Revista de Informática Teórica e Aplicada - RITA - ISSN 2175-2745
Vol. 33, Num. 2 (2026) 129-136

RESEARCH ARTICLE

SAR Images Despeckling Using Convolutional Neural Networks with Stochastic Distances

Filtragem de Ruído Speckle em Imagens SAR Usando Redes Neurais Convolucionais com Distâncias Estocásticas

Pedro Henrique Salles Arouck de Souza[1]*, José Hiroki Saito[1], Francisco Fambrini[2]

Abstract: This work presents a study on speckle noise filtering in SAR images using convolutional neural networks. The objective is to develop an accessible and efficient solution to improve the quality of images degraded by this type of noise, facilitating their analysis and interpretation. The proposed methodology involves the using synthetic SAR-like data, the training of a convolutional neural network, and the evaluation of results using traditional metrics such as PSNR and SSIM. Additionally, stochastic distances were integrated into the loss function of the convolutional neural network, enabling a more detailed analysis of the preservation of the statistical properties of the filtered images. The results indicate that the model effectively reduces speckle noise while preserving the structural details of the image. Furthermore, comparisons with recent studies on the topic showed that the proposed solution achieves competitive performance.
Keywords: convolutional neural networks — noise — speckle — SAR images

Resumo: Este trabalho apresenta um estudo sobre a filtragem do ruído speckle em imagens SAR utilizando redes neurais convolucionais. O objetivo é desenvolver uma solução acessível e eficiente para melhorar a qualidade de imagens degradadas por esse tipo de ruído, facilitando sua análise e interpretação. A metodologia proposta envolve o uso de dados sintéticos semelhantes a SAR, o treinamento de uma rede neural convolucional e a avaliação dos resultados por meio de métricas tradicionais, como PSNR e SSIM. Além disso, distâncias estocásticas foram integradas à função de perda da rede neural convolucional, possibilitando uma análise mais detalhada da preservação das propriedades estatísticas das imagens filtradas. Os resultados indicam que o modelo reduz efetivamente o ruído speckle ao mesmo tempo em que preserva os detalhes estruturais da imagem. Ademais, comparações com estudos recentes sobre o tema mostraram que a solução proposta atinge desempenho competitivo.
Palavras-Chave: redes neurais convolucionais — ruído — speckle — imagens SAR

[1] *Centro Universitário Campo Limpo Paulista (UNIFACCAMP), Brazil*
[2] *Universidade Virtual do Estado de São Paulo (UNIVESP), Brazil*
***Corresponding author**: pedrohenrique.sales@hotmail.com
DOI: http://dx.doi.org/10.22456/2175-2745.150936 • **Received:** 14/10/2025 • **Accepted:** 03/11/2025

1. Introduction

Synthetic Aperture Radar (SAR) is an imaging technique employed for Earth observation applications due to its capability to acquire high-resolution imagery independent of weather conditions and solar illumination. SAR imagery plays a role in several domains, including environmental monitoring, agriculture management, urban planning, defense surveillance, and disaster response. Despite its benefits, SAR images are inherently contaminated by multiplicative noise known as speckle, which arises from the coherent summation of reflected signals. This noise degrades the interpretability of SAR data, reducing its usability for visual inspection and automated processing tasks, such as segmentation and classification [1].

Speckle noise reduction, or despeckling, is essential to enhance SAR images' quality, improving both visual analysis and computational performance for downstream applications. Traditional despeckling techniques, such as multilook averaging and spatial-domain filtering (e.g., Lee, Frost, and Kuan filters), typically reduce noise at the expense of spatial resolution, often causing undesired artifacts like blurring and the loss of essential structural details. Over the last decade, more advanced methods have emerged. Non-local approaches, such as Non-Local Means (NLM) and Block-Matching 3D (BM3D), have demonstrated good results by leveraging similarities among image patches. More recently, Convolutional Neural Networks (CNNs) have gained prominence in the despeckling domain, driven by their capabilities in modeling complex image structures and extracting relevant spatial features directly from noisy observations [2].

Despite their performance, these methods still face challenges. One significant challenge is obtaining appropriate datasets for training and benchmarking despeckling algorithms, as highlighted by recent surveys on SAR despeckling methods [3]. CNN-based despeckling techniques often rely on conventional distance metrics, like the Euclidean distance, to measure similarity between patches or features, which may not accurately reflect the statistical distribution of SAR data [4]. As SAR speckle is multiplicative and modeled by specific statistical distributions, conventional metrics might not capture the inherent stochastic nature of SAR data effectively [5].

Motivated by these challenges, this paper explores the integration of stochastic distance measures into the training process of CNNs for speckle noise reduction in SAR images. Instead of relying solely on conventional similarity metrics, we incorporate stochastic distances derived from statistical models commonly used to represent SAR data as feedback in the CNN loss function. By doing so, we aim to align the learning process more closely with the statistical nature of speckle-contaminated imagery.

While not introducing an entirely new technique, this study broadens the perspective on how modern deep learning frameworks can benefit from established statistical tools, encouraging further research into hybrid strategies for SAR image despeckling. In practical terms, improving the quality of SAR imagery benefits a wide range of fields, from environmental monitoring and precision agriculture to defense and disaster management, reinforcing the societal relevance of this line of research.

The remainder of this paper is structured as follows: Section 2 provides the theoretical background on SAR imagery, speckle noise, CNNs, and stochastic distances. Section 3 presents the proposed despeckling methodology, detailing dataset preparation, CNN architecture, and implementation of the selected stochastic distances. Experimental results obtained with synthetic SAR images are discussed in Section 4, highlighting the effectiveness and robustness of the proposed approach compared to other techniques. Finally, Section 5 concludes the paper and outlines future research directions.

2. Theoretical Background

2.1 Synthetic Aperture Radar Imaging and Speckle Noise

Synthetic Aperture Radar imaging involves transmitting electromagnetic pulses toward Earth's surface and analyzing the received echoes to generate high-resolution images. Unlike optical systems, SAR can operate day and night under various weather conditions due to its active sensing nature and use of microwave frequencies [1].

However, due to the coherent processing of backscattered signals, SAR images are inherently contaminated by speckle noise. Speckle arises as a multiplicative noise resulting from constructive and destructive interference of echoes returning from numerous scatterers within a resolution cell, as illustrated in Fig. 1. This phenomenon produces granular artifacts that significantly degrade image quality, complicating interpretation, segmentation, and classification tasks [4].

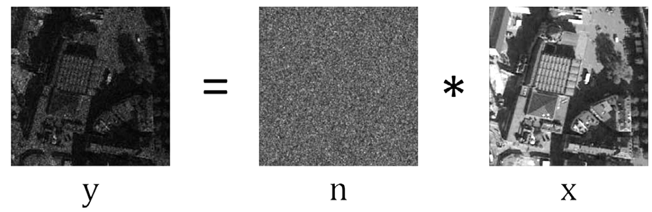

Figure 1. Example of speckle noise in a SAR image, where x = original scene, n = speckle noise, and y = resulting SAR image [4].

Speckle noise in SAR images follows a Gamma distribution, whose probability density function (PDF) is defined as [4]:

$$\rho(n) = \frac{L^L n^{L-1} \exp(-nL)}{\Gamma(L)}, \tag{1}$$

where:

- $\rho(n)$: represents the PDF of the speckle noise, indicating the likelihood of the noise taking a specific value n;
- n: is the random noise component, representing the multiplicative intensity variation in a SAR image pixel;
- L: is the number of looks, which defines the degree of speckle smoothing;
- $\Gamma(\cdot)$: is the Gamma function.

The number of looks reflects the level of signal integration in SAR image formation, representing how many independent observations of the same area are averaged to reduce speckle noise. It is directly related to the variance of the speckle, which is inversely proportional to L. Therefore, higher L values decrease the noise variance, producing smoother images that facilitate both visual interpretation and automated analysis [1].

2.2 Despeckling Techniques

Classical spatial-domain filters, such as the Lee, Frost, and Kuan filters, utilize local statistics to smooth noise but typically compromise spatial resolution, resulting in blurred edges and loss of fine details [3]. The multilook approach effectively reduces speckle by averaging multiple SAR looks, but at the cost of reduced spatial resolution [5].

To address these limitations, modern methods have leveraged non-local similarity between image patches, such as NLM and BM3D [5]. NLM filters images by comparing patches throughout the image and using weighted averages based on patch similarity. BM3D further enhances this idea by grouping similar patches in 3D arrays and collaboratively filtering them [6].

Recently, deep learning-based methods, particularly convolutional neural networks, have emerged as tools for speckle reduction, demonstrating a good performance in preserving spatial structures and achieving high-quality results. CNNs are capable of learning complex patterns directly from noisy input data, thereby surpassing traditional despeckling techniques in terms of effectiveness and flexibility [2].

2.3 Stochastic Distances

Stochastic distances provide statistical measures to compare probability distributions. They are suitable for SAR image analysis, where data follow well-known statistical distributions such as Gamma, or Exponential-Polynomial (EP), distributions [6].

The stochastic distances adopted in this work were derived from the (h,ϕ)-divergence family and correspond to closed-form expressions previously obtained by Penna and Mascarenhas [5].

The following stochastic distances are commonly used and relevant to SAR speckle analysis:

- **Kullback–Leibler (KL)**: Measures the relative entropy between two distributions, indicating how one probability distribution diverges from a second, expected distribution.

$$d = \frac{(\ln(a_1+1)-\ln(a_2+1))(a_1-a_2)}{2(a_1+1)(a_2+1)}. \qquad (2)$$

- **Rényi**: A generalization of KL divergence parameterized by an order α, controlling the sensitivity to distribution tails.

$$d = -2\ln(1-d_{\mathrm{H}}). \qquad (3)$$

- **Bhattacharyya (BH)**: Quantifies the amount of overlap between two distributions; often used to measure class separability.

$$d = -\ln(1-d_{\mathrm{H}}). \qquad (4)$$

- **Jensen–Shannon (JS)**: A symmetric and smoothed version of KL divergence that always produces a finite and bounded value.

$$d = \frac{\ln(2a_2+2)-\ln(a_1+a_2+2)}{2(a_1+1)} + \frac{\ln(2a_1+2)-\ln(a_1+a_2+2)}{2(a_2+1)}. \qquad (5)$$

- **Hellinger (H)**: Is is the similarity between two probability distributions based on the geometric mean, bounded between 0 and 1.

$$d = \frac{a_1+a_2-2\sqrt{(a_1+1)(a_2+1)}}{(2a_1+2)(a_2+1)}. \qquad (6)$$

- **Arithmetic–Geometric (AG)**: Captures differences by comparing arithmetic and geometric means of two distributions; useful for its symmetry and interpretability.

$$d = -\frac{\left[\ln(a_2+1)+\ln(4a_1+4)-2\ln(a_1+a_2+2)\right](a_1+a_2+2)}{4(a_1+1)(a_2+1)}. \qquad (7)$$

- **Triangular (TRI)**: Emphasizes absolute differences and penalizes discrepancies between distributions linearly, offering stability for small variations.

$$d = \frac{(a_1-a_2)^2}{(a_1+1)(a_2+1)(a_1+a_2+2)}. \qquad (8)$$

- **Harmonic Mean (HM)**: Derived from the harmonic average, this distance highlights extreme values in distribution discrepancies and is especially sensitive to outliers.

$$d = -\ln\left(1-\frac{d_{\mathrm{T}}}{2}\right). \qquad (9)$$

These distances, d, quantify dissimilarities between two probability distributions estimated from local image patches, whose means are denoted by a_1 and a_2 [5]. In particular, d_{H} refers to the Hellinger distance, while d_{T} denotes the Triangular distance. Their incorporation into loss functions or similarity measures enables neural networks to better reflect the multiplicative nature of speckle noise and the statistical behavior of SAR imagery.

2.4 CNN Loss Functions and Similarity Metrics

Convolutional Neural Networks are specialized neural networks designed primarily for processing grid-like structured data, such as images. CNNs leverage spatial relationships through convolutional operations, extracting hierarchical features at multiple abstraction levels. Typically, CNN architectures consist of layers including convolution, activation (e.g., ReLU), pooling, and fully-connected layers [4]. Their ability to model complex spatial patterns and adaptively learn feature representations directly from the data has made CNNs the leading technique in image processing tasks, including image classification, segmentation, and denoising.

CNN performance depends on the choice of loss function, which guides network training by quantifying the discrepancy between predicted and target images. Traditionally, Mean Squared Error (MSE) and Mean Absolute Error (MAE) are employed due to their computational simplicity [4]. However, these metrics might not reflect the statistical complexity of SAR images, particularly when dealing with speckle noise.

By integrating stochastic distance measures, CNNs can better capture the statistical characteristics of SAR imagery, enhancing their capability to differentiate between actual image features and noise-induced patterns [5]. This strategy can improve model generalization and produce visually and quantitatively superior results compared to conventional CNN training approaches.

2.5 Evaluation Metrics

To quantitatively assess the despeckling performance, three complementary metrics were adopted: peak signal-to-noise ratio (PSNR), structural similarity index (SSIM) and equivalent number of looks (ENL).

These measures evaluate the fidelity, perceptual similarity, and residual noise characteristics of the filtered images.

The PSNR quantifies the ratio between the maximum possible signal power and the noise power, providing an objective indication of overall image fidelity. Higher PSNR values indicate lower residual noise and better preservation of radiometric information [7]. It is expressed as:

$$PSNR = 10\log_{10}\left(\frac{g_{\max}^2}{MSE}\right), \tag{10}$$

where $g_{\max}$ denotes the maximum possible pixel value.

The MSE between the original and despeckled images is given by:

$$MSE = \frac{1}{M \times N}\sum_{i,j}(z_{i,j} - f_{i,j})^2, \tag{11}$$

with M and N being the image dimensions in pixels, $z_{i,j}$ represents the gray-level intensity of the noisy input image at pixel position (i,j), while $f_{i,j}$ denotes the corresponding intensity in the filtered image.

The SSIM measures perceptual similarity between two images, considering luminance, contrast, and structural information, and is defined as:

$$SSIM = \frac{(2\bar{z}\bar{f} + C_1)(2\sigma_{zf} + C_2)}{(\bar{z}^2 + \bar{f}^2 + C_1)(\sigma_z^2 + \sigma_f^2 + C_2)}, \tag{12}$$

where $\bar{z}$ and $\bar{f}$ are the local means, σ_z and σ_f the standard deviations, and σ_{zf} the cross-covariance between the noisy and filtered images. Constants $C_1 = (0.01 \times r)^2$ and $C_2 = (0.03 \times r)^2$ depend on the dynamic range r (255 for 8-bit images). SSIM values range from -1 (completely dissimilar) to 1 (identical), with higher scores indicating better structural preservation [7].

Finally, for homogeneous regions, the ENL quantifies the amount of residual speckle noise and is defined as:

$$ENL = \frac{\mu^2}{\sigma^2}, \tag{13}$$

where μ and σ^2 denote the mean and variance of the intensities within a homogeneous region. A higher ENL indicates a smoother region with less speckle [5].

3. Methodology

3.1 Dataset Preparation

For training and validating our despeckling approach, we used the RESISC45 dataset, a publicly available remote sensing image dataset containing 31,500 optical RGB images distributed across 45 scene classes [8]. To simulate SAR-like conditions, images were converted to grayscale and corrupted with synthetic speckle noise, applied on-the-fly during data loading [3].

We followed a 70-15-15% split strategy: 22,050 images were used for training, 4,725 for validation, and 4,725 for testing. The speckle noise was modeled as multiplicative, defined by:

$$I_{\text{noisy}} = I_{\text{clean}} + I_{\text{clean}} \times N, \tag{14}$$

where N follows a uniform distribution with zero mean and variance inversely proportional to the number of looks L.

The effectiveness of loss functions based on stochastic distances is directly dependent on the volume of data and the training duration. As the number of samples increases, whether through larger patches or a broader training set, and the optimization process is extended over more epochs, estimates of distributional divergences become more accurate [9].

Using larger images, training for more epochs, and employing a robust dataset enhances the ability of these loss functions to capture distributional similarities, leading to better generalization and visually more faithful results.

3.2 CNN Denoising Model

The denoising architecture used is a custom Convolutional Neural Network implemented in Python using PyTorch. The model comprises:

- Input: Single-channel grayscale image.
- Convolutional layers: One initial convolution layer (32 filters), followed by seven blocks of Conv-BatchNorm-ReLU-Dropout (with dropout rate of 0.5), and a final 1-channel convolution with sigmoid activation.
- All convolutions use 3×3 kernels with padding to preserve dimensions.

This architecture is tailored for low-level vision tasks such as denoising, balancing learning capacity and computational efficiency.

Fracastoro *et al.* [2] review deep learning approaches for SAR despeckling and emphasize that convolutional architectures have become central to modern methods, providing effective noise suppression while maintaining efficiency and fine spatial detail.

In the same direction, Passah *et al.* [4] propose a deep fully convolutional CNN with multiple skip connections for

SAR speckle reduction, trained on synthetically speckled data, showing that such architectures successfully learn nonlinear mappings from noisy to clean intensities.

Following this idea, the present work adopts a lightweight fully convolutional model (nine convolutional layers, no pooling or upsampling), keeping architectural complexity modest while focusing the methodological contribution on the loss formulation based on stochastic distances.

3.3 Training and Loss Function

The training process uses the Adam optimizer [10] with a learning rate of 10^{-3}, batch size 5, and 10 epochs. Dataloader classes efficiently handle preprocessing and on-the-fly noise injection using PyTorch's 'DataLoader' module.

To guide training, we adopt a composite objective that combines a pixel-wise fidelity term with multiple stochastic distances computed between the predicted and target images, treated as discrete probability distributions. This aligns optimization with both per-pixel accuracy and distributional similarity relevant to speckle statistics.

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{MSE}}(\hat{I}, I) + \sum_{d \in \mathcal{D}} w_d \, D_d\big(p(\hat{I}),\, p(I)\big) \tag{15}$$

here, $\hat{I}$ denotes the network output, I the clean reference, and $p(\cdot)$ a non-negative, normalized discrete distribution (with a small stability term ε). The set $\mathcal{D} = \{\text{KL}, \text{JS}, \text{H}, \text{AG}, \text{BH}, \text{HM}, \text{TRI}, \text{Rényi}\}$ contains the stochastic distances used.

Traditional data-fidelity terms for SAR despeckling are usually based on the MSE loss functions or its variants,as discussed by Fracastoro *et al.* [2], most despeckling networks include it in their loss functions, sometimes combined with aditional terms, and there is currently no clear evidence favoring one over the other for SAR denoising.

In our experiments, we observed that the MSE component yields more stable convergence when the amount of training data or the number of epochs is limited.

The stochastic distances require more data and iterations to become effective, since they rely on local statistical estimates that stabilize only after sufficient sampling. When this condition is met, they contribute to preserving structural and textural details while maintaining overall fidelity.

Learnable weights were adopted for KL, JS, and Hellinger, while the remaining terms were added with unit weight. This composite objective keeps the optimization aligned with both pixel-wise fidelity and distributional similarity relevant to speckle statistics.

All terms are computed with tensor operations and non-negative normalization (with ε) to ensure full differentiability.

We evaluate two approaches:

- A baseline using the traditional MSE loss.
- A stochastic loss combining MSE with multiple stochastic distances, as described above.

3.4 Experimental Setup

The experiments were carried out using the publicly available NWPU-RESISC45 dataset [11], which contains 31,500 remote sensing images divided into 45 scene classes, each class including 700 images. All images are RGB with a spatial resolution of 256×256 pixels. For the purposes of this study, the images were converted to grayscale and normalized to the range $[0, 1]$.

To simulate synthetic SAR-like data, multiplicative noise was added to each image according to the model $y = x \times n$, where n follows a Gamma distribution with a single-look parameter ($L = 1$). The corresponding clean images were used as ground truth for computing the evaluation metrics.

The model was trained for 10 epochs using the Adam optimizer with a learning rate of 10^{-3} and a batch size of 5.

3.5 Implementation Details

All scripts were implemented in Python. The system is composed of modular files:

- `data.py`: has the custom dataset and preprocessing pipeline, including grayscale conversion and speckle noise injection.
- `model.py`: contains the CNN architecture definition.
- `loss.py`: implements the combined loss functions including stochastic distances.
- `train.py` and `eval.py`: manage training, validation, and testing loops.
- `checkpoint.py`: handles saving and loading model weights.
- `visualize.py`: generates visual comparisons between noisy, ground truth, and denoised images.

The code is modular and reproducible, allowing controlled experimentation with loss functions, data fractions, network depths, and distance formulations. All experiments were executed on a local GPU workstation using PyTorch.

The codebase enables systematic testing of different loss formulations, data fractions, network depths, and stochastic distance combinations. The full implementation,is publicly available for research and reproducibility purposes at https://github.com/pedroarouck/SAR-Images-Despeckling

4. Results and Discussions

4.1 Quantitative Analysis

We quantitatively assessed our proposed CNN approach integrating stochastic distances against the baseline CNN approach trained with the traditional Mean Squared Error loss function.

The numerical results summarized in Table 1 illustrate the quantitative effect of integrating stochastic distances into the loss function. The column labeled *MSE* corresponds to the baseline CNN trained only with the mean squared error loss, whereas the column *Stochastic* represents the proposed model trained with the composite stochastic loss.

It is worth noting that the stochastic distance values reported in Table 1 are presented for indicative purposes only. They represent the average distance values computed between the despeckled outputs and their corresponding reference images, and therefore serve as an internal statistical measure of similarity rather than a direct image-quality metric. Smaller distance values indicate closer distributional agreement between the filtered and reference images.

Metric	Baseline (MSE)	Proposed (Stochastic)
KL Divergence	0.0094	0.0042
Rényi Divergence	0.0067	0.0030
Hellinger Distance	0.0489	0.0325
Bhattacharyya Distance	0.0024	0.0011
Jensen–Shannon Divergence	0.0024	0.0011
Arithmetic–Geometric	0.0024	0.0011
Triangular	0.0095	0.0042
Harmonic Mean	0.0473	0.0341
PSNR (dB)	21.3237	25.4187
SSIM	0.4888	0.5864

Table 1. Comparison of image metrics between baseline CNN (trained with MSE only) and the proposed CNN (trained with the composite loss described in the Equation 15, combining MSE with multiple stochastic distances).

These results indicate an improvement in all evaluated metrics, including traditional metrics such as PSNR and SSIM, as well as stochastic distances, demonstrating the effectiveness of integrating statistical measures directly into the loss function.

4.2 Qualitative (Visual) Analysis

Visual results from our experiments are presented in Figure 2. The baseline method produced denoised images with significant residual noise and artifacts. In contrast, the CNN trained with stochastic distances reduced noise while better preserving important structural details, enhancing overall visual quality. This confirms quantitatively measured improvements and highlights the practical benefit of our proposed method.

4.3 Comparison with Other Techniques

For further benchmarking, we compared our method's performance against the despeckling techniques presented by Dabhi *et al.* (2020) [3], as summarized in Table 2.

Here, the ENL quantifies speckle reduction quality by the ratio of the local mean squared to its variance; PPB refers to the Probabilistic Patch-Based filter, which weights patches via a maximum-likelihood criterion under a Gamma noise model and SAR-BM3D denotes the SAR-adapted Block-Matching 3D algorithm, an extension of the classic BM3D framework to multiplicative speckle via local Gamma-based shrinkage in grouped 3D blocks.

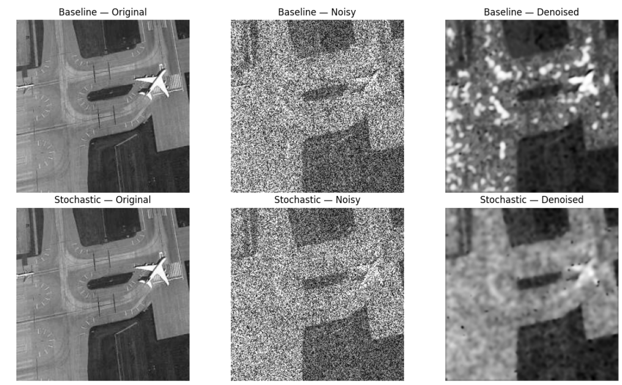

Figure 2. Visual comparison between baseline CNN (MSE loss) and proposed CNN (stochastic distances loss).

Table 2. Benchmarking against state-of-the-art despeckling techniques from Virtual SAR paper.

Method	ENL	PSNR (dB)	SSIM
Noisy	11.237	18.182	0.294
Lee filter	20.852	22.965	0.457
Kuan filter	20.950	22.989	0.458
Frost filter	21.080	22.945	0.453
PPB	30.351	25.830	0.594
SAR-BM3D	34.990	26.025	0.617
CNN (Virtual SAR)	31.362	27.119	0.661
Baseline (Ours)	–	21.324	0.489
Proposed (Ours)	–	25.419	0.586

Although the original images used for training and testing differ from those adopted in the works compared in Table 2, the procedure for generating synthetic SAR data was similar. In all cases, speckle noise was added following the multiplicative model $y = x \times n$, where n follows a Gamma distribution with a single-look parameter.

For the evaluation metrics, the corresponding noise-free images were used as ground truth, as also done in the benchmarked methods. Therefore, the numerical results reported here can be comparable in principle, since both our approach and the reference techniques rely on equivalent noise modeling and assessment criteria, even though different image sets were employed.

In Table 2, the ENL values quantify speckle reduction in homogeneous areas. Higher ENL values correspond to smoother and more uniform regions, reflecting lower residual speckle. Classical filters such as Lee, Kuan, and Frost show moderate ENL gains but tend to oversmooth textures. More advanced methods like PPB and SAR-BM3D achieve higher ENL.

The proposed method shows comparable performance to advanced methods like PPB and SAR-BM3D and surpasses classical spatial-domain filtering methods. Although our SSIM and PSNR are slightly below the best CNN-based approach from Virtual SAR, our results demonstrate that incorporating stochastic distances directly into the training process bridges the performance gap, validating the effectiveness of this novel integration strategy.

4.4 General Discussion

The integration of stochastic distances into the CNN training loss function has demonstrated benefits, improving both quantitative metrics and visual quality of denoised SAR images. However, certain limitations remain, such as residual noise and potential over-smoothing of delicate features. Future research could focus on optimizing the combination of stochastic distances or employing adaptive weighting strategies within the loss function to further refine results.

The proposed method also faces practical challenges, particularly related to overfitting, which can degrade the CNN's performance on unseen data. This issue is exacerbated by the inherent difficulty of obtaining large-scale annotated SAR datasets for training robust models. As a limitation, the experiments in this study were conducted on synthetic SAR-like data derived from the RESISC45 dataset. While this provides a controlled and reproducible setup, future work should include evaluation on real SAR imagery to better assess generalization in operational contexts.

It is also noted that incorporating stochastic distances slightly increases training time due to the additional divergence computations. Thus, balancing performance improvement with computational feasibility remains a crucial challenge for future research.

5. Conclusion

In this study, we explored the integration of stochastic distance measures into the loss function of a CNN for SAR image despeckling. By leveraging statistical similarity metrics specifically tailored to SAR data, our method improved noise reduction performance compared to a baseline CNN trained with traditional MSE loss. The experimental results demonstrated marked improvements across several quality metrics, including PSNR, and SSIM. Visual analysis further validated the effectiveness of our approach, revealing superior preservation of image details alongside effective speckle noise suppression.

The main contribution of this work is not a new CNN architecture, but rather the demonstration that stochastic distances, when incorporated into the training objective, can enhance despeckling performance by aligning the optimization process with the statistical properties of SAR imagery. This highlights the potential of combining well-established statistical tools with modern deep learning techniques.

We acknowledge some limitations: the evaluation was restricted to synthetic SAR-like data, which provided a controlled and reproducible setup but does not fully reflect the complexity of real SAR acquisitions. In addition, incorporating stochastic distances slightly increased training cost due to the additional divergence computations.

Future work can extend this line of research by validating the approach on real SAR images, exploring adaptive weighting strategies for different distances, and assessing the trade-off between accuracy and computational feasibility in larger-scale experiments. Looking ahead, stochastic distances could also be embedded into more advanced neural architectures, such as U-Nets or Vision Transformers, opening possibilities for large-scale and real-time despeckling. Beyond methodological improvements, the potential applications in agriculture, climate monitoring, and security underline the importance of developing robust and reliable despeckling techniques for the next generation of Earth observation systems.

Author Contributions

Pedro Henrique Salles Arouck de Souza: Conducted the research as part of a Master's dissertation in Computer Science, developed the methodology and implemented the proposed algorithm. Prof. José Hiroki Saito:Supervisor of the dissertation, responsible for the organization, guidance, and revision of the work. Prof. Francisco Fambrini: Member of the dissertation defense committee, contributed to the adjustment of the experimental design using more relevant images and supported the adaptation of the text for the article submission.

References

[1] ARSET. *An Introduction to Synthetic Aperture Radar (sar) and Its Applications.* [S.l.], 2024. Applied Remote Sensing Training Program (ARSET). Disponível em: ⟨https://appliedsciences.nasa.gov/get-involved/training/english/arset-introduction-synthetic-aperture-radar-sar-and-its-applications⟩.

[2] FRACASTORO, G. et al. Deep learning methods for synthetic aperture radar image despeckling: An overview of trends and perspectives. *IEEE Geoscience and Remote Sensing Magazine*, v. 9, n. 2, p. 29–51, 2021.

[3] DABHI, S. et al. *Virtual SAR: A Synthetic Dataset for Deep Learning based Speckle Noise Reduction Algorithms*. 2020.

[4] PASSAH, A. et al. Sar image despeckling using deep cnn. *IET Image Processing*, v. 15, n. 6, p. 1285–1297, 2021.

[5] PENNA, P. A. A.; MASCARENHAS, N. D. A. Sar speckle nonlocal filtering with statistical modeling of haar wavelet coefficients and stochastic distances. *IEEE Transactions on Geoscience and Remote Sensing*, v. 57, n. 9, p. 7194–7208, 2019.

[6] EVANGELISTA, R. C. et al. A new bayesian poisson denoising algorithm based on nonlocal means and stochastic distances. *Pattern Recognition*, v. 122, p. 108363, 2022. ISSN 0031-3203.

[7] SANTOS, C. A. N. *Redução de ruído speckle em imagens de ultrassom com filtragem não-local e distâncias estocásticas*. Dissertação (Master's thesis) — UFSCAR, 2017.

[8] CHENG, G.; HAN, J.; LU, X. Remote sensing image scene classification: Benchmark and state of the art. *Proceedings of the IEEE*, v. 105, n. 10, p. 1865–1883, 2017.

[9] BRIGATO, L.; IOCCHI, L. A close look at deep learning with small data. In: *2020 25th International Conference on Pattern Recognition (ICPR)*. [S.l.: s.n.], 2021. p. 2490–2497.

[10] KINGMA, D. P.; BA, J. Adam: A method for stochastic optimization. In: BENGIO, Y.; LECUN, Y. (Ed.). *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*. [S.l.: s.n.], 2015.

[11] CHENG, G.; HAN, J.; LU, X. Remote sensing image scene classification: Benchmark and state of the art. *CoRR*, abs/1703.00121, 2017.