SCIENCE CHINA
Information Sciences

• LETTER •

August 2026, Vol. 69, Iss. 8, 180307:1–180307:2
https://doi.org/10.1007/s11432-025-5025-1

Special Topic: Large Model-Driven Aerospace Microwave Imaging

GeoDiff-SAR: a geometric prior guided diffusion model for SAR image generation

Fan ZHANG[1], Xuanting WU[1], Fei MA[1*], Qiang YIN[1] & Yuxin HU[2]

[1]*College of Information Science and Technology, Beijing University of Chemical Technology, Beijing 100029, China*
[2]*Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100190, China*

Received 30 December 2025/Revised 21 April 2026/Accepted 26 June 2026/Published online 27 July 2026

Citation Zhang F, Wu X T, Ma F, et al. GeoDiff-SAR: a geometric prior guided diffusion model for SAR image generation. Sci China Inf Sci, 2026, 69(8): 180307, https://doi.org/10.1007/s11432-025-5025-1

Synthetic aperture radar (SAR) image generation is increasingly used to mitigate data scarcity, but controllable generation under sparse observation angles poses a challenge. Recent studies on SAR image generation have improved texture realism, yet explicit geometry-aware control is still limited [1, 2]. This issue is particularly severe for aircraft SAR, where changes in the azimuth affect the layover, shadow, and dominant scattering centers. In this study, we focus on a narrower yet verifiable claim than that presented in the original full manuscript: three-dimensional (3D) model-derived geometric priors can enable controllable intermediate-azimuth completion under sparse-angle training.

Figure 1 summarizes the revised framework and the supporting evidence retained in this study. Figure 1(a) shows the overall geometry-guided generation pipeline. GeoDiff-SAR receives a structured prompt such as "SAR image, Pilatus PC 12, Ku-band, 0.5 m resolution, VV polarization, azimuth 5°" and retrieves the corresponding 3D model. Measured azimuths that do not coincide with the 5° grid are quantized to the nearest discrete azimuth label for prompting. Figure 1(b) visualizes the multibounce ray-tracing process used to construct the geometric prior. Rather than performing rigorous electromagnetic simulation, the ray-casting branch is designed only to supply geometry-aware scattering cues for generation, thereby offering a practical balance between physical relevance and computational cost [3].

Geometry-guided prior. As shown in Figure 1(b), the geometric branch simulates multibounce scattering cues and occlusion relationships by recursive ray tracing with fixed heuristic parameters that are matched to the current acquisition setting: $\alpha_{\text{edge}} = 2.0$, $\alpha_{\text{vert}} = 2.0$, $\alpha_{\text{struct}} = 1.6$, $\tau_{\text{area}} = 0.01$, $\tau_{\text{vert}} = 0.3$, $\zeta = 0.8$, $\mu = 0.03$, $K_{\max} = 3$, $\tau_{\min} = 0.03$, $\sigma_{\text{scan}} = 0.3$, and $\sigma_{\text{reflect}} = 0.1$. These parameters are manually selected once for the present dataset and then kept fixed across all categories, polarizations, and viewpoints. For a ray state $S_k = (\mathbf{p}_k, \mathbf{d}_k, E_k)$, the scattering intensity is approximated by

$$I_{\text{scatter}} = I_k \mathrm{e}^{-\mu L_{\text{path}}} \mathcal{W}_{\text{base}} \mathcal{H}_{\text{edge}} \mathcal{H}_{\text{orient}} \mathcal{H}_{\text{struct}},$$

where the four factors in the above equation encode the base response, edge enhancement, side-looking orientation, and structural scattering, as indicated by the subscripts. The reflected direction is then updated as follows:

$$\mathbf{d}_{k+1} = \text{Normalize}(\mathbf{d}_{\text{spec}} + \zeta \mathbf{u}), \quad \mathbf{u} \sim \mathcal{U}(-1, 1)^3,$$

which injects a roughness-controlled diffuse component. The resulting point cloud is encoded and fused with the native SD3.5 text-conditioning branches, namely, one T5 encoder and two CLIP encoders. LoRA is used to adapt stable diffusion 3.5 medium to the SAR domain while keeping the backbone largely frozen [4, 5]. No paired SAR image is required during inference. In other words, the prior branch is not a substitute for full electromagnetic reconstruction; rather, it is a geometry-aware condition generator tailored to sparse-angle controllable synthesis.

Condition fusion. The multimodal condition is formed by adaptive weighted fusion,

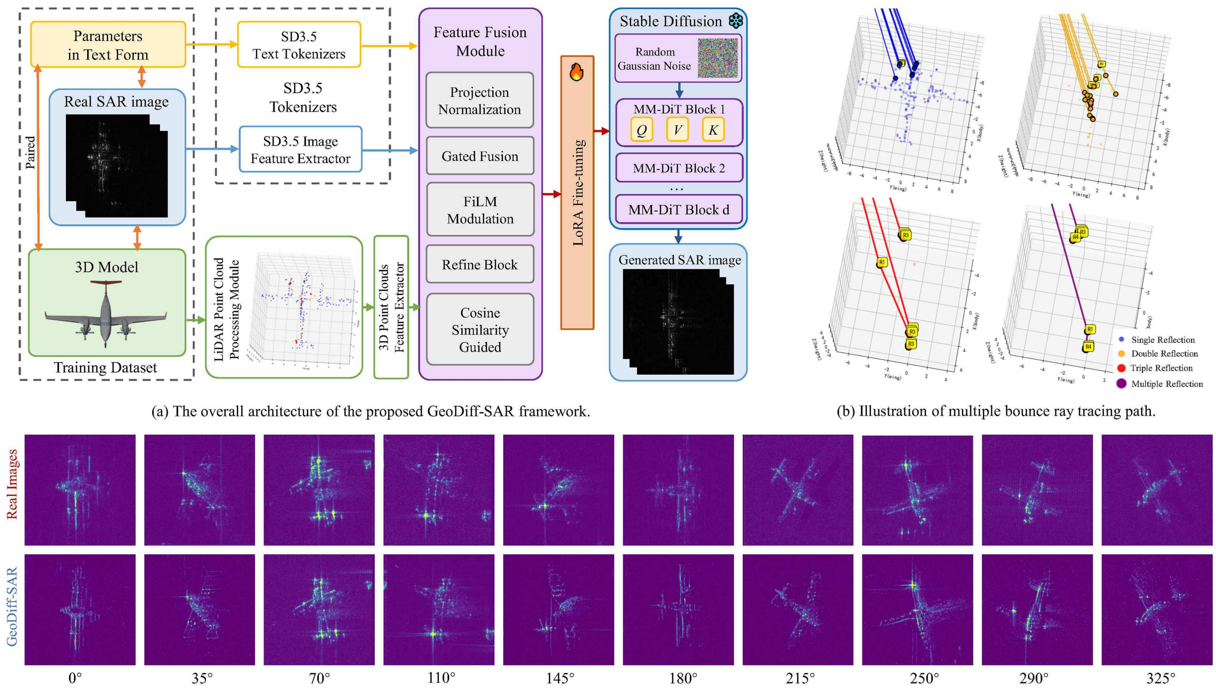
$$C_{\text{fused}} = \alpha_t \tilde{F}_t + \alpha_p \tilde{F}_p + \alpha_i \tilde{F}_i,$$

where $\tilde{F}_t$, $\tilde{F}_p$, and $\tilde{F}_i$ denote the text, geometric, and auxiliary image features, and $\alpha_t + \alpha_p + \alpha_i = 1$. Here, $\tilde{F}_i$ is used only as an auxiliary visual anchor during training to stabilize texture-semantic alignment. During inference, because no paired SAR image is available, generation is driven by the text and geometric branches; this is consistent with the intended text-plus-3D conditioning scenario. The diffusion model is optimized by the standard noise-prediction objective.

$$\mathcal{L}_{\text{simple}} = \mathbb{E}\left[\|\epsilon - \epsilon_\theta(z_t, t, C_{\text{fused}})\|_2^2\right].$$

Dataset and protocol. The real SAR aircraft dataset contains 8536 images from 4 aircraft categories, namely, Cessna 208 (3468), Kodiak 100 (3468), King Air 350i (800), and Pilatus PC 12 (800), with 4 polarizations and full azimuth coverage from 0° to 355° at a 5° interval. The training split contained only 1870 samples at 10° intervals, while the 6666-image test split retained the full 5° sampling. Therefore, GeoDiff-SAR was evaluated on intermediate-azimuth completion: the model was trained with $0°, 10°, 20°, \ldots$ and asked to generate the missing $5°, 15°, 25°, \ldots$ views. All quantitative results were computed by comparing generated samples against held-out real SAR images under identical

* Corresponding author (email: mafei@mail.buct.edu.cn)



(a) The overall architecture of the proposed GeoDiff-SAR framework.

(b) Illustration of multiple bounce ray tracing path.

(c) Visual comparison between real SAR samples (top row) and images generated by GeoDiff-SAR (bottom row) across ten distinct azimuth angles.

Figure 1 (Color online) Overview and visual evidence of GeoDiff-SAR. (a) Overall geometry-guided generation framework, where a 3D model-derived point-cloud prior is fused with textual conditions to guide diffusion-based SAR image generation; (b) illustration of the multibounce ray-tracing paths used to construct the geometric prior; (c) visual comparison between real (top) and GeoDiff-SAR (bottom) generated SAR images across representative azimuths, showing that the generated images preserve view-dependent structures and dominant scattering patterns under sparse-angle training.

semantic conditions. To partially address public-benchmark generality, we further conducted a lightweight sanity check on MSTAR using five vehicle classes (2S1, BMP2, BTR70, D7, and T72) under fixed 17° depression and HH polarization. This public split follows the same sparse-angle protocol and contains 677 training samples at 10° intervals and 1581 test samples at 5° intervals, without any changes in the geometric-prior parameters. Representative visual examples on the aircraft set are shown in Figure 1(c), where the real and generated samples exhibit consistent view-dependent target structure.

Evaluation. As perceptual metrics developed for natural images are not ideally suited to SAR imagery, we place greater emphasis on the structural similarity index measure (SSIM) and azimuth consistency than on metrics such as FID and LPIPS. Azimuth consistency is defined by

$$\mathrm{Acc}_{\mathrm{azi}} = \frac{1}{N}\sum_{n=1}^{N}\mathbf{1}[\hat{y}_n = y_n],$$

where y_n is the commanded azimuth label, and $\hat{y}_n$ is the label predicted by a held-out azimuth evaluator trained only on real SAR images. On the private aircraft set, GeoDiff-SAR achieves SSIM $= 0.812$ and azimuth consistency $= 0.940$, compared with 0.738 and 0.782, respectively, for SD3.5m without the 3D geometric prior. On the public MSTAR sanity check, the same framework reaches SSIM $= 0.878$ and azimuth consistency $= 0.917$, exceeding the baseline values of 0.733 and 0.804, respectively. Together with the real-versus-generated comparisons shown in Figure 1(c), these gains indicate that the injected prior improves viewpoint adherence rather than only SAR-like texture synthesis.

Sensitivity and efficiency. The results of a minimal sensitivity study further support the selected setting. With $\zeta = 0.8$, varying $K_{\max} \in \{1, 2, 3\}$ yields (SSIM, azimuth consistency) of $(0.736, 0.824)$, $(0.784, 0.866)$, and $(0.812, 0.940)$. With $K_{\max} = 3$, varying $\zeta \in \{0, 0.4, 0.8\}$ yields $(0.739, 0.871)$, $(0.758, 0.902)$, and $(0.812, 0.940)$. Therefore, the final configuration $K_{\max} = 3$ and $\zeta = 0.8$ is retained. Averaged over 100 trials, prior construction takes about 48 s per azimuth on CPU, whereas diffusion inference takes about 13 s per image on an RTX 4090 with 30 sampling steps. Hence, the main overhead lies in geometric-prior construction.

Discussion. We deliberately avoided stronger claims such as arbitrary continuous-angle synthesis or strict physical reconstruction. The current evidence supports a narrower conclusion: 3D model-derived geometric priors can guide SAR image generation under sparse-angle training and substantially improve controllable intermediate-azimuth completion. The prior branch remains heuristic and should be interpreted as a lightweight guidance mechanism rather than a high-fidelity EM solver. Moreover, although the main high-resolution aircraft dataset is not publicly available, the added MSTAR sanity check shows that the same pipeline retains clear gains on a public benchmark. Broader cross-sensor and cross-scene validation will be explored in future work.

Acknowledgements This work was supported by National Natural Science Foundation of China (Grant Nos. 62271034, 62331026) and Natural Science Foundation of Shandong Province (Grant No. ZR2024ZD19).

References

1 Huang Z, Zhang X, Tang Z, et al. Generative artificial intelligence meets synthetic aperture radar: a survey. IEEE Geosci Remote Sens Mag, 2026, 14: 6–48
2 Zhang X, Zhuang Y, Guo Q, et al. Ph-GAN: physics-inspired GAN for generating SAR images under limited data. In: Proceedings of IEEE ICCV, 2025. 29075–29085
3 Auer S, Hinz S, Bamler R. Ray-tracing simulation techniques for understanding high-resolution SAR images. IEEE Trans Geosci Remote Sens, 2010, 48: 1445–1456
4 Esser P, Kulal S, Lang A, et al. Scaling rectified flow transformers for high-resolution image synthesis. In: Proceedings of ICML, 2024
5 Hu E J, Shen Y, Wallis P, et al. LoRA: low-rank adaptation of large language models. In: Proceedings of ICLR, 2022