\bar{r}_{c,i} : \varphi_{c,i} = \varphi\})\,. \tag{2}$$

Here, $\mathcal{W}_\theta$ denotes image-plane rotation, while $\mathrm{Agg}(\cdot)$ denotes aggregation over the aligned observations. How $r_{c,i}$ and $\mathrm{Agg}(\cdot)$ are implemented for each benchmark are detailed in the Appendix B.1.

The above construction defines the CSS at each observed depression angle. For a query depression φ_q that lies between two observed depression angles φ_1 and φ_2, we obtain its spatial support by linear interpolation:

$$\mathrm{CSS}_c(\varphi_q) = (1-\lambda)\mathrm{CSS}_c(\varphi_1) + \lambda \mathrm{CSS}_c(\varphi_2), \qquad \lambda = \frac{\varphi_q - \varphi_1}{\varphi_2 - \varphi_1}. \tag{3}$$

The support is then aligned to the query azimuth to form the QSS:

$$\mathrm{QSS}_c(\mathbf{v}_q) = \mathcal{W}_{\theta_q}[\mathrm{CSS}_c(\varphi_q)]\,. \tag{4}$$

3.3 Scattering When Reliable: Compatible Scattering Transfer

For the scattering term $A_c(\mathbf{v}_q)$, sparse target observations provide much weaker cross-view constraints, so its query-dependent variation is left to the generator unless reliable additional evidence is available. To realize *scattering when reliable*, we propose *Compatible Scattering Transfer* (CST), which introduces relative, query-dependent scattering modulation from a compatible reference (target h) with denser angular coverage.

As illustrated in Figure 2, CST proceeds in two steps: extracting the reference's relative response changes $D_{h,i}$ (equation 5), and applying them to the target's spatial support to construct the CST response $R_{h\to c}^{\mathrm{CST}}(\mathbf{v}_q)$ (equation 6).

We first aggregate the inverse-aligned reference maps at each measured depression φ into $\mathrm{CSS}_h(\varphi)$ following Sec. 3.2. For each sample i in compatible reference h with the depression angle φ, its relative response change is

$$D_{h,i} \triangleq \log \frac{\bar{r}_{h,i} + \epsilon_0}{\mathrm{CSS}_h(\varphi) + \epsilon_0}, \tag{5}$$

where $\epsilon_0 > 0$ is a small stabilizing constant. Positive and negative values indicate responses stronger or weaker than the reference's azimuth-aggregated response, respectively.

These angle-indexed changes define $D_h(c, \mathbf{v})$, the reference's relative response change at view $\mathbf{v}$, registered to target c's canonical coordinates. At the query view $\mathbf{v}_q = (\theta_q, \varphi_q)$, CST modulates the target's CSS before query alignment:

$$R_{h\to c}^{\mathrm{CST}}(\mathbf{v}_q) \triangleq \mathcal{W}_{\theta_q}\left[\mathrm{CSS}_c(\varphi_q) \odot \exp(W_c(\varphi_q) \odot D_h(c, \mathbf{v}_q))\right], \tag{6}$$

where $W_c(\varphi_q)$ is a spatial weight map that controls the modulation strength and restricts it to the target support. The target c supplies the spatial support, while the reference h contributes relative, view-dependent strengthening or weakening of responses within that support. Implementation details, including reference-view interpolation, spatial filtering, clipping, and strength calibration, are provided in Appendix B.2.

3.4 Query-View Generation

Having constructed QSS and, when available, CST, we instantiate Φ-Flow by using them as spatial conditions for a conditional flow-matching generator. QSS always anchors generation to the target-derived query-view geometry, whereas CST is introduced only as an optional residual that captures additional query-dependent scattering variation. Accordingly, for a view $\mathbf{v}$, we define

$$Z_c(\mathbf{v}) = [\mathrm{QSS}_c(\mathbf{v}), \mathbf{0}], \qquad Z^{\mathrm{CST}}_{h\to c}(\mathbf{v}) = [\mathrm{QSS}_c(\mathbf{v}), R^{\mathrm{CST}}_{h\to c}(\mathbf{v}) - \mathrm{QSS}_c(\mathbf{v})]. \tag{7}$$

The first channel provides query-aligned spatial support, while the second encodes the relative response modulation transferred by CST. Class c, view $\mathbf{v}$, and time t are embedded separately.

During training, let $I_{c,i} = I_c(\mathbf{v}_i)$ be a measured training image at observed view $\mathbf{v}_i$. Given Gaussian noise x_0 and $t \sim \mathcal{U}(0,1)$, we set $x_t = (1-t)x_0 + tI_{c,i}$ and optimize

$$\mathcal{L}_{\mathrm{FM}} = \mathbb{E}\Big[\|u_\psi(x_t, t, Z(\mathbf{v}_i), c, \mathbf{v}_i) - (I_{c,i} - x_0)\|_2^2\Big]. \tag{8}$$

At inference, the condition is constructed at the requested unseen view $\mathbf{v}_q$, and the learned flow is integrated from Gaussian noise to synthesize $I_c(\mathbf{v}_q)$.

4 Experiments

We evaluate Φ-Flow on MSTAR Diemunsch & Wissinger (1998) and ATRNet-STAR Liu et al. (2026a) under sparse azimuth sampling and unobserved-depression settings. Complete dataset descriptions and sparse-view protocols are provided in Appendix A. Our analyses focus on three questions: 1) whether explicit query-view geometry can be replaced by model capacity, 2) what information makes QSS effective, and 3) when additional scattering evidence provides complementary value. We further compare Φ-Flow with representative SAR generative models across these sparse-view regimes. We report generation quality, geometric consistency, and downstream task performance, with metric definitions provided in Appendix B.4.

4.1 Can Model Capacity Replace QSS?

We first conduct a controlled capacity ablation to ask whether increasing generator capacity can compensate for the absence of explicit query-view geometry. We vary the U-Net base width b from 16 to 128 (denoted as `b16`, `b64`, and `b128`) and compare models with (`w/`) and without (`w/o`) QSS under identical data and optimization settings across six MSTAR sparse-view protocols, with azimuth spacings $\Delta\theta \in \{30^\circ, 45^\circ, 60^\circ, 72^\circ, 90^\circ, 120^\circ\}$. The results are recorded as the average on the six sparse protocols. We evaluate support IoU (sIoU) and orientation error as geometric metrics, with their definitions provided in Appendix B.4.

Table 1a and Figure 3a show a clear separation between the two model families (`w/` and `w/o` QSS) across all sparsity levels. Once QSS is provided, further capacity scaling produces only minor changes in sIoU and orientation error, indicating that model capacity and explicit spatial information address different sources of difficulty. Larger generators cannot reliably recover the missing query-view geometry from sparse observations, whereas QSS makes this geometry explicit, directly supporting our geometry-first principle and allowing the generator to focus on the remaining scattering and appearance variation.

4.2 What Makes QSS Effective?

Having established that model capacity cannot replace QSS, we next perform an internal ablation to identify which components of QSS drives its benefit. All variants in Table 1b use the same `b64` architecture and training budget. Starting from the no-QSS baseline, we progressively introduce:

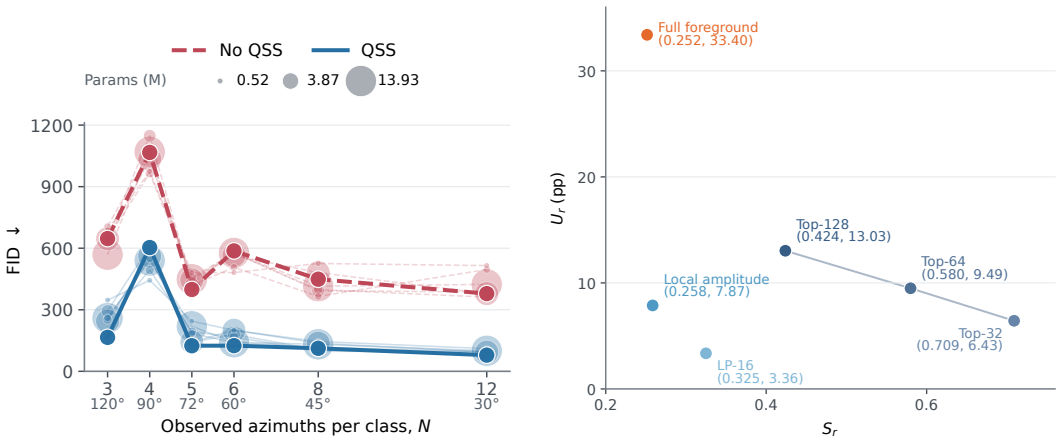
(a) Capacity comparison.

Condition	Params (M)	FID ↓	sIoU ↑	Orient. err. ↓
w/o QSS (b128)	13.929	583.82	0.011	39.11°
w/ QSS (b16)	0.519	248.48	0.398	8.90°
w/ QSS (b64)	3.867	**201.62**	0.422	8.58°
w/ QSS (b128)	13.929	232.07	**0.424**	**8.55°**

(b) Spatial conditioning at b64.

Condition	FID ↓	sIoU ↑	Orient. err. ↓	ATR (%) ↑
w/o QSS	587.87	0.011	39.74°	34.80
CSS	209.82	0.348	19.08°	47.41
Aligned mask	203.98	0.421	**8.53°**	56.97
w/ QSS	**201.62**	**0.422**	8.58°	**58.15**

Table 1: Capacity and spatial-conditioning ablations averaged over six sparse-view protocols.



(a) (b)

Figure 3: (a) Increasing model capacity does not close the gap caused by removing QSS; bubble area represents parameter count, and thick trajectories denote b64. (b) Responses with greater generation utility (U_r ↑) appear harder to infer (S_r ↓) from sparse target views under the tested angular rules.

1) CSS, which provides class-specific spatial support in the canonical frame; 2) an aligned support mask, which preserves the query-aligned support while suppressing the response information in QSS; 3) the full QSS, , which retains both query alignment and the aggregated response amplitude. This design separates the contributions of spatial support, query alignment, and response information in QSS.

The ablation in Table 1b shows that CSS is already beneficial even without query alignment, but aligning it to the requested view produces the dominant improvement in geometric consistency. The aligned support mask performs nearly identically to full QSS in support IoU and orientation error, indicating that most of QSS's geometric benefit comes from placing the target support in the correct query frame. Retaining the response amplitude provides a smaller additional improvement in generation quality and downstream recognition. Therefore, the effectiveness of QSS is driven primarily by explicit query-aligned geometry, rather than by the amplitude information contained in the support map.

4.3 When Does Scattering Help Beyond QSS?

4.3.1 Which Scattering Responses Are Worth Modeling Beyond QSS?

QSS provides coarse spatial support and orientation but does not explicitly model view-dependent scattering. We therefore ask: which response information can be inferred from sparse observations beyond QSS (inferability), and how much it could improve generation if accurately available (utility). To probe this *"inferability vs utility"*, we examine six representations spanning coarse response variation to full foreground details indexed by r: a low-frequency response (LP-16), local amplitude variation on fixed support, localized dominant responses (Top-32, Top-64, and Top-128), and the full foreground response. All comparisons are based on QSS-only Φ-Flow. (Details see Appendix D.1)

	Φ-Flow	Φ -Flow$_{\mathrm{CST}}$
	$\Delta\theta$ =30°	
Acc. ↑	**83.73 ± 1.16**	83.02 ± 0.96
Az. MAE ↓	25.51 ± 0.96	**18.67 ± 0.80**
	$\Delta\theta$ =60°	
Acc. ↑	73.47 ± 2.31	**73.55 ± 0.84**
Az. MAE ↓	35.05 ± 1.51	**29.05 ± 1.19**

Acc. in %; Az. MAE in degrees.

Table 2: CST at unseen azimuths.

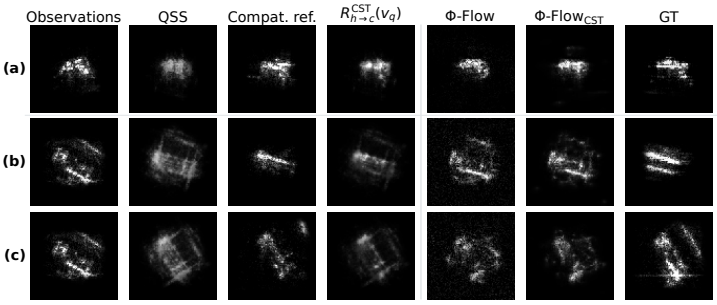


Figure 4: CST examples at unseen azimuths. (a) 8228-5 → ZL40F at θ_q =90°; (b) J6P → KR230 at θ_q =285°; and (c) J6P → KR230 at θ_q =325°.

First, to assess which response representations can be inferred from sparse target views more faithfully than QSS alone, we define $S_r = 1 - \left\langle \mathrm{MAE}(\widehat{R}_r, R_r)/\,\mathrm{MAE}(\mathrm{CSS}_c, R_r) \right\rangle$. Here, R_r denotes the measured query representation, and $\widehat{R}_r$ is inferred from sparse observations using simple angular rules. S_r is calculated in the canonical frame before query-alignment. Thus, S_r indicates how much additional query-dependent response variation can be inferred from sparse target observations.

Second, assuming that such a response representation could be faithfully inferred at the query view, we ask whether it would provide complementary value beyond QSS for downstream tasks. To isolate this potential utility from inference error, the oracle experiment directly supplies the measured query-view representation and defines U_r as the resulting ATR gain over QSS-only Φ-Flow. Details for this oracle experiment are provided in Appendix D.2.

Figure 3b jointly compares these two dimensions. Top-32 is relatively easy to infer from sparse observations, but provides only limited additional downstream utility beyond QSS. Within the Top-K family, retaining more response detail (from K=32 to 128) consistently increases potential downstream utility while reducing inferability from sparse observations. The full foreground response further extends this trend, providing the largest complementary utility while is the most difficult to infer from sparse observation. These results suggest that the response information more valuable beyond QSS is also harder to infer from sparse observations using the tested angular rules. These findings motivate introducing additional angular evidence beyond the sparse target observations. CST realizes this idea by drawing query-dependent response variation from a compatible, more densely observed reference while retaining the target-derived QSS.

4.3.2 CST Provides Viewpoint Evidence at Unseen Azimuths

We evaluate CST on four ATRNet-STAR vehicle categories grouped into two reciprocal target–reference pairs: FAW J6P with Dongfeng Tianjin KR230, and Changlin 8228-5 with SDLG ZL40F. The paired categories have similar cross-view scattering-response behavior, with each serving in turn as target c and compatible reference h. At 30° depression, target views are sampled at 30° or 60° azimuth intervals, while the compatible reference retains its available measurements on the native 5° grid. Data and evaluation details are given in the Appendix A.2.

Across both azimuth protocols in Table 2, CST consistently reduces azimuth error, while classification accuracy changes little. This indicates that the transferred cross-view response variation primarily contributes complementary viewpoint-sensitive evidence. This is consistent with the intended separation between spatial support and scattering response. QSS preserves the target's coarse spatial information, whereas CST adds query-dependent scattering modulation from the compatible reference. Figure 4 shows that, under the same QSS, CST alters the dominant local response layout toward the measured query view. Together, these results indicate that scattering information that is difficult to infer from sparse target views can still be useful when reliable angular evidence is provided by a compatible reference.

Table 3: Four-class CST evaluation on the withheld 30° target-depression plane. $\Delta\theta$ denotes the azimuth sampling interval of the target training images at each observed depression angle (15° and 45°). bold marks the better mean within each model pair.

$\Delta\theta$	Model	Acc. (%) ↑	Az. MAE (°) ↓	Dep. MAE (°) ↓
30°	Φ-Flow	75.50 ± 1.06	39.66 ± 1.38	7.39 ± 1.81
	Φ-Flow$_{\mathrm{CST}}$	**76.68 ± 3.22**	**33.42 ± 1.33**	**6.91 ± 1.00**
60°	Φ-Flow	**73.80 ± 2.20**	53.64 ± 2.72	5.68 ± 1.24
	Φ-Flow$_{\mathrm{CST}}$	73.58 ± 3.69	**34.15 ± 4.31**	**5.39 ± 0.74**

Table 4: MSTAR-10 comparison under azimuth sampling intervals of 30°, 60°, and 120°. Results are mean ± standard deviation over five seeds; bold marks the best mean.

		30°			60°			120°		
Method	Para. (M)	FID ↓	Acc. ↑	Az. MSE ↓	FID ↓	Acc. ↑	Az. MSE ↓	FID ↓	Acc. ↑	Az. MSE ↓
Sparse only	–	–	67.84 ± 3.49	0.244 ± 0.015	–	37.35 ± 5.12	0.452 ± 0.043	–	29.41 ± 2.49	0.681 ± 0.031
Repeated sparse	–	–	74.87 ± 1.59	0.224 ± 0.020	–	44.22 ± 1.94	0.463 ± 0.018	–	34.39 ± 1.64	0.748 ± 0.002
ACGAN	12.45	416.83	65.34 ± 4.88	0.266 ± 0.013	678.13	44.38 ± 3.01	0.526 ± 0.018	1833.52	18.62 ± 10.13	0.523 ± 0.031
StyleGAN	23.22	199.83	79.60 ± 0.87	0.176 ± 0.009	246.14	45.82 ± 3.15	0.459 ± 0.013	394.93	32.47 ± 1.45	0.696 ± 0.020
Φ-GAN	12.45	160.41	80.07 ± 0.40	0.203 ± 0.019	313.57	46.12 ± 2.51	0.472 ± 0.035	487.86	31.22 ± 2.16	0.722 ± 0.025
cDDPM	107.90	152.25	79.81 ± 2.86	0.175 ± 0.017	257.00	45.52 ± 2.37	0.452 ± 0.039	298.89	31.07 ± 0.91	0.742 ± 0.039
LDM	234.47	98.82	56.59 ± 1.79	0.266 ± 0.021	280.17	43.00 ± 0.85	0.572 ± 0.022	304.54	32.99 ± 2.31	0.733 ± 0.045
FM	237.26	105.87	**80.75 ± 2.85**	0.194 ± 0.015	193.72	44.78 ± 1.60	0.520 ± 0.028	231.02	32.82 ± 2.77	0.741 ± 0.056
SAR-NV	1286.34	184.51	51.22 ± 2.43	0.353 ± 0.015	766.61	43.36 ± 3.97	0.423 ± 0.004	456.73	32.08 ± 2.80	0.647 ± 0.022
Φ-Flow	3.87	**78.90**	80.23 ± 1.06	**0.175 ± 0.021**	**125.15**	**61.06 ± 1.15**	**0.335 ± 0.017**	**165.49**	**45.25 ± 2.40**	**0.479 ± 0.029**

4.3.3 CST Transfers Response Across an Unobserved Depression Plane

We further evaluate CST when the query depression is entirely unobserved for the target, where the target's 30° depression sequence is withheld. Its QSS is obtained from the 15° and 45° measurements following Sec. 3.2. CST uses the compatible reference's measured response at 30° depression, with circular interpolation only for missing azimuths. All real target images at 30° are excluded from conditioning-input construction and FM training.

Table 3 shows that the most evident CST benefit appears in azimuth estimation under the 30° and 60° target sampling settings, while classification and depression error change only modestly. This extends the evaluation to an unseen target depression: QSS provides spatial support through interpolation of the target's observed depression planes, while the compatible reference supplies additional azimuth-dependent scattering evidence at the query depression.

4.4 Comparison with Existing SAR Generative Models

We compare Φ-Flow with ACGAN Odena et al. (2017), StyleGAN Karras et al. (2021), Φ-GAN Zhang et al. (2025), conditional DDPM Dhariwal & Nichol (2021), LDM Rombach et al. (2022), SAR-NV Rahimi & Yu (2026), and conditional FM Lee et al. (2024) on MSTAR and ATRNet-STAR. All methods use the same sparse observations and synthetic-data budget within each protocol, with complete data splits and training settings provided in Appendix A.

Tables 4 and 5 show that Φ-Flow consistently improves sparse-view generation across datasets and sparsity levels. On MSTAR, it achieves the best FID at all evaluated azimuth intervals and becomes increasingly advantageous in downstream recognition as angular sampling becomes sparser. The same trend holds on ATRNet-STAR, where Φ-Flow consistently yields the highest recognition accuracy under sparse azimuth sampling and on the completely withheld depression plane. These results indicate that explicitly providing query-view geometry is particularly beneficial when the available angular evidence becomes limited.

The qualitative results in Figure 5 further illustrate this advantage under large angular gaps. Competing methods tend to produce distorted geometry or remain close to observed-view appearances at

Table 5: ATRNet-STAR comparison under azimuth sampling intervals of 60°, 90°, and 120°, with additional depression interpolation from 15° and 45° to 30°. Results are mean ± standard deviation over five seeds; bold marks the best mean.

	Sparse azimuth						Depression interpolation	
	60°		90°		120°		15°,45° → 30°	
Method	Acc. (%) ↑	Az. MSE ↓	Acc. (%) ↑	Az. MSE ↓	Acc. (%) ↑	Az. MSE ↓	Acc. (%) ↑	Az. MSE ↓
Sparse only	45.16 ± 1.19	0.596 ± 0.039	36.68 ± 3.21	0.547 ± 0.035	47.78 ± 1.10	0.697 ± 0.041	55.50 ± 5.26	0.503 ± 0.027
Repeated sparse	44.38 ± 3.27	0.565 ± 0.016	36.21 ± 1.74	0.581 ± 0.022	41.76 ± 1.73	0.789 ± 0.028	56.17 ± 3.52	0.575 ± 0.052
ACGAN	36.36 ± 1.30	0.661 ± 0.032	30.57 ± 1.43	0.655 ± 0.033	9.84 ± 2.60	**0.500 ± 0.000**	50.33 ± 3.61	0.573 ± 0.032
StyleGAN	42.73 ± 2.06	0.590 ± 0.032	35.95 ± 1.29	0.625 ± 0.033	30.12 ± 1.91	0.756 ± 0.035	50.83 ± 5.14	0.588 ± 0.017
Φ-GAN	40.28 ± 2.60	0.591 ± 0.032	30.62 ± 2.50	0.609 ± 0.040	15.47 ± 13.73	0.514 ± 0.032	56.17 ± 4.27	0.579 ± 0.047
cDDPM	38.79 ± 2.45	0.569 ± 0.032	25.29 ± 3.23	0.715 ± 0.028	29.25 ± 1.56	0.843 ± 0.035	57.50 ± 4.68	0.572 ± 0.049
LDM	33.17 ± 1.93	0.603 ± 0.021	30.34 ± 3.51	0.599 ± 0.023	27.11 ± 2.49	0.638 ± 0.007	51.17 ± 6.63	0.581 ± 0.044
FM	40.65 ± 1.50	0.648 ± 0.018	31.42 ± 1.50	0.615 ± 0.034	28.47 ± 3.75	0.669 ± 0.038	49.67 ± 0.95	0.602 ± 0.015
SAR-NV	29.62 ± 2.91	0.773 ± 0.032	26.22 ± 1.28	0.761 ± 0.033	39.61 ± 1.66	0.724 ± 0.017	55.50 ± 4.23	0.721 ± 0.047
Φ-Flow	**57.68 ± 2.27**	**0.524 ± 0.028**	**49.86 ± 3.26**	**0.509 ± 0.021**	**50.16 ± 3.20**	0.604 ± 0.035	**74.33 ± 3.20**	**0.487 ± 0.037**

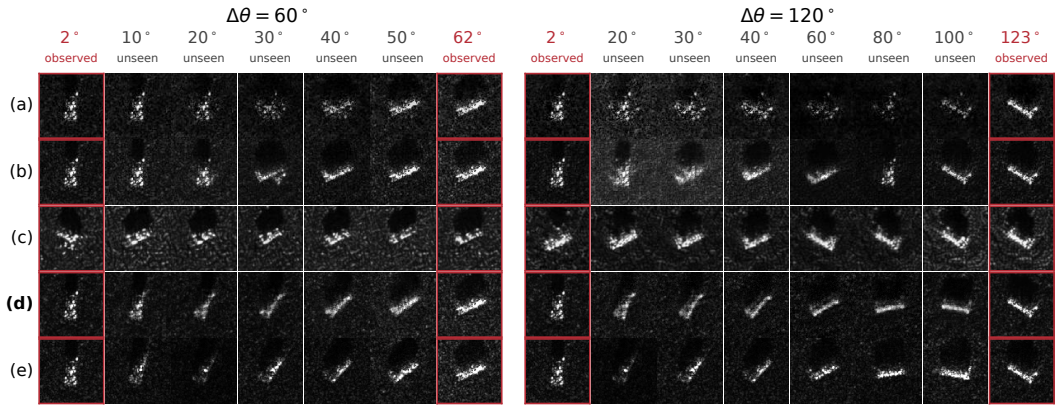


Figure 5: Qualitative comparison on MSTAR at $\Delta\theta = 60°$ (left) and $120°$ (right). Rows: (a) Φ-GAN, (b) FM, (c) SAR-NV, (d) ours, and (e) ground truth. Red frames mark observed views; others are unseen. All images share the same grayscale range.

unseen azimuths, whereas Φ-Flow better follows the requested orientation while maintaining coherent target responses. Notably, these gains are achieved with fewer than 4M inference parameters.

5 Conclusion

We propose a geometry-first, scattering-when-reliable principle for sparse-view SAR generation and instantiate it with Φ-Flow, which makes query-view geometry explicit through QSS and introduces additional reliable scattering information through optional CST. Experiments reveal two key findings: explicit query-aligned geometry cannot be replaced by model scaling, and query-dependent scattering variation with greater complementary value beyond QSS tends to be harder to infer from sparse target observations. The proposed CST provides a practical way to exploit such information when reliable angular evidence is available, improving viewpoint fidelity at unseen views. Overall, Φ-Flow achieves SOTA performance with a lightweight model, supporting the principle of making reliably constrained information explicit while leaving weakly constrained variation to the generator.

Limitation and future work. Our validation is restricted to vehicle targets. Whether the proposed principle and Φ-Flow generalize to other stationary targets, such as aircraft, whose sparse-view geometry and scattering variation may follow different patterns, remains to be investigated.

AI USE STATEMENT

Generative AI tools (ChatGPT and Codex) were used for coding assistance, debugging, literature search, and language editing. All reported numerical results were obtained from the authors' own experimental pipelines. The authors reviewed the final manuscript, results, and figures and take full responsibility for the content of the paper.

ETHICS STATEMENT

This work uses existing SAR target-recognition datasets and involves neither human participants nor personally identifiable information. SAR generation and recognition nevertheless have dual-use implications, including surveillance and military applications. Synthetic SAR images may be mistaken for measurements or used to augment downstream recognition systems; their provenance should therefore be preserved and their synthetic status clearly identified. The reported results are controlled research benchmarks and do not establish reliability or safety for operational deployment.

REPRODUCIBILITY STATEMENT

Section 3 specifies the problem formulation, CSS, QSS, CST, and the flow-matching objective. Appendices A and B document the datasets, splits, sample budgets, preprocessing, evaluation protocols, response construction, calibration, model architecture, optimization, inference procedure, random seeds, and metric definitions. Appendices C–F provide the supporting ablations, randomized controls, response diagnostics, compatible-reference evaluations, and complete per-protocol comparisons. Downstream results are reported over five random seeds with generated images held fixed, and paired generator comparisons use controlled initialization and sampling as described in Appendix B.

REFERENCES

Chenjie Cao, Yunuo Cai, Qiaole Dong, Yikai Wang, and Yanwei Fu. Leftrefill: Filling right canvas based on left reference through generalized text-to-image diffusion model. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 7705–7715. IEEE, 2024.

Mihai Datcu, Zhongling Huang, Andrei Anghel, Juanping Zhao, and Remus Cacoveanu. Explainable, physics-aware, trustworthy artificial intelligence: A paradigm shift for synthetic aperture radar. *IEEE Geoscience and Remote Sensing Magazine*, 11(1):8–25, 2023. doi: 10.1109/MGRS.2023.3237465.

Prafulla Dhariwal and Alexander Nichol. Diffusion models beat gans on image synthesis. *Advances in neural information processing systems*, 34:8780–8794, 2021.

Joseph R Diemunsch and John Wissinger. Moving and stationary target acquisition and recognition (mstar) model-based automatic target recognition: Search technology for a robust atr. In *Algorithms for synthetic aperture radar Imagery V*, volume 3370, pp. 481–492. SPIE, 1998.

Noam Elata, Bahjat Kawar, Yaron Ostrovsky-Berman, Miriam Farber, and Ron Sokolovsky. Novel view synthesis with pixel-space diffusion models. In *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 26756–26766. IEEE, 2025.

Dongdong Guan, Rui Feng, Yuzhen Xie, Huaiyue Ding, Yang Cui, and Deliang Xiang. Sar vehicle data generation with scattering features for target recognition. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 19:5520–5538, 2026. doi: 10.1109/JSTARS.2026.3652520.

Rui Guo, Tianyao Huang, Maokun Li, Haiyang Zhang, and Yonina C. Eldar. Physics-embedded machine learning for electromagnetic data imaging: Examining three types of data-driven imaging methods. *IEEE Signal Processing Magazine*, 40(2):18–31, 2023. doi: 10.1109/MSP.2022.3198805.

Bingyu Han, Mou Wang, Shunjun Wei, Kun Chen, Zeyang Dai, Jin Li, Xiaowo Xu, Xiaoling Zhang, Zongyong Cui, Di Jiang, Yuanyuan Zhou, and Pengcheng Gao. Scattering center prior-guided diffusion for unknown-azimuth sar image generation. *Remote Sensing*, 18(14), 2026. ISSN 2072-4292. doi: 10.3390/rs18142417. URL https://www.mdpi.com/2072-4292/18/14/2417.

Zhongling Huang, Yihan Zhuang, Zipei Zhong, Feng Xu, Gong Cheng, and Junwei Han. X-fake: Juggling utility evaluation and explanation of simulated sar images. *IEEE Transactions on Image Processing*, 34:7830–7844, 2025. doi: 10.1109/TIP.2025.3634988.

Zhongling Huang, Xidan Zhang, Zuqian Tang, Feng Xu, Mihai Datcu, and Junwei Han. Generative artificial intelligence meets synthetic aperture radar: A survey. *IEEE Geoscience and Remote Sensing Magazine*, 14(1):6–48, 2026. doi: 10.1109/MGRS.2024.3483459.

Zihao Jiang, Chao Liu, Zhangzeyu Xing, Yamei Wang, and Shuwen Xiao. Physically consistent sar image generation for unseen aspect angles via attributed scattering center evolution. *Remote Sensing*, 18(13), 2026. ISSN 2072-4292. doi: 10.3390/rs18132247. URL https://www.mdpi.com/2072-4292/18/13/2247.

Tero Karras, Miika Aittala, Samuli Laine, Erik Härkönen, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. Alias-free generative adversarial networks. In *Proceedings of the 35th International Conference on Neural Information Processing Systems*, NIPS '21, Red Hook, NY, USA, 2021. Curran Associates Inc. ISBN 9781713845393.

Sangyun Lee, Zinan Lin, and Giulia Fanti. Improving the training of rectified flows. *Advances in neural information processing systems*, 37:63082–63109, 2024.

Zhengxin Lei, Feng Xu, Jiangtao Wei, Feng Cai, Feng Wang, and Ya-Qiu Jin. Sar-nerf: Neural radiance fields for synthetic aperture radar multiview representation. *IEEE Transactions on Geoscience and Remote Sensing*, 62:1–15, 2024. doi: 10.1109/TGRS.2024.3452505.

Yongxiang Liu, Weijie Li, Li Liu, Jie Zhou, Bowen Peng, Yafei Song, Xuying Xiong, Wei Yang, Tianpeng Liu, Zhen Liu, et al. Atrnet-star: A large dataset and benchmark towards remote sensing object recognition in the wild. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2026a.

Zinan Liu, Shengren Niu, Xiaolan Qiu, Lingxiao Peng, Yuying Shang, Lihua Zhong, and Chibiao Ding. A differentiable method for novel view sar image generation via 3d gaussian splatting. *ISPRS Journal of Photogrammetry and Remote Sensing*, 231:167–195, 2026b. ISSN 0924-2716. doi: https://doi.org/10.1016/j.isprsjprs.2025.10.034. URL https://www.sciencedirect.com/science/article/pii/S0924271625004204.

Norman Müller, Katja Schwarz, Barbara Rössle, Lorenzo Porzi, Samuel Rota Bulò, Matthias Nießner, and Peter Kontschieder. Multidiff: Consistent novel view synthesis from a single image. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 10258–10268. IEEE, 2024.

Augustus Odena, Christopher Olah, and Jonathon Shlens. Conditional image synthesis with auxiliary classifier gans. In *International conference on machine learning*, pp. 2642–2651. PMLR, 2017.

Jihyong Oh and Munchurl Kim. Peacegan: A gan-based multi-task learning method for sar target image generation with a pose estimator and an auxiliary classifier. *Remote Sensing*, 13(19), 2021. ISSN 2072-4292. doi: 10.3390/rs13193939. URL https://www.mdpi.com/2072-4292/13/19/3939.

Amir Rahimi and Stella Yu. Noise-tolerant novel-view sar synthesis via denoising diffusion. *IEEE Transactions on Geoscience and Remote Sensing*, 2026.

Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 10684–10695, 2022.

Xiaokun Sun, Xinwei Li, Deliang Xiang, and Canbin Hu. Sar vehicle image generation with integrated deep imaging geometric information. *International Journal of Applied Earth Observation and Geoinformation*, 132:104028, 2024. ISSN 1569-8432. doi: https://doi.org/10.1016/j.jag.2024.104028. URL https://www.sciencedirect.com/science/article/pii/S1569843224003820.

Yuanshuang Sun, Yinghua Wang, Liping Hu, Yuanyuan Huang, Hongwei Liu, Siyuan Wang, and Chen Zhang. Attribute-guided generative adversarial network with improved episode training strategy for few-shot sar image generation. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 16:1785–1801, 2023. doi: 10.1109/JSTARS.2023.3239633.

Chenwei Wang, Jifang Pei, Xiaoyu Liu, Yulin Huang, Deqing Mao, Yin Zhang, and Jianyu Yang. Sar target image generation method using azimuth-controllable generative adversarial network. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 15:9381–9397, 2022. doi: 10.1109/JSTARS.2022.3218369.

Xuanting Wu, Fan Zhang, Fei Ma, Yingbing Liu, Lingxiao Peng, Qiang Yin, and Yongsheng Zhou. Geodiff-sar ii: 3d-driven foundation diffusion models for sar generation via decoupled control, 2026. URL https://arxiv.org/abs/2605.21116.

Deliang Xiang, Ye Liu, Jianda Cheng, Xinyu Lu, Yuzhen Xie, and Dongdong Guan. Sar target recognition with image generation and azimuth angle feature constraints. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 18:18561–18580, 2025. doi: 10.1109/JSTARS.2025.3586614.

Jiawei Zhang, Zhen Liu, Weidong Jiang, Yongxiang Liu, Xiaolin Zhou, and Xiang Li. Application of deep generative networks for SAR/ISAR: a review. 56(10):11905–11983, 2023. ISSN 0269-2821, 1573-7462. doi: 10.1007/s10462-023-10469-5.

Xidan Zhang, Yihan Zhuang, Qian Guo, Haodong Yang, Xuelin Qian, Gong Cheng, Junwei Han, and Zhongling Huang. ϕ-gan: Physics-inspired gan for generating sar images under limited data. In *2025 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 29075–29085, 2025. doi: 10.1109/ICCV51701.2025.02700.

Zixuan Zhang, Heng Zhang, Wei Xiang, Fengli Xue, Congrui Yang, Sheng Chang, and Yunkai Deng. A visibility-aware neural gaussian framework for sar image generation and data augmentation. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 19:24255–24278, 2026. doi: 10.1109/JSTARS.2026.3709914.

Xiao Xiang Zhu, Sina Montazeri, Mohsin Ali, Yuansheng Hua, Yuanyuan Wang, Lichao Mou, Yilei Shi, Feng Xu, and Richard Bamler. Deep learning meets sar: Concepts, models, pitfalls, and perspectives. *IEEE Geoscience and Remote Sensing Magazine*, 9(4):143–172, 2021. doi: 10.1109/MGRS.2020.3046356.

APPENDIX

This appendix specifies the data and evaluation protocols, implementation details, and additional results. It follows the main paper's distinction between target-derived Query-Aligned Spatial Support (QSS) and optional Compatible Scattering Transfer (CST). The cross-model benchmark uses QSS only; Its ten-category ATRNet-STAR benchmark uses different scene and sampling protocols from the CST studies, with some vehicle categories in common.

A DATASETS AND EVALUATION PROTOCOLS

For every downstream condition, a separate predictor is trained for target classification and viewing-angle estimation using five random seeds, 0–4. We report the mean and sample standard deviation across these runs, keeping the generated images fixed.

A.1 MSTAR AND THE TEN-CATEGORY ATRNET-STAR BENCHMARK

MSTAR. The ten categories are 2S1, BMP-2, BRDM-2, BTR-60, BTR-70, D7, T-62, T-72, ZIL-131, and ZSU-234. Sparse generator training samples are drawn from the 2,747-image 17° source pool. The downstream test set contains 2,425 real images at 15° and is not used for generator training. The six azimuth intervals are 30, 45, 60, 72, 90, and 120°, corresponding to 121, 81, 60, 50, 40, and 30 observed images across the ten categories. These are actual list sizes, not counts inferred from an ideal angular grid. Capacity, alignment, and response-information studies use all six protocols; the complete cross-model comparison uses 30, 45, 60, and 120°.

Table 6: MSTAR dense source and real downstream-test counts by category. Sparse training uses subsets of the source pool.

Category	Source 17°	Test 15°	Category	Source 17°	Test 15°
2S1	299	274	D7	299	274
BMP-2	233	195	T-62	299	273
BRDM-2	298	274	T-72	232	196
BTR-60	256	195	ZIL-131	299	274
BTR-70	233	196	ZSU-234	299	274
			Total	2,747	2,425

The dense 17° source list specifies 2,747 generation queries and the reference pool for generation metrics. It includes observed as well as unobserved source azimuths. This generation evaluation must be distinguished from downstream recognition on the disjoint 15° test set. Only the response-information diagnostic deliberately reveals measured query responses, as specified in Section D.

ATRNet-STAR cross-model benchmark. All experiments use X-band VH images. The ten categories are Hongqi H5, Chang'an CS75 Plus, Dongfeng EQ6608LTV, Yutong ZK6120HY1, Dongfeng Tianjin KR230, SHACMAN DeLong X3000, SHACMAN DeLong M3000, CNHTC HOWO, Changlin 8228-5, and SDLG ZL40F. Each sandstone source plane contains 12 available azimuths per category before sparse subsampling.

For azimuth generation, training uses sandstone images at 15° depression, with 60, 90, or 120° azimuth spacing. The real test set contains 923 images at the same depression from urban (591), factory (267), woodland (29), and bare-soil (36) scenes. For depression generation, training and testing both use sandstone, but the complete query-depression plane is withheld. We evaluate interpolation from 15 and 45° to 30°, and extrapolation from 15 and 30° to 45° and from 30 and 45° to 15°. Both observed planes use 60° source-azimuth spacing. Each test set contains 120 real images at the requested generation depression, which is absent from the real training observations. Table 7 summarizes the sample budgets.

Downstream controls. *Sparse only* trains the downstream predictor on the selected real observations alone. For each learned generator, the augmented downstream training set combines these same

Table 7: Cross-model benchmark sample counts, pooled over ten categories. The predictor's augmented training set combines the generator's real training observations with the synthetic images. Each depression task uses two observed source planes.

Study	Azimuth spacing	Augmented predictor training			Predictor real test
		Generator real train	Synthetic added	Total	
MSTAR	30°	121	2,747	2,868	2,425
	45°	81	2,747	2,828	2,425
	60°	60	2,747	2,807	2,425
	120°	30	2,747	2,777	2,425
ATRNet-STAR azimuth	60°	60	3,600	3,660	923
	90°	40	3,600	3,640	923
	120°	30	3,600	3,630	923
ATRNet-STAR depression	60° per source plane	120	3,600	3,720	120

real observations with generated images covering the full azimuth range: 2,747 synthetic images for MSTAR or 3,600 for the ten-category ATRNet-STAR benchmark. *Repeated sparse* matches the total size of this combined training set by duplicating the sparse real observations, without adding new views. Within each protocol, all generators use the same real observations and contribute the same number of synthetic images.

A.2 CST SUBSETS AND COMPATIBLE PAIRS

All CST experiments use the urban X-band VH subset of ATRNet-STAR, with reference sequences retaining their available measurements on the native 5° azimuth grid. Target–reference pairs are selected using measured response compatibility and then fixed for subsequent experiments.

Compatibility screening compares canonically aligned, spatially smoothed log-response changes and image cosine similarity over common measured azimuths. A 90°-shifted reference provides a control for whether the matching reflects the viewing direction. The comparisons are evaluated on two interleaved angular subsets, starting at 0 and 5° with 10° spacing. Full measured target sequences are available during screening. Once the pairs are fixed, only the designated target observations construct CSS, calibrate CST, and supervise FM. Pair selection does not use generated-image or downstream performance.

The main study retains two screened pairs: FAW J6P with Dongfeng Tianjin KR230, and Changlin 8228-5 with SDLG ZL40F. The first pair comprises two trucks; the second comprises a road roller and a wheel loader. Both pairs are evaluated in both transfer directions. Figure 6 shows their optical appearances and measured SAR responses.

The additional set contains four screened target–reference pairs, excluding the main-study categories and keeping target and reference identities disjoint. Figure 10 shows these pairs, which remain fixed across unseen-azimuth generation and depression interpolation.

For unseen-azimuth generation, the main study uses target observations at 30° depression with nominal 30 or 60° azimuth spacing. The additional set uses 15° depression with 30, 60, or 120° azimuth spacing. In both studies, real target images at the remaining measured azimuths form the test set. Observed target azimuths are excluded from these test queries.

For depression interpolation, target observations come from the 15 and 45° planes, and the complete 30° target plane is reserved for evaluation. Target images at 30° are excluded from CSS construction, CST-strength calibration, and FM training. The target's query-plane CSS is interpolated from its two observed-plane CSSs, while the reference supplies measurements at 30°. Source azimuths use the native grid or 30 and 60° spacing in the main study; the additional set also includes 120° spacing. Construction and calibration details are provided in Appendix B.2.



Figure 6: The two reciprocal CST pairs in the main study. SAR panels show measured urban X-band VH images at 30° depression. Optical images show the corresponding vehicles from the official ATRNet-STAR catalogue and serve only as illustrations.

Table 8: CST sample counts, pooled over four target categories. Real target training images are used by both the generator and the downstream predictor. Augmented predictor training adds 288 generated target images. Dense reference measurements are not included in these counts.

Subset	Task	Azimuth spacing	Real target train	Predictor augmented train	Real target test
Main pairs	Azimuth	30°	46	334	225
		60°	23	311	248
	Dep. interpolation	30°	90	378	271
		60°	47	335	271
Additional set	Azimuth	30°	40	328	182
		60°	22	310	200
		120°	12	300	210
	Dep. interpolation	30°	88	376	269
		60°	46	334	269
		120°	24	312	269

In the main study, the reciprocal transfer directions are trained as separate two-target generators. Within each direction, target and reference identities are disjoint; their roles exchange in the other direction. For each generation condition, downstream training combines sparse real target observations with generated target views pooled across both directions. The additional set instead trains a four-target generator with fixed target and reference roles. In both studies, the downstream network predicts four-class identity and viewpoint and is evaluated on real held-out target images. Dense reference sequences provide generator conditioning and are not directly added to downstream training.

Table 8 summarizes the sample counts. The real target observations serve as generator training data and as the real component of downstream training. Each augmented downstream training set additionally contains 288 generated target images on a 5° query grid.

Optical illustrations are cropped from the actual-vehicle catalogue in the official ATRNet-STAR repository. The figure-source package records the catalogue source and crop coordinates. These images illustrate vehicle identity only; compatibility screening and generation use SAR measurements.

B Implementation and Metric Definitions

B.1 Response maps, CSS, and QSS

Starting from the preprocessed SAR images, we construct response maps $r_{c,i}$ on a 128×128 grid. Inverse azimuth alignment gives $\bar{r}_{c,i}$, which are aggregated by target class and depression to form CSS.

For MSTAR, we retain the 128 strongest pixels of each image and render an amplitude-weighted sum of Gaussian kernels centered at their locations, with standard deviation 0.7 pixels. Dividing this rendered map by its maximum gives $r_{c,i}$. After inverse alignment, Agg takes the pixelwise mean and peak-normalizes the result:

$$\mathrm{CSS}_c(\varphi) = \mathcal{N}(\mathrm{mean}_i\, \bar{r}_{c,i})\,, \tag{9}$$

where the mean includes only observations at depression φ, and $\mathcal{N}$ divides a map by its maximum value, with a minimum denominator of 10^{-8}.

For ATRNet-STAR, we use the preprocessed SAR image itself as $r_{c,i}$. After inverse alignment, Agg combines the pixelwise mean and maximum across observed views:

$$\mathrm{CSS}_c(\varphi) = 0.8\ \mathrm{mean}_i\, \bar{r}_{c,i} + 0.2\ \max_i \bar{r}_{c,i}. \tag{10}$$

Both operations use observed views at the same depression φ. The maximum is taken independently at each spatial pixel. No additional peak normalization is applied after aggregation.

Inverse alignment and query rotation use bilinear sampling with zero padding and opposite rotation directions. At a query depression between two observed planes, CSS is linearly interpolated between their CSSs; for 15 and 45° observations queried at 30°, this is their midpoint. Outside the observed range, the QSS-only benchmark uses the nearest observed-plane CSS. Rotating the resulting CSS to the query azimuth gives QSS. During training, the generator receives each training image's observed view; during generation, it receives the requested view $\mathbf{v}_q$ and the corresponding QSS.

B.2 CST construction and calibration

Reference observations undergo the same response construction, inverse alignment, and aggregation as target observations, giving $\bar{r}_{h,i}$ and $\mathrm{CSS}_h(\varphi)$. At each reference depression φ, CSS aggregates the available reference views at that depression. The log ratio defining $D_{h,i}$ uses an offset ϵ_0 of 0.02. We apply a size-preserving spatial average filter of size 5×5, followed by clipping between -1.5 and 1.5. The filter smooths neighboring pixels within each map, without averaging across viewing angles.

At the query depression, the reference change is retrieved directly when the requested azimuth is measured. Otherwise, circular linear interpolation uses the two neighboring measured azimuths, accounting for wrap-around at 360°. Registration then maps this change to the target's canonical coordinates through centered anisotropic scaling, yielding $D_h(c, \mathbf{v}_q)$. The sampling-grid scale factors are ℓ_h/ℓ_c horizontally and w_h/w_c vertically, where ℓ and w denote vehicle length and width. Sampling uses bilinear interpolation and zero padding.

The spatial weight map is $W_c(\varphi) = q_c M_c(\varphi)$, where $M_c(\varphi)$ is a binary target support mask and q_c is a calibrated scalar strength. The mask is obtained by thresholding $\mathrm{CSS}_c(\varphi)$ at 0.05 and dilating the result with a 7×7 max filter. Thus, W_c equals q_c inside the mask and zero outside it. The implemented CST response includes intensity clipping before query alignment:

$$R^{\mathrm{CST}}_{h\to c}(\mathbf{v}_q) = \mathcal{W}_{\theta_q}\left[\mathrm{clip}_{[0,1]}\left(\mathrm{CSS}_c(\varphi_q) \odot \exp(W_c(\varphi_q) \odot D_h(c, \mathbf{v}_q))\right)\right]. \tag{11}$$

Outside the mask, the multiplicative factor is one, leaving the target CSS unchanged.

We select q_c from 0, 0.25, 0.5, 0.75, and 1 using leave-one-observed-azimuth-out validation. For each held-out observation, target CSS and its support mask are rebuilt from the remaining observations at the same depression. Each candidate strength produces a reconstructed response at the held-out view, which is compared with the measured target response. The reconstruction MAE assigns weight 5 to pixels exceeding 0.05 in either response and weight 1 elsewhere. We select the strength with the lowest mean validation error, resolving ties toward the smaller value.

For depression interpolation, calibration uses only the observed target planes at 15 and 45°, together with reference responses at the corresponding depression. Errors from both planes are pooled to select a single q_c. At the query depression of 30°, target CSS is the midpoint of the two observed-plane CSSs, and its support mask is derived from this interpolated CSS. CST uses the reference sequence measured at 30°; reference responses are not interpolated across depression. The withheld 30° target images are excluded from CSS construction, strength calibration, and FM training.

The generator's first spatial channel always contains QSS. Its second channel is zero for the QSS-only model and $R^{\mathrm{CST}}_{h\to c}(\mathbf{v}_q) - \mathrm{QSS}_c(\mathbf{v}_q)$ for the CST-enabled model. The two generators are trained separately with matched architecture, initial weights, target observations, and optimization budget. Their generation queries and noise are also matched.

B.3 Training and downstream evaluation

The flow-matching velocity network u_ψ uses a U-Net with class, time, and view embeddings. The standard configuration has base width 64 and approximately 3.87M parameters. MSTAR capacity experiments use widths 16, 24, 32, 48, 64, 96, and 128. No-QSS controls retain the same class and view embeddings and replace the spatial channels with zeros. All Φ-Flow runs use AdamW with learning rate and weight decay 0.0001, batch size 8, gradient-norm clipping at 1, and 50 Euler steps at inference. MSTAR and fixed-depression CST training use 2,000 epochs; the ten-category ATRNet-STAR benchmark uses 3,000 epochs. CST depression interpolation uses approximately 14,400 optimizer updates, rounded to complete epochs.

Matched generator comparisons use seed 42 and identical saved initial weights. For MSTAR and the ten-category ATRNet-STAR benchmark, downstream predictors are trained for 100 epochs with batch size 32, learning rate 0.001, zero weight decay, and angular-loss weight 20. Inputs are resized to 88×88. Generated PNG values are read directly without an additional per-image contrast stretch.

For CST, the augmented downstream training set combines the observed real target images with 288 generated target images per four-category set, covering the full azimuth range on a 5° grid. Predictors are trained for 3,000 optimizer updates with batch size 32. Each training example is sampled with equal probability from the real or synthetic pool, then uniformly across target categories within that pool. Dense reference sequences are used to construct CST conditions, not appended as additional real training images for the predictor. Accuracy and viewing-angle errors are evaluated on the real held-out target images.

B.4 Metrics

Generation. MSTAR FID compares generated images with the dense 17° real source pool using the same Inception-v3 pipeline for all methods. We do not report ATRNet-STAR FID: the benchmark's designated real query plane contains only 120 images in total, making that estimate less stable than the MSTAR comparison. SSIM is computed on the central 80×80 image crop.

Support IoU is computed from generated and reference Top-128 Gaussian response maps, independently peak-normalized, centrally cropped, and thresholded at 0.20. This is strong-response support overlap, not a manually annotated target segmentation. Generated-target orientation error compares the intensity-weighted principal axes of these response maps modulo 180°. It measures image-domain orientation and is distinct from a downstream predictor's full-azimuth error.

Downstream angles. Azimuth MAE is the mean absolute circular difference modulo 360°; depression MAE uses the absolute depression-angle difference. In the cross-model tables, azimuth and depression MSE are computed on the two-dimensional sine/cosine angle encodings, averaged over both components and test images. These MSEs are dimensionless, not squared degree errors. A zero sine/cosine prediction has MSE 0.5, so this loss is not interchangeable with angular MAE. Accuracy is the fraction of correctly classified real test images.

C Additional Evidence for Query-Aligned Spatial Support

C.1 Capacity and Angular Sampling

Figure 7 compares generators with and without QSS across seven network widths and six azimuth sampling protocols. The two families differ most clearly in support IoU and orientation error. Once QSS is provided, increasing width produces relatively small changes in these spatial metrics, while FID, SSIM, and downstream performance remain sensitive to the sampling protocol. These results support the main finding that increasing model capacity does not substitute for query-aligned spatial conditioning.

The 90° protocol also differs from a simple ordering by observation count: its four nominal views contain only two distinct axes modulo 180°, whereas the three views at 120° contain three. This geometric distinction helps explain why interval and sampling configuration should be retained when interpreting the curves; it does not make opposite-view scattering responses interchangeable.

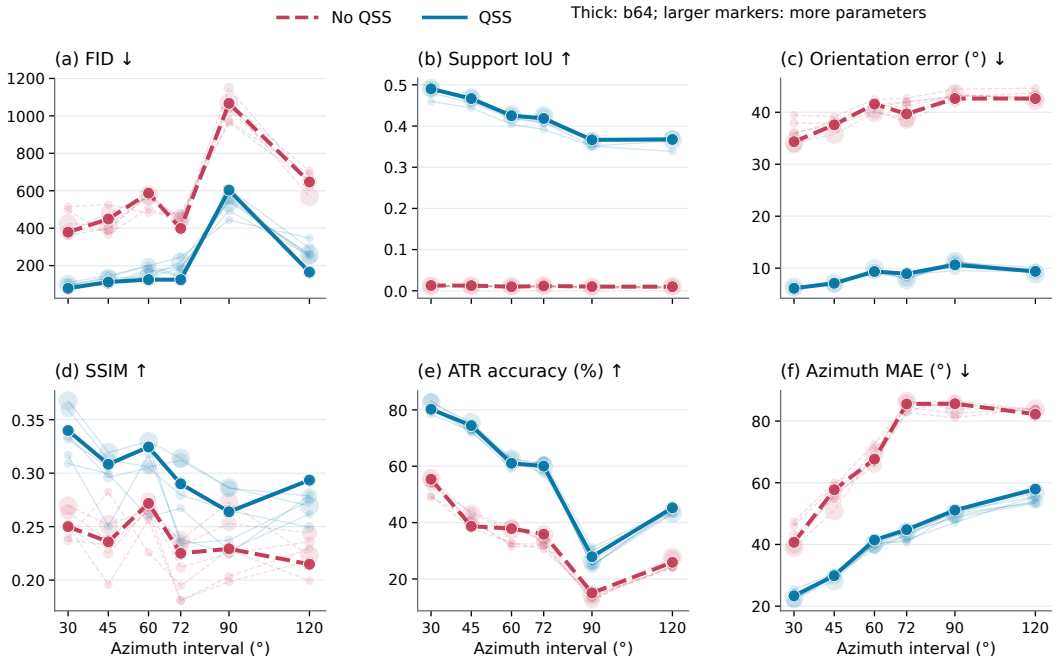


Figure 7: Capacity scaling across six MSTAR azimuth protocols. Thin curves show all seven widths; thick curves highlight width 64. Marker area increases with parameter count. Generation metrics use one paired generator run; downstream metrics average five predictor seeds.

C.2 Alignment, Amplitude, and Randomized Conditions

All spatial ablations use the same width-64 generator. The aligned-mask control is implemented as a soft threshold, $\mathrm{sigmoid}((\mathrm{QSS} - a)/b)$, with threshold a of 0.12 and temperature b of 0.025. It preserves support while suppressing most amplitude variation; it is not an exactly binary tensor. Figure 8 shows that query alignment accounts for most of the orientation improvement, while retaining QSS amplitude gives a smaller, less uniform downstream increment.

Randomized controls supply either another class's support or an incorrect azimuth, while leaving the true class/view embeddings unchanged. Wrong classes exclude the true class; wrong azimuths have a random sign and an offset between 30 and 150°. Perturbations are resampled deterministically by sample and epoch, preventing a fixed permutation from becoming an alternative condition encoding. Table 9 shows that correctly matched support consistently gives higher accuracy.

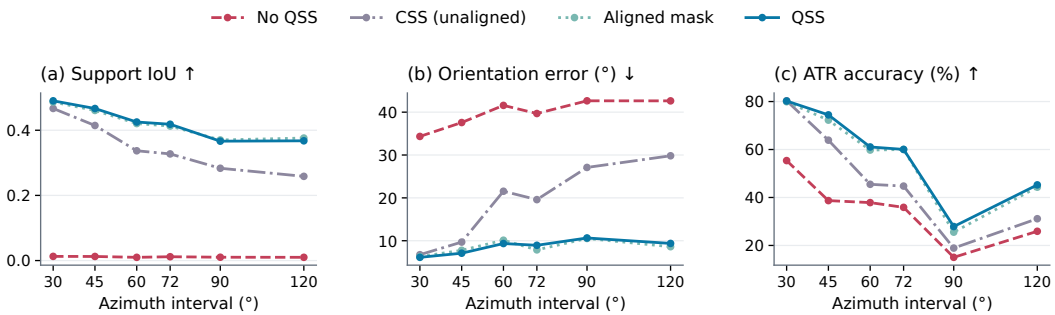


Figure 8: Protocol-wise spatial conditioning at width 64. Unaligned CSS is useful, but aligning it to the query substantially improves orientation. The aligned soft-mask control closely tracks QSS in support and orientation. Accuracy averages five downstream seeds.

Table 9: Randomized spatial conditions. Accuracy (%) is mean ± sample standard deviation over five downstream seeds. Wrong classes or azimuths are resampled for each training example, rather than fixed to a learnable permutation.

Azimuth interval	Wrong azimuth	Wrong class	QSS
30°	78.09 ± 2.13	79.40 ± 1.41	$\mathbf{80.23 \pm 1.06}$
45°	64.67 ± 2.63	63.99 ± 1.45	$\mathbf{74.44 \pm 1.39}$
60°	49.81 ± 2.24	52.00 ± 1.25	$\mathbf{61.06 \pm 1.15}$
72°	46.27 ± 1.97	52.20 ± 1.87	$\mathbf{60.04 \pm 2.69}$
90°	17.39 ± 1.07	17.87 ± 0.98	$\mathbf{27.88 \pm 2.07}$
120°	37.56 ± 1.49	37.10 ± 1.22	$\mathbf{45.25 \pm 2.40}$

D Response Information Beyond QSS

D.1 Representations and Target-Only Estimation

We evaluate response information along two separate axes: its estimation from sparse target observations and its downstream value when supplied accurately to the generator. Both experiments use the six MSTAR protocols with azimuth intervals of 30, 45, 60, 72, 90, and 120°. The six response representations indexed by r are constructed from the prepared SAR images in image coordinates as follows:

- **Top-32, Top-64, and Top-128:** retain the locations and amplitudes of the corresponding number of strongest image pixels, apply Gaussian splatting with width 0.7 pixels, and peak-normalize the map.
- **LP-16:** subtract QSS from the measured Top-128 representation, adaptively average-pool the residual to 16×16, bilinearly upsample it to the original resolution, and add it back to QSS. Clip the result between 0 and 1 and peak-normalize it.
- **Local amplitude:** average the measured image and QSS with a 9×9 spatial filter. Divide the former by the latter, with the denominator floored at 0.001, and clip the resulting gain between 0 and 3. Multiply QSS by this gain and clip the map between 0 and 1. This modulates amplitude on the existing support without independently adding new support.
- **Full foreground:** multiply the measured-image residual relative to QSS by a soft QSS-derived mask, then add it back to QSS. The mask is obtained by 17×17 max-pool dilation of QSS followed by a sigmoid with threshold 0.05 and temperature 0.02. The resulting representation is clipped between 0 and 1, without peak normalization.

For target-only estimation, each extracted response representation is inverse-aligned to the canonical frame. Angular estimation operates on these representations, rather than on the original SAR images. The estimates are not passed through the representation extractor again. Representations from repeated measurements at the same azimuth within a class are averaged before fitting. The four

candidate angular rules are static averaging of the observed representations, nearest-view transfer, circular linear interpolation between adjacent observed azimuths, and pixelwise first-order harmonic fitting with constant, sine, and cosine terms. Harmonic fitting uses ridge coefficient 0.25; the constant term receives one twentieth of the penalty applied to the other terms. Estimated values are bounded between 0 and 1.

For each representation and protocol, we select one angular rule by leave-one-observed-azimuth-out validation. Within each class, all measurements at the held-out azimuth are excluded from the estimator's inputs. Candidate rules are ranked by their mean validation score over all held-out class–azimuth groups, with top-response recall breaking ties. The protocol's CSS remains fixed. After selection, the chosen rule uses all observed representations of each class to estimate those at unseen azimuths. Unseen query images are used only to evaluate these estimates.

For an unseen query i, let $R_{r,i}$ be its measured response representation in the canonical frame and $\widehat{R}_{r,i}$ its estimate. We compute $s_{r,i} = 1 - \mathrm{MAE}(\widehat{R}_{r,i}, R_{r,i}) / \max\{\mathrm{MAE}(\mathrm{CSS}_{c_i}, R_{r,i}), \epsilon\}$, with ϵ =0.00000001. Both MAEs are evaluated over the complete map, including for Top-K representations. Scores are averaged within each class, then equally over the ten classes and six protocols to obtain S_r. This is an average of per-query relative error reductions, not a ratio of pooled MAEs. Larger S_r indicates a larger improvement over CSS for that representation; different representations have different spatial contents and therefore different error baselines.

D.2 Oracle-conditioned generation and downstream utility

This experiment evaluates the value of response information when it is available accurately, independently of the target-only estimators above. For each representation and protocol, we train a separate b64 FM model from the same initial weights as QSS-only Φ-Flow. Training uses the same sparse observed images, with each image's measured response representation provided as a condition. In image coordinates, the first spatial channel contains QSS and the second contains the measured representation minus QSS; the QSS-only baseline instead uses a zero second channel. The architecture, QSS, optimization settings, and training budget are otherwise unchanged.

During generation, the trained model receives the response representation extracted from the measured query image, with generation queries and query-wise noise matched to the QSS-only baseline. Thus, measured responses are supplied during both training and generation, but generator training still uses only the sparse observed images. No angular estimate is used in this branch. This oracle experiment measures potential value under accurate response conditions rather than performance with a deployable response predictor.

For downstream evaluation, the generated images augment the same sparse real training set. The recognition and viewpoint models are evaluated on the real MSTAR test images at 15° depression, using the same five downstream seeds as the QSS-only baseline. For each protocol, we compute the difference in mean accuracy between oracle-conditioned and QSS-only generation, expressed in percentage points; its equal-weight average over the six protocols is U_r. We similarly report the reduction in downstream azimuth MAE, with positive values indicating lower error. These downstream gains are measured using the oracle-generated images, not images conditioned on the estimates used to calculate S_r.

Across these six representations, the mean recognition gains and azimuth-MAE reductions have the same ordering. Increasing localized response detail, and ultimately supplying the full foreground response, provides greater downstream value under measured response conditions. The tested target-only estimators achieve smaller relative improvements over CSS for these more informative responses. The distinction between response estimation and downstream value therefore extends to viewpoint estimation, motivating additional angular evidence from a compatible reference in CST.

E CST on an Additional Compatible Target Set

We further evaluate CST on four additional target–reference pairs, shown in Figure 10. Target identities are disjoint from reference identities within this set, and the associations remain fixed across sampling protocols. The experiments compare Φ-Flow and Φ-Flow$_{\mathrm{CST}}$ using the same target observations and training settings.

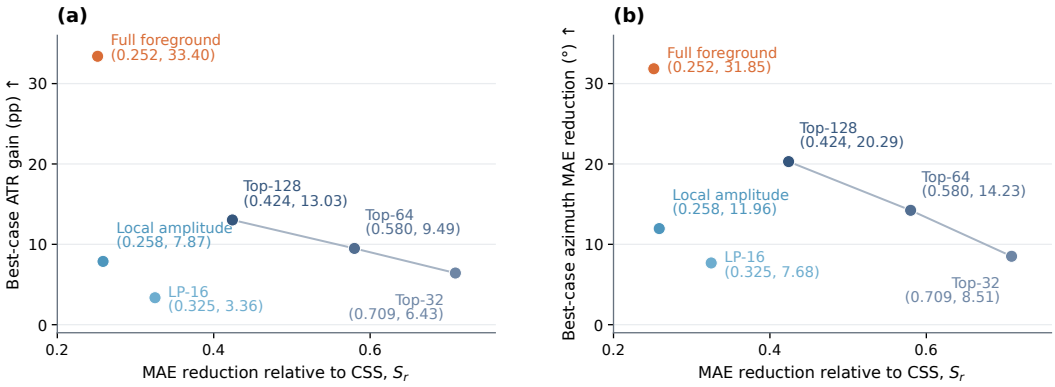


Figure 9: Response-estimation improvement versus downstream value under measured response conditions. Each point denotes one representation. The horizontal axis shows relative MAE reduction over CSS; the vertical axes show ATR gain (left) and azimuth-MAE reduction (right) over QSS-only Φ-Flow. All values are averaged over six protocols. Lines connect only the Top-K family.

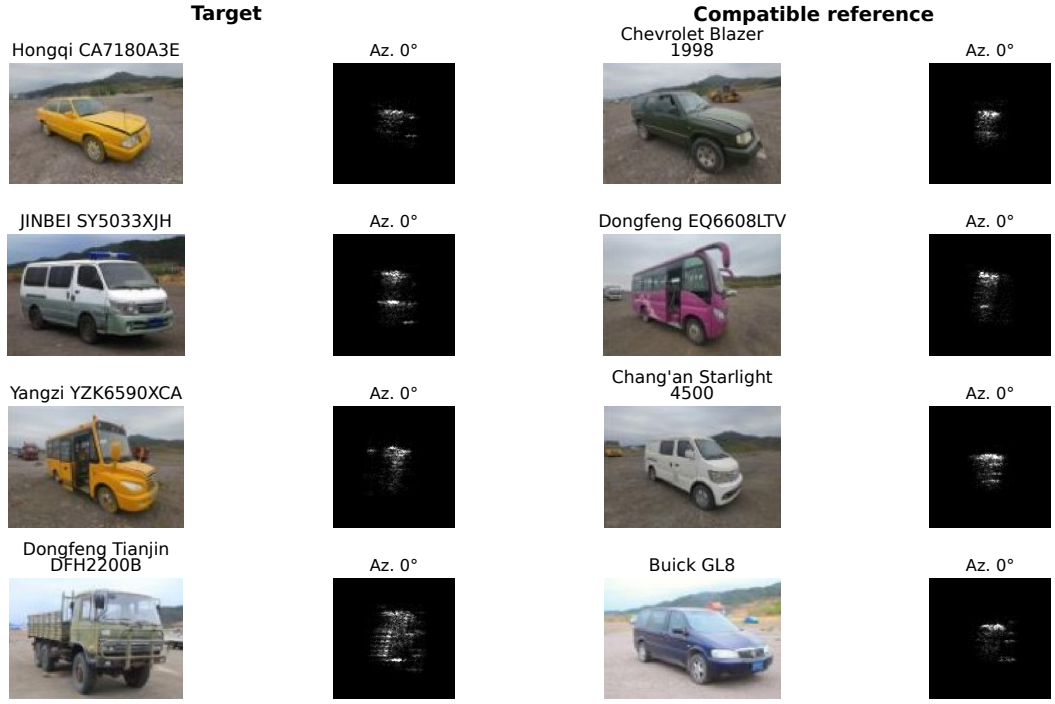


Figure 10: Additional compatible target–reference pairs. Each left-hand target receives response variation from the right-hand reference. SAR panels show measured urban X-band VH images at 15° depression and the labeled azimuth. Optical images are from the official ATRNet-STAR vehicle catalogue and are used only for illustration.

E.1 Unseen-azimuth generation

Target observations are sampled at 30, 60, or 120° azimuth intervals on the 15° depression plane. References retain their available measurements on the native 5° grid. Downstream evaluation uses real target images at azimuths excluded from the observed training set.

Table 10 shows lower mean azimuth MAE with CST at all three sampling intervals. Recognition accuracy, however, decreases at 30 and 60° and changes little at 120°. As in the main CST experiment, the more consistent benefit is additional viewpoint evidence rather than improved classification.

Table 10: CST at unseen azimuths for the additional target set. Results are mean ± sample standard deviation. Azimuth MAE is in degrees; bold marks the better mean within each matched pair.

Azimuth spacing	Model	Acc. (%) ↑	Az. MAE ↓
30°	Φ-Flow	**83.63 ± 1.37**	38.64 ± 3.96
	Φ-Flow$_{\mathrm{CST}}$	80.66 ± 1.19	**37.41 ± 2.19**
60°	Φ-Flow	**71.60 ± 1.98**	56.02 ± 3.75
	Φ-Flow$_{\mathrm{CST}}$	70.90 ± 3.07	**51.15 ± 4.35**
120°	Φ-Flow	69.90 ± 3.15	60.41 ± 1.43
	Φ-Flow$_{\mathrm{CST}}$	**70.29 ± 1.98**	**56.17 ± 2.07**

Table 11: CST depression interpolation for the additional target set. Target observations are at 15 and 45°; evaluation uses real target images at the withheld 30° plane. Spacing refers to azimuth sampling on each observed source plane. Results are mean ± sample standard deviation. Angle MAEs are in degrees; bold marks the better mean within each matched pair.

Azimuth spacing	Model	Acc. (%) ↑	Az. MAE ↓	Dep. MAE ↓
30°	Φ-Flow	77.32 ± 1.82	36.25 ± 4.78	**6.28 ± 1.23**
	Φ-Flow$_{\mathrm{CST}}$	**78.81 ± 1.95**	**30.37 ± 2.13**	6.50 ± 0.94
60°	Φ-Flow	69.44 ± 3.20	49.88 ± 3.51	5.79 ± 1.11
	Φ-Flow$_{\mathrm{CST}}$	**72.64 ± 2.21**	**41.88 ± 2.97**	**5.00 ± 0.79**
120°	Φ-Flow	**65.72 ± 2.70**	52.98 ± 5.25	4.72 ± 1.19
	Φ-Flow$_{\mathrm{CST}}$	65.50 ± 1.79	**42.66 ± 5.38**	**4.66 ± 1.21**

Reference-derived response variation can help distinguish viewing directions without necessarily improving the separation of target categories.

E.2 Depression interpolation

The complete 30° target plane is withheld. QSS is constructed by midpoint interpolation between the target CSSs at 15 and 45°, while CST uses the reference response measured at 30°. Target source-plane azimuths use 30, 60, and 120° spacing. CST strength is calibrated only on the observed target planes, as described in Section B.2.

Table 11 shows lower mean azimuth MAE with CST at every evaluated source-azimuth spacing. Accuracy improves at 30 and 60°, but not at 120°; depression MAE changes in both directions. The reference therefore contributes most consistently to azimuth estimation on the unobserved target plane. Here, QSS supplies spatial support through interpolation of the target's own CSS, while CST adds response variation measured at the query depression. These results extend the viewpoint benefit to this additional target set without implying a uniform gain in recognition or depression estimation.

F Complete Cross-Model Comparisons

Tables 12–14 expand the main comparison with all specified protocols and five-seed downstream variability. All Φ-Flow entries use QSS only, without dense references. Data splits and the sample-matched repeated-sparse control are specified in Section A.

Table 12: MSTAR-10 comparison at four azimuth spacings. Accuracy is in percent; azimuth MSE uses sine/cosine encodings. Downstream values are mean ± sample standard deviation over five seeds; FID is from one generator run. Bold marks the best mean.

		30°			45°		
Method	Params (M)	FID ↓	Acc. ↑	Az. MSE ↓	FID ↓	Acc. ↑	Az. MSE ↓
Sparse only	–	–	67.84 ± 3.49	0.244 ± 0.015	–	52.56 ± 4.51	0.330 ± 0.014
Repeated sparse	–	–	74.87 ± 1.59	0.224 ± 0.020	–	59.06 ± 4.52	0.337 ± 0.027
ACGAN	12.45	416.83	65.34 ± 4.88	0.266 ± 0.013	427.75	56.12 ± 5.42	0.375 ± 0.020
StyleGAN	23.22	199.83	79.60 ± 0.87	0.176 ± 0.009	189.83	68.75 ± 1.31	0.282 ± 0.020
Φ-GAN	12.45	160.41	80.07 ± 0.40	0.203 ± 0.019	272.76	65.18 ± 2.17	0.280 ± 0.017
cDDPM	107.90	152.25	79.81 ± 2.86	0.175 ± 0.017	207.43	62.28 ± 3.69	$\mathbf{0.230 \pm 0.020}$
LDM	234.47	98.82	56.59 ± 1.79	0.266 ± 0.021	187.25	48.07 ± 0.75	0.500 ± 0.029
FM	237.26	105.87	$\mathbf{80.75 \pm 2.85}$	0.194 ± 0.015	127.68	68.29 ± 1.09	0.295 ± 0.022
SAR-NV	1286.34	184.51	51.22 ± 2.43	0.353 ± 0.015	184.52	43.95 ± 2.43	0.540 ± 0.016
Φ-Flow	3.87	**78.90**	80.23 ± 1.06	$\mathbf{0.175 \pm 0.021}$	**112.15**	$\mathbf{74.44 \pm 1.39}$	0.232 ± 0.010
		60°			120°		
Method	Params (M)	FID ↓	Acc. ↑	Az. MSE ↓	FID ↓	Acc. ↑	Az. MSE ↓
Sparse only	–	–	37.35 ± 5.12	0.452 ± 0.043	–	29.41 ± 2.49	0.681 ± 0.031
Repeated sparse	–	–	44.22 ± 1.94	0.463 ± 0.018	–	34.39 ± 1.64	0.748 ± 0.002
ACGAN	12.45	678.13	44.38 ± 3.01	0.526 ± 0.018	1833.52	18.62 ± 10.13	0.523 ± 0.031
StyleGAN	23.22	246.14	45.82 ± 3.15	0.459 ± 0.013	394.93	32.47 ± 1.45	0.696 ± 0.020
Φ-GAN	12.45	313.57	46.12 ± 2.51	0.472 ± 0.035	487.86	31.22 ± 2.16	0.722 ± 0.025
cDDPM	107.90	257.00	45.52 ± 2.37	0.452 ± 0.039	298.89	31.07 ± 0.91	0.742 ± 0.039
LDM	234.47	280.17	43.00 ± 0.85	0.572 ± 0.022	304.54	32.99 ± 2.31	0.733 ± 0.045
FM	237.26	193.72	44.78 ± 1.60	0.520 ± 0.028	231.02	32.82 ± 2.77	0.741 ± 0.056
SAR-NV	1286.34	766.61	43.36 ± 3.97	0.423 ± 0.004	456.73	32.08 ± 2.80	0.647 ± 0.022
Φ-Flow	3.87	**125.15**	$\mathbf{61.06 \pm 1.15}$	$\mathbf{0.335 \pm 0.017}$	**165.49**	$\mathbf{45.25 \pm 2.40}$	$\mathbf{0.479 \pm 0.029}$

Table 13: ATRNet-STAR ten-category cross-scene azimuth comparison. Training uses sparse sandstone observations; evaluation uses 923 real images from other scenes at the same 15° depression. Results are mean ± sample standard deviation over five downstream seeds.

	60°		90°		120°	
Method	Acc. (%) ↑	Az. MSE ↓	Acc. (%) ↑	Az. MSE ↓	Acc. (%) ↑	Az. MSE ↓
Sparse only	45.16 ± 1.19	0.596 ± 0.039	36.68 ± 3.21	0.547 ± 0.035	47.78 ± 1.10	0.697 ± 0.041
Repeated sparse	44.38 ± 3.27	0.565 ± 0.016	36.21 ± 1.74	0.581 ± 0.022	41.76 ± 1.73	0.789 ± 0.028
ACGAN	36.36 ± 1.30	0.661 ± 0.032	30.57 ± 1.43	0.655 ± 0.033	9.84 ± 2.60	$\mathbf{0.500 \pm 0.000}$
StyleGAN	42.73 ± 2.06	0.590 ± 0.032	35.95 ± 1.29	0.625 ± 0.033	30.12 ± 1.91	0.756 ± 0.035
Φ-GAN	40.28 ± 2.60	0.591 ± 0.032	30.62 ± 2.50	0.609 ± 0.040	15.47 ± 13.73	0.514 ± 0.032
cDDPM	38.79 ± 2.45	0.569 ± 0.032	25.29 ± 3.23	0.715 ± 0.028	29.25 ± 1.56	0.843 ± 0.035
LDM	33.17 ± 1.93	0.603 ± 0.021	30.34 ± 3.51	0.599 ± 0.023	27.11 ± 2.49	0.638 ± 0.007
FM	40.65 ± 1.50	0.648 ± 0.018	31.42 ± 1.50	0.615 ± 0.034	28.47 ± 3.75	0.669 ± 0.038
SAR-NV	29.62 ± 2.91	0.773 ± 0.032	26.22 ± 1.28	0.761 ± 0.033	39.61 ± 1.66	0.724 ± 0.017
Φ-Flow	$\mathbf{57.68 \pm 2.27}$	$\mathbf{0.524 \pm 0.028}$	$\mathbf{49.86 \pm 3.26}$	$\mathbf{0.509 \pm 0.021}$	$\mathbf{50.16 \pm 3.20}$	0.604 ± 0.035

Table 14: ATRNet-STAR ten-category generation on withheld depression planes. Both source planes use 60° azimuth spacing. Each task uses 120 sparse training images, 3,600 synthetic images in augmented runs, and 120 real query-plane test images. MSEs use sine/cosine encodings, not squared degrees. Results are mean ± sample standard deviation over five downstream seeds.

Method	Acc. (%) ↑	Az. MSE ↓	Dep. MSE ↓
Interpolation: observed 15 and 45; query 30°			
Sparse only	55.50 ± 5.26	0.503 ± 0.027	0.0181 ± 0.0060
Repeated sparse	56.17 ± 3.52	0.575 ± 0.052	0.0267 ± 0.0022
ACGAN	50.33 ± 3.61	0.573 ± 0.032	0.0005 ± 0.0003
StyleGAN	50.83 ± 5.14	0.588 ± 0.017	0.0003 ± 0.0001
Φ-GAN	56.17 ± 4.27	0.579 ± 0.047	0.0002 ± 0.0001
cDDPM	57.50 ± 4.68	0.572 ± 0.049	0.0024 ± 0.0027
LDM	51.17 ± 6.63	0.581 ± 0.044	$\mathbf{0.0001 \pm 0.0001}$
FM	49.67 ± 0.95	0.602 ± 0.015	0.0003 ± 0.0003
SAR-NV	55.50 ± 4.23	0.721 ± 0.047	0.0003 ± 0.0001
Φ-Flow	$\mathbf{74.33 \pm 3.20}$	$\mathbf{0.487 \pm 0.037}$	0.0004 ± 0.0004
Extrapolation: observed 15 and 30; query 45°			
Sparse only	47.50 ± 3.12	$\mathbf{0.404 \pm 0.043}$	0.0748 ± 0.0144
Repeated sparse	52.00 ± 4.11	0.472 ± 0.031	0.0564 ± 0.0138
ACGAN	44.17 ± 5.03	0.552 ± 0.048	0.0036 ± 0.0019
StyleGAN	49.67 ± 1.92	0.536 ± 0.022	0.0029 ± 0.0009
Φ-GAN	46.50 ± 3.70	0.557 ± 0.052	0.0038 ± 0.0038
cDDPM	49.67 ± 2.67	0.614 ± 0.048	0.0018 ± 0.0011
LDM	40.33 ± 4.55	0.574 ± 0.040	$\mathbf{0.0007 \pm 0.0004}$
FM	44.67 ± 3.36	0.537 ± 0.024	0.0015 ± 0.0011
SAR-NV	45.33 ± 2.47	0.558 ± 0.020	0.0083 ± 0.0048
Φ-Flow	$\mathbf{59.83 \pm 2.85}$	0.448 ± 0.038	0.0062 ± 0.0037
Extrapolation: observed 30 and 45; query 15°			
Sparse only	44.67 ± 5.55	0.513 ± 0.028	0.0676 ± 0.0121
Repeated sparse	49.83 ± 4.98	0.507 ± 0.018	0.0570 ± 0.0120
ACGAN	45.17 ± 3.84	0.597 ± 0.033	0.0152 ± 0.0060
StyleGAN	44.50 ± 3.36	0.622 ± 0.040	0.0070 ± 0.0025
Φ-GAN	42.50 ± 3.49	0.589 ± 0.051	0.0080 ± 0.0031
cDDPM	51.67 ± 3.82	0.608 ± 0.038	0.0192 ± 0.0105
LDM	37.17 ± 3.36	0.627 ± 0.058	0.0065 ± 0.0038
FM	46.67 ± 2.70	0.539 ± 0.029	0.0094 ± 0.0060
SAR-NV	50.00 ± 3.12	0.680 ± 0.034	$\mathbf{0.0029 \pm 0.0022}$
Φ-Flow	$\mathbf{64.83 \pm 1.90}$	$\mathbf{0.505 \pm 0.042}$	0.0167 ± 0.0139

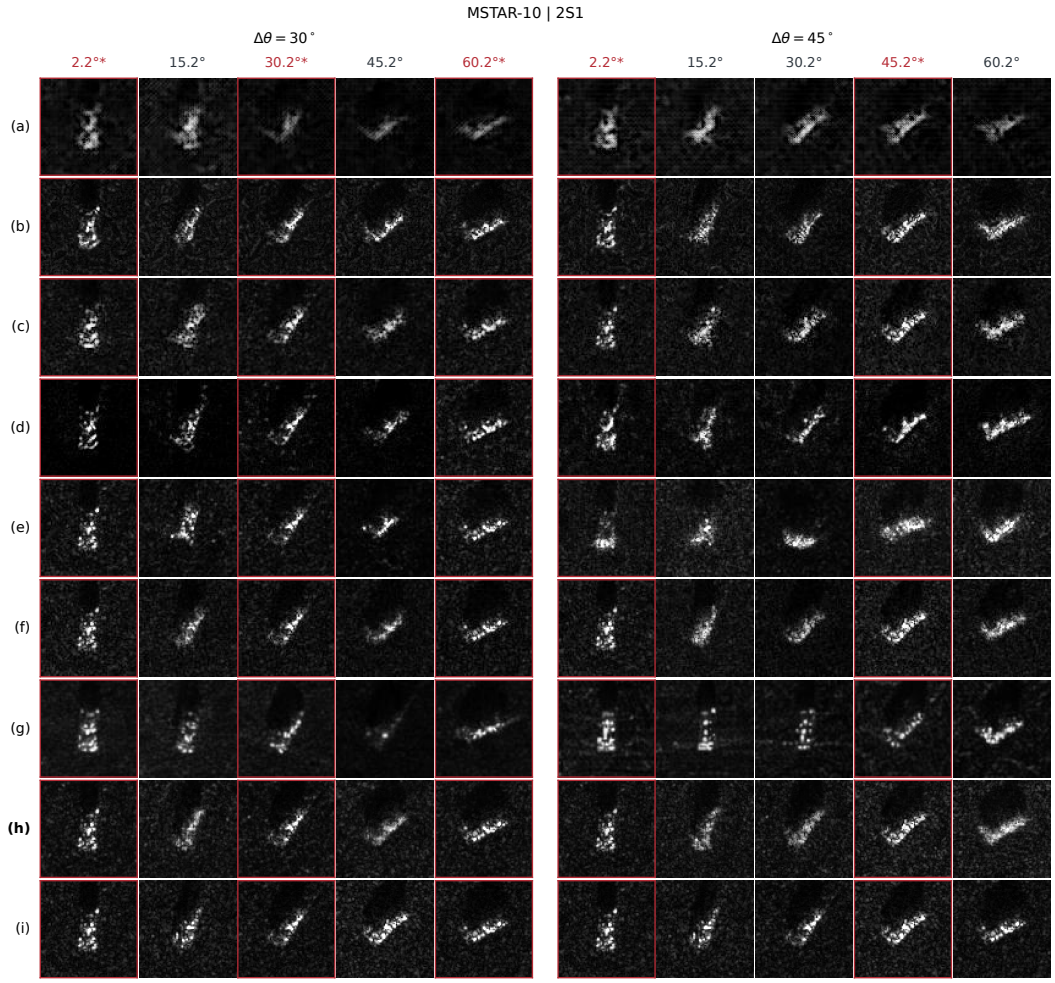


Figure 11: MSTAR generation with 30° (left) and 45° (right) azimuth spacing. Rows: (a) ACGAN; (b) StyleGAN; (c) Φ-GAN; (d) cDDPM; (e) LDM; (f) FM; (g) SAR-NV; (h) Φ-Flow; (i) real target. Red frames and asterisks mark observed training views; the remaining columns are unobserved views.

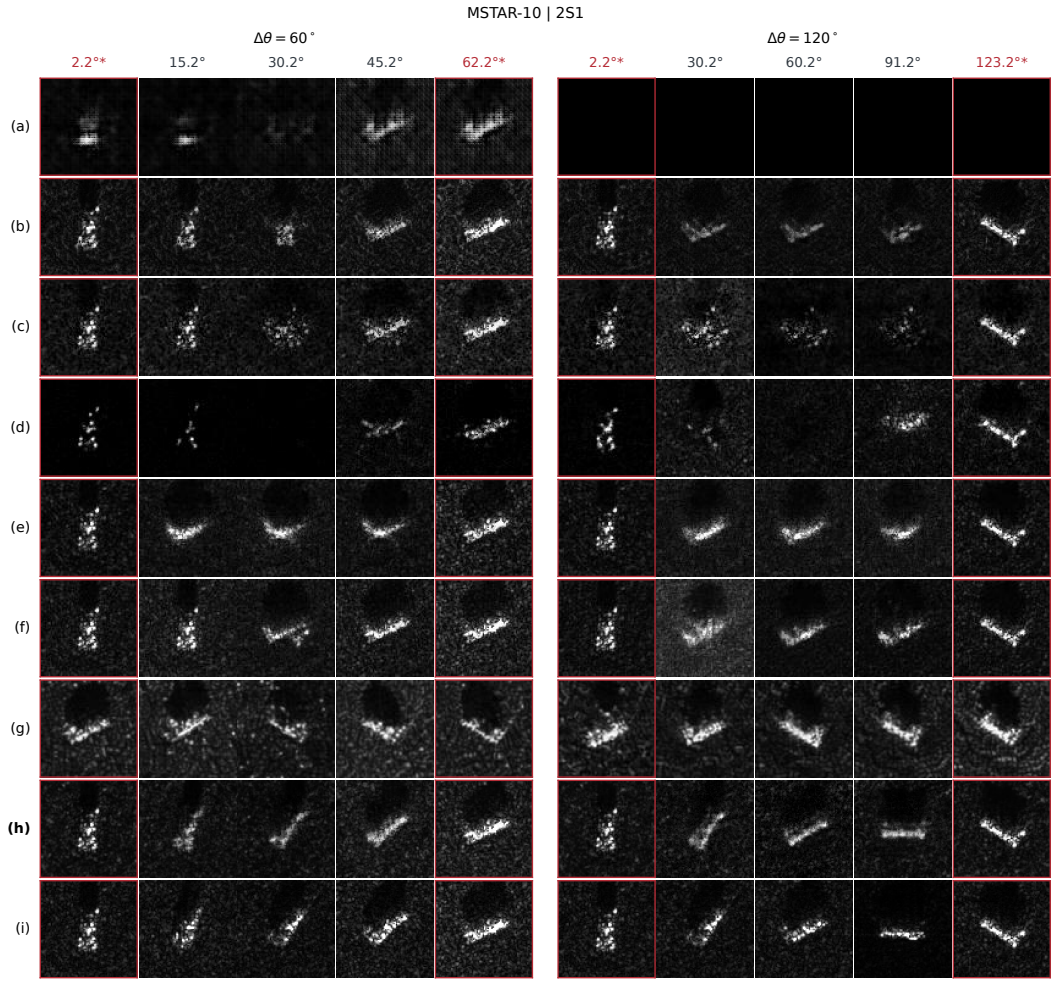


Figure 12: MSTAR generation with 60° (left) and 120° (right) azimuth spacing. Rows: (a) ACGAN; (b) StyleGAN; (c) Φ-GAN; (d) cDDPM; (e) LDM; (f) FM; (g) SAR-NV; (h) Φ-Flow; (i) real target. Red frames and asterisks mark observed training views.

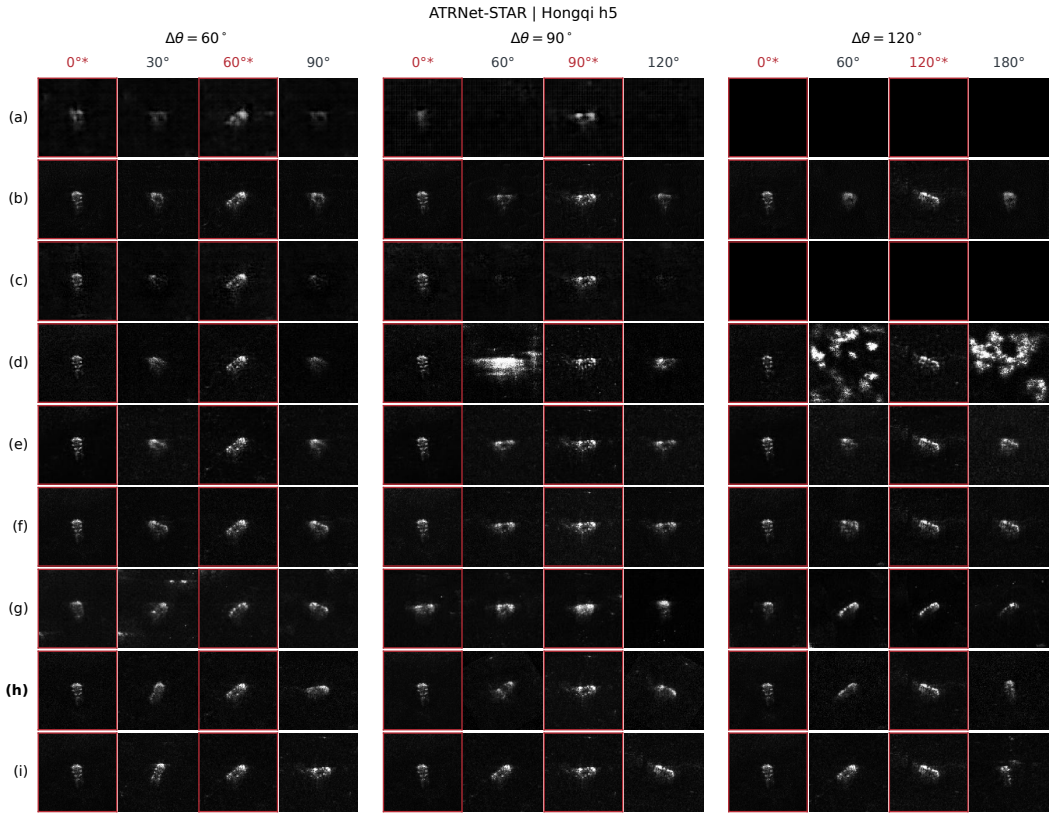


Figure 13: ATRNet-STAR generation with 60, 90, and 120° azimuth spacing from left to right. Real references are sandstone images at 15° depression, matching the generation setting. Rows: (a) ACGAN; (b) StyleGAN; (c) Φ-GAN; (d) cDDPM; (e) LDM; (f) FM; (g) SAR-NV; (h) Φ-Flow; (i) real target. Red frames and asterisks mark observed training views.