



Article

Despeckling of SAR Images Using Residual Twin CNN and Multi-Resolution Attention Mechanism

Blaž Pongrac * and Dušan Gleich

Faculty of Electrical Engineering and Computer Science, University of Maribor, 2000 Maribor, Slovenia

* Correspondence: blaz.pongrac1@um.si

Abstract: The despeckling of synthetic aperture radar images using two different convolutional neural network architectures is presented in this paper. The first method presents a novel Siamese convolutional neural network with a dilated convolutional network in each branch. Recently, attention mechanisms have been introduced to convolutional networks to better model and recognize features. Therefore, we propose a novel design for a convolutional neural network using an attention mechanism for an encoder–decoder-type network. The framework consists of a multiscale spatial attention network to improve the modeling of semantic information at different spatial levels and an additional attention mechanism to optimize feature propagation. Both proposed methods are different in design but they provide comparable despeckling results in subjective and objective measurements in terms of correlated speckle noise. The experimental results are evaluated on both synthetically generated speckled images and real SAR images. The methods proposed in this paper are able to despeckle SAR images and preserve SAR features.

Keywords: synthetic aperture radar; speckle; speckle suppression; despeckling; deep learning; convolutional neural network



Citation: Pongrac, B.; Gleich, D. Despeckling of SAR Images Using Residual Twin CNN and Multi-Resolution Attention Mechanism. *Remote Sens.* **2023**, *15*, 3698. <https://doi.org/10.3390/rs15143698>

Academic Editor: Masoud Mahdianpari

Received: 25 May 2023
Revised: 27 June 2023
Accepted: 20 July 2023
Published: 24 July 2023



Copyright: © 2023 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) is an air-born radar imaging sensor capable of producing radar images under all weather conditions and at any time of the day. It provides backscattering information on electromagnetic echoes produced by radar. Many different applications using radar imaging technology have been used widely in polarimetry and interferometry for the monitoring of the Earth's surface in real time. SAR is a coherent imaging system that is affected by speckle noise, which hinders details of the observed scene and makes automatic image interpretation difficult. Speckle noise is a consequence of many scattered echoes with a shifted phase within the same resolution cell. The sum of all echoes within the resolution cell causes strong fluctuations in intensity from one cell to another. SAR data are generally complex-valued; therefore, information of interest can be explored within the SAR amplitude data or a phase part. In recent years, many different SAR satellites, including Cosmo SkyMed, Sentinel 1, ALOS-2, and RadarSAT2, have been orbiting around the Earth, and the data can be accessed very easily. Speckle removal is an essential task before performing SAR data processing. Many different algorithms have been proposed for suppressing speckle since 1980 [1].

Speckle can be suppressed by using multi-looking techniques or spatial and frequency averaging, applying Bayesian inference or time-frequency analysis, averaging neighbor pixels, and, more recently, using convolutional networks. Speckle is multiplicative noise; therefore, models for despeckling should estimate speckle and preserve all spatial features, edges, and strong scatterers. The estimated speckle should have a Gamma distribution with a mean equal to 1.

Different methods of despeckling exist in the literature that depend on the modeling of SAR data. Adaptive spatial domain filters, such as Lee [2], Kuan [3], and Frost [4],

deploy a weighted average of the central pixel intensity and the average intensity of its neighboring pixels in the moving window. An advanced version of the Lee and Frost filters was proposed that operates similarly but categorizes the coefficient of variation into homogeneous regions, heterogeneous regions, and isolated points [5].

Model-based despeckling involves maximizing the posterior (MAP) probability density function (PDF), which consists of PDF data and a model [6]. A group of despeckling techniques involves discrete wavelet transform (DWT). These techniques involve reducing noise by applying thresholding to the coefficients of the log-transformed single-look image's DWT. Several works have been proposed within wavelet and the second generation of wavelet transforms, which use the threshold within the wavelet subbands [7], or apply Bayesian inference to the subbands [8]. The authors of [9] proposed a weighted-average algorithm of similar pixel values. Pixel similarity is defined as the Euclidean distance between patches. A more general definition of pixel similarity is based on a noise distribution model, as proposed in [9]. A scattering-based version of the SAR block-matching 3D (SARBM3D) filter [8] was presented in [10]. The authors of [10] modified the original algorithm from [8] and exploited the already available information on the imaged scene, and the authors of [11] proposed a despeckling evaluation framework.

Deep learning has also been investigated for image-denoising tasks in recent years. Deep learning methods have shown a lot of success in classification [12] and low-level computer vision problems, such as segmentation [13], denoising [14], and super-resolution [15]. State-of-the-art image restoration with high-quality results and real-time processing capability using a convolutional neural network (CNN) was proposed in [16]. The deep learning approach can also be applied to the denoising of optical images, mainly in the context of additive white Gaussian noise (AWGN). A dramatic improvement in denoising can be achieved using advanced regularization and learning methods, such as Rectifier Linear Unit (ReLU), batch normalization, and residual learning [17]. The authors of [14] presented a feed-forward denoising convolutional neural network (DCNN) following a residual learning approach [17] paired with batch normalization (BN) [18]. Residual learning methods perform better since they focus on predicting the residual image (i.e., noise image) instead of directly producing a noise-free image. The residual learning approach helps improve the training of CNNs. CNNs learn better when asked to produce an output that is significantly different from the input [14,19]. The authors of [20] proposed an image-despeckling convolutional neural network (ID-CNN), which assumes multiplicative speckle noise, recovering the filtered image through a component-wise division-residual layer.

Recent research has proposed different machine learning approaches for processing SAR images, such as the SAR Dilated Residual Network (SAR-DRN) [21], which employs dilated convolutions and a combination structure of skip connections with residual learning, the CNN-based deep encoder–decoder architecture of the U-Net to capture speckle statistical features [22], and a non-local despeckling method for SAR images. In this method, the weight of the target pixel is estimated using a convolutional neural network [23]. The despeckling of SAR images within the contourlet wavelet transform using a CNN-based structure was proposed in [24]. The CNN-based despeckling of polarimetric SAR data using a complex-value CNN was proposed in [25]. Recently, CNNs have been updated with attention mechanisms [26], where a residual attention network maps features better within an encoder–decoder network. A second-order channel attention that refines convolutional features using second-order statistics was proposed in [27]. The despeckling performance was improved by incorporating an attention mechanism [28–31]. Almost all despeckling methods introduce several artifacts, which are consequences of either image modeling or the spatial relation between image data. The SAR evaluation procedures are well defined in [11].

Recently, published methods for SAR image despeckling have used convolutional neural networks [32–36]. The authors of [32] proposed a residual network known as SAR-DRDNet, which consists of non-local and detail recovery parts and uses the global information of the SAR image and multiscale contextual information of the pixels. A

wavelet-based thresholding method known as MSPB [33] uses the pixel neighborhood and a bilateral filter for noise suppression, together with an intelligent Bayesian thresholding rule. SAR-DDPM [34] is a denoising diffusion probabilistic model with a Markov chain. The despeckled image is obtained through a reverse process that predicts the added noise iteratively using a noise predictor conditioned on the speckled image. The authors of [35] employed an overcomplete CNN architecture to focus on learning low-level features by restricting the receptive field. The proposed network consists of an overcomplete branch that focuses on the local structures and an undercomplete branch that focuses on the global structures. Transformer-based SAR image despeckling [36] comprises a Transformer-based encoder, which allows the network to learn global dependencies between different image regions, aiding in better despeckling. The network is trained end-to-end with synthetically generated speckled images using a composite loss function. The SAR-CAM method [37] improves the performance of the encoder–decoder CNN architecture by using various attention modules to capture multiscale information.

A self-calibrated dilated convolutional neural network for SAR image despeckling called SARSCDCNN was proposed in [38], which consists of several self-calibrated blocks. The features are extracted within two branches, representing the contextual features within the original space and the features within the long-range space. A down-up sampling operation and convolutions with hybrid dilated rates are used. Multi-temporal features and a similarity estimation approach to despeckling were proposed in [39]. A single-image-capable despeckling method using similarity-based block-matching and a noise-referenced encoder–decoder convolutional neural network exploits similarity-based block-matching within one noisy SAR image. The method uses a Siamese network to share parameters between two branches.

Traditional CNNs try to increase the receptive field size as the network goes deeper, thus extracting global features. However, the speckle is relatively small, and increasing the receptive field does not help extract speckle features. The goal of newly developed methods is to incorporate newly designed features into deep convolutional networks, such as spatial attention mechanisms and transformers.

This paper presents two different approaches to the modeling of SAR data statistics, both using a CNN. The first method uses residual learning with additive noise, similar to [14]. The adaptation of optical imaging processing to SAR image processing is achieved by handling multiplicative speckle noise and using a dual or Siamese CNN. Each sub-network consists of a structure and loss function, which ensures that each sub-network is the most similar to the corresponding clean SAR image. The minimal configuration of each sub-network assures minimal training time using the proposed training database. The second method uses an encoder–decoder structure, similar to the U-Net structure, and adds a multi-resolution attention mechanism for advanced speckle modeling. The attention mechanism consists of several additional sub-networks: a Dilated Residual Network (DRN), an Attention Supervision Network (ASN), and a h1Multi-resolution Attention Mechanism (MAM). The experimental results of the proposed methods show very good results in speckle estimation and provide visual results similar to those of the SARBM3D method for synthetic and real data. In addition, the proposed methods provide better results compared to the SAR-CAM and overcomplete neural network methods in objective and subjective measurements.

2. Related Works

2.1. Gaussian Denoiser

Despeckling using a deep convolutional network, as proposed in [14], is used to estimate Gaussian noise statistics. The architecture of the algorithm is depicted in Figure 1. The algorithm is designed in such a manner that it assumes an additive speckle model. Therefore, the algorithm estimates the speckle noise, which is subtracted from the original image, as shown in Figure 1. The supervised algorithm is trained using synthetically generated speckled images with known noise parameters. The goal was to design a

network that minimizes errors between the estimated and synthetically generated speckle. Different approaches to the generation of synthetic speckle images have been used in the literature. The simplest one adds synthetically generated speckle [40] to the optical images as multiplicative noise [14,20]. The averaging method using time series over a scene with constant backscatter over time is the most reliable method for providing speckle-free images [19]. The main problem in SAR images is that strong scatterers and double-bounce scatterers do not contain speckle, and this model should be used in speckle modeling.

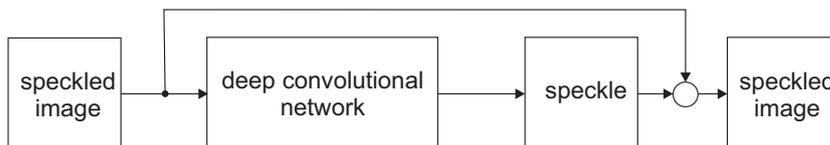


Figure 1. Architecture of the denoising convolutional neural network [14].

The network consists of 17 fully convolutional layers with no pooling. The first layer uses Conv+ReLU with 64 filters of size 3×3 to generate 64 feature maps, followed by rectified linear units for non-linearity. Layers 2–16 use Conv+BN+ReLU with 64 filters of size $3 \times 3 \times 64$, where batch normalization is added between the convolution and the ReLU. The last layer uses Conv with a filter of size $3 \times 3 \times 64$ to reconstruct the output. The loss function that is minimized during the training step is the l_2 loss (i.e., the sum of squared errors, averaged over the whole training set).

The goal of despeckling using a CNN is to determine the link between speckled and despeckled images, which is defined as

$$x = CNN(y, \theta) \tag{1}$$

where CNN represents the denoising CNN , and $\theta = \{W1, W2, b_1, b_2\}$ represents the parameters of the network. Equation (1) can be transformed for a CNN with residual learning R as

$$y - x = R(y, \theta) \tag{2}$$

Residual learning R is given by

$$R(y, \theta) = W_2 f(W_1 y + b_1) + b_2 \tag{3}$$

where θ and $f(\cdot)$ represent the non-linearity at the rectified linear units $ReLU = \max(0, \cdot)$. A single hidden-layer CNN can be extended by introducing M convolutional layers with W_M weights. Residual learning $R(y; \theta)$ is then given by

$$\begin{aligned} R(y, \theta) &= W_M * a_{M-1} + b_M \\ a_m &= ReLU \cdot BN(W_m * a_{m-1} + b_m) \end{aligned} \tag{4}$$

where b_m represents the estimated biases, a_m represents an activation, BN is the batch normalization, and $m \in (1, M)$ is the current convolutional layer. Each convolutional layer's size is defined as $p \times p \times C_m \times d_m$, where $p \times p$ is the size of the convolutional filter, $C_m = d_{m-1}$ is the number of channels in the m -th layer, and d_m is the number of convolutional filters in the m -th layer. A loss function $l(\theta)$ estimates the discrepancy D between the target output and actual network output. The loss function $l(\theta)$ is estimated as

$$l(\theta) = \sum_{i=1}^N D(R(\log y_i, \theta), \log(y_i / x_i)) \tag{5}$$

where x_i and y_i represent the discrepancy between the targets in N training pairs, and the actual output is $D(x, y) = \|\log x - \log y\|_2^2$. A mini-batch stochastic gradient descent algorithm is utilized to minimize the loss function $l(\theta)$. Other perceptual-based loss

functions could be used to achieve better visual quality. However, the use of different loss functions is beyond the scope of this paper and is not discussed further.

The Adam optimizer [41] is utilized to train the designed CNN. The same parameters are used throughout the designed CNN architecture to reduce the number of parameters and the plausibility of overfitting.

Chierchia et al. [19] used a log transformation and applied a very similar structure to SAR image despeckling using an additive model. A multiplicative model applied to a deep convolutional network was proposed in [20], as depicted in Figure 2, where the smoothing of images is improved by the joint minimization of both the Euclidean loss (EL) and total variation (TV) loss functions.

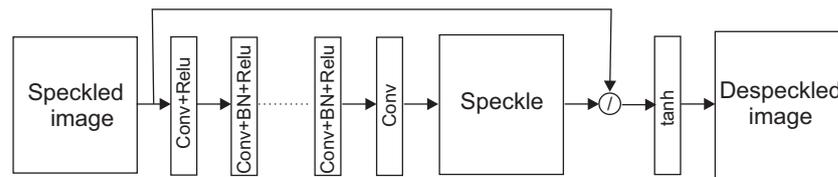


Figure 2. SAR image despeckling assuming a multiplicative noise model [20].

2.2. SAR Dilated Residual Network

A novel network for SAR image despeckling, called SAR-DRN, was proposed, which is trained in an end-to-end fashion using a combination of dilated convolutions and skip connections with a residual learning structure [21]. Instead of relying on a pre-determined image, a priori knowledge, or a noise description model, the main advantage of using a deep neural network strategy for SAR image despeckling is that the model can acquire and update the network parameters directly from the training data and the corresponding labels, which does not require the manual adjustment of critical parameters, and can automatically learn the complex internal non-linear relations with trainable network parameters from the massive training simulative data. The holistic neural network model for SAR image despeckling contains seven dilated convolution layers and two skip connections, as illustrated in Figure 3.

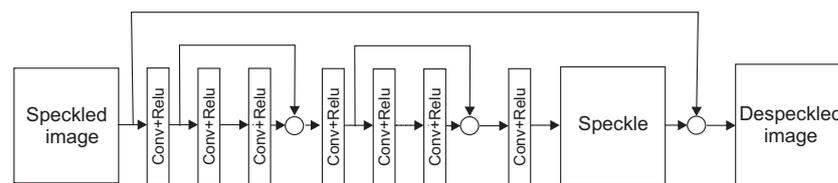


Figure 3. SAR-DRN for SAR image despeckling [20].

Dilated Filter

Dilated convolutions can both enlarge the receptive field and maintain the filter size with a lightweight structure. Furthermore, skip connections are added to the despeckling model to maintain the image details and avoid the vanishing gradient problem. Dilated convolution is a technique that expands the kernel by inserting zeros between its consecutive elements. In simpler terms, it is the same as regular convolution, but it involves pixel skipping, allowing it to cover a larger area of the input. An additional parameter l (dilation factor) indicates how much the input is expanded. Based on the value of this parameter, $(l - 1)$ pixels are skipped in the kernel. Therefore, regular convolution has a dilation factor of $l = 1$. Dilated convolution helps expand the area of the input image area without pooling. The objective is to capture more information from the output obtained with every convolution operation. This method offers a wider field of view at the same computational cost. We determine the value of the dilation factor l by observing how much information is obtained with each convolution for varying values of l . By using this method, we can obtain more information without increasing the number of kernel parameters. When we keep the value of $l = 2$, we skip 1 pixel ($l - 1$ pixels) while mapping the filter onto the

input, thus capturing more information in each step. If the traditional 3×3 convolution filter is used, the network will either have a receptive field with dimensions of 15×15 with the same network depth (i.e., 7) or have a depth of 16 with the same receptive field (i.e., 33×33).

2.3. U-Shaped Denoising Network

Based on the principle of denoising auto-encoders, the authors of [22] proposed a despeckling method, where speckled images and their corresponding speckle-free reconstructions are involved in training the CNN. The same approach can be applied to SAR image despeckling by extracting latent representations from the speckled images and generating their corresponding reconstructions. A modified version of the U-Net, initially designed for biomedical image segmentation tasks [42], can compress the information by extracting relevant features from the input image at different scales. The U-Net employs a symmetric encoder–decoder structure (U-shape), with skip connections going from the contracting path on the left to the expansive one on the right. This way, details in the SAR image can be captured at various scales. Figure 4 depicts the architecture of the original U-Net proposed in [42].

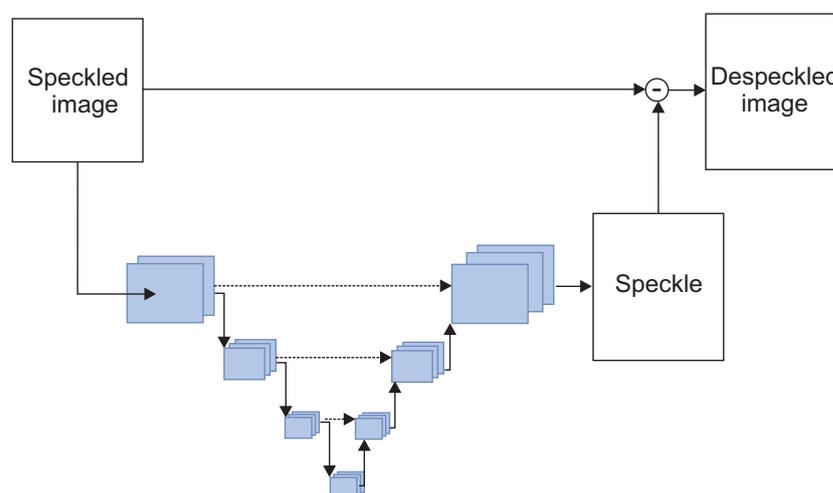


Figure 4. Architecture of U-shaped CNN [22].

3. The Proposed Deep Despeckling Architecture

3.1. The Proposed Architecture of the Siamese-Based Dilated Deep CNN

The combination of a Siamese network and the SAR-DNR method allows for the exploitation of both the inherent properties of speckle noise learned by the network and the denoising capabilities of the SAR-DNR algorithm. This novel method, which is basically a fusion of two existing networks, can lead to improved despeckling results by effectively preserving important image information while suppressing unwanted noise. The Siamese network can learn discriminative features that capture the statistical properties of speckle noise. By training the network on pairs of noisy and noise-free SAR images, it can learn to distinguish between noise and the underlying signal. This feature learning capability enables the network to extract relevant information for despeckling. The Siamese network can capture complex non-linear relationships between the input noisy SAR image and the desired denoised output. By employing deep neural network architectures, the Siamese network can learn intricate mappings that effectively suppress speckle noise while preserving important image details. The Siamese network can capture contextual information within the SAR imagery by considering local and global image patches. This allows the network to leverage the spatial dependencies and correlations present in SAR images, enabling better noise estimation and removal. The trained Siamese network can adapt to the different noise patterns and noise levels encountered in SAR imagery. The combination of a Siamese network and the SAR-DNR method allows for the exploitation of both the inherent

properties of speckle noise learned by the network and the denoising capabilities of the SAR-DNR algorithm. This fusion of approaches leads to improved despeckling results by effectively preserving important image information while suppressing unwanted noise. A twin-based network using a SAR-DNR-like network structure is depicted in Figure 5.

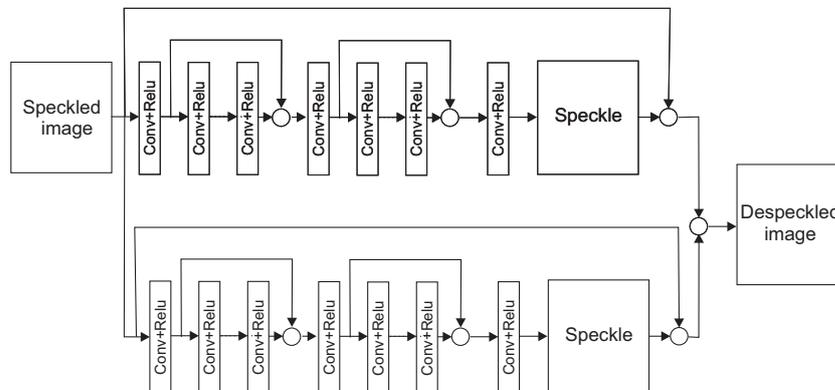


Figure 5. Architecture of the Siamese-based Dilated Residual Convolutional Neural Network (SDR-CNN).

The Siamese-based Neural Network (SNN) [43] consists of twin networks, which accept distinct inputs but are joined by an energy function at the top. This function estimates the metric between the highest-level feature representation on each side. The parameters between the twin networks are tied. Weight-tying guarantees that two extremely similar images could not possibly be mapped by their respective networks to very different locations in feature space because each network computes the same function. Also, the network is symmetric, so that whenever we present two distinct images to the twin networks, the top conjoining layer will compute the same metric as if we were to present the same two images to the opposite twins.

3.1.1. Loss Function

The performance of a CNN relies heavily on loss functions. While per-pixel *EL* is a reliable quantitative measure, it could reduce image details and produce artifacts in the final output image. The minimization of pixel-wise errors cannot be based solely on the per-pixel loss; therefore, differences in detailed features must be considered. To address this problem, we combine the *TV* regularization term and *EL* into a final loss function that evaluates the visual quality of each subband. The *TV* regularization term and *EL* are estimated as

$$TV = \sum_{i=1}^W \sum_{j=1}^H \sqrt{(x(i, j + 1) - x(i, j))^2 + (x(i + 1, j) - x(i, j))^2} \tag{6}$$

$$EL = \frac{1}{WH} \sum_{i=1}^W \sum_{j=1}^H (x(i, j) - y(i, j))^2$$

where $x(i, j)$ is the input speckled SAR image, and $y(i, j)$ is the corresponding despeckled image with width W and height H . The combined loss function, which can be used for estimating the difference between the speckled and corresponding despeckled images, is given by

$$L_{CS} = TV + EL \tag{7}$$

The common loss function can be estimated using the *TV* and penalty terms for both sub-networks:

$$L(d, Y) = \frac{1}{2} Y d^2 + (1 - Y) \frac{1}{2} \max(0, m - d)^2 \tag{8}$$

$$d = TV_{low} + TV_{high} + EL_{low} + EL_{high} \tag{9}$$

where the subscripts *low* and *high* represent the two branches of the network.

3.1.2. Training the Designed CNN

The two CNN models were trained on two datasets: a synthetic SAR image dataset and a real SAR image dataset. The synthetic SAR image dataset was generated by adding a single-look speckle to 2500 optical images. The real SAR image dataset consisted of a stack of 150 images obtained from 10 different scenes using the Sentinel platform. All the images used in the experiment were either cut out from existing SAR images or resized to dimensions of 256×256 , and 90% of the images in each dataset were used for learning. The TensorFlow framework was used to implement the entire network, and the Adam optimizer was employed to train it [41]. The ReLU activation function was used, with convolution kernels of 3×3 for feature extraction and 1×1 for feature aggregation. The final convolution layer generated a grayscale image. The learning rate was initially set to 0.0002, which decreased by a quarter every 25 epochs. The CNN training took approximately nine hours to complete using an NVIDIA RTX 2080Ti for the 13 sub-networks.

Figure 6a,b show real SAR images of the same area acquired by the ALOS-2 satellite and TerraSAR-X. The images were taken in HH polarization, with a ground resolution of 4 (Alos-2) and 1 (TerraSAR-X) meters. The images depict an urban area with houses, a river, bridges, a forest, and homogeneous areas.

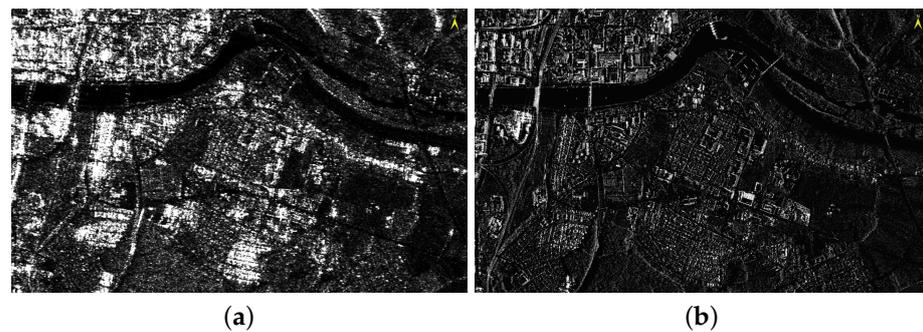


Figure 6. Geocoded SAR images of an urban area. (a) ALOS 2 image with a ground resolution of 4 m. (b) TerraSAR-X image with a ground resolution of 1 m.

3.2. The Proposed Architecture of the Dilated Deep CNN with an Attention Mechanism

In this paper, we propose a combination of different deep learning approaches for despeckling image enhancement. The architecture of this method is depicted in Figure 7. It consists of a Dilated Residual Network (DRN), U-net, Attention Supervision Network (ASN), and hIMulti-resolution Attention Mechanism (MAM). The algorithm converts a speckled image into feature maps using the first set of convolutional layers, and the U-net extracts fine features at different scales. The extracted features are enhanced once again using the ASN, which extracts important features and neglects non-important features; therefore, it gives larger weights to important features using training sets. The MAM is used to enhance contextual features at different scales. All the features are reconstructed in the final convolutional layer, which provides a speckle-free image.

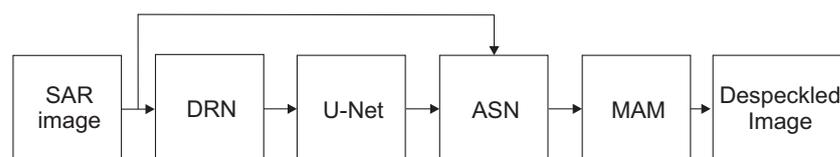


Figure 7. Structure of the Attention-Based Convolutional Neural Network (ABCNN).

To introduce additional speckle modeling, the DRN is inserted before the U-net to extract features additionally. The DRN used is shown in Figure 8, and it is described in detail in the previous sections. The ASN shown in Figure 9 uses the extracted features of

the U-net, generated into the residual image using 1×1 convolution, and adds them to the original image. This step is very important since it adds additional supervision using the original image. The resulting sum is further filtered using 1×1 convolution and a Sigmoid function, which generates an attention mask.

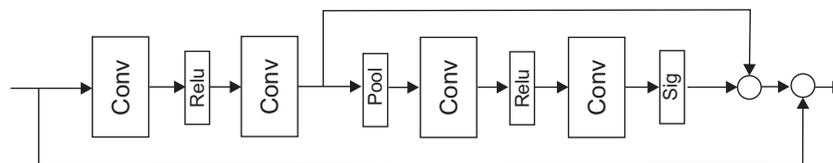


Figure 8. Structure of the DRA network used within the proposed network.

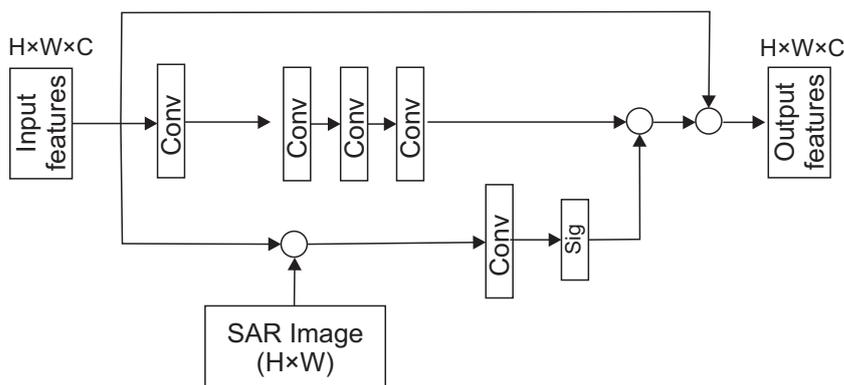


Figure 9. Structure of the Attention Supervision Module (ASM).

The information hindered by the speckle can be recovered if the speckle can be detected across different resolutions. A MAM can be used to detect and extract features and multiscale features. The MAM block can be designed using dilation convolution and an Efficient Channel Attention (ECA) module [44]. The main role of the residual block is to overcome the problem where the learning rate becomes low and the accuracy cannot be effectively improved due to the deepening of the network. The MAM is depicted in Figure 10a, where a dilated convolution is used with factors of 1, 3, and 5 for each convolution branch. The residual block within the MAM network is shown in Figure 10b. The ECA module collects the aggregated features and averages them using Global Average Pooling (GAP) and then estimates the weights using one-dimensional CNN with a kernel dimension of 3. The MAM consists of a series of MAM modules. Five of them were considered, as shown in Figure 10a [44].

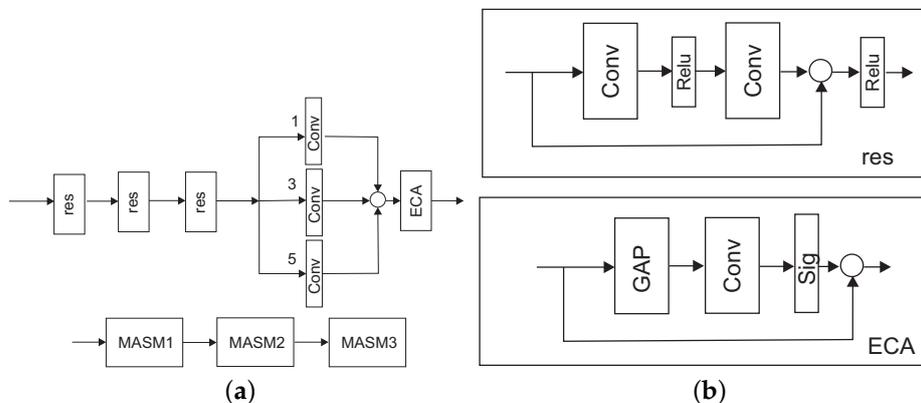


Figure 10. Multi-resolution attention mechanism (MAM). (a) Structure of the MAM. (b) Structure of the residual network within the MAM and structure of the ECA network.

Loss Function

The loss function that uses the L_2 norm is not effective because it introduces additional artifacts in the reconstructed data [45]. The loss function can be formed using a logarithm of despeckling gain and L_{TV}

$$L = L_{DG} + (1 - L_{SSIM}(Y - Y_{orig})) \quad (10)$$

where Y represents the despeckled image, Y_{orig} represents the speckle-free image, and L_{DG} is given by

$$L_{DG} = \log_{10} \sum_{i=1}^N \sum_{j=1}^M \frac{\|\hat{x}_{i,j} - x_{i,j}\|_2^2}{\|\hat{y}_{i,j} - x_{i,j}\|_2^2} \quad (11)$$

where N and M represent the width and height of the SAR image. Within the learning process, the Adam optimizer is used to update the network weights.

4. Experimental Results

This section evaluates the two proposed methods, Siamese-Based Dilated Residual Convolutional Neural Network (SDRCNN) and Attention-Based CNN (ABCNN), shown in Figures 5 and 7, respectively. Two scenarios were considered for the efficiency evaluation of the proposed methods using synthetically generated speckled images and real SAR images. First, we evaluated the performance of the proposed despeckling methods using synthetically speckled images, and later, we added the real-valued data to adapt the model to the real SAR data. This procedure can be done because the model can be further trained. In this paper, we generated synthetic speckle and added it to the synthetically generated images as multiplicative noise. A total of 800 images were used with known speckle. In [19], multi-temporal data were used to average the same scene and train the network to estimate the speckle. The second step was to use temporally averaged images from homogeneous areas. The strong scatterers from buildings within high-resolution SAR data do not contain any speckle. Therefore, those datasets were also included in the learning process. The proposed SDRCNN and ABCNN methods were compared to the SARBM3D [8], DCNN [20], overcomplete convolutional neural network (OCNN) [35], and SAR image despeckling using a continuous attention module (SAR-CAM) [37] methods.

4.1. Experimental Settings for the SNN Method

Figure 5 shows that the SNN comprises 17 fully convolutional layers without pooling. The first layer was built with 64 filters of size 3×3 to generate 64 feature maps, followed by rectified linear units for non-linearity (ReLU). For layers 2–16, 64 filters of size $3 \times 3 \times 64$ were used, with batch normalization added between the convolution and the ReLU. The final layer used a filter of size $3 \times 3 \times 64$ for output reconstruction. During the training step, the loss function l_2 was minimized, which involved the sum of squared errors averaged over the entire training set.

4.2. Synthetic Example

The despeckling efficiency was estimated using the despeckling evaluation methodology suggested in [11]. The authors of [11] proposed a method for evaluating despeckling techniques using specially designed metrics on five different datasets. The experiments conducted in this study employed synthetically generated images, including a homogeneous image, as well as square, building, corner, and Digital Elevation Model (DEM) images, as illustrated in Figures 11a, 12a, 13a, 14a and 15a.

4.2.1. Homogeneous Area

The synthetic SAR homogeneous image, shown in Figure 11a, was despeckled, and the image quality was assessed based on the mean of intensity (MoI), mean of ratio (MoR), variance of ratio (VoR), equivalent number of looks (ENL), and despeckling gain (DG), as proposed in [11]. MoI is an indication of possible bias; MoR shows how well the

speckle was estimated; VoR reports under-smoothing ($\text{VoR} < 1$) and over-smoothing ($\text{VoR} > 1$, only for non-flat areas); DG shows the despeckling rejection, with higher values providing better speckle reduction; and ENL represents the averaging during the SAR data formation. The measurements for the SAR homogeneous images are reported in Table 1. The proposed ABCNN method provided the best results in all measurements, followed by the proposed SDRCNN method. The SARBM3D and DCNN methods provided similar results for the objective measurements of the MoR, equivalent number of Looks (ENL), and DG. The OCNN and SAR-CAM methods were able to smooth the homogeneous image but introduced a shift in the mean of the despeckled image and did not estimate speckle noise well, as reported in Table 1. The despeckled homogeneous images using all methods are shown in Figure 11c–h.

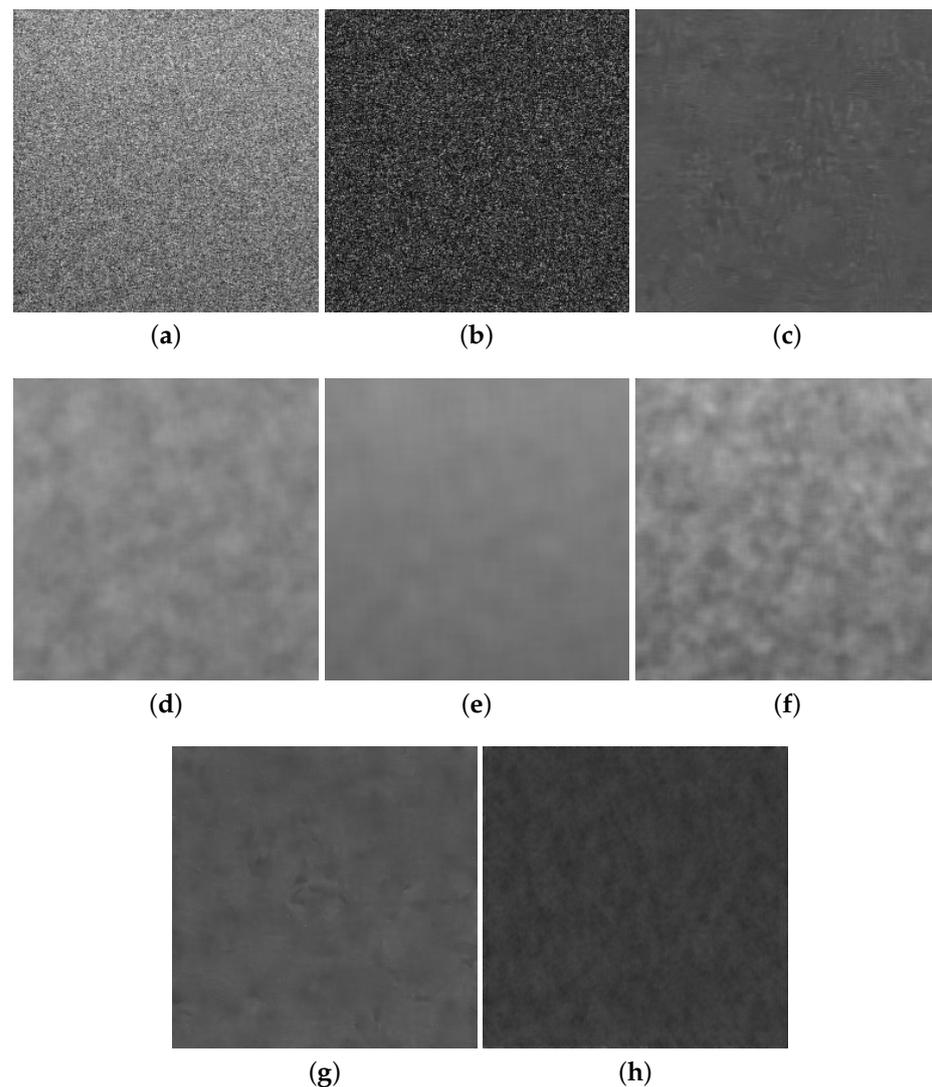
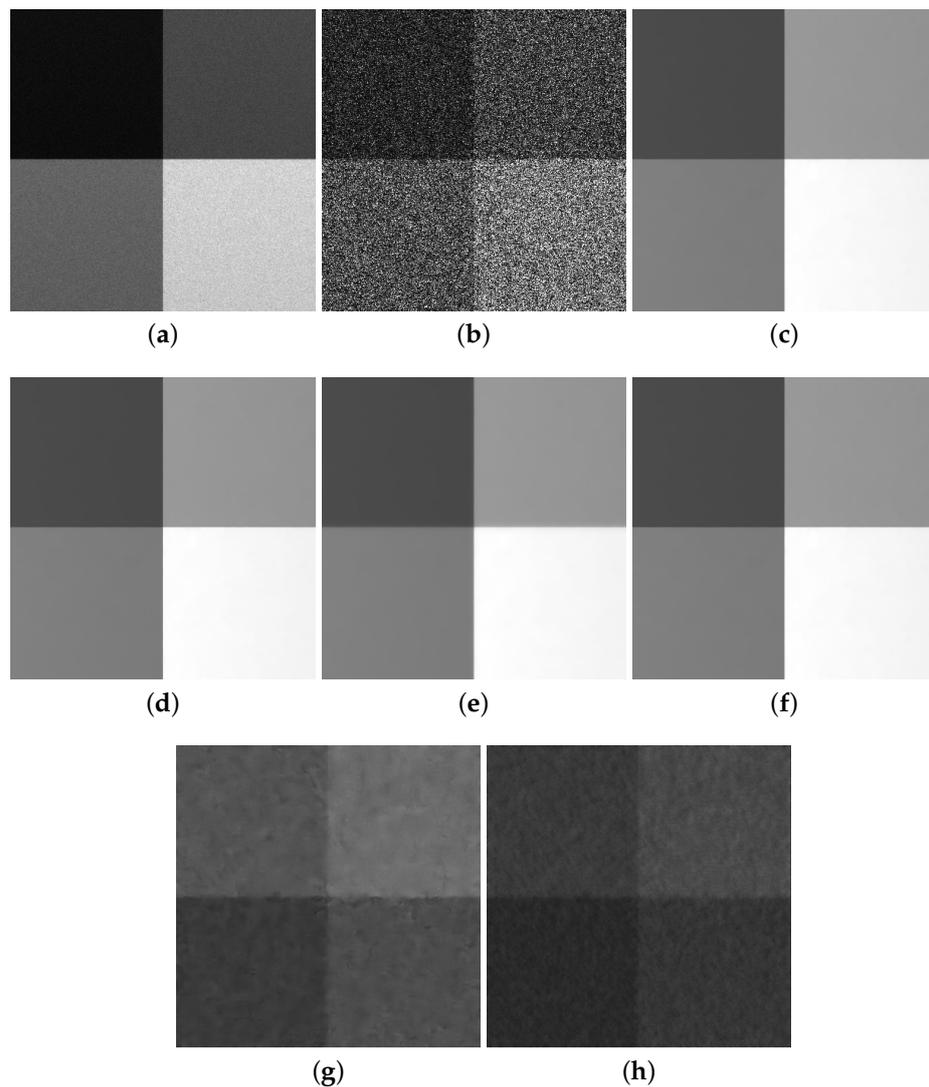


Figure 11. Synthetic SAR homogeneous images. (a) Original image. (b) Speckled image. (c) Image despeckled using the SDRCNN method. (d) Image despeckled using the ABCNN method. (e) Image despeckled using the SARBM3D method. (f) Image despeckled using the DCNN method. (g) Image despeckled using the OCNN method. (h) Image despeckled using the SAR-CAM method.

Table 1. Image quality measurements based on the MoI, MoR, VoR, ENL, and DG for homogeneous images.

Method	MoI	MoR	VoR	ENL	DG
clean	1.00	0.998	0.987	436.97	-
noisy	0.9980	-	-	1	0
SDRCNN	0.998	0.998	0.928	271	23.3
ABCNN	0.998	0.998	0.928	273	23.2
SARBM3D	0.998	0.997	0.912	150	21.65
DCNN	0.998	0.998	0.923	247	22.5
OCNN	1.12	0.893	0.788	541	14.4
SAR-CAM	0.78	1.27	1.362	194	12.6

**Figure 12.** Synthetic SAR square images. (a) Original image. (b) Speckled image. (c) Image despeckled using the SDRCNN method. (d) Image despeckled using the ABCNN method. (e) Image despeckled using the SARBM3D method. (f) Image despeckled using the DCNN method. (g) Image despeckled using the OCNN method. (h) Image despeckled using the SAR-CAM method.

4.2.2. Square Image

The proposed despeckling methods' efficiency in preserving edges was determined using a synthetic SAR image of four homogeneous areas using different methods, as shown in Figure 12a. Edge Smearing (ES) and Pratt's figure of merit (FOM) were used as indirect measures to evaluate the ability to recognize edges in the filtered image [11]. A higher

value of FOM shows better edge preservation. The square image had two vertical edges and two horizontal edges of varying contrast—one area had a higher contrast, whereas the other had a lower contrast. Table 2 shows the measurements for the synthetic SAR square image. The DCNN method resulted in the most blurred edges, followed by the SARBM3D, SDRCNN, and ABCNN methods. The values reported were rather similar, indicating that the CNN-based methods were superior to the SARBM3D method. The OCNN and SAR-CAM methods did not estimate speckle well; therefore, the despeckled images shown in Figure 12g,h still had noise that was basically artifacts caused by both methods. Both methods introduced a mean shift and smeared edges. These results were confirmed by the objective measurements shown in Table 2. The despeckled images are illustrated in Figure 12c–h.

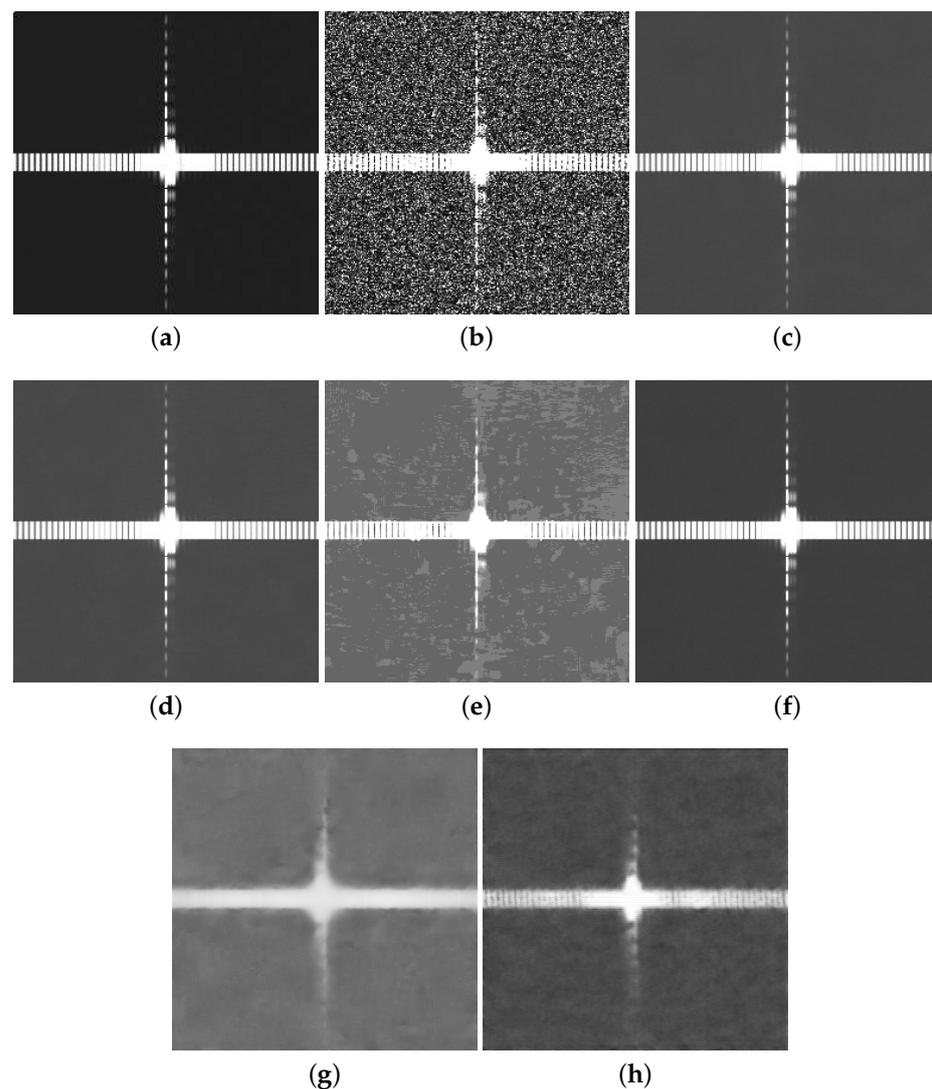
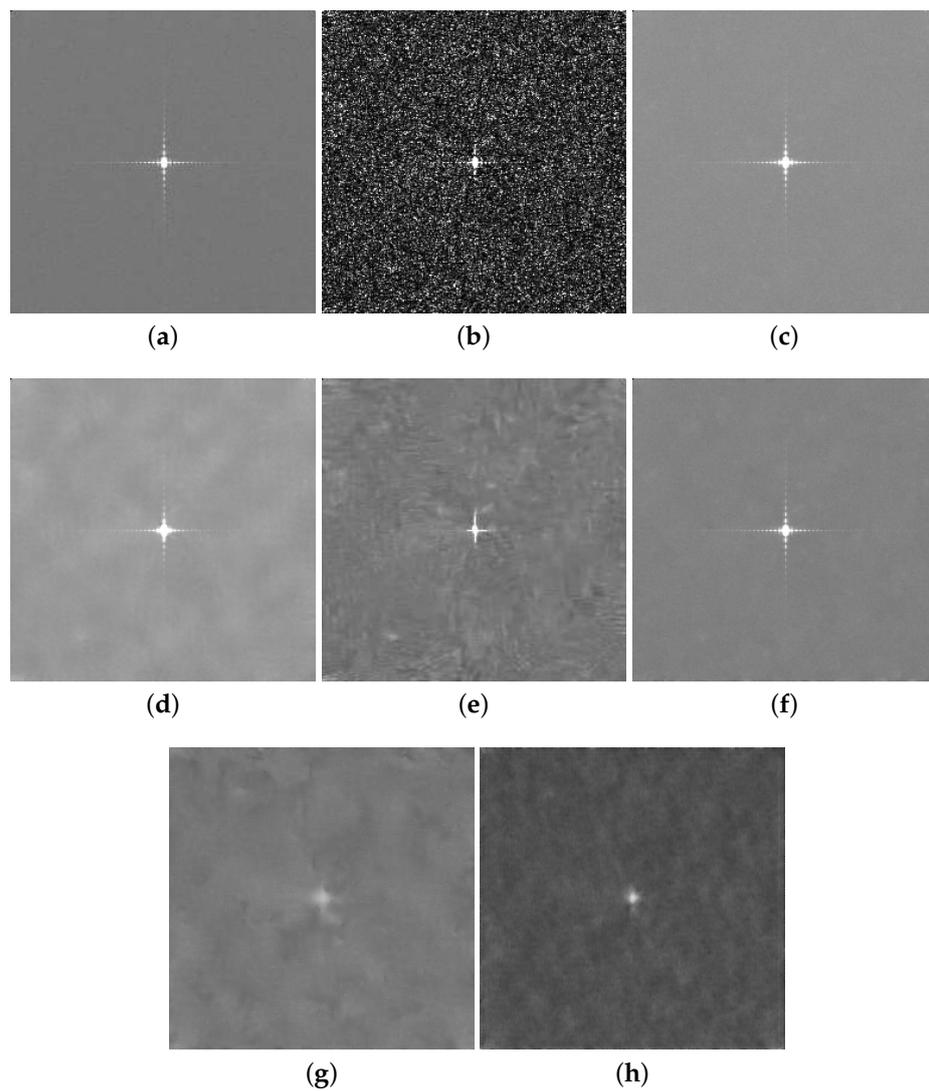


Figure 13. Synthetic SAR image depicting a building. (a) Original image. (b) Speckled image. (c) Image despeckled using the SDRCNN method. (d) Image despeckled using the ABCNN method. (e) Image despeckled using the SARBM3D method. (f) Image despeckled using the DCNN method. (g) Image despeckled using the OCNN method. (h) Image despeckled using the SAR-CAM method.

Table 2. Edge preservation measurements based on ES and FOM for the square image.

Method	ES (Up)	ES (Down)	FOM
clean	-	-	0.993
noisy	0.01	0.029	-
SDRCNN	0.071	0.21	0.888
ABCNN	0.071	0.21	0.885
SARBM3D	0.036	0.113	0.847
DCNN	0.070	0.21	0.881
OCNN	0.024	0.18	0.099
SAR-CAM	0.016	0.13	0.098

**Figure 14.** Synthetic SAR image depicting a corner reflector. (a) Original image. (b) Speckled image. (c) Image despeckled using the SDRCNN method. (d) Image despeckled using the ABCNN method. (e) Image despeckled using the SARBM3D method. (f) Image despeckled using the DCNN method. (g) Image despeckled using the OCNN method. (h) Image despeckled using the SAR-CAM method.

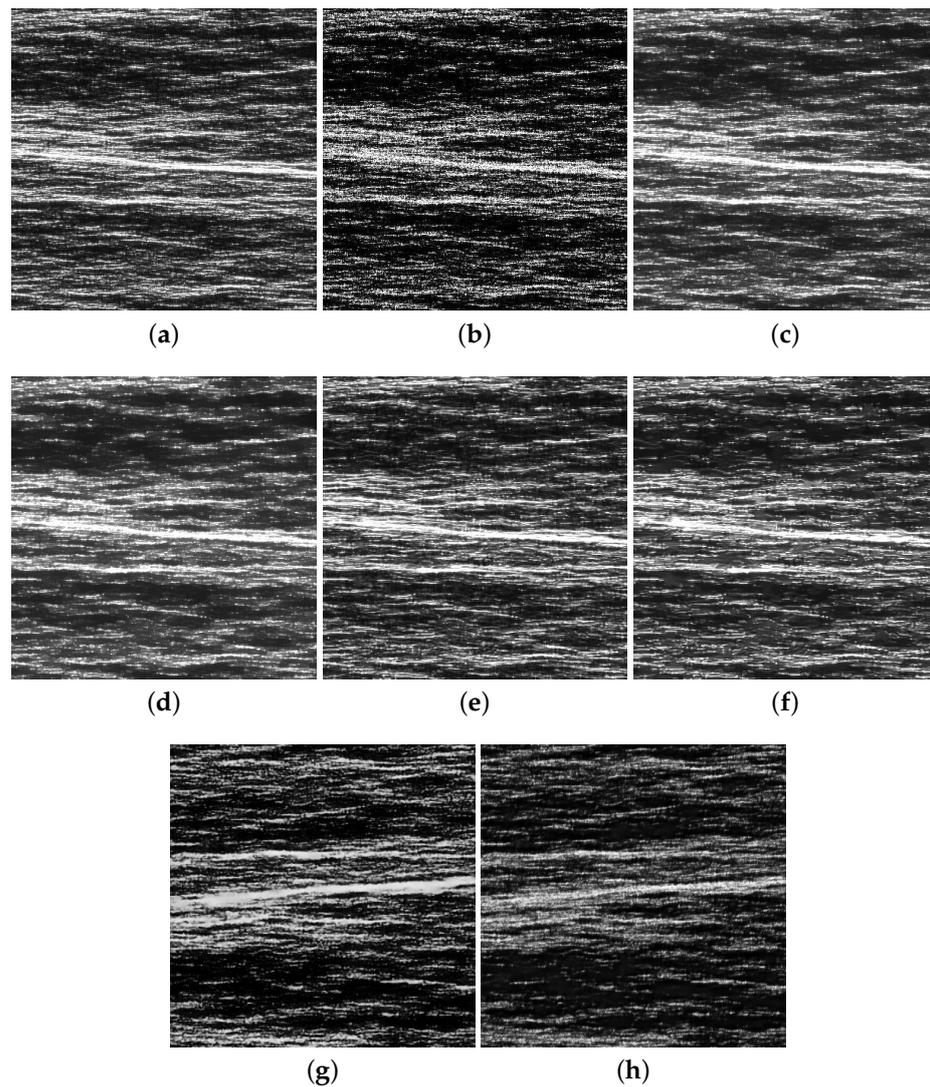


Figure 15. Synthetic SAR DEM image. (a) Original image. (b) Speckled image. (c) Image despeckled using the SDRCNN method. (d) Image despeckled using the ABCNN method. (e) Image despeckled using the SARBM3D method. (f) Image despeckled using the DCNN method. (g) Image despeckled using the OCNN method. (h) Image despeckled using the SAR-CAM method.

4.2.3. Building

A synthetic SAR image of a building was constructed, as shown in Figure 13a. Two parameters were assessed: radiometric precision C_{DR} , as the estimated contrast difference between the average intensity estimated on the double reflection segment and the average intensity of the SAR image's background, and Building Smearing (BS), which measures the distortion of the radiometric building profile in the range direction [11]. Table 3 shows the C_{DR} and BS measurements for the proposed despeckling methods. The SDRCNN and ABCNN methods preserved building features well by estimating BS close to zero. The SARBM3D method preserved building features well but had problems with correlated speckle noise over a homogeneous area. The OCNN and SAR-CAM methods smeared high-intensity scatterers, and noise remained within the homogeneous areas. The despeckled images are depicted in Figure 13c–h.

Table 3. Measurements for the building image.

Method	C_{DR}	BS
clean	127.96	-
SDRCNN	66.05	0.22
ABCNN	66.04	0.23
SARBM3D	65.91	1.46
DCNN	65.99	0.26
OCNN	55.83	0.31
SAR-CAM	51.69	0.33

4.2.4. Corner Reflector

Figure 14a,b depict corner reflectors with and without speckle, respectively. The contrast values C_{NN} and C_{BG} are indicators of radiometric preservation and were used to evaluate the proposed despeckling methods for synthetic SAR images depicting corner reflectors. C_{NN} is determined by calculating the logarithmic ratio between the intensity observed at the corner reflector site and the average intensity of the surrounding area formed by the eight-connected nearest neighbors. C_{BG} is computed as the ratio between the intensity observed at the corner reflector site and the average background intensity. The despeckled images should demonstrate values comparable to those obtained for the input image. Table 4 shows the C_{NN} and C_{BG} in the despeckled images using the proposed despeckling methods. The SDRCNN, ABCNN, SARBM3D, and DCNN methods delivered nearly identical outcomes since they all exhibited good preservation of point scatterers. The OCNN and SAR-CAM methods did not preserve the impulse response caused by the simulated corner reflector. The SAR-CAM method introduced a slight shift in the despeckled image and noise remained within the homogeneous areas. The despeckled images can be seen in Figure 14c–h.

Table 4. C_{NN} and C_{BG} measurements for a synthetic SAR image depicting a corner reflector.

Method	C_{NN}	C_{BG}
clean	7.75	36.56
noisy	7.77	36.50
SDRCNN	7.48	35.98
ABCNN	7.47	35.81
SARBM3D	7.39	35.46
DCNN	7.41	35.98
OCNN	3.21	18.27
SAR-CAM	4.23	14.56

4.3. DEM

The despeckling of the synthetic DEM image shown in Figure 15a was evaluated using the MoI, MoR, VoR, and DG, as well as the coefficient of variation C_x , providing information on the edge and texture preservation. Table 5 shows the measurements for the DEM image. The SDRCNN, ABCNN, DCNN, and SARBM3D methods obtained the best results. The coefficient of variation and despeckling gain show that the CNN-based techniques produced the best outcomes for the DEM image, followed by the SARBM3D method. The OCNN and SAR-CAM methods preserved the details of the DEM image. The OCNN over-smoothed the image and introduced bias, and the SAR-CAM method was unable to estimate speckle noise and also introduced bias. The despeckled images are illustrated in Figure 15c–h.

In conclusion, the efficiency of CNN-based methods depends primarily on the training images used. The proposed CNN despeckling methods successfully preserved homogeneous areas and edge preservation, but the SARBM3D method introduced some additional noise within the homogeneous areas. The SAR-CAM and OCNN methods introduced slight biases into the despeckled image, and both methods were not able to estimate speckle

within the homogeneous areas, which is reflected in the VoR, MoR, and DG. The point scatterers were not preserved, and all the SAR features were over-smoothed or smeared.

Table 5. Measurements for the DEM image.

Method	MoI	MoR	VoR	C_x	DG
clean	1.0	1.001	0.999	2.40	-
noisy	1.003	-	-	3.54	0
SDRCNN	0.999	0.973	0.91	2.73	6.68
ABCNN	0.999	0.973	0.90	2.72	6.68
SARBM3D	0.968	0.933	0.756	2.42	5.46
DCNN	0.999	0.971	0.90	2.70	6.61
OCNN	1.26	0.885	0.763	1.02	3.51
SAR-CAM	0.89	0.817	0.835	0.97	4.65

4.4. Real SAR Images

The proposed CNN despeckling methods were tested using two real SAR images: a mosaic of real SAR images and a high-resolution SAR image. The mosaic of real SAR images comprised 16 distinct patches using eight other SAR images obtained by the TerraSAR-X satellite in Spotlight and Stripmap mode. The SAR patches represented varied scenes with sizes of 200×200 pixels. Figure 16a shows the complete 800×800 -pixel mosaic image.

Figure 16b–g show the despeckled images using the proposed SDRCNN, ABCNN, SARBM3D, DCNN, OCNN, and SAR-CAM methods, respectively. The differences between the SDRCNN and ABCNN methods are barely noticeable. The images produced by the SDRCNN and ABCNN methods effectively suppressed speckles within the homogeneous areas while preserving strong scatterers. This can also be said for homogeneous areas with high scatterers such as grassland and woodlands and artificial objects such as buildings. The despeckled images generated by the SDRCNN and ABCNN methods preserved the textured regions. However, the homogeneous areas appeared significantly blurred. It was observed that in the case of the mosaic SAR image, the DCNN method yielded a despeckled image with more speckle in homogeneous areas compared to the ABCNN and SDRCNN methods. The DCNN-based methods occasionally introduced artifacts that appeared in the training datasets while strong scatterers were well-preserved. The SARBM3D method did not estimate speckle well and yielded a despeckled image with speckle in homogeneous areas. The OCNN method removed noise well over the homogeneous areas but over-smoothed point scatterers and smeared edges. It also introduced bias in the despeckled images. The SAR-CAM method did not over-smooth the textural features and point scatterers but did not estimate speckle well within homogeneous areas and also introduced bias.

The effectiveness of the proposed methods was evaluated using the ratio images between the original image shown in Figure 16a and the despeckled images shown in Figure 16b–g. The ratio images shown in Figure 17a,b confirm effective speckle estimation using the proposed SDRCNN and ABCNN methods. The ENLs for the SDRCNN and ABCNN methods' ratio images were 1.12 and 1.13, with a mean value of 0.997 and 0.998, respectively. The ratio images also demonstrate that the edges within the homogeneous areas were represented accurately. The SARBM3D and DCNN methods yielded slightly worse results than the SDRCNN and ABCNN methods. Nevertheless, the ratio images shown in Figure 17c,d confirm effective speckle estimation using the proposed SARBM3D and DCNN methods. The OCNN and SAR-CAM methods failed to estimate the speckle in the test images. It can be concluded that the CNN-based methods demonstrated comparable results in both the subjective and objective measurements, as shown in Table 6. The MoI, MoR, and VoR values are presented for the entire SAR image. The ENL was estimated from the homogeneous area of the SAR image. The estimated C_x values are shown in the first row and last column in Figure 16a.

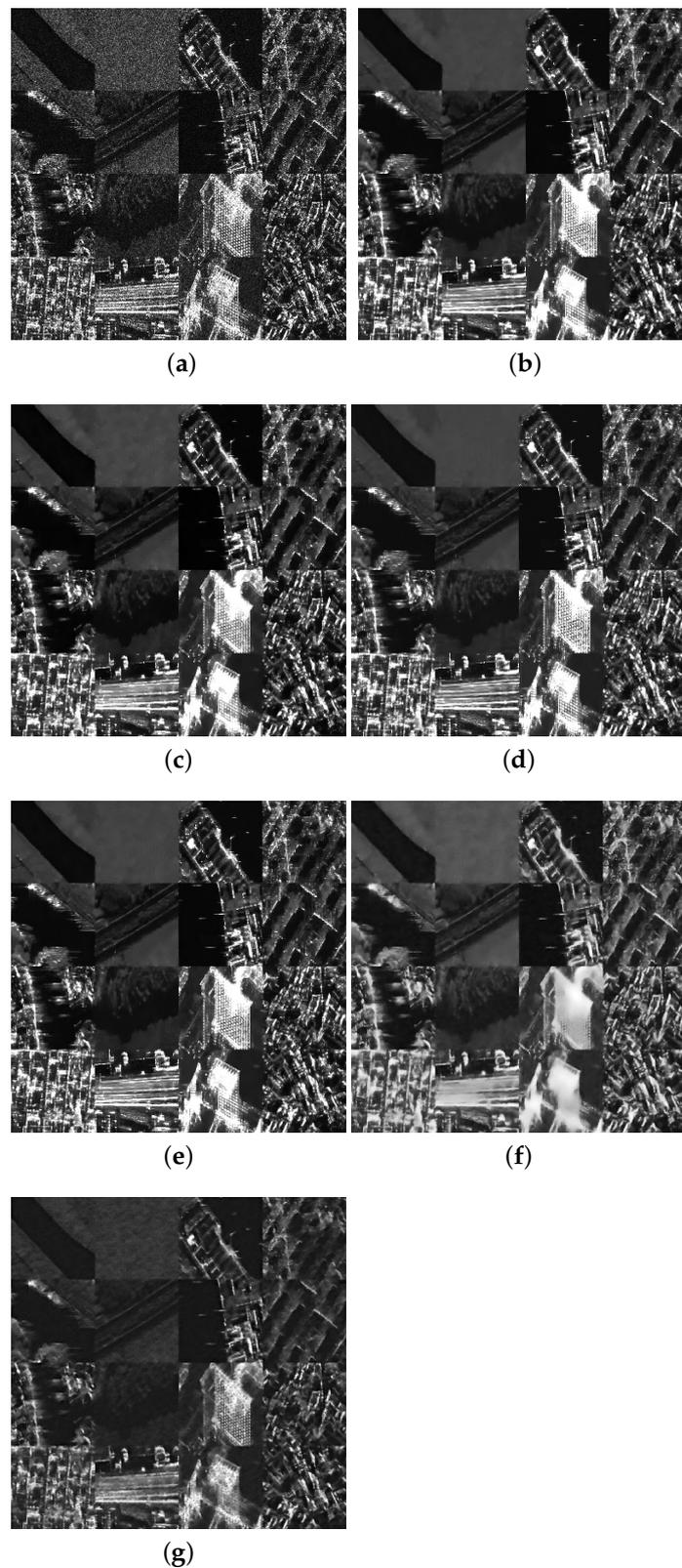


Figure 16. Real SAR images. (a) Mosaic of SAR images, 800×800 pixels in size. (b) Despeckled using the SDRCNN method. (c) Despeckled using the ABCNN method. (d) Despeckled using the SARBM3D method. (e) Despeckled using the DCNN method. (f) Image despeckled using the OCNN method. (g) Image despeckled using the SAR-CAM method.

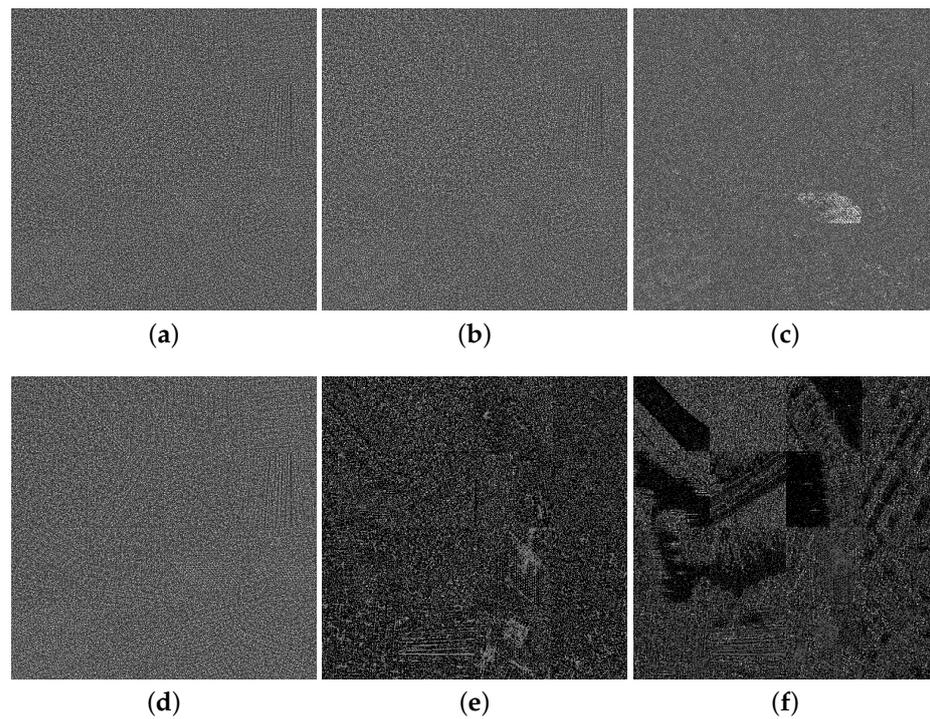


Figure 17. Ratio images between the original SAR image shown in Figure 16a and the despeckled images. (a) Original SAR image compared to the despeckled image using the SDRCNN method shown in Figure 16b. (b) Original SAR image compared to the despeckled image using the ABCNN method shown in Figure 16c. (c) Original SAR image compared to the despeckled image using the SARBM3D method shown in Figure 16d. (d) Original SAR image compared to the despeckled image using the DCNN method shown in Figure 16e. (e) Original SAR image compared to the despeckled image using the OCNN method shown in Figure 16f. (f) Original SAR image compared to the despeckled image using the SAR-CAM method shown in Figure 16g.

Table 6. Objective measurements of the quality criteria for the mosaic SAR image shown in Figure 16a.

Method	MoI	MoR	VoR	ENL	C_x
Original image	182.9	-	-	-	1.93
SDRCNN	182.9	0.99	3.13	488.2	1.92
ABCNN	182.9	0.99	3.05	489.2	1.92
SARBM3D	186.7	1.01	3.72	468.7	1.26
DCNN	182.9	0.99	3.02	476.1	1.91
OCNN	153.6	0.83	0.46	588.2	0.707
SAR-CAM	159.5	0.99	0.54	520.1	0.7059

The high-resolution SAR image used for the proposed despeckling methods' evaluation, which is shown in Figure 18a, was obtained from the TerraSAR-X satellite in Spotlight mode with a 300 MHz chirp bandwidth. Its size is 1024×1024 pixels and it features both a forest and an urban industrial area. The despeckled images are shown in Figure 18b–g for the proposed SDRCNN, ABCNN, SARBM3D, DCNN, OCNN, and SAR-CAM methods, respectively. The SDRCNN and ABCNN methods removed speckle effectively while preserving the integrity of the strong scatterers. The ratio images in Figure 19a,b compare the original image and its SDRCNN and ABCNN despeckled counterparts, respectively. The ratio images of the SDRCNN and ABCNN methods demonstrated ENL rates of 1.1 and 1.09 and mean values of 0.995 and 0.994, respectively. The ratio images in Figure 19c,d compare the original image and its SARBM3D and DCNN despeckled counterparts, respectively. The SARBM3D method could not handle correlated speckles effectively, as evidenced by

the associated despeckled and ratio images. The proposed DCNN methods proved highly adept at estimating speckles in high-resolution SAR images accurately while preserving the edge sharpness in homogeneous areas. The OCNN and SAR-CAM methods cannot be compared with the other methods presented in this paper, because the OCNN method over-smoothed the image and the SAR-CAM method underestimated the speckle, which was still noticeable within the homogeneous areas. The point scatterers and textural features were preserved well by the SAR-CAM method. The OCNN and SAR-CAM methods failed to produce comparable results in the objective measurements presented in Table 6.

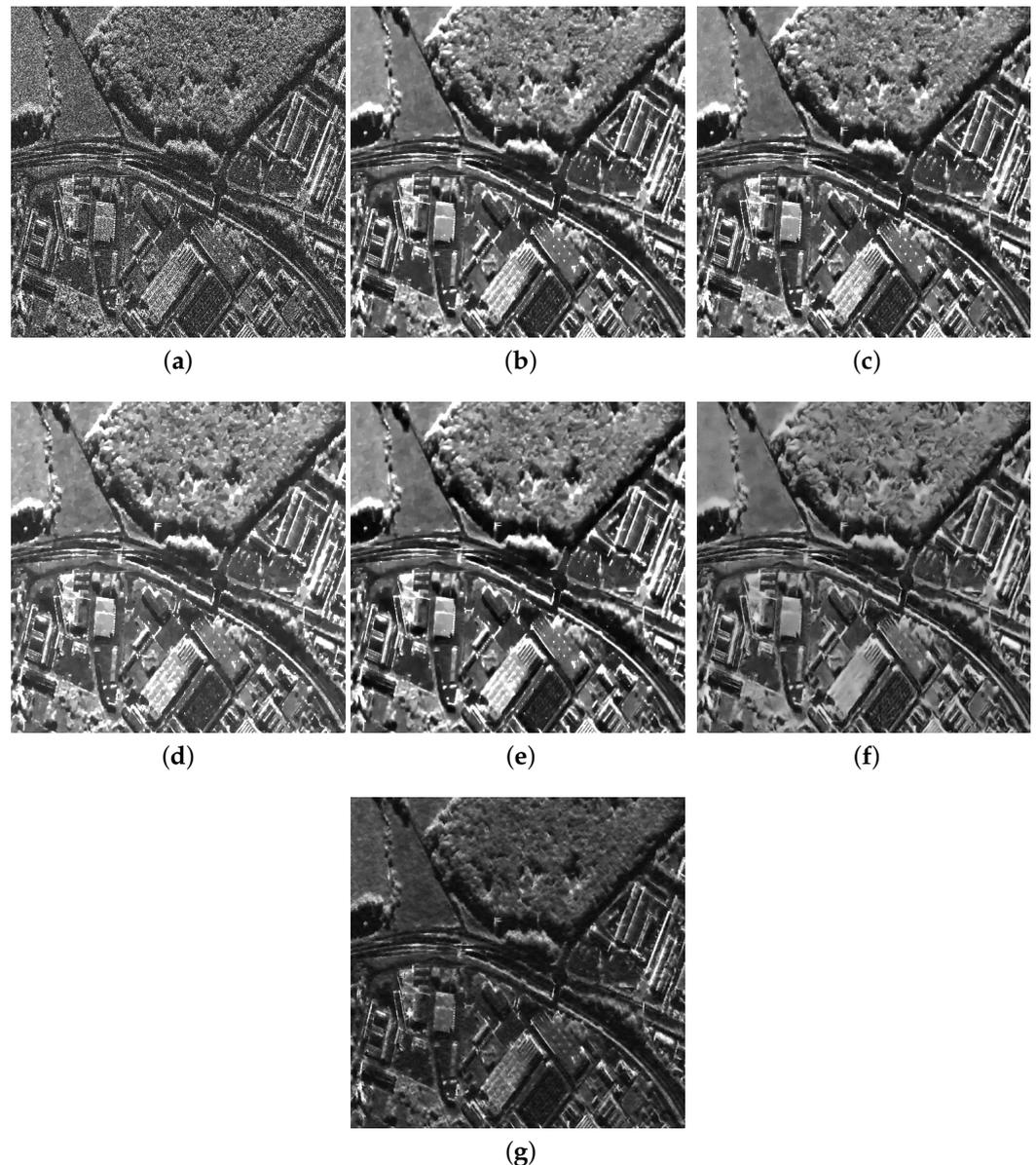


Figure 18. Real SAR image. (a) Original SAR image ©DLR 2012 (1024 × 1024 pixels). (b) Despeckled using the SDRCNN method. (c) Despeckled using the ABCNN method. (d) Despeckled using the SARBM3D method. (e) Despeckled using the DCNN method. (f) Image despeckled using the OCNN method. (g) Image despeckled using the SAR-CAM method.

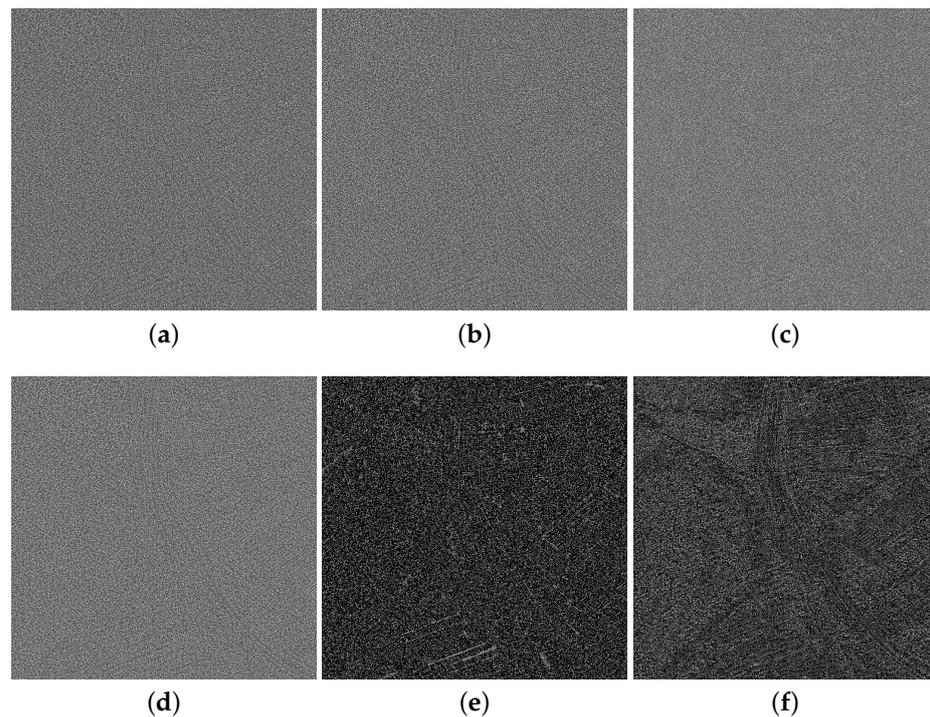


Figure 19. Ratio images between original the SAR image shown in Figure 18a and the despeckled images. (a) Original SAR image compared to the despeckled image using the SDRCNN method shown in Figure 18b. (b) Original SAR image compared to the despeckled image using the ABCNN method shown in Figure 18c. (c) Original SAR image compared to the despeckled image using the SARBM3D method shown in Figure 18d. (d) Original SAR image compared to the despeckled image using the DCNN method shown in Figure 18e. (e) Original SAR image compared to the despeckled image using the OCN method shown in Figure 18f. (f) Original SAR image compared to the despeckled image using the SAR-CAM method shown in Figure 18g.

The high-resolution SAR image used for the proposed despeckling methods' evaluation, which is shown in Figure 18a, was obtained from the TerraSAR-X satellite in Spotlight mode with a 30 MHz chirp bandwidth. Its size is 1024×1024 pixels and it features both a forest and an urban industrial area. The despeckled images are shown in Figure 18b–e for the proposed SDRCNN, ABCNN, SARBM3D, and DCNN methods, respectively. The SDRCNN and ABCNN methods removed speckle effectively while preserving the integrity of the strong scatterers. The ratio images in Figure 19a,b compare the original image and its SDRCNN and ABCNN despeckled counterparts, respectively. The ratio image of the SDRCNN and ABCNN methods demonstrated ENL rates of 1.1 and 1.09 and mean values of 0.995 and 0.994, respectively. The ratio images in Figures 19c,d compare the original image and its SARBM3D and DCNN despeckled counterparts, respectively. The SARBM3D method could not effectively handle correlated speckle, as evidenced by the associated despeckled and ratio images. The analysis of the despeckling quality was summarized using the objective measurements and is presented in Table 7. The MoI, MoR, and VoR averages for the entire SAR image are presented. The ENL was estimated from the homogeneous area of the SAR image. All the CNN-based methods proved highly adept at accurately estimating speckle in high-resolution SAR images while preserving edge sharpness in homogeneous areas. The results in Table 7, the despeckled images shown in Figure 18f,g, and the ratio images shown in Figure 19e,f clearly indicate that the OCN and SAR-CAM methods introduced biases and over-smoothed the original SAR image.

Table 7. Objective measurements of the quality criteria for the SAR high-resolution image shown in Figure 18a.

Method	MoI	MoR	VoR	ENL	C_x
Original image	32.62	-	-	-	-
SDRCNN	32.46	1.001	1.36	318.1	-
ABCNN	32.45	1.001	1.35	319.8	-
SARBM3D	36.69	0.86	1.75	336.8	-
DCNN	32.50	0.99	1.37	322.2	-
OCNN	46.6	0.85	0.28	423.1	-
SAR-CAM	25.8	1.18	0.46	488.3	-

5. Discussion

In this paper, we propose two methods based on a CNN to improve the overall performance of existing methods. The proposed methods differ in their network structures. The first one exploits a Siamese structure with a DRN network. This method introduces a minor novelty over existing approaches because it “just” combines two existing methods. The second proposed method uses recent trends in object detection and combines a U-shaped network with a multi-resolution attention mechanism. The methods were compared with the overcomplete CNN and SAR-CAM methods. The SAR-CAM method uses some of the components within the proposed network’s structure. We chose synthetic and real SAR images to compare the despeckling efficiency of five methods. The results obtained from the synthetically generated images showed that the OCNN and SAR-CAM methods can estimate speckle noise, but they are not capable of preserving high SAR dynamics such as point scatterers, edges, and changes of contrast. They could not preserve textural features. The reason for the very bad efficiency of the OCNN and SAR-CAM methods is that the authors of these methods converted SAR images into 8-bpp images and decreased the SAR image dynamics. Therefore, the methods do not have any practical value, except for competing in PSNR with existing methods. Considering the examples using synthetically generated images, we can conclude that the homogeneous areas were over-smoothed and some bias in the SAR image’s mean value was introduced by the OCNN and SAR-CAM methods. This may be a consequence of the fact that we did not scale the presented images to an 8-bit dynamic. The SARBM3D, SNN, and ABCNN methods achieved very similar results in the objective and subjective measurements, as reported in Tables 1–7. The methods achieved different despeckling efficiencies due to their different network structures. Also important is the training of the network so that it includes diverse SAR features.

When used to despeckle the real SAR image, the proposed methods achieved results that were comparable to those of the SARBM3D method, more successfully modeled homogeneous areas and estimated speckle noise in homogeneous areas, and effectively preserved textural features. The OCNN and SAR-CAM methods over-smoothed the real SAR image, destroying the point features and textural features, as shown in the Experimental Results section.

Further work could include recent advances in SAR image processing, where the attention mechanism can have different structures and can be exploited by transformer-based structures. Recent advances have shown that the standard approaches to adaptive mean weighting and nonlocal means can be combined successfully with CNN-based methods. To achieve improvements in efficiency, it is important to find a mechanism that is able to assign higher importance to the strong features and restrain the non-important ones.

6. Conclusions

This paper proposed and evaluated two methods for SAR image despeckling. In the first method, the dilated convolutional network for SAR image despeckling was extended using a twin network to better estimate and preserve SAR image features. The second

method combined a residual network and U-net and additionally enhanced the image features using a multi-resolution attention mechanism containing a Multiscale Attention Mechanism with ECA. The methods were designed differently and used different loss functions but were trained on the same datasets. The datasets used for training and validation consisted of a dataset containing synthetic images with speckle noise and a multi-temporal dataset containing images from the Sentinel 1, ALOS-2, and TerraSAR-X satellites. This paper demonstrated that although both proposed methods have different architectures, they both provide very good results in measurements of despeckled image quality. The methods were compared with the SARBM3D and dilated CNN methods. All the methods could effectively deal with correlated speckle noise, except for the SARBM3D method, which could not estimate this noise in the case of real SAR images. According to the experimental evaluation, the suggested approaches can effectively assess speckle and they exhibit superior capabilities over conventional techniques and DCNNs.

Author Contributions: Conceptualization, B.P. and D.G.; methodology, B.P. and D.G.; software, B.P. and D.G.; validation, B.P. and D.G.; formal analysis, D.G.; investigation, B.P.; resources, B.P. and D.G.; data curation, B.P.; writing—original draft preparation, D.G.; writing—review and editing, B.P.; visualization, B.P.; supervision, D.G.; project administration, D.G.; funding acquisition, D.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Slovenian Research Agency (ARRS) research program number P2-0065.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

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

Abbreviations

The following abbreviations are used in this manuscript:

SAR	Synthetic Aperture Radar
MAP	Maximum a Posteriori
PDF	Probability Density Function
DWT	Discrete Wavelet Transform
SARBM3D	SAR block-matching three-dimensional
CNN	Convolutional Neural Network
AWGN	Additive White Gaussian Noise
ReLU	Rectifier Linear Unit
DCNN	Denosing Convolutional Neural Network
BN	Batch Normalization
ID-CNN	Image Despeckling Convolutional Neural Network
SAR-DRN	SAR Dilated Residual Network
TV	Total Variation
SNN	Siamese Neural Network
EL	Euclidean loss
DRN	Dilated Residual Network
ASN	Attention Supervision Network
MAM	Multi-resolution Attention Mechanism
ECA	Efficient Channel Attention
SDRCNN	Siamese-based Dilated Residual Convolutional Neural Network
ABCNN	Attention-Based CNN
DEM	Digital Elevation Model
VoR	variance of ratio
MoR	mean of ratio
DG	despeckling gain
ENL	equivalent number of looks

ES	Edge Smearing
FOM	Figure of Merit
BS	Building Smearing
OCNN	overcomplete convolutional neural network
SAR-CAM	SAR image despeckling using a continuous attention module

References

- 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. [\[CrossRef\]](#)
- Lee, J. Digital image enhancement and noise filtering by using local statistics. *IEEE Trans. Pattern Anal. Mach. Intell.* **1980**, *PAMI-2*, 165–168. [\[CrossRef\]](#)
- Kuan, D.T.; Sawchuk, A.A.; Strnad, T.C.; Chavel, P. Adaptive noise smoothing filter for Images with Signal dependent noise. *IEEE Trans. Pattern Anal. Mach. Intell.* **1985**, *7*, 165–177. [\[CrossRef\]](#) [\[PubMed\]](#)
- Frost, V.; Stiles, J.; Shanmugan, K.; Holtzman, J. A model for radar images and its application to adaptive digital filtering of multiplicative noise. *IEEE Trans. Pattern Anal. Mach. Intell.* **1982**, *PAMI-4*, 157–165. [\[CrossRef\]](#) [\[PubMed\]](#)
- Lee, J.S.; Wen, J.H.; Ainsworth, T.; Chen, K.S.; Chen, A. Improved Sigma Filter for Speckle Filtering of SAR Imagery. *IEEE Trans. Geosci. Remote Sens.* **2009**, *47*, 202–213.
- Molina, D.E.; Gleich, D.; Datcu, M. Evaluation of Bayesian Despeckling and Texture Extraction Methods Based on Gauss Markov and Auto-Binomial Gibbs Random Fields: Application to TerraSAR-X Data. *IEEE Trans. Geosci. Remote Sens.* **2012**, *50*, 2001–2025. [\[CrossRef\]](#)
- Argenti, F.; Bianchi, T.; Lapini, A.; Alparone, L. Fast MAP Despeckling Based on Laplacian Gaussian Modeling of Wavelet Coefficients. *IEEE Geosci. Remote Sens. Lett.* **2012**, *50*, 13–17. [\[CrossRef\]](#)
- 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. [\[CrossRef\]](#)
- Deledalle, C.A.; Denis, L.; Tupin, F. Iterative Weighted Maximum Likelihood Denoising with Probabilistic Patch-Based Weights. *IEEE Trans. Image Process.* **2009**, *18*, 2661–2672. [\[CrossRef\]](#)
- Martino, G.D.; Simone, A.D.; Iodice, A.; Poggi, G.; Riccio, D.; Verdoliva, L. Scattering-Based SARBM3D. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2016**, *9*, 2131–2144. [\[CrossRef\]](#)
- Martino, G.D.; Poderico, M.; Poggi, G.; Riccio, D.; Verdoliva, L. Benchmarking Framework for SAR Despeckling. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 1596–1615. [\[CrossRef\]](#)
- Hinton, G.E.; Salakhutdinov, R.R. Reducing the Dimensionality of Data with Neural Networks. *Science* **2006**, *313*, 504–507. [\[CrossRef\]](#)
- Diao, W.; Sun, X.; Zheng, X.; Dou, F.; Wang, H.; Fu, K. Efficient Saliency-Based Object Detection in Remote Sensing Images Using Deep Belief Networks. *IEEE Geosci. Remote Sens. Lett.* **2016**, *13*, 137–141. [\[CrossRef\]](#)
- Zhang, K.; Zuo, W.; Chen, Y.; Meng, D.; Zhang, L. Beyond a Gaussian Denoiser: Residual Learning of Deep CNN for Image Denoising. *IEEE Trans. Image Process.* **2017**, *26*, 3142–3155. [\[CrossRef\]](#)
- Shi, W.; Caballero, J.; Huszár, F.; Totz, J.; Aitken, A.P.; Bishop, R.; Rueckert, D.; Wang, Z. Real-Time Single Image and Video Super-Resolution Using an Efficient Sub-Pixel Convolutional Neural Network. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 1874–1883.
- Dong, C.; Loy, C.C.; He, K.; Tang, X. Image super-resolution using deep convolutional networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2016**, *38*, 295–307. [\[CrossRef\]](#) [\[PubMed\]](#)
- He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. *arXiv* **2016**, arXiv:1512.03385.
- Ioffe, S.; Szegedy, C.B.N. Accelerating Deep Network Training by Reducing Internal Covariate Shift. *arXiv* **2016**, arXiv:1502.03167.
- Chierchia, G.; Cozzolino, D.; Poggi, G.; Verdoliva, L. SAR image despeckling through convolutional neural networks. In Proceedings of the 2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Fort Worth, TX, USA, 23–28 July 2017; pp. 5438–5441.
- Wang, P.; Zhang, H.; Patel, V. SAR Image Despeckling Using a Convolutional Neural Network. *arXiv* **2017**, arXiv:1706.00552v2.
- Zhang, Q.; Yang, Z.; Yuan, Q.; Li, J.; Ma, X.; Shen, H.; Zhang, L. Learning a Dilated Residual Network for SAR Image Despeckling. *arXiv* **2017**, arXiv:1709.02898.
- Lattari, F.; Leon, B.G.; Asaro, F.; Rucci, A.; Prati, C.; Matteucci, M. Deep Learning for SAR Image Despeckling. *Remote Sens.* **2019**, *11*, 1532. [\[CrossRef\]](#)
- Cozzolino, D.; Verdoliva, L.; Scarpa, G.; Poggi, G. Nonlocal CNN SAR Image Despeckling. *Remote Sens.* **2020**, *12*, 1006. [\[CrossRef\]](#)
- Liu, G.; Kang, H.; Wang, Q.; Tian, Y.; Wan, B. Contourlet-CNN for SAR Image Despeckling. *Remote Sens.* **2021**, *13*, 764. [\[CrossRef\]](#)
- Mullissa, A.G.; Persello, C.; Reiche, J. Despeckling Polarimetric SAR Data Using a Multi-Stream Complex-Valued Fully Convolutional Network. *arXiv* **2021**, arXiv:2103.07394.
- Wang, F.; Jiang, M.; Qian, C.; Yang, S.; Li, C.; Zhang, H.; Wang, X.; Tang, X. Residual Attention Network for Image Classification. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 6450–6458. [\[CrossRef\]](#)

27. Dai, T.; Cai, J.; Zhang, Y.; Xia, S.T.; Zhang, L. Second-Order Attention Network for Single Image Super-Resolution. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019; pp. 11057–11066. [[CrossRef](#)]
28. Li, J.; Li, Y.; Xiao, Y.; Bai, Y. HDRANet: Hybrid Dilated Residual Attention Network for SAR Image Despeckling. *Remote Sens.* **2019**, *11*, 2921. [[CrossRef](#)]
29. Liu, Z.; Lai, R.; Guan, J. Spatial and Transform Domain CNN for SAR Image Despeckling. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 4002005. [[CrossRef](#)]
30. Liu, S.; Lei, Y.; Zhang, L.; Li, B.; Hu, W.; Zhang, Y.D. MRDDANet: A Multiscale Residual Dense Dual Attention Network for SAR Image Denoising. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5214213. [[CrossRef](#)]
31. Shen, H.; Zhou, C.; Li, J.; Yuan, Q. SAR Image Despeckling Employing a Recursive Deep CNN Prior. *IEEE Trans. Geosci. Remote Sens.* **2021**, *59*, 273–286. [[CrossRef](#)]
32. Wu, W.; Huang, X.; Shao, Z.; Teng, J.; Li, D. SAR-DRDNet: A SAR image despeckling network with detail recovery. *Neurocomputing* **2022**, *493*, 253–267. [[CrossRef](#)]
33. Singh, P.; Shankar, A.; Diwakar, M. MSPB: Intelligent SAR despeckling using wavelet thresholding and bilateral filter for big visual radar data restoration and provisioning quality of experience in real-time remote sensing. In *Environment, Development and Sustainability*; Springer: Berlin/Heidelberg, Germany, 2022; pp. 1573–2975.
34. Perera, M.V.; Nair, N.G.; Bandara, W.G.C.; Patel, V.M. SAR Despeckling Using a Denoising Diffusion Probabilistic Model. *IEEE Geosci. Remote Sens. Lett.* **2023**, *20*, 4005305. [[CrossRef](#)]
35. Perera, M.V.; Bandara, W.G.C.; Valanarasu, J.M.J.; Patel, V.M. SAR Despeckling Using Overcomplete Convolutional Networks. *arXiv* **2022**, arXiv:2205.15906
36. Perera, M.V.; Bandara, W.G.C.; Valanarasu, J.M.J.; Patel, V.M. Transformer-based SAR Image Despeckling. *arXiv* **2022**, arXiv:2201.09355.
37. Ko, J.; Lee, S. SAR Image Despeckling Using Continuous Attention Module. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2022**, *15*, 3–19. [[CrossRef](#)]
38. Yuan, Y.; Wu, Y.; Tang, C.; Fu, Y.; Wu, Y.; Jiang, Y.; Zhao, Y. Self-calibrated dilated convolutional neural networks for SAR image despeckling. *Int. J. Remote Sens.* **2022**, *43*, 6483–6508. [[CrossRef](#)]
39. Wang, C.; Yin, Z.; Ma, X.; Yang, Z. SAR Image Despeckling Based on Block-Matching and Noise-Referenced Deep Learning Method. *Remote Sens.* **2022**, *14*, 931. [[CrossRef](#)]
40. Goodman, J. Some fundamental properties of speckle. *J. Opt. Soc. Am.* **1976**, *66*, 1145–1150. [[CrossRef](#)]
41. Kingma, D.P.; Ba, J. Adam: A Method for Stochastic Optimization. *arXiv* **2014**, arXiv:1412.6980.
42. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Munich, Germany, 5–9 October 2015; pp. 234–241.
43. Bromley, J.; Guyon, I.; Lecun, Y.; Säckinger, E.; Shah, R. Signature verification using a siamese time delay neural network. *Int. J. Pattern Recognit. Artif. Intell.* **1993**, *7*, 217–222. [[CrossRef](#)]
44. Wang, Q.; Wu, B.; Zhu, P.F.; Li, P.; Zuo, W.; Hu, Q. ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020 ; pp. 11531–11539.
45. Zhao, H.; Gallo, O.; Frosio, I.; Kautz, J. Loss Functions for Image Restoration With Neural Networks. *IEEE Trans. Comput. Imaging* **2017**, *3*, 47–57. [[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.