SAR-DisentDM: A Semantic-Disentangled Diffusion Model for Limited-Data SAR Image Synthesis

Yue Yang[1], Song Tang[2], Qijun Zhao[1,2*], Hailun Zhang[1], Xiwen Wang[1], Zijian Deng[1]

[1]College of Computer Science, Sichuan University, Chengdu, China
[2]National Key Laboratory of Fundamental Science on Synthetic Vision, Sichuan University, Chengdu, China
yueyang7@scu.edu.cn, tangsong@stu.scu.edu.cn, qjzhao@scu.edu.cn, tamakoko@stu.scu.edu.cn,
wangxiwen124@stu.scu.edu.cn, dengzijian@stu.scu.edu.cn

Abstract

The high cost of synthetic aperture radar (SAR) data acquisition motivates SAR image generation research. However, the data scarcity and SAR's inherent azimuth sensitivity make generative models suffer from severe azimuth overfitting. Most existing methods require supplementary data to work effectively, limiting their practicality. In this paper, we propose SAR-DisentDM, a novel semantic-disentangled diffusion model for limited-data SAR image generation, without requiring any auxiliary resources. We develop a physics-aware diffusion architecture that explicitly models semantic knowledge of SAR images, including intrinsic characteristics, contextual diversity, and measurement randomness. A key innovation is the attention-guided semantic disentanglement (AGSD) module, designed to decouple category-specific features from azimuth-variable scattering patterns. This is achieved by aid of a dual disentangled loss with time-step-adaptive optimization. Furthermore, we introduce an azimuth angle perturbation augmentation (AAPA) mechanism, to enhance the model's robustness to minor azimuth angle errors. Extensive evaluations validate that SAR-DisentDM enables controllable SAR image synthesis with designated attributes, significantly improving representation and generalization abilities under limited data. Synthetic imagery from our approach boosts automatic target recognition (ATR) accuracy beyond state-of-the-art methods.

Introduction

Synthetic aperture radar (SAR) enables all-weather, day-night imaging, driving its use in automatic target recognition (ATR) for surveillance, reconnaissance, urban management, and disaster assessment (Curlander and McDonough 1991; Joo, Sungmin, and Taesup 2019; Goldberg 2024; Yang et al. 2025). While deep learning has advanced SAR ATR performance (Li et al. 2022; Chen et al. 2022; Tan et al. 2025; Wang et al. 2025; Liu et al. 2025), fundamental data scarcity remains a major constraint. As depicted in Fig. 1(a), SAR data acquisition is costly and time-intensive, resulting in sparse observation apertures and scarce imagery. Crucially, SAR targets exhibit unique electromagnetic (EM) scattering characteristics that change significantly with the azimuth angle. This leads to a pronounced discrepancy between the actual target rotation and its corresponding rotation in the SAR image, as shown in Fig. 1(b). Consequently, conventional geometric augmentations (e.g., rotation) are often ineffective for SAR image.

Generative data augmentation offers a promising solution. Deep generative models (DGMs) like generative adversarial networks (GANs) (Goodfellow et al. 2014; Odena et al. 2017; Zhu et al. 2023) can synthesize SAR images and potentially generate samples of unseen azimuth angles. However, the strong azimuth sensitivity of SAR images makes it difficult to learn both azimuth-invariant representations and azimuth-dependent rules. This issue is exacerbated by data scarcity. With limited training samples, the model may incorrectly associate a target's scattering features with the specific azimuth angles contained in the small training set, leading to severe azimuth overfitting. Existing SAR image generation methods primarily employ GANs, and rely on costly auxiliary data such as semantic maps or simulated SAR data (Song et al. 2021; Sun et al. 2023; Bao et al. 2024), limiting their practical deployment.

Diffusion models (DMs) (Ho et al. 2020; Nichol and Dhariwal 2021; Dhariwal and Nichol 2021; Rombach et al. 2022) recently surpass GANs in training stability and conditioning flexibility. Inspired by this, we propose SAR-DisentDM, a novel semantic-disentangled diffusion model, requiring no auxiliary resources. Specifically, we employ a conditioned latent diffusion model (LDM) (Rombach et al. 2022) as our backbone, and explicitly integrate key SAR semantic information, including intrinsic characteristics, contextual diversity and measurement randomness, into the framework. Then an attention-guided semantic disentanglement (AGSD) module is proposed to decouple category-specific features from azimuth-variable patterns. This is achieved by aid of a dual disentangled loss with time-step-adaptive optimization strategy. Moreover, to enhance model's tolerance for minor angular value uncertainties arising from measurement errors or other factors, we introduce an azimuth angle perturbation augmentation (AAPA) mechanism in conditioning. As Fig. 1(c)-(f) illustrate, SAR-DisentDM enables precise and controllable attribute-specified image synthesis. Extensive experiments confirm that our synthetic images boost ATR accuracy.

The contributions of this work are summarized as follows:

*Corresponding author.

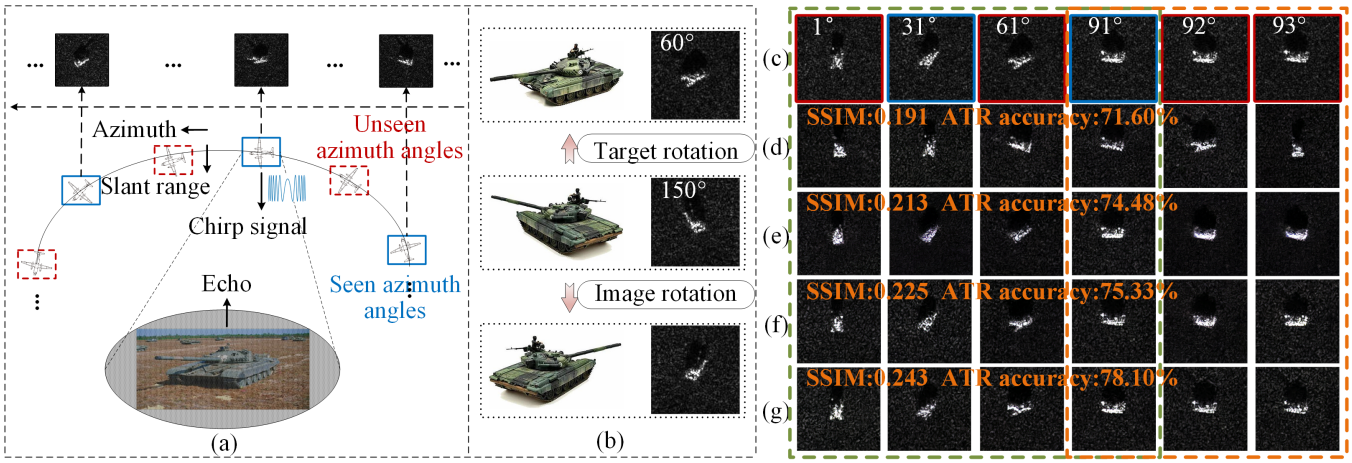

Figure 1: Addressing azimuth-controllable SAR image generation under data scarcity. (a) Goal: generating SAR images at unseen azimuth angles using limited data. (b) SAR-specific challenge: target rotation $\neq$ SAR image rotation due to EM scattering characteristics. (c)-(f) Method comparison: (c) Ground-truth SAR images (red boxes: unseen angles; blue boxes: seen angles); (d) ACGAN (Odena et al. 2017): fails to provide azimuth control, leading to incorrect categories/angles; (e) AGGAN (Sun et al. 2023) & (f) LDM (Rombach et al. 2022): realize basic azimuth control but obtain inaccurate categories/angles; (g) SAR-DisentDM (ours): reduces azimuth overfitting, achieving precise category and azimuth angle control.

(1) We propose a novel method termed SAR-DisentDM to tackle an untouched problem, i.e., semantic disentanglement for SAR image with limited data. To the best of our knowledge, this work is the first study on this problem. (2) A novel attention-guided semantic disentanglement (AGSD) module is presented to solve azimuth-target coupling. (3) We develop an azimuth angle perturbation augmentation (AAPA) mechanism to further combat azimuth overfitting and enhance model's robustness. (4) Extensive validation shows that our synthetic images yield significant SAR-ATR accuracy improvements, outperforming previous benchmarks.

Related Work

Synthetic Image Generation

Advancements in DGMs have enabled high-quality synthetic image generation. While GANs produce high-fidelity images (Goodfellow et al. 2014; Odena et al. 2017; Zhu et al. 2023), they suffer from unstable training and mode collapse under limited data. DMs (Ho et al. 2020; Nichol and Dhariwal 2021; Dhariwal and Nichol 2021; Rombach et al. 2022) exceed GANs in training stability and conditioning flexibility, achieving impressive results in many generative tasks, particularly in LDM (Rombach et al. 2022). However, current DM applications mainly focus on generating synthetic human faces (Kim et al. 2023; Melzi et al. 2023; Boutros et al. 2023; Xu et al. 2024; Otroshi and Marcel 2025). For instance, IDiff-Face (Boutros et al. 2023) employs a DM backbone with contextual partial dropout (CPD) on identity embeddings to enhance diversity, though at the cost of identity fidelity. Later ID^3 (Xu et al. 2024) utilizes an identity-preserving loss to balance appearance diversity with intra-class consistency. Despite these successes, applying such advanced generative techniques to SAR data remains challenging. SAR's intrinsic azimuth-sensitive scattering patterns differ from the human-interpretable features of natural images. This highlights the urgent need for SAR-specific generative methods capable of learning both azimuth-invariant representations and azimuth-dependent scattering patterns.

SAR Image Generation

Early SAR image generation relied on numerical simulation or template-based methods, which either incurred high computational costs or failed to model complex scenes (Ding et al. 2017; Yang, Zhao, and Wan 2024). Recent DGM-based approaches leverage their ability to learn complex SAR data distributions. Certain researches have focused on enhancing GAN architectures, aiming to reduce complexity or improve feature representation (Hu et al. 2021; Shi et al. 2022; Zou et al. 2020; Cao et al. 2022; Wang et al. 2022a). Another key direction involves employing SAR-specific properties, such as incorporating statistical priors to suppress speckle artifacts (Guo et al. 2017; Wang et al. 2018; Peng et al. 2024), employing SAR frequency domain features to prevent low-frequency interference (Ying et al. 2025), or integrating physical models (e.g., causal or scattering models) to enforce physical consistency (Guo, Xu, and Xu 2023; Huang et al. 2024; Zhang et al. 2025). However, performance degrades significantly in low-data regimes due to azimuth overfitting. Recent efforts aim to address data scarcity via semantic map guidance (requiring costly annotations) (Song et al. 2021), meta-optimized episodic training (needing abundant base classes) (Sun et al. 2023), and cross-modal translation (dependent on multi-modal fusion) (Bao et al. 2024). Nevertheless, these methods' dependence on substantial supplementary data limits their practical utility. In this paper, we introduce the semantic diffusion-based SAR generation framework. Our approach disentangles azimuth-target semantics and incorpo-

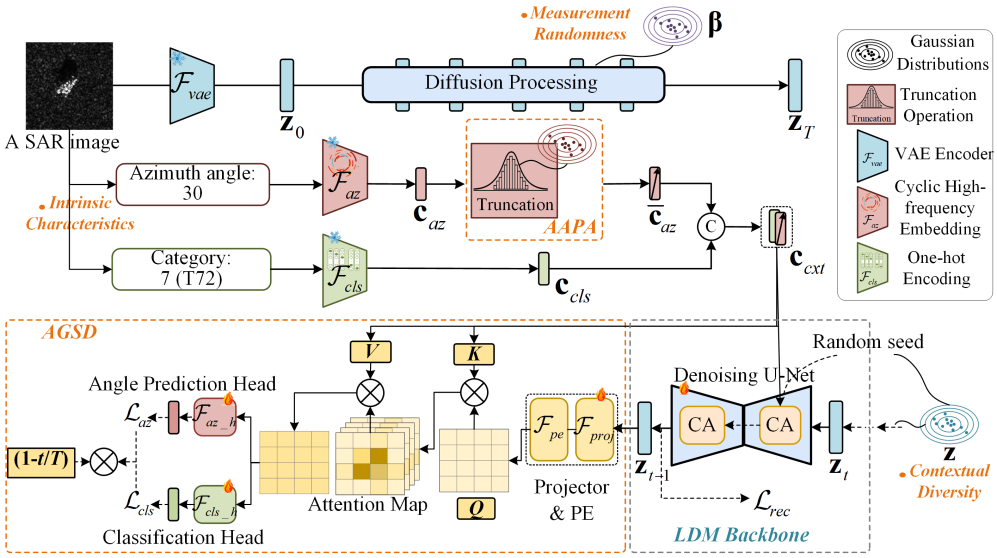

Figure 2: The pipeline of our SAR-DisentDM for SAR image synthesis with limited data. The LDM backbone is leveraged to incorporate semantic knowledge of SAR images: intrinsic characteristics (including category and azimuth angle) injected as conditions, contextual diversity enhanced through stochastic sampling, and measurement randomness modeled via diffusion noise scheduling. AGSD decouples category-specific and azimuth-sensitive features within the latent space. AAPA improves model's robustness to minor errors of angular values.

rates azimuth perturbation augmentation, effectively overcoming the aforementioned limitations without auxiliary resources.

Methodology

SAR-DisentDM aims to generate diverse SAR images with category consistency using limited training samples. As Fig. 2 shows, our framework employs an LDM backbone incorporating SAR semantic representations. During training, the model is fed with SAR images, paired with their ground-truth azimuth angles and categories. For inference, only target azimuth angles and categories are required to generate designated SAR images. To solve azimuth-target coupling, we propose AGSD, which enforces attention separation between target identity and angular features. Furthermore, we introduce AAPA, which promotes model's robustness to small azimuth uncertainties.

Preliminaries

Denoising diffusion probabilistic model (DDPM) (Ho et al. 2020) is a typical DM designed to learn data distribution by variational inference based on Markov chain. One could add Gaussian noise to the input data during forward process while mitigating the noise by training a denoising network during reverse process. In the forward (diffusion) procedure, supposing a clean SAR image $\mathbf{x}_0$, we obtain a series of noise samples $\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_T$, i.e.,

$$\mathbf{x}_t = \sqrt{\alpha_t}\mathbf{x}_{t-1} + \sigma_t \boldsymbol{\epsilon}, \tag{1}$$

for $\forall t \in \{1, 2, ...T\}$, where T is the number of diffusion time-steps, $\boldsymbol{\epsilon}$ follows a normal distribution, α_t and σ_t control the noise strength with $\sigma_t = \sqrt{1-\alpha_t}$.

The noisy SAR image $\mathbf{x}_t$ is denoised in the reverse (denoising) stage by training a model $\boldsymbol{\epsilon}_\theta(\mathbf{x}_t, t)$, mathematically formulated as

$$\mathbf{x}_{t-1} = \frac{1}{\sqrt{\alpha_t}}(\mathbf{x}_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}\boldsymbol{\epsilon}_\theta(\mathbf{x}_t, t)) + \sigma_t \mathbf{g}, \tag{2}$$

where $\bar{\alpha}_t = \prod_{s=1}^{t} \alpha_s$, and $\mathbf{g}$ obeys $\mathcal{N}(\mathbf{0}, \mathbf{I})$. A U-Net structure (Rombach et al. 2022) is employed for the denoising network $\boldsymbol{\epsilon}_\theta(\mathbf{x}_t, t)$ to predict $\boldsymbol{\epsilon}$, by minimizing loss function:

$$\mathcal{L}_{ddpm} = \mathbb{E}_{\mathbf{x}_t, t, \boldsymbol{\epsilon}\sim\mathcal{N}(0,1)}\left[\|\boldsymbol{\epsilon}_\theta(\mathbf{x}_t, t) - \boldsymbol{\epsilon}\|_2^2\right]. \tag{3}$$

Semantic Knowledge Latent Diffusion Modeling

SAR-DisentDM builds upon an LDM and operates in the compressed latent space of a pre-trained VAE (denoted as $\mathcal{F}_{vae}$) (Rombach et al. 2022). The input image $\mathbf{x}_0$ is encoded as latent representation $\mathbf{z}_0 = \mathcal{F}_{vae}(\mathbf{x}_0)$, mitigating dimensionality challenges under limited data. SAR semantic knowledge is embedded through three components. (i) Intrinsic characteristics: essential attributes including target azimuth $\mathbf{c}_{az}$ and category $\mathbf{c}_{cls}$, representing the primary Information of Interest (IoI) for SAR ATR. They are injected as conditions via cross-attention (CA). (ii) Contextual diversity: variations from target deformations and observation scenes, causing minor imaging fluctuations within a class. This is modeled through stochastic latent sampling during generation, namely $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$, denoting random seeds and Gaussian initial noise. (iii) Measurement randomness: inherent uncertainty from sensor noise, measurement and imaging errors (Yang, Ma, and Yang 2026). This is captured by the noise scheduling vector $\boldsymbol{\beta} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ within the diffusion process.

Implementation details: we respectively imply one-hot encoding for $\mathbf{c}_{cls}$ and cyclic high-frequency embedding (CHE) for $\mathbf{c}_{az}$ due to its periodicity $\theta \in [0, 2\pi]$ (Guo, Xu, and Xu 2023), expressed as

$$\mathbf{c}_{az} = [\sin(\theta), \cos(\theta), ..., \sin(L\theta), \cos(L\theta)], \quad (4)$$

where L is set as 5 to balance feature dimensionality and representational capacity. By concatenating $\mathbf{c}_{az}$ and $\mathbf{c}_{cls}$ as $\mathbf{c}_{cxt}$, the conditioned denoising U-Net $\hat{\boldsymbol{\epsilon}}_\theta(\mathbf{z}_t, t, \mathbf{c}_{cxt})$ executes reverse sampling:

$$\mathbf{z}_{t-1} = \frac{1}{\sqrt{\alpha_t}}(\mathbf{z}_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}\hat{\boldsymbol{\epsilon}}_\theta(\mathbf{z}_t, t, \mathbf{c}_{cxt})) + \sigma_t \mathbf{g}. \quad (5)$$

Attention-Guided Semantic Disentanglement

To address azimuth-target coupling, model must distinct their relevant components in latent space. This motivates cross-attention, which adaptively weights interactions between $\mathbf{z}_t$ and $\mathbf{c}_{cxt}$. Task-specific supervision further enforces attention to distinct latent aspects. Our proposed AGSD is achieved through the following three core components.

Attention Extraction. A convolutional feature projector and a 2D positional encoding are firstly applied for the latent representation $\mathbf{z}_t$, expressed as $\bar{\mathbf{z}}_t = \mathcal{F}_{pe}(\mathcal{F}_{proj}(\mathbf{z}_t))$. The resulting features are transformed into queries $\mathbf{Q}$, while keys $\mathbf{K}$ and values $\mathbf{V}$ are derived from conditional representations $\mathbf{c}_{cxt}$:

$$\mathbf{Q} = \mathbf{W}_Q \cdot \bar{\mathbf{z}}_t, \mathbf{K} = \mathbf{W}_K \cdot \mathbf{c}_{cxt}, \mathbf{V} = \mathbf{W}_V \cdot \mathbf{c}_{cxt}, \quad (6)$$

where $\mathbf{W}_Q, \mathbf{W}_K, \mathbf{W}_V$ are the learnable projection matrices. The multi-head cross-attention maps at time-step t are computed as

$$\mathbf{A}_t^h = \mathrm{Softmax}\{\mathbf{Q}^h(\mathbf{K}^h)^T/\sqrt{d}\} \quad (h = 1, ..., H), \quad (7)$$

where d is the query/key dimension, H denotes the head number. Attention values are then calculated by concatenating multi-head maps and projecting them:

$$\mathbf{A}_t = \mathrm{Concat}(\mathbf{A}_t^1\mathbf{V}, ..., \mathbf{A}_t^H\mathbf{V})\mathbf{W}_O, \quad (8)$$

where $\mathbf{W}_O$ is the output projection. After LayerNorm and Feed-Forward Networks (FFNs) processing, the attention-refined feature $\hat{\mathbf{z}}$ is yielded. It is then fed into task-disentangled heads, i.e., the classification head and angle prediction head, which are formally defined as

$$\begin{aligned} \hat{\mathbf{c}}_{cls} &= \mathbf{W}_{cls} \cdot \mathrm{AAP}(\hat{\mathbf{z}}) + \mathbf{b}_{cls}, \\ \hat{\mathbf{c}}_{az} &= \mathrm{Tanh}(\mathbf{W}_{az} \cdot \mathrm{AAP}(\hat{\mathbf{z}}) + \mathbf{b}_{az}), \end{aligned} \quad (9)$$

where AAP denotes adaptive average pooling, Tanh activation constrains angle predictions to [-1,1] for normalization. $\mathbf{W}_{cls}, \mathbf{W}_{az}, \mathbf{b}_{cls}, \mathbf{b}_{az}$ are weight matrices and bias vectors.

Dual Disentangled Loss. Classification and angle prediction are supervised using dual disentangled loss:

$$\begin{aligned} \mathcal{L}_{cls} &= \mathrm{CE}(\hat{\mathbf{c}}_{cls}, \mathbf{c}_{cls}^{gt}), \\ \mathcal{L}_{az} &= \mathrm{MSE}(\hat{\mathbf{c}}_{az}, \mathbf{c}_{az}^{gt}), \end{aligned} \quad (10)$$

where CE and MSE are the cross entropy and ℓ_2-norm losses, respectively. $\mathbf{c}_{cls}^{gt}$ and $\mathbf{c}_{az}^{gt}$ represent the ground-truth category and angle, respectively.

Time-Step-Adaptive Optimization. Since azimuth and category are not inferred in early steps, we incorporate a time-step-adaptive weighting scheme for the reconstruction and dual disentangled losses. The total loss is given by

$$\begin{aligned} \mathcal{L}_{total} &= \mathcal{L}_{rec} + (1 - t/T)(\lambda_1\mathcal{L}_{cls} + \lambda_2\mathcal{L}_{az}), \\ \mathcal{L}_{rec} &= \mathbb{E}_{\mathbf{z}_t, t, \mathbf{c}_{cxt}, \boldsymbol{\epsilon}\sim\mathcal{N}(0,1)}\left[\|\hat{\boldsymbol{\epsilon}}_\theta(\mathbf{z}_t, t, \mathbf{c}_{cxt}) - \boldsymbol{\epsilon}\|_2^2\right]. \end{aligned} \quad (11)$$

In Eq. (11), λ_1 and λ_2 are balancing coefficients. The term $(1 - t/T)$ provides a time-step-adaptive weighting that linearly changes from 0 to 1. This loss allows $\hat{\boldsymbol{\epsilon}}_\theta$ to play different roles depending on t. When $t \approx T$, $\hat{\boldsymbol{\epsilon}}_\theta$ is predicting $\mathbf{z}_t$ from random noise, the AGSD plays a minor role ($(1 - t/T) \approx 0$), and we let the model fully exploit the reconstruction loss $\mathcal{L}_{rec}$. Gradually as t increases, the loss progressively guides the denoising trajectory using semantic disentanglement. This strategy exerts granular control over feature formation by modulating the emphasis between reconstruction and disentanglement during denoising.

Azimuth Angle Perturbation Augmentation

Dropout is commonly used in preventing model overfitting (Boutros et al. 2023). However, it risks discarding critical information in low-dimensional azimuth embeddings $\mathbf{c}_{az}$. To address this, we introduce AAPA by applying controlled perturbations to $\mathbf{c}_{az}$. Specifically, AAPA preserves original azimuth information through orthogonal perturbation:

$$\hat{\mathbf{c}}_{az} = \mathbf{c}_{az} \cdot \cos\theta_{\mathcal{T}} + \boldsymbol{\delta} \cdot \sin\theta_{\mathcal{T}}, \quad (12)$$

where $\boldsymbol{\delta}$ is a randomly sampled orthogonal unit vector. The perturbation angle $\theta_{\mathcal{T}}$ follows a truncated Gaussian distribution with mean $\mu = 0$, standard deviation η (controlling intensity), and interval (v_a, v_b) (ensuring physical rationality). The probability density function (PDF) is defined as

$$\begin{aligned} f(x; \mu, \eta, v_a, v_b) =& 1/\eta \cdot \phi\left((x - \mu)/\eta\right) \\ & \left[\Phi\left((v_a - \mu)/\eta\right) - \Phi\left((v_b - \mu)/\eta\right)\right]^{-1}, \end{aligned} \quad (13)$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ are standard normal PDF and cumulative distribution function (CDF). By introducing azimuthal variations via small perturbations, model's robustness to angular errors is improved.

Experimental Setups

Datasets

We evaluate our model on two SAR image datasets: (1) MSTAR 10-class target dataset (AFRL 2016), containing ten military vehicle types captured at 0.3×0.3 m (128×128 pixels). (2) SAR-Airplane database (Wang et al. 2022b), featuring two aircraft types with 72 samples per class (128×128 pixels). MSTAR targets span full azimuth range ($0^\circ \sim 360^\circ$, $1^\circ \sim 2^\circ$ intervals) at 17° and 15° depression angles. SAR-Airplane provides 5° azimuth increments over 0° to 355°. To emulate limited-data scenarios, we uniformly sample subsets from the original datasets based on the target azimuth angles. Specifically, we use azimuth intervals of 10° for MSTAR (17° depression-angle images) and 20° for SAR-Airplane to train the generative models. Once trained, we sample 2746 and 720 synthetic images for MSTAR and SAR-Airplane datasets, respectively.

Methods	MSTAR			SAR-Airplane		
	SSIM↑	FID↓	COSS↑	SSIM↑	FID↓	COSS↑
ACGAN	0.19	588.1	0.54	0.31	926.4	0.26
AGGAN	0.21	485.5	0.62	0.61	810.3	0.34
LDM	0.23	477.8	0.63	0.62	808.7	0.35
IDiff-Face	0.23	472.3	0.63	0.62	793.1	0.37
ID^3	0.23	461.5	0.64	0.63	786.4	0.38
SAR-DisentDM	**0.24**	**433.6**	**0.65**	**0.66**	**725.2**	**0.43**

Table 1: Synthetic image quality assessment: SOTA DGMs vs. SAR-DisentDM on MSTAR and SAR-Airplane datasets.

Implementation Detail

We employ a DM trained for 30000 diffusion steps using a linear noise schedule. Training is performed with a batch size of 32 using float16 precision. Optimization uses Adam with an initial learning rate of $1e^{-4}$. Input images (128×128 pixels) are compressed into a 32×32×3 latent space via a pre-trained VQ-VAE (Razavi et al. 2019). Hyperparameters are set to λ_1=0.17, λ_2=0.5, and perturbation parameters (v_a, v_b)=(-0.2, 0.2), η=0.13.

Experimental Results & Discussion

Representation Capability Assessment

Image Synthesis Ability. We first verify SAR-DisentDM's image synthesis capability. Fig. 3 compares generated samples with real images on MSTAR and SAR-Airplane datasets. The model generates two image groups, demonstrating effective representation of diverse samples. The subtle inter-group variations are both expected and beneficial, as they arise from inherent diffusion stochasticity (noise addition/removal) and random seeds, which are critical mechanisms for enhancing output diversity.

Azimuth Attribute Disentanglement. To assess azimuth attribute disentanglement, we conduct an angle-interpolation experiment. We gradually interpolate azimuth attributes from initial to target angles and generate corresponding samples. Fig. 4 presents interpolated samples for two MSTAR classes (BMP-2: $[92°, 119°]$, BTR-60: $[40°, 67°]$). The results clearly highlight the model's azimuth disentanglement ability. The interpolated samples exhibit high visual similarity to the corresponding target at respective angles, consistent with patterns observed in the real data. Critically, the proposed model effectively prevents overfitting to specific angles, ensuring that the target orientation changes coherently as the azimuth vector is modified.

Class Attribute Disentanglement. A target interpolation experiment is carried out to evaluate class attribute disentanglement ability. Fixing the azimuth angle, we linearly interpolate class embedding from one class (e.g., 2S1) to another, and generate images from the interpolated vectors. Fig. 5 shows two sets of results, where columns represent source class (left), target class (right), and intermediate synthetic images with transformation ratios $[0, 1]$. For example, in Fig. 5(a), as transformation ratio increases, the target size in the synthetic images decreases, and the associated shadow evolves accordingly, while the azimuth information remains fixed at $45°$. Although these generated images represent objects not present in the real world, the model is able to generate convincing and realistic samples. These findings demonstrate that SAR-DisentDM effectively disentangles class attributes while preserving azimuth information.

Synthetic Image Quality Assessment

We evaluate synthetic image quality using three metrics: Structural Similarity Index (SSIM), Frechet Inception Distance (FID) and Cosine Similarity (COSS). Table 1 compares our method against SOTA approaches, including ACGAN (Odena et al. 2017), AGGAN (Sun et al. 2023), LDM (Rombach et al. 2022), IDiff-Face (Boutros et al. 2023) and ID^3 (Xu et al. 2024). ACGAN conditions only on class labels, lacking azimuth control and yielding inferior performance. While AGGAN and LDM incorporate azimuth condition, their naive concatenation of angular and class embeddings entangles geometric and identity attributes, reducing precision in target/azimuth representation. IDiff-Face and ID^3 improve upon previous methods via CPD or enhanced loss strategies, but they fails to resolve fundamental attribute entanglement limitations. Our SAR-DisentDM achieves superior image quality. The designed AGSD and AAPA produce samples with precise target contours and accurate azimuthal transitions. This validates our model's effectiveness in learning complex data distributions from limited samples.

Utility of Synthetic Images for SAR-ATR

SAR-ATR performance is evaluated using synthetic data produced by various generative models. Six recognition architectures are benchmarked: VGG16 (Simonyan and Zisserman 2014), ResNet34 (RN34) (He et al. 2016), ResNet50 (RN50) (He et al. 2016), EfficientNet (EN) (Tan and Le 2019), ConvNext (CN) (Woo et al. 2023), and Vision Transformer (ViT) (Dosovitskiy 2021). These recognition models are pre-trained on synthetic data (for feature learning) and fine-tuned with limited real samples (for adaptation). For fairness, they are evaluated on unified test sets, i.e., $15°$ depression-angle MSTAR data or pre-partitioned SAR-Airplane test data. All ATR models adopt standard configurations, trained with a batch size of 64 for 50 epochs. Performance is reported using overall accuracy (%) indicator.

The experimental results are listed in Table 2. It is observed that ACGAN provides marginal gains over Baseline (trained solely on limited real data), because it lacks azimuth modeling, leading to imbalanced angular distributions and spurious feature correlations. AGGAN and LDM obtain better scores by incorporating both class and angular conditions, but naive attribute concatenation causes feature entanglement, resulting in limited angular diversity and distribution gaps. IDiff-Face suffers from partial information loss with 10-dimensional azimuth embeddings under CPD strategy. ID^3 shows minor gains over IDiff-Face through customized loss design. In contrast, SAR-DisentDM achieves SOTA performance, attaining an average accuracy of 78.10%/93.30% for MSTAR and SAR-Airplane datasets, surpassing ID^3 by +2.05%/+1.09%. This

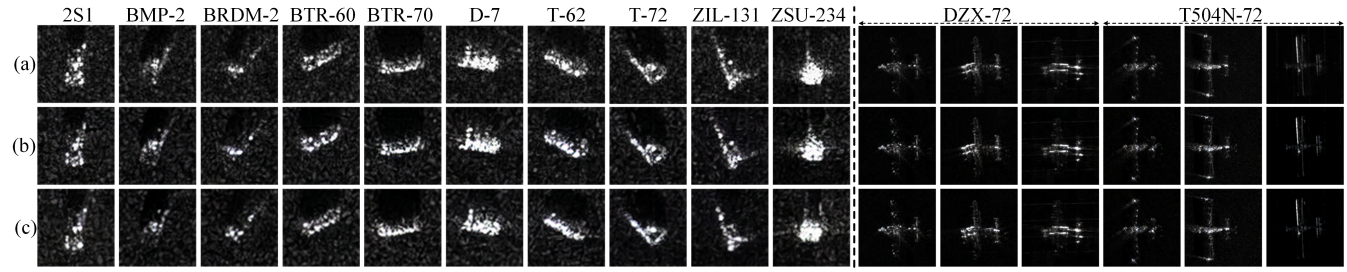

Figure 3: Illustrations of synthetic images on MSTAR and SAR-Airplane datasets. (a) Real images. (b)-(c) Our synthetic images (round #1 & round #2). Both image groups are synthesized with identical intrinsic attributes but varied contextual diversity and measurement randomness. The generated images accurately depict the targets, and subtle differences appear in the background texture. (MSTAR images are enlarged for clarity.)

Methods	MSTAR							SAR-Airplane						
	VGG16	RN34	RN50	EN	CN	ViT	Avg.	VGG16	RN34	RN50	EN	CN	ViT	Avg.
Baseline	68.35	63.45	71.42	64.65	80.61	48.38	66.14	86.36	90.91	90.34	80.68	94.89	50.00	82.20
ACGAN	69.19	68.33	72.98	72.26	82.49	64.35	71.60	86.53	91.86	91.77	90.54	96.01	78.24	89.17
AGGAN	72.42	71.47	75.83	73.62	84.35	69.21	74.48	90.66	92.75	92.34	95.69	96.17	82.61	91.70
LDM	74.25	71.86	77.39	73.53	84.81	70.14	75.33	91.13	93.10	92.31	96.52	96.14	83.12	92.05
IDiff-Face	74.52	72.21	77.45	73.58	84.83	71.43	75.67	91.21	93.25	92.38	96.61	96.22	83.16	92.14
ID^3	74.84	72.80	77.87	74.14	85.17	71.50	76.05	91.24	93.27	92.38	96.73	96.38	83.25	92.21
SAR-DisentDM	**77.28**	**73.36**	**78.92**	**77.03**	**86.51**	**75.51**	**78.10**	**92.05**	**96.02**	**92.77**	**97.16**	**97.73**	**84.09**	**93.30**

Table 2: SAR-ATR accuracy (%) comparison: SOTA DGMs vs. SAR-DisentDM on MSTAR and SAR-Airplane datasets.

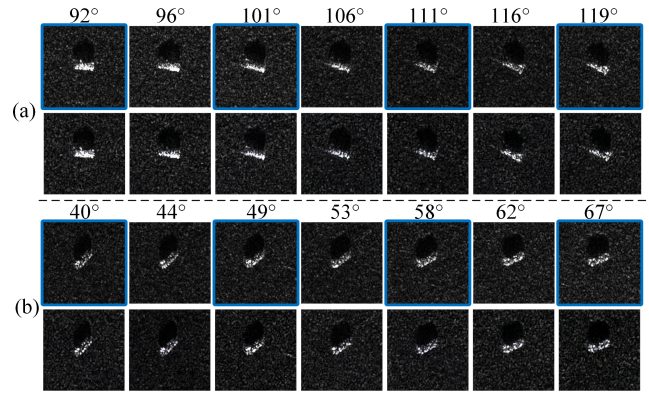

Figure 4: Illustrations of azimuth attribute disentanglement. (a) BMP-2. (b) BTR-60. Top: real images; bottom: synthetic images. Blue boxes highlight training samples.

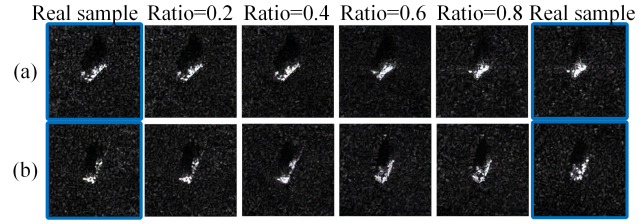

Figure 5: Illustrations of class attribute disentanglement. (a) Transformation from 2S1 to D7 (azimuth angle = 45°). (b) Transformation from 2S1 to T72 (azimuth angle = 200°).

C_1	C_2	VGG16	RN34	RN50	EN	CN	ViT	Avg.
×	×	74.25	71.86	77.39	73.53	84.81	70.14	75.33
√	×	75.04	72.41	77.84	74.42	84.93	71.48	76.02
×	√	76.13	72.71	78.45	76.14	86.26	74.05	77.29
√	√	**77.28**	**73.36**	**78.92**	**77.03**	**86.51**	**75.51**	**78.10**

Table 3: Ablation studies on MSTAR dataset (C_1 and C_2 respectively stand for AAPA and AGSD).

improvement originates from its specific design: 1) the AGSD ensures identity discrimination through azimuth-target decoupling; and 2) the AAPA increases robustness to azimuth angle errors by perturbation augmentation.

Ablation Studies

We conduct ablation studies to demonstrate the efficacy of key components in SAR-DisentDM: the AAPA and the AGSD. As summarized in Table 3, incorporating AAPA results in consistently higher ATR accuracy across the tested ATR models, indicating its effectiveness in introducing greater angle variation, thereby reducing the generative model from overfitting to specific azimuth angles. Meanwhile, adding AGSD further enhances ATR accuracy. Specifically, AGSD provides an average accuracy gain of approximately 2.0% for the six ATR models, attributable to its role in extracting discriminative identity features. These ablation studies confirm the critical role of both the AAPA and

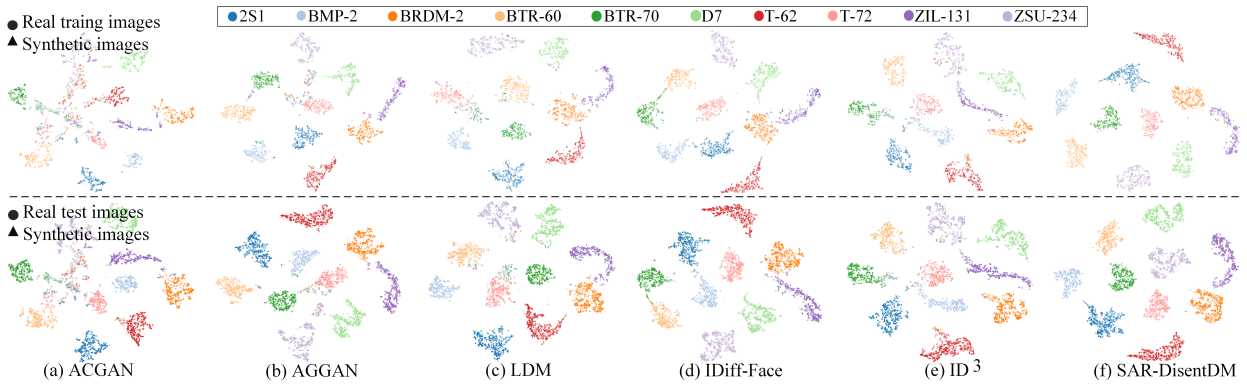

Figure 6: t-SNE visualizations of SAR data generated by six SOTA DGMs (a-f) on MSTAR dataset. Top: real training images vs. synthetic images; bottom: real test images vs. synthetic images.

Tra. Size	VGG16	RN34	RN50	EN	CN	ViT	Avg.
N_f/5	79.04	80.01	82.26	73.89	87.13	59.80	77.02
	82.52	84.78	84.77	83.68	94.25	79.25	84.88
N_f/10	68.35	63.45	71.42	64.65	80.61	48.38	66.14
	77.28	73.36	78.92	77.03	86.51	75.51	78.10
N_f/25	55.78	53.85	67.43	52.26	68.72	27.04	54.18
	76.08	78.92	76.21	71.39	79.56	49.14	71.88
N_f/40	41.34	36.31	44.00	42.33	47.72	15.13	37.81
	49.71	42.04	56.25	54.22	56.65	39.94	49.80

Table 4: SAR-ATR accuracy (%) vs. training data size for generation on MSTAR dataset. Top: Baseline (ATR models trained with real data only); bottom: Ours (pre-trained with synthetic data, fine-tuned with real data).

Syn. Size	VGG16	RN34	RN50	EN	CN	ViT	Avg.
$N_f \times 0.5$	71.62	68.55	72.37	67.39	83.31	52.25	69.25
$N_f \times 1$	77.28	73.36	78.92	77.03	86.51	75.51	78.10
$N_f \times 2$	77.52	73.71	79.53	77.66	87.14	76.80	78.73
$N_f \times 5$	77.73	73.83	79.61	77.70	87.42	77.81	79.02

Table 5: SAR-ATR accuracy (%) vs. synthetic data size for ATR training on MSTAR dataset.

AGSD in boosting the performance of SAR-DisentDM.

Discussion

We investigate how training data size affects generative models, by progressively increasing azimuth sampling intervals: 5°, 10°, 25°, 40°. This reduces training data size to $N_f/5, N_f/10, N_f/25, N_f/40$, where N_f is the full 17° depression-angle MSTAR data size. Training samples are evenly distributed per interval. As Table 4 depicted, decreasing training data size lowers classification accuracy across all SAR-ATR models. In particular, SAR-DisentDM significantly outperforms the Baseline (trained solely on sparse real samples), achieving average accuracy boosts of 7.86%, 11.96%, 17.70%, and 11.99% for the respective intervals, demonstrating superior performance under data scarcity.

Subsequently, we analyze the impact of synthetic data volume on ATR performance, by scaling ATR training sets to $N_f \times 0.5$, $N_f \times 1$, $N_f \times 2$, and $N_f \times 5$. As observed in Table 5, identification accuracy generally improves with increasing synthetic data volume; however, these gains begin to saturate once the data reaches a certain size. For instance, expanding the dataset to $N_f \times 5$ yields negligible improvement over $N_f \times 2$. This saturation likely occurs because the additional images, generated under identical azimuth conditions but with varying noise, contribute less discriminative information than increased azimuth diversity. This also implies that the generative model's representational capacity is bounded above under limited samples.

Finally, as shown in Fig. 6, we visualize the t-SNE distribution of SAR data generated by five DGMs and our proposed SAR-DisentDM on MSTAR dataset. From the figures, one can observe that compared to benchmark methods, the distribution of our generated data aligns better with real training/testing samples from the corresponding categories. This confirms that SAR-DisentDM produces a more discriminative representation by effectively capturing class-specific features and generalizing across azimuth angles.

Conclusions

This work addresses limited-data SAR image generation, where data scarcity and azimuth sensitivity cause azimuth-target coupling and overfitting. We propose SAR-DisentDM, a semantic knowledge guided diffusion framework featuring: 1) an AGSD module decoupling category-specific/azimuth-variant features via adaptive dual loss, and 2) an AAPA mechanism enhancing robustness to azimuth angle errors. Comprehensive experiments are conducted on two SAR image datasets, and evaluation against several SOTA approaches validates the effectiveness of our proposed SAR-DisentDM. Future work will enforce azimuth correlations by incorporating EM constraints into the diffusion process, further improving the interpretability and generalization capability of SAR image synthesis.

Acknowledgements

This work is supported by the National Natural Science Foundation of China (No. 62301464, 62176170, 62176169), and the Sichuan Science and Technology Program (No. 2024NSFSC1427, 2025ZNSFSC0469).

References

AFRL. 2016. The Air Force Moving and Stationary Target Recognition Database. Available: https://www.sdms.afrl.af.mil/datasets/mstar/.

Bao, J.; Yu, W. M.; Yang, K.; Liu, C.; and Cui, T. J. 2024. Improved few-shot SAR image generation by enhancing diversity. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 17: 3394–3408.

Boutros, F.; Grebe, J. H.; Kuijper, A.; and Damer, N. 2023. IDiff-Face: synthetic-based face recognition through fizzy identity-conditioned diffusion model. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 19650–19661.

Cao, C.; Cui, Z.; Wang, L.; Wang, J.; Cao, Z.; and Yang, J. 2022. A demand-driven SAR target sample generation method for imbalanced data learning. *IEEE Transactions on Geoscience and Remote Sensing*, 60: 1–15.

Chen, J.; Qiu, X.; Ding, C.; and Wu, Y. 2022. SAR image classification based on spiking neural network through spike-time dependent plasticity and gradient descent. *ISPRS Journal of Photogrammetry and Remote Sensing*, 188: 109–124.

Curlander, J. C.; and McDonough, R. N. 1991. *Synthetic aperture radar*, volume 11. New York: Wiley.

Dhariwal, P.; and Nichol, A. 2021. Diffusion models beat gans on image synthesis. *Advances in Neural Information Processing Systems*, 34: 8780–8794.

Ding, B.; Wen, G.; Huang, X.; Ma, C.; and Yang, X. 2017. Data augmentation by multilevel reconstruction using attributed scattering center for SAR target recognition. *IEEE Geoscience and Remote Sensing Letters*, 14(6): 979–983.

Dosovitskiy, A. 2021. An image is worth 16x16 words: Transformers for image recognition at scale. In *International Conference on Learning Representations*, 1–22.

Goldberg, C. 2024. Statistically principled deep learning for SAR image segmentation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, 23742–23743.

Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; and Bengio, Y. 2014. Generative adversarial nets. *Advances in Neural Information Processing Systems*, 27: 1–9.

Guo, J.; Lei, B.; Ding, C.; and Zhang, Y. 2017. Synthetic aperture radar image synthesis by using generative adversarial nets. *IEEE Geoscience and Remote Sensing Letters*, 14(7): 1111–1115.

Guo, Q.; Xu, H.; and Xu, F. 2023. Causal adversarial autoencoder for disentangled SAR image representation and few-shot target recognition. *IEEE Transactions on Geoscience and Remote Sensing*, 61: 1–14.

He, K.; Zhang, X.; Ren, S.; and Sun, J. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 770–778.

Ho, J.; Jain, A.; Abbeel, P.; and et al. 2020. Denoising diffusion probabilistic models. *Advances in Neural Information Processing Systems*, 33: 6840–6851.

Hu, X.; Feng, W.; Guo, Y.; and Wang, Q. 2021. Feature learning for SAR target recognition with unknown classes by using CVAE-GAN. *Remote Sensing*, 13(18): 3554.

Huang, Z.; Zhang, X.; Tang, Z.; Xu, F.; Datcu, M.; and Han, J. 2024. Generative artificial intelligence meets synthetic aperture radar: a survey. *IEEE Geoscience and Remote Sensing Magazine*, 1(1): 2–44.

Joo, S.; Sungmin, C.; and Taesup, M. 2019. DoPAMINE: double-sided masked CNN for pixel adaptive multiplicative noise despeckling. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, 4031–4038.

Kim, M.; Liu, F.; Jain, A.; and Liu, X. 2023. DCFace: synthetic face generation with dual condition diffusion model. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 12715–12725.

Li, C.; Du, L.; Li, Y.; and Song, J. 2022. A novel SAR target recognition method combining electromagnetic scattering information and GCN. *IEEE Geoscience and Remote Sensing Letters*, 19: 1–5.

Liu, Y.; Li, W.; Liu, L.; Zhou, J.; Peng, B.; Song, Y.; Xiong, X.; Yang, W.; Liu, T.; Liu, Z.; and Li, X. 2025. ATRNet-STAR: a large dataset and benchmark towards remote sensing object recognition in the wild. *arXiv preprint arXiv:2501.13354*.

Melzi, P.; Rathgeb, C.; Tolosana, R.; Vera-Rodriguez, R.; Lawatsch, D.; Domin, F.; and Schaubert, M. 2023. GANDiffFace: controllable generation of synthetic datasets for face recognition with realistic variations. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 3086–3095.

Nichol, A. Q.; and Dhariwal, P. 2021. Improved denoising diffusion probabilistic models. In *International Conference on Machine Learning*, 8162–8171.

Odena, A.; Olah, C.; Shlens, J.; and et al. 2017. Conditional image synthesis with auxiliary classifier gans. In *International Conference on Machine Learning*, 2642–2651.

Otroshi, H.; and Marcel, S. 2025. HyperFace: generating synthetic face recognition datasets by exploring face embedding hypersphere. In *International Conference on Learning Representations*, 1–13.

Peng, G.; Liu, M.; Chen, S.; Tao, M.; Li, Y.; and Xing, M. 2024. A directional generation algorithm for SAR image based on azimuth-guided statistical generative adversarial network. *IEEE Transactions on Signal Processing*, 72: 5406–5421.

Razavi, A.; Van Den Oord, A.; Vinyals, O.; and et al. 2019. Generating diverse high-fidelity images with VQ-VAE-2. In *Advances in Neural Information Processing Systems*, volume 32, 1–10.

Rombach, R.; Blattmann, A.; Lorenz, D.; Esser, P.; and Ommer, B. 2022. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 10684–10695.

Shi, X.; Xing, M.; Zhang, J.; and Sun, G. 2022. ISAGAN: a high-fidelity full-azimuth SAR image generation method. In *China International SAR Symposium (CISS)*, 1–4.

Simonyan, K.; and Zisserman, A. 2014. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

Song, Q.; Xu, F.; Zhu, X. X.; and Jin, Y. Q. 2021. Learning to generate SAR images with adversarial autoencoder. *IEEE Transactions on Geoscience and Remote Sensing*, 60: 1–15.

Sun, Y.; Wang, Y.; Hu, L.; Huang, Y.; Liu, H.; Wang, S.; and Zhang, C. 2023. 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.

Tan, H.; Zhang, Z.; Shi, X.; Yang, X.; Li, Y.; Bai, X.; and Zhou, F. 2025. Few-shot SAR ATR via multilevel contrastive learning and dependence matrix-based measurement. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 18: 8175–8188.

Tan, M.; and Le, Q. 2019. Efficientnet: rethinking model scaling for convolutional neural networks. In *International Conference on Machine Learning*, 6105–6114.

Wang, C.; Pei, J.; Liu, X.; Huang, Y.; Mao, D.; Zhang, Y.; and Yang, J. 2022a. 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.

Wang, C.; Xu, R.; Huang, Y.; Pei, J.; Huang, C.; Zhu, W.; and Yang, J. 2025. Limited-data SAR ATR causal method via dual-invariance intervention. *IEEE Transactions on Geoscience and Remote Sensing*, 63: 1–19.

Wang, K.; Zhang, G.; Leng, Y.; and Leung, H. 2018. Synthetic aperture radar image generation with deep generative models. *IEEE Geoscience and Remote Sensing Letters*, 16(6): 912–916.

Wang, R.; Zhang, H.; Han, B.; and et al. 2022b. Multiangle SAR dataset construction of aircraft targets based on angle interpolation simulation. *Journal of Radars*, 11: 637–651.

Woo, S.; Debnath, S.; Hu, R.; Chen, X.; Liu, Z.; Kweon, I. S.; and Xie, S. 2023. Convnext v2: co-designing and scaling convnets with masked autoencoders. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 16133–16142.

Xu, J.; Li, S.; Wu, J.; Xiong, M.; Deng, A.; Ji, J.; Huang, Y.; Mu, G.; Feng, W.; Ding, S.; et al. 2024. ID^3: identity-preserving-yet-diversified diffusion models for synthetic face recognition. *Advances in Neural Information Processing Systems*, 37: 77777–77798.

Yang, Y.; Gui, S.; Zhao, Q.; and Wan, Q. 2025. Deep-unfolded fusion optimization with motion correlation learning for ViSAR multi-GMTIm. *IEEE Transactions on Instrumentation and Measurement*, 1–18.

Yang, Y.; Ma, F.; and Yang, L. 2026. On the anchor aided phase interferometer calibration and source localization. *Signal Processing*, 240: 110385.

Yang, Y.; Zhao, Q.; and Wan, Q. 2024. Polarimetric inverse scattering using LSTM-aided association-learning sparse reconstruction framework. *Signal Processing*, 224: 109586.

Ying, Z.; Ke, W.; Zhai, Y.; Long, Z.; Zhou, J.; Zhu, H.; and Chen, C. L. P. 2025. DiffuSAR: frequency domain-aware diffusion model for SAR image generation. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 18: 11851–11866.

Zhang, X.; Zhuang, Y.; Guo, Q.; Yang, H.; Qian, X.; Cheng, G.; Han, J.; and Huang, Z. 2025. Phi-GAN: physics-inspired GAN for generating SAR images under limited data. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 29075–29085.

Zhu, T.; Chen, J.; Zhu, R.; and Gupta, G. 2023. StyleGAN3: generative networks for improving the equivariance of translation and rotation. *arXiv preprint arXiv:2307.03898*.

Zou, L.; Zhang, H.; Wang, C.; Wu, F.; and Gu, F. 2020. MW-ACGAN: generating multiscale high-resolution SAR images for ship detection. *Sensors*, 20(22): 6673–6683.