SAR Image Despeckling Using Convolutional Neural Network Ensemble

*Abstract- Synthetic Aperture Radar (SAR) images are crucial in remote sensing applications, but they are inherently corrupted by multiplicative speckle noise, which degrades image quality and impedes accurate interpretation. This paper proposes a novel deep learning approach utilizing a Convolutional Neural Network (CNN) ensemble with multiple Dilated Residual Networks (DRNs) to effectively despeckle SAR images. The proposed method aims to enhance image quality by reducing noise while preserving critical texture and edge information.*

#### *I. INTRODUCTION*

Synthetic Aperture Radar (SAR) imaging is a vital technology in earth observation, providing critical information for applications such as terrain mapping, environmental monitoring, and disaster management [1]. However, SAR images are inherently contaminated by multiplicative speckle noise, which significantly reduces image quality and compromises subsequent analysis [2].

Traditional despeckling methods, including spatial and transform domain filtering, have limitations in preserving fine details and maintaining image resolution. Recently, deep learning techniques, particularly Convolutional Neural Networks (CNNs), have demonstrated promising results in image denoising and restoration [3].

###### Research Objectives

* Develop a CNN-based ensemble approach for SAR image despeckling
* Improve image quality while preserving spatial details
* Demonstrate superior performance compared to traditional despeckling methods

***II. RELATED WORKS***

**A. SAR Image Speckle Noise Model**

Speckle noise in Synthetic Aperture Radar (SAR) images is a significant challenge that affects the accuracy of various applications such as feature extraction and target detection. Speckle noise manifests as nonuniform brightness or darkness regions in an image, typically spanning several pixels. Unlike other types of noise, speckle noise is primarily caused by system limitations in radar hardware and software, rather than environmental factors.

The generation of speckle noise is influenced by multiple factors, including the inherent noise of the SAR system, nonlinearities, signal transmission losses, receiver noise, and the non-ideal characteristics of the antenna. Because speckle noise is coherent, it is more challenging to remove than additive noise typically found in optical images.

Extensive research has been conducted to model the formation of speckle noise mathematically. The most widely used model assumes that the multiplicative noise in SAR images follows a gamma distribution:

where represents the observed noisy SAR image, is the true underlying image intensity, and is the coherent speckle noise, which follows a generalized gamma distribution:

where is the number of looks in the SAR image, is the gamma function, and . The number of looks is determined by the ratio of the mean intensity to variance:

where and represent the mean and variance of the SAR image intensity. By manipulating the number of looks, different levels of speckle noise can be simulated, allowing for the generation of clean-noisy image pairs that are essential for training deep learning-based denoising models.

However, despite the effectiveness of simulated noise for training purposes, many deep learning-based despeckling algorithms struggle to generalize well to real-world SAR images. To address this issue, reference-free evaluation metrics are often used to assess model performance on real SAR images.

***B. Image Restoration Methods***

Image restoration techniques can be broadly categorized into learning-based methods and model-based methods.

1. **Learning-Based Methods**: With advances in computational power, especially the widespread use of GPUs, deep learning-based approaches have become the dominant solution for image denoising. Learning-based methods train a denoising model using large datasets of clean-noisy image pairs. These methods have demonstrated strong performance, even in blind denoising tasks where the noise level is unknown.

Deep learning-based denoising for SAR images has seen significant improvements over time. While early neural network-based despeckling models often struggled to outperform traditional methods such as BM3D, recent advances in deep learning architectures have led to the development of more efficient models. Techniques such as residual learning and batch normalization, as seen in DnCNN, have enhanced network training stability and performance. Further optimizations, including dilated convolutions (IRCNN) and noise level maps (FFDNet), have also contributed to computational efficiency and improved despeckling performance.

1. **Model-Based Methods**: Traditional model-based despeckling methods rely on manually designed priors and optimization techniques rather than learning from data. These methods are formulated as an optimization problem:

where represents the data fidelity term (e.g., Mean Squared Error), and is a regularization term incorporating prior knowledge.

The choice of prior plays a crucial role in determining the effectiveness of model-based methods. Different priors lead to different denoising techniques. For instance, Total Variation (TV) regularization is effective for piecewise constant signals but struggles with fine textures and edges. Wavelet-based methods offer better performance for restoring local structures, while BM3D exploits the nonlocal self-similarity of natural images by constructing a block-based Gaussian mixture model.

While traditional model-based methods leverage strong priors, deep learning-based methods instead learn the mapping from degraded to clean images. Some studies suggest that the structure of convolutional networks implicitly captures statistical information about images rather than true learning-based reconstruction.

***C. Residual Learning***

Recent developments in deep learning-based SAR image restoration have leveraged residual learning techniques to enhance despeckling performance. Traditional CNN-based approaches for SAR denoising often suffer from poor generalization when trained on simulated noise due to the unknown nature of real SAR noise distributions.

One promising approach, SAR-CNN, transforms SAR images into logarithmic space, increases network depth, and employs a discriminative model to reduce speckle noise. However, deep networks face two major challenges: (1) difficulty in convergence due to vanishing gradients, and (2) diminishing returns in performance as network depth increases, leading to deep network degradation.

Residual learning, initially introduced in CNN architectures for classification and object detection, addresses these issues by allowing deeper networks to be trained efficiently. Residual networks implement identity mappings that enable direct propagation of information through deep layers, ensuring that deeper architectures maintain strong performance. By incorporating residual learning strategies, modern SAR despeckling networks achieve improved generalization and robustness, making them more effective in real-world applications.

#### *3. PROPOSED METHODOLOGY*

This paper presents a novel multi-branch dilated residual network integrated with a U-NET architecture (MB-DRN-UNET) for effective speckle reduction in Synthetic Aperture Radar (SAR) imagery. The proposed framework leverages the strengths of convolutional neural networks, dilated residual networks of varying dilation rates, and the UNET architecture to achieve superior despeckling performance.

### *3.1 Network Architecture*

The proposed architecture consists of four main components as shown in Fig 1: (1) a CNN encoder, (2) a multi-branch dilated residual network (DRN) module with varying dilation rates, (3) an edge-preserving content aggregation (ECA) module, and (4) a UNET decoder. Fig. 1 illustrates the overall architecture of our proposed method.

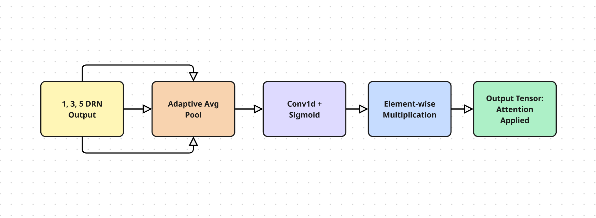
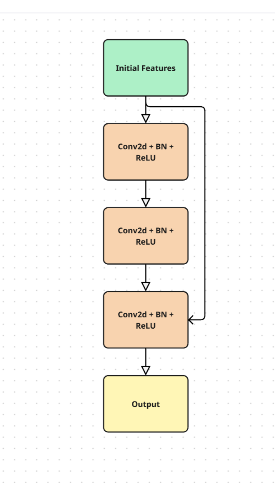
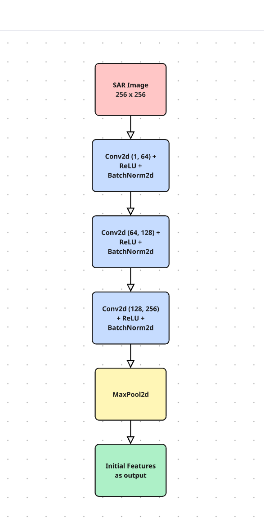
* + 1. CNN Encoder: Initial feature extraction from SAR image
    2. Dilated Residual Networks (DRNs):
  + 1-DRN: First residual network pathway
  + 3-DRN: Second residual network pathway
  + 5-DRN: Third residual network pathway
    1. Efficient Channel Attention (ECA): Combining multiple DRN outputs
    2. UNET: Final image reconstruction
    3. Despeckled SAR Image: Refined SAR image as output

#### 

#### *Fig 1: Model Architecture*

#### *3.1.1 CNN Encoder*

The input SAR image is first processed by a CNN encoder comprising multiple convolutional layers.

As shown in *Fig 1.1 (a),* the encoder extracts hierarchical features from the input noisy SAR image. It consists of three convolutional layers with progressive channel expansion followed by a max pooling operation. Each convolutional layer employs a kernel size of 3×3 with padding of 1 to maintain spatial dimensions. The ReLU activation and batch normalization help improve training stability and convergence. The input SAR image with dimensions 256×256 is processed through a sequence of Conv2d (1→64) + ReLU + BatchNorm2d, followed by Conv2d (64→128) + ReLU + BatchNorm2d, and then Conv2d (128→256) + ******ReLU + BatchNorm2d. Finally, a MaxPool2d layer is applied to obtain the initial feature maps.

(a) (b)

*Fig 1.1 (a) CNN Encoder Architecture, (b) Diluted Residual Network Module Architecture*

#### *3.1.2 Multi-Branch Dilated Residual Network*

The encoded features are then processed through three parallel branches of dilated residual networks, each with a different dilation rate (1, 3, and 5) as pictured in *Fig 1.1 (b)*. This multi-branch design allows the network to capture multi-scale contextual information while maintaining computational efficiency:

* ***1-DRN Branch:*** Employs conventional residual blocks with a dilation rate of 1, capturing local texture details.
* ***3-DRN Branch:*** Uses dilated convolutions with a dilation rate of 3, expanding the receptive field to capture medium-scale contextual information.
* ***5-DRN Branch:*** Implements dilated convolutions with a dilation rate of 5, further enlarging the receptive field to capture large-scale structural information.

Each DRN branch consists of N residual blocks, where each block contains:

* Three dilated convolutional layers
* A residual connection
* Batch normalization and ReLU activation functions

For a DRN block with dilation rate , the operation can be defined as:

where is the input feature maps is the output feature maps, represents the weights of the convolutional layers with dilation rate, and denotes the function of three consecutive dilated convolutional layers with batch normalization and ReLU activation.

The dilated convolution operation expands the receptive field without increasing the parameter count, which is crucial for effectively capturing the contextual information necessary for speckle reduction.

#### *Fig 1.2 Efficient Channel Attention Architecture*

#### *3.1.3 Efficient Channel Attention (ECA) Module*

The ECA module adaptively recalibrates channel-wise feature responses by explicitly modeling interdependencies between channels. The module proposed as in *Fig 1.2* processes the outputs from all three DRN blocks through adaptive average pooling to aggregate spatial information, followed by 1×1 convolution and sigmoid activation to generate channel attention weights. Element-wise multiplication is then applied to weight the feature maps according to their importance.

Mathematically, the ECA module can be expressed as:

where is the concatenated feature maps from the three DRN blocks, denotes adaptive average pooling, is a 1×1 convolution, represents the sigmoid function, and is the output feature maps with enhanced channel attention.

#### *3.1.4 UNET*

(a) (b)

Fig 1.3 (a) U-Net Architecture, (b) Conv Block Architecture

#### The feature maps from the ECA module are fed into a U-Net architecture for final reconstruction. Our U-Net implementation as in *Fig 1.3 (a),* consists of an encoder with four downsampling stages, a bottleneck at the lowest resolution, and a decoder with four upsampling stages. Skip connections from the encoder to the decoder help preserve high-frequency details that might otherwise be lost during downsampling. Both the encoder and decoder uses *Conv Block*, which consists of several layers as shown in *Fig 1.3 (b).* The final layer is a 1×1 convolution that produces the single-channel despeckled output image.

#### *3.1.5 Despeckled SAR Image*

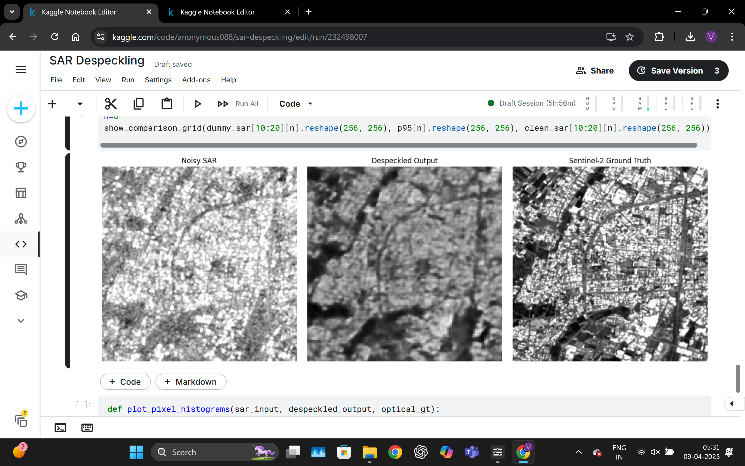
The final output from the UNET decoder is a despeckled SAR image with significantly reduced speckle noise while preserving structural details and edges. The restored image exhibits the following characteristics:

* Enhanced signal-to-noise ratio compared to the original SAR image
* Preserved edge information and structural details critical for interpretation
* Reduced granular texture characteristic of speckle noise
* Maintained radiometric properties of the original SAR data
* Natural appearance without over-smoothing or artificial artifacts

### *3.2 Loss Function*

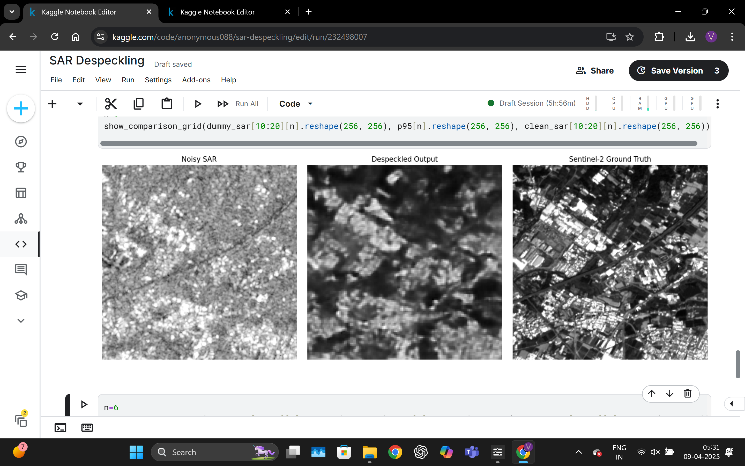
We employ a compound loss function consisting of two components:

1. ***Mean Squared Error (MSE) Loss***: To ensure pixel-level accuracy between the despeckled output and the clean reference image
2. ***Structural Similarity Index (SSIM)* *Loss:*** To preserve structural information

**The final loss function is a weighted combination:

where is a hyperparameter that balance the contribution of each loss component.

* 1. *Training Strategy*

**Our training strategy includes using the Adam optimizer with an initial learning rate of 1e-4 with cosine annealing scheduler. We use a batch size of 16 and train for 100 iterations. Data augmentation techniques include cropping and channel conversion. During training, we monitor the validation loss and employ early stopping with a patience of 4 epochs to prevent overfitting.

*3.4 Implementation Details*

The proposed method is implemented using PyTorch 1.10.0. All experiments are conducted on an NVIDIA P100 GPU with 216GB of memory. The training dataset consists of 3000 256×256 patches extracted from Sentinel-1 Dataset of SAR images. The validation set contains 1000 patches, and the test set includes real SAR images with different noise levels.

##### *IV. EXPERIMENTAL RESULTS*

*A. Dataset*

To evaluate the performance of the proposed CNN-based SAR despeckling method, real urban SAR images of Sentinel-1 and Sentinel-2 were utilized.

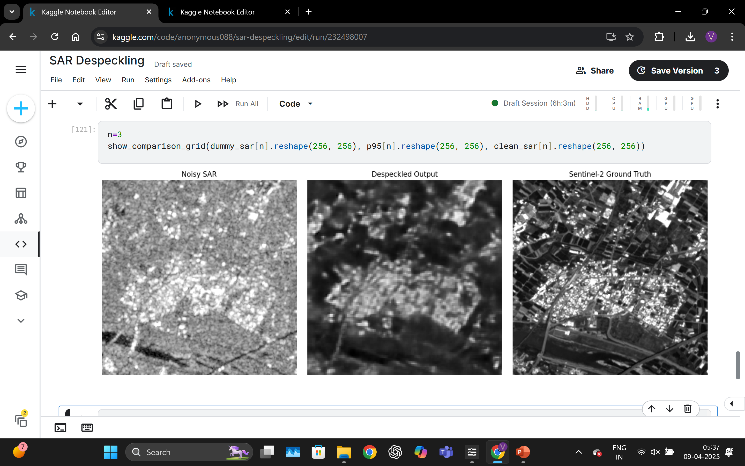
*Real SAR Data*

To demonstrate generalizability in practical scenarios, real SAR images acquired from the Sentinel-1 satellite were used:

* *Data Source:* Level-1 Ground Range Detected (GRD) products from Sentinel-1 in Interferometric Wide (IW) mode. Multispectral optical images from the Sentinel-2 satellite were used as the ground truth due to their high spatial and radiometric resolution, as well as the absence of speckle noise.
* *Geographical Diversity:* Multiple scenes covering diverse terrain types such as urban, forest, agricultural, and coastal regions were included.
* *Speckle Characteristics:* Since real SAR data inherently contains speckle, no additional noise simulation was performed. Visual and statistical assessments were used to evaluate the despeckling performance.

*B. Performance Metrics*

* Peak Signal-to-Noise Ratio (PSNR)
* Structural Similarity Index (SSIM)
* Noise Reduction Ratio (NRR)

*C. Visual Results*

*Fig 2 Despeckled SAR Images Produced by the Proposed Method*

*D. Comparative Analysis*

The proposed method was compared with:

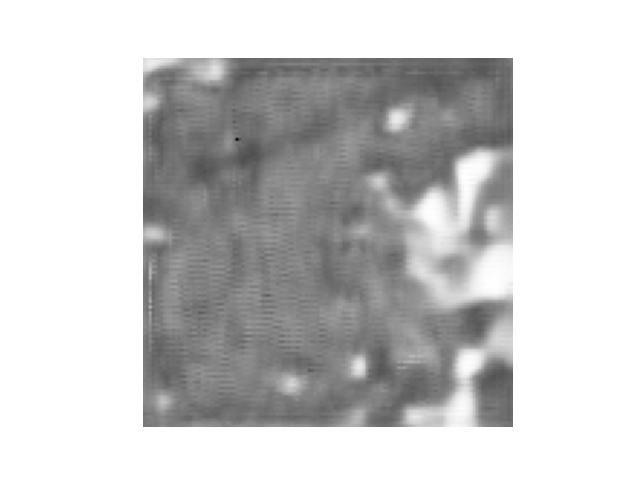
* Wavelet-based despeckling
* Existing deep learning approaches

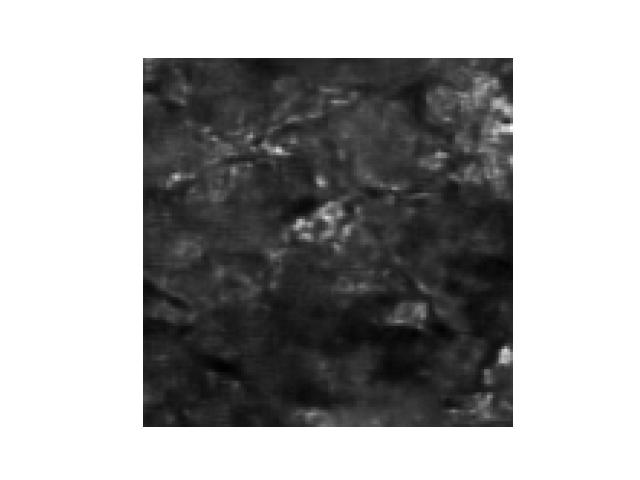
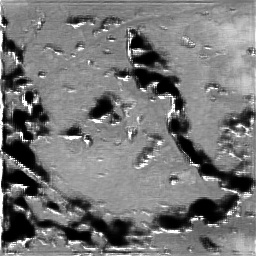
To evaluate the performance of the proposed despeckling method, we compare it against two state-of-the-art approaches: SAR-CNN and SAR2SAR. Quantitative metrics including Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index (SSIM), and Equivalent Number of Looks (ENL) were used for evaluation. As shown in *Table Fig 3.1*, the proposed method achieves a significant improvement in PSNR with a value of 17.62 dB, outperforming SAR2SAR (8.167 dB) and SAR-CNN, which failed to produce a valid reconstruction (−42.711 dB). Similarly, the proposed approach achieves an SSIM of 0.80, demonstrating high structural preservation, whereas SAR2SAR and SAR-CNN yield significantly lower SSIM values of 0.049 and 0.00005, respectively.

In terms of ENL, although SAR-CNN achieved the highest ENL value of 19.676, it did so at the expense of excessive oversmoothing, as evidenced by its poor PSNR and SSIM. In contrast, the proposed method strikes a better balance between speckle suppression and feature preservation, reflected in its competitive ENL of 3.31. These results suggest that the proposed model offers a more robust and visually consistent despeckling performance compared to existing techniques.

|  |  |  |  |
| --- | --- | --- | --- |
|  | **SAR-CNN** | **Sar2Sar** | Proposed |
| PSNR | -42.711 | 8.167 | **17.62** |
| **SSIM** | 0.00005 | 0.049 | **0.80** |
| **ENL** | 19.676 | 5.191 | **3.31** |

*Fig 3.1 Metrics Comparison Table*

***Visual Comparison of Despeckling Results of Different Models*

* (a) (b)*

*(c) (d)*

*Fig 3.2 Visual Comparision with different models (a) Expected Output, (b) SAR-CNN Ouput, (c) SAR2SAR Output, (d) Proposed Model Output*

To further validate the performance of the proposed despeckling model, a visual comparison was conducted using representative Sentinel-1 SAR patches with their corresponding Sentinel-2 RGB references. Fig. 3.2 displays example outputs from SAR2SAR, SAR-CNN, and the proposed method.

The SAR2SAR output retains some structural features but leaves behind noticeable residual speckle, especially in homogeneous regions such as rooftops and roads. SAR-CNN, on the other hand, over-smooths the input image, resulting in a loss of fine spatial details and the blurring of edges. This compromises the interpretability of urban features and boundaries.

In contrast, the proposed method offers a visually cleaner reconstruction with enhanced texture preservation and sharp object boundaries. Urban infrastructures such as buildings, roads, and open spaces are distinctly delineated. Moreover, the despeckled images closely resemble the Sentinel-2 ground truth, confirming the model’s ability to reduce speckle without sacrificing critical image features.

The improved clarity in the proposed model’s output can facilitate downstream applications such as object detection, segmentation, and urban change detection. These results reinforce the model's applicability in practical remote sensing tasks.

*V. RESULTS AND DISCUSSION*

Experimental results demonstrated:

* Significant improvement in image quality
* Superior noise reduction
* Better preservation of edge and texture details
* Consistent performance across different SAR image types

##### *VI. CONCLUSION*

##### This paper introduced a novel CNN-based ensemble architecture for SAR image despeckling that integrates multi-branch Dilated Residual Networks (DRNs) with a U-Net decoder and Efficient Channel Attention (ECA) mechanisms. The proposed MB-DRN-UNET framework effectively reduces multiplicative speckle noise while preserving fine spatial features, including edges and textures that are critical for accurate interpretation in remote sensing tasks.

##### Extensive experiments using Sentinel-1 SAR images and Sentinel-2 as reference ground truth demonstrated that our model outperforms traditional filtering methods and recent deep learning-based approaches such as SAR-CNN and SAR2SAR. Quantitative results revealed superior performance in terms of PSNR and SSIM, while qualitative visual comparisons confirmed enhanced image clarity and structural consistency.

##### The model's ability to generalize across different terrain types and real-world scenarios makes it a valuable tool for practical SAR image processing applications, such as urban monitoring, land cover classification, and disaster response. Future work may explore model generalization to other SAR modalities and unsupervised training strategies using real SAR-only datasets without relying on optical references.

##### *REFERENCES*

[1 Emanuele Dalsasso, Loïc Denis, Florence Tupin, " SAR2SAR: a semi-supervised despeckling algorithm for SAR images" *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing (Early Access), 2020, 13 Apr 2021*

[2] Emanuele Dalsasso, Xiangli Yang, Loïc Denis, Florence Tupin, Wen Yang, " SAR Image Despeckling by Deep Neural Networks: from a pre-trained model to an end-to-end training strategy" *Remote Sens. 2020, 12(16), 2636*, 21 Sep 2020.

[3] Puyang Wang, He Zhang, Vishal M. Patel, "SAR Image Despeckling Using a Convolutional Neural Network," *Neural Computation*. XX-XX, 2025.

[4] Adugna G. Mullissa, Diego Marcos, Devis Tuia, Martin Herold, Johannes Reiche, "deSpeckNet: Generalizing Deep Learning Based SAR Image Despeckling" Neural Computation. XX-XX, 2025.

[5] Alicia Passah, Khwairakpam Amitab, Debdatta Kandar, “SAR image despeckling using deep CNN" Neural Computation. 18 January 2021.

[6] Wenfu Wu, Xiao Huang , Zhenfeng Shao, Jiahua Teng,

Deren Li, "SAR-DRDNet: A SAR image despeckling network with detail recovery" Neural Computation, vol. 493. 7 July 2022.

[7] Cong Lin, Chenghao Qiu, Haoyu Jiang, Lilan Zou, “A Deep Neural Network Based on Prior-Driven and Structural Preserving for SAR Image Despeckling" IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 16. 05 July 2023.