ORIGINAL RESEARCH PAPER

SAR image despeckling using deep CNN

Alicia Passah  |  Khwairakpam Amitab  |  Debdatta Kandar

Department of Information Technology,
North-Eastern Hill University, Shillong, Meghalaya

Correspondence
Debdatta Kandar, Department of Information Technology, North-Eastern Hill University, Shillong, Meghalaya.
Email: dkandar@nehu.ac.in

Abstract
Synthetic aperture radar (SAR) images are contaminated with noise called speckle that is multiplicative in nature. The presence of speckles in SAR images makes it impossible to understand and interpret for extensive range of applications. However, certain characteristics of SAR and the ability to function irrespective of weather conditions, are making such images worth processing in order to be able to extract relevant information. A despeckling model is proposed that uses deeper convolutional neural networks, which was never used before, as far as authors are concerned, for diminishing speckle in noisy SAR images. Multiple skip connections from the ResNet model are also employed in authors’ proposed architecture. In order to maintain uniformity, a formula to be followed while applying skip connections is also derived. A hybrid loss function is developed to train the network more consistently to achieve the desired output. Experiments on simulated SAR images using the NWPU-RESISC benchmark are conducted and tested on real TerraSAR-X images and compared with the state-of-the-art techniques. Results show that the proposed method achieved considerable improvements compared to state-of-the-art methods with PSNRs 27.02, 24.60, and 22.01 for looks 10, 4, and 1, respectively.

1 | INTRODUCTION

Synthetic aperture radar is an all-weather remote sensing imaging radar used to construct two-dimensional images or three-dimensional recreation of objects present on the ground [1]. The ability of SAR to function in all types of weather circumstances has made it popular in diverse applications like military defense, disaster management, pollution detection, illegal mining detection, battlefield surveillance, exploration of resources, traffic monitoring, urban planning etc. [2]. The SAR modality has a number of characteristics that distinguish it from optical imagery. It is an active imaging system, in the sense that it has its own illumination, and is generally mounted on a spacecraft or aircraft and even in guided missiles these days. The SAR system captures the image by transmitting electromagnetic waves and processing the received echo to reconstruct the image of the target. These reflected echoes are mostly interfered by signals resulting from uneven surfaces and mutual coupling arrays causing the system to add vague and irrelevant information termed as speckle noise on the reconstructed image.

It may also be mentioned that the performance of a SAR radar can be characterised by its target detectability and imaging quality. But in the real environment, SAR’s performance is greatly affected by the multi-path effect and clutter. To encounter such issues, researchers are looking for the development of low probability of intercept (LPI) signals [3], and such signals are characterised by their auto-correlation function or ambiguity function. As per studies, the desired signal should have a high peak side-lobe level (PSL) ratio and gradually decreasing side lobe level [4]. At the same time, one of the promising solutions is to have multiple antenna technologies like multiple input multiple output (MIMO) or phased array to encounter such environmental effects [5, 6]. However, collocated antenna systems give rise to mutual coupling issue between array elements, which ultimately degrades the performance of a SAR radar. Due to the coupling effect, the side lobes in the auto-correlation function gets perturbed leading to the reduction in PSL values, which affects the target detectability of SAR systems. The undesired signals resulting from array element mutual coupling degrades the measured data acquired by SAR radar. Hence the images produced from such corrupted signals are contaminated with unwanted effects such as artifacts and blurring [7, 8]. One possible solution to reduce this problem is by employing a distributed antenna arrangement, which may
be at the cost of coherent signal processing gain. The collocated antenna can be arranged by taking advantage of the benefits of coherent processing.

However, in this work, our primary focus is on the SAR image itself rather than the process of capturing it. Therefore, it may be worth mentioning that the images obtained by SAR inherently contain speckle noise, which ultimately degrades the quality and features of the images, thereby making the interpretation and understandings of SAR images difficult and therefore, a challenging task. The importance of SAR have led researchers to incorporate modern techniques like deep learning to improve the quality of images acquired by SAR. Hence, image denoising is still a dynamic subject since it is an essential step in a wide range of viable applications.

Considering the presence of noise in SAR images, many researchers have proposed several filtering techniques, of which most of them are based on statistical methods such as Lee filter [9], Kuan filter [10], Frost filter [11], BM3D [12], SAR-BM3D [13]. However, due to their local processing nature, these methods resulted in blurring or artifacts and therefore failed to preserve essential features for subsequent processing [14]. Lately, deep learning have been known popularly for its ability to show certain improvements in diverse applications related to classification [15, 16], detection [17–19], segmentation [20, 21]. The use of deep learning has no doubt reduce the work load as it eliminates the process of handcrafting the necessary features for processing. This is because deep learning is known to automatically extract relevant features without the need of human intervention [22]. This has led researchers to diverge into deep learning and explore its application in SAR interpretation because handcrafting features from SAR images is nearly impossible as even humans could not recognise the images properly due to the presence of speckle noises. Few works on SAR interpretation using deep learning includes [23–29]. These works are all based on deep learning comprising of the convolutional neural network (CNN) specifically designed for SAR purposes and resulted in better accuracies as compared to conventional methods. The improvements shown in these works have urged researchers in recent years to positively apply deep learning, particularly CNNs in SAR denoising as well. The performance of a CNN model usually depends on the way the network is tuned based on its configurations, which include the depth, number of units in each depth, form of non-linear functions, type of optimiser used, and hyperparameters such as learning rate, momentum, decay etc. Few state-of-the-art SAR denoising methods worth mentioning include the SAR-CNN [30] that first transformed the SAR image into a logarithmic space and then mapped it to a denoised image via a 17-layered CNN. Another similar work, referred to as SAR-DRN [31] uses dilated convolutions to learn a non-linear mapping between the noisy and clean SAR images. SAR-DRN also uses the skip connections to prevent the vanishing gradient problem and to retain relevant image information. ID-CNN [32] is another recent work whose architecture uses a division residual layer to calculate noise. It is an end-to-end based architecture that considers multiplicative noise without having to shift the image to a logarithmic space. These works have shown improvements in the field of SAR image denoising, and it may be observed that these works have inherited the properties of the famous ResNet architecture [33]. Motivated by these works, we have proposed a denoising architecture by exploring more on the behavior of ResNet and its usefulness in the application of SAR image despeckling. The following objectives are the main contributions of this paper.

- A deep CNN based end-to-end architecture for denoising SAR images is proposed.
- Quite a number of the famous ResNet’s skip connections are utilised in the architecture so as to preserve the relevant features of the image and to prevent gradients from vanishing during backpropagation.
- A formula pertaining to the use of skip connections in the architecture is developed.
- A hybrid loss function is also developed and adopted in training the proposed model.

The rest of the paper is organised as follows. Section 2 discusses the related works, which include the speckle noise, CNN, ResNet and state-of-the-art techniques used for denoising. Section 3 represents the proposed method describing the architecture and configurations used for denoising. Section 4 shows the experimental results and comparisons. Section 5 discusses the conclusion and future plan of work.

2 | RELATED WORK

In this section, the literature review related to denoising of SAR images is presented. Firstly, we described the speckle noise, followed by the description of the convolutional neural network and ResNet. Finally, we present the relevant state-of-the-art techniques used for SAR image denoising.

2.1 | Speckle noise

In this sub-section, a brief explanation of the type of noise referred to as speckle, which is dominant in SAR imagery, is discussed. Speckle noises are the grainy features present in images that occurs due to the influence of the rough surfaces of various ground targets on the electromagnetic waves. Also, the ability of SAR to operate in any weather conditions like rain, lightning, fog etc. makes it prone to speckle noises. In other words, the reflected echo may be out of phase depending on the roughness of the target surface and the path in which the echo travels before reaching the SAR receiver. The existence of speckle noise in image acquired by SAR, causes difficulty in understanding the various aspects of the image, which in turn, resulted in performance deterioration for target surveillance. To be specific, the type or nature of the target in SAR images is hard to be examined.

The existence of speckle noise in SAR images, however, makes it difficult to interpret or process them. Therefore, it becomes necessary to suppress the noise in SAR images in order to enhance the performance in various computer vision applications.
to apply it in denoising techniques as well. Ongoing investigations that use CNN to denoise SAR images resulted in compelling performances as compared to other conventional filtering techniques [30, 32]. Thus, with this motivation, we proposed a CNN-based architecture to enhance the clarity of SAR images by eliminating the speckle noise. It is also noteworthy to highlight that the configurations of CNN models differ depending on the type of application it is meant for. Next, we briefly discuss the different basic computational layers of a CNN in the following subsection.

2.2.1 Computational layers of CNN

The architecture of a CNN is organised as layers, whereby each layer performs computational operations on the inputs. A pictorial representation of a basic CNN architecture is presented in Figure 2. An input image is usually convolved to a set of feature maps with the help of convolutional layers whose dimensions are reduced using pooling layers. The final feature information is finally aggregated using the fully connected layer resulting in the final output. Some of the generally used mathematical models in the computational layers are discussed in the following [36].

Consider an image \( x \in \mathbb{R}^{h \times w \times c} \), where \( b \) is the height, \( w \) the width, and \( c \) is the number of channels. This image \( x \) along with a set of parameters \( W \), acts as input to a block in order to generate a new image \( y \in \mathbb{R}^{d \times w' \times c'} \) as output, that is, \( y = f(x; W) \). The entire process of generating the output goes through the following mathematical computations at respective layers.

1) **Convolution layer:** This layer generates the convolution of the input \( x \) with a bank of filters \( W \in \mathbb{R}^{b \times d \times c} \), also called kernels, and adds a bias \( b \in \mathbb{R}^{d'} \) [41]. This computation can be formulated as

\[
y_{i'j'k'} = f(b_{d'} + \sum_{i=1}^{b} \sum_{j=1}^{d} \sum_{d'=1}^{c} W_{i'd'}x_{i'j'+i,j'd+} \times x_{i'j'+i,j'd+}) \tag{4}
\]

where \( f(.) \) indicates a nonlinear activation function for which in the case of rectified linear unit (ReLU) [42]

\[
y_{ijk} = \max\{0, x_{ijk}\} \tag{5}
\]

2) **Pooling or subsampling layer:** This layer is used to downsample the output of the convolutional layer along each image channel in a \( b \times w' \) subwindow. One of such subsampling method is the max pooling method which can be written as [43].

\[
y_{i'j'k} = \max_{1 < i < b, 1 < j < w'} x_{i'j'+i,j'+} \tag{6}
\]

3) **Softmax layer:** This layer, when applied, is usually used to implement multi-class logistic regression mainly for classification purposes and is primarily implemented just before

**FIGURE 1** Visual representation of the noise model in SAR image
4) **Loss function:** Given the true label class $l$ for an input $x$, this function is used to compute the error with respect to the output of the network, and based on this, the network parameters are re-adjusted until the error is reduced. One such function is the log loss function which is computed as

$$y = -\sum_{i=1}^{h} \sum_{j=1}^{w} \log x_{ij}.$$  

(8)

### 2.3 ResNet or residual learning

One of the most pivotal works of recent years in computer vision called ResNet [33] is briefly explained in this section. ResNet, which is an acronym for Residual Network, is one of the known architectures of deep learning introduced in the year 2015. This model could train a network with 152 layers while at the same time it could handle the vanishing gradient issue, which normally occurs in deeper networks. The ability of ResNet to cope the vanishing gradient issue is because of the skip connections introduced in its architecture [33]. The skip connections also known as identity mappings does not require extra parameters. Instead, it only combines the output of the previous layer to the forthcoming layers. Figure 3 shows the basic building block of a ResNet model. As seen from Figure 3, the activation from the preceding layer is being appended to the activation of the next layer, and this small connection enabled ResNet to train deeper networks. ResNet has therefore, proved its capability to train deeper networks with less complexity compared to other architectures such as VggNet [46].

ResNet has recently been applied to various image denoising models such as ID-CNN [32], SAR-CNN [30], SAR-DRN [31]. The application of ResNet in these works has shown considerable improvement in performance, and this has motivated us towards exploring the ResNet along with its novel skip connections for possible application in SAR image denoising. The works in [30, 32, 47] have also utilised the skip connection from the ResNet architecture. However, they did not take full advantage of the characteristics offered by the traditional ResNet.

### 2.4 Existing noise filtering techniques

In this subsection, we highlighted some of the noise filtering methods proposed in the literature. Concerning the existence of noise in images, many researchers have come up with several noise removing techniques from statistical to deep learning approaches. Lee filter [9] is a filtering technique that calculates the central pixel value based on the statistical distribution of the pixel values of a moving kernel. The lee filtering method works based on variance, in the sense that it performs smoothing operation only when the variance of the intended area is low, resulting in the ability to smooth noise and preserve edges adaptively. The filter results in blurring and may bring about the loss of detailed target information when applied to SAR images [48]. Another statistical method named Kuan filter [10] converts the noise model from multiplicative to additive model, and it resembles the Lee filter, except that it uses an alternate
function for weighting. The filtered pixel replacement value is calculated based on the local statistics. Due to the local processing nature, the Kuan filter also results in artifacts and blurring, which, when applied to SAR images, might misinterpret its meaning [48]. Frost filter [11] is used to reduce noise, retaining the edge, and shape features by utilising exponentially damped convolution kernel. The trade-off of using a high value of damp factors resulted in better retaining of sharp edges. However, the noise reduction is less and vice versa. This trade-off makes Frost filter challenging when applied to SAR images. Another method named NLMeans [49] was brought forward, which follows a selective data-driven way of setting the weights by taking the average weight of the values of resembling pixels. The calculation of similar pixels is based upon the distance between the plot surrounding the central pixel and the plot surrounding a given nearby pixel. Due to the involvement of similar patch searching, the NLMeans filter has low computational efficiency resulting in its narrow application [31]. The BM3D [12] is a statistical image denoising method that clubs patches of an image into 3D arrays based on their resemblance and does combine filtering, thereby resulting in the estimation of 2D images for these grouped patches. SAR-BM3D [13] adopts the technique of BM3D for SAR image denoising by modifying the way of checking image patch resemblance based on probabilistic measure and taking into consideration the noise model, suppression of noise is performed based on the local linear minimum mean squared error. While the mechanism in [9, 10, 49] resulted in blurring and generation of artifacts, the techniques in [12, 13], are computationally intensive and therefore are less recommended for SAR systems [50].

Recent works using deep learning include the SAR-CNN [30], which map a SAR image to the corresponding noise image via a 17-layered CNN model. The predicted noise is then subtracted from the input image to obtain the filtered clean output. This technique, however, is a two-step process because the inputs are first log-transformed, and the output is then exponentiated back to the original image space. In comparison to SAR-BM3D, the SAR-CNN is capable of preserving more details in the despeckled image [30]. Another deep learning approach proposed for SAR image denoising is the SAR-DRN [31], which is a 7-layered network that uses dilated convolutions in place of the normal convolutions. It employs three skip connections: one from the first layer to the third layer and second from the fourth layer to the seventh layer, while the third one from the input directly to the output of the last layer. The SAR-DRN utilises the skip connections in order to prevent vanishing gradient and to maintain image details for better predictions. The key concept of SAR-DRN is the use of dilated convolutions resulting in larger receptive fields, preserving the size of the filter and the depth of the model. Though SAR-DRN outperforms other filters like SAR-BM3D, yet it produced artifacts contaminated images [50]. DnCNN [47] is another deep learning approach for image denoising that consists of 17 convolutional layers designed to handle Gaussian noise. It is an end-to-end mapping from the noisy image to the residual image irrespective of the gaussian noise model. The DnCNN incorporates residual learning, whereby the predicted residue is subtracted from the noisy input to generate the filtered image. The DnCNN, however requires two steps in producing the filtered image output as its model does not directly output the denoised image. One of the recent work named ID-CNN [32] was proposed for SAR image denoising whereby an 8-layered CNN is used to learn an end-to-end mapping from the noisy input to a clean output by strictly obeying multiplicative speckle model. Considering the noise model, the ID-CNN uses the division residual layer to divide the noisy input by the output of the last layer in order to predict a clean image. In addition to this, the ID-CNN combines the euclidean loss along with total variation loss as its loss function to train the model, and this resulted in the improvement of denoising accuracy. The ID-CNN, have shown better performance as compared to the above-mentioned techniques [32]. Table 1 shows some of the major properties of few existing deep learning based works on SAR image denoising. To further improve the performance of the state-of-art techniques, we proposed a model that uses a deeper network along with multiple skip connections and a hybrid loss function that are discussed in the next section.

3 | PROPOSED METHOD

In this section, we present our proposed CNN for denoising of SAR images. Firstly, the architecture that we consider for our work incorporates the ResNet architecture, as it can be best used when a model needs to learn as many useful and relevant features as it can, especially when these features gets distorted in the process, as in the case of denoising of images, whereby important features may get suppressed along with the unwanted speckles. Therefore, our network enables the re-concatenation of the noise information to the later layers while retaining the image information with the help of skip connections. Also, these connections act as gradient highways when weights adjustment are back-propagated to the earlier layers resulting in faster convergence and non-vanishing gradient. The detailed architecture of the proposed model is discussed in the next subsection, followed by a discussion on the skip connections used in our architecture.

| TABLE 1 | Major properties of the existing deep-learning networks used for SAR image denoising |
|-----------------|---------------|-----------------|
| Architecture    | SAR-CNN       | SAR-DRN         | ID-CNN         |
| No. of SC       | 17-Layered    | 7-Layered       | 8-Layered      |
| Loss function   | Custom        | MSE             | E + TV         |
| Type of convolution | Normal     | Dilated          | Normal         |
| Noise model     | M             | A               | M              |
| Input log transform | Yes         | No              | No             |

E = euclidean, TV = total variation, M = multiplicative, A = additive, SC = skip connections
3.1 Proposed architecture

Deeper networks are known to be significantly more proficient in creating deep representations by learning more abstract information about the input with the help of stacked layers. The benefit of a deep network is that it can learn a hierarchy of simple to sophisticated features. Therefore deeper the network, better is the feature portrayal. Also, considering the ImageNet [51] competitions’ results, the winning networks became deeper and deeper from initially 8 layers [52] to lately 152 layers [33]. The competition has proved the capability of deeper networks and hence can be applied to various applications. The application of neural networks in SAR image denoising, deeper convolutional networks have not been explored despite its several benefits. Therefore in our proposed architecture, we have used deeper networks for denoising and the proposed method have shown considerable improvement in performance as compared to state-of-the-art techniques. The works in [30–32] have also used the famous property of ResNet called skip connections in their architectures but did not take advantage of the properties offered by ResNet in the use of numerous skip connections for deeper networks by considering deeper architectures. Therefore, the architecture of the proposed model consists of 24 convolutional layers with 64 feature maps Convolved (Conv) by a 3 × 3 kernel window. However, we have also incorporated 12 skip connections that skips layers based on a formula we derived. The use of this traditional skip connection approach enables the model to learn the residue without forgetting the relevant features or information in order to be able to predict a desired clear image. Figure 4 shows the proposed architecture of our method. The architecture is made up of two types of layers:

- Type I: Conv + ReLU
- Type II: Conv + BN + ReLU

The first layer is made of Type I, and based on the strategy in [46], the kernel of size 3 × 3 is used for convolving the input to 64 features map. The 22 intermediate layers are made of Type II with 64 feature maps convolved by 3 × 3 kernels. The last layer is again made of Type I and works as a fully connected layer resulting in a single feature map with the help of a 3 × 3 kernel. In order to make the dimensions of the input and the output of each layer equivalent, we have added borders of zeros to every input of a convolutional layer. Turning to the activation function, the most used non-linear activation function in recent years is the ReLU [42] because of its simplicity and its linear-like behaviour that enables a network to be easily optimised. Therefore, we have also opted ReLU to be the activation function in every layer of our proposed network. To enable the layers in our network to independently learn a bit more by itself irrespective of what it learns from the previous layers, we have also used Batch Normalization (BNorm) [53] before every non-linear activation, except in the first and the last layer.

3.2 Skip connections

With the increase in the depth of the proposed network, we have also incorporated skip connections in our network in order to prevent the gradient from vanishing while it is being backpropagated to the earlier layers. The use of these skip connections has, however, helped our network in retaining the relevant information from the previous layers to the posterior layers resulting in the generation of better results. In order to maintain uniformity in the application of skip connections to our proposed architecture, we have derived a mathematical formula pertaining to the use of skip connections in our network. The derived mathematical formula is expressed as

$SC(L_n) = I(L_{n-1}) + O(L_n)$,  \hspace{1cm} (9)

where $n > 0$ and $n \% 2 = 0$. The skip connection at the $n^{th}$ layer $L_n$ indicates the connection of the input to $(n - 1)^{th}$ layer and the output of the $n^{th}$ layer.

We have also used one super-stretch skip from the input to the last layer, whereby it acts as a division residual layer that divides the input image by the output of the last layer in order to gain the despeckled image. We have used the division residual layer because, as per the prior information about the nature of speckle that is inherent in SAR images, the clean image can be obtained by dividing the noisy image by the noise component. Hence the residual layer is expected to yield the desired despeckled information, after which a hyperbolic tangent (Tanh) is then applied to this output as an activation function.
### 3.3 Loss function

The ability of a CNN to work the way it is modeled depends on the estimation of the errors generated during the learning process. The model learns with respect to these errors and is calculated with the help of a function called the loss function that depicts how much a network misses the required target. If the prediction of the model is inaccurate, the loss function will yield a higher value, whereas it will generate a lower value for near accurate predictions. The mean squared error and the mean absolute error have mostly been used as loss functions in previous CNN based image restoration. The model in SAR-DRN [31] have used the popular mean squared error as their loss function in optimising their network to learn the denoised SAR image, whereas, with the motive of allowing image restoration as well as smoothness on the resulted image, ID-CNN [32] has combined the euclidean loss (L2) with the total variation loss as their loss function which resulted in better learning of the model.

L2 is the most used loss function in image restoration tasks because of its simplicity and stable solution. However, it does not seem to correlate strongly with image quality and produces artifacts. Therefore, L2 is more sensitive to anomalies but gives a progressively steady solution. The other popular loss known as the L1 or the mean absolute error loss is vigorous to outliers and hence used to enhance the image quality by reducing such artifacts [54]. However, the derivatives of L1 are not continuous. Therefore, L1 may not converge to the best solution. Our objective is to design a model that converges to the best or near optimal solution, and that is less affected by outliers. Therefore, we venture into loss functions that incorporate the advantages offered by both L1 and L2 and observed that the pseudo Huber loss function [55] is one such loss function that combines the best properties of both L1 and L2. The pseudo Huber loss is a modified version of the Huber loss [56]. It is known to be less receptive to outliers and usually bends around the minima that diminishes the gradient. The pseudo Huber loss is computed as

\[
P_\delta = \sum_{i=1}^{n} \delta^2 \sqrt{1 + \left( \frac{X_i - \hat{X}_i}{\delta} \right)^2} - 1, \tag{10}
\]

where \(X\) is the true output and \(\hat{X}\) is the predicted output, and \(\delta\) is a parameter that measures the steepness of the loss function.

On the other hand, the total variation loss [57] is best known for simultaneously preserving edges and smoothing away noise when used for denoising. Therefore, we have fused this loss with the pseudo Huber loss as the loss function to train our model. The total variation calculated on an image with a height \(H\) and width \(W\) is determined as

\[
TV = \sum_{x=1}^{W} \sum_{y=1}^{H} \sqrt{(\hat{X}_{x+1,y} - \hat{X}_{x,y})^2 + (\hat{X}_{x,y+1} - \hat{X}_{x,y})^2}. \tag{11}
\]

Therefore, the comprehensive loss function is exemplified as

\[
Loss = P_\delta + \lambda TV, \tag{12}
\]

where \(\lambda\) is a regularisation parameter whose value strictly reflects on the ability of TV to achieve the right amount of noise removal. It may be mentioned that we have investigated on the working of the pseudo Huber loss independently and realised that it resulted in a very low loss as compared to the total variation that resulted in a large loss. Hence to keep these losses at par with each other, we tuned its dependent parameters \(\delta\) and \(\lambda\) in a manner so as to yield the best outcome out of them. The values chosen for these parameters, along with the implementation and results of our proposed method is discussed in the next section.

### 4 IMPLEMENTATION RESULTS AND DISCUSSION

In this section, we discuss the implementation and performance of our proposed work. Results on both synthetic and real SAR images are also highlighted and compared with five denoising algorithms, which comprises of the known filtering works named Lee Filter [9], NLMeans [49] and state-of-the-art methods called SAR-DRN [31], SAR-CNN [30], and IDCNN [32]. For deep learning-based works such as IDCNN, we have strictly implemented, as mentioned in their paper. Similarly, for SAR-CNN we have implemented by setting the architecture and parameters precisely as mentioned therein, except that we have used MSE as their loss function. Likewise, for SAR-DRN, we have implemented the way it is specified in their paper, including parameters and architecture. Since the number of epochs for training has not been mentioned in most of these works, therefore, we have set the number of epochs to 50 in all our experiments, including the proposed work. It may be mentioned that our proposed method is based on an end-to-end method, and the inputs to our model need not have to be transformed into different image space.

The proposed model is learning both the noise as well as the relevant information that resulted in the contribution of generating an output close to the desired image. The difference between our method and the above previous deep learning-based works is that, in our work, we have considered a deeper model for denoising of SAR images with multiple skip connections, and to the best of our knowledge, deeper models have not been ventured for denoising of SAR images. Also, a different loss function that was never used before for training a denoising model has been adopted. The proposed method has shown considerable improvements compared to state-of-the-art techniques.

We have implemented our proposed method using the benchmark for remote sensing image scene classification (RESISC), created by North Western Polytechnical University (NWPU) [58]. From this benchmark, we have prepared a dataset that consists of 3500 images, in which 3419 random images are used for training our model, and the remaining 81 images...
are saved for testing. Figure 5 depicts few images present in the dataset. We have resized the images to 256 × 256 before using them in training our model. After our model is trained and tested on synthetic images, we have also extended the test on real SAR images, which comprises of images acquired by the SAR sensor mounted on TerraSAR-X Radar, European Space Agency (ESA) [59].

It may be mentioned that there is a need for a dataset containing both clean and noisy SAR images in order to successfully train neural networks like CNN for SAR image despeckling purposes. However, the unavailability of ground truth images which are the clean SAR images, is a major concern while training the CNN for despeckling models. Hence, to handle this critical issue, while not eliminating the use of CNN in the proposed method, the optical images which are the RESISC datasets [58], have been artificially simulated with speckles, adhering to the characteristics of SAR. The addition of speckle to the RESISC datasets is achieved by following the noise model shown in Equation (1). The resulting noisy synthetic images and their corresponding clean labels are then used to train the proposed CNN model in a supervised manner. Therefore, the main advantage of using CNN is its capability to learn the features closest to that of SAR even if non-SAR image is used while training the model, provided the training images have similar characteristics with the intended targets. The speckle that is artificially added to the training dataset follows a gamma distribution having probability density function shown in Equation (2). A sample distribution of the speckle with look 4 is shown in Figure 6. Figure 1 depicts how we diffused speckle to one of the training samples following the distribution shown in Figure 6.

While training our model, the adaptive moment (ADAM) [60] is used as an optimiser for optimising the network with a learning rate of 0.0002. Batch size of 10 is used in training our model, and the regularisation parameters \( \lambda \) and \( \delta \) are set to \( \lambda = 8 \) and 1.5, respectively. Larger values of \( \delta \) results in slow convergence of our model, while smaller values of \( \delta \) enables outliers to affect our model. Hence, a proper value of \( \delta \) is chosen as stated above that best suits our model. The results of our implementation on synthetic as well as real SAR images are shown in the next subsection.

4.1 Results on estimated or synthetic SAR images

Experiments on synthetic images has been carried out, and the performances of various experiments including the proposed method has been measured with the help of several popular metrics used for image restoration such as the structural similarity index measure (SSIM) [61], the peak signal to noise ratio (PSNR), the universal quality index (UQI) [62] and the mean squared error (MSE). The benchmark images have been synthetically added noise before using them into our experiments. We have also tested our model on various noise levels 1, 4, and 10, respectively, and compared with state-of-the-art methods. The results of our implementation on various denoising work along with the proposed method are shown in Table 2 and the graphical representations of the simulation results with respect to PSNR, SSIM and UQI are shown in Figures 7, 8 and 9, respectively. It is observed that our proposed method achieved higher PSNRs and UQIs in almost all the looks. Our method also achieved higher SSIMs in all the looks as compared to other filters, and our proposed method resulted in minimum errors as compared to other methods. Sample results on synthetic images are also shown in Figure 10. It is observed that the proposed method shows considerable improvements in the suppression of noise, resulting in the enhancement of synthetic SAR images. Though the results of the proposed method shown in Figure 10 appear to be closer to that of IDCNN, however; our method preserve more image information as compared to other methods. Sample results on synthetic images are also shown in Figure 10. It is observed that the proposed method shows considerable improvements in the suppression of noise, resulting in the enhancement of synthetic SAR images. Though the results of the proposed method shown in Figure 10 appear to be closer to that of IDCNN, however; our method preserve more image information as compared to IDCNN. To justify this claim, we have used histograms since histograms are widely used for picturing a general understanding of the intensity distribution of an image [63–65]. Therefore, we have plotted histograms of the bottom most image of Figure 10. The histograms plotted comprises of a noisy, clean, a filtered image of...
TABLE 2 Synthetic data results and comparisons

| Filter Type | Metrics Used | Looks = 1 | Looks = 4 | Looks = 10 |
|-------------|--------------|-----------|-----------|------------|
| Lee         | PSNR         | 13.44     | 18.49     | 20.47      |
|             | SSIM         | 0.279     | 0.500     | 0.610      |
|             | UQI          | 0.855     | 0.934     | 0.962      |
|             | MSE          | 0.0472    | 0.016     | 0.010      |
| NLMeans     | PSNR         | 7.577     | 14.66     | 18.15      |
|             | SSIM         | 0.156     | 0.447     | 0.655      |
|             | UQI          | 0.687     | 0.917     | 0.974      |
|             | MSE          | 0.1943    | 0.039     | 0.010      |
| SAR-DRN     | PSNR         | 22.30     | 24.32     | 24.53      |
|             | SSIM         | 0.519     | 0.621     | 0.636      |
|             | UQI          | 0.893     | 0.936     | 0.941      |
|             | MSE          | 0.0074    | 0.0041    | 0.0038     |
| SAR-CNN     | PSNR         | 21.67     | 24.03     | 26.03      |
|             | SSIM         | 0.583     | 0.722     | 0.819      |
|             | UQI          | 0.975     | 0.985     | 0.991      |
|             | MSE          | 0.0073    | 0.0039    | 0.0026     |
| ID-CNN      | PSNR         | 20.33     | 24.23     | 26.18      |
|             | SSIM         | 0.550     | 0.776     | 0.844      |
|             | UQI          | 0.967     | 0.984     | 0.991      |
|             | MSE          | 0.0101    | 0.0041    | 0.0026     |
| Proposed    | PSNR         | 22.01     | 24.60     | 27.02      |
|             | SSIM         | 0.633     | 0.784     | 0.853      |
|             | UQI          | 0.974     | 0.987     | 0.995      |
|             | MSE          | 0.0071    | 0.0039    | 0.0023     |

4.2 Results on actual or real SAR images

In this subsection, we present the results achieved while testing our model on real SAR images acquired by TerraSAR-X radar. Figure 12 shows sample SAR images predicted by our model as well as those produced by other techniques. The results are examined with the help of the equivalent number of looks.
(ENL) [66] metric that depicts the noise suppression ability of the model on these images. The ENL metric is a measure that is regarded as the quantitative evaluation index used for original SAR image denoising experiments in the absence of ground truth images [67]. Thus the ENL metric can be used to infer the speckle reduction capability and is computed as shown in Equation (13).

\[
ENL = \frac{\text{mean}^2}{\text{variance}}
\]  

(13)

**TABLE 3** Real SAR data ENL results and comparisons

| Filters Types | Chip No. 1 | Chip No. 2 | Chip No. 3 | Chip No. 4 |
|---------------|------------|------------|------------|------------|
| Lee           | 110.44     | 89.23      | 10.14      | 17.07      |
| NLMeans       | 158.39     | 251.84     | 696        | 118.47     |
| SAR-DRN       | 76.54      | 81.53      | 87.06      | 67.86      |
| SAR-CNN       | 160.06     | 170.91     | 132.91     | 116.35     |
| ID-CNN        | 132.31     | 129.82     | 98.60      | 151.03     |
| Proposed      | **167.20** | **162.04** | **219.68** | **157.01** |
From Equation (13), it can be seen that the ENL is obtained by dividing the mean of the homogeneous region by the variance of the same homogeneous region of the despeckled image. Four random patches from homogeneous regions drawn in white boxes, as shown in Figure 12, are extracted for examining the ENL on SAR images predicted by each filter. The results are presented in Table 3. It is observed that for chips 1 and 4, the proposed technique achieved higher ENL as compared to the other state-of-the-art methods. Even for chips 2 and 3, the proposed method has higher ENL in comparison to other methods except when compared to NLMeans. However, the higher ENL values of chips 1 and 2 in NLMeans is due to the fact of its ability to smooth images, especially for regions around the center of the images while failing to preserve the edges as can be seen in Figure 12. It is also observed in Figure 12 that our proposed method has the right balance of smoothing the images while also preserving necessary edges in the denoised SAR images.

4.3 Comparisons

This subsection highlights the comparison of our proposed method with state-of-the-art methods used for comparing our work. As seen from the experimental results discussed in the previous section, the proposed method outperforms the recent state-of-the-art work named IDCNN. Therefore, the incorporation of deeper CNN in our architecture has proven its contribution in the improvement of SAR image denoising performance. Though one of the works named SAR-CNN has used quite a number of convolutional layers in its denoising architecture, it has not taken advantage of ResNets’ skip connections by using them numerously. Unlike IDCNN, SAR-DRN, and SAR-CNN, our proposed network consists of numerous skip connections that have played its role in enabling the network to learn the speckle without forgetting the image information, while simultaneously handling the vanishing gradient issue during the process of backpropagation. In addition to this, none of the existing works have used the pseudo Huber loss while training the model for denoising of SAR images. Incorporating the pseudo Huber loss in the proposed loss function has also helped train our model to converge to a near optimal solution. Table 4 highlights a comparison summary of the proposed method with a few state-of-the-art techniques.

5 CONCLUSION

In this paper, we have used deeper networks for despeckling of SAR images and have incorporated numerous skip connections from ResNet in our architecture. As far as we are concerned, our work is the only method that uses very deep convolutional neural networks in denoising of SAR images. We have also used the pseudo Huber loss as a part of the loss function to train our model, and this has helped our network to converge to a near optimal solution, while at the same time reducing artifacts. The advantage of the proposed model is that it can learn more features from deeper networks with a controlled number of parameters due to the presence of numerous skip connections. Moreover, the usage of multiple skip connections in the proposed model can also simultaneously handle the vanishing gradient issue during backpropagation. The proposed method has been tested on both synthetic as well as real SAR images and compared with various state-of-the-art methods. The proposed method has shown considerable improvements in restoring noisy SAR images, and it is an end-to-end based method. As future works, we will be investigating numerous parameters of the network to further enhance the denoising performance in SAR images and taking into consideration the inhomogeneity of SAR images, we will be adopting some other noise suppression metrics along with ENL for evaluating the performance on real SAR images.

REFERENCES

1. Moreira, A., et al.: A tutorial on synthetic aperture radar. IEEE Geosci. Remote Sens. Mag. 1(1), 6–43 (2013)
2. Chaturvedi, S.K.: Study of synthetic aperture radar and automatic identification system for ship target detection. J. Ocean. Eng. Sci. 4(2), 173–182 (2019)
3. Yirong, S., Zengping, C.: Evaluation and simulation of lpi radar signals’ low probability of exploitation. In: 2017 2nd International Conference on Image, Vision and Computing (ICIVC), pp. 842–846.IEEE, Piscataway (2017)
4. DeFord, J.F., Gandhi, O.P.: Phase-only synthesis of minimum peak sidelobe patterns for linear and planar arrays. IEEE Trans. Antennas Propag. 36(2), 191–201 (1988)
5. Li, Y., et al.: MIMO radar transmit beampattern design for DOA estimation with sidelobe suppression. Int. J. Antennas Propag. 2016, 1–10 (2016)
6. Yi, L., Yu, Y.: Dual-frequency MIMO antennas with reduced mutual coupling. In: 2014 Asia-Pacific Microwave Conference, pp. 507–509.IEEE, Piscataway (2014)
7. Bolomey, J., et al.: Rapid near-field antenna testing via arrays of modulated scattering probes. IEEE Trans. Antennas Propag. 36(6), 804–814 (1988)
8. Case, J.T., et al.: Correcting mutual coupling and poor isolation for real-time 2D microwave imaging systems. IEEE Trans. Instrum. Meas. 63(5), 1310–1319 (2014)
9. Lee, J.: Digital image enhancement and noise filtering by use of local statistics. IEEE Trans. Pattern Anal. Mach. Intell. PAMI-2(2), 165–168 (1980)
10. Kuan, D.T., et al.: Adaptive noise smoothing filter for images with signal-dependent noise. IEEE Trans. Pattern Anal. Mach. Intell. PAMI-7(2), 165–177 (1985)
11. Frost, V.S., et al.: A model for radar images and its application to adaptive digital filtering of multiplicative noise. IEEE Trans. Pattern Anal. Mach. Intell. PAMI-4(2), 157–166 (1982)
12. Dahov, K., et al.: Image Denoising with Block-Matching and 3D Filtering. In: Nasrabadi, N.M., Rivzi, S.A., Dougherty, E.R., Astola, J.T., Egiazarian, K.O. (eds.) Image Processing: Algorithms and Systems, Neural Networks, and Machine Learning, vol. 6064, pp. 354–365. International Society for Optics and Photonics, SPIE, Washington DC (2006)
13. Parrilli, S., et al.: A non-local SAR image denoising algorithm based on LLMMS wavelet shrinkage. IEEE Trans. Geosci. Remote Sens. 50(2), 606–616 (2012)
14. Patel, V.M., et al.: Separated component-based restoration of speckled SAR images. IEEE Trans. Geosci. Remote Sens. 52(2), 1019–1029 (2014)
15. Raj, A.P.S.S., Vajravelu, S.K.: DDLA: Dual deep learning architecture for classification of plant species. IET Image Proc. 13, 2176–2182 (2019)
16. Cheng, G., et al.: Exploring hierarchical convolutional features for hyperspectral image classification. IEEE Trans. Geosci. Remote Sens. 56(11), 6712–6722 (2018)
17. Lee, S.L., Zare, M.R.: Biomedical compound figure detection using deep learning and fusion techniques. IET Image Proc. 12, 1031–1037 (2018)
18. Zhou, L., et al.: Deep multiple instance learning for automatic detection of diabetic retinopathy in retinal images. IET Image Proc. 12, 565–571 (2018)
19. Sabokrou, M., et al.: Deep-anomaly: Fully convolutional neural network for deep anomaly detection in crowded scenes. Comput. Vision Image Understanding 172, 68–97 (2018)
20. Huang, M., et al.: Medical image segmentation using deep learning with feature enhancement. IET Image Proc., (Early Access) 1–10 (2020)
21. Kallenberg, M., et al.: Unsupervised deep learning applied to breast density segmentation and mammographic risk scoring. IEEE Trans. Med. Imaging 35(5), 1322–1331 (2016)
22. LeCun, Y., et al.: Backpropagation applied to handwritten zip code recognition. Neural Comput. 1(4), 541–551 (1989)
23. Zhong, C., et al.: Classification for SAR scene matching areas based on convolutional neural networks. IEEE Geosci. Remote Sens. Lett. 15(9), 1377–1381 (2018)
24. Cao, X., et al.: SAR image change detection based on deep denoising and CNN. IET Image Proc. 13(9), 1509–1515 (2019)
25. Chen, S., et al.: Target classification using the deep convolutional networks for SAR images. IEEE Trans. Geosci. Remote Sens. 54(8), 4806–4817 (2016)
26. Wang, C., et al.: Sea ice classification with convolutional neural networks using sentinel-1 scatter images. In: IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium, pp. 7125–7128. IEEE, Piscataway (2018)
27. Shang, R., et al.: Target classification based on deep memory convolutional neural networks and transfer parameters. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 11(8), 2834–2846 (2018)
28. Fukukawa, H.: Deep learning for end-to-end automatic target recognition from synthetic aperture radar imagery, arXiv preprint arXiv:180108558, 2018
29. Pei, J., et al.: SAR automatic target recognition based on multiview deep learning framework. IEEE Trans. Geosci. Remote Sens. 56(4), 2196–2210 (2018)
30. Chierchia, G., et al.: SAR image despeckling through convolutional neural networks. In: 2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), pp. 5438–5441. IEEE, Piscataway (2017)
31. Zhang, Q., et al.: Learning a dilated residual network for SAR image despeckling. Remote Sens. 10(2), 196 (2018)
32. Wang, P., et al.: SAR image despeckling using a convolutional neural network. IEEE Signal Process. Lett. 24(12), 1763–1767 (2017)
33. He, K., et al.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778. IEEE, Piscataway (2016)
34. Tur, M., et al.: When is speckle noise multiplicative? Appl. Opt. 21(7), 1157–1159 (1982)
35. Ranjani, J.J., Thiruvengadam, S.J.: Dual-tree complex wavelet transform based SAR despeckling using interscale dependence. IEEE Trans. Geosci. Remote Sens. 48(6), 2723–2731 (2010)
36. Ding, J., et al.: Convolutional neural network with data augmentation for SAR target recognition. IEEE Geosci. Remote Sens. Lett. 13(3), 364–368 (2016)
37. Cho, J.H., Park, C.G.: Multiple feature aggregation using convolutional neural networks for SAR image-based automatic target recognition. IEEE Geosci. Remote Sens. Lett. 15(12), 1882–1886 (2018)
38. Wang, Z., et al.: SAR target detection based on SSD with data augmentation and transfer learning. IEEE Geosci. Remote Sens. Lett. 16(1), 150–154 (2019)
39. Saha, S.: A comprehensive guide to convolutional neural networks. https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53, (2018), pp. 12–15. Accessed 02 Feb 2020
40. Lecun, Y., et al.: Gradient-based learning applied to document recognition. Proc. IEEE 86(11), 2278–2324 (1998)
41. LeCun, Y., et al.: Convolutional networks and applications in vision. In: Proceedings of 2010 IEEE International Symposium on Circuits and Systems, pp. 253–256. IEEE, Piscataway (2010)
42. Nair, V., Hinton, G.E.: Rectified linear units improve restricted Boltzmann machines. In: Proceedings of the 27th International Conference on Machine Learning, pp. 807–814. Omnipress, Madison (2010)
43. Stanford University: CS231n convolutional neural networks for visual recognition. https://cs231n.github.io/convolutional-networks/, (2020), pp. 2–7. Accessed 12 Feb 2020
44. Bishop, C.M.: Probability Distributions. Pattern Recognition and Machine Learning, Information Science and Statistics.1. Springer, Verlag New York (2006)
45. Christophe Pere: What are loss functions? https://towardsdatascience.com/what-is-loss-function-1e2605ab904, (2017), pp. 6–17. Accessed 14 Jan 2020
46. Simonoy, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint, arXiv:14091556, 2014
47. Zhang, K., et al.: Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising. IEEE Trans. Image Process. 26(7), 3142–3155 (2017)
48. Wang, J., et al.: Ground target classification in noisy SAR images using convolutional neural networks. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 11(11), 4180–4192 (2018)
49. Buades, A., et al.: A non-local algorithm for image denoising. In: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’05), vol. 2, pp. 60–65. IEEE, Piscataway (2005)
50. Lattari, F., et al.: Deep learning for SAR image despeckling. Remote Sens. 11(13), 1532 (2019)
51. Russakovsky, O., et al.: ImageNet large scale visual recognition challenge. Int. J. Comput. Vision (IJCV) 115(3), 211–252 (2015)
52. Krizhevsky, A., et al.: Imagenet classification with deep convolutional neural networks. In: Advances in Neural information processing systems, pp. 1097–1105. Curran Associates, Red Hook (2012)
53. Ioffe, S., Szegedy, C.: Batch normalization: Accelerating deep network training by reducing internal covariate shift. arXiv preprint arXiv:150203167, 2015, pp. 1–11
54. Zhao, H., et al.: Loss functions for image restoration with neural networks. IEEE Trans. Comput. Imaging 3(1), 47–57 (2017)
55. Hartley, R., Zisserman, A.: Iterative Estimation Methods: Multiple View Geometry in Computer Vision, 2nd edn. Cambridge University Press, New York (2003)
56. Huber, P.J.: Robust estimation of a location parameter. Ann. Math. Statist. 35(1), 73–101 (1964)
57. Strong, D., Chan, T.: Edge-preserving and scale-dependent properties of total variation regularization. Inverse Prob. 19(6), S165–S187 (2003)
58. Cheng, G., et al.: Remote sensing image scene classification: Benchmark and state of the art. Proc. IEEE 105(10), 1865–1883 (2017)
59. ESA online dissemination homepage_2020: Tpm-ds.esa.int. https://tpm-ds.esa.int/oads/access/, (2020). Accessed 14 Nov 2020
60. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint, arXiv:1412.6980, 2014, pp. 1–15
61. Wang, Z., et al.: Image quality assessment: From error visibility to structural similarity. IEEE Trans. Image Process. 13(4), 600–612 (2004)
62. Wang, Z., Bovik, A.C.: A universal image quality index. IEEE Signal Process. Lett. 9(3), 81–84 (2002)
63. Sutton, E.: Histograms and the zone system. Illustrated Photography 6–12 (2016)
64. Bovik, A.C.: 2.1 - Basic Gray-Level Image Processing. In: Bovik, A. (ed.) Handbook of Image and Video Processing (Communications, Networking and Multimedia), pp. 21–37, 2nd edn. Academic Press, Burlington (2005)
65. Liu, S., et al.: Group sparse representation based dictionary learning for SAR image despeckling. IEEE Access 7, 30809–30817 (2019)
66. Oliver, C., Quegan, S. Fundamental Properties of SAR Images. Understanding Synthetic Aperture Radar Images. SciTech Publishing, Raleigh (2004)
67. Di Martino, G., et al.: Benchmarking framework for SAR despeckling. IEEE Trans. Geosci. Remote Sens. 52(3), 1596–1615 (2014)

How to cite this article: Passah A, Amitab K, Kandar D. SAR image despeckling using deep CNN. IET Image Process. 2021;15:1285–1297. https://doi.org/10.1049/ipr2.12104