# Sim2Real SAR Image Restoration: Metadata-Driven Models for Joint Despeckling and Sidelobes Reduction

Antoine De Paepe[1], Pascal Nguyen[2], Michael Mabelle[2], Cédric Saleun[1], Antoine Jouadé[1], and Jean-Christophe Louvigne[1]

[1]Direction Générale de l'Armement Maîtrise de l'Information, Bruz, France.
[2]Agence Ministérielle pour l'Intelligence Artificielle de Défense, Bruz, France.

***Abstract*—Synthetic aperture radar (SAR) provides valuable information about the Earth's surface under all weather and illumination conditions. However, the inherent phenomenon of speckle and the presence of sidelobes around bright targets pose challenges for accurate interpretation of SAR imagery. Most existing SAR image restoration methods address despeckling and sidelobes reduction as separate tasks. In this paper, we propose a unified framework that jointly performs both tasks using neural networks (NNs) trained on a realistic SAR simulated dataset generated with MOCEM. Inference can then be performed on real SAR images, demonstrating effective simulation to real (Sim2Real) transferability. Additionally, we incorporate acquisition metadata as auxiliary input to the NNs, demonstrating improved restoration performance.**

***Index Terms*—SAR, Image Restoration, Despeckling, Sidelobes Reduction, Metadata Injection, Sim2Real**

## NOTATION AND TERMINOLOGY

Throughout this paper, the term SAR image refers broadly to SAR data, including both single look complex (SLC) images and their amplitude representations. When necessary for clarity, we explicitly distinguish between the two. For Section II, all vectors are represented as column vectors. The symbol '$\top$' denotes the matrix transpose operation. For a given column vector $\boldsymbol{s} = [s_1, s_2, \cdots, s_n]^\top \in \mathcal{S}^n$, with $\mathcal{S}$ being either $\mathbb{R}$ or $\mathbb{C}$, $s_k$ denotes the $k$th entry of $\boldsymbol{s}$. For the specific case where $\mathcal{S} \triangleq \mathbb{C}$, we write $\boldsymbol{s} = \boldsymbol{s}^{\text{re}} + i\boldsymbol{s}^{\text{im}}$, with $\boldsymbol{s}^{\text{re}} \in \mathbb{R}^n$ denoting the real part and $\boldsymbol{s}^{\text{im}} \in \mathbb{R}^n$ denoting the imaginary part. We define $\boldsymbol{D}_{\boldsymbol{s}} \triangleq \text{diag}(\boldsymbol{s}) \in \mathcal{S}^{n\times n}$ as the diagonal matrix whose diagonal entries are given by the components of the vector $\boldsymbol{s}$.

## I. INTRODUCTION

SAR imaging has emerged as a crucial remote sensing technology, extensively applied across a range of domains, including environmental monitoring, disaster management, and military surveillance. Its capability to acquire high-resolution imagery under various conditions renders SAR a valuable tool for observing and analyzing the Earth's surface.

Despite its many advantages, SAR imagery is often compromised by inherent noise—particularly speckle—as well as sidelobes surrounding bright targets, which can obscure critical details. To improve the quality of SAR image interpretation, it is essential to address the challenges posed by both speckle noise and the presence of sidelobes.

A variety of sidelobes reduction techniques have been developed, each targeting different aspects of SAR signal processing. The most widely used methods are windowing functions [1], such as Hamming, Hanning, and Kaiser windows, which reduce sidelobes amplitude by smoothing the radar signal during processing, albeit at the cost of some resolution loss. More sophisticated approaches aim to minimize sidelobes levels without significantly compromising resolution or signal strength, such as spatial variant apodization (SVA) [2]. However, SVA presents several drawbacks: it distorts statistics in speckle-dominated regions, spreads point-like targets across multiple pixels, and introduces a negative bias, making homogeneous areas appear less reflective [3]. Recent advancements in deep learning have introduced convolutional neural network (CNN)-based [4] and bidirectional recurrent neural network (RNN)-based [5] models to mitigate these issues.

Parallel to sidelobes reduction techniques, SAR image despeckling has been extensively studied and is now commonly addressed through two main categories of methods: variational approaches and learning-based techniques.

Variational methods typically formulate the problem as the minimization of a cost function balancing a data fidelity term—which accounts for the statistical properties of speckle under the Goodman speckle model [6]—and a regularization term, which incorporates prior knowledge of the underlying image [7, 8].

Deep learning-based methods are usually trained on natural image pairs augmented with synthetic speckle noise to effectively learn the despeckling task. SAR-CNN [9] and ID-CNN [10] were specifically designed to handle the multiplicative nature of speckle noise, inspired by the DnCNN architecture [11]. Additionally, UNet architectures [12], initially developed for medical image segmentation, have proven highly effective in natural image restoration tasks [13, 14, 15], and their adaptability has been demonstrated in SAR despeckling applications [16, 17]. These networks excel at capturing multi-scale features, which is crucial for restoring fine details in noisy SAR images. Generative adversarial networks (GANs) have also emerged as a powerful tool for SAR image restoration [18, 19]. A more recent development in SAR image restoration is the application of conditional diffusion processes [20, 21] for despeckling SAR images. Finally, self-supervised learning approaches [22, 23], which eliminate the reliance on ground truth (GT) data, are emerging as effective solutions for SAR image restoration.

Despite the promise of deep learning methods, most are primarily designed to reduce spatially uncorrelated speckle due to their training strategies. However, in real SAR images, speckle is often spatially correlated through the SAR transfer function. This domain gap can lead to suboptimal restoration in practical applications. Few methods have been developed to address this challenge, such as MuLoG-DRUNet [24], a plug-and-play (PnP) method that stands out as the first robust technique to reduce spatially correlated speckle.

In this paper, we propose a novel approach for joint despeckling and sidelobes reduction for SAR image restoration using deep learning-based methods. We use MOCEM [25] to create a dataset of SAR images with access to GT, specifically designed to replicate the speckle characteristics and effects of the SAR transfer function observed in real-world scenarios. This dataset is used for a supervised learning task, and its consistency with real SAR acquisitions enables effective Sim2Real SAR image restoration. Furthermore, we propose incorporating SAR acquisition metadata to enhance the restoration process. By integrating metadata into the deep learning framework, we provide contextual information that allows the NNs to tailor its restoration strategy according to specific acquisition conditions.

## II. METHODOLOGY

### A. Problem Formulation

The SAR acquisition process involves a sensor, typically mounted on an aircraft or satellite, which acts as an active system that emits electromagnetic waves toward the Earth's surface and captures the reflected signals as a set of complex-valued measurements. These measurements are coherently combined to synthesize a long virtual aperture, forming the SLC image $\boldsymbol{y} \in \mathcal{Y} \triangleq \mathbb{C}^n$ of the scene reflectivity $\boldsymbol{x} \in \mathcal{X} \triangleq \mathbb{R}^n_+$ being imaged. For interpretability and practical processing, SAR images are also expressed in terms of the amplitude of the SLC, defined as $\tilde{\boldsymbol{y}} \triangleq |\boldsymbol{y}|$, with $\tilde{\boldsymbol{y}} \in \tilde{\mathcal{Y}} \triangleq \mathbb{R}^n_+$. Owing to the impulse response of the SAR system and the coherent nature of radar signals, SAR images are often degraded by both sidelobes and spatially correlated speckle. We denote by $\boldsymbol{H} \in \mathbb{R}^{n \times n}$ the linear spatial-domain operator associated with the SAR transfer function. Under the Goodman model, it is commonly assumed that for a restoration task, the measurement $\boldsymbol{y}$, in the absence of electronic noise, follows a circular Gaussian distribution [22] with independent entries given by

$$(\boldsymbol{y}|\boldsymbol{H}, \boldsymbol{x}) \sim \mathcal{CN}(\mathbf{0}_{\mathcal{X}}, \boldsymbol{H}\boldsymbol{D}_{\boldsymbol{x}}\boldsymbol{H}^\top), \tag{1}$$

where $\mathcal{CN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes a complex circular Gaussian distribution with mean $\boldsymbol{\mu} \in \mathbb{R}^n$ and covariance matrix $\boldsymbol{\Sigma} \in \mathbb{R}^{n \times n}$. This model can be reparametrized to exhibit the multiplicative nature of speckle as

$$\boldsymbol{y} = \boldsymbol{H}(\mathcal{C}(\boldsymbol{x}^{\frac{1}{2}}) \odot \boldsymbol{n}), \quad \boldsymbol{n} \sim \mathcal{CN}(\mathbf{0}_{\mathcal{X}}, \boldsymbol{I}_{\mathcal{X}}), \tag{2}$$

where $\mathcal{C} : \mathbb{R}^n \to \mathbb{C}^n$ is the complexification operator such that $\mathcal{C}(\boldsymbol{x}) = \boldsymbol{x} + i\boldsymbol{x}$, and $\boldsymbol{n}$ represents the multiplicative speckle noise in its complex form.

When the transfer function $\boldsymbol{H}$ is known, image restoration can be achieved by finding the maximum a posteriori (MAP) estimate of $\boldsymbol{x}$, i.e.,

$$\max_{\boldsymbol{x} \in \mathcal{X}} p(\tilde{\boldsymbol{y}}|\boldsymbol{H}, \boldsymbol{x}) \cdot p(\boldsymbol{x}), \tag{3}$$

where the conditional probability density function (PDF) $p(\tilde{\boldsymbol{y}}|\boldsymbol{H}, \boldsymbol{x})$ is given by (1), and $p(\boldsymbol{x})$ is a prior PDF on $\boldsymbol{x}$, which is generally unknown and replaced (in its post-log form) by a regularizer promoting piecewise smoothness. However, solving the inverse problem in (3) relies on knowledge of $\boldsymbol{H}$, which may be unavailable in many real-world applications—particularly when the processing pipeline or acquisition geometry is proprietary or undisclosed. In such cases, traditional model-based restoration approaches become inapplicable or unreliable due to the absence of an accurate system model, especially in the joint task of despeckling and sidelobes reduction.

To address this limitation, we consider an alternative strategy that leverages a large collection of known SAR transfer functions to train NN $f_{\boldsymbol{\theta}} : \tilde{\mathcal{Y}} \to \mathcal{X}$, parameterized by $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, which can generalize to unseen or implicitly characterized systems. The training process is formulated as the minimization problem

$$\min_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \mathbb{E}_{(\boldsymbol{x}, \tilde{\boldsymbol{y}}) \sim p(\boldsymbol{x}, \tilde{\boldsymbol{y}})} [\ell(f_{\boldsymbol{\theta}}(\tilde{\boldsymbol{y}}), \boldsymbol{x})], \tag{4}$$

where $p(\boldsymbol{x}, \tilde{\boldsymbol{y}}) = \int p(\tilde{\boldsymbol{y}}|\boldsymbol{H}, \boldsymbol{x}) p(\boldsymbol{H}) p(\boldsymbol{x}) d\boldsymbol{H}$ is the joint PDF of $\boldsymbol{x}$ and $\tilde{\boldsymbol{y}}$, marginalized over $p(\boldsymbol{H})$, the PDF of possible SAR transfer functions, and $\ell : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is a loss function. In this way, the model implicitly learns to invert the composite effect of the transfer function $\boldsymbol{H}$ and the speckle noise. This approach enables model-free SAR image restoration, particularly suited to operational contexts.

### B. Sim2Real: Learning with Simulations for Real-World Restoration

In real-world scenarios, obtaining a large dataset of paired samples $(\boldsymbol{x}, \tilde{\boldsymbol{y}}) \sim p(\boldsymbol{x}, \tilde{\boldsymbol{y}})$ for supervised learning is nearly impossible. It therefore becomes necessary to overcome this limitation through simulation by generating synthetic training datasets. We propose using simulated SAR images obtained with the MOCEM [25] software. This simulation tool for modeling SAR acquisition enables complete replication of the SAR imaging chain, producing high-fidelity SAR images from computer aided design (CAD) models while incorporating fundamental electromagnetic material properties. In particular, it provides access to the observed SLC image $\boldsymbol{y} \in \mathcal{Y}$, an intermediate SLC image $\boldsymbol{z} \in \mathcal{Z}$ obtained before the system transfer function $\boldsymbol{H}$ is applied (corresponding to $\mathcal{C}(\boldsymbol{x}^{\frac{1}{2}}) \odot \boldsymbol{n}$ in (2)), and the scene reflectivity $\boldsymbol{x} \in \mathcal{X}$. Although $\boldsymbol{x}$ does not represent a physically measurable quantity, it serves as an effective GT candidate because it is easily interpretable by human operators. The different images are illustrated in Figure 1.

(a) Scene reflectivity $\boldsymbol{x}$.

(b) Amplitude of SLC $\boldsymbol{z}$.

(c) Amplitude of SLC $\boldsymbol{y}$.

Fig. 1: Images produced with MOCEM.

The MOCEM software also provides, for each scene, detailed metadata $\boldsymbol{m} = [m_1, \cdots, m_d]^\top \in \mathcal{M} \triangleq \mathbb{R}^d$, where $d$ is the number of metadata parameters per scene, including radar geometry, squint angle, acquisition resolution, noise level, and other acquisition-specific parameters. This supplementary information can be leveraged to further enhance model performance.

Our approach is to train restoration NNs exclusively on simulated SAR images generated by MOCEM, taking advantage of both the known GT reflectivity scenes and the associated metadata. Once trained, these models are directly applied to real SAR acquisitions for inference, enabling image restoration despite the absence of paired real data. This Sim2Real strategy bridges the gap between synthetic training and operational deployment by exploiting the physical fidelity of simulation to generalize effectively to real-world SAR imaging scenarios.

### *C. Input Variations and Custom Loss Functions for Improved Performance*

For the supervised learning task, the minimization problem (4) for training the NN $f_{\boldsymbol{\theta}}$ can be generalized as

$$\min_{\boldsymbol{\theta}} \mathbb{E}_{(\boldsymbol{v},\boldsymbol{w})\sim p(\boldsymbol{v},\boldsymbol{w})}[\ell(f_{\boldsymbol{\theta}}(\boldsymbol{v}),\boldsymbol{w})] \tag{5}$$

where $\boldsymbol{v} \in \mathcal{V}$ is the input of the NN, $\boldsymbol{w} \in \mathcal{W}$ is the target and $p(\boldsymbol{v}, \boldsymbol{w})$ is the joint distribution. Defining the set of possible inputs, targets, loss functions, and model architectures can have a significant impact on restoration performance. As a model choice, considering the effectiveness of the DRUNet architecture [26], originally designed for Gaussian denoising in a PnP ADMM framework, we chose to retain it and adapt it for SAR image restoration. We also implemented a DRUNet enhanced with Squeeze-and-Excitation (SE) blocks—modules that perform a squeeze operation via global pooling followed by an multilayer perceptron (MLP) generating channel-wise excitation weights to recalibrate feature maps [27]—which we call SEDRUNet. We retain $\boldsymbol{v} = (\tilde{\boldsymbol{y}}, \boldsymbol{y}^{\text{re}}, \boldsymbol{y}^{\text{im}})$, stacking the amplitude $\tilde{\boldsymbol{y}}$ alongside both the real and imaginary parts, $\boldsymbol{y}^{\text{re}}$ and $\boldsymbol{y}^{\text{im}}$, as the input of the NNs. Although the amplitude information is contained in the real and imaginary parts of $\boldsymbol{y}$, we experimented (results not shown in the paper) and found that using information from these three channels improves performance. The target is set to $\boldsymbol{w} = \boldsymbol{x}$. We propose training the DRUNet architectures using different loss functions $\ell$: the mean absolute error (MAE), commonly used in image restoration tasks; the perceptual loss (PL) [28], which compares high-level feature representations extracted from a pretrained network to preserve perceptual quality; and the edge preserving loss (EPL) [29], which emphasizes the retention of sharp edges and fine structures by penalizing differences in image gradients. For the following, we denote by DRUNet EPL (resp. DRUNet PL) the models trained using the corresponding loss function.

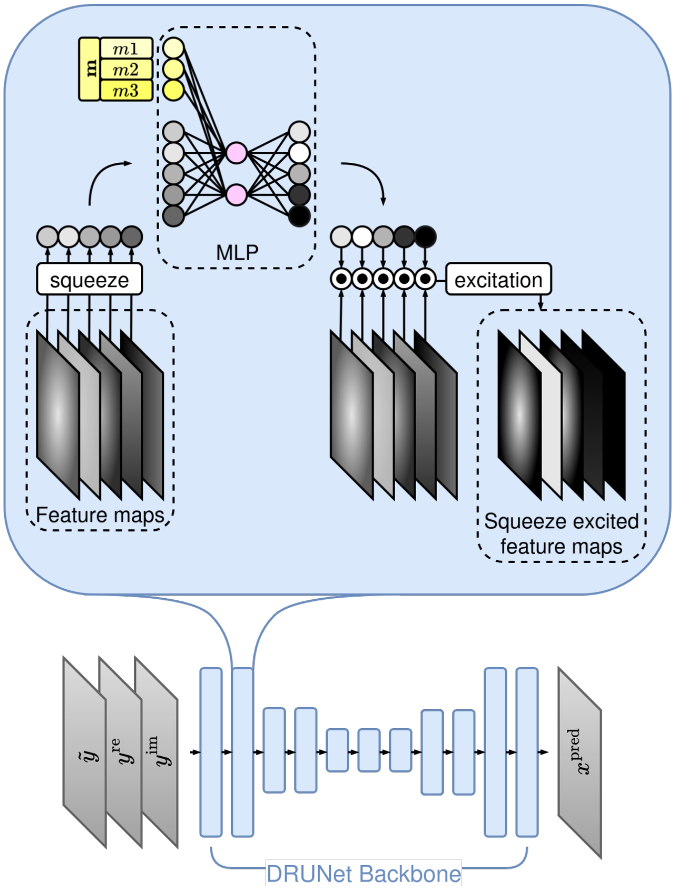


Fig. 2: Overview of the architecture of M-SEDRUNet.

Finally, we also leveraged the available metadata in the dataset to enhance the model's restoration. In that context, $\boldsymbol{v} = (\tilde{\boldsymbol{y}}, \boldsymbol{y}^{\text{re}}, \boldsymbol{y}^{\text{im}}, \boldsymbol{m})$ and $\boldsymbol{w} = \boldsymbol{x}$ remain the same as before. We propose two different methods to incorporate the metadata. The naive one, which we call M-DRUNet, consists of passing the metadata through input maps, similar to image metadata maps. For each metadata parameter $m_i$, a map of the same spatial dimension as the input amplitude $\tilde{\boldsymbol{y}}$ is created and filled with the corresponding single metadata value. The second approach, called M-SEDRUNet, incorporates metadata through SE blocks, inspired by the work of Plutenko et al. [30]. Metadata are then injected into the MLPs of the SE blocks, enabling it to adapt dynamically to context. Figure 2 shows the metadata injection principle for M-SEDRUNet.

## III. EXPERIMENTS

### *A. Data Preparation*

As previously mentioned, we used a dataset of SAR images simulated using the MOCEM software. The GT images of size $n = 256 \times 256$ are composed of background maps into which various CAD models are inserted. To ensure scene diversity, a total of 78 distinct CAD models—representing both military and civilian objects such as airplanes, boats, antennas, land vehicles, and buildings—are randomly placed with varying scales and orientations onto background images. The background images themselves depict a range of environments, including forests, roads, fields, and airports. From these composite scenes, MOCEM is used to simulate realistic SAR acquisitions. It is important to note that, in this case, MOCEM accurately models the electromagnetic scattering behavior only for the CAD objects; the background images do not reflect true radar backscattering properties and are included solely to enhance contextual realism. To ensure diversity in acquisition scenarios, the simulation process involves uniform sampling of scene and sensor parameters, including radar acquisition settings such as squint angle, spatial resolution, and noise level. This variability contributes to the generation of a rich and diverse dataset, better preparing the trained NNs for real-world generalization. The final generated dataset consists of 5,000 4-tuples $(\boldsymbol{y}, \boldsymbol{z}, \boldsymbol{x}, \boldsymbol{m}) \in \mathcal{Y} \times \mathcal{Z} \times \mathcal{X} \times \mathcal{M}$, which were split into 4,631 training samples and 369 validation samples. The split was performed by isolating specific acquisition parameter ranges, as shown in Table I, to ensure that the validation set contains conditions not present in the training set. This design explicitly tests the NNs ability to generalize to unseen radar acquisition configurations. Note that all images are standardized according to the statistics of the training split. The same standardization procedure is applied individually to each metadata variable.

| Metadata \ Dataset | Training set | Validation set |
|---|---|---|
| Bearing (°) | $[0, 150[\cup]155, 360]$ | $[150, 155]$ |
| Incidence (°) | $[15, 55[\cup]55.7, 75]$ | $[55, 55.7]$ |
| Squint (°) | $[0, 15[\cup]15.5, 45]$ | $[15, 15.5]$ |
| Resolution ($\mathrm{m}^2$) | $[0.2, 0.35[\cup]0.36, 0.6]$ | $[0.35, 0.36]$ |
| Noise level (db) | $[-40, -32[\cup]-31.7, -20]$ | $[-32, -31.7]$ |

TABLE I: Partitioning of the training and validation sets based on non-overlapping metadata intervals of simulated acquisitions.

Finally, we used real UMBRA[1] and CAPELLA[2] SLC images as a test set. These images have statistics properties similar to those seen in training, thereby reducing the domain gap.

### *B. Experimental Settings*

For performance evaluation, we compare our proposed methods specifically with SARCAM [31], a deep learning architecture designed for SAR image despeckling, as well as with AdaIR [32] and IRNeXt [33], originally developed for general image restoration. They are included in our comparison as they are well suited to the joint SAR despeckling and sidelobes reduction problem. In addition, we consider MERLIN [22] and MuLoG-DRUNet [24], well-established and effective approaches for real SAR data despeckling. Their inclusion allows us to benchmark our method against robust despeckling baselines, even though they do not explicitly target sidelobes reduction. Note that both models are evaluated only on real SAR images using publicly available pretrained weights.

All models are trained using the Adam optimizer with an initial learning rate of $1 \times 10^{-4}$, coupled with an exponential decay schedule. Training is conducted for 200 to 400 epochs, depending on when convergence is observed on the validation set. The batch size is adjusted according to available GPU memory. To improve generalization, data augmentation techniques such as random flips and rotations are applied. We use classical image restoration metrics such as peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM) to evaluate image quality. In addition, we include feature similarity index (FSIM) [34] and learned perceptual image patch similarity (LPIPS) [35], which capture perceptual fidelity and better align with human visual perception. To evaluate despeckling performance on real SAR images, we use the equivalent number of looks (ENL), which measures the level of speckle noise reduction in homogeneous SAR image regions. No quantitative metric for sidelobes reduction is used, as no established metric currently exists.

### *C. Results on Simulated SAR Images*

| Model \ Metric | PSNR ↑ | SSIM ↑ | FSIM ↑ | LPIPS ↓ |
|---|---|---|---|---|
| SARCAM | 29.915 | 0.897 | 0.877 | 0.347 |
| IRNeXt | 30.023 | 0.899 | 0.874 | 0.342 |
| AdaIR | 30.040 | 0.899 | 0.878 | **0.333** |
| DRUNet | 29.794 | 0.897 | 0.876 | 0.338 |
| SEDRUNet | 30.012 | 0.899 | 0.874 | 0.339 |
| DRUNet PL | 29.919 | 0.897 | 0.879 | 0.341 |
| DRUNet EPL | 29.836 | 0.896 | **0.880** | 0.340 |
| M-DRUNet | 30.272 | 0.899 | 0.878 | 0.339 |
| M-SEDRUNet | **30.461** | **0.901** | 0.874 | 0.335 |

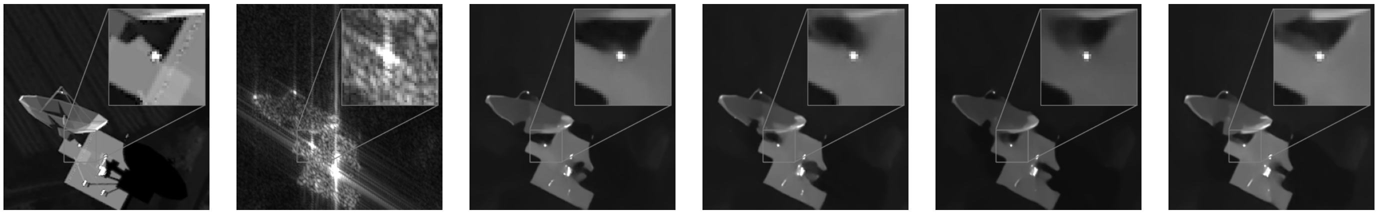
TABLE II: Quantitative evaluation on the simulated SAR validation set.

In terms of PSNR, the best-performing models are those incorporating metadata, which highlights the impact of metadata integration on the image restoration task. On the other hand, according to perceptual metrics such as SSIM, FSIM,

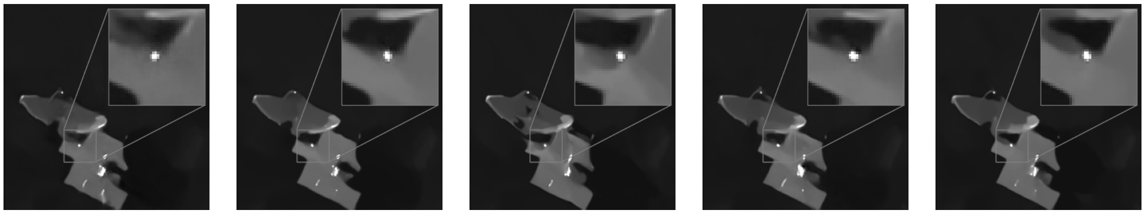
---

[1] https://umbra.space/open-data/

[2] https://www.capellaspace.com/earth-observation/gallery



(a) GT (b) SAR (c) DRUNet (d) DRUNet EP (e) DRUNet PL (f) SEDRUNet



(g) SARCAM (h) IRNeXt (i) AdaIR (j) M-DRUNet (k) M-SEDRUNet

Fig. 3: GT and and restored simulated SAR images.

and LPIPS, all models demonstrate comparable performance. The comparison between DRUNet and SEDRUNet and their metadata-driven counterparts demonstrates that metadata injection enhances model performance, even though the improvement margin remains modest. From a visual standpoint, the restored images appear similar across all models (see Figure 3). Speckle noise is completely suppressed, and sidelobes are clearly eliminated in all restored outputs.

### *D. Results on Real SAR Images*

Table III summarizes the quantitative evaluation of all models on real SAR imagery using the ENL metric. The despeckling-oriented methods MERLIN and MuLoG-DRUNet exhibit strong performance, with MuLoG-DRUNet achieving the best overall ENL, confirming their effectiveness on real data. The metadata-driven M-SEDRUNet attains the second-highest ENL, surpassing all DRUNet-based baselines (DRUNet, DRUNet EPL, DRUNet PL) and demonstrating the benefit of incorporating acquisition metadata for improved generalization to real imagery. IRNeXt and AdaIR also yield competitive results, indicating that general-purpose restoration models can adapt well to SAR despeckling. Overall, these results highlight that metadata-aware architectures can approach the performance of specialized despeckling methods while simultaneously addressing sidelobes reduction.

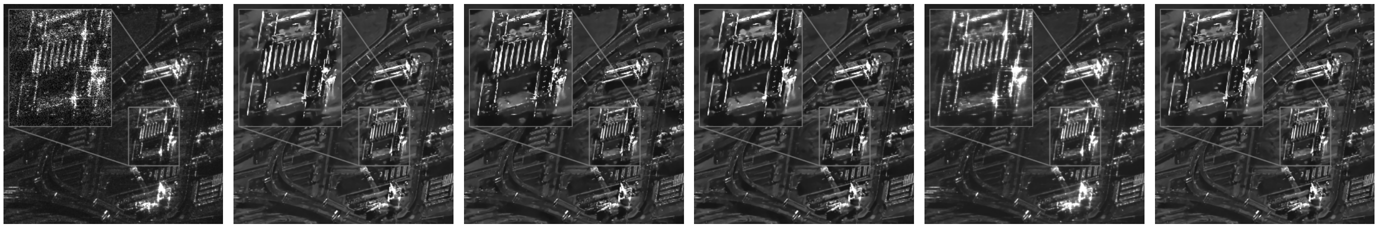
| Metric<br>Model | ENL ↑ |
|---|---|
| MERLIN | 257.6 |
| MuLoG DRUNet | **367.7** |
| SARCAM | 121.9 |
| IRNeXt | 242.1 |
| AdaIR | 173.1 |
| DRUNet | 83.35 |
| SEDRUNet | 175.7 |
| DRUNet PL | 73.71 |
| DRUNet EPL | 60.15 |
| M-DRUNet | 121.2 |
| M-SEDRUNet | 286.2 |

TABLE III: Quantitative evaluation on real SAR images.

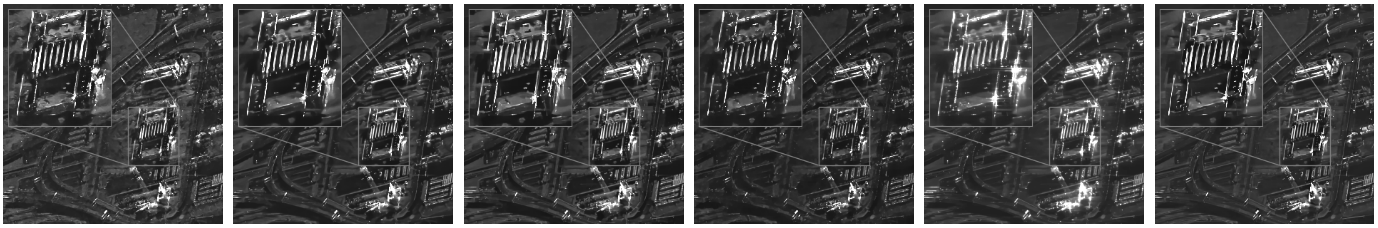
While ENL does not capture other critical aspects of SAR image restoration, such as the suppression of sidelobes or detail preservation in structured areas, relying solely on ENL may provide an incomplete assessment of a model's performance. Visual inspection therefore remains crucial for evaluating the overall quality of SAR image restoration.

Regarding the results shown in Figure 4, the despeckling-only methods MERLIN and MuLoG-DRUNet exhibit strong speckle suppression performance on real SAR images but fail to remove sidelobe artefacts, as they are not designed for joint restoration. In contrast, the proposed approaches, along with general-purpose models such as IRNeXt and AdaIR, demonstrate both despeckling and sidelobes reduction capabilities. While most methods effectively smooth homogeneous regions, IRNeXt tends to over-smooth certain areas, achieving strong sidelobes suppression at the cost of texture fidelity. DRUNet-based models generally preserve more structural details, though some loss is observed in regions with repetitive patterns, particularly for DRUNet PL. The metadata-enhanced variants maintain restoration quality comparable to their non-metadata counterparts, offering a balanced trade-off between noise suppression and detail preservation. Additional visual results are presented in Figure 5.

Even if the effects of metadata are not always directly observable in the restored images, an interesting outcome is that metadata-driven models can implicitly interpret the underlying physics of the acquisition parameters for the restoration task. For example, injecting a resolution parameter that differs from the one used by the sensor to acquire the original SAR image can influence the restored image's effective resolution. Figure 6 clearly illustrates that providing a higher resolution parameter leads to visibly enhanced image detail. This capability, while promising, has some drawbacks. In particular, it can introduce new sidelobes or artefacts when the injected metadata significantly deviate from the original acquisition parameters. Nonetheless, this adaptive functionality opens new possibilities for dynamic image analysis and restoration tuning. Similar behaviors are observed when varying the incidence angle.

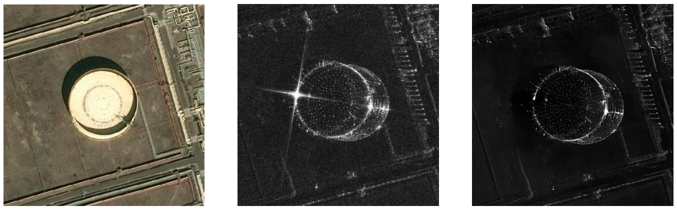


(a) SAR (b) DRUNet (c) DRUNet EPL (d) DRUNet PL (e) MERLIN (f) SEDRUNet



(g) SARCAM (h) IRNeXt (i) AdaIR (j) M-DRUNet (k) MuLoG DRUNet (l) M-SEDRUNet

Fig. 4: Restored images from real UMBRA SLC.



(a) (b) (c)

Fig. 5: Optical GT (a) alongside CAPELLA SAR image (b) and restored image with SEDRUNet (c).

(a) (b) (c)

Fig. 6: UMBRA SAR image (a) and restorations with M-SEDRUNet using the real sensor resolution set to 0.40m$^2$ (b) and a different one set to 0.25m$^2$ (c).

## IV. DISCUSSION

This study demonstrates the potential of the Sim2Real approach, as well as the integration of SAR acquisition parameters into the NNs to enhance performance. However, several limitations must still be addressed.

The injection of metadata into DRUNet and SEDRUNet has shown improved performance, particularly in simulated environments where GT data are accessible. Comparisons using identical architectural baselines confirm the benefits of this approach. However, we have not yet explored metadata injection in other advanced models such as AdaIR and IRNeXt. Integrating metadata into these architectures could potentially further enhance their performance and remains an avenue for future investigation.

While the current results confirm that metadata can guide the restoration process by modulating the network's behavior according to acquisition conditions, the quantitative characterization of this control remains limited. Specifically, the trade-off between the degree of metadata influence and the risk of artefact introduction has not yet been systematically evaluated. Future work should therefore investigate how varying the weighting or conditioning strength of metadata impacts quantitative metrics and perceptual fidelity.

One major challenge is the domain gap between the simulated images used during training and real SAR images. For certain SAR image providers, such as TerraSAR-X[3], the statistical properties of real images deviate from those seen in training, leading to poor image restoration results. Currently, no effective solution has been identified to reconcile these statistical discrepancies, highlighting the limitations of supervised training strategies in this context. This motivates the exploration of unsupervised approaches. Recently, deep posterior sampling (DPS) approaches [36] have gained attention for solving inverse problems, where prior information is learned using diffusion models. Methods such as BlindDPS [37] aim to jointly estimate the forward operator and the latent clean image $\boldsymbol{x}$. Applying such techniques in the SAR domain may enable the joint estimation of both the SAR transfer function and the underlying image, offering a promising direction for future work.

## V. CONCLUSION

In this work, we presented a unified deep learning framework for simultaneous despeckling and sidelobes reduction in SAR images. Leveraging the realistic SAR simulated dataset

---

[3] https://earth.esa.int/eogateway/missions/terrasar-x-and-tandem-x/sample-data

generated by MOCEM, which includes access to GT, we enabled supervised training that effectively addresses both inherent noise and artefacts in SAR imagery. Our results demonstrate that incorporating acquisition metadata as auxiliary input improves restoration performance and enables the guided control of image restoration.

## ACKNOWLEDGMENTS

This study has been carried out with financial support from the French Direction Générale de l'Armement and the French Agence Ministérielle pour l'Intelligence Artificielle de Défense.

## REFERENCES

[1] A. Nuttall. Some windows with very good sidelobe behavior. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 29(1):84–91, 1981.

[2] H.C. Stankwitz, R.J. Dallaire, and J.R. Fienup. Nonlinear apodization for sidelobe control in sar imagery. *IEEE Transactions on Aerospace and Electronic Systems*, 31(1):267–279, 1995.

[3] Rémy Abergel, Loïc Denis, Saïd Ladjal, and Florence Tupin. Subpixellic methods for sidelobes suppression and strong targets extraction in single look complex sar images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 11(3):759–776, 2018.

[4] Sen Yuan, Ze Yu, Chunsheng Li, and Shusen Wang. A novel sar sidelobe suppression method based on cnn. *IEEE Geoscience and Remote Sensing Letters*, 18(1):132–136, 2021.

[5] Shuyi Liu, Yan Jia, Yongqing Liu, Limin Zhai, and Xiangkun Zhang. A new birnn-sva method for side lobe suppression. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 17:1167–1175, 2024.

[6] J. W. Goodman. *Laser Speckle and Related Phenomena.*, chapter Statistical properties of laser speckle patterns. Springer Berlin., 1984.

[7] José M. Bioucas-Dias and Mário A. T. Figueiredo. Multiplicative noise removal using variable splitting and constrained optimization. *IEEE Transactions on Image Processing*, 19(7):1720–1730, 2010.

[8] Charles-Alban Deledalle, Loic Denis, Sonia Tabti, and Florence Tupin. Mulog, or how to apply gaussian denoisers to multi-channel sar speckle reduction? *IEEE Transactions on Image Processing*, 26(9):4389–4403, September 2017.

[9] G. Chierchia, D. Cozzolino, G. Poggi, and L. Verdoliva. Sar image despeckling through convolutional neural networks. In *2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pages 5438–5441, 2017.

[10] 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.

[11] Kai Zhang, Wangmeng Zuo, Yunjin Chen, Deyu Meng, and Lei Zhang. Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. *IEEE Transactions on Image Processing*, 26(7):3142–3155, July 2017.

[12] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In Nassir Navab, Joachim Hornegger, William M. Wells, and Alejandro F. Frangi, editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pages 234–241, Cham, 2015. Springer International Publishing.

[13] Javier Gurrola-Ramos, Oscar Dalmau, and Teresa E. Alarcón. A residual dense u-net neural network for image denoising. *IEEE Access*, 9:31742–31754, 2021.

[14] Fan Jia, Wing Hong Wong, and Tieyong Zeng. Ddunet: Dense dense u-net with applications in image denoising. In *2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, pages 354–364, 2021.

[15] Chi-Mao Fan, Tsung-Jung Liu, and Kuan-Hsien Liu. Sunet: Swin transformer unet for image denoising. In *2022 IEEE International Symposium on Circuits and Systems (ISCAS)*. IEEE, May 2022.

[16] Jindong Jiang, Lunan Zheng, Fei Luo, and Zhijun Zhang. Rednet: Residual encoder-decoder network for indoor rgb-d semantic segmentation, 2018.

[17] 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.

[18] Puyang Wang, He Zhang, and Vishal M. Patel. Generative adversarial network-based restoration of speckled sar images. In *2017 IEEE 7th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, pages 1–5, 2017.

[19] Ruijiao Liu, Yangyang Li, and Licheng Jiao. Sar image specle reduction based on a generative adversarial network. In *2020 International Joint Conference on Neural Networks (IJCNN)*, pages 1–6, 2020.

[20] 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.

[21] Xuran Hu, Ziqiang Xu, Zhihan Chen, Zhengpeng Feng, Mingzhe Zhu, and LJubisa Stankovic. Sar despeckling via regional denoising diffusion probabilistic model, 2024.

[22] 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*, 60:1–13, 2022.

[23] Liang Chen, Yifei Yin, Hao Shi, Qingqing Sheng, and Wei Li. A self-supervised sar image despeckling strategy based on parameter-sharing convolutional neural networks. 2023.

[24] Cristiano Ulondu Mendes, Loïc Denis, Charles Deledalle,

and Florence Tupin. Robustness to spatially-correlated speckle in plug-and-play polsar despeckling. *IEEE Transactions on Geoscience and Remote Sensing*, 2024.

[25] Christian COCHIN, Philippe POULIGUEN, Benoit DELAHAYE, Daniel le HELLARD, Philippe GOSSELIN, and Franck AUBINEAU. Mocem - an 'all in one' tool to simulate sar image. In *7th European Conference on Synthetic Aperture Radar*, pages 1–4, 2008.

[26] Kai Zhang, Yawei Li, Wangmeng Zuo, Lei Zhang, Luc Van Gool, and Radu Timofte. Plug-and-play image restoration with deep denoiser prior, 2021.

[27] Jie Hu, Li Shen, Samuel Albanie, Gang Sun, and Enhua Wu. Squeeze-and-excitation networks, 2019.

[28] Justin Johnson, Alexandre Alahi, and Li Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In *European conference on computer vision*, pages 694–711. Springer, 2016.

[29] Ram Krishna Pandey, Nabagata Saha, Samarjit Karmakar, and AG Ramakrishnan. Msce: An edge-preserving robust loss function for improving super-resolution algorithms. In *International Conference on Neural Information Processing*, pages 566–575. Springer, 2018.

[30] Iaroslav Plutenko, Mikhail Papkov, Kaupo Palo, Leopold Parts, and Dmytro Fishman. Metadata improves segmentation through multitasking elicitation. In *MICCAI Workshop on Domain Adaptation and Representation Transfer*, pages 147–155. Springer, 2023.

[31] 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, 2021.

[32] Yuning Cui, Syed Waqas Zamir, Salman Khan, Alois Knoll, Mubarak Shah, and Fahad Shahbaz Khan. Adair: Adaptive all-in-one image restoration via frequency mining and modulation. *arXiv preprint arXiv:2403.14614*, 2024.

[33] Yuning Cui, Wenqi Ren, Sining Yang, Xiaochun Cao, and Alois Knoll. Irnext: Rethinking convolutional network design for image restoration. 2023.

[34] Lin Zhang, Lei Zhang, Xuanqin Mou, and David Zhang. Fsim: A feature similarity index for image quality assessment. *IEEE transactions on Image Processing*, 20(8):2378–2386, 2011.

[35] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 586–595, 2018.

[36] Hyungjin Chung, Jeongsol Kim, Michael T Mccann, Marc L Klasky, and Jong Chul Ye. Diffusion posterior sampling for general noisy inverse problems. *arXiv preprint arXiv:2209.14687*, 2022.

[37] Hyungjin Chung, Jeongsol Kim, Sehui Kim, and Jong Chul Ye. Parallel diffusion models of operator and image for blind inverse problems. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6059–6069, 2023.