Speckle Noise Removal in SAR Images using Autoencoders

Bianca Balzarini[1] and Juliana Gambini[2,3,1]

[1] Instituto Tecnológico de Buenos Aires, Buenos Aires, Argentina
bbalzarini@itba.edu.ar
[2] Universidad Nacional de Hurlingham, LIDEC
juliana.gambini@unahur.edu.ar
[3] Universidad Tecnológica Nacional, Regional Buenos Aires, CPSI

Abstract. Synthetic Aperture Radar (SAR) images are fundamental for Earth observation due to their ability to acquire data independently of weather and lighting conditions. However, these images are affected by speckle noise, a multiplicative granular pattern that degrades their quality and makes them difficult to automatically interpret. This work presents an approach for speckle removal using convolutional autoencoders. The main innovation lies in the use of the $\mathcal{G}_I^0$ distribution to synthetically generate pairs of noisy and clean images for supervised training, adequately modeling the statistical characteristics of actual speckle noise. In addition, we developed a flexible file-based configuration system that facilitates experimentation with different network architectures. Experimental outcomes reveal that deeper network architectures achieve better results with no evidence of overfitting. Models trained on synthetic data showed good performance on synthetic test images, but limitations in generalizing to actual SAR images. To address this problem, we explored training with pairs of actual images obtained through time averaging, achieving significant improvements and performance comparable to other known models.

Keywords: aperture synthetic radar, speckle noise, autoencoders.

Eliminación de Ruido *Speckle* en Imágenes SAR utilizando Autoencoders

Abstract. Las imágenes de Radar de Apertura Sintética (SAR: Synthetic Aperture Radar) son fundamentales para la observación de la Tierra debido a su capacidad de adquisición independiente de las condiciones climáticas y de iluminación. Sin embargo, estas imágenes se ven afectadas por el ruido *speckle*, un patrón granular multiplicativo que degrada su calidad, y dificulta su interpretación. Este trabajo presenta

un enfoque para la eliminación de ruido *speckle* mediante el uso de autoencoders convolucionales. La principal innovación radica en el uso de la distribución $\mathcal{G}_I^0$ para generar sintéticamente pares de imágenes ruidosas y limpias destinadas al entrenamiento supervisado, modelando adecuadamente las características estadísticas del ruido *speckle* real. Complementariamente, desarrollamos un sistema flexible basado en archivos de configuración que facilita la experimentación con diferentes arquitecturas de redes. Los experimentos revelan que las arquitecturas más profundas obtienen consistentemente mejores resultados sin evidencia de overfitting. Los modelos entrenados con datos sintéticos mostraron buen desempeño en imágenes de prueba sintéticas, pero limitaciones en la generalización a imágenes SAR reales. Para abordar este problema, exploramos el entrenamiento con pares de imágenes reales obtenidas mediante promediado temporal, logrando mejoras significativas y desempeño comparable a otros ya conocidos.

Keywords: radar de apertura sintética (SAR), ruido speckle, autoencoders.

1 Introduction

Synthetic Aperture Radar (SAR) is a highly effective remote sensing technology capable of generating high-resolution images of the Earth's surface regardless of weather or lighting conditions. This makes SAR imagery an invaluable tool for applications ranging from environmental monitoring to terrain change detection and disaster management. However, a significant drawback of SAR images is the inherent presence of speckle noise. Unlike the additive noise common in optical imagery, speckle is a multiplicative, non-Gaussian phenomenon resulting from the coherent interference of radar waves scattered by surface elements. This granular noise severely degrades the visual quality of SAR images and complicates subsequent analysis and interpretation tasks.

Classical speckle mitigation techniques include adaptive filters (Lee, Frost) (Lee, 1980; Frost et al., 1982), wavelets (L. Zhang et al., 2017), and non-local methods like SAR-BM3D (Parrilli et al., 2012). These struggle to balance noise reduction with detail preservation. Deep learning, particularly CNNs and Autoencoders, has revolutionized image denoising (Q. Zhang et al., 2018; Lattari et al., 2019; Vitale et al., 2021; Cardona-Mesa et al., 2025) using supervised learning with large datasets of paired noisy-clean images. Since perfectly noise-free SAR images are hard to aquire, synthetic data generation is essential, either by injecting simulated noise or averaging temporal images (Vásquez-Salazar et al., 2024), though the latter is scarce.

The efficacy of supervised deep learning depends heavily on simulated noise realism. Existing literature often uses statistical models as Gamma or Rayleigh, which are computationally convenient but limited. The Rayleigh distribution,

for instance, is accurate only for homogeneous areas like pasturelands (Lattari et al., 2019), failing to capture complex statistical properties in heterogeneous terrain (urban, vegetation). This lmits generalization capability of the trained networks.

To address this gap, this paper proposes the use of $\mathcal{G}_I^0$ distribution for speckle simulation. The $\mathcal{G}_I^0$ distribution is renowned for accurately modeling speckle across diverse surface roughness and heterogeneous clutter (Frery et al., 1997). Despite proven accuracy in SAR statistical modeling, its integration into synthetic deep-learning training pipelines remains underexplored. This work investigates how $\mathcal{G}_I^0$-based data preparation enhances autoencoder robustness and generalization, validated on both synthetic and actual SAR images using standard metrics.

Our main contributions are: (i) integrating the $\mathcal{G}_I^0$ model into a synthetic data pipeline for autoencoder training; (ii) a systematic evaluation of 25 autoencoder configurations driven by flexible config files; and (iii) empirical evidence that fine-tuning on actual SAR pairs substantially improves generalization.

This paper is organized as follows: Section 2 describes the methodology (data generation and architecture); Section 3 presents experiments and comparisons; Section 4 provides the conclusions and outlines future work.

2 Materials and Methods

2.1 SAR Images

For this study, two types of datasets were used: synthetically generated images and actual SAR images. The synthetic dataset was created based on the $\mathcal{G}_I^0$ distribution, which models the statistical properties of speckle noise. A random variable z following the $\mathcal{G}_I^0$ distribution can be expressed as the product of two independent random variables:

$$z = x \cdot y \tag{1}$$

where x represents the noise-free backscatter, modeled by an inverse Gamma distribution ($x \sim \Gamma^{-1}(-\alpha, 1/\gamma)$), and y represents the speckle noise, modeled by a Gamma distribution ($y \sim \Gamma(L, 1/L)$). The parameter α is related to the texture or roughness of the surface, while L is the number of looks of the SAR system. For instance, α values close to zero (e.g., -1 to -3) correspond to heterogeneous urban areas, while more negative values (e.g., -10 to -20) represent homogeneous surfaces like calm water. This model allowed us to generate pairs of clean (x) and noisy (z) images for training the proposed models. To create a robust training dataset, we generated synthetic images structured in quadrants, where each quadrant simulated a different surface texture by assigning a randomly chosen value for the α parameter, as shown in Figure 1.

A key SAR concept is the *Equivalent Number of Looks* (ENL), which refers to the number of independent images of the same scene that are averaged to reduce speckle noise. It quantifies smoothness in homogeneous regions: higher ENL indicates better noise reduction.

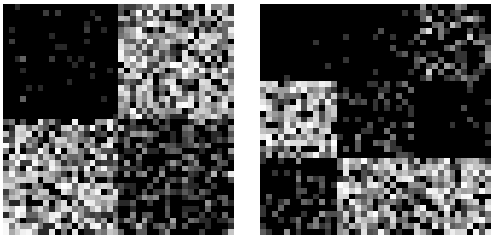
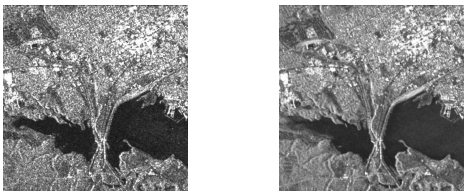


Fig. 1: Examples of artifically generated SAR images split into quadrants, with different values of α in each one. All with $L = \gamma = 1$. Four and Nine quadrants, left and right, respectively

The actual SAR dataset used in this work are provided by Vásquez-Salazar et al., 2024, and contains paired noisy-clean images via temporal averaging, with diverse textures (urban, vegetation, water) suitable for generalization assessment.



(a) Noisy SAR image.

(b) Clean SAR image.

Fig. 2: Example of an actual SAR image pair (noisy and clean) from the dataset.

2.2 Methods

Autoencoder Architecture Our despeckling method uses a convolutional autoencoder: a neural network designed to learn a compressed representation (encoding) of its input and then reconstruct it (decoding). The network is composed of an encoder, which maps the input image to a lower-dimensional latent space, and a decoder, which reconstructs the image from the latent representation. The architecture is symmetric, with the network trained to reconstruct clean images from noisy inputs by minimizing MSE.

Our implementation uses a convolutional architecture. The encoder consists of a series of convolutional layers that apply filters to the input image, extracting hierarchical features and reducing spatial dimensions. The decoder uses transposed convolutional layers (also known as deconvolutions) to upsample the feature maps and reconstruct the image. The general structure of the autoencoder

is symmetrical. Figure 3 shows a scheme of a particular architecture. A key aspect of our training process is that the network takes a noisy image as input and is trained to reconstruct the corresponding clean (noise-free) version, which is provided as the target output.

The statistical figures of merit based on the ratio image and the GLCM are used as complementary evaluation criteria rather than direct loss terms, since they are not differentiable and cannot be straightforwardly optimized by gradient descent.

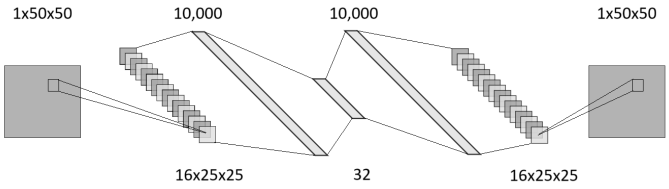


Fig. 3: Scheme of a convolutional autoencoder. The input image is progressively downsampled by the encoder to a latent representation and then upsampled by the decoder to the original dimensions.

Tested Architectures We developed a config-file driven framework to test 25 autoencoder configurations without code modification. It also had the ability to automatically validate the dimensional consistency of the network layers. Key hyperparameters are shown in Table 1: *Encoding Dim* (smallest dimension), *Samples* (training images), *LR* (learning rate), *ELR* (exp. decay, $\gamma = 0.95$), *RLROP* (Reduce LR on Plateau), *Mixed Data* (whether varied partitions were used). All models use Adam, MSE loss, batch size 64, ReLU for hidden layers, and Sigmoid for the output layer. We explored the impact of depth, latent dimensionality, and training duration.

Evaluation Metrics To quantitatively evaluate the despeckling filter, we used statistical metrics based on the multiplicative speckle noise model, where the noisy image z is the product of the clean image x and speckle noise y ($z = x \cdot y$). Under this model, the ratio image $I = z/\hat{x}$, where $\hat{x}$ is the filtered image, should contain only speckle noise and no residual structural information. Accordingly, we used two metrics proposed by Gomez et al., 2017.

The first metric, based on first-order statistics ($r_{\mathrm{ENL},\mu}$), assesses mean preservation and smoothness in homogeneous regions. It assumes that the ratio image mean μ_I should be close to 1 and that filtering should increase smoothness, quantified by the Equivalent Number of Looks ($\mathrm{ENL} = (\mu/\sigma)^2$), where μ and σ are the mean and standard deviation of a homogeneous region. The metric combines

Table 1: Summary of all tested convolutional autoencoder configurations.

Config.	Conv Layers	Encoding Dim	Epochs	Samples	LR	Scheduler	Mixed Data
Base and Reference Architectures							
A1	1+MaxPool	32	5	50K	0.001	None	No
A2	1	32	30	50K	0.001	ELR	No
A3	2+MaxPool	512	500	50K	0.001	ELR	No
Depth Variations							
1 Convolutional Layer							
A4	1	32	500	50K	0.001	None	No
A5	1	32	70	50K	0.001	None	No
A6	1	32	500	50K	0.0005	None	No
2 Convolutional Layers							
A7	2	128	500	50K	0.001	ELR	No
A8	2	128	70	50K	0.001	ELR	No
A9	2	256	500	50K	0.001	ELR	No
A10	2	256	70	50K	0.001	ELR	No
A11	2	64	100	100K	0.001	RLROP	No
A12	2	64	40	100K	0.001	RLROP	No
3+ Convolutional Layers							
A13	3	128	500	50K	0.001	ELR	No
A14	3	128	70	50K	0.001	ELR	No
A15	3	256	500	50K	0.001	ELR	No
A16	3	256	70	50K	0.001	ELR	No
A17	3+Dense	128	500	50K	0.001	None	No
A18	3+Dense	128	70	50K	0.001	None	No
A19	3	64	100	100K	0.001	RLROP	No
A20	4+Dense	64	100	100K	0.001	RLROP	Yes
Special Training Runs							
A21	1	32	15	50K	0.001	ELR	No
A22	1	32	500	50K	0.001	ELR	No
A23	1	32	500	100K	0.001	ELR	Yes
Experimental Architectures							
A24	Conv2DT*	32	20	50K	0.001	None	No
A25	Conv2DT*	32	11	50K	0.001	None	No

these criteria over N selected homogeneous regions:

$$r_{\mathrm{ENL},\mu} = \frac{1}{2N} \sum_{i=1}^{N} \left(\frac{|\mathrm{ENL}_z(i) - \mathrm{ENL}_I(i)|}{\mathrm{ENL}_z(i)} + |1 - \mu_I(i)| \right). \tag{2}$$

A lower $r_{\mathrm{ENL},\mu}$ indicates better speckle suppression with brightness preservation.

The second metric, based on second-order statistics (δ_h), measures residual structural information in the ratio image through the invariance of its texture to random pixel permutations. Since an ideal ratio image contains only speckle noise, its texture should remain unchanged after shuffling. Using the Gray-Level Co-occurrence Matrix (GLCM) homogeneity descriptor (h), δ_h is defined as the percentage difference between the homogeneity of the original ratio image (h_0) and the mean homogeneity of its randomly permuted versions ($\bar{h}_m$):

$$\delta_h = 100 \cdot \left| 1 - \frac{\bar{h}_m}{h_0} \right|. \tag{3}$$

A value near zero indicates successful filtering, with negligible geometric information remaining in the ratio image.

Finally, we define the overall metric as $\mathcal{M} = r_{\mathrm{ENL},\mu} + \delta_h$, where lower values indicate better despeckling.

3 Results

The extensive experimentation, summarized in Table 1, allowed for a deep analysis of the different architectural configurations. In this section, we present the quantitative, qualitative, and comparative results of the most relevant models.

3.1 Performance on Synthetic Data

The models were initially evaluated based on their Mean Squared Error (MSE) on a synthetic test set of 1000 images, kept separate from the training data. Table 2 summarizes the performance of all the architectures, ranked by the statistical figure of merit $\mathcal{M}$, along with MSE values and other statistical metrics ($r_{\mathrm{ENL},\mu}$, δ_h, and $\mathcal{M}$) for reference.

The results indicate that deeper architectures with three convolutional layers (A13, A15, A16) achieved the best performance regarding MSE, with values around 0.0023. This suggests that a hierarchical feature extraction is crucial for effectively modeling and removing speckle noise. Notably, the dimensionality of the latent space in these top models (128 or 256) also played a significant role.

Furthermore, we observed that training duration did not linearly correlate with better performance. For instance, model A16, trained for only 70 epochs, achieved a performance nearly identical to A15 (trained for 500 epochs), suggesting that well-designed architectures can converge rapidly. This is a key finding, as it points towards more efficient training strategies. Several models, particularly shallower ones with long training times (e.g., A4), exhibited signs of overfitting, with a test MSE more than 25% higher than their final training MSE.

In short, although classical bias–variance intuition links more parameters to higher variance, depth changes the effective parametrization and the network's inductive bias. Deep architectures tend to learn hierarchical, compositional representations that favor simpler functions and, together with implicit regularization from the training dynamics (e.g., SGD), initialization and normalization, often generalize better than shallower networks with a similar raw parameter count. In our experiments, shallower architectures required denser parameterizations to cover the same spatial context, increasing memorization and overfitting C. Zhang et al., 2017.

While MSE primarily measures reconstruction accuracy, the statistical metrics offer insights into the filter's ability to preserve the underlying statistical properties of the SAR image. For instance, configurations A24, A2, and A22 achieved the best statistical performance (lowest $\mathcal{M}$ values of 2.576, 3.206, and 3.414 respectively), even though they did not rank highest in terms of MSE. This highlights that a low reconstruction error does not always guarantee optimal

statistical preservation, and vice-versa. Conversely, models exhibiting significant overfitting in MSE (e.g., A6, A4) also generally showed suboptimal statistical performance, reinforcing the correlation between generalization capacity and filtering quality.

Table 2: Performance metrics for all autoencoder configurations, ordered by the statistical figure of merit $\mathcal{M}$. Test and final train MSE values are included for reference, along with first and second order statistical metrics ($r_{\mathrm{ENL},\mu}$, δ_h, and $\mathcal{M}$). The 'Test vs. Train' column indicates overfitting (red) or underfitting (green).

Rank	Config.	Test MSE	Train MSE	Test vs. Train	$r_{\mathrm{ENL},\mu}$	δ_h	$\mathcal{M}$
1	A24	0.002505	0.002520	↘ 0.60%	0.233	2.343	2.576
2	A2	0.002594	0.002314	↗ 12.10%	0.218	2.988	3.206
3	A22	0.002799	0.002245	↗ 24.68%	0.232	3.182	3.414
4	A25	0.002586	0.002566	↗ 0.78%	0.241	3.273	3.515
5	A16	0.002348	0.002387	↘ 1.63%	0.201	3.329	3.530
6	A10	0.002516	0.002313	↗ 8.78%	0.207	3.340	3.548
7	A14	0.002421	0.002380	↗ 1.72%	0.204	3.402	3.606
8	A11	0.002643	0.002197	↗ 20.30%	0.238	3.453	3.691
9	A15	0.002319	0.002318	↗ 0.04%	0.194	3.574	3.768
10	A8	0.002568	0.002383	↗ 7.76%	0.222	3.629	3.851
11	A13	0.002346	0.002412	↘ 2.74%	0.213	3.687	3.900
12	A18	0.002485	0.002174	↗ 14.31%	0.230	3.873	4.103
13	A4	0.002781	0.002164	↗ 28.51%	0.273	3.838	4.111
14	A9	0.002547	0.002283	↗ 11.56%	0.217	3.963	4.179
15	A17	0.002528	0.002108	↗ 19.92%	0.313	3.979	4.293
16	A7	0.002433	0.002318	↗ 4.96%	0.208	4.291	4.499
17	A12	0.002575	0.002216	↗ 16.20%	0.236	4.232	4.468
18	A23	0.011212	0.011238	↘ 0.23%	0.305	4.688	4.993
19	A20	0.003691	0.003071	↗ 20.19%	0.343	4.946	5.289
20	A6	0.003444	0.002666	↗ 29.18%	0.643	6.863	7.506
21	A5	0.002879	0.002774	↗ 3.79%	0.344	7.443	7.787
22	A21	0.004343	0.004349	↘ 0.14%	0.376	7.627	8.003
23	A19	0.005331	0.005353	↘ 0.41%	-	-	-
24	A1	0.005695	0.005349	↗ 6.47%	-	-	-

Note: Configuration A3 does not present data because we could not carry out the training as the trainable parameters (weights and biases) were too many and the available computing power was not sufficient. For configurations A1 and A19, statistical metrics could not be calculated due to convergence issues during training.

3.2 Visual Assessment and Architectural Insights

Quantitative metrics are essential, but a visual assessment is crucial to understand the filter's behavior. Figure 4 shows the result of applying the best-performing model, A15, to a sample synthetic test image.

The autoencoder effectively reduces speckle and reconstructs the mean gray levels of homogeneous regions. However, the filtered images exhibit smooth, spatially correlated artifacts, reflecting the difficulty of convolutional networks in

distinguishing removable speckle noise from the intrinsic random backscatter texture that should be preserved. This limitation was consistently observed across all high-performing architectures.

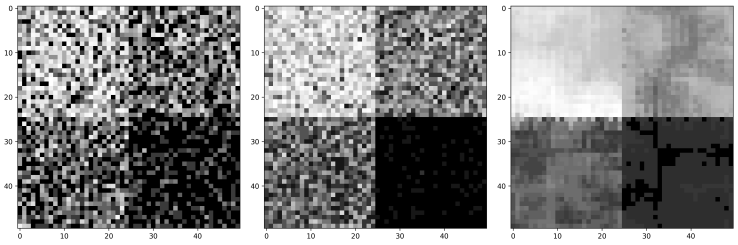


Fig. 4: Visual result of the A15 filter on a synthetic test image. From left to right: noisy input, clean target, and filtered output. The model effectively removes noise but introduces artificial smoothness.

3.3 Performance on Actual SAR Images

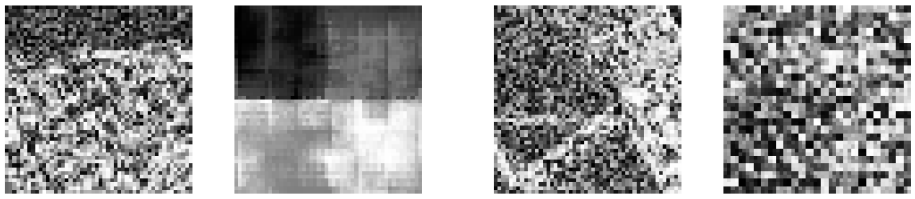
To assess the generalization capabilities of models trained on synthetic data, they were applied to actual SAR images. The results, shown in Figure 5, reveal a critical difference in performance based on the training data strategy.

Although Model A15 achieved the lowest test MSE on synthetic data, it exposed an important limitation when applied to real images (Figure 5a). The model effectively reduced speckle while preserving the overall gray-level distribution, maintaining the darker appearance of the upper quadrants. However, it also reproduced the square partition structure present in the synthetic training data, indicating that it learned the artificial spatial patterns instead of only the underlying noise characteristics. This result shows that, even with a realistic statistical model such as $\mathcal{G}_I^0$, insufficient structural diversity in the training set limits the model's ability to generalize to real SAR images.

In contrast, model A20 was trained on a mixed dataset of synthetic images with different partition granularities. As shown in Figure 5b, this strategy yielded significantly better generalization. Although its MSE on synthetic data was higher, the A20 model does not impose a rigid quadrant structure and preserves the natural texture of the actual SAR image more faithfully. This highlights a key insight: a lower error on a narrow synthetic test set does not guarantee superior performance on real-world data, and training data variety is crucial for robust generalization.

3.4 Performance of Models Trained on Actual SAR Data

The limitations of synthetic data motivated a second phase of experiments using paired real SAR images for training. To reduce computational cost and increase



(a) Model A15 (trained on uniform partitions) fails to generalize.

(b) Model A20 (trained on mixed partitions) generalizes better.

Fig. 5: Comparison of generalization on an actual SAR image. In both (a) and (b), the original actual image is on the left and the filtered output is on the right. Model A15 superimposes an artificial structure, while Model A20 preserves the image's natural texture more faithfully.

the number of training samples, the original 512×512 images were divided into smaller, non-overlapping patches, with input sizes listed in Table 3. Completely flat patches, where the maximum and minimum pixel intensities were equal, were discarded to avoid training on uninformative regions. Five deeper, fully convolutional architectures (B1-B5) were then developed, with their architectures and performance summarized in Table 3.

Table 3: Architectures and performance of models trained with actual SAR images.

Config.	**Architecture Type**	**Encoding Dim.**	**Image Size**	**Test MSE**	$r_{\mathbf{ENL},\mu}$	δ_h	$\mathcal{M}$
B1	Hybrid (Conv + Dense)	1600	512×512	0.03217	0.159	49.346	49.505
B2	Fully Convolutional (5 layers)	1500	256×256	0.00905	0.487	30.591	31.078
B3	Fully Convolutional (6 layers)	1500	256×256	0.00907	0.298	30.721	31.019
B4	Fully Convolutional (4 layers)	1500	128×128	0.01014	0.778	27.388	28.166
B5	Hybrid (Conv + Pooling)	1500	256×256	**0.00887**	**0.540**	**29.319**	**29.859**

Model B5 emerged as the top-performing architecture, achieving the lowest test MSE and a strong statistical figure of merit ($\mathcal{M}$). The visual results, shown in Figure 6, confirm a substantial improvement over models trained on synthetic data. The filtered image is significantly cleaner, with well-defined edges and textures that are much more consistent with the clean target image. This demonstrates the clear superiority of training directly on actual data for this task. Furthermore, visual inspection of the ratio images ($I = z/\hat{x}$) for the B5 model corroborated the quantitative statistical metrics. These ratio images lacked evident geometric structures or fine textures, behaving essentially as pure random noise (as illustrated in Figure 7). This suggests that the B5 autoencoder largely isolated the speckle component while preserving most underlying physical struc-

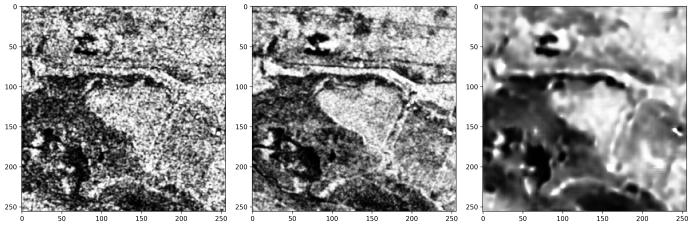


Fig. 6: Result of the best-performing model (B5), trained on actual SAR data. Left: noisy input. Center: clean target. Right: filtered output. The model achieves excellent noise reduction while preserving structural detail.

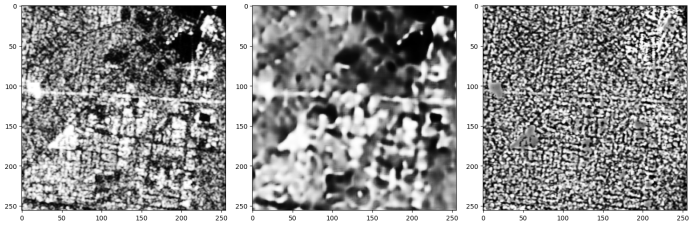


Fig. 7: Visual inspection of the ratio image for the B5 model. Left: Original noisy SAR image. Center: Filtered output. Right: Ratio image ($I = z/\hat{x}$). The lack of prominent geometric structure in the ratio image indicates successful speckle removal.

tures of the scene, although some loss of fine detail is still visible in the filtered output.

3.5 Statistical Evaluation and Comparative Analysis

A final comparative analysis was performed to benchmark our models against established methods. The Lee filter, a classic adaptive filter that uses local image statistics to estimate the clean signal, was used as a baseline. As shown in Table 4, the Lee filter achieved a figure of merit $\mathcal{M} = 5.05$ on our synthetic dataset with its optimal window size (3x3). In contrast, our best autoencoder configuration (A24) achieved $\mathcal{M} = 2.58$, demonstrating a clear quantitative improvement. It is particularly noteworthy that the A24 model features an atypical architecture: it employs a transposed convolution within the encoder, deviating from traditional symmetric compression principles. The fact that this unconventional design achieved the best statistical preservation of speckle properties opens an interesting avenue for future research beyond standard symmetrical autoencoders. This holds true even for the Lee filter's optimal window size (3x3), confirming the viability and superior performance of the deep learning approach.

Furthermore, we compare our statistical results with other state-of-the-art works. In Chan et al., 2022, methods based on Shannon Entropy and Renyi

Table 4: Statistical evaluation for a selection of models trained on synthetic data, compared to the Lee filter. Lower values are better.

Filter	$r_{\mathbf{ENL},\mu}$	δ_h	$\mathcal{M}$
Autoencoder A24 (Best $\mathcal{M}$)	**0.233**	**2.343**	**2.576**
Autoencoder A2	0.218	2.988	3.206
Autoencoder A15 (Best MSE)	0.194	3.574	3.768
Autoencoder A21 (Worst $\mathcal{M}$)	0.376	7.627	8.003
Lee Filter (3x3)	0.399	4.654	5.053
Lee Filter (7x7)	0.509	10.837	11.346

Entropy yield δ_h values of 0.113 and 0.173. Our best value (2.34 for A24) is larger, which reflects differences in dataset characteristics rather than a failure of the method.

The δ_h metric is based on GLCM homogeneity and therefore depends on the spatial patterns in the image. Changes in SAR resolution, terrain heterogeneity, and image intensity distribution alter the GLCM statistics and shift the absolute homogeneity values. Thus, absolute δ_h numbers are baseline-dependent: they are tied to the dataset properties and are not directly comparable across different image collections.

Within a single dataset, the relevant comparison is the relative change in δ_h before and after filtering. A truly fair cross-study comparison would require applying the same reference filter to our dataset, which we leave for future work. Meanwhile, our conclusions are supported by MSE on synthetic test data and the Equivalent Number of Looks (ENL), metrics that are less sensitive to texture-dependent baselines and still show consistent improvement with the proposed autoencoder approach.

4 Conclusions and Future Work

In this work, we explored the use of autoencoders for speckle noise reduction in SAR images, presenting original contributions in both the training data generation methodology and the development of flexible tools for experimenting with different architectures. We demonstrated that it is possible to use the $\mathcal{G}_I^0$ distribution to generate synthetic training data for training autoencoders capable of filtering speckle noise, although with limitations in their generalization to actual images. Our experiments consistently reveal that deeper and more complex architectures yield better results.

Our findings highlight a key dichotomy in the use of synthetic data: the $\mathcal{G}_I^0$ model offers a controllable environment for architecture development, but a gap remains with real SAR images. Thus, the strongest use of synthetic data is as a pre-training step to provide a statistical prior, followed by fine-tuning on a smaller real SAR dataset. This two-stage strategy can combine synthetic fidelity with real data adaptation and should be validated using both reconstruction error and speckle-preserving metrics.

Future research will focus on exploring more sophisticated architectures, such as U-Net and attention-based models, and on formalizing the hybrid training strategy suggested by our results.We also plan to refine synthetic generation with greater spatial diversity and richer texture variations. In addition, although this work compares the proposed autoencoder approach with the classical Lee filter, a more complete evaluation should include modern speckle reduction algorithms such as the FANS filter. A rigorous comparison with other state-of-the-art methods remains a priority for future work. Finally, we will extend the proposed methodology to more complex data, including polarimetric and multi-temporal SAR images, to assess its performance in a broader range of applications. While the use of Mean Squared Error (MSE) as the loss function provided a solid baseline for this study, we acknowledge its limitations in capturing the perceptual and textural nuances of SAR imagery. As noted, optimizing for MSE can lead to a 'regression to the mean,' resulting in blurred reconstructions that may lack high-frequency details. Our decision was pragmatic, aimed at first validating our data generation and architectural hypotheses in a controlled manner. Thus, our work separates the optimization objective (MSE) from the evaluation criteria: MSE is used to train stable reconstructions, while the speckle-specific metrics verify that the output maintains the desired statistical behavior.

For future work, we plan to explore more advanced loss functions. A promising direction is the development of a custom perceptual loss tailored for SAR images, which would require pre-training a feature extractor on a large SAR dataset. This type of loss compares learned features rather than individual pixels, helping to preserve textural and structural details that MSE alone tends to smooth.

Another direction is to incorporate terms that measure the match between distributions, such as those used in Generative Adversarial Networks (GANs) or Wasserstein-based losses. Implementing this would require a discriminative component or GAN-style architecture capable of learning the statistical distribution of real clean SAR images, which is more complex than our current MSE-based framework.

The positive results from our current MSE-based models provide a strong foundation for these future explorations, which we believe will yield significant improvements in reconstruction quality. In conclusion, this work establishes a solid foundation for using autoencoders for speckle noise reduction in SAR images, demonstrating the potential of combining statistical models with deep learning to address classic radar image processing problems.

References

Cardona-Mesa, A., Vásquez-Salazar, R., Diaz-Paz, J., Sarmiento-Maldonado, H., Gómez, L., & Travieso-González, C. (2025). Optimization of autoencoders for speckle reduction in SAR imagery through variance analysis and quantitative evaluation. *Mathematics*, *13*(3). https://doi.org/10.3390/math13030457

Chan, D., Gambini, J., & Frery, A. C. (2022). Entropy-based non-local means filter for single-look SAR speckle reduction. *Remote Sensing*, *14*(3).

Frery, A., Muller, H.-J., Yanasse, C., & Sant'Anna, S. (1997). A model for extremely heterogeneous clutter. *IEEE Transactions on Geoscience and Remote Sensing*, *35*(3), 648–659. https://doi.org/10.1109/36.581981

Frost, V. S., Stiles, J. A., Shanmugan, K. S., & Holtzman, J. C. (1982). A model for radar images and its application to adaptive digital filtering of multiplicative noise. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *PAMI-4*(2), 157–166.

Gomez, L., Ospina, R., & Frery, A. (2017). Unassisted quantitative evaluation of despeckling filters. *Remote Sensing*, *9*(4). https://doi.org/10.3390/rs9040389

Lattari, F., Leon, B. G., Asaro, F., Rucci, A., Prati, C., & Matteucci, M. (2019). Deep learning for SAR image despeckling. *Remote Sensing*, *11*(13).

Lee, J. (1980). Digital image enhancement and noise filtering by use of local statistics. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *PAMI-2*(2), 165–168.

Parrilli, S., Poderico, M., Angelino, C. V., & Verdoliva, L. (2012). A nonlocal SAR image denoising algorithm based on LLMMSE wavelet shrinkage. *IEEE Transactions on Geoscience and Remote Sensing*, *50*(2), 606–616.

Vásquez-Salazar, R., Cardona-Mesa, A., Gómez, L., Travieso-González, C., Garavito-González, A., & Vásquez-Cano, E. (2024). Labeled dataset for training despeckling filters for SAR imagery. *Data in Brief*, *53*. https://doi.org/10.1016/j.dib.2024.110065

Vitale, S., Ferraioli, G., & Pascazio, V. (2021). Multi-objective CNN-based algorithm for SAR despeckling. *IEEE Transactions on Geoscience and Remote Sensing*, *59*(11).

Zhang, C., Bengio, S., Hardt, M., Recht, B., & Vinyals, O. (2017). Understanding deep learning requires rethinking generalization [arXiv:1611.03530]. *International Conference on Learning Representations (ICLR)*.

Zhang, L., Zhou, X., Wang, Z., Tan, C., & Liu, X. (2017). A nonmodel dualtree wavelet thresholding for image denoising through noise variance optimization based on improved chaotic drosophila algorithm. *International Journal of Pattern Recognition and Artificial Intelligence*, *31*(08).

Zhang, Q., Yuan, Q., Li, J., Yang, Z., & Ma, X. (2018). Learning a dilated residual network for SAR image despeckling. *Remote Sensing*, *10*(2).