Towards Realistic SAR Super-Resolution: A Two-Stage Enhancement Framework with Latent Diffusion Models

Yejun Lee, Sohee Son*
SI Analytics
Daejeon, Republic of Korea
{yejun.lee, shson}@si-analytics.ai

Abstract

Synthetic Aperture Radar (SAR) image super-resolution (SR) remains challenging due to the low spatial resolution of SAR observations and the presence of severe speckle noise. Unlike natural images, SAR imagery often exhibits a substantial domain gap between low-resolution (LR) and high-resolution (HR) data, arising from sensor-dependent imaging characteristics, resolution differences, and complex degradation processes that are difficult to model explicitly. Since real paired LR-HR SAR data are rarely available, existing SAR SR methods commonly rely on predefined synthetic degradation operators, which can limit their generalization to real LR data acquired from different sensors. In this paper, we propose a two-stage SAR image SR framework based on Latent Diffusion Models (LDMs), designed to reconstruct realistic HR SAR images under practical settings. To alleviate the domain gap caused by the lack of paired LR-HR data, we incorporate a Blind-Spot Network (BSN) during the LR-to-HR reconstruction process. The proposed framework separately restores structural details and speckle noise in two distinct stages, leveraging the strong generative power of LDMs to produce realistic speckle patterns. In addition, we propose a new evaluation strategy based on Fréchet Wavelet Distance (FWD) to assess the similarity between restored images and real HR SAR datasets from both structural and speckle perspectives. Extensive experiments demonstrate that the proposed two-stage framework yields outputs more similar to real SAR data than a single-stage approach, and validate the effectiveness of the proposed metric.

1. Introduction

Synthetic Aperture Radar (SAR) imagery offers the unique advantage of all-weather, day-and-night observation, owing to its active sensing nature and the physical properties of microwaves, which allow penetration through clouds and thin surface obstructions. These properties make SAR data useful for many remote sensing applications, including object detection [16, 23] and terrain classification [36]. However, SAR images remain difficult to interpret because they often exhibit low spatial resolution and severe speckle noise, which can degrade both visual interpretation and downstream recognition performance. To mitigate these limitations, recent studies have explored both SAR image super-resolution (SAR SR) [8, 24, 37] and speckle suppression [17, 21, 27].

Compared with despeckling (hereafter, *denoising* for clarity), SAR SR has received relatively limited attention. The practical goal of SAR SR is to reconstruct real high-resolution (HR) SAR images from real low-resolution (LR) inputs while preserving not only structural information but also realistic speckle characteristics. In practice, this remains difficult because paired LR-HR SAR data are rarely available. As a result, many existing approaches rely on synthetic LR images generated by bicubic downsampling or train denoising models with gamma-distributed noise (hereafter, *gamma noise*). Yet speckle statistics depend on imaging conditions such as band, polarization, and terrain geometry, so generic synthetic degradation often fails to reflect real SAR observations. In this sense, realistic SAR SR is constrained by three closely related challenges: severe speckle noise, complex and poorly understood degradation, and limited evaluation metrics.

First, unlike EO imagery, where textures and structures often reflect underlying terrain or object boundaries, SAR images are dominated by high-frequency speckle patterns that persist across diverse land-cover types. As illustrated in Figure 1-(b) and (c), the Sentinel-2 image exhibits relatively smooth and interpretable patterns, whereas the corresponding Sentinel-1 image is visually dominated by speckle. In the context of SR, such speckle noise becomes part of the learning target. As the resolution gap increases, this can lead to unstable training dynamics and degraded reconstruction fidelity.

*Corresponding author. Email: shson@si-analytics.ai

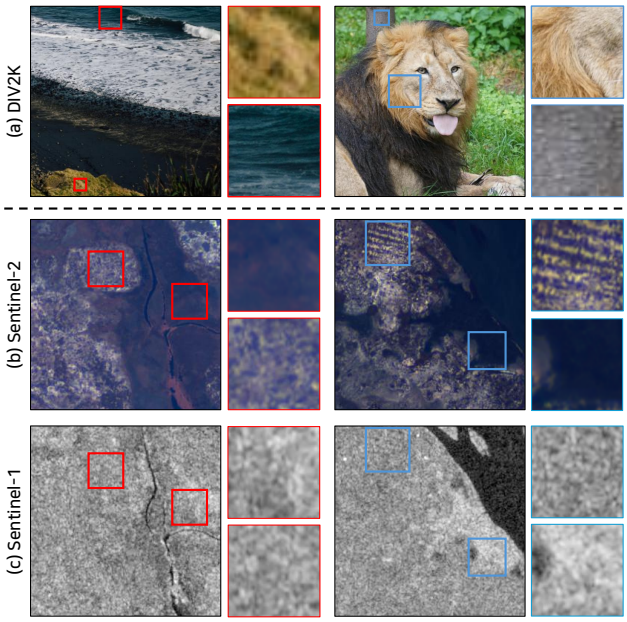

Figure 1. Visual comparison of natural, EO, and SAR images. Unlike optical images, SAR images contain weaker structural cues and are dominated by speckle across the scene.

Second, the degradation relationship between LR and HR SAR data remains poorly understood and is often sensor-dependent. In the natural image domain, Real-ESRGAN [29] showed that carefully designed synthetic degradation pipelines can substantially improve real-world SR, even without real paired LR-HR data. However, its degradation model is tailored to natural-image artifacts such as blur, noise, and compression, rather than SAR-specific imaging effects. Likewise, recent SAR studies often approximate degradation with bicubic downsampling and simulated gamma noise, but such simplifications do not capture the full physical variability of SAR acquisition. Bridging the domain gap between different SAR sensors or acquisition modes therefore remains a central challenge for realistic SAR SR.

Third, evaluation is itself difficult. A super-resolved SAR image should resemble real HR SAR not only in structure but also in speckle characteristics. However, most existing evaluation metrics in SAR restoration do not fully capture quality or distributional realism. PSNR, a common SR metric, is closely tied to mean squared error and therefore tends to favor overly smooth outputs. As a result, outputs with weaker speckle may receive better scores even when they are less faithful to the real SAR distribution. Visual inspection is often used as a complementary tool, but it is subjective and does not provide a reliable quantitative basis for comparing methods.

In this paper, we propose a two-stage enhancement framework based on Latent Diffusion Models (LDMs) [22] to reconstruct realistic HR SAR images from real LR inputs. This two-stage design enables the recovery of fine speckle patterns that are difficult to reconstruct in a single step. Given the complexity of the degradation process, we introduce a blind-spot network (BSN) [11, 35]-based denoiser to simplify the reconstruction. The denoised output from the denoiser serves as a bridge between the real LR SAR inputs and the HR targets, helping the LDM focus on restoring high-frequency speckle characteristics. We further introduce an evaluation strategy based on Fréchet Wavelet Distance (FWD) [25], which measures similarity to real HR SAR data from both structural and speckle perspectives by separating low- and high-frequency components through a Haar wavelet transform.

Our contributions can be summarized as follows:

- We propose a two-stage SAR SR framework that separately restores structure and speckle while leveraging a BSN to bridge the LR-HR domain gap.
- We introduce an FWD-based evaluation strategy to assess both structural similarity and speckle realism with respect to real SAR data.
- Extensive experiments show that the proposed two-stage approach produces outputs closer to real HR SAR imagery than single-stage baselines, both quantitatively and qualitatively.

2. Related works

2.1. SAR image restoration

SAR denoising has long relied on spatial filters [6, 12, 14] and wavelet-based methods [4, 30], which often trade speckle suppression against structural preservation. Recent approaches increasingly use deep networks, including CNN-based denoising [27], diffusion-based restoration [21], attention-based models [17], and joint denoising-SR frameworks [24].

SAR SR has received less attention than denoising. Early work applied FSRCNN to MSTAR [5, 18], while later methods introduced two-stage GAN-based restoration [8], complex-valued SR [1], optics-guided joint SR and denoising [37], and unified despeckling-SR frameworks [24]. Despite this progress, most SAR SR methods still rely on synthetic low-quality data generated by predefined degradation processes such as bicubic downsampling and simulated speckle injection [8, 24, 37]. As a result, realistic cross-sensor SAR SR under unknown or weakly paired LR-HR settings remains relatively less studied.

2.2. Evaluation metrics

Quantitative evaluation is difficult in SAR denoising and SR because clean SAR images or perfectly aligned real HR references are rarely available. As a result, many studies rely

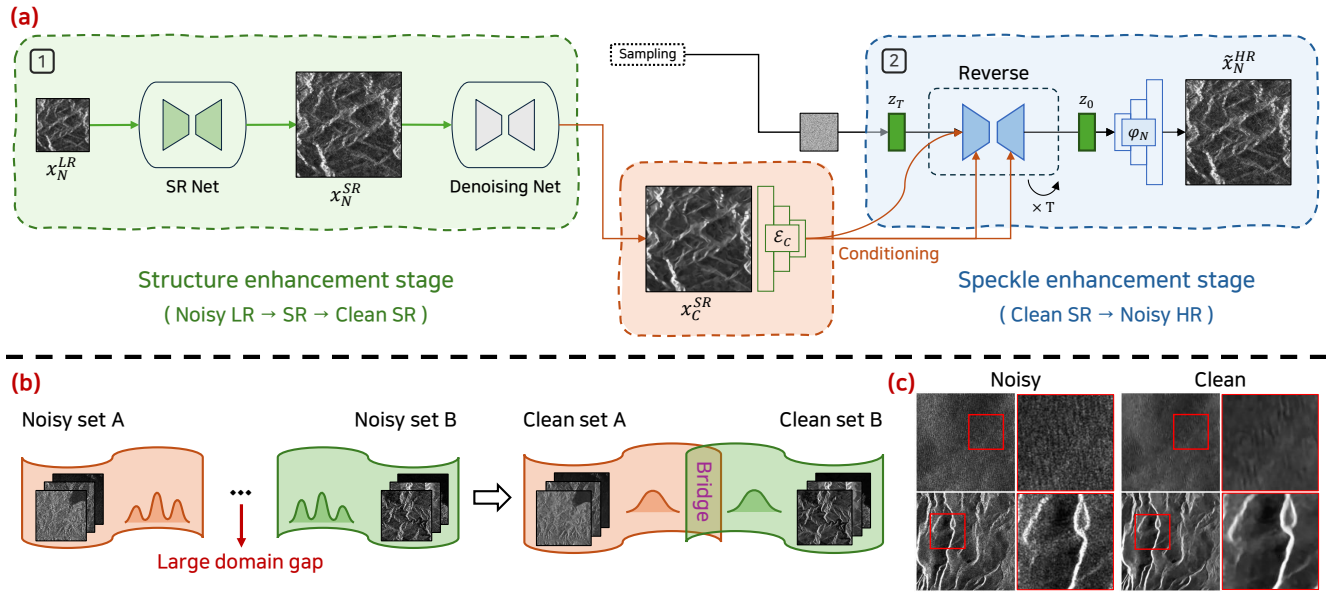

Figure 2. (a) Overview of the proposed framework. (b) Denoised SAR images can serve as an effective bridge between datasets exhibiting different speckle characteristics. (c) The denoising output of the BSN-based denoiser.

on synthetic test data and distortion-based metrics. PSNR is widely used, but it often favors overly smooth outputs and does not necessarily reflect perceptual quality [3, 13]. ENL is also common for denoising without ground truth, yet it may reward oversmoothing and does not fully capture structural preservation [7, 32].

Recent restoration work has also adopted no-reference image quality assessment metrics such as MUSIQ [10], MANIQA [31], and CLIP-IQA [26]. However, these metrics are largely designed for natural images and may not reliably reflect SAR structure and speckle. This motivates evaluation strategies that separately assess structural fidelity and speckle realism. In this work, we build on Fréchet Wavelet Distance (FWD) [25] and adapt it to SAR evaluation through wavelet-based separation of structural and speckle components.

3. Method

3.1. Proposed framework

Our framework consists of an SR network for stage 1, a denoising network, and a latent diffusion model (LDM) for stage 2. Inference is performed in two steps: structure enhancement followed by speckle enhancement. An overview is shown in Figure 2-(a).

3.1.1. Stage 1. Structure enhancement stage

The first stage restores the structural information of LR SAR images using a backbone SR network. We train this stage in a supervised manner with HR-LR pairs generated by a predefined degradation process. Since the real degradation between LR and HR SAR data is difficult to model accurately, we do not attempt to reproduce the full sensor gap. Instead, we apply a moderate synthetic degradation so that the network focuses on recovering HR-consistent structures. Given the Sentinel resolution gap considered in this work, LR inputs often appear strongly blurred. We therefore adopt a blur-based degradation module with diverse kernels and stochastic operations, following the strategy of Real-ESRGAN [29]. The training objective is

$$\mathcal{L}_{SR} = \mathcal{L}_{rec} + \lambda_{GAN} \cdot \mathcal{L}_{GAN} + \lambda_{FDL} \cdot \mathcal{L}_{FDL} \quad (1)$$

where $\mathcal{L}_{rec}$ is an L1 reconstruction loss. Since L1 alone tends to produce oversmoothed outputs, we additionally use a GAN loss to sharpen structures and a frequency distribution loss (FDL) [20] to encourage consistency in the frequency domain. In our experiments, the loss weights are empirically set to $\lambda_{GAN} = 10^{-3}$ and $\lambda_{FDL} = 10^{-5}$.

3.1.2. Stage 2. Speckle enhancement stage

The second stage uses an LDM to restore realistic HR speckle characteristics. Directly learning an LR-to-HR mapping without paired data is highly challenging due to the differences between real LR inputs and HR targets in both structural detail and speckle statistics. To mitigate this issue, we adopt a blind-spot denoising network (BSN) [11] to generate clean SAR images, which helps bridge the domain gap between LR and HR SAR images (see Fig. 2-(b)). Recent advances in BSN-based self-supervised denoising networks have demonstrated effective performance without

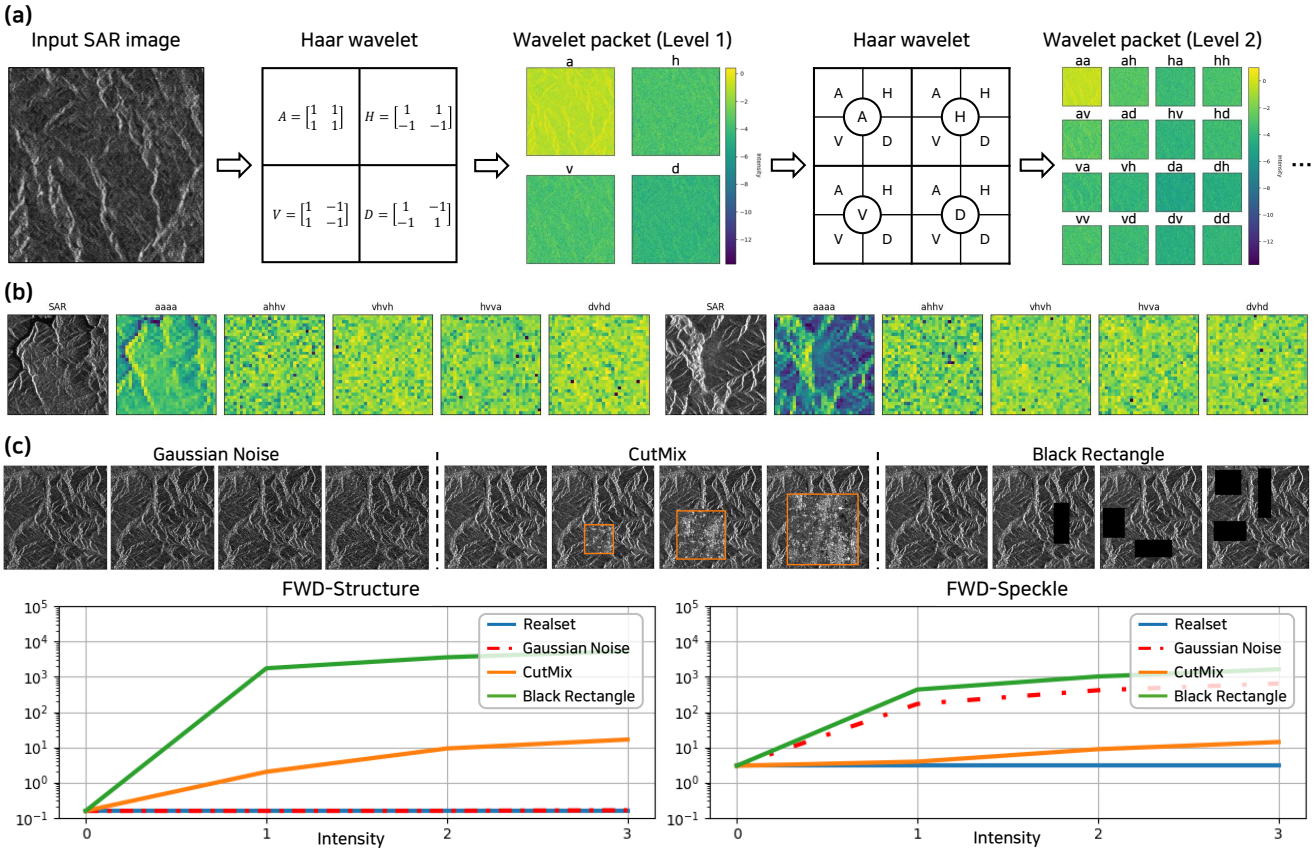

Figure 3. Wavelet-based analysis used for SAR evaluation: (a) Haar wavelet decomposition, (b) low- and high-frequency packets, and (c) controlled failure cases with their effects on FWD-Structure and FWD-Speckle.

requiring ground truth [11, 19, 35]. We empirically validate that applying MM-BSN [35] to SAR data can effectively suppress speckle noise (as shown in Fig. 2-(c)).

The LDM contains two autoencoders, $AE_N(\cdot) = \varphi_N(\varepsilon_N(\cdot))$ for noisy SAR data and $AE_C(\cdot) = \varphi_C(\varepsilon_C(\cdot))$ for clean SAR data, together with a denoising U-Net ϵ_θ. The model learns the conditional distribution $p(z_N^{HR}|z_C^{HR})$, where $\varepsilon_N(I_N^{HR}) = z_N^{HR}$ and $\varepsilon_C(I_C^{HR}) = z_C^{HR}$ are the noisy and clean HR latent codes. Specifically, a noisy HR image $I_N^{HR} \in \mathbb{R}^{1\times H\times W}$ and a clean HR image $I_C^{HR} \in \mathbb{R}^{1\times H\times W}$ are mapped to $z_N^{HR} \in \mathbb{R}^{6\times \frac{H}{4}\times \frac{W}{4}}$ and $z_C^{HR} \in \mathbb{R}^{3\times \frac{H}{4}\times \frac{W}{4}}$, respectively. We found that a 6-channel latent space for noisy SAR data is beneficial for modeling realistic speckle. The denoising U-Net is trained with

$$\mathcal{L}_{LDM} := \mathbb{E}_{\varepsilon_N(I_N^{HR}), z_C^{HR}, \epsilon\sim\mathcal{N}(0,1), t}\left[||\epsilon - \epsilon_\theta(z_{N,t}^{HR}, t, z_C^{HR})||_2^2\right] \quad (2)$$

to recover z_N^{HR} from its noisy version $z_{N,t}^{HR}$ conditioned on z_C^{HR}. During inference, z_C^{HR} is replaced with the latent embedding of the stage 1 output, denoted by z_C^{SR}.

3.2. Data preprocessing

SAR images contain 16-bit intensity values that reflect surface backscatter. Because both intensity distributions and speckle statistics vary across land-cover types, preserving the global intensity range helps the generative model capture land-cover-dependent speckle more faithfully. Patch-wise normalization can remove relative intensity differences across samples and distort the dataset-level distribution. We therefore normalize all data using global min-max values computed over the entire dataset. To reduce the effect of extreme outliers, pixel intensities are clipped to the 1st and 99th percentiles before normalization.

3.3. Evaluation

We build our evaluation on Fréchet Wavelet Distance (FWD), a domain-agnostic metric based on Haar wavelet decomposition. Haar wavelets separate an image into low- and high-frequency components across multiple levels, which is useful for disentangling structural content from speckle-like high-frequency patterns in SAR images. As illustrated in Figure 3-(b), low-frequency packets primar-

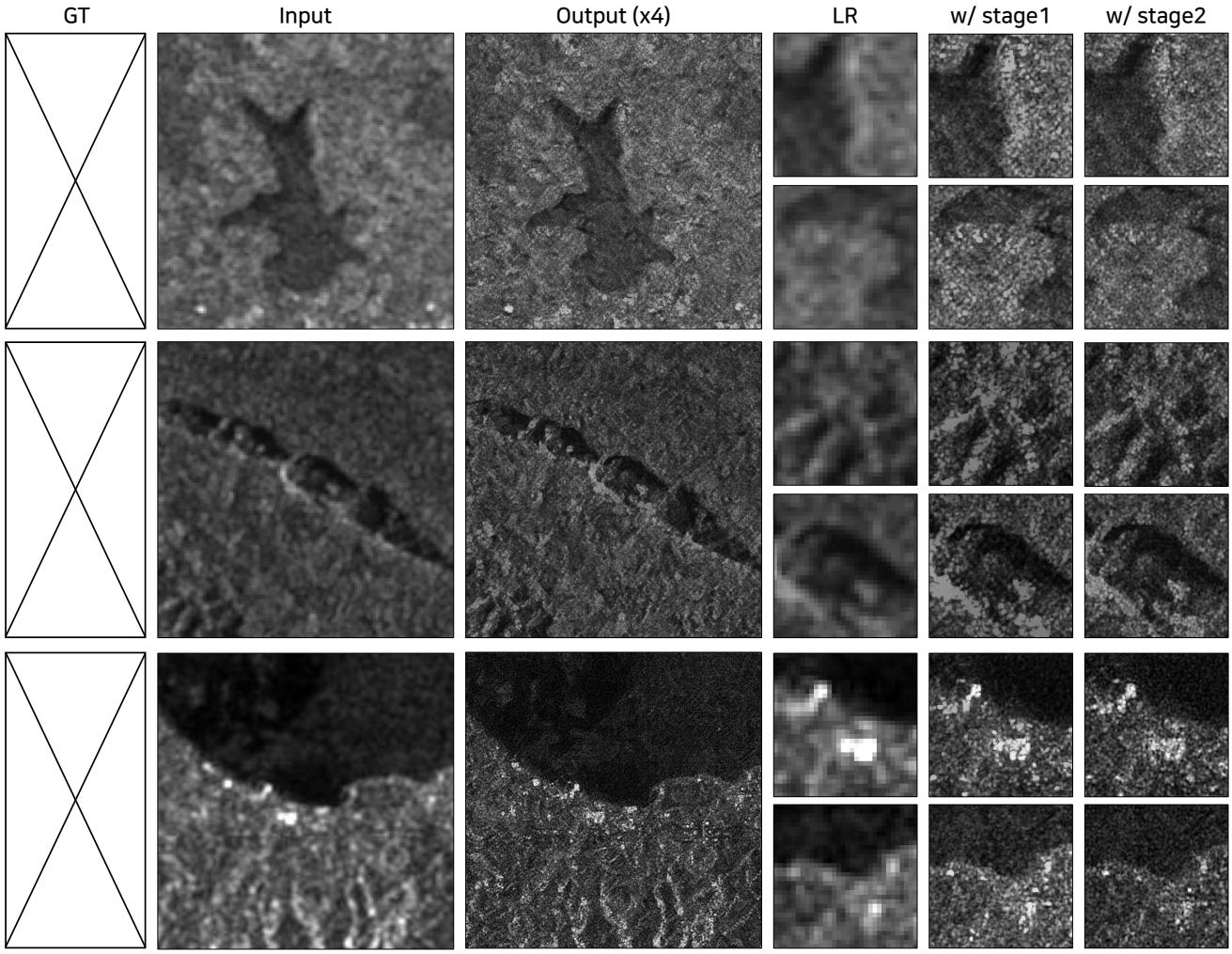

Figure 4. Qualitative results of the proposed framework on real LR SAR images. The stage 1 output restores fine structural details from the LR input, while the stage 2 output further recovers fine speckle details.

ily capture structure with reduced speckle, whereas high-frequency packets predominantly encode speckle characteristics. We therefore define FWD-Structure as the Fréchet Distance computed on the level-4 low-frequency packet (the 'aaaa' packet) and FWD-Speckle as the Fréchet Distance computed on the remaining high-frequency packets. These two variants allow us to assess structural fidelity and speckle realism separately with respect to the real HR SAR dataset. To sanity-check this metric, we evaluate FWD-Structure and FWD-Speckle under controlled failure cases (Figure 3-(c)). Specifically, we compare two sets of 15,000 randomly sampled real SAR images while applying degradations of varying intensity to one set. Gaussian noise largely perturbs speckle while preserving structure. CutMix [33] introduces strong structural distortion by replacing random regions between samples, while black rectangles severely degrade both structure and speckle. These cases help illustrate that the two metrics respond differently to structural and speckle perturbations.

4. Experiments

4.1. Implementation details

4.1.1. Dataset

The dataset used in our experiments consists of dual-polarized (VV/VH) C-band SAR images acquired by Sentinel-1 and accessed through ASF Data Search [2]. To simplify the task, we use only the VV polarization and restrict the data to descending orbits. Sentinel-1 provides multiple acquisition modes with different spatial resolutions; in this work, we treat Interferometric Wide (IW) mode images as the HR domain and Extra Wide (EW) mode images as the LR domain. For training, only IW-mode images are used,

Table 1. FWD comparison across different SR networks with and without stage 2. Inference was performed on real LR SAR images, and FWD scores were measured against real HR images for each stage output.

	FWD-Structure ↓	FWD-Speckle ↓
Bicubic	199.44	10262.47
ESRGAN	186.99	109.19
+ stage 2	**176.60**	**103.97**
Restormer	190.13	211.21
+ stage 2	**178.29**	**99.19**
DRCT	215.61	440.42
+ stage 2	**205.03**	**284.69**

Table 2. Quantitative results on 1,000 randomly selected synthetic test samples. For stage 2, multiple outputs are generated for each input, and the table reports the quantitative results for each sampled output, as well as for the averaged image across the four generated samples.

		PSNR	SSIM
Stage 1		23.4196	0.4389
Stage 2	Sample 1	22.5012	0.3676
	Sample 2	22.4960	0.3673
	Sample 3	22.5012	0.3676
	Sample 4	22.4965	0.3674
	Mean of samples	**24.0388**	**0.4580**

while inference is performed on EW-mode images. A total of 64 scenes covering diverse land-cover types were selected. From these, we extracted 135,406 non-overlapping patches of size 512×512, which were used for training and validation in both stage 1 and stage 2. The test set comprises EW-mode images from the Japan region, which does not overlap with the training data.

4.1.2. Models

For stage 1, we trained multiple SR backbones, including ESRGAN [28], Restormer [34], and DRCT [9], to compare performance under identical conditions. All models were trained using the Adam optimizer with a learning rate of 0.0001, a batch size of 4, and a total of 400,000 iterations. The denoiser that bridges stage 1 and stage 2 is MM-BSN [35], an enhanced variant of AP-BSN [15] that introduces a multi-mask strategy to further disrupt spatial noise correlation. In our experiments, we use a combination of masks including "o", "r", "c", "a45", and "a135". For stage 2, we adopt an LDM built on a KL-regularized autoencoder, and incorporated a hybrid conditioning mechanism to provide strong guidance during the denoising diffusion process. We adopt the LDM training configuration from the original implementation [22].

4.2. Qualitative results

Figure 4 presents the qualitative results produced by our proposed framework. Since the test samples are collected from real-world SAR data, no perfectly aligned ground-truth images are available for direct comparison (leftmost column in Figure 4). From the results, we observe that the stage 1 model successfully reconstructs fine structural details. However, due to the limited diversity of synthetic degradation used during training, the SR network in stage 1 struggles to recover realistic speckle textures. In contrast, the stage 2 outputs preserve the structural content reconstructed in stage 1 while recovering detailed speckle characteristics more effectively.

4.3. Quantitative results

Table 1 presents the FWD-based evaluation results for various SR backbones and the outputs after applying stage 2. Bicubic upsampling produces overly smooth outputs with little realistic speckle content, which is reflected in particularly large FWD-Speckle. Compared with bicubic interpolation, most of SR backbones improve FWD-Structure and also reduces FWD-Speckle, suggesting a smaller distribution gap to real HR SAR data. Notably, applying stage 2 further reduces both metrics across all backbones, indicating that the second stage improves speckle realism while maintaining or refining the structural content restored by stage 1.

Table 2 shows quantitative results on 1,000 synthetic test samples generated by downsampling real HR SAR images and applying the same blur degradation used during training. For each input, stage 1 produces a single deterministic output, whereas stage 2 produces multiple sampled outputs. Each sampled output is evaluated individually. We also compute a pixel-wise average across the four stage 2 outputs and evaluate the resulting smoothed image using PSNR and SSIM. Although the stage 1 output achieves higher PSNR and SSIM than each individual stage 2 sample, the averaged stage 2 output yields even higher scores. This observation is consistent with the tendency of PSNR and SSIM to favor smoother reconstructions, even when such smoothing may suppress high-frequency speckle details.

4.4. Limitations

While the proposed framework demonstrates improved realism in SAR image SR, it exhibits strong dependence on the performance of the first-stage network. Enhancing the structure restoration stage is therefore crucial, and its improvement can directly boost the effectiveness of the entire framework. Finally, since the framework is based on diffusion models, it incurs considerable computational costs when applied to large-scale satellite imagery, which may limit its practicality in some operational settings.

5. Conclusion

We presented a two-stage framework for realistic SAR image super-resolution that separates structural restoration from speckle enhancement. The first stage recovers HR-consistent structure, and the second stage uses an LDM guided by a BSN-based bridge to restore realistic speckle patterns. We also introduced an FWD-based evaluation strategy that measures similarity to real HR SAR data from structural and speckle perspectives. Experiments on Sentinel-1 data show that the proposed two-stage design produces outputs more similar to real HR SAR data than the corresponding stage-1 outputs.

Acknowledgement This work was supported by Korea Research Institute for defense Technology planning and advancement(KRIT) grant funded by the Korea government(DAPA(Defense Acquisition Program Administration)) (KRIT-CT-22-040, Heterogeneous Satellite constellation based ISR Research Center, 2022)

References

[1] Pia Addabbo, Mario Luca Bernardi, Filippo Biondi, Marta Cimitile, Carmine Clemente, Nicomino Fiscante, Gaetano Giunta, Danilo Orlando, and Linjie Yan. Super-resolution of synthetic aperture radar complex data by deep-learning. *IEEE Access*, 11:23647–23658, 2023. 2

[2] Alaska Satellite Facility. ASF Data Search. https://search.asf.alaska.edu/. Accessed: 2026-05-07. 5

[3] Yochai Blau and Tomer Michaeli. The perception-distortion tradeoff. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 6228–6237, 2018. 3

[4] Min Dai, Cheng Peng, Andrew K Chan, and Dmitri Loguinov. Bayesian wavelet shrinkage with edge detection for sar image despeckling. *IEEE Transactions on Geoscience and Remote Sensing*, 42(8):1642–1648, 2004. 2

[5] Chao Dong, Chen Change Loy, and Xiaoou Tang. Accelerating the super-resolution convolutional neural network. In *European conference on computer vision*, pages 391–407. Springer, 2016. 2

[6] Victor S Frost, Josephine Abbott Stiles, K Sam Shanmugan, and Julian C Holtzman. A model for radar images and its application to adaptive digital filtering of multiplicative noise. *IEEE Transactions on pattern analysis and machine intelligence*, (2):157–166, 1982. 2

[7] Luis Gomez, Raydonal Ospina, and Alejandro C. Frery. Unassisted quantitative evaluation of despeckling filters. *Remote Sensing*, 9(4):389, 2017. 3

[8] Feng Gu, Hong Zhang, Chao Wang, and Fan Wu. Sar image super-resolution based on noise-free generative adversarial network. In *IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium*, pages 2575–2578. IEEE, 2019. 1, 2

[9] Chih-Chung Hsu, Chia-Ming Lee, and Yi-Shiuan Chou. Drct: Saving image super-resolution away from information bottleneck. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6133–6142, 2024. 6

[10] Junjie Ke, Qifei Wang, Yilin Wang, Peyman Milanfar, and Feng Yang. Musiq: Multi-scale image quality transformer. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 5148–5157, 2021. 3

[11] Alexander Krull, Tim-Oliver Buchholz, and Florian Jug. Noise2void-learning denoising from single noisy images. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2129–2137, 2019. 2, 3, 4

[12] DARWINT Kuan, ALEXANDERA Sawchuk, TIMOTHYC Strand, and Pierre Chavel. Adaptive restoration of images with speckle. *IEEE transactions on acoustics, speech, and signal processing*, 35(3):373–383, 1987. 2

[13] Christian Ledig, Lucas Theis, Ferenc Huszar, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, and Wenzhe Shi. Photo-realistic single image super-resolution using a generative adversarial network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4681–4690, 2017. 3

[14] Jong-Sen Lee. Speckle analysis and smoothing of synthetic aperture radar images. *Computer graphics and image processing*, 17(1):24–32, 1981. 2

[15] Wooseok Lee, Sanghyun Son, and Kyoung Mu Lee. Ap-bsn: Self-supervised denoising for real-world images via asymmetric pd and blind-spot network. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 17725–17734, 2022. 6

[16] Weijie Li, Wei Yang, Yuenan Hou, Li Liu, Yongxiang Liu, and Xiang Li. Saratr-x: Toward building a foundation model for sar target recognition. *IEEE Transactions on Image Processing*, 34:869–884, 2025. 1

[17] Shuaiqi Liu, Shikang Tian, Yuhang Zhao, Qi Hu, Bing Li, and Yu-Dong Zhang. Lg-dbnet: Local and global dual-branch network for sar image denoising. *IEEE Transactions on Geoscience and Remote Sensing*, 62:1–15, 2024. 1, 2

[18] Zhenyu Luo, Junpeng Yu, and Zhenhua Liu. The super-resolution reconstruction of sar image based on the improved fsrcnn. *The Journal of Engineering*, 2019(19):5975–5978, 2019. 2

[19] Andrea Bordone Molini, Diego Valsesia, Giulia Fracastoro, and Enrico Magli. Speckle2void: Deep self-supervised sar despeckling with blind-spot convolutional neural networks. *IEEE Transactions on Geoscience and Remote Sensing*, 60: 1–17, 2021. 4

[20] Zhangkai Ni, Juncheng Wu, Zian Wang, Wenhan Yang, Hanli Wang, and Lin Ma. Misalignment-robust frequency distribution loss for image transformation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2910–2919, 2024. 3

[21] Malsha V Perera, Nithin Gopalakrishnan Nair, Wele Gedara Chaminda Bandara, and Vishal M Patel. Sar despeckling using a denoising diffusion probabilistic model. *IEEE Geoscience and Remote Sensing Letters*, 20:1–5, 2023. 1, 2

[22] 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*, pages 10684–10695, 2022. 2, 6

[23] Xian Sun, Yixuan Lv, Zhirui Wang, and Kun Fu. Scan: Scattering characteristics analysis network for few-shot aircraft classification in high-resolution sar images. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–17, 2022. 1

[24] Shunsuke Takao. Md-glow: Multi-task despeckling glow for sar image enhancement. In *Proceedings of the Winter Conference on Applications of Computer Vision*, pages 536–543, 2025. 1, 2

[25] Lokesh Veeramacheneni, Moritz Wolter, Hildegard Kuehne, and Juergen Gall. Fr\'echet wavelet distance: A domain-agnostic metric for image generation. *arXiv preprint arXiv:2312.15289*, 2023. 2, 3

[26] Jianyi Wang, Kelvin CK Chan, and Chen Change Loy. Exploring clip for assessing the look and feel of images. In *Proceedings of the AAAI conference on artificial intelligence*, pages 2555–2563, 2023. 3

[27] Puyang Wang, He Zhang, and Vishal M Patel. Sar image despeckling using a convolutional neural network. *IEEE Signal Processing Letters*, 24(12):1763–1767, 2017. 1, 2

[28] Xintao Wang, Ke Yu, Shixiang Wu, Jinjin Gu, Yihao Liu, Chao Dong, Yu Qiao, and Chen Change Loy. Esrgan: Enhanced super-resolution generative adversarial networks. In *Proceedings of the European conference on computer vision (ECCV) workshops*, pages 0–0, 2018. 6

[29] Xintao Wang, Liangbin Xie, Chao Dong, and Ying Shan. Real-esrgan: Training real-world blind super-resolution with pure synthetic data. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 1905–1914, 2021. 2, 3

[30] Hua Xie, Leland E Pierce, and Fawwaz T Ulaby. Despeckling sar images using a low-complexity wavelet denoising process. In *IEEE international geoscience and remote sensing symposium*, pages 321–324. IEEE, 2002. 2

[31] Sidi Yang, Tianhe Wu, Shuwei Shi, Shanshan Lao, Yuan Gong, Mingdeng Cao, Jiahao Wang, and Yujiu Yang. Maniqa: Multi-dimension attention network for no-reference image quality assessment. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 1191–1200, 2022. 3

[32] Xuezhi Yang, Kewei Wu, and Yiming Tang. A new metric for measuring structure-preserving capability of despeckling of sar images. *ISPRS Journal of Photogrammetry and Remote Sensing*, 94:143–159, 2014. 3

[33] Sangdoo Yun, Dongyoon Han, Seong Joon Oh, Sanghyuk Chun, Junsuk Choe, and Youngjoon Yoo. Cutmix: Regularization strategy to train strong classifiers with localizable features. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 6023–6032, 2019. 5

[34] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, and Ming-Hsuan Yang. Restormer: Efficient transformer for high-resolution image restoration. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5728–5739, 2022. 6

[35] Dan Zhang, Fangfang Zhou, Yuwen Jiang, and Zhengming Fu. Mm-bsn: Self-supervised image denoising for real-world with multi-mask based on blind-spot network. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4189–4198, 2023. 2, 4, 6

[36] Pan Zhang, Baochai Peng, Chaoran Lu, Quanjin Huang, and Dongsheng Liu. Asanet: Asymmetric semantic aligning network for rgb and sar image land cover classification. *ISPRS Journal of Photogrammetry and Remote Sensing*, 218:574–587, 2024. 1

[37] Zhicheng Zhao, Qing Gao, Jinquan Yan, Chenglong Li, and Jin Tang. Hsfmamba: Hierarchical selective fusion mamba network for optics-guided joint super-resolution and denoising of noise-corrupted sar images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2025. 1, 2