



Article

SAR Image Despeckling Based on Block-Matching and Noise-Referenced Deep Learning Method

Chen Wang^{1,2,3,4} , Zhixiang Yin^{1,2,3,4,*} , Xiaoshuang Ma^{1,2,3,4} and Zhutao Yang⁵

¹ School of Resources and Environmental Engineering, Anhui University, Hefei 230601, China; chen.wang@ahu.edu.cn (C.W.); mxs.88@whu.edu.cn (X.M.)

² Information Materials and Intelligent Sensing Laboratory of Anhui Province, Anhui University, Hefei 230601, China

³ Anhui Province Key Laboratory of Wetland Ecosystem Protection and Restoration, Anhui University, Hefei 230601, China

⁴ Engineering Center for Geographic Information of Anhui Province, Hefei 230601, China

⁵ Central Southern China Electric Power Design Institute Co., Ltd., Wuhan 430071, China; yzt5651@csepedi.com

* Correspondence: yinzx@ahu.edu.cn; Tel.: +86-15256595329

Abstract: The noise2noise-based despeckling method, capable of training the despeckling deep neural network with only noisy synthetic aperture radar (SAR) image, has presented very good performance in recent research. This method requires a fine-registered multi-temporal dataset with minor time variance and uses similarity estimation to compensate for the time variance. However, constructing such a training dataset is very time-consuming and may not be viable for a certain practitioner. In this article, we propose a novel single-image-capable speckling method that combines the similarity-based block-matching and noise referenced deep learning network. The denoising network designed for this method is an encoder–decoder convolutional neural network and is accommodated to small image patches. This method firstly constructs a large number of noisy pairs as training input by similarity-based block-matching in either one noisy SAR image or multiple images. Then, the method trains the network in a Siamese manner with two parameter-sharing branches. The proposed method demonstrates favorable despeckling performance with both simulated and real SAR data with respect to other state-of-the-art reference filters. It also presents satisfying generalization capability as the trained network can despeckle well the unseen image of the same sensor. The main advantage of the proposed method is its application flexibility. It could be trained with either one noisy image or multiple images. Furthermore, the despeckling could be inferred by either the ad hoc trained network or a pre-trained one of the same sensor.

Keywords: block matching; convolutional neural network (CNN); deep learning; speckle filtering; synthetic aperture radar (SAR); image similarity



Citation: 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. <https://doi.org/10.3390/rs14040931>

Academic Editor: Jean-Christophe Cexus

Received: 20 December 2021

Accepted: 11 February 2022

Published: 14 February 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 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

As a coherent system, synthetic aperture radar (SAR) is naturally affected by speckles, which is a random structure caused by the interference of waves scattered from detection objects. Speckle is a real SAR measurement that brings information about the detection object. However, its uncertainty, variance, and noise-like appearance affect the image interpretation, making despeckling indispensable. Despeckling research has been constantly driven by new image processing development. Recently, deep learning methods have shown very promising results. The latest developments of supervised deep learning denoising methods, such as denoising convolutional neural network (DnCNN) [1] and fast and flexible denoising convolutional neural network (FFDNet) [2], have improved the denoising results of ordinary images compared with the traditional methods. Inspired by this, many researchers thus then have tried to apply supervised deep learning methods to SAR despeckling. Such a method requires corresponding clean images as references to help

the deep model learn the mapping from noisy image to denoised image. However, there is no “clean” SAR image. To solve this problem, some researchers have tried to build an approximate clean reference from noisy images by using the multi-look fusion [3]. On the other hand, other researchers have used the optical image or pre-denoised multi-temporal fusion images as clean references and calculated simulated noisy images for training [4]. However, both the approximate clean reference method and simulated noise reference method are problematic. Multi-look fusion cannot achieve comparable denoising results of the state-of-art denoising filter, thus greatly limiting the denoising potential of the deep model. Simulating noise reference, on the other hand, creates a domain gap [5], which causes despeckling performance mismatch between the synthesized image and the real SAR image.

Due to the lack of clean ground truth, the self-supervised deep learning despeckling method, which requires no clean reference, is an appealing research direction. Researchers in the image denoising domain have proposed multiple self-supervised deep learning denoising methods. Some of these methods require only a single noisy image to train the network, such as deep image prior [6], noise2void [7], and noise2self [8]. The application to SAR despeckling has just begun. Currently, various blind-spot methods have been proposed, which evaluate the value of a pixel by its surrounding pixels as the receptive field. By using the SAR image statistics, Molini et al. [9] adapted Laine et al.’s [10] model to SAR despeckling and achieved a fair result. Joo et al. [11] extended the neural adaptive image denoiser (N-AIDE) for multiplicative noise, and their results out-performed dilated residual network-based SAR filter (SAR-DRN) [12] on a simulated SAR dataset.

Using only one noisy image is valuable when the independent noise realization pair is not available for training. On the other hand, using multiple noise realization of the same ground truth may have the potential to produce better denoising results, as multiple noise realizations provide the filter with richer information concerning the data distribution and the noise feature. Lehtinen et al. [13] proposed a noise2noise method, which trains a deep learning model with multiple noisy image pairs that refer to the same clean target and follow an independent noise distribution. Directly applying the noise2noise method to multi-temporal SAR despeckling is problematic, as the multi-temporal SAR image of the same region cannot be regarded as independent noise realizations of the same clean image without the ground truth concerning the temporal variance. Addressing this problem, Dalsasso et al. [14] used a simulated-based and supervised trained despeckling model to acquire pre-estimations of the despeckled images. Then, they used the difference between the pre-estimations to compensate for the temporal change and by this way formalized a series of noise2noise training images pairs. Their model can be evaluated as a hybrid of supervised and self-supervised, thus may not be fully domain gap free. In our previous work [15], we proposed a fully self-supervised noise2noise SAR despeckling method (NR_SAR_DL) that uses bi-temporal SAR amplitude image blocks of the same region as training pairs and weights the training loss by the estimated pixel similarity between the training images. This method yields state-of-arts results.

The similarity loss-weighting training employed in the training process in [15] can only reduce but not fully eliminate the influence of the temporal variance, as no similarity lower bound was employed for both the training image pair and its pixel pair. As a result, the noisy images with major changes in their ground truths are not excluded from the training process and thus may taint the neural network. The influence is minor when the time variance is limited but evident when the variance is significant. Another drawback is that the multi-temporal SAR images used for training should achieve a fine subpixel registration, which is often a very time-consuming task.

Extending the original method of [15] by setting a threshold for the similarity of the training image block pair shows promising results to improve the overall despeckling performance. Such a method answers the question of whether an image pair contains two independent noise realizations of the same clean image by two quantitative parameters: an

estimation of their similarity value based on the statistical feature, and the selection of the similarity threshold that separates the similar and the non-similar.

Some prominent similarity, like the one used in NR_SAR_DL [15], does not consider the location and the time. Thus, similar block pairs can be selected from not only those time sequence images that refer to the same location as in the method of [15] but also those of different locations and different times. Inspired by the irrelevance of location and time in the similarity calculation, we developed a similar Block Matching and Noise-Referenced Deep Learning network filter (BM-NR-DL). The main idea of this method is to search and construct a large number of image block pairs that are deemed similar by their similarity estimation and then train the deep learning despeckling network with them. It uses only real SAR images to train the deep model and could be fully self-supervised. Compared with other self-supervised methods, the proposed method has two main innovations:

(1) The proposed method in this article for the first time in SAR despeckling combines the self-supervised noise2noise method with the similarity-based block-matching method. The block can be searched in a single SAR image or multiple images, thus providing a uniform yet flexible training solution. Furthermore, when training on time sequence images of the same region, the images no longer need to be fine registered.

(2) The U-network [16] style model employed in both the original noise2noise [13] and the NR_SAR_DL solution [15] requires a minimal size of image blocks for training, such as 64×64 . However, large image blocks impede the block-matching procedure. The proposed method inserts a transposed convolutional layer before the encoder and minimizes the encoding layer number without damaging the despeckling performance. The model can now be trained with a block size significantly smaller than those of [13] and [15], making block-matching possible.

The test with both simulated and real SAR single-look intensity data shows that the proposed method outperforms other state-of-art despeckling methods. The proposed method has great application flexibility, as it can be trained with either one SAR image or multiple images. The method also presents good generalization capability. The network trained using SAR images acquired by a certain sensor can readily despeckle other images acquired by the same sensor. Moreover, the proposed method has good scalability in both the training and despeckling process. For training, due to the training flexibility provided by this method, the set of image pairs used in the training process could be readily scaled up. For despeckling, the deep learning inference process is much more computationally efficient than other traditional methods, such as block-matching and 3D filtering (BM3D) and probabilistic block-based weights (PPB). The despeckling process can be easily scaled up taking advantage of the parallel computing.

The article is organized as follows: Section 2 demonstrates the related research. Section 3 describes the main feature of the proposed method. Section 4 presents the validation of the proposed model and the comparison with other despeckling methods. Section 5 discusses the performance of the proposed model and points to potential future research directions. Section 6 concludes the research.

2. Related Work

Deep Learning and SAR despeckling: Denoising is one of the image restoration processes whose purpose is to recover the clean image from its noisy observation. The noisy observation can be expressed as a function of the signal and corruption:

$$\hat{X} = f(X, n) \quad (1)$$

where \hat{X} is the noisy observation, X is the clean image, and n is the corruption. The function $f(\cdot)$ is largely domain specific. In SAR despeckling, it is often described as a multiplicative noise function in most circumstances and noise-free in extreme intensity conditions. The ill-posed problem of such a process requires image priors to limit and regulate the solution space. Many explicitly, handcraft priors, such as total variant (TV) [17], Markov random fields (MRFs) [18], and nonlocal means (NLM) [19], are popular among the state-of-art

denoising algorithms. The classical SAR despeckling methods, including Lee filter [20], Kuan filter [21], Frost filter [22], PPB [23], and SAR-BM3D [24], also use handcraft priors. However, the features of both the noise and the original image are complex and may be dedicated by the application instance, thus impeding the construction of widely applicable yet precise priors. Ma et al. [25] presented variational methods by introducing non-local regularization functions. Deep learning methods, on the other hand, implicate priors by their network structures with performances better than those with explicit priors [6]. Meanwhile, deep learning methods can learn priors from data with proper training [26]. The implicit priors with learnable priors together may explain the better despeckling results of the deep learning methods than the traditional ones, which are demonstrated in recent studies [12,27].

The deep learning despeckling often consists of a training process and an inference process. The training process aims to find the parameter set θ of the network $f_{\theta}(\cdot)$ that minimizes the sum of the loss calculated by loss function L between the network output image and its corresponding training reference

$$\operatorname{argmin}_{\theta} \sum_i^M L[f_{\theta}(\hat{X}_i), R_i] \quad (2)$$

where M is the number of training samples, and R_i is the i -th training reference. Due to the large number of training samples used in the learning process, each optimization step often employs only a small subset of the training samples, achieving stochastic gradient descent of loss. Once the network is well trained, it infers the clean estimation X' from noisy images \hat{X} .

$$X' = f_{\theta}(\hat{X}) \quad (3)$$

The two main factors that influence the despeckling performance are the network structure, which dedicates the handcraft prior, and the learning method, which dedicates the learnable prior. Currently, the most common network structure employed for SAR despeckling is the convolutional neural network (CNN) with an encoder–decoder structure [28], Siamese structure [15], N3 non-local layer [3], residual architecture [28], or Generative Adversarial Network [29]. The input of the network can be original SAR images [30] or their log transformation [31], and the output of the network can be the residual content, the denoised image, or the weights for further denoising process [3]. On the other hand, self-supervised learning has merged as a promising future trend, as it avoids the domain gap introduced by supervised learning methods, which employ simulated data in their training. The existing self-supervised methods require either one single noise image or multiple noisy image pairs. Compared with only a single noise instance, using multiple noisy images with independent noise distribution provides more information concerning the noise and the original signal. It may lead to better denoising performance, as demonstrated in [7].

Noise2Noise and SAR despeckling: For the deep learning denoising method that only uses noisy images, noise2noise [13] achieves state-of-the-art results. This method trains the model with a series of noisy image pairs, and each image pair contains two independent and zero-mean noisy realizations of the same scene, one as training input and one as training reference. The main obstacle to the application of the noise2noise SAR despeckling is the construction of the noisy SAR image pair that refers to the same ground truth, since the ground truth is not directly available. Different observations of the same land area cannot be deemed fully stationary, as the land feature is constantly changing. Addressing this problem, some researchers try to construct the image pairs by using synthesized speckled images [14,32,33], either directly for network training [32,33] or as an initial process [14] to evaluate and compensate for the change that occurs between multi-temporal images acquired over the same area. The training process using synthesized noisy images may inevitably bring in the domain gap. On the other hand, the NR-SAR-DL previously proposed by our research group [15] employs no synthesized noisy image. As

demonstrated in Figure 1, this method directly uses multi-temporal observations of the same area as the noise2noise training set. The denoising network is an encoder–decoder convolutional network with skip-connection and trained in a Siamese way with two branches. The method compensates the temporal change by weighting the training loss with the pixel level similarity estimation.

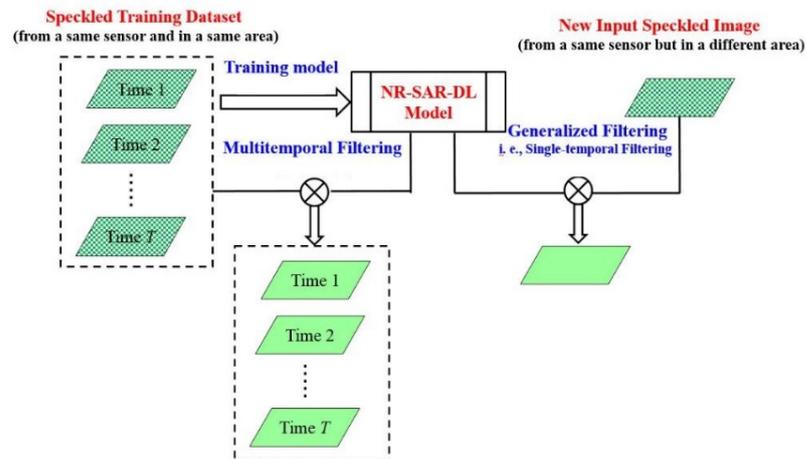


Figure 1. Basic framework of the NR-SAR-DL filter [15].

Block-matching in SAR despeckling algorithm: The self-similarity presented inside noisy images provides the possibility to construct a denoising algorithm by matching similar blocks and then using the nonlocal filter as presented in the BM3D [34]. As the speckled SAR image also presents self-similarity, there are several despeckle algorithms that are built on block-matching and have good performance, such as SAR-BM3D [24] and Fast adaptive nonlocal SAR (FANS) [35]. Block-matching can also be carried out among multi-temporal SAR images [36]. Block-matching has recently been embedded in some deep learning denoising methods [37–39] and despeckling methods [40,41]. However, the network involved in these methods still requires supervised training with clean references. Combining block-matching and self-supervised noise-referenced network training has only emerged in the latest pre-print with feasible results [42].

3. Methodology

Changing the training reference to the noisy image whose ground truth is different but similar can still fulfill the premise of noise2noise once the noise and the difference are zero-expectation and independent. Regarding the applicability of constructing noise2noise training pairs by using noisy images of similar blocks, we proposed a novel SAR despeckling method that combines block-matching and noise-referenced deep learning (BM-NR-DL). The basic idea and framework of this method are shown in Figure 2. This method trains the denoising network with a large number of noise2noise training samples constructed by similarity-based noisy-block-matching in a single SAR image or a group of SAR images. Firstly, the index blocks are randomly selected in the noisy image or the group of noisy images. Then their similar blocks are searched in the dataset to construct the noise2noise training pair with the index block. Finally, the network is trained in a Siamese style with two parameter-sharing branches. In each training step, one image in the image pair is the input of one branch and the reference of the other. The training procedure consists of two passes. The first pass uses the original noisy image to calculate the similarity and produces preliminary denoised images. The second pass then updates the similarity estimation with the information provided by the first pass and refines the denoising result. The network trained in the second pass produces the final denoising output. The principle of this method is to ensure that the ground-truth difference between the training input and reference is

small by setting a similarity threshold for block-matching, and the difference expectation is close to zero by switching the input and reference in a Siamese style training.

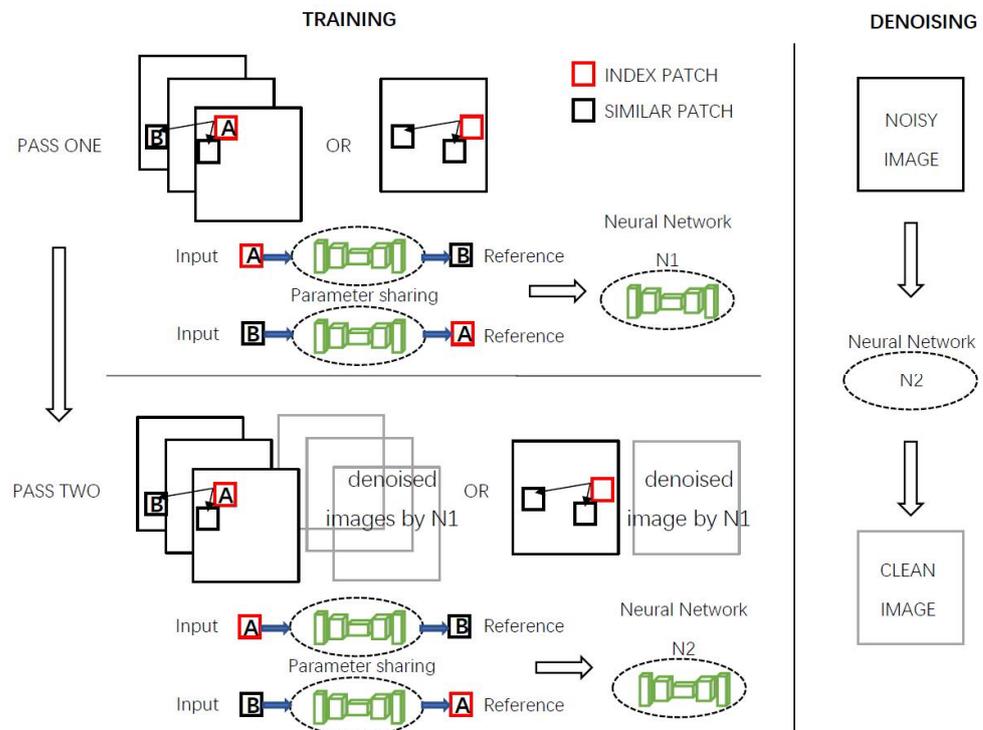


Figure 2. The workflow of the proposed BM-NR-DL method. Each noise2noise training pair contains an index patch A and a similar patch B. There are two training passes, and the second pass searches the similarity patches by considering the denoised result of the first pass. The similarity patch search could be carried out in either one noisy image or multiple ones.

3.1. Block Matching and Noise2noise Pairing

In each training pass, p index blocks are firstly randomly chosen from the noisy data, which could be a single SAR image or multiple SAR images of the same sensor. Then, for each index block, the method constructs a group of candidate similar blocks from the noisy data by k -th nearest neighbor search within these noisy images with a similarity estimation S and a similarity threshold of T .

For a randomly chosen index block, the mean of its k -th nearest neighbor blocks may not be zero, as it is influenced heavily by its domestic signal distribution, the similarity estimation S , the similarity threshold T , and the value of k . However, when calculating from the stochastic point of view with a small T , which is a certainty in a high-dimension k -th nearest neighbor (KNN) search, the probability of a specific block to appear in the index part is equal to its probability to appear in the neighbor part and thus may result in a zero-mean difference. However, when the sample size is limited, there could be an imbalance. Considering this, our method then establishes noise2noise training pairs by using a Siamese network structure. In the training process, we simultaneously set each index block as the input and each of its candidate similar neighbor as the reference and set each candidate similar neighbor as the input and the corresponding index block as the reference. Choosing the block size in a block matching process faces a dilemma: small block size is sensitive to noise, while big block size increases the similarity estimation error and computation burden. Some of the current block-matching despeckling methods choose a block size of 8×8 , such as that of [24,36]; this is similar to BM3D, whose optimized block-size ranges from 4 to 12 [43]. The block size used in the BM-NR-DL filter has a similar range as BM3D, but with a lower bound of 6.

3.2. Denoising Neural Network

The denoising neural network is based on the convolutional neural network and consists of an encoder part, which captures the image features from different levels of abstraction, and a decoder part with skip connections, which concatenates these features to establish the denoised output. Such an encoder–decoder structure is designed to first capture the texture information with the receptive field at different scales and then recover the image with the captured information. Previous investigation has proven that such a structure is a good image prior for low-level image statistics that can be used in various inverse problems such as denoising [6]. It also showed good performance in our previous deep learning denoising attempt [15]. Table 1 presents the detailed structure of the network. Compared with the ordinary U-net neural network, the encoder part of the proposed network consists of transpose convolutional layers, so the encoder part can deal with small input image blocks while still maintaining enough pooling layers. The proposed network can be trained with blocks as small as 6×6 . During the training, two branches of the denoising network are grouped into a Siamese network with parameter sharing. Each branch takes one image in the image pair as input and outputs the denoised image. The loss is the sum of the L2 loss between the denoised output of each branch with its noisy counterpart and the L2 loss between the two denoised outputs:

$$L(\theta, I_A, I_B, \lambda) = \|f_{\theta}(I_B), I_A\|_2 + \|f_{\theta}(I_A), I_B\|_2 + \lambda \|f_{\theta}(I_B), f_{\theta}(I_A)\|_2 \quad (4)$$

where θ is the denoising network parameter set, I_A and I_B represent the two images in an image pair, and λ weighs the relative importance of the loss within each branch and the loss between branches. The basic idea behind parameter-sharing is that the two images in the image pairs are independent noisy realization of similar ground truth. The network corresponding to each branch should achieve an overall optimization with each noisy image and produce similar output.

Table 1. Parameters of the denoising deep network.

Step	Feature _{out}	Function
Input	1	Input SAR image batch
Upinput	64	transposed convolution 4×4 , stride: 4
Conv1	256	2 steps convolution 3×3
Pool1	256	maxpooling 2×2
Conv2	256	convolution 3×3
Pool2	256	maxpooling 2×2
Bottleneck	256	convolution 3×3
Up1	256	Upsampling scale factor: 2
Cat1	512	Concatenate Pool1 output
Conv3	256	2 steps convolution 3×3
Up2	256	Upsampling scale factor: 2
Cat2	256 + 64	Concatenate UpInput output
Conv4	32	3 steps convolution 3×3
Conv5	1	convolution 3×3
AvgPool	1	average pool scale factor: 4

This method contains no batch normalization, since in our experiment, introducing batch normalization deteriorated the denoising performance. The activation function is leaky rectified linear unit [44] for all the convolution layers except the last one, which uses liner activation function. In the inference stage, only one branch of the trained network is needed to despeckle the SAR image, as each branch shares the same set of parameters.

3.3. Two-Pass Noise-Referenced Training

The proposed method contains two training passes. The first pass searches similar blocks with threshold T_1 and trains the denoising network N_1 based on the original cor-

rupted SAR images. The block similarity measure S_1 is designed for Gamma distributed data [23,24] for single-looked amplitude data

$$S_1 = \sum_k \log \left(\frac{x(s, h)}{x(r, h)} + \frac{x(r, h)}{x(s, h)} \right) \quad (5)$$

where x denotes amplitude of a pixel in the noisy block; s , r refer to the index pixel of the two compared blocks; and h is used to scan the entire block. The first pass then produces preliminary despeckled images for the next pass.

The second pass takes advantage of the preliminary despeckled images. Inspired by [23,24], it trains the network N_2 with similar block pairs from the original noisy image with threshold T_2 and similarity S_2 :

$$S_2 = \sum_k \left[\log \left(\frac{x(s, k)}{x(t, k)} + \frac{x(t, k)}{x(s, k)} \right) + \eta \frac{|x'(s, k) - x'(t, k)|^2}{x'(s, k)x'(t, k)} \right] \quad (6)$$

where x' denotes the amplitude of a pixel in the preliminary despeckled block produced by the first pass, and η denotes the weights of the preliminary despeckled image in the similarity measure. Finally, the denoising process uses N_2 to infer the clean estimation.

4. Experiments and Results

We conducted experiments on both the simulated and real SAR data to gauge the performance of the proposed despeckling method. The training was carried out with either a single SAR image or time sequence images of the same sensor. We compared the despeckling results of the proposed model with the following four state-of-the-art methods: for single image-based despeckling, the probabilistic block-based (PPB) filter [23], and the dilated residual network-based SAR (SAR-DRN) filter [12]; and for multitemporal image-based despeckling, the 3D block matching-based multitemporal SAR (MSAR-BM3D) filter [36], and the noise-referenced deep learning (NR-SAR-DL) filter [15]. All these methods were implemented by the source codes provided by the authors of the respective articles. The proposed method was implemented in the Pytorch package [45] and run on a workstation with an INTEL 10920x CPU, two NVIDIA 2080Ti GPUs, and 128G RAM. The experiment employed the Adam method [46] to optimize the network in the training process, with the parameter settings as $\beta_1 = 0.9$ and $\beta_2 = 0.999$. The learning rate was set to 0.003 without decay during the training. The block size was set to 13×13 pixels, and η was set to 2 for similarity calculation in the second pass for all the dataset except Sentinel-1, where η was set to 1. The loss weight λ was set to 0 for the first pass and 0.5 for the second pass. Due to the calculation capacity limitation, each test used 10,000 index blocks, and the similarity blocks were searched within 90×90 pixels from the center of the index blocks. For each index block, 32 of the most similar blocks were searched to construct training pairs. We chose the similarity threshold T to eliminate the 10% outliers in the training pairs produced in the similar search process. However, fine tuning T based on the denoising result was still required. The source codes of the proposed method and the experiment dataset can be retrieved from the link <https://github.com/githubeastmad/bm-nr-dl> (accessed on 19 December 2021).

Figure 3 demonstrates the training curve for a TerraSAR image. During the training, the loss converged very quickly but would not decrease close to zero, since the training output was a clean image, and the reference was its noisy counterpart. The training on the TerraSAR image cost around 9 min for each pass with 1500 iterations and a batch size of 300 to achieve a decent denoising result. For comparison, the PPB filter cost around 4 min running on the same machine.

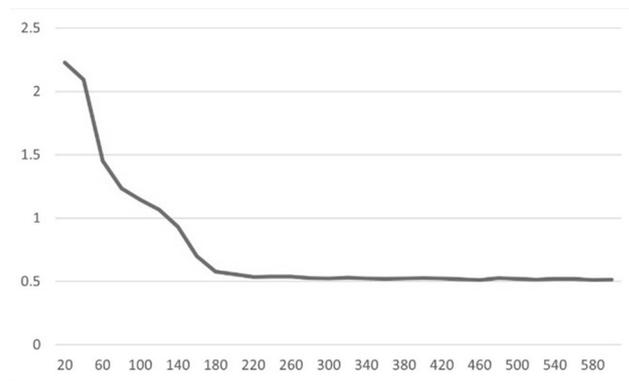


Figure 3. The training curve of the TerraSAR training pass one. The x axis is the iteration with a batch size of 300. The y axis is the average loss of every consecutive 20 iterations with log10 transformation.

4.1. Experiments with Simulated Images

Using the same optical image, we synthesized single-look speckled SAR images for experiments. Two tests were carried out: the first test trained the network on a single-look speckled image and despeckled the same image. The second test trained the network on bi-temporal images and despeckled the trained image. The despeckled results of the proposed model were quantitatively compared with other models by the peak signal-to-noise ratio (PSNR) [47], the structural similarity (SSIM) index [48], and the equivalent number of looks (ENL) [49]. Table 2 lists the assessment results, and Figure 4 shows the filtering results of the different methods. Table 2 also includes the rank-sum of each filter, which is the sum of the filter performance ranking in the three metrics. In the experiments, the PPB filter presented an over-smoothed result with considerable fine linear structure loss. The SAR-DRN filtered image still had notable speckles. It also presented a lower range of gray-level compared with the clean image. The MSAR-BM3D filtered image had a block effect in the homogeneous region. The linear structure in the urban region was also blurred to some degree. NR-SAR-DL introduced some undesirable point-like artifacts. It also had a subtle block effect in the homogeneous area. The proposed BM-NR-DL filter, either trained using a single noisy image or bi-temporal images, yielded good despeckling results with visually pleasing output. The despeckled image presented no significant artifacts or blurs. The quantitative evaluation listed in Table 2 supports the previous observations.

Table 2. Quantitative assessment results for the simulated dataset.

Method	Filter	PSNR	SSIM	ENL	Rank-Sum
Single noisy image	PPB	19.30	0.64	2202.53	7
	SAR-DRN	19.04	0.63	53.93	3
	BM-NR-DL	21.43	0.69	860.80	8
Two noisy images	MSAR-BM3D	23.80	0.71	805.76	7
	NR-SAR-DL	23.66	0.63	257.17	3
	BM-NR-DL	24.64	0.71	726.66	8

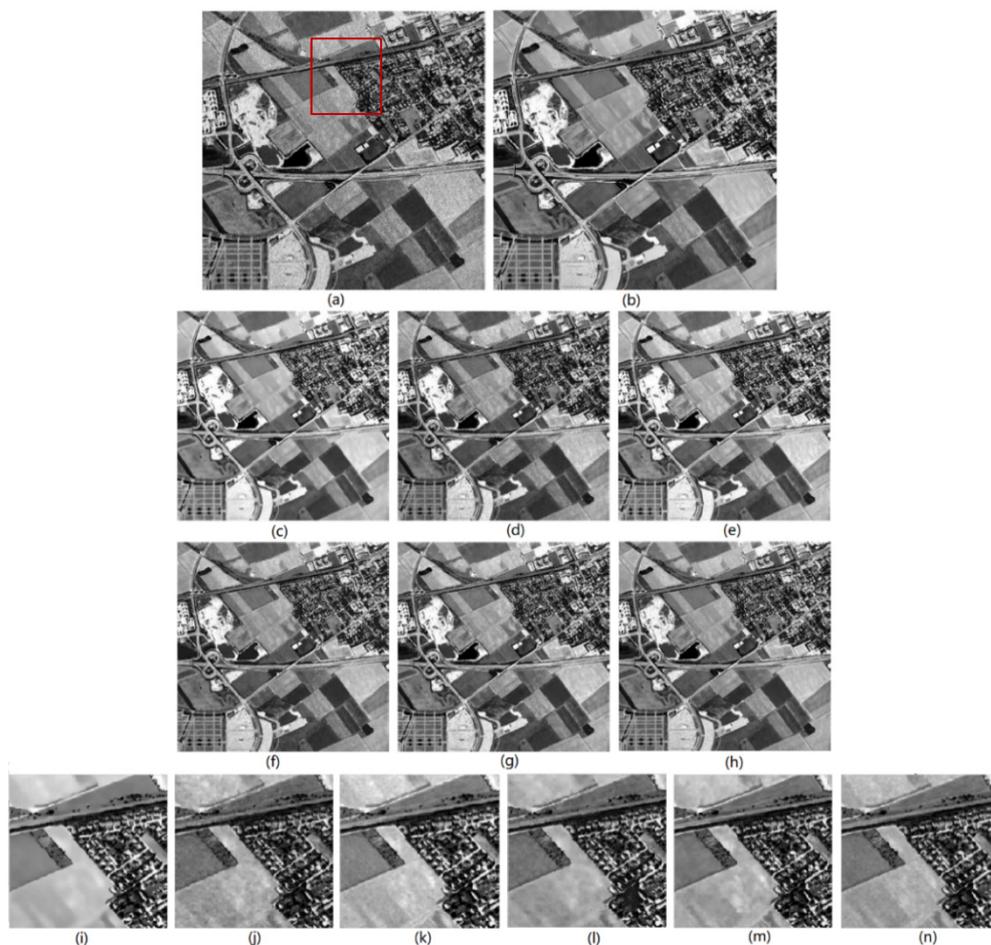


Figure 4. The despeckling result of the simulated SAR dataset. (a) Speckled image. (b) Clean image. (c–h) The PPB filter, the SAR-DRN filter, the BM-NR-DL filter trained with one image, the NR-SAR-DL filter, the MSAR-BM3D filter, and the BM-NR-DL filter trained with bi-temporal images, respectively. (i–n) Sub images of the highlight part in (a) and cropped from (c–h), respectively.

4.2. Experiments with Real SAR Images

We then inspected the despeckling performance of the proposed model using real single-look SAR datasets from three different sensors: TerraSAR, Sentinel-1, and ALOS. The TerraSAR images were in Ruhr, Germany on 20 February and 4 March, 2008. The Sentinel-1 images were in Wuhan, China on 26 April and 8 May, 2019. The ALOS images were acquired in 2007 and 2010, which contained major changes. The multi-temporal data had not been fine-registered. Two tests were carried out for each dataset: the first test trained the despeckling deep model with only one speckled image and then despeckled the same image. The second test trained the network with two speckled images and despeckled the two images. Since no ground-truth was available, we compared the performance of the proposed method with others by the four following quantitative metrics: the ENL, the edge-preservation degree based on the ratio of average (EPD-ROA) [50], the mean of ratio (MOR), and the target-to-clutter ratio (TCR) [51]. The EPD-ROA metric is calculated as follows:

$$EPD - ROA = \frac{\sum_{i=1}^M |x'_A(i)/x'_B(i)|}{\sum_{i=1}^M |x_A(i)/x_B(i)|} \quad (7)$$

where $x'_A(i)$ and $x'_B(i)$ are the adjacent pixel pair of a certain direction in the despeckled image. The EPD-ROA metric describes how the denoising filter retains edges along certain directions. Its value should be close to 1 if the edge preservation is good. The MOR metric is the ratio between the despeckled image and the original one. It describes how the denoising

filter preserves the radiometric information and, if the preservation is good, should be close to 1. The TCR is calculated with image patches as follows:

$$TCR = |20 \log_{10} \max(\mathbf{X}')/E(\mathbf{X}') - 20 \log_{10} \max(\hat{\mathbf{X}})/E(\hat{\mathbf{X}})| \quad (8)$$

where \mathbf{X}' and $\hat{\mathbf{X}}$ denote the despeckled and noisy image patches with a high return point, respectively. The TCR measures the radiometric information preservation of the region around the high return point and should be close to 0.

Table 3 lists the assessment results, and Figures 5–7 show the despeckling results of the experiments with ALOS, Sentinel-1, and the TerraSAR dataset, respectively. Table 3 also includes the rank-sum of each filter, which is the sum of the filter performance ranking in the four metrics. For the TerraSAR dataset, the PPB, BM-NR-DL, and NR-SAR-DL presented superior speckling surpassing capacity, as indicated by ENL. However, the PPB over-smoothed the image and yielded unnatural despeckling results. It also introduced undesirable artifacts in the despeckled image. The SAR-DRN retained well the fine structures but was poor in speckle surpassing. NR-SAR-DL achieved very good speckling surpassing capacity but failed to retain some strong point targets, as demonstrated in the TCR value for the selected highlight. MSAR-BM3D yielded natural despeckling output with good highlight retaining capacity. However, its result presented some artifacts and had a significant block effect. The BM-NR-DL filter showed the best speckle suppression, good fine structure maintenance, and good strong point target retainment. The despeckling results of the model trained with single image presented higher ENL and TCR compared with those of the model trained with two images. However, the later model preserved better the fine structures and radiometric information.

Table 3. Quantitative assessment results for the TerraSAR, Sentinel-1, and ALOS datasets.

	Method	Filter	ENL	EPD	MOR	TCR	Rank-Sum
TerraSAR	Single noisy images	PPB	111.68	0.66	0.85	1.86	7
		SAR-DRN	6.89	0.74	0.93	2.17	8
		BM-NR-DL	171.71	0.66	0.96	6.05	9
	Two noisy images	NR-SAR-DL	130.17	0.66	0.99	0.58	9
		MSAR-BM3D	57.98	0.67	0.96	3.35	6
		BM-NR-DL	59.22	0.66	0.99	2.57	9
Sentinel-1	Single noisy images	PPB	102.85	0.70	0.86	3.10	11
		SAR-DRN	9.30	0.69	0.38	8.51	5
		BM-NR-DL	76.06	0.68	0.87	8.03	8
	Two noisy images	NR-SAR-DL	67.68	0.71	0.79	7.95	7
		MSAR-BM3D	61.74	0.70	0.88	1.73	8
		BM-NR-DL	79.62	0.69	0.88	5.02	9
ALOS	Single noisy images	PPB	683.94	0.96	0.98	6.46	9
		SAR-DRN	9.17	1.07	0.61	1.73	6
		BM-NR-DL	276.29	0.96	0.93	4.03	9
	Two noisy images	NR-SAR-DL	29.68	0.95	0.92	2.21	7
		MSAR-BM3D	205.32	0.96	0.91	2.67	7
		BM-NR-DL	303.99	0.96	0.99	2.81	10

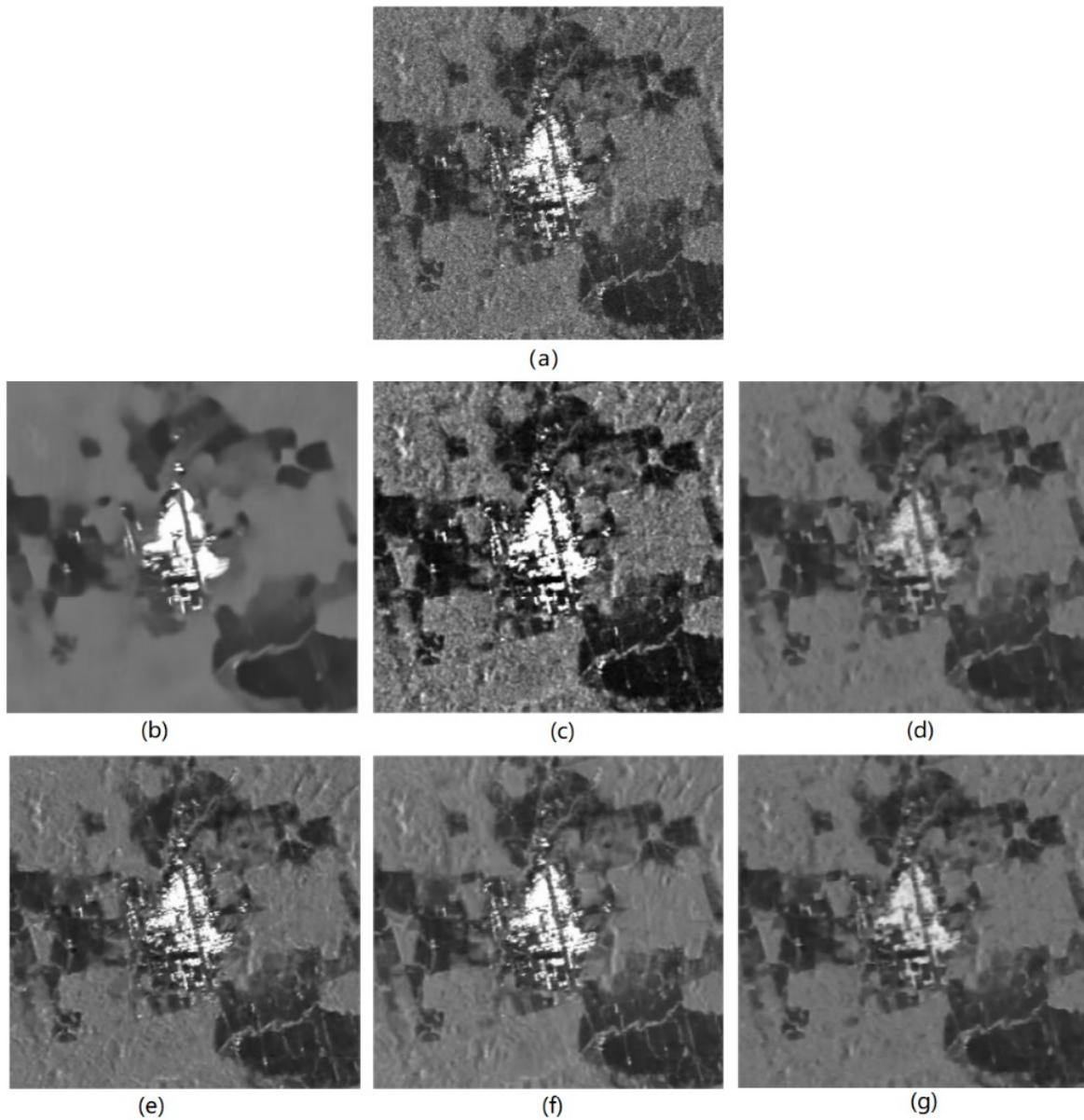


Figure 5. The despeckling result of the ALOS dataset. (a) Speckled image. (b–g) The images filtered by the PPB filter, the SAR-DRN filter, the BM-NR-DL filter trained with one image, the NR-SAR-DL filter, the MSAR-BM3D filter, and the BM-NR-DL filter trained with bi-temporal images, respectively.

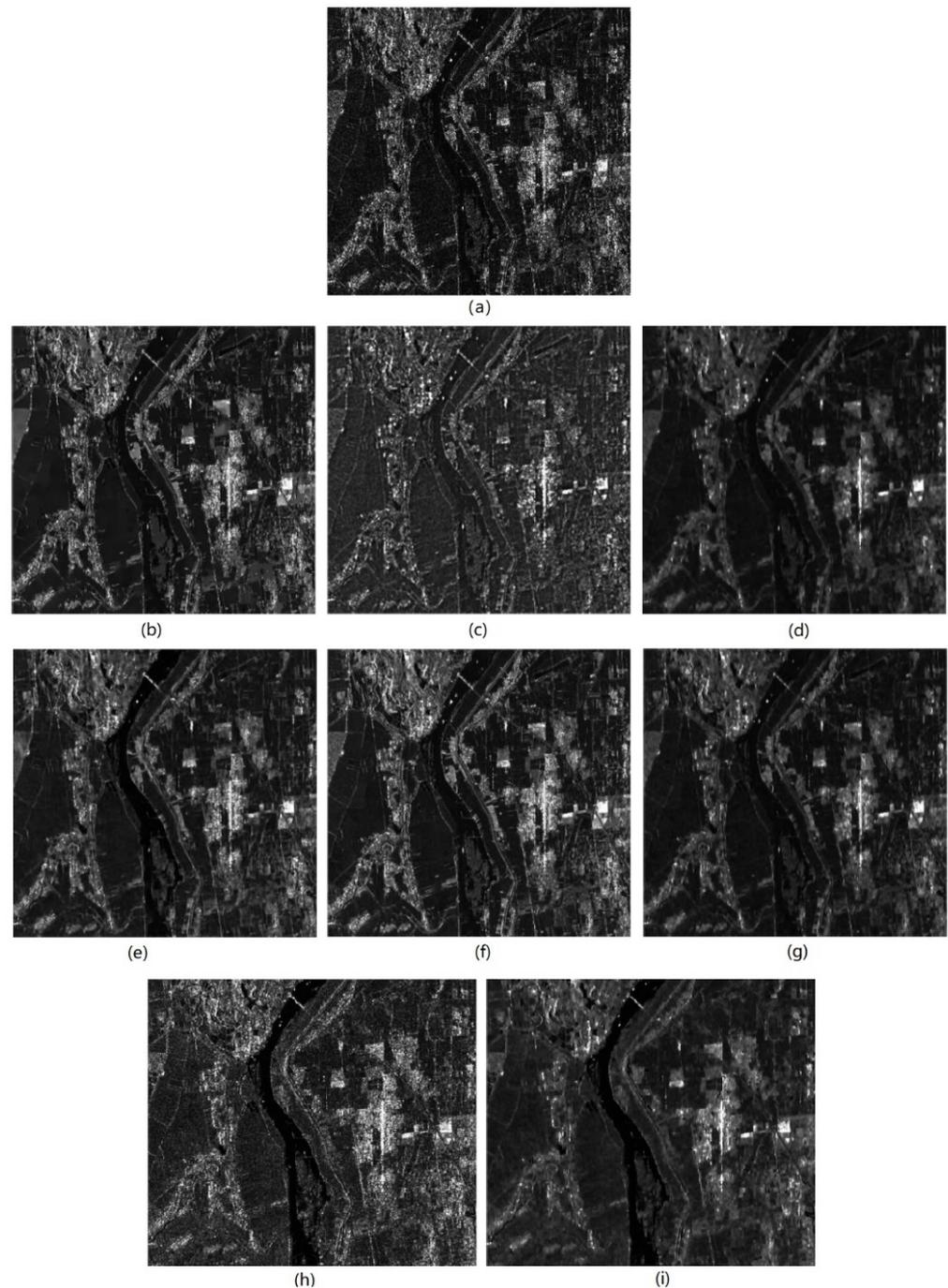


Figure 6. The despeckling result of the Sentinel-1 dataset. (a) Speckled image. (b–g) The images filtered by the PPB filter, the SAR-DRN filter, the BM-NR-DL filter trained with one image, the NR-SAR-DL filter, the MSAR-BM3D filter, and the BM-NR-DL filter trained with bi-temporal images, respectively. (h) Speckled image of a same region but another time phase. (i) The image filtered by the BM-NR-DL filter pre-trained with bi-temporal images.

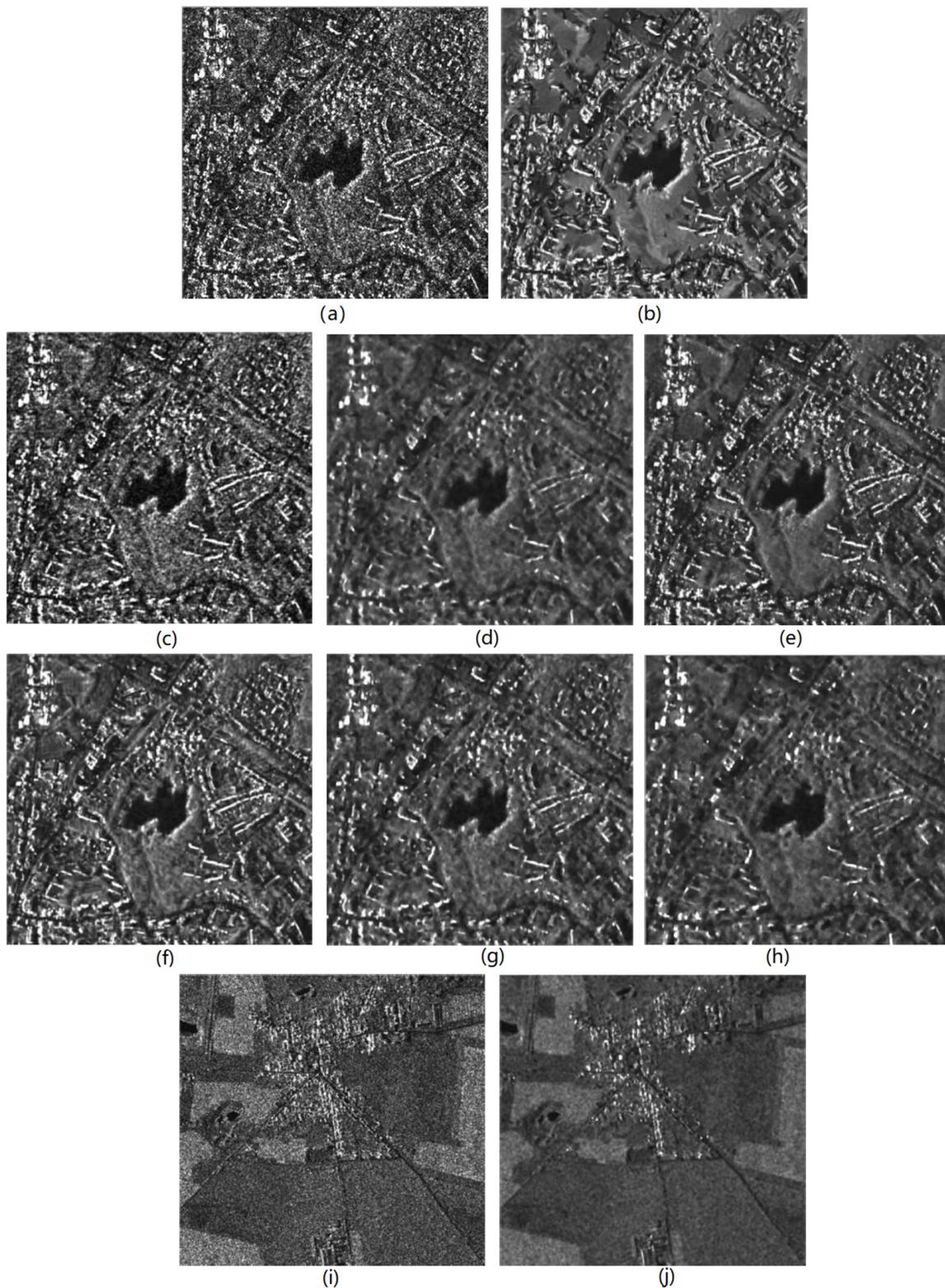


Figure 7. The despeckling result of the TerraSAR dataset. (a) Speckled image. (b–g) The images filtered by the PPB filter, the SAR-DRN filter, the BM-NR-DL filter trained with one image, the NR-SAR-DL filter, the MSAR-BM3D filter, and the BM-NR-DL filter trained with bi-temporal images, respectively. (h) Another time phase image filtered by the BM-NR-DL filter pre-trained with bi-temporal images. (i) Speckled image of a new region. (j) The image filtered by the BM-NR-DL filter pre-trained with bi-temporal images.

For the Sentinel-1 and ALOS datasets, though PPB presented the highest ENL, visual inspection showed that its results missed fine structures and were over-smooth. On the other hand, the SAR-DRN filter failed to eliminate the speckles to a large degree. It also greatly distorted the radiometric information of the speckled image. The despeckling results of the MSAR-BM3D method had artifacts and presented block-effects, similar to the results with simulated and TerraSAR images. NR-SAR-DL showed good performance in all the four metrics. However, BM-NR-DL performances were even better, except for EPD and TCR in the Sentinel-1 dataset. Visual inspection also discovered that the BM-NR-DL method preserved more fine structures and had much less radiometric distortion compared with the NR-SAR-DL method. The only exception was that the BM-NR-DL despeckled ALOS image had a lower contrast compared with that despeckled by the NR-SAR-DL method. However, by post-processing, the contrast could be easily maintained, and the fine structures were still well preserved. The high return preserving capability of BM-NR-DL was acceptable. In sum, BM-NR-DL achieved a good balance between speckle surpassing and structure preserving. Unlike other methods, which present an over sharpening feeling, the despeckled image of the BM-NR-DL was more natural.

4.3. Generalization Experiments

Generalization capacity, which refers to the performance of a denoising method on a non-training dataset, is not a necessity for BM-NR-DL, as this method can be self-supervised. However, exploiting the generalization capacity of the despeckling network is valuable, since using a pre-trained network has a great performance advantage compared with training an ad hoc model for each new noisy image.

We used the model trained in the previous experiments with real SAR data to despeckle the image that was not used in the training process to understand the generalization capability of the denoising network. The preliminary evaluation with images acquired by a different sensor of the training image was unsatisfactory. As a result, generalization experiments were carried out only with the images of the same sensor. For TerraSAR, the single-image-trained network was used to despeckle another TerraSAR image of the same region but with significant temporal alternation. In addition, the two-image-trained network was used to despeckle another TerraSAR image of a new region. For Sentinel-1, the two-image-trained network was used to despeckle another Sentinel-1 image of the same region but with significant temporal alternation. In the test of both datasets, we used the PPB filter as the reference. The images are shown in Figure 6h,i and Figure 7h,j, and the quantitative analysis is shown in Table 4.

Table 4. Quantitative assessment results of the generalization experiments.

Dataset	Experiment	ENL	EPD	MOR	TCR
TerraSAR	T1 trained, T3 despeckling	97.03	0.60	0.90	4.15
	PPB T3	108.88	0.66	0.86	1.95
	T1, T2 trained, new region despeckling	95.5581	0.7178	0.93	6.05
	PPB new region	262.28	0.7206	0.88	3.56
Sentinel-1	T1, T2 trained, T3 despeckling	110.91	0.64	0.99	10.33
	PPB T3	108.94	0.69	0.85	3.64

The experiment showed positive results. The visual inspection of the despeckling result showed that the speckle was effectively eliminated without significant loss in the fine structure in accordance with the quantitative evaluation. We spotted no obvious block effects and very few artifacts. In contrast, the PPB filter produced an over-smooth result with noticeable artifacts. The quantitative evaluation indicated a slight decrease in the EPD value for the TerraSAR dataset when compared with the despeckling result of the training

images. The previous ad hoc model despeckling result showed a higher EPD value than PPB in contrast to a slightly lower value than PPB in the current experiment. However, there was no perceivable deterioration in the edge perception.

The observation of the previously mentioned experiment demonstrated that the generalization capability of the denoising network is very promising when used for a given sensor. The great generalization capacity of the proposed model brings application flexibility. The noisy image can now be despeckled either by an ad hoc trained network or by the network that is previously well-trained with the images of the same sensor. The ad-hoc trained network can yield promising despeckling results. On the other hand, using the pre-trained network can process the noisy image much quicker with very limited performance reduction. In our experiment, the despeckling of a $2400 * 2400$ Sentinel-1 image only costs around 10 s, while PPB requires over 400 s.

5. Discussion

The proposed BM-NR-DL despeckling method combines block-matching and noise-referenced SAR deep learning despeckling. To our knowledge, it is the first such attempt in this domain. The proposed model shows good speckling results in the previous experiments with both synthesized and real SAR data. Its speckling result is visually pleasing with a balance between speckle suppression and detail preservation, and its quantitative evaluation metrics are generally superior to other reference methods. These experiments also demonstrate the applicability of the BM-NR-DL in various training and despeckling settings. The noisy image patches for network training could be acquired from either a single SAR image or multiple images of the same sensor. Synthesized noisy data or a “clean” image are no longer necessary, thus avoiding the domain gap introduced by synthesized data. The despeckling network also has good generalization capability. The despeckling network can now be specifically trained for the targeted noisy images, or it can be pre-trained with other images of the same sensor, thus saving calculation time for a specific mission.

Despite the overall promising results, the high return preservation of the proposed method can only be regarded as fair. By carefully inspecting the highlights in the tested images, we discover that the structure of the highlighting is largely preserved in the despeckled result; yet the radiometric information has been distorted to some degree. In the experiment, the highlighting despeckling performance is very sensitive to the training settings, such as the block-matching searching radius and the nature of the training noisy image. Extending the search radius has a positive impact on the performance of high return preservation. This observation may indicate that the current block-matching process has not searched and constructed enough similar noisy image pairs that contains high returns. In addition, some of the training settings, such as λ in the training loss calculation and the similarity threshold T , are set according to the empirical test result. It may require multiple fine-tuning manually, which could be time-consuming. The training of the proposed deep network also requires more time compared with classical denoising approaches. Thus, practitioners may expect a longer denoising time if training an ad hoc network for the best denoising result. On the other hand, denoising with a pre-trained network is significantly quicker than classical denoising approaches.

Potential further investigations include using the lookup tables for a wider range of block searching, such as that of [36], and introducing auto-adaptive similarity estimation according to different image blocks, such as that of [25]. Moreover, using the lookup tables and variational methods may raise an interesting research question for future work: instead of randomized index block selection in the current BM-NR-DL method, how can training image pairs be constructed that lead to better performance and generalization capability? Discovering the rule of parameter setting could also facilitate this process and may lead to an automatic solution in the future.

6. Conclusions

The noise-referenced speckling deep learning network has shown promising results in recent research. It can train with only real SAR data, thus avoiding domain gaps. However, its training process requires a fine registered multitemporal dataset for image pair extraction. The training cannot deal well with images that have major changes and may still require temporal variation compensation based on pixel-level similarity evaluation. Inspired by the fact that the similarity index could be irrelevant to the block's location and acquiring time, in this article, we propose a block-matching noise-referenced deep learning method for self-supervised network training and SAR despeckling. The main idea of the proposed method is to construct a series of noisy block pairs by similarity-based block matching in one or several SAR images as the network training input. Then the method uses these noisy block pairs to train a Siamese network with two parameter-sharing encoder–decoder convolutional neural network branches through two passes: first for the preliminary despeckling, and second for the final training with a refined similarity estimation. The despeckling inference requires only one branch of the Siamese network.

The experiments with both the simulated and real SAR dataset show promising despeckling results. The despeckled image is visually pleasing without the feeling of over sharpening. The fine structure is well preserved without significant artifacts. The proposed method also demonstrates a good despeckling capability for non-training images from the same sensor. Despite the good despeckling performance, another main advantage of the proposed method is its training flexibility. The network can be trained with either one SAR image or multiple images of a certain sensor. Moreover, the network can be trained ad hoc or pre-trained with images other than the targeted one.

Author Contributions: Conceptualization, C.W.; methodology, C.W. and X.M.; software, C.W. and Z.Y. (Zhixiang Yin); validation and analysis, C.W., Z.Y. (Zhixiang Yin), X.M. and Z.Y. (Zhutao Yang); writing—original draft preparation, C.W.; writing—review and editing, Z.Y. (Zhixiang Yin), X.M. and Z.Y. (Zhutao Yang). All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the National Natural Science Foundation of China under Grants 41901410, 42104036, 41701390, and 41906168, by the Open Research Fund Program of the Key Laboratory of Digital Mapping and Land Information Application Engineering, NASG, China, under Grant ZRZYBWD201905, by the Natural Science Foundation of Anhui Province under Grant 1908085QD161, and by the Hefei Municipal Natural Science Foundation No. 2021041.

Data Availability Statement: The original code and the test dataset are available online at <https://github.com/githubbeastmad/bm-nr-dl> (accessed on 20 December 2021).

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

List of Some Symbols

X	Clean image
\tilde{X}	Noisy/speckled image
X'	Denoised image
$f_{\theta}(\cdot)$	Network with the parameter set θ
S	Similarity index
$L(\cdot)$	Loss function
T	Similarity threshold

References

1. Zhang, K.; Zuo, Y.; 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](#)] [[PubMed](#)]
2. Zhang, K.; Zuo, W.; Zhang, L. FFDNet: Toward a fast and flexible solution for CNN-based image denoising. *IEEE Trans. Image Process.* **2017**, *27*, 4608–4622. [[CrossRef](#)] [[PubMed](#)]
3. Cozzolino, D.; Verdoliva, L.; Scarpa, G.; Poggi, G. Nonlocal CNN SAR image despeckling. *Remote Sens.* **2020**, *12*, 1006. [[CrossRef](#)]
4. Zhou, Y.; Shi, J.; Yang, X.; Wang, C.; Kumar, D.; Wei, S.; Zhang, X. Deep multi-scale recurrent network for synthetic aperture radar images despeckling. *Remote Sens.* **2019**, *11*, 2462. [[CrossRef](#)]

5. Gretton, A.; Smola, A.; Huang, J.; Schmittfull, M.; Borgwardt, K.; Schölkopf, B. Covariate shift and local learning by distribution matching. In *Dataset Shift in Machine Learning*; Quiñero-Candela, J., Sugiyama, M., Schwaighofer, A., Lawrence, N.D., Eds.; MIT Press: Cambridge, MA, USA, 2009; pp. 131–160.
6. Ulyanov, D.; Vedaldi, A.; Lempitsky, V. Deep Image Prior. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 9446–9454.
7. Krull, A.; Buchholz, T.O.; Jug, F. Noise2void-learning Denoising from Single Noisy Images. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 16–20 June 2019; pp. 2129–2137.
8. Batson, J.; Royer, L. Noise2self: Blind Denoising by Self-supervision. In Proceedings of the International Conference on Machine Learning, Long Beach, CA, USA, 10–15 June 2019; pp. 524–533.
9. Molini, A.B.; Valsesia, D.; Fracastoro, G.; Magli, E. Towards Deep Unsupervised SAR Despeckling with Blind-Spot Convolutional Neural Networks. In Proceedings of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Waikoloa, HI, USA, 26 September–2 October 2020.
10. Laine, S.; Karras, T.; Lehtinen, J.; Aila, T. High-quality self-supervised deep image denoising. In Proceedings of the Advances in Neural Information Processing Systems 32, Vancouver, BC, Canada, 8–14 December 2019; pp. 6970–6980.
11. Joo, S.; Cha, S.; Moon, T. DoPAMINE: Double-sided Masked CNN for Pixel Adaptive Multiplicative Noise Despeckling. In Proceedings of the AAAI Conference on Artificial Intelligence, Honolulu, HI, USA, 27 January–1 February 2019; Volume 33, pp. 4031–4038.
12. Zhang, Q.; Yuan, Q.; Li, J.; Yang, Z.; Ma, X. Learning a dilated residual network for SAR image despeckling. *Remote Sens.* **2018**, *10*, 1–18. [[CrossRef](#)]
13. Lehtinen, J.; Munkberg, J.; Hasselgren, J.; Laine, S.; Karras, T.; Aittala, M.; Aila, T. Noise2noise: Learning Image Restoration without Clean Data. In Proceedings of the International Conference on Machine Learning, Stockholm, Sweden, 10–15 July 2018; pp. 2965–2974.
14. Dalsasso, E.; Denis, L.; Tupin, F. SAR2SAR: A semi-supervised despeckling algorithm for SAR images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2020**, *14*, 4321–4329. [[CrossRef](#)]
15. Ma, X.; Wang, C.; Yin, Z.; Wu, P. SAR image despeckling by noisy reference-based deep learning method. *IEEE Trans. Geosci. Remote Sens.* **2020**, *12*, 8807–8818. [[CrossRef](#)]
16. 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.
17. Rudin, L.I.; Osher, S.; Fatemi, E. Nonlinear total variation based noise removal algorithms. *Phys. D Nonlinear Phenom.* **1992**, *60*, 259–268. [[CrossRef](#)]
18. Osher, S.; Burger, M.; Goldfarb, D.; Xu, J.; Yin, W. An iterative regularization method for total variation-based image restoration. *Multiscale Model Simul.* **2005**, *4*, 460–489. [[CrossRef](#)]
19. Buades, A.; Coll, B.; Morel, J.M. A Non-local Algorithm for Image Denoising. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Diego, CA, USA, 20–26 June 2005; Volume 2, pp. 60–65.
20. Lee, J.S. A simple speckle smoothing algorithm for synthetic aperture radar images. *IEEE Trans. Syst. Man. Cybern.* **1983**, *13*, 85–89. [[CrossRef](#)]
21. Kuan, D.; Sawchuk, A.; Strand, T.; Chavel, P. Adaptive restoration of images with speckle. *IEEE Trans. Acoust. Speech Signal Process.* **1987**, *35*, 373–383. [[CrossRef](#)]
22. Frost, V.S.; Stiles, J.A.; Shanmugan, K.S.; Holtzman, J.C. A model for radar images and its application to adaptive digital filtering of multiplicative noise. *IEEE Trans. Pattern Anal. Mach. Intell.* **1982**, *4*, 157–166. [[CrossRef](#)] [[PubMed](#)]
23. Deledalle, C.A.; Denis, L.; Tupin, F. Iterative weighted maximum likelihood denoising with probabilistic patch-based weights. *IEEE Trans. Image Proc.* **2009**, *18*, 2661–2672. [[CrossRef](#)] [[PubMed](#)]
24. 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](#)]
25. Ma, X.; Shen, H.; Zhao, X.; Zhang, L. SAR image despeckling by the use of variational methods with adaptive nonlocal functionals. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 3421–3435. [[CrossRef](#)]
26. Shen, H.; Zhou, C.; Li, J.; Yuan, Q. SAR image despeckling employing a recursive deep CNN prior. *IEEE Trans. Geosci. Remote Sens.* **2020**, *59*, 273–286. [[CrossRef](#)]
27. Chierchia, G.; Cozzolino, D.; Poggi, G.; Verdoliva, L. SAR Image Despeckling through Convolutional Neural Networks. In Proceedings of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Fort Worth, TX, USA, 23–28 July 2017; pp. 5438–5441.
28. Gu, F.; Zhang, H.; Wang, C.; Zhang, B. Residual Encoder-decoder Network Introduced for Multisource SAR Image Despeckling. In Proceedings of the SAR in Big Data Era: Models, Methods and Applications (BIGSAR DATA), Beijing, China, 13–14 November 2017; pp. 1–5.
29. Liu, R.; Li, Y.; Jiao, L. SAR Image Speckle Reduction Based on a Generative Adversarial Network. In Proceedings of the International Joint Conference on Neural Networks (IJCNN), Glasgow, UK, 19–24 July 2020; pp. 1–6.
30. Li, J.; Li, Y.; Xiao, Y.; Bai, Y. HDRANet: Hybrid dilated residual attention network for SAR image despeckling. *Remote Sens.* **2019**, *11*, 2921. [[CrossRef](#)]

31. Yue, D.X.; Xu, F.; Jin, Y.Q. SAR despeckling neural network with logarithmic convolutional product model. *Int. J. Remote Sens.* **2018**, *39*, 7483–7505. [CrossRef]
32. Yuan, Y.; Sun, J.; Guan, J.; Feng, P.; Wu, Y. A Practical Solution for SAR Despeckling with Only Single Speckled Images. *arXiv* **2019**, arXiv:1912.06295. Available online: <https://arxiv.org/abs/1912.06295> (accessed on 19 December 2021).
33. Zhang, G.; Li, Z.; Li, X.; Xu, Y. Learning synthetic aperture radar image despeckling without clean data. *J. Appl. Remote Sens.* **2020**, *14*, 1–20. [CrossRef]
34. Dabov, K.; Foi, A.; Katkovnik, V.; Egiazarian, K. Image denoising by sparse 3D transform-domain collaborative filtering. *IEEE Trans. Image Process.* **2007**, *16*, 2080–2095. [CrossRef]
35. Cozzolino, D.; Parrilli, S.; Scarpa, G.; Poggi, G.; Verdoliva, L. Fast adaptive nonlocal SAR despeckling. *IEEE Geosci. Remote Sens. Lett.* **2014**, *11*, 524–528. [CrossRef]
36. Chierchia, G.; El Gheche, M.; Scarpa, G.; Verdoliva, L. Multitemporal SAR image despeckling based on block-matching and collaborative filtering. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 5467–5480. [CrossRef]
37. Yang, D.; Jian, S. Bm3d-net: A convolutional neural network for transform-domain collaborative filtering. *IEEE Signal Process. Lett.* **2017**, *25*, 55–59. [CrossRef]
38. Cruz, C.; Foi, A.; Katkovnik, V.; Egiazarian, K. Nonlocality-reinforced convolutional neural networks for image denoising. *IEEE Signal Process. Lett.* **2018**, *25*, 1216–1220. [CrossRef]
39. Ahn, B.; Yoonsik, K.; Guyong, P.; Cho, N.I. Block-matching Convolutional Neural Network (bmcnn): Improving Cnn-based Denoising by Block-matched Inputs. In Proceedings of the 2018 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), Honolulu, HI, USA, 12–15 November 2018; pp. 516–525.
40. Zeng, T.; So, H.K.-H.; Lam, E.Y. Computational image speckle suppression using block matching and machine learning. *Appl. Opt.* **2019**, *58*, B39–B45. [CrossRef]
41. Denis, L.; Deledalle, C.A.; Tupin, F. From Patches to Deep Learning: Combining Self-similarity and Neural Networks for SAR Image Despeckling. In Proceedings of the IEEE International Geoscience and Remote Sensing Symposium, Yokohama, Japan, 28 July–2 August 2019; pp. 5113–5116.
42. Niu, C.; Ge, W. Noise2Sim—Similarity-based Self-Learning for Image Denoising. *arXiv* **2020**, arXiv:2011.03384. Available online: <https://arxiv.org/abs/2011.03384> (accessed on 19 December 2021).
43. Lebrun, M. An analysis and implementation of the BM3D image denoising method. *Image Process. OnLine* **2012**, *2*, 175–213. [CrossRef]
44. Maas, A.L.; Hannun, A.Y.; Ng, A.Y. Rectifier Nonlinearities Improve Neural Network Acoustic Models. In Proceedings of the International Conference on Machine Learning, Atlanta, GA, USA, 16–21 June 2013; Volume 30.
45. Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; Killeen, T.; Lin, Z.; Gimelshein, N.; Antiga, L.; et al. Pytorch: An imperative style, high-performance deep learning library. *arXiv* **2019**, arXiv:1912.01703. Available online: <https://arxiv.org/abs/1912.01703> (accessed on 19 December 2021).
46. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980. Available online: <https://arxiv.org/abs/1412.6980> (accessed on 19 December 2021).
47. Zhao, W.; Deledalle, C.-A.; Denis, L.; Maitre, H.; Nicolas, J.-M.; Tupin, F. Ratio-based multitemporal SAR images denoising: RABASAR. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 3552–3565. [CrossRef]
48. Wang, Z.; Bovik, A.C.; Sheikh, H.R.; Simoncelli, E.P. Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Process.* **2004**, *13*, 600–612. [CrossRef] [PubMed]
49. Lee, J.S.; Pottier, E. *Polarimetric Radar Imaging: From Basics to Applications*; CRC Press: Boca Raton, FL, USA, 2009.
50. Feng, H.; Hou, B.; Gong, M. SAR image despeckling based on local homogeneous-region segmentation by using pixel-relativity measurement. *IEEE Trans. Geosci. Remote Sens.* **2011**, *49*, 2724–2737. [CrossRef]
51. Argenti, F.; Lapini, A.; Bianchi, T.; Alparone, L. A tutorial on speckle reduction in synthetic aperture radar images. *IEEE Geosci. Remote Sens. Manag.* **2013**, *1*, 6–35. [CrossRef]