remote sensing

MDPI

Article

Comparative Analysis of Despeckling Filters Based on Generative Artificial Intelligence Trained with Actual Synthetic Aperture Radar Imagery

Ahmed Alejandro Cardona-Mesa [1], Rubén Darío Vásquez-Salazar [2], Carlos M. Travieso-González [3,*] and Luis Gómez [4]

[1] Faculty of Sciences and Humanities, Institución Universitaria Digital de Antioquia, 55th Av, 42-90, Medellín 050012, Colombia; amhed.cardona@iudigital.edu.co
[2] Faculty of Engineering, Politécnico Colombiano Jaime Isaza Cadavid, 48th Av, 7-151, Medellín 050022, Colombia; rdvasquez@elpoli.edu.co
[3] Signals and Communications Department, IDeTIC, Universidad de Las Palmas de Gran Canaria, 35001 Las Palmas de Gran Canaria, Spain
[4] Electronic Engineering and Automatic Control Department, IUCES, Universidad de Las Palmas de Gran Canaria, 35001 Las Palmas de Gran Canaria, Spain; luis.gomez@ulpgc.es
* Correspondence: carlos.travieso@ulpgc.es

Abstract: The speckle is a granular undesired pattern present in Synthetic-Aperture Radar (SAR) imagery. Despeckling has been an active field of research during the last decades, with approaches from local filters to non-local filters that calculate the new value of a pixel according to characteristics of other pixels that are not close, the more advanced paradigms based on deep learning, and the newer based on generative artificial intelligence. For the latter, it is necessary to have a large enough labeled dataset for training and validation. In this study, we propose using a dataset designed entirely from actual SAR imagery, calculated by multitemporal fusion operations to generate a ground truth reference, which will yield the models to be trained with the actual speckle patterns in the noisy images. Then, a comparative analysis of the impacts of including the generative capacity in the models is performed through visual and quantitative assessment. From the findings, it is concluded that the use of generative artificial intelligence with actual speckle exhibits notable efficiency compared to other approaches, which makes this a promising path for research in the context of SAR imagery.

Keywords: synthetic-aperture radar (SAR); remote sensing; deep learning; despeckling; generative artificial intelligence

check for updates

Academic Editor: Dusan Gleich

Received: 13 December 2024
Revised: 29 January 2025
Accepted: 24 February 2025
Published: 27 February 2025

Citation: Cardona-Mesa, A.A.; Vásquez-Salazar, R.D.; Travieso-González, C.M.; Gómez, L. Comparative Analysis of Despeckling Filters Based on Generative Artificial Intelligence Trained with Actual Synthetic Aperture Radar Imagery. *Remote Sens.* **2025**, *17*, 828. https://doi.org/10.3390/rs17050828

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Synthetic-Aperture Radar (SAR) imagery is compromised by the speckle phenomenon, which arises from the coherent illumination utilized in the imaging process as well as the intricate backscatter mechanisms associated with the reflected radar signals. This speckle complicates the interpretation and processing operations when SAR imagery is used. The speckle also complicates the evaluation of filtering techniques, as ground truth data for accurate validation are not accessible [1]. For artificial intelligence, and more especially deep learning (DL), it is necessary to count on a large enough dataset containing labeled data to effectively train these models. In some cases, thousands or millions of samples are required. Particularly in the despeckling process, pairs of noisy and ground truth images are required to train a model that predicts noiseless images. In [2], a dataset with these

pairs of images obtained from Sentinel-1 is available, containing 1600 pairs of noisy and ground truth images of 512 × 512 pixels.

The use of DL to despeckle SAR images has been widely used, with significant advancements achieved through convolutional neural networks (CNNs) and autoencoders. Among the prominent methods is the multi-objective network (MONet), which integrates a modified cost function leveraging the Kullback–Leibler divergence for improved performance [3,4]. Another widely cited approach is the Semantic Conditional U-Net (SCUNet), which employs a U-shaped architecture to effectively capture spatial features while preserving fine details [5,6]. Autoencoders, such as those introduced in prior research, utilize encoder–decoder structures to compress and reconstruct SAR images, enabling the attenuation of speckle while maintaining edge information [7,8]. Additionally, hybrid models combining total variation and anisotropic diffusion techniques with CNNs have demonstrated enhanced denoising capabilities, balancing noise reduction and detail preservation [9,10]. These DL-based approaches have consistently outperformed traditional filtering methods, setting a foundation for further advancements in speckle noise mitigation for SAR imagery.

Also, several GenAI models have been proposed to despeckle SAR images, with Generative Adversarial Networks (GANs) and transformer-based architectures leading this innovation. GANs, such as Pix2Pix and Wasserstein GAN (WGAN), have been utilized to generate despeckled images by training adversarial networks that learn the distribution of speckle-free SAR imagery [11,12]. Conditional GANs (C-GANs) further improve performance by conditioning the generation process on specific input features, enabling better alignment between noisy and despeckled outputs [13]. Transformers, like the SWIN transformer and Restormer, have introduced attention mechanisms to the despeckling process, excelling at capturing long-range dependencies in SAR imagery [14–16]. The U-Former architecture, inspired by U-Net but enhanced with self- and cross-attention modules, combines hierarchical feature extraction with refined edge preservation [17]. These models have demonstrated superior performance in both visual and quantitative assessments, marking a paradigm shift in SAR despeckling research.

However, most of these models have not been trained on actual SAR imagery, since they use the traditional protocol proposed in [18], in which optical images are considered as noiseless data and then they are corrupted by using a model designed by using the Gamma distribution law. With this operation, pairs of noisy and ground truth data are artificially generated. However, it has been proven that models trained on synthetic datasets usually perform poorly in practice [19].

In our approach, we propose leveraging state-of-the-art models, including transformers and GenAI, to be trained and validated using a dataset composed of actual SAR images. This methodology enables the model to converge effectively during training by learning from the authentic patterns and statistical distribution of the speckle noise inherent in SAR data. To evaluate the impact of integrating GenAI into the despeckling process, we compare its performance against the results of the autoencoder (AE) presented in [20]. The evaluation is conducted both visually and quantitatively, utilizing established performance metrics.

This paper is organized as follows. In Section 2, the materials and methods are presented, including the speckle model, the different approaches for despeckling SAR images and the description of the GenAI models considered in this study. In Section 3, the experimental results are presented, describing both visual inspection and quantitative measurements using the despeckling models trained by using the dataset including actual SAR imagery. In Section 4, a discussion and interpretation of the results is presented. Finally, in Section 5, some relevant conclusions and future work are outlined.

2. Materials and Methods

The methodological framework of this study is designed to address the challenges of despeckling Synthetic-Aperture Radar (SAR) images using advanced artificial intelligence techniques. This section outlines the speckle modeling approach, the selection of despeckling filters, and the evaluation metrics employed, providing a comprehensive foundation for understanding the comparative analysis performed in this work.

2.1. Speckle Model

Speckle noise is an inherent characteristic of SAR imagery, caused by the coherent nature of the radar signal. This noise appears as a grainy texture that can obscure important details and make image interpretation challenging. Unlike random noise in optical images, speckle is multiplicative, meaning it is dependent on the signal itself and varies across different regions of the image. Effective speckle reduction techniques are crucial for improving the usability of SAR images in applications such as environmental monitoring, agriculture, and disaster management. To understand and mitigate speckle, it is essential to model its behavior mathematically, which helps in designing appropriate filtering techniques.

An SAR image Y can be expressed in terms of the noiseless image X and the speckle noise N according to Equation (1).

$$Y = X \cdot N, \tag{1}$$

where N has a distribution according to the Gamma distribution law characterized by the density in Equation (2).

$$f_Z(z; L, \sigma^2) = \frac{L^L}{\sigma^{2L}\Gamma(L)} z^{L-1} e^{-Lz/\sigma^2}, \tag{2}$$

where L is the Equivalent Number of Looks (ENL) of the SAR image and $\Gamma(L)$ is the Gamma distribution law [21]. For ease of implementation in programming languages, with $a = L$ and $b = \sigma^2/a$, the above expression can be rewritten as Equation (3).

$$f_Z(z; a, b) = \frac{1}{b^a \Gamma(a)} z^{a-1} e^{z/b}. \tag{3}$$

While the Gamma distribution remains a widely accepted model for describing speckle noise in SAR images, it is crucial to consider the variability introduced by different imaging conditions, such as sensor characteristics, resolution, and scene heterogeneity. Advanced noise modeling approaches have also explored incorporating contextual information or non-stationary models to better represent the complex nature of the speckle in real-world scenarios. These refinements ensure that speckle models not only provide a theoretical foundation but also align closely with empirical observations, enabling the design of more effective despeckling algorithms tailored to specific SAR datasets and applications.

2.2. Despeckling Filters

The objective of despeckling a noisy SAR image is to smooth homogeneous regions, while, at the same time, the edges, texture, and details of the image are preserved [22]. The problem of despeckling SAR data continues to be an active area of research, with ongoing efforts to develop effective solutions despite the initial filtering algorithms proposed more than 30 years ago [23].

Newer filters are artificial intelligence models, and some examples are the multi-objective network (MONet) proposed in [24] which uses a modified cost function that includes the Kullback–Leibler divergence; the autoencoder proposed in [20] using a convolutional neural network (CNN) used to learn efficient codings of images with an encoder–

decoder structure; and the Semantic Conditional U-Net (SCUNet) with U-shaped DL architecture proposed in [25], among others.

The despeckling filters considered in this study are the Lee filter since it is one of the most traditional and well-known approaches for despeckling SAR imagery; FANS [26] since it is also a well-known despeckling non-local filter with an outstanding performance on SAR images; the AE proposed in [20] since it was trained on the same dataset used in this study, demonstrating outperformance compared to FANS in the actual SAR images; models based on Generative Adversarial Networks (GANs) originally proposed in [27], considered one of the first approaches to generative artificial intelligence; and finally, the transformer-based models that use attention modules proposed in [28], which created a new paradigm in the field of GenAI.

Despeckling filters play a crucial role in enhancing the quality of SAR images by reducing noise while preserving important features such as edges and textures. Traditional filters, such as the Lee and FANS filters, are widely used due to their simplicity and efficiency, but they may struggle to maintain fine details in complex scenes. On the other hand, modern deep learning-based approaches, including autoencoders and generative models, offer more advanced capabilities by learning patterns from large datasets and adapting to different noise conditions. The recent introduction of transformer-based models further improves despeckling by capturing long-range dependencies in the image, resulting in a better balance between noise reduction and detail preservation. Choosing the most suitable despeckling approach depends on the specific requirements of the application, such as processing speed, computational resources, and the level of detail needed for analysis.

2.2.1. Lee Filter

The Lee filter is one of the most widely used techniques for reducing speckle noise in SAR images due to its simplicity and effectiveness. It works by analyzing the local statistical properties of an image, such as the mean and variance, to identify and reduce noise while preserving important details. The main idea behind the Lee filter is to distinguish between homogeneous areas, where noise reduction is applied more aggressively, and regions with significant features, where details are preserved. This adaptability makes the Lee filter particularly useful in applications that require both noise suppression and edge retention, such as terrain mapping and environmental monitoring.

Possibly the most common despeckling approach is the Lee filter proposed in [29], whose operation is given by Equation (4):

$$I_{ij}^{Lee} = \bar{I}_{ij} + \frac{\sigma_{ij}^2}{\bar{I}_{ij}^{-2}\sigma_{\mu}^2 + \sigma_{ij}^2}(I_{ij} - \bar{I}_{ij}), \tag{4}$$

where I_{ij}^{Lee} is the despeckled image, $\bar{I}_{ij}$ stands for the local mean in the scanning window centered on the ij-th pixel, I_{ij} denotes the central element in the window, and σ_{ij}^2 is the variance of the pixel values in the current window.

Despite its simplicity and long-standing use, the Lee filter remains a foundational technique in SAR despeckling due to its ability to balance noise reduction and detail preservation. However, its performance is highly dependent on the choice of window size and the homogeneity of the image region being processed, which can lead to over-smoothing in textured areas or insufficient noise reduction in highly heterogeneous regions. Recent studies have explored adaptive variations of the Lee filter, incorporating local statistical analysis or combining it with modern techniques, such as wavelet-based methods, to address these limitations [30]. These advancements highlight the filter's ongoing relevance while ac-

knowledging its constraints in handling the complex speckle patterns in contemporary SAR imagery.

2.2.2. FANS Non-Local Filter

Other types of filters appeared later and are considered as non-local since they replaced the value of a pixel with the average of similar pixels that have no reason to be spatially close [31]. The Fast Adaptive Nonlocal SAR (FANS) [26] is a well-known example of this type of non-local filter.

The FANS non-local filter represents a significant step forward in SAR despeckling by leveraging the similarity between distant pixels, a principle that traditional local filters fail to capture. Its ability to identify and average pixels with similar characteristics, regardless of their spatial proximity, allows for superior noise reduction in homogeneous regions and better preservation of fine details. However, its reliance on computationally intensive matching processes can lead to higher processing times, especially for high-resolution SAR images. Recent developments have focused on optimizing the FANS algorithm by incorporating machine learning techniques or parallel processing to improve its efficiency without sacrificing performance. These advancements solidify FANS as a robust choice for SAR despeckling, particularly in applications where preserving image texture and structural details are critical.

FANS represents an advancement over the previously proposed SAR-BM3D [32], including principles of non-local filtering and wavelet-domain shrinkage for the denoising of additive white Gaussian noise. This approach incorporates modifications to its processing steps to address the unique characteristics of SAR. Upon its publication, FANS demonstrated superior performance compared to other leading reference techniques, particularly regarding PSNR in simulated speckled images, as well as enhancements in perceived image quality.

2.2.3. Autoencoder Without GenAI

Autoencoders are a type of deep learning model commonly used for image processing tasks, including despeckling SAR images. They work by learning to compress and then reconstruct an image, removing noise in the process while preserving important details such as edges and textures. Unlike traditional filters, autoencoders do not rely on fixed mathematical formulas; instead, they learn patterns from data, making them highly adaptable to different noise levels and image characteristics. This makes autoencoders a powerful tool for improving the quality of SAR images, enabling better interpretation and analysis in various applications such as environmental monitoring and urban planning.

The methodology proposed in [20] uses the same dataset for both the training and evaluation of filter performance since the ground truth images are not corrupted with simulated speckle. With this approach, the model learns genuine speckle patterns generated by the satellite sensor and not from corrupted images, generating a better approximation to denoising results. The cost function used during the training process is binary cross-entropy/log loss, according to Equation (5).

$$LogLoss = -\frac{1}{N}\sum_{i=1}^{N} y_i \cdot \log(p(y_i)) + (1 - y_i) \cdot \log(1 - p(y_i)), \quad (5)$$

where N is the total number of items of training data, y_i is the i-th output of the model, $p(y_i)$ is the predicted probability of y_i.

The structure of the AE (Figure 1) consists of several stages, starting from an encoder that compresses the images from 512 × 512 in the input layer to 256 × 256, then in the latent space layer the images are compressed even more to 128 × 128, and finally, in the decoder

layer, the images are restored to their original size. During this compression–decompression process, the speckle is eliminated or attenuated while preserving edges and important details of the SAR information.

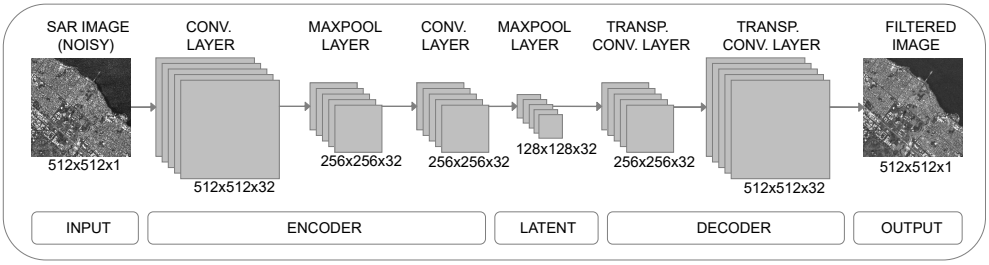


Figure 1. Structure of the autoencoder composed of input, encoder, latent layer (bottleneck), decoder, and output [33].

The best results described in that study show a consistent outperformance when tested on actual SAR images by measuring the ENL [34], PSNR [35], MSE [35], SSIM [36], and M estimator [37].

2.2.4. Generative Adversarial Networks (GANs)

Generative Adversarial Networks (GANs) are an advanced deep learning technique that can be used to improve the quality of SAR images by effectively reducing speckle noise. GANs consist of two neural networks that work together in a competitive manner: a generator, which tries to create despeckled images that resemble real ones, and a discriminator, which evaluates how realistic the generated images are. Through this continuous process of competition and improvement, GANs learn to produce high-quality despeckled images that maintain important details and structures. This innovative approach has shown promising results in many image processing applications, offering a powerful alternative to traditional despeckling methods.

As stated above, GANs were proposed in [27] as an adversarial process in which two models are trained simultaneously, namely a generator G and a discriminator D that estimates the probability that a sample came from training data rather than G, while G is trained to maximize the probability of D making a mistake. They play the following two-player minimax game with value function $V(G, D)$, according to, Equation (6).

$$\min_G \max_D V(D, G) = \mathbb{E}_{x \sim p_{data}(x)}[\log D(x)] + \mathbb{E}_{z \sim p_z(z)}[1 - \log D(G(z))]. \quad (6)$$

More variations of the GANs have been proposed like the Conditional GANs (C-GANs) [38], and even Deep Convolutional GAN (DCGAN), Laplacian GAN (LapGAN), Wasserstein GAN (WGAN), Information Maximizing GAN (InfoGAN), Energy-based GAN (EBGAN), Boundary Equilibrium GAN (BEGAN), Progressive Growing GAN (PGGAN), Big GAN (BigGAN), Style-based Generative Architecture for GANs, Pix2Pix, and more studied in [39], which shows GANs as a very active topic of research.

Including this approach of GANs, two models will be considered, an autoencoder and a U-Net, namely GenAE and GenU-Net, respectively.

2.2.5. Transformers

Transformers are a cutting-edge deep learning technology originally developed for natural language processing tasks, but they have recently been adapted to improve image processing, including the despeckling of SAR images. Unlike traditional approaches that focus on local image features, transformers use an advanced mechanism called attention, which allows them to analyze and understand relationships between distant parts

of an image. This ability to capture long-range dependencies makes transformers particularly effective in preserving important details such as edges and textures while removing speckle noise. In the context of SAR despeckling, transformers analyze the entire image simultaneously, identifying patterns and structures that help in distinguishing noise from meaningful information. Recent transformer-based models, such as the SWIN transformer and U-Former, have introduced improvements that allow them to efficiently process high-resolution SAR images with remarkable accuracy. These models are designed to divide images into smaller patches and process them in a hierarchical manner, ensuring that both fine details and overall structures are well preserved. Thanks to their flexibility and powerful feature extraction capabilities, transformers have become a promising approach for enhancing the quality of SAR images, making them suitable for applications in remote sensing, disaster management, and environmental monitoring.

Before the transformer architectures were proposed in [28], the dominant sequence transduction models were considered the autoencoders. Some of them were improved through attention mechanisms for the natural language processing (NLP) tasks. The attention function as a set of queries simultaneously, packed together into a matrix Q. The keys and values are also packed together into matrices K and V. The matrix of outputs is calculated according to Equation (7).

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V, \tag{7}$$

where Q is the query matrix, K transposed is the key matrix, V is the value matrix, d_k is the dimension of the keys and queries in the attention mechanism, and $softmax$ is a function that scales values into probabilities.

In computer vision, the task of filtering an image can be considered as a transduction model, except that words must not be encoded, but in computer vision, 2D images are represented by arrays.

There are challenges when adapting these transformer structures from language to the context of vision, especially for the large variations in the scale of visual entities and the high resolution of pixels in images compared to words in text [40], but in [41], a methodology to split images into 16×16 patches for encoding purposes was successfully proposed.

The SWIN transformer consists of a hierarchical transformer whose representation is computed with shifted windows, with greater efficiency by limiting self-attention computation to non-overlapping local windows while also allowing for cross-window connection [40]. This structure is intended to be a general-purpose backbone for computer vision. According to its results, the SWIN outperforms the results of previously documented vision transformers (ViT) in applications such as image classification, object detection, and semantic segmentation.

The Restormer architecture proposed in [42] makes several redesigns in the multi-head attention blocks and feed-forward network of well-known transformers, remaining applicable to large images while achieving state-of-the-art results in deblurring and denoising applications.

The U-former architecture proposed in [43] uses a U-shaped structure inspired by [44]. This architectural structure demonstrates proven effectiveness, as the subsequent layers integrate information from preceding layers, enabling the synthesis of both fundamental and advanced learning representations inherent to the earlier layers. Besides the U-shaped structure, this transformer includes multi-head attention mechanisms at two levels: one multi-head self-attention module to calculate the attention map along both time and frequency-axis to generate time and frequency sub-attention maps for leveraging global interactions between encoder features, and another multi-head cross-attention mod-

ule inserted in the skip connections to allow a fine recovery in the decoder by filtering out uncorrelated features.

2.3. Metrics

In addition to the visual inspection of despeckled images, we also require the quantitative assessment of the models used in this study. For this purpose, well-known metrics in the SAR field are considered: Equivalent Number of Looks (ENL), Mean Squared Error (MSE), the Structural Similarity Index (SSIM), and Peak Signal-to-Noise Ratio (PSNR).

2.3.1. Equivalent Number of Looks

The Equivalent Number of Looks (ENL) is a measure used to evaluate the amount of speckle noise present in an SAR image. It provides an indication of how much the noise has been reduced by a despeckling filter. A higher ENL value means that the image has less noise and appears smoother, making it easier to analyze. ENL is calculated by comparing the average brightness of a uniform area in the image to the variability within that area. If the variability is low, it means the noise has been effectively reduced. This metric is especially useful when assessing the performance of different despeckling techniques, helping researchers determine which method provides the best balance between noise reduction and detail preservation.

The ENL serves as an indicator of noise levels present in an image. A higher ENL value corresponds to a reduced level of noise in the image. The ENL is quantitatively determined by taking the square of the mean pixel intensity within a homogeneous region and dividing it by the square of the standard deviation of the pixel intensities in the same region. It is calculated according to Equation (8).

$$ENL = \frac{\mu^2}{\sigma^2}. \tag{8}$$

2.3.2. Mean Squared Error

Mean Squared Error (MSE) is a commonly used metric to evaluate the quality of despeckling in SAR images by measuring the difference between the original and filtered images. It calculates the average squared difference between corresponding pixels in the two images. A lower MSE value indicates that the despeckled image is very similar to the original, meaning that noise has been effectively reduced while preserving important details. Since MSE emphasizes larger errors more than smaller ones, it helps identify areas where the despeckling process might have caused significant changes. This metric is useful for comparing the performance of different despeckling techniques and ensuring that important image features remain intact.

The MSE is calculated from the averaged squared errors in the pixels of two images and is calculated according to Equation (9). Values close to zero indicate less error.

$$MSE(x,y) = \frac{1}{N}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}(x(i,j) - y(i,j))^2. \tag{9}$$

2.3.3. Structural Similarity Index

The Structural Similarity Index (SSIM) is a metric used to measure the similarity between two images by considering their structural information, contrast, and brightness. Unlike simpler metrics like MSE, SSIM focuses on how well the structural details of the image are preserved after despeckling. It provides a value between 0 and 1, where values closer to 1 indicate that the despeckled image closely resembles the original in terms of important visual features. SSIM is particularly useful because it takes into account how the

human eye perceives images, making it a valuable tool for assessing the effectiveness of despeckling techniques in preserving meaningful patterns and textures.

The SSIM is a global measurement of the structural similarity between two images by using the luminance, the contrast, and the structure. The SSIM index is calculated according to Equation (10). Values close to 1 indicate higher similarity.

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + c_1)\cdot(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)\cdot(\sigma_x^2 + \sigma_y^2 + c_2)}, \tag{10}$$

where μ_x and μ_y are the average pixel values of images x and y, respectively, σ_x and σ_y are the standard deviations of x and y, σ_{xy} is the covariance between x and y, and c_1 and c_2 are constants to stabilize the division.

2.3.4. Peak Signal-to-Noise Ratio

Peak Signal-to-Noise Ratio (PSNR) is a widely used metric to evaluate the quality of despeckled SAR images by measuring how much noise remains in the image compared to its original, noise-free version. It is expressed in decibels (dB), with higher values indicating better image quality and lower noise levels. PSNR compares the maximum possible pixel intensity in the image to the average difference between the original and despeckled images. A high PSNR value means that the despeckling process has been effective in removing noise while preserving important image details. This metric is commonly used because it provides a simple and objective way to assess the performance of despeckling filters across different images.

The PSNR is a measurement of the ratio between the maximum possible power and the power of the noise in a signal. Higher values of PSNR indicate a better filtering process and it is calculated according to Equation (11).

$$PSNR = 10 \cdot \log_{10}\left(\frac{MAX^2}{MSE}\right), \tag{11}$$

where MAX is the maximum possible pixel value of the images. In our case, $MAX = 255$.

2.3.5. Pratt's Figure of Merit

The Pratt's Figure of Merit (PFOM) serves as a quantitative assessment of edge detection accuracy, achieved by analyzing the correspondence between actual edges and those extracted from two distinct images, alongside the spatial distances between the identified edges [45]. The metric yields values ranging from 0 to 1, with 0 indicating minimal edge similarity and 1 denoting complete congruence between the edges. The PFOM is calculated according to Equation (12).

$$PFOM = \frac{1}{max(I_l, I_A))}\sum_{i=1}^{I_A}\frac{1}{1 + \alpha_f \times d^2(i)}, \tag{12}$$

where I_A and I_l determine points of real edges and extracted edges, respectively, $d(i)$ is the distance between the i-th point on the image with estimated edges and nearest edge point on the real image, and α_f is a constant fixed at $1/9$.

2.3.6. Ratio Analysis

Ratio analysis is a simple yet effective way to evaluate the performance of despeckling filters in SAR images. It involves comparing the despeckled image to a reference image by calculating their ratio, which helps in identifying how much noise remains after filtering. Ideally, if a despeckling method works perfectly, the ratio should only contain noise without any patterns or structures from the original image. This technique allows researchers to

visually and quantitatively assess whether the filtering process has successfully removed speckle while preserving important details. By analyzing ratio images, it becomes easier to detect any remaining noise and fine-tune the despeckling methods for better results.

The analysis of ratio images has been a recent approach to design filters and assess the quality of a despeckling process; some examples are [23,46–48].

A filtered SAR image $\hat{X}$ will resemble its noise-free reference X. In the case of an ideal filter, $\hat{X} = X$. However, after the despeckling process, some of the speckle will remain in the filtered image. Solving from Equation (1), it is possible to obtain the noise N as a ratio expression of the ground truth and the SAR image, according to Equation (13).

$$N = \frac{Y}{X} \tag{13}$$

The analysis of the resulting ratio image N is a novel technique to assess the quality of the filters. In the ideal case, N should contain pure speckle and lack of structures or artifacts. The Shannon entropy or the Jensen–Shannon divergence [49,50] can be used to quantitatively analyze this remaining speckle, as proposed in [48].

2.4. Dataset with Actual SAR Imagery

The dataset used in this study was proposed in [2], specially designed with actual SAR imagery from Sentinel-1, obtained with the C band at 5.4 GHz with a resolution of 10 m at a height of 693 km from Ecuador. The main focus of the dataset is to train supervised machine learning models, including deep learning approaches, since it has pairs of images of the same location, properly rescaled and coregistered before the multitemporal fusion performed using a mean operation. This mean operation performs an averaged fusion of ten images into a single one, generating a significant increase in the Equivalent Number of Looks, departing from a value of $ENL = 16.490$ in the noisy images, while the resulting fusion (generated ground truth) resulted in $ENL = 70.605$. These values were calculated by using Equation (8).

This increase shows that the multitemporal mean operation decreased the level of speckle present in the ground truth images, so this is a quantitative reference for assessing the despeckling filters in this study. The models applied on SAR images should increase the ENL at least, or preferably, over the value calculated for the ground truth reference. These values described above must be considered only for reference, since they must be calculated for every single image used in this study by selecting a homogeneous region through a visual inspection.

Some samples of these pairs of labeled data obtained from the dataset are shown in Figure 2, where some regions of interest to highlight and compare the speckle present in the images are zoomed in red bounding boxes.

The dataset utilized in this study comprises a diverse collection of images representing various types of land cover. As detailed in [2], these images encompass regions characterized by distinct features such as water bodies, shorelines, rural landscapes, and densely populated urban areas. The urban scenes further include a variety of man-made structures, such as bridges, buildings, highways, and other infrastructural elements. The selected region from which these images were obtained exhibits a range of significant and unique attributes that contribute to the dataset's high degree of heterogeneity and diversity. These characteristics are crucial for ensuring the robustness and generalizability of the models trained and evaluated using this dataset.

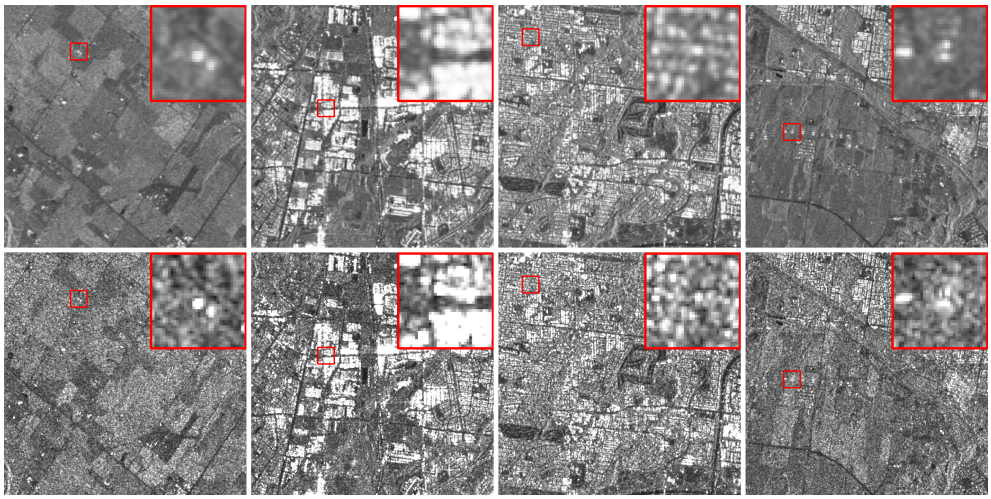


Figure 2. Four samples of images from the dataset. Top: Denoised reference images (ground truth). Bottom: Actual SAR with speckle (noisy). Zoom of regions of interest in red bounding boxes.

The use of a dataset derived from actual SAR imagery represents a significant step forward in ensuring the relevance and applicability of despeckling models in real-world scenarios. By leveraging multitemporal fusion to generate ground truth references, this dataset captures the inherent speckle patterns and complexities of SAR data, which are often oversimplified in synthetic datasets. This realistic training and evaluation framework not only enhances the reliability of model performance but also bridges the gap between experimental results and operational requirements.

3. Experimental Results

The experimental analysis aims to evaluate the performance of various despeckling techniques, including traditional, deep learning, and transformer-based models. This section presents the outcomes of visual inspections and quantitative assessments, highlighting the strengths and limitations of each method across a diverse set of SAR imagery samples.

3.1. Despeckling

The despeckling process was performed on the noisy (actual SAR) images selected from the dataset, as shown in Figure 2. These samples were despeckled with the Lee filter, FANS, and the AE. Then, the GenAI capabilities were added to an autoencoder and U-Net structures (namely GenAE, GenU-Net), and finally, a U-former and a Restormer were applied to the images.

The autoencoder with a U-Net structure was implemented to enhance despeckling performance by leveraging its encoder–decoder architecture with skip connections. The encoder consists of four convolutional layers with filter sizes of 3×3, each followed by batch normalization and ReLU activation. The number of filters in the encoder layers follows a sequence of 16, 32, 64, and 128, with max-pooling applied after each convolution. The decoder mirrors this structure, using transposed convolution layers to restore the original image dimensions. Skip connections are employed to retain fine details from the encoder. The model was trained using the Adam optimizer with a learning rate of 0.0001, a batch size of 8, and a binary cross-entropy (BCE) loss function. Training was conducted for 100 epochs with early stopping based on validation loss to prevent overfitting. For the transformer-based models, U-Former and Restormer, we used their standard configurations with minor adjustments to optimize for SAR despeckling tasks. The U-Former model was trained using a learning rate of 0.0001, a batch size of 8, and the AdamW optimizer, leveraging hierarchical multi-head self-attention layers for enhanced feature extraction.

Similarly, the Restormer model was trained with a learning rate of 0.0001, a batch size of 8, and a patch size of 16 × 16, incorporating channel attention and long-range dependency mechanisms to improve noise suppression. Both models were trained for 50 epochs with an early stopping criterion to ensure convergence.

The models were trained on a workstation equipped with a Ryzen 9 7900X CPU, 64 GB DDR5 RAM, and RTX4070 GPU. Training the AE, GenAE, and GenU-Net models involved minimal computational cost using a GPU, requiring less than 1 h to reach an optimal solution. In contrast, the U-former and Restormer models exhibited significantly higher computational demands, taking approximately 50 and 22 h to converge to the desired outcomes, respectively. This disparity highlights the need for advanced processing units or cloud-based infrastructure with high-performance specifications when utilizing these advanced and complex models. On the other hand, once trained, the application of these models to new images becomes highly efficient, with the despeckling of a single image completed in just a few seconds. The visual results after applying all the aforementioned despeckling models are illustrated in Figure 3.

The despeckling process showcased a wide range of performance across the filters, with each technique exhibiting distinct trade-offs between noise reduction and detail preservation. Traditional methods, such as the Lee filter, displayed strong edge preservation but struggled with residual speckle in homogeneous regions. Non-local approaches, like FANS, effectively reduced noise in smooth areas but occasionally introduced artifacts and weakened edge clarity. Deep learning-based models, particularly the autoencoder, demonstrated their capacity to achieve homogeneous despeckling while maintaining critical structural elements. The introduction of generative and transformer-based methods further elevated performance, delivering cleaner images with minimal speckle and enhanced feature retention. The visual results highlighted the superior adaptability of these advanced models, particularly in challenging scenarios where both fine texture and high-contrast edges were present. These findings underscore the importance of aligning despeckling techniques with the specific requirements of the application, whether prioritizing noise reduction, structural fidelity, or computational efficiency.

As an additional analysis for the despeckling process, applying Equation (13) with the filtered and the SAR images, the ratio images N were obtained, as shown in Figure 4.

The results obtained from the despeckling process reveal significant differences in the performance of the various filtering techniques, highlighting the trade-offs between noise reduction and detail preservation. Traditional filters such as the Lee and FANS filters demonstrated their well-known strengths in preserving edges and computational efficiency; however, they struggled to effectively remove speckle noise in highly textured or heterogeneous regions. Deep learning-based methods, particularly the autoencoder (AE) and its generative variant (GenAE), showed improved noise suppression while maintaining a good level of structural integrity, yet some over-smoothing was observed in complex areas. The generative models introduced additional learning capabilities, leveraging adversarial training to produce despeckled images that visually appear more natural, but their quantitative performance metrics indicated occasional artifacts and instability in homogeneous regions. In contrast, the transformer-based approaches, specifically U-Former and Restormer, exhibited superior performance across all evaluated aspects, achieving a fine balance between despeckling effectiveness and structural fidelity. Their ability to model long-range dependencies allowed for more precise noise reduction while maintaining critical details such as edges and textures, making them particularly suitable for applications requiring high-precision analysis. Despite their higher computational demands during training, these models provided consistent and reliable results once deployed, indicating their potential for operational use in real-world SAR applications. The findings emphasize

the importance of selecting an appropriate despeckling method based on the specific requirements of the application, such as computational constraints, desired level of detail retention, and target operational environment.

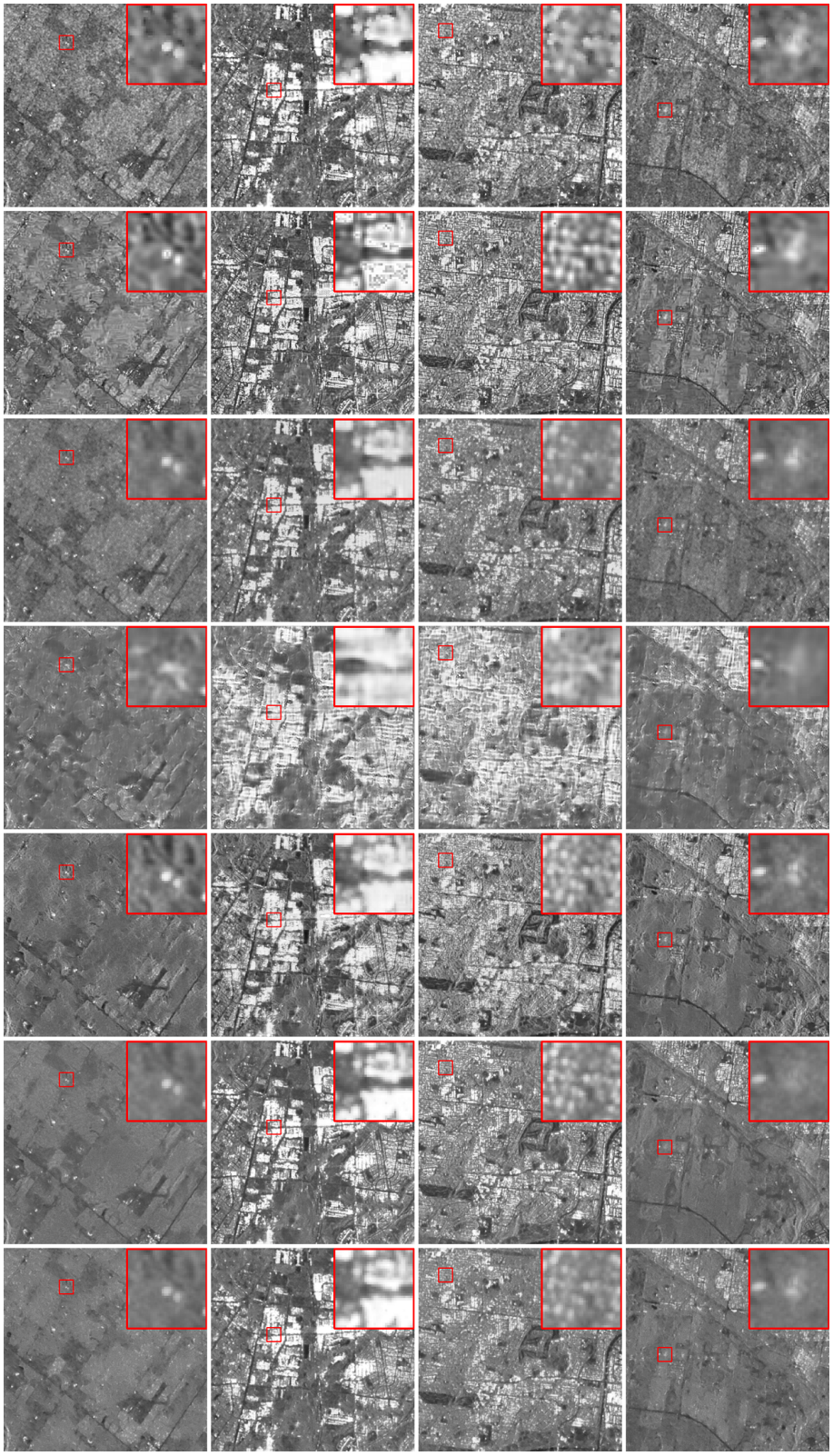


Figure 3. Four samples despeckled with different filters. From left to right: four samples of different locations. From top to bottom: Lee, FANS, AE, GenAE, GenU-Net, U-former, and Restormer. Zoom of regions of interest in red bounding boxes.

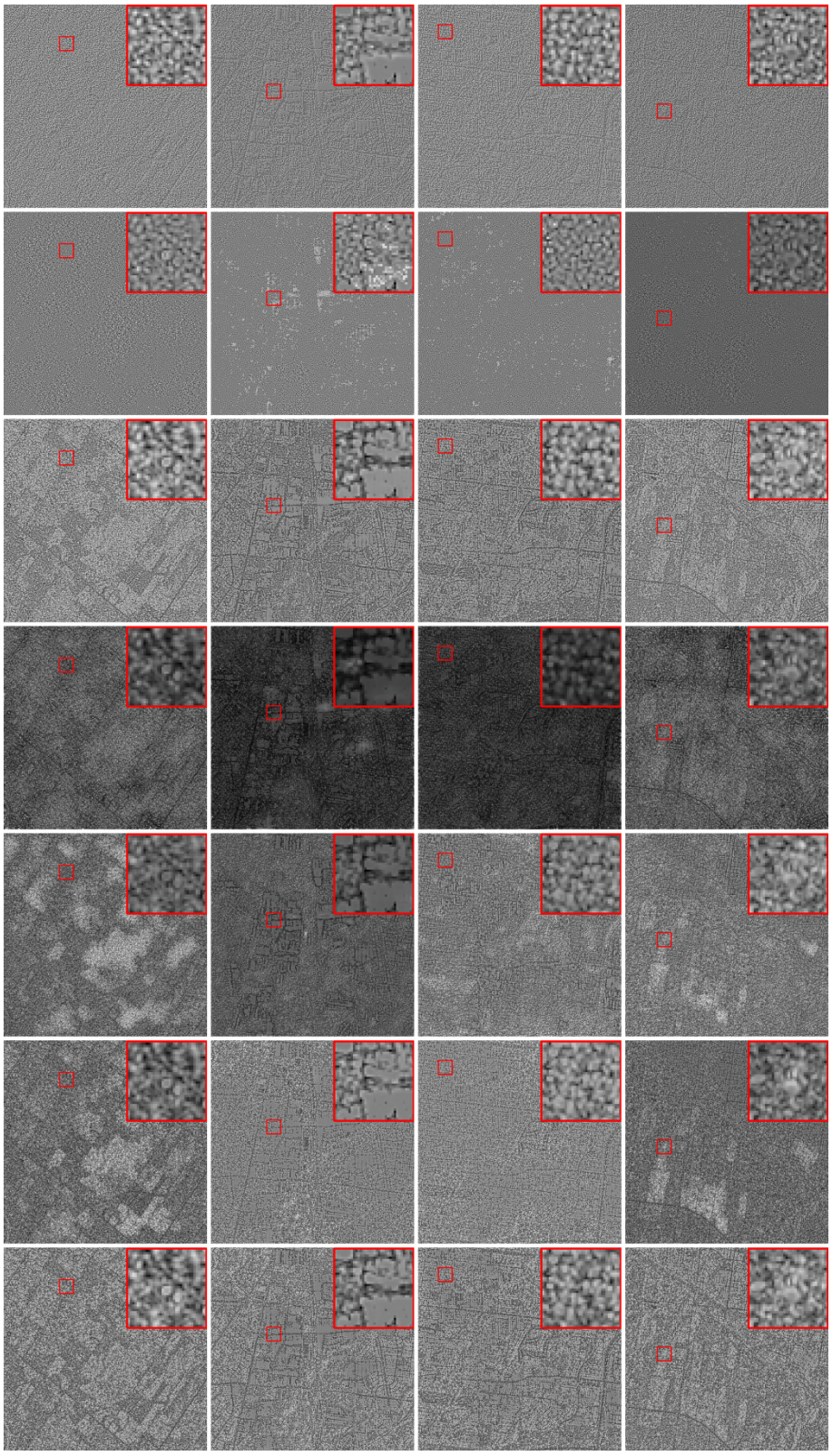


Figure 4. Ratio images of four samples despeckled with different filters. From left to right: four samples of different locations. From top to bottom: Lee, FANS, AE, GenAE, GenU-Net, U-former, and Restormer. Zoom of regions of interest in red bounding boxes.

3.2. Quantitative Assessment

The quantitative assessment was performed by calculating the ENL of all the images, including ground truth and actual SAR. Then, the other metrics, namely MSE, SSIM, PSNR,

and PFOM were calculated on noisy and despeckled images concerning the ground truth reference. The results of these measurements are shown in Table 1.

Table 1. Quantitative assessment of four different samples despeckled. The best metric results are in boldface. The second position results are underlined.

Validation Sample	Image	ENL	MSE	SSIM	PSNR	PFOM
1	Ground truth	231.27	—	—	—	—
	Noisy	22.97	1053.15	0.34	17.91	0.70
	Filtered Lee	94.95	540.88	0.52	20.80	**0.83**
	Filtered FANS	231.13	541.29	0.56	20.80	0.80
	Filtered AE	383.99	254.83	0.64	24.07	0.82
	Filtered GenAE	337.71	368.75	0.49	22.46	0.70
	Filtered GenU-Net	270.14	320.45	0.62	23.07	0.77
	Filtered U-former	**855.29**	**189.49**	**0.67**	**25.35**	0.81
	Filtered Restormer	704.18	229.98	0.66	24.51	0.80
2	Ground truth	196.96	—	—	—	—
	Noisy	18.95	893.34	0.61	18.62	0.73
	Filtered Lee	51.44	458.51	0.72	21.52	0.86
	Filtered FANS	71.74	571.22	0.70	20.56	0.79
	Filtered AE	81.84	424.64	0.71	21.85	0.83
	Filtered GenAE	108.10	1930.63	0.39	15.27	0.81
	Filtered GenU-Net	111.58	497.36	0.71	21.16	0.83
	Filtered U-former	**230.52**	**347.99**	**0.77**	**22.72**	**0.87**
	Filtered Restormer	131.72	381.79	0.74	22.31	0.83
3	Ground truth	98.86	—	—	—	—
	Noisy	13.72	847.47	0.65	18.85	0.80
	Filtered Lee	34.66	436.66	0.71	21.73	**0.90**
	Filtered FANS	31.64	493.80	0.74	21.20	0.85
	Filtered AE	87.60	422.93	0.69	21.87	0.82
	Filtered GenAE	**401.66**	2044.12	0.35	15.03	0.82
	Filtered GenU-Net	74.00	443.50	0.73	21.66	0.89
	Filtered U-former	106.45	**325.49**	**0.76**	**23.01**	0.87
	Filtered Restormer	94.99	401.22	0.71	22.10	0.81
4	Ground truth	319.97	—	—	—	—
	Noisy	19.12	850.93	0.49	18.83	0.66
	Filtered Lee	83.10	439.55	0.61	21.70	0.78
	Filtered FANS	138.26	455.29	0.65	21.55	0.82
	Filtered AE	263.44	318.86	0.65	23.09	0.82
	Filtered GenAE	**1807.24**	1065.46	0.43	17.86	0.75
	Filtered GenU-Net	345.85	304.66	0.67	23.29	**0.84**
	Filtered U-former	649.79	**247.51**	**0.69**	**24.19**	0.81
	Filtered Restormer	677.21	327.77	0.65	22.98	0.81

This table presents a comprehensive comparison of despeckling performance across various filters and models, evaluated using key metrics such as ENL, MSE, SSIM, PSNR, and PFOM. The ENL indicates significant improvements in noise reduction for most advanced models, particularly the U-former and Restormer, which consistently achieve

the highest values, signifying superior noise attenuation. Meanwhile, the MSE and PSNR metrics further validate the effectiveness of these models in closely approximating the ground truth, with lower MSE and higher PSNR scores indicating better fidelity to the reference data. The SSIM highlights the ability of each filter to preserve structural details, where transformer-based models outperform traditional filters and generative models. Lastly, the PFOM metric underscores the edge-preservation capabilities, with the Lee filter demonstrating competitive performance, albeit at the cost of higher residual noise. Together, these metrics provide a holistic view of each filter's strengths and limitations, underscoring the transformative impact of modern architectures like transformers in SAR despeckling tasks.

The quantitative assessment of the despeckling filters provided valuable insights into their effectiveness in balancing noise reduction and structural preservation. The results showed that traditional methods, such as the Lee and FANS filters, achieved moderate improvements in terms of noise suppression, as indicated by increased ENL values. However, their MSE and SSIM scores highlighted their limitations in preserving fine details and textures, which are crucial for accurate SAR image interpretation. Deep learning-based approaches, particularly the autoencoder (AE), demonstrated a significant enhancement in quantitative metrics, achieving lower MSE and higher SSIM and PSNR values compared to traditional methods. The GenAE and GenU-Net models showed further improvements in PSNR and structural preservation but at the cost of occasional over-smoothing, which could impact the accuracy of fine-feature analysis. On the other hand, transformer-based models, such as U-Former and Restormer, consistently outperformed all other approaches across all metrics, exhibiting the highest ENL values, lowest MSE, and superior SSIM and PSNR scores. Their ability to capture both local and global features allowed for efficient noise reduction without compromising edge definition, making them highly suitable for operational applications that require high-fidelity despeckling. PFOM results further confirmed the effectiveness of these models in maintaining edge sharpness, which is crucial for applications such as terrain classification and change detection. The quantitative findings underscore the importance of selecting a despeckling technique that aligns with the specific requirements of SAR-based applications, whether prioritizing noise reduction, detail preservation, or computational efficiency.

4. Discussion

The discussion synthesizes the key findings from the experimental results, emphasizing the implications of the visual and quantitative analyses. By critically examining the trade-offs between noise reduction and detail preservation, this section provides insights into the relative performance of each despeckling approach and its potential applications in SAR image analysis.

4.1. Visual Inspection

The reference images from the dataset shown in Figure 2 exhibit a first row of ground truth images with a low presence of speckle since the bottom row exhibits a considerable speckle level (noise). An ideal despeckling filter would take as input the noisy image and exactly replicate the ground truth reference.

The visual results in Figure 3 highlight key differences in the performance of despeckling filters, emphasizing the trade-offs between noise reduction and detail preservation. Traditional approaches like the Lee filter, while successful in maintaining edge sharpness, fail to sufficiently attenuate speckle, leaving significant noise artifacts in homogeneous regions. This limitation underscores the inherent constraints of local filtering methods, which rely heavily on statistical assumptions within small windows and struggle to adapt

to complex speckle patterns found in real-world SAR imagery. Similarly, the FANS filter, despite leveraging non-local characteristics, introduces visual artifacts that disrupt the natural texture of the scene and degrade structural continuity, particularly in regions with intricate features.

Deep learning-based models provide a marked improvement, as seen in the results from the autoencoder (AE). The AE demonstrates an ability to homogenize regions effectively, significantly reducing speckle while maintaining a degree of structural integrity. However, it generates a degree of over-smoothing, which compromises the sharpness of edges and fine details critical for SAR interpretation. The generative autoencoder (GenAE) builds on this foundation by preserving more details and edges yet exhibits inconsistencies in homogeneous regions where some residual speckle remains visible. These results suggest that while deep learning can outperform traditional filters, there is still a need to balance homogenization with the retention of critical features.

The transformer-based models, including U-former and Restormer, deliver the most promising results, with their outputs closely resembling the ground truth reference. These models excel in attenuating speckle without compromising the natural textures and sharp edges within the image. The ability of transformers to capture long-range dependencies and adaptively allocate attention across different regions of the image enables them to achieve this superior balance. This performance indicates the transformative potential of attention-based mechanisms in SAR despeckling, particularly for applications requiring high fidelity in structural and textural information. The visual superiority of these models is further supported by their consistent results across diverse image samples, reinforcing their robustness and reliability.

The ratio images in Figure 4 show that in all the cases they preserve structure and patterns that should not be present in the case of an ideal despeckling process. The FANS filter is probably the one with less structure in these ratio images, while the GenAE filter generates darker ratio images with white blurry patches. The ratio images of the AE, the U-former and the Restormer are similar in values and structure.

Visual inspection of despeckled and ratio images not only provides qualitative insights into the performance of different filters but also highlights their strengths and limitations in preserving critical features such as edges, textures, and fine details. While some filters excel at reducing speckle noise, they may inadvertently smooth out essential structures, which can compromise the interpretability of SAR imagery for downstream applications. The inclusion of zoomed-in regions of interest further emphasizes the subtle differences in filter performance, particularly in challenging areas such as high-contrast boundaries or textured regions. This qualitative analysis complements quantitative metrics, offering a comprehensive evaluation of each filter's ability to balance noise attenuation with feature preservation, ultimately guiding the selection of optimal despeckling techniques for specific use cases.

4.2. Quantitative Analysis

The measurement of the metrics selected in this study applied on the four samples shown in Table 1 demonstrate a high superiority of the U-former model concerning the other despeckling filters. This model outperformed and obtained the first position in most of the measurements in the four SAR samples. The Restormer model also exhibited high performance, obtaining the second position in several of the measurements.

In some particular cases, like samples three and four, the ENL was outperformed when GenAE was applied. However, this could be caused by over-filtering in some particular regions, generating high values of this number of looks but smoothing important edge

information. This hypothesis is confirmed by its low performance in other metrics like SSIM, PSNR, and PFOM.

The edge preservation information, mainly measured by the PFOM metric, indicates that the best results were obtained by the Lee filter. However, this metric cannot be used as a quality indicator of the filtering process by itself, since the visual inspection and other metrics indicate that the remaining speckle of this filter was high, which could explain the edge preservation results.

The quantitative metrics in Table 1 provide a multi-dimensional assessment of despeckling performance, highlighting the distinct strengths and weaknesses of each filter. The ENL reveals that transformer-based models, particularly the U-former and Restormer, achieve the highest noise reduction, as evidenced by their significantly higher ENL values compared to other methods. This suggests superior homogenization of uniform regions without introducing substantial artifacts. However, a closer examination of the MSE indicates that while the GenAE occasionally outperformed other methods in specific samples, its performance was inconsistent across the dataset. This inconsistency, paired with its relatively low SSIM scores, implies a trade-off where over-filtering leads to loss of fine details and edge sharpness, even as it achieves noise suppression.

Metrics such as SSIM and PSNR further underscore the comprehensive effectiveness of transformer-based architectures. These models consistently achieved the highest SSIM values, indicating superior preservation of structural and textural features. The PSNR results align with these findings, showcasing their ability to produce despeckled images with minimal distortion relative to the ground truth. Interestingly, PFOM, a metric focused on edge preservation, highlights the strong performance of the Lee filter in maintaining sharp edges. However, when considered alongside other metrics, the Lee filter's high PFOM values are likely due to residual speckle accentuating edge intensity rather than accurate despeckling. This interplay between metrics demonstrates the need for a balanced approach that minimizes noise while preserving critical structural information, a challenge effectively addressed by advanced models like U-former and Restormer.

The results presented in Table 1 provide crucial insights into the effectiveness of various despeckling techniques, highlighting their strengths and limitations across different quantitative metrics. The performance indicators, such as ENL, MSE, SSIM, PSNR, and PFOM, offer a comprehensive evaluation of each model's capability to reduce speckle noise while preserving critical image details. The superior performance of transformer-based models, particularly U-Former and Restormer, underscores their potential to set new benchmarks in SAR despeckling by achieving a balanced trade-off between noise suppression and structural fidelity. These results are of significant importance for advancing AI-based despeckling models, as they provide empirical evidence that can guide the refinement of existing architectures and inspire the development of new hybrid approaches. By analyzing the trends observed in the results, researchers can identify areas for improvement, such as enhancing the generalization capability of generative models, optimizing computational efficiency, and addressing challenges related to over-smoothing or loss of fine details. Ultimately, the insights gained from this study contribute to the ongoing evolution of artificial intelligence in remote sensing, paving the way for more accurate and reliable SAR image analysis that can support critical applications in environmental monitoring, disaster response, and land use planning.

4.3. Integrated Analysis of Visual and Quantitative Results

The results obtained from both visual inspection and quantitative metrics provide complementary insights into the performance of despeckling filters. Visual inspection revealed that traditional methods, such as the Lee and FANS filters, often leave residual

speckle or introduce artifacts, limiting their applicability in scenarios where high visual fidelity is critical. On the other hand, advanced models like the U-former and Restormer demonstrated a remarkable ability to balance speckle reduction with the preservation of edge sharpness and fine textures, closely approximating the visual quality of ground truth images.

From a quantitative perspective, the superiority of transformer-based models is evident in their consistently high performance across all metrics, particularly ENL and SSIM. These results indicate that these models not only reduce noise effectively but also maintain the structural integrity of the images. However, the ENL values for certain models, such as the GenAE, were sometimes inflated due to over-smoothing, which diminished their ability to retain critical edge details. This trade-off underscores the importance of interpreting quantitative metrics alongside visual results to fully assess filter performance.

Interestingly, the PSNR values highlighted the capability of the U-former and Restormer to minimize distortion relative to the ground truth. This metric, coupled with Pratt's Figure of Merit (PFOM), underscores the edge-preservation abilities of these models. While the Lee filter achieved high PFOM scores, visual inspection confirmed that this was primarily due to the retention of residual speckle, rather than true edge accuracy. This finding emphasizes the need for multi-metric evaluations to capture the nuanced trade-offs between noise reduction and detail preservation.

Overall, the combined visual and quantitative analysis underscores the transformative potential of transformer-based models in SAR despeckling tasks. Their ability to adaptively focus on both homogeneous and textured regions enables them to address the limitations of traditional and earlier deep learning approaches. Future research should delve deeper into optimizing these models to further enhance their generalization capabilities across diverse SAR datasets and imaging conditions.

4.4. Impact and Practical Applications

The experimental findings underline the transformative potential of advanced despeckling techniques, particularly transformer-based models, for real-world applications in SAR imagery analysis. The superior noise attenuation and structural preservation achieved by these models can significantly enhance the interpretability of SAR data in critical domains such as environmental monitoring, urban planning, and disaster management. For instance, the ability of models like U-former and Restormer to maintain edge clarity and fine textures is crucial for detecting subtle changes in land cover or infrastructure, which are often masked by speckle noise in traditional SAR analyses.

Furthermore, the quantitative superiority of transformer-based architectures, reflected in metrics such as ENL and SSIM, positions them as reliable tools for automated SAR workflows. In applications involving large-scale or multitemporal datasets, these models can reduce the need for extensive manual post-processing, accelerating decision-making processes. However, the computational demands of such advanced techniques must also be considered, particularly in resource-constrained scenarios, where lightweight models like autoencoders may offer a viable alternative despite their limitations in preserving fine details.

This study also highlights the importance of training models with datasets that closely mimic real-world speckle patterns, as opposed to synthetic datasets that may not accurately represent operational conditions. By leveraging a dataset derived from actual SAR imagery, the models demonstrated robust performance, reinforcing the need for realistic training data in developing reliable despeckling solutions. Future research should explore the integration of these techniques with operational SAR systems to validate their performance

in dynamic and unpredictable environments, further bridging the gap between theoretical advancements and practical utility.

The application of these GenAI models for despeckling in real-world SAR imagery demonstrates significant promise. These models have been trained to recognize and learn the authentic patterns and statistical characteristics of the speckle noise intrinsic to SAR data. As a result, when new SAR imagery is acquired from satellites, the models can efficiently execute the despeckling process. This not only enhances the quality of the data but also facilitates subsequent processing steps critical to specific applications, including classification tasks, change detection, multitemporal analysis, and other related operations.

5. Conclusions and Future Work

The use of the labeled dataset containing SAR images derived from actual speckle patterns proved critical for achieving robust and reliable despeckling performance. By training models on a dataset generated from real SAR imagery rather than synthetic speckle, this study demonstrated significant improvements in the model's ability to generalize to operational conditions, ensuring that noise reduction aligns with the true characteristics of SAR data. Future expansions of this dataset that include more diverse imaging conditions and environmental contexts will further strengthen its utility in advancing despeckling research.

The comparative analysis of despeckling filters highlighted the limitations of traditional methods, such as the Lee and FANS filters, in handling complex speckle patterns present in real-world SAR imagery. While these methods demonstrated strong edge preservation in quantitative metrics, visual inspection revealed significant residual speckle and artifacts, underscoring the need for advanced approaches like deep learning and transformer-based models.

The results of the transformer were consistent over all the metrics, showcasing their effectiveness in balancing noise reduction and detail preservation. Transformer-based models, such as U-former and Restormer, achieved the highest scores in key metrics like ENL, SSIM, and PSNR, demonstrating their capability to produce despeckled images that closely resemble ground truth while retaining structural and textural integrity.

The superiority of the transformers lies in their attention mechanisms, which enable enhanced edge preservation and feature extraction compared to traditional and other deep learning methods. Their ability to capture long-range dependencies and adapt to both homogeneous and textured regions has established transformers as a benchmark for SAR despeckling, particularly for high-resolution imagery requiring fine-detail fidelity.

Solely the addition of GenAI was not sufficient to surpass transformer performance, highlighting the need for synergistic approaches. While models like GenAE showed promise in reducing speckle, they often over-smoothed critical details, demonstrating that the integration of GenAI with attention-based frameworks may provide a more balanced and effective solution for future developments in SAR despeckling.

This study underscores the importance of multi-metric evaluations to comprehensively assess despeckling performance across diverse conditions. Metrics such as ENL, SSIM, PSNR, and PFOM, along with the ratio images, provided valuable insights, but their interpretation required alignment with visual inspection results to fully understand the trade-offs between noise reduction, edge preservation, and overall image fidelity.

The training process of GenAI models is inherently time-consuming and computationally demanding, which necessitates the use of advanced processing units or cloud-based infrastructure equipped with high-performance technical specifications. As a result, this stage can be particularly taxing, both in terms of time and computational expense. However, once these models are adequately trained, the process of applying them to new images be-

comes considerably more efficient, even for large images or datasets. This makes the models highly adaptable for iterative processing, allowing them to be applied to large volumes of new data without a significant increase in computational burden. Consequently, this efficiency facilitates the implementation of the models in real-world scenarios, including applications that require near-real-time processing.

Future Work

The dataset is crucial for the training and validation of the models, so expanding its diversity and incorporating additional scenarios is a relevant improvement. Although the current dataset based on multitemporal fusion offers an effective framework for training despeckling models, its scope can be broadened by including images with varying resolutions, acquisition geometries, and environmental contexts. This enhancement will allow the development of models that generalize better across diverse operational scenarios, particularly for applications requiring high adaptability and robustness.

The implementation of transformers and more complex and conceptual structures can further improve the performance of despeckling filters, particularly in challenging environments. Future studies could explore the integration of hybrid architectures that combine the attention mechanisms of transformers with the feature extraction capabilities of convolutional networks. Additionally, leveraging self-supervised or unsupervised learning frameworks may mitigate the dependency on labeled datasets, addressing one of the primary limitations of supervised learning in SAR despeckling.

Advancing the interpretability and computational efficiency of despeckling models remains a key area for exploration. Future research should optimize transformer-based models to reduce their computational demands, making them more feasible for real-time applications and resource-constrained environments. Moreover, explainable AI techniques can be incorporated to better understand these models' decision-making processes, ensuring their reliability and trustworthiness in critical applications like environmental monitoring and disaster management.

In addition to expanding the dataset, future research should focus on evaluating the performance of despeckling filters across diverse environmental conditions and categorized surface zones. Applying these models to different geographical regions, such as urban areas, agricultural fields, forests, and coastal environments, will provide valuable insights into their adaptability and robustness. Investigating how environmental factors, including seasonal variations, terrain complexity, and land cover heterogeneity, influence despeckling effectiveness will enhance the practical applicability of these models for real-world SAR image analysis. This approach can help optimize the filters to better accommodate the specific characteristics and challenges posed by various Earth surface categories.

Author Contributions: A.A.C.-M.: Conceptualization, Methodology, Writing—Original Draft; R.D.V.-S.: Methodology, Writing—Original Draft; C.M.T.-G.: Conceptualization, Methodology, Formal Analysis, Writing—Review and Editing; L.G.: Conceptualization, Methodology, Formal Analysis, Writing—Review and Editing. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Politécnico Colombiano Jaime Isaza Cadavid (Colombia) through the project No. 2024-00066-001 titled "Detección de cambios en cobertura vegetal y cuerpos de agua por medio de procesamiento de imágenes de radar de apertura sintética (SAR) con técnicas de inteligencia artificial para el análisis de implicaciones ambientales en la región del bajo Cauca Antioqueño".

Data Availability Statement: All scripts and images used in this paper will be provided on demand.

Acknowledgments: Many thanks to Politécnico Colombiano Jaime Isaza Cadavid (Colombia) for funding the project and to the University of Las Palmas de Gran Canaria (Spain) for its support of the project as a co-executor. Special thanks to the platforms Google Earth Engine (GEE), ASF Data Search Vertex of the University of Alaska, and the European Space Agency (ESA) for the open-access availability Sentinel-1 SAR images used in this study.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

SAR	Synthetic-Aperture Radar
MSE	Mean Squared Error
SSIM	Structural Similarity Index
PSNR	Peak Signal-to-Noise Ratio
PFOM	Pratt's Figure of Merit
CNN	Convolutional Neural Network
DL	Deep Learning
GenAI	Generative Artificial Intelligence
AE	Autoencoder
NLP	Natural Language Processing
ViT	Vision Transformer
GAN	Generative Adversarial Network
GRD	Ground Range Detected
VV	Vertical Vertical
VH	Vertical Horizontal

References

1. Argenti, F.; Lapini, A.; Bianchi, T.; Alparone, L. A Tutorial on Speckle Reduction in Synthetic Aperture Radar Images. *IEEE Geosci. Remote Sens. Mag.* **2013**, *1*, 6–35. [CrossRef]
2. Vásquez-Salazar, R.D.; Cardona-Mesa, A.A.; Gómez, L.; Travieso-González, C.M.; Garavito-González, A.F.; Vásquez-Cano, E. Labeled dataset for training despeckling filters for SAR imagery. *Data Brief* **2024**, *53*, 110065. [CrossRef] [PubMed]
3. Vitale, S.; Ferraioli, G.; Pascazio, V.; Schirinzi, G. InSAR-MONet: Interferometric SAR phase denoising using a multiobjective neural network. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–14. [CrossRef]
4. Vitale, S.; Aghababaei, H.; Ferraioli, G.; Pascazio, V.; Schirinzi, G. A Multi-Objective Approach for Multi-Channel SAR Despeckling. In Proceedings of the 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS, Brussels, Belgium, 11–16 July 2021 ; pp. 419–422.
5. Vijay Kumar, S.; Sun, X.; Wang, Z.; Goldsbury, R.; Cheng, I. A U-net approach for insar phase unwrapping and denoising. *Remote Sens.* **2023**, *15*, 5081. [CrossRef]
6. Zhang, G.; Li, Z.; Li, X.; Liu, S. Self-supervised despeckling algorithm with an enhanced U-net for synthetic aperture radar images. *Remote Sens.* **2021**, *13*, 4383. [CrossRef]
7. Zhao, G.; Peng, Y. Semisupervised SAR image change detection based on a siamese variational autoencoder. *Inf. Process. Manag.* **2022**, *59*, 102726. [CrossRef]
8. Frontera-Pons, J.; Brigui, F.; de Milly, X. Unsupervised SAR change detection with despeckling autoencoders. In Proceedings of the 2023 IEEE International Radar Conference (RADAR), Sydney, Australia, 6–10 November 2023; pp. 1–6.
9. Qi, H.; Tan, S.; Li, Z. Anisotropic weighted total variation feature fusion network for remote sensing image denoising. *Remote Sens.* **2022**, *14*, 6300. [CrossRef]
10. Singh, P.; Shankar, A. A novel optical image denoising technique using convolutional neural network and anisotropic diffusion for real-time surveillance applications. *J. Real Time Image Process.* **2021**, *18*, 1711–1728. [CrossRef]
11. Araujo, G.F.; Machado, R.; Pettersson, M.I. Synthetic SAR data generator using Pix2pix cGAN architecture for automatic target recognition. *IEEE Access* **2023**, *11*, 143369–143386. [CrossRef]
12. Xu, X.; Fan, W.; Wang, S.; Zhou, F. WBIM-GAN: A Generative Adversarial Network Based Wideband Interference Mitigation Model for Synthetic Aperture Radar. *Remote Sens.* **2024**, *16*, 910. [CrossRef]
13. Mao, C.; Huang, L.; Xiao, Y.; He, F.; Liu, Y. Target recognition of SAR image based on CN-GAN and CNN in complex environment. *IEEE Access* **2021**, *9*, 39608–39617. [CrossRef]

14. Wang, C.; Zheng, R.; Zhu, J.; Xu, W.; Li, X. A Practical SAR Despeckling Method Combining SWIN Transformer and Residual CNN. *IEEE Geosci. Remote Sens. Lett.* **2023**, *21*, 4001205. [CrossRef]
15. Huang, M.; Luo, S.; Wang, S.; Guo, J.; Wang, J. HTCNet: Hybrid Transformer-CNN for SAR Image Denoising. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2024**, *17*, 19380–19394. [CrossRef]
16. Perera, M.V.; Bandara, W.G.C.; Valanarasu, J.M.J.; Patel, V.M. Transformer-based SAR image despeckling. In Proceedings of the IGARSS 2022—2022 IEEE International Geoscience and Remote Sensing Symposium, Kuala Lumpur, Malaysia, 17–22 July 2022; pp. 751–754.
17. Yan, H.; Fang, C.; Qiao, Z. A multi-attention Uformer for low-dose CT image denoising. *Signal Image Video Process.* **2024**, *18*, 1429–1442. [CrossRef]
18. Lee, J.S.; Jurkevich, L.; Dewaele, P.; Wambacq, P.; Oosterlinck, A. Speckle filtering of synthetic aperture radar images: A review. *Remote Sens. Rev.* **1994**, *8*, 313–340. [CrossRef]
19. Wang, C.; Zheng, R.; Zhu, J.W.; He, X.; Li, X. Unsupervised SAR Despeckling by Combining Online Speckle Generation and Unpaired Training. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2023**, *16*, 10175–10190. [CrossRef]
20. Vásquez-Salazar, R.D.; Cardona-Mesa, A.A.; Gómez, L.; Travieso-González, C.M. A New Methodology for Assessing SAR Despeckling Filters. *IEEE Geosci. Remote Sens. Lett.* **2024**, *21*, 1–5. [CrossRef]
21. Frery, A.; Wu, J.; Gomez, L. *SAR Image Analysis—A Computational Statistics Approach: With R Code, Data, and Applications*; Wiley-IEEE Press: New York, NY, USA , 2022; ISBN: 978-1-119-79546-9.
22. Li, J.; Wang, Z.; Yu, W.; Luo, Y.; Yu, Z. A Novel Speckle Suppression Method with Quantitative Combination of Total Variation and Anisotropic Diffusion PDE Model. *Remote Sens.* **2022**, *14*, 796. [CrossRef]
23. Ferraioli, G.; Pascazio, V.; Schirinzi, G. Ratio-Based Nonlocal Anisotropic Despeckling Approach for SAR Images. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 7785–7798. [CrossRef]
24. Vitale, S.; Ferraioli, G.; Pascazio, V. Multi-Objective CNN-Based Algorithm for SAR Despeckling. *IEEE Trans. Geosci. Remote Sens.* **2020**, *59*, 9336–9349. [CrossRef]
25. Wang, W.; Kang, Y.; Liu, G.; Wang, X. SCU-net: Semantic segmentation network for learning channel information on remote sensing images. *Comput. Intell. Neurosci.* **2022**, *2022*, 8469415. [CrossRef]
26. Parrilli, S.; Poderico, M.; Angelino, C.V.; Verdoliva, L. A Nonlocal SAR Image Denoising Algorithm Based on LLMMSE Wavelet Shrinkage. *IEEE Trans. Geosci. Remote Sens.* **2012**, *50*, 606–616. [CrossRef]
27. Goodfellow, I.J.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.C.; Bengio, Y. Generative Adversarial Nets. In Proceedings of the Neural Information Processing Systems, Montreal, QC, Canada, 8–11 December 2014.
28. Vaswani, A.; Shazeer, N.M.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention is All you Need. In Proceedings of the Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017.
29. Lee, J.S. Digital image enhancement and noise filtering by use of local statistics. *IEEE Trans. Pattern Anal. Mach. Intell.* **1980**, *1980*, 165–168. [CrossRef]
30. Salehi, H.; Vahidi, J. A novel hybrid filter for image despeckling based on improved adaptive wiener filter, bilateral filter and wavelet filter. *Int. J. Image Graph.* **2021**, *21*, 2150036. [CrossRef]
31. Buades, A.; Coll, B.; Morel, J.M. Non-Local Means Denoising. *Image Process. Online* **2011**, *1*, 208–212. [CrossRef]
32. Poderico, M.; Parrilli, S.; Poggi, G.; Verdoliva, L. Sigmoid shrinkage for BM3D denoising algorithm. In Proceedings of the 2010 IEEE International Workshop on Multimedia Signal Processing, Saint-Malo, France 4–6 October 2010; pp. 423–426.
33. Cardona-Mesa, A.A.; Vásquez-Salazar, R.D.; Diaz-Paz, J.P.; Sarmiento-Maldonado, H.O.; Gómez, L.; Travieso-González, C.M. Optimization of autoencoders for speckle reduction in sar imagery through variance analysis and quantitative evaluation. *Mathematics* **2025**, *13*, 457. [CrossRef]
34. Qin, X.; Zhang, Y.; Li, Y. Efficient Maximum-Likelihood Estimation of Equivalent Number of Looks for PolSAR Image. *IEEE Geosci. Remote Sens. Lett.* **2023**, *20*, 1–5. [CrossRef]
35. Sara, U.; Akter, M.; Uddin, M.S. Image Quality Assessment through FSIM, SSIM, MSE and PSNR—A Comparative Study. *J. Comput. Commun.* **2019**, *7*, 8–18. [CrossRef]
36. Bakurov, I.; Buzzelli, M.; Schettini, R.; Castelli, M.; Vanneschi, L. Structural similarity index (SSIM) revisited: A data-driven approach. *Expert Syst. Appl.* **2021**, *189*, 116087. [CrossRef]
37. Gómez, L.; Ospina, R.; Frery, A.C. Statistical Properties of an Unassisted Image Quality Index for SAR Imagery. *Remote Sens.* **2019**, *11*, 385. [CrossRef]
38. Mirza, M.; Osindero, S. Conditional Generative Adversarial Nets. *arXiv* **2014**, arXiv:1411.1784.
39. Handalage, U. Generative Adversarial Networks—A Review of Its Variants and Applications. 2021. Available online: https://www.researchgate.net/profile/Upulie-Handalage/publication/355395578_Generative_Adversarial_Networks_-_A_Review_of_Its_Variants_and_Applications/links/61bed0eba6251b553acb88da/Generative-Adversarial-Networks-A-Review-of-Its-Variants-and-Applications.pdf (accessed on 12 December 2024). [CrossRef]

40. Liu, Z.; Lin, Y.; Cao, Y.; Hu, H.; Wei, Y.; Zhang, Z.; Lin, S.; Guo, B. SWIN Transformer: Hierarchical Vision Transformer using Shifted Windows. In Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, BC, Canada, 11–17 October 2021; pp. 9992–10002.
41. Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. An Image is Worth 16 × 16 Words: Transformers for Image Recognition at Scale. *arXiv* **2020**, arXiv:2010.11929.
42. Zamir, S.W.; Arora, A.; Khan, S.H.; Hayat, M.; Khan, F.S.; Yang, M.H. Restormer: Efficient Transformer for High-Resolution Image Restoration. In Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, 18–24 June 2022; pp. 5718–5729.
43. Xu, X.; Hao, J. U-Former: Improving Monaural Speech Enhancement with Multi-head Self and Cross Attention. In Proceedings of the 2022 26th International Conference on Pattern Recognition (ICPR), Montreal, QC, Canada, 21–25 August 2022; pp. 663–669.
44. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. *arXiv* **2015**, arXiv:1505.04597.
45. Balochian, S.; Baloochian, H. Edge detection on noisy images using Prewitt operator and fractional order differentiation. *Multimed. Tools Appl.* **2022**, *81*, 9759–9770. [CrossRef]
46. Wu, J.; Déniz, L.G.; Frery, A.C. A Non-Local Means Filters for Sar Speckle Reduction with Likelihood Ratio Test. In Proceedings of the IGARSS 2022—2022 IEEE International Geoscience and Remote Sensing Symposium, Kuala Lumpur, Malaysia, 17–22 July 2022; pp. 2319–2322.
47. Déniz, L.G.; Buemi, M.E.; Jacobo-Berlles, J.; Mejail, M. A New Image Quality Index for Objectively Evaluating Despeckling Filtering in SAR Images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2016**, *9*, 1297–1307.
48. Gómez, L.; Cardona-Mesa, A.A.; Vásquez-Salazar, R.D.; Travieso-González, C.M. Analysis of Despeckling Filters Using Ratio Images and Divergence Measurement. *Remote Sens.* **2024**, *16*, 2893. [CrossRef]
49. Lin, J. Divergence measures based on the Shannon entropy. *IEEE Trans. Inf. Theory* **1991**, *37*, 145–151. [CrossRef]
50. Menéndez, M.; Pardo, J.; Pardo, L.; Pardo, M. The Jensen-Shannon divergence. *J. Frankl. Inst.* **1997**, *334*, 307–318. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.