A Deep Learning Iterative Framework for Sentinel-1 Stripmap Enhancement Based on Azimuth Doppler Decomposition

Anonymous CVPR submission

Paper ID *****

Abstract

Synthetic Aperture Radar (SAR) imagery enables all-weather, day-and-night Earth observation; however, it remains difficult to interpret due to speckle noise and other intrinsic imaging artifacts. Sentinel-1 (S1) constitutes one of the most widely used spaceborne SAR missions, offering systematic global coverage, high temporal resolution, dual-polarization imaging, and free data availability. Among S1 modes, Stripmap (SM) provides the highest resolution, yet speckle noise and spatial constraints often hinder applications requiring finer spatial detail. This motivates the need for effective image enhancement strategies. In this work, we propose a self-supervised enhancement framework for S1 SM imagery based on azimuth subaperture decomposition. The method exploits the physical consistency between subaperture reconstructions and the corresponding full-aperture image to generate paired training data without external sensors, simulated ground truth, or multi-temporal stacks. The proposed framework integrates single- and multi-frame learning and incorporates an iterative inference scheme that progressively refines image quality. Experiments on real S1 SM data show that the proposed approach consistently outperforms the widely adopted self-supervised deep learning baseline MERLIN, in terms of PSNR and SSIM, while MERLIN attains higher ENL, highlighting a trade-off between structural fidelity and speckle smoothing. Overall, the results demonstrate that subaperture-based supervision provides a physically grounded, reproducible, and operationally viable approach for SAR image enhancement using S1 data. It is worth noting that the proposed approach can be extended to other SAR platforms, polarizations, and acquisition modes.

1. Introduction

Synthetic Aperture Radar (SAR) has become a cornerstone of operational Earth observation due to its ability to acquire imagery independent of daylight and weather conditions [48]. However, SAR data is inherently degraded by speckle noise, which arises from the coherent summation of echoes from elementary scatterers within a resolution cell [28]. Furthermore, the side-looking acquisition geometry induces significant radiometric variability driven by local incidence angles and terrain slope [49]. These effects complicate visual interpretation and can degrade the performance of downstream applications [16, 26, 59].

Among open-access SAR missions, Sentinel-1 (S1) provides globally consistent C-band acquisitions supporting a broad range of applications [48]. S1 operates in multiple acquisition modes, with Stripmap (SM) offering the finest spatial resolution (approximately $5\,\mathrm{m} \times 5\,\mathrm{m}$ in range $\times$ azimuth) at a narrow swath width of 80 km [21]. Despite its comparatively higher resolution, SM imagery remains strongly affected by speckle, and its spatial detail is often insufficient to delineate small or densely packed man-made structures [6, 12]. Consequently, recent work has focused on SAR image enhancement techniques that aim to remain consistent with the underlying sensor physics [16, 59].

Most recent S1-enhancement studies rely on supervised learning with proxy references, either by using cross-sensor supervision from high-resolution commercial SAR data [2, 5, 50] or by supervising across different S1 acquisition modes [3]. While effective, these strategies suffer from practical bottlenecks: they require additional acquisitions, precise co-registration, and often rely on proprietary or limited-access data. In parallel, semi- and self-supervised deep learning approaches such as SAR2SAR [15] and MERLIN [16] have demonstrated that speckle reduction can be learned directly from SAR data without clean ground truth. However, these methods are tailored specifically for despeckling and do not explicitly address broader enhancement, where noise suppression and spatial resolution improvement are considered jointly.

Azimuth subaperture decomposition is a well-established SAR processing technique that partitions the full azimuth Doppler spectrum into discrete sub-bands and reconstructs an image from each via inverse transformation [4]. Restricting the azimuth bandwidth coarsens

the effective azimuth resolution while preserving the original acquisition geometry and underlying scattering mechanisms. Consequently, the resulting subaperture images exhibit reduced effective spatial resolution and distinct speckle realizations compared to the full-aperture image, even though the ground-sampling-distance (GSD) remains unchanged.

Recent learning-based approaches have exploited these properties in different ways. Wang *et al*. [51] use subaperture images as auxiliary representations for guiding complex-valued neural networks in SAR target recognition, leveraging the fact that each subaperture emphasizes different Doppler components of the same scene. Similarly, Ristea *et al*. [45] employ azimuth subaperture decomposition to generate multi-view training data for self-supervised representation learning. This strategy effectively creates diverse views of a scene without altering the spatial support or requiring external annotations.

In the context of image restoration, An *et al*. [4] utilize subaperture decomposition for SAR despeckling by treating subaperture images as multiple observations of an identical scene generated under the same imaging mechanism. Their formulation demonstrates that subaperture data can effectively suppress speckle while remaining consistent with SAR signal physics. Conversely, Dong *et al*. [20] integrate subaperture decomposition into a deep learning framework for complex-valued super-resolution (SR), leveraging multi-scale sub-band information to enhance azimuth detail reconstruction. However, the literature currently treats these tasks separately: subapertures are employed either for despeckling or for resolution enhancement. To date, no approach has explicitly leveraged this physics-grounded property for constructing self-supervised training pairs to jointly address speckle reduction and spatial detail recovery.

To address these limitations, we propose a framework that leverages azimuth subaperture decomposition to construct self-supervised input–target pairs directly from S1 SM acquisitions. By exploiting the physical consistency between azimuth-band-limited subaperture reconstructions and the corresponding full-aperture image, the proposed approach enables sensor-consistent enhancement without requiring external reference sensors, simulated ground truth, or multi-temporal stacks. The proposed framework could also be applied to other sensors, acquisition modes, and polarizations.

This paper makes the following contributions:

- We introduce a self-supervised data generation strategy for S1 SM imagery based on azimuth subaperture decomposition, enabling physics-consistent training without external supervision.
- We investigate single- and multi-frame SAR enhancement within a unified learning framework.
- We propose an iterative inference scheme that exposes the trade-off between speckle suppression and structural detail preservation.

2. Related Work

2.1. Speckle Reduction for SAR Images

Classical Statistical and Nonlocal Methods. Early despeckling approaches relied on local statistics and simplified speckle models to design adaptive smoothing filters, such as the Lee and Kuan filters [32, 34]. These methods exploit spatial correlations via sliding-window estimators. However, while effective at speckle suppression, they often degrade edge sharpness and fine textures. The introduction of Non-Local Means [8] established the value of nonlocal patch similarity, leading to probabilistic frameworks [17] and collaborative filtering methods like SAR-BM3D [43]. Despite their effectiveness in preserving fine details, these classical methods are often limited by high computational complexity and strong parameter dependence [44].

Deep Learning-based Despeckling. Recent despeckling literature is dominated by deep learning, ranging from supervised to self-supervised approaches [24, 59]. Supervised methods typically depend on pseudo-ground truth obtained via multi-temporal averaging, multi-looking, or algorithmic reconstruction [11, 14, 33, 42]. Since "clean" reference images are rarely available in real-world SAR scenarios, recent research has pivoted toward self-supervised learning by adapting principles from natural image denoising methods like Noise2Noise and blind-spot learning [31, 35]. These approaches include SAR2SAR [15], MERLIN [16], and blind-spot despeckling in the SAR domain (Speckle2Void) [41], alongside Bernoulli-sampling variants [56]. While these methods mitigate the need for paired data, they hinge on assumptions, such as temporal consistency in multi-image settings and spatial noise independence in blind-spot formulations, that may be violated in dynamic scenes or structured targets. This limitation motivates the integration of physics-aware priors into learning frameworks [24].

2.2. Learning-based Spatial Resolution Enhancement in SAR

Super-Resolution in Computer Vision. Traditional image SR methods, such as bicubic interpolation, are limited in their ability to reconstruct sharp edge structures [40]. Over the past decade, deep learning-based single-image SR (SISR) approaches have gained significant attention by learning a mapping between low- and high-resolution images [54]. Early architectures like SRCNN [19] applied shallow convolutional networks to pre-upsampled inputs, focusing on refining high-frequency details rather than explicitly increasing sampling distance within the network itself. Subsequent models, such as

EDSR [38], significantly improved reconstruction accuracy through deeper residual networks and learnable upsampling. More recently, attention mechanisms have pushed the state of the art by modeling long-range spatial dependencies, resulting in enhanced detail recovery. Prominent examples include RCAN [57], SAN [13], and ESRT [39].

Beyond single-image techniques, Multi-Frame SR (MFSR) exploits complementary information across multiple frames to improve reconstruction. CNN-based methods like EDVR [52] utilize spatiotemporal attention for feature fusion, while recurrent frameworks such as BasicVSR++ [10] leverage bidirectional propagation to capture long-term context. Recent Video Restoration Transformers (e.g., VRT, RVRT) further enhance this by using self-attention to model spatiotemporal features explicitly [36, 37].

SAR Super-Resolution and Cross-Sensor Supervision. SAR SR has primarily relied on GAN-based translation and reconstruction approaches, such as S1-to-TerraSAR-X mapping [5] and various SR-specific architectures [50, 58]. Some studies also explored optical-guided strategies to inject spatial details from co-registered optical imagery [55]. While visually appealing, cross-sensor or cross-modality supervision suffers from domain gaps such as differing carrier frequencies, incidence angles, processing chains, and scene-dependent scattering which complicate validation and risk introducing modality-specific artifacts inconsistent with SAR physics [59]. Furthermore, most existing approaches primarily focus on spatial resolution enhancement and do not explicitly address speckle suppression, which remains intrinsic to coherent SAR imaging. To bridge the gap between data-driven models and physical constraints, hybrid formulations such as compressed sensing and algorithm-unrolling have been explored [26]. However, the consensus in recent literature suggests that SAR SR requires supervision that is both sensor-consistent and faithful to SAR physics, prioritizing reconstruction fidelity over photorealistic synthesis [7, 59].

Amieva *et al*. [3] has addressed this challenge by proposing a joint SR and despeckling framework for S1 Interferometric Wide Swath (IW) imagery, utilizing S1 SM as a reference. While this strategy maintained sensor consistency, the cross-mode supervision (IW→SM) still introduced inevitable domain shifts that limited the fidelity of the reconstruction.

2.3. Physics-grounded Supervision via Azimuth Subapertures

Azimuth Subaperture Decomposition and Effective Resolution. In SAR imaging, spatial resolution is determined by different physical mechanisms in the range and azimuth dimensions. The slant-range resolution is primarily controlled by the transmitted bandwidth B [20], approximated as:

$$\rho_r \approx \frac{c}{2B}, \tag{1}$$

where c is the speed of light. Conversely, the azimuth resolution is intrinsically linked to the Doppler bandwidth B_D accumulated over the synthetic aperture. For a platform velocity v and physical antenna length L, this relationship is given by:

$$B_D \approx \frac{2v}{L}, \qquad \rho_a \approx \frac{v}{B_D} \approx \frac{L}{2}, \tag{2}$$

Azimuth subaperture decomposition partitions the full azimuth Doppler spectrum into K distinct sub-bands and reconstructs an image from each, obtaining subaperture looks that represent the same scene with reduced effective azimuth resolution [4, 20] (See Fig. 1). Specifically, if the k-th subaperture retains a fraction $\alpha_k = B_{D,k}/B_D$ of the total Doppler bandwidth, its effective azimuth resolution approximately degrades to:

$$\rho_{a,k} \approx \frac{\rho_a}{\alpha_k}, \tag{3}$$

This degradation does not modify the spatial sampling of the final ground-projected images since identical geocoding and resampling operators are applied. Both subaperture and full-aperture products can therefore share the same ground-range grid. The difference appears as reduced high-Doppler support and altered speckle statistics rather than spatial decimation [4].

Consequently, learning a mapping from subaperture inputs to full-aperture references can be formulated as a *same-grid* enhancement problem, which does not require SR-specific upsampling modules. In this setting, fully convolutional encoder-decoder architectures (e.g., U-Net variants) are particularly well-suited, as they naturally process fixed-size inputs while capturing multi-scale contextual features necessary for detail recovery.

Subaperture-aware Learning. Recent literature has begun to exploit subaperture decomposition as a source of complementary views for representation learning and downstream SAR understanding tasks [45, 46]. In high-level tasks, subaperture cues have been utilized for Doppler-driven ship detection and velocity estimation [27], while early explorations into subaperture-based supervision for target recognition were conducted by Wang *et al*. [51]. In the context of restoration, subapertures have been employed to denoise full-aperture imagery while mitigating the resolution loss associated with classical multi-looking [4], and into fusion strategies for complex-valued SR frameworks [20]. Unlike these approaches, our work leverages azimuth subaperture reconstruction as a scalable, physics-grounded mechanism to synthesize paired LR/HR supervision on the same grid. This strategy is specifically tailored for S1 SM enhancement, prioritizing conservative detail recovery and

speckle reduction consistent with the sensor's native imaging physics.

3. Methodology

3.1. Dataset

We developed a self-supervised dataset using S1 SM Single Look Complex (SLC) dual-polarization (VV + VH) data. The SM acquisition mode was selected due to its continuous azimuth illumination and larger intrinsic azimuth bandwidth compared to the standard IW mode, enabling higher azimuth resolution. However, SM acquisitions are less frequent and typically driven by specific tasking requests [21].

Overall, the dataset consists of 10 acquisitions selected from the limited available archive to ensure geographical diversity across continental regions, covering a total area of 162,305.4 km^2. To ensure geometric consistency and maximize data availability, we focused exclusively on ascending orbits, leaving the integration of descending passes for future research.

Pre-processing. All SLC products were processed using the Sentinel Application Platform (SNAP) [1]. Specifically, we performed precise orbit correction, radiometric calibration, and terrain correction, adhering to the standard preprocessing workflow for S1 data following the procedures outlined by Filipponi *et al.* [23]. The resulting calibrated and georeferenced backscatter intensities were converted to a decibel (dB) scale to transform multiplicative speckle into approximately additive noise [18]. To ensure training stability across heterogeneous scenes, we mitigated the impact of outliers by clipping the intensity values for each polarization channel at the 0.1^{th} and 99.9^{th} percentiles computed over the entire dataset. Finally, the clipped values were linearly rescaled to the $[0, 1]$ range to facilitate gradient-based optimization.

Splitting Protocol and Patch Extraction. To mitigate spatial leakage caused by high correlation between neighboring patches, we partitioned the dataset at the acquisition level. In this regard, entire scenes were assigned to either the training, validation, or test sets, ensuring that the model is evaluated on geographically unseen regions. For training and validation, we extracted non-overlapping 96×96 patches from each acquisition, discarding any tiles containing no-data values. Conversely, testing is performed on full scenes using a sliding window with 50% overlap. Final dataset statistics are provided in Tab. 1.

3.2. Subapertures Generation

Training inputs are generated from SLC products via azimuth subaperture decomposition. Sub-band boundaries are derived from product metadata by constructing the azimuth frequency grid and evenly partitioning the azimuth bandwidth into three non-overlapping intervals. The complex SLC data is transformed into the azimuth Doppler frequency domain via a 1-D Fast Fourier Transform (FFT). We then perform spectrum normalization and a de-weighting step to compensate for the azimuth apodization (Hamming-type weighting). Each interval is then extracted, re-centered in the spectral domain, and transformed back to the spatial domain via an Inverse FFT (IFFT) [4], yielding one complex-valued look per interval. Finally, a circular shift is applied along the azimuth direction to spatially re-center the reconstructed looks. Fig. 1 presents a simplified schema of the described processing. Although the azimuth bandwidth could be divided into an arbitrary number of intervals, increasing the number of subapertures reduces the effective Doppler support of each look and consequently degrades azimuth resolution. For this reason, we adopt a three-look, non-overlapping partition of the azimuth Doppler spectrum per acquisition. The choice of this configuration was based on preliminary experiments.

Table 1. Dataset extension statistics.

Split	# Acquisitions	Area (km^2)	# Patches
Train	5	84,401.3	301,573
Validation	3	49,434.5	210,467
Test	2	28,469.7	–

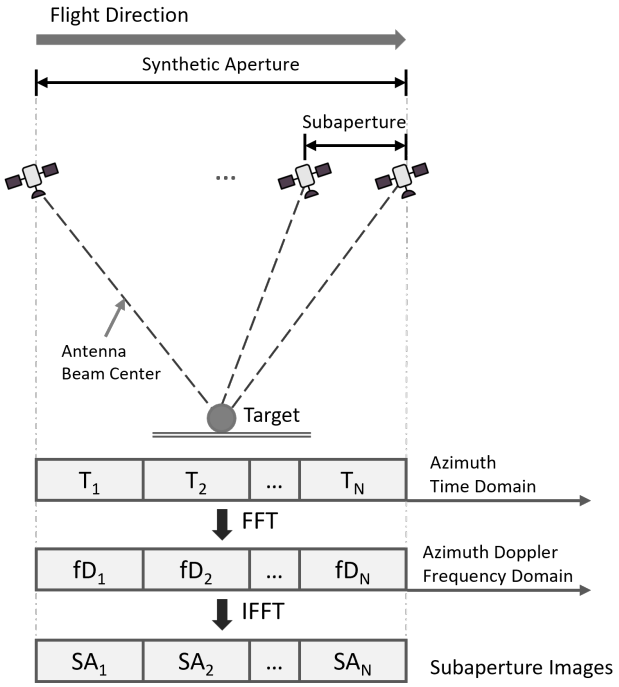


Figure 1. Simplified schema of subaperture decomposition process.

Following decomposition, the subaperture products undergo the same pre-processing workflow described in Sec. 3.1.

3.3. Problem Setup

Training. We consider subapertures as inputs and the corresponding full-aperture intensity as reference. We evaluate two distinct configurations, as summarized in Fig. 2:

- *Single-input (SI):* This configuration takes a single subaperture patch as input to predict the full-aperture reference. We employ *MA-Net* [22], a standard U-Net-like architecture with multi-scale attention, as the backbone for pixel-wise intensity regression.
- *Multi-frame (MF):* This setup jointly processes all three subaperture patches to predict the full-aperture reference. We adopt a commonly used multi-frame fusion architecture, *FPN-ConvLSTM* [9], which integrates multi-scale feature extraction via a Feature Pyramid Network (FPN) with ConvLSTM layers to fuse information across the subaperture sequence before decoding the final intensity prediction.

Inference. Given the distinct architectural designs of the two frameworks, the inference strategy varies slightly between them. For *SI*, inference is performed independently on each subaperture, resulting in three separate reconstructions that can be analyzed individually or averaged. In contrast, *MF* processes the three subapertures jointly to produce a single reconstruction (see Fig. 2). In operational settings, the SI configuration processes a single full-aperture S1 SM acquisition, whereas the MF configuration jointly processes three co-registered S1 SM acquisitions.

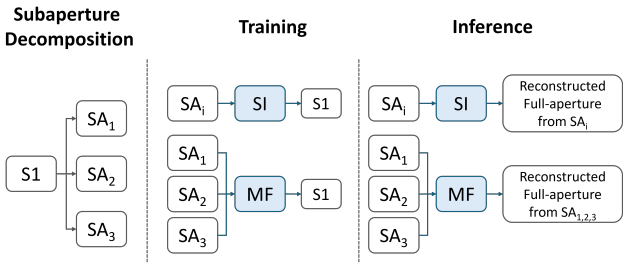


Figure 2. Summarized Problem Setup.

3.4. Iterative Refinement

Given that both proposed frameworks produce an enhanced full-aperture reconstruction from synthesized subapertures, we investigate the potential for further enhancement gains through iterative refinement. Specifically, we evaluate a recursive inference scheme where the enhanced output from an initial pass is treated as a "pseudo-subaperture" and fed back into the *SI* architecture. While the *MF* model requires a triplet of distinct subapertures and thus cannot be directly iterated, its initial output can serve as the starting point for subsequent refinement steps using the *SI* model. This approach allows us to explore whether the learned denoising and reconstruction priors can be recursively applied to progressively refine the enhanced output. Fig. 3 illustrates this iterative procedure.

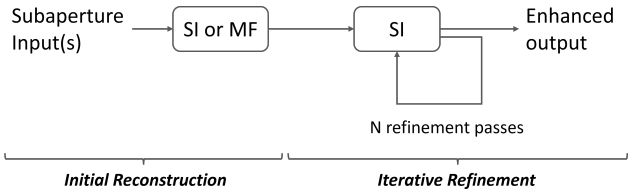


Figure 3. Iterative Refinement Process.

3.5. Experimental Setup

Both configurations are trained using the Adam optimizer [29] for 100 epochs, using a OneCycle learning rate scheduler [47] with a maximum learning rate of 10^{-3}. Due to hardware memory constraints, *SI* is trained with a batch size of 512, while *MF* utilizes a batch size of 32. All experiments are conducted on an NVIDIA RTX A5500 GPU (24 GB VRAM).

The training objective compares the network prediction $\hat{y}$ with the full-aperture reference y using a weighted combination of an ℓ_2 fidelity, Structural Similarity Index (SSIM), and a Kernel Density Estimation (KDE)-based distribution matching term [25]:

$$\mathcal{L} = \alpha \, \|\hat{y} - y\|_2^2 + \beta \, (1 - \mathrm{SSIM}(\hat{y}, y)) + \gamma \, \mathcal{L}_{\mathrm{KDE}}(\hat{y}, y). \quad (4)$$

where α, β, and γ weight the ℓ_2, SSIM, and KDE terms, respectively. We set $\alpha = 0.2$, $\beta = 0.3$, and $\gamma = 0.5$ based on preliminary experiments.

Data augmentation is limited to dihedral transformations (combination of rotations by multiples of 90° and flips), applied consistently to both inputs and targets. To improve robustness against the Doppler partitioning scheme, *SI* randomly samples one of the three subapertures as input during each training step, while *MF* is trained with random permutations of the input subaperture order. Furthermore, we apply a patch-wise histogram matching step to align the backscatter intensity distributions of the subaperture inputs with the full-aperture reference. The purpose of this operation is to ensure that the input distributions are consistent with those used during the model's operational phase, which relies on full-aperture input images.

3.6. Evaluation Framework

We evaluate our models using Peak Signal-to-Noise Ratio (PSNR) and SSIM [53] to assess reconstruction fidelity

against the full-aperture reference. Since the full-aperture reference is affected by speckle, no noise-free ground truth is available; thus, PSNR and SSIM measure consistency rather than absolute fidelity. To quantify despeckling performance, we utilize the Equivalent Number of Looks (ENL) [30], averaged over 20 manually selected homogeneous ROIs within the test set. It is worth mentioning that these ROIs have been manually selected to ensure no structural features were present. All metrics are reported as averages across the test acquisitions.

Metrics are computed according to the inference configuration:

- *SI*: Inference is performed independently on each of the three subapertures; metrics are calculated per output and subsequently averaged.
- *MF*: The three subapertures are processed jointly to generate a single reconstruction, with metrics computed directly on this output.
- *Iterative Inference:* Metrics are computed at every refinement step, according to the corresponding input for the iterative process (*SI* or *MF*).

Finally, we compare our approach against *MERLIN* [16] as a baseline using two distinct strategies. First, we apply *MERLIN* to each subaperture independently and average the results, following the *SI* protocol as a Full-Reference assessment. Second, to provide a broader comparison, we apply *MERLIN* directly to the full-aperture images. Since no noise-free ground truth exists for the full-aperture data, this constitutes a No-Reference analysis; consequently, we only report the ENL to evaluate the degree of speckle reduction without calculating fidelity metrics. Qualitative visual comparisons are additionally provided.

4. Results and Discussion

Following the framework described in Sec. 3.6, this section addresses the following research questions:

1. *Comparative Evaluation of SI, MF, and MERLIN*: How do the proposed *SI* and *MF* configurations compare in terms of reconstruction fidelity and despeckling, and how do they perform relative to the *MERLIN* baseline?
2. *Iterative Refinement*: What is the impact of recursive *SI* inference, and how does the trade-off between speckle reduction and structural detail preservation evolve across multiple refinement passes?
3. *Qualitative Assessment in an Operational Scenario*: Are the learned enhancement characteristics preserved on real full-aperture S1 SM imagery?

The following sections present the corresponding analyses to address each research question.

4.1. Question 1: Comparative Evaluation of SI, MF, and MERLIN

Tab. 2 reports the quantitative results for the proposed *SI* and *MF* configurations, alongside the *MERLIN* baseline, in terms of SSIM, PSNR, and ENL metrics computed for both polarizations (VV and VH). Note that the best results for each metric are highlighted in **boldface**.

Table 2. *SI* and *MF* compared with *MERLIN*.

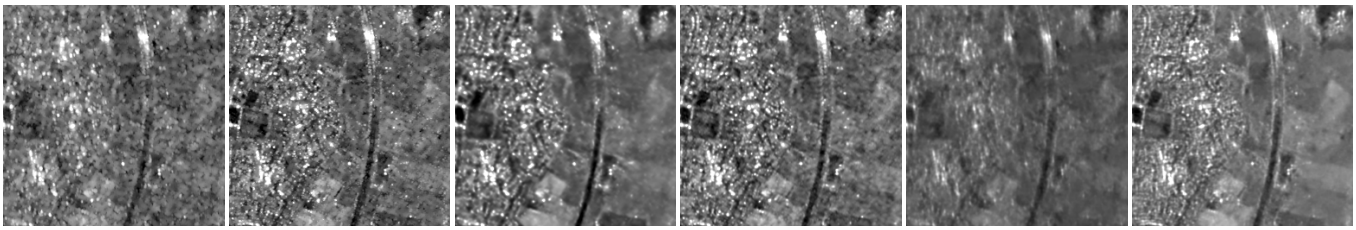
	SSIM ↑		PSNR ↑		ENL ↑	
Method	VV	VH	VV	VH	VV	VH
SI	70.5	62.4	29.2	27.6	28	26
MF	**84.2**	**78.4**	**30.3**	**28.2**	14	13
$\text{MERLIN}_{\text{Sub}}$	63.5	45.4	19.6	12.2	66	50
$\text{MERLIN}_{\text{Full}}$	–	–	–	–	**294**	**81**

Overall, both *SI* and *MF* substantially outperform the *MERLIN* baseline applied to subaperture inputs ($MERLIN_{\text{Sub}}$), yielding significant gains in reconstruction fidelity across both polarizations. The *MF* configuration achieves the highest performance, reaching a PSNR of $30.3/28.2$ dB and an SSIM of $0.842/0.784$. While the *SI* model yields lower fidelity metrics, it achieves a higher ENL than *MF*, suggesting more aggressive speckle suppression. In contrast, $MERLIN_{\text{Full}}$ produces markedly higher ENL values (294/81), reflecting substantial smoothing when applied directly to full-aperture imagery. However, this increase in ENL is not necessarily accompanied by improved structural fidelity. Fig. 4 reinforces these quantitative findings through a qualitative comparison, highlighting the trade-off between noise reduction and structural preservation across the different methods.

4.2. Question 2: Iterative Refinement

This section investigates the impact of recursive *SI* inference and the resulting trade-off between speckle reduction and structural preservation. Quantitative results for varying refinement passes, using initial reconstructions generated by either *SI* or *MF*, are summarized in Tab. 3.

Across both polarizations, increasing the number of refinement passes results in a consistent increase in ENL and a corresponding decrease in PSNR and SSIM. This trend explicitly quantifies the speckle–detail trade-off: while repeated passes enhance image smoothness, they progressively degrade reconstruction fidelity. For instance, when the initial reconstruction is produced by *SI*, two additional passes boost ENL from $28/26$ to $105/102$ (VV/VH), yet PSNR drops from $29.2/27.6$ dB to $26.3/24.7$ dB, and SSIM from $0.705/0.624$ to $0.662/0.571$. At these high pass counts, ENL values approach or exceed the $MERLIN_{Full}$ benchmark; however, the substantial degradation in fidelity



SA_1 | Full Aperture Reference (PSNR / SSIM) | *SI* (26.80 dB / 69.81) | *MF* (31.24 dB / 86.59) | $MERLIN_{Sub}$ (25.22 dB / 59.28) | $MERLIN_{Full}$

Figure 4. Visual comparison (VV polarization) between the *MERLIN* baseline and the proposed *SI* and *MF* methods. Only the first subaperture (SA_1) is shown due to space constraints.

Table 3. Effect of additional *SI* refinement passes. "Init." denotes the model used to generate the initial reconstruction (*SI* or *MF*). "Passes" denotes the number of additional *SI* refinement passes.

Init.	Passes	SSIM ↑		PSNR ↑		ENL ↑	
		VV	VH	VV	VH	VV	VH
SI	-	70.5	62.4	29.2	27.6	28	26
	1	68.8	60.3	27.8	26.3	52	55
	2	66.2	57.1	26.3	24.7	105	102
	3	63.6	53.9	25.0	23.3	201	207
	4	61.4	51.2	24.0	22.2	**274**	**395**
MF	-	**84.2**	**78.4**	**30.3**	**28.2**	14	13
	1	78.9	73.1	29.2	27.8	31	31
	2	72.8	66.2	27.1	26.0	60	68
	3	67.8	60.3	25.4	24.3	113	124
	4	63.9	55.5	24.1	22.9	213	235

indicates a increased risk of over-smoothing and loss of critical spatial features.

A particularly favorable operating point emerges when applying a single *SI* refinement pass to the *MF* output. This configuration effectively shifts the high-fidelity *MF* result toward stronger despeckling. Specifically, one SI pass on the *MF* initial reconstruction increases ENL from $14/13$ to $31/31$, while maintaining a robust SSIM of $0.789/0.731$.

Overall, applying *SI* refinement passes to the *MF* output yields a more balanced reconstruction than the *SI* model alone. The *MF* formulation provides complementary information across subaperture inputs, establishing a strong baseline that subsequent *SI* refinement further regularizes. Qualitative examples in Fig. 5 support this, showing that the first few refinement passes significantly reduce speckle with limited structural alterations, while excessive iterations eventually compromise fine details.

4.3. Question 3: Qualitative Assessment in an Operational Scenario

Since no reference data is available for full-aperture inputs, quantitative evaluation cannot be performed in this operational setting. Therefore, Fig. 6 presents qualitative results obtained by applying *MERLIN* and our best-performing configuration MF_1 directly to full-aperture S1 SM imagery. For a fair comparison with *MERLIN*, the same S1 SM acquisition was replicated three times to match the multi-frame structure of *MF*, preventing the exploitation of multi-temporal information.

The proposed MF_1 exhibits improved structural sharpness while avoiding the stronger smoothing observed in $MERLIN_{Full}$.

Moreover, to complement the presented results and better assess achievable reconstruction quality, full-acquisition inference examples are provided as supplementary material at [url].

5. Conclusions and Future Work

In this work, we have presented a self-supervised framework for enhancing S1 SM imagery by leveraging azimuth Doppler subaperture decomposition. By treating the full-aperture intensity as a sensor-consistent reference, we developed two distinct architectures: Single-Input (SI) and Multi-Frame (MF), that effectively learn to reconstruct high-fidelity images while suppressing speckle noise.

Our results demonstrate that the *MF* configuration, which exploits diversity across the azimuth Doppler spectrum, achieves higher structural consistency with the full-aperture reference. Conversely, the *SI* model provides stronger despeckling capabilities. We further explored iterative refinement, showing that recursive passes through the *SI* architecture substantially increase ENL but eventually lead to diminishing returns, where high-frequency details are compromised. Finally, compared to the baseline self-supervised enhancement method *MERLIN*, the proposed approach achieves superior overall performance.

For future work, we aim to extend this framework to other sensors, acquisition modes, polarizations, and flight directions, as well as to incorporate multi-temporal sequences to further improve the robustness of the self-supervised priors. We also plan to investigate the effect of varying the number of azimuth subapertures used for supervision, analyzing how this design choice influences reconstruction fidelity and speckle suppres-

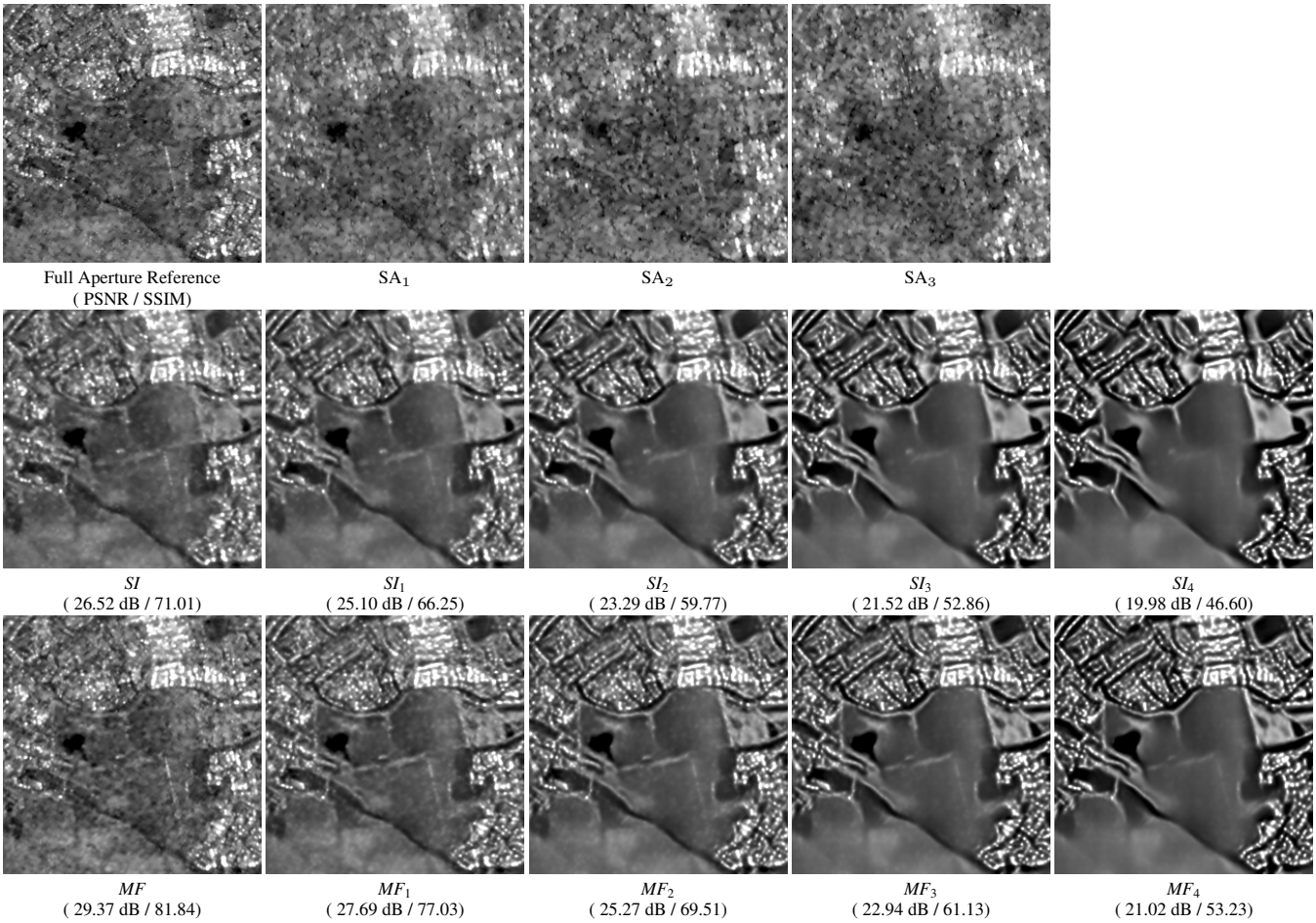


Figure 5. Qualitative comparison (VV polarization) of subaperture inputs, full-aperture reference, and reconstructions. The second and third rows correspond to *SI* and *MF* models; subindices denote inference refinement passes (1–4).

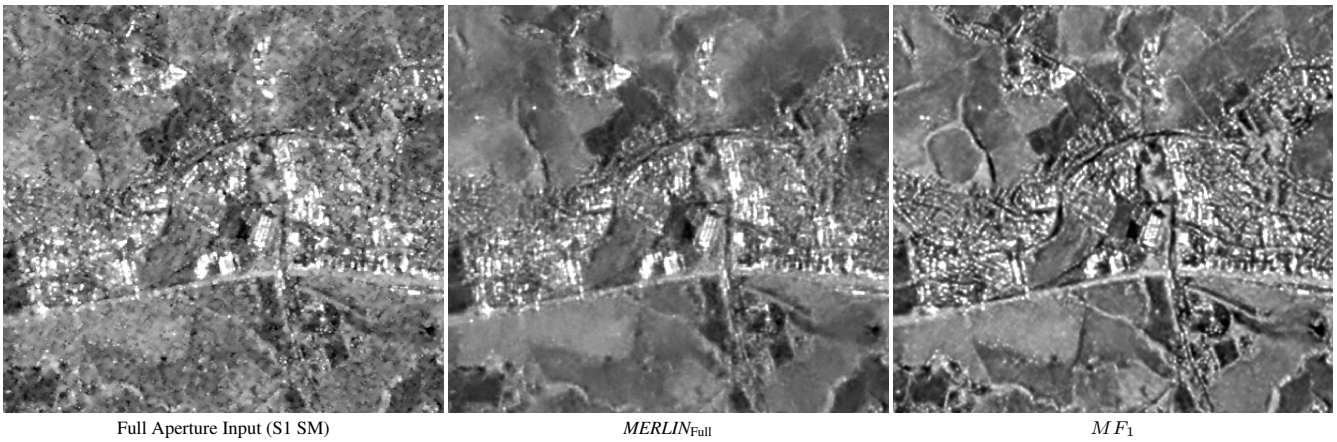


Figure 6. Visual results on real full-aperture S1 SM imagery (VV polarization), comparing the proposed MF_1 versus $MERLIN_{\text{Full}}$.

sion. Furthermore, we will assess the impact of these enhanced products on downstream tasks such as maritime object detection and land-cover classification, assessing whether the improved visual quality translates into higher automated analysis performance. Finally, we will explore alternative strategies for constructing more robust reference targets to enable a more comprehensive quantitative evaluation of enhancement performance.

References

[1] European Space Agency. Snap - esa sentinel application platform. 4

[2] Juan Francisco Amieva, Christian Ayala, and Mikel Galar. Deep learning for enhancing sentinel-1 imagery. In *2025 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 2025. 1

[3] Juan Francisco Amieva, Christian Ayala, and Mikel Galar. A deep learning approach to jointly super-resolve and despeckle sentinel-1 imagery. *Acta Astronautica*, 238:1396–1407, 2026. 1, 3

[4] Xinwei An, Hongcheng Zeng, Zhaohong Li, Wei Yang, Wei Xiong, Yamin Wang, and Yanfang Liu. Sar images despeckling using subaperture decomposition and non-local low-rank tensor approximation. *Remote Sensing*, 17(15):2716, 2025. 1, 2, 3, 4

[5] Dongyang Ao, Corneliu Octavian Dumitru, Gottfried Schwarz, and Mihai Datcu. Dialectical gan for sar image translation: From sentinel-1 to terrasar-x. *Remote Sensing*, 10(10):1597, 2018. 1, 3

[6] Christian Ayala, Rubén Sesma, Carlos Aranda, and Mikel Galar. A deep learning approach to an enhanced building footprint and road detection in high-resolution satellite imagery. *Remote Sensing*, 13(16):3135, 2021. 1

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

[8] Antoni Buades, Bartomeu Coll, and J-M Morel. A non-local algorithm for image denoising. In *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)*, pages 60–65. Ieee, 2005. 2

[9] Jorge Andres Chamorro Martinez, Laura Elena Cué La Rosa, Raul Queiroz Feitosa, Ieda Del'Arco Sanches, and Patrick Nigri Happ. Fully convolutional recurrent networks for multidate crop recognition from multitemporal image sequences. *ISPRS Journal of Photogrammetry and Remote Sensing*, 171:188–201, 2021. 5

[10] Kelvin CK Chan, Shangchen Zhou, Xiangyu Xu, and Chen Change Loy. Basicvsr++: Improving video super-resolution with enhanced propagation and alignment. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5972–5981, 2022. 3

[11] Giovanni Chierchia, Davide Cozzolino, Giovanni Poggi, Giuseppe Scarpa, and Luisa Verdoliva. SAR image despeckling through convolutional neural networks. In *2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 2017. 2

[12] Marco Chini, Ramona Pelich, Radu Hostache, Patrick Matgen, and Carlos Lopez-Martinez. Towards a 20 m global building map from sentinel-1 SAR data. *Remote Sensing*, 10 (11):1833, 2018. 1

[13] Tao Dai, Jianrui Cai, Yongbing Zhang, Shu-Tao Xia, and Lei Zhang. Second-order attention network for single image super-resolution. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 11065–11074, 2019. 3

[14] Emanuele Dalsasso, Xiangli Yang, Loïc Denis, Florence Tupin, and Wen Yang. SAR image despeckling by deep neural networks: From a pre-trained model to an end-to-end training strategy. *Remote Sensing*, 12(16), 2020. 2

[15] Emanuele Dalsasso, Loïc Denis, and Florence Tupin. SAR2SAR: A semi-supervised despeckling algorithm for SAR images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 14:4321–4329, 2021. 1, 2

[16] Emanuele Dalsasso, Loïc Denis, and Florence Tupin. As if by magic: Self-supervised training of deep despeckling networks with MERLIN. *IEEE Transactions on Geoscience and Remote Sensing*, 60, 2022. 1, 2, 6

[17] Charles-Alban Deledalle, LoÏc Denis, and Florence Tupin. Iterative weighted maximum likelihood denoising with probabilistic patch-based weights. *IEEE Transactions on Image Processing*, 18(12):2661–2672, 2009. 2

[18] Charles-Alban Deledalle, Loïc Denis, and Florence Tupin. How to compare noisy patches? patch similarity beyond gaussian noise. *International Journal of Computer Vision*, 99(1):86–102, 2012. 4

[19] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Image super-resolution using deep convolutional networks. *IEEE transactions on pattern analysis and machine intelligence*, 38(2):295–307, 2015. 2

[20] Ganggang Dong, Yao Wang, Hongwei Liu, and Songlin Liu. Complex-valued sar image super-resolution via subaperture learning and fusion. *IEEE Transactions on Geoscience and Remote Sensing*, 63:1–14, 2025. 2, 3

[21] European Space Agency (ESA). Sentinel-1 user handbook. ESA Standard Document, Issue 1 Rev 0, 2013. 1, 4

[22] Tongle Fan, Guanglei Wang, Yan Li, and Hongrui Wang. Ma-net: A multi-scale attention network for liver and tumor segmentation. *IEEE Access*, 8:179656–179665, 2020. 5

[23] Federico Filipponi. Sentinel-1 grd preprocessing workflow. In *International Electronic Conference on Remote Sensing*, page 11. MDPI, 2019. 4

[24] Giulia Fracastoro, Enrico Magli, Giovanni Poggi, Giuseppe Scarpa, Diego Valsesia, and Luisa Verdoliva. Deep learning methods for synthetic aperture radar image despeckling: An overview of trends and perspectives. *IEEE Geoscience and Remote Sensing Magazine*, 9(2):29–51, 2021. 2

[25] Alejandro Gomez-Alanis, Jose A. Gonzalez-Lopez, and Antonio M. Peinado. A kernel density estimation based loss function and its application to asv-spoofing detection. *IEEE Access*, 8:108530–108543, 2020. 5

[26] Sanhita Guha, Mihai Datcu, and Joachim Ender. Sar super-resolution using physics-aware adaptive compressed sensing. In *IGARSS 2022-2022 IEEE International Geoscience and Remote Sensing Symposium*, pages 52–55. IEEE, 2022. 1, 3

[27] Muhammad Amjad Iqbal, Andrei Anghel, and Mihai Datcu. Subaperture decomposition analysis for accurate ship detection and velocity estimation in synthetic aperture radar imagery. *Remote Sensing Letters*, 16(1):55–65, 2025. 3

[28] Piet Dewaele PatrickWambacq Jong-Sen Lee, L Jurkevich and André Oosterlinck. Speckle filtering of synthetic aperture radar images: A review. *Remote Sensing Reviews*, 8(4): 313–340, 1994. 1

[29] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *International Conference on Learning Representations (ICLR)*, 2015. 5

[30] Jaekyun Ko and Sanghwan Lee. Sar image despeckling using continuous attention module. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 15:3–19, 2022. 6

[31] 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

[32] Darwin T. Kuan, Alexander A. Sawchuk, Timothy C. Strand, and Pierre Chavel. Adaptive noise smoothing filter for images with signal-dependent noise. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-7(2):165–177, 1985. 2

[33] Federico Lattari, Claudio González, D. León, F. Asaro, and Matteo Matteucci. Deep learning for SAR image despeckling. *Remote Sensing*, 11(13), 2019. 2

[34] Jong-Sen Lee. Digital image smoothing and the sigma filter. *Computer Vision, Graphics, and Image Processing*, 24(2): 255–269, 1983. 2

[35] Jaakko Lehtinen, Jacob Munkberg, Jon Hasselgren, Samuli Laine, Tero Karras, Miika Aittala, and Timo Aila. Noise2noise: Learning image restoration without clean data. *arXiv preprint arXiv:1803.04189*, 2018. 2

[36] Jingyun Liang, Yuchen Fan, Xiaoyu Xiang, Rakesh Ranjan, Eddy Ilg, Simon Green, Jiezhang Cao, Kai Zhang, Radu Timofte, and Luc V Gool. Recurrent video restoration transformer with guided deformable attention. *Advances in Neural Information Processing Systems*, 35:378–393, 2022. 3

[37] Jingyun Liang, Jiezhang Cao, Yuchen Fan, Kai Zhang, Rakesh Ranjan, Yawei Li, Radu Timofte, and Luc Van Gool. Vrt: A video restoration transformer. *IEEE Transactions on Image Processing*, 33:2171–2182, 2024. 3

[38] Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee. Enhanced deep residual networks for single image super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 136–144, 2017. 3

[39] Zhisheng Lu, Juncheng Li, Hong Liu, Chaoyan Huang, Linlin Zhang, and Tieyong Zeng. Transformer for single image super-resolution. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 457–466, 2022. 3

[40] Carmine Massarelli. Super-resolution of sentinel-2 satellite images: A comparison of different interpolation methods for spatial knowledge extraction. *Machine Learning and Knowledge Extraction*, 8(1), 2026. 2

[41] 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, 2022. 2

[42] Adugna G Mullissa, Diego Marcos, Devis Tuia, Martin Herold, and Johannes Reiche. Despecknet: Generalizing deep learning-based sar image despeckling. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–15, 2020. 2

[43] Sara Parrilli, Mariana Poderico, Cesario Vincenzo Angelino, and Luisa Verdoliva. A nonlocal sar image denoising algorithm based on llmmse wavelet shrinkage. *IEEE Transactions on Geoscience and Remote Sensing*, 50(2):606–616, 2012. 2

[44] Malsha V. Perera, Wele Gedara Chaminda Bandara, Jeya Maria Jose Valanarasu, and Vishal M. Patel. Transformer-based sar image despeckling. In *IGARSS 2022 - 2022 IEEE International Geoscience and Remote Sensing Symposium*, pages 751–754, 2022. 2

[45] Nicolae-Cătălin Ristea, Andrei Anghel, Mihai Datcu, and Bertrand Chapron. Guided deep learning by subaperture decomposition: Ocean patterns from sar imagery. In *IGARSS 2022-2022 IEEE International Geoscience and Remote Sensing Symposium*, pages 6825–6828. IEEE, 2022. 2, 3

[46] Nicolae-Cătălin Ristea, Andrei Anghel, Mihai Datcu, and Bertrand Chapron. Guided unsupervised learning by subaperture decomposition for ocean sar image retrieval. *IEEE Transactions on Geoscience and Remote Sensing*, 61:1–11, 2023. 3

[47] Leslie N. Smith and Nicholay Topin. Super-convergence: very fast training of neural networks using large learning rates. In *Defense + Commercial Sensing*, 2018. 5

[48] Ramon Torres, Paul Snoeij, Malcolm Davidson, David Bibby, Steve Lokas, et al. The sentinel-1 mission and its application capabilities. In *2012 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pages 1703–1706. IEEE, 2012. 1

[49] Andreas Vollrath, Adugna Mullissa, and Johannes Reiche. Angular-based radiometric slope correction for sentinel-1 on google earth engine. *Remote Sensing*, 12(11), 2020. 1

[50] Longgang Wang, Mana Zheng, Wenbo Du, Menglin Wei, and Lianlin Li. Super-resolution sar image reconstruction via generative adversarial network. In *2018 IEEE International Symposium on Antennas and Propagation and USNC-URSI Radio Science Meeting (ISAPE)*, 2018. 1, 3

[51] Ruonan Wang, Zhaocheng Wang, Yu Chen, Hailong Kang, Feng Luo, and Yingxi Liu. Target recognition in sar images using complex-valued network guided with sub-aperture decomposition. *Remote Sensing*, 15(16):4031, 2023. 2, 3

[52] Xintao Wang, Kelvin CK Chan, Ke Yu, Chao Dong, and Chen Change Loy. Edvr: Video restoration with enhanced deformable convolutional networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, pages 0–0, 2019. 3

[53] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004. 5

[54] Wenming Yang, Xuechen Zhang, Yapeng Tian, Wei Wang, Jing-Hao Xue, and Qingmin Liao. Deep learning for single image super-resolution: A brief review. *IEEE Transactions on Multimedia*, 21(12):3106–3121, 2019. 2

[55] LI Yanshan, ZHOU Li, XU Fan, and CHEN Shifu. Ogsrn: Optical-guided super-resolution network for sar image. *Chinese Journal of Aeronautics*, 35(5):204–219, 2022. 3

[56] Ye Yuan, Yanxia Wu, Yan Fu, Yulei Wu, Lidan Zhang, and Yan Jiang. An advanced sar image despeckling method by bernoulli-sampling-based self-supervised deep learning. *Remote Sensing*, 13(18), 2021. 2

[57] Yulun Zhang, Kunpeng Li, Kai Li, Lichen Wang, Bineng Zhong, and Yun Fu. Image super-resolution using very deep residual channel attention networks. In *Proceedings of the European conference on computer vision (ECCV)*, pages 286–301, 2018. 3

[58] Ce Zheng, Xue Jiang, Ye Zhang, Xingzhao Liu, Bin Yuan, and Zhixin Li. Self-normalizing generative adversarial network for super-resolution reconstruction of sar images. In *IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium*, pages 1911–1914, 2019. 3

[59] 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. 1, 2, 3