UDC 004.8:528.8:621.396.96
DOI: 10.15587/2706-5448.2026.358316

Raed Al-Senaikh,
Oleksii Rubel

DEVELOPMENT OF A HYBRID METHOD VGG16-FrostNet FOR ADAPTIVE DESPECKLING OF SYNTHETIC APERTURE RADAR (SAR) IMAGES USING ATTENTION MECHANISM AND DIFFERENTIABLE FROST FILTER

The object of research is the process of suppressing multiplicative speckle noise in synthetic aperture radar (SAR) images, which significantly complicates their analysis. The problem addressed is the lack of end-to-end hybrid methods capable of spatial adaptation by integrating a mathematical model of local statistics (the Frost filter) directly into the neural network computation graph. This research is aimed at automating the process of adaptive SAR image filtering by developing the hybrid VGG16-FrostNet method. These research tasks were addressed by formulating a differentiable mathematical model of the classical Frost filter for integration into a neural network, developing an architecture based on a pretrained VGG16 (Visual Geometry Group) backbone (blocks 1–2), and integrating the Convolutional Block Attention Module (CBAM), which predicts a spatially varying damping coefficient map A_{map} within 0.5–10.0 for each pixel. The developed hybrid architecture includes a residual branch for detail recovery and was optimized end-to-end using a comprehensive loss function combining L1, Edge Loss (Sobel), SSIM, and attention regularization. The model was trained on synthetic data with gamma-distributed speckle (equivalent looks between 3.0 and 6.0) under typical SAR conditions. On the test set, experimental evaluation yielded a mean PSNR of 34.18 dB and SSIM of 0.97. The gain relative to the noisy image constituted 9.45 dB, and 3.36 dB in PSNR compared to the classical Frost filter with an optimal static coefficient. Edge indicators EPI = 0.8903 and FOM = 0.8340 substantiate reliable preservation of structural boundaries. It was established that the developed hybrid method provides spatially adaptive damping with interpretable attention maps, enabling its deployment in automated SAR data processing pipelines.

Keywords: *SAR images, speckle noise, suppression, Frost filter, VGG16, CBAM, deep learning.*

Received: 19.01.2026
Received in revised form: 07.04.2026
Accepted: 20.04.2026
Published: 30.04.2026

How to cite
Al-Senaikh, R., Rubel, O. (2026). Development of a hybrid method VGG16-FrostNet for adaptive despeckling of synthetic aperture radar (SAR) images using attention mechanism and differentiable Frost filter. Technology Audit and Production Reserves, 2 (2 (88)), 25–33. https://doi.org/10.15587/2706-5448.2026.358316

1. Introduction

Synthetic aperture radar (SAR) systems are a key tool for Earth observation because they can acquire images regardless of illumination and weather conditions. The fundamental principles and applications of SAR technology are comprehensively covered in [1]. However, SAR images inherently exhibit multiplicative speckle noise caused by the coherent nature of radar sensing and the interference of returns from multiple scatterers within a single resolution cell [2].

Speckle markedly degrades the perceived image quality and impairs automated interpretation, including classification, segmentation, and change detection tasks [3]. The impact of speckle on various image analysis tasks has been extensively studied, with recent comprehensive analyses provided in [4]. Suppressing speckle while preserving structural details is therefore still an open problem.

Classical adaptive despeckling filters, including the Frost filter [5], the Lee filter [6], the Kuan filter [7], and Gamma MAP [8], rely on local image statistics under a multiplicative noise model. In the Frost filter, the weights are computed as $w_{ij} = exp(-A \cdot C^2_I \cdot d_{ij})$, where A denotes the damping coefficient, $C_I = \sigma_I/\mu_I$ – the local coefficient of variation, and d_{ij} – the distance from the window center. The parameter A critically controls the trade-off between noise attenuation and detail preservation: larger values better preserve edges, whereas smaller values produce stronger smoothing in homogeneous areas [2]. The ongoing relevance of these local-statistic methods is evidenced by recent studies demonstrating substantial performance gains through improved classical algorithms, such as the modified Kuan filter for Sentinel-1 imagery [9].

A key limitation of these filters is the need for manual parameter tuning, with optimal settings strongly dependent on scene content and noise level. This requirement hinders their use in automated high-throughput processing pipelines typical of contemporary satellite missions such as Sentinel-1 [10]. To mitigate this issue, prior approaches have attempted to predict the expected filtering efficiency from statistical image descriptors. In [11, 12], this idea was explored for DCT-based filters, showing that local statistics can estimate potential quality improvements and that the relationship between noise characteristics and filtering outcomes is predictable. Later studies extended the methodology to SAR imagery and introduced neural-network-based predictors.

An initial framework for predicting filtering efficiency of DCT-based filters applied to SAR images using machine learning was proposed in [13]. This approach was further refined in [14] by incorporating more complex image features. More recently, the methodology was extended to adapt filter parameters to varying image content for Sentinel SAR imagery [15], and the approach was subsequently applied to optimal parameter selection for the Frost filter on Sentinel SAR images [16]. However, these methods primarily estimate relative quality gains rather than directly selecting spatially adaptive filtering parameters.

In recent years, deep learning has improved SAR image despeckling. Convolutional neural network (CNN) approaches such as SAR-CNN [17], DnCNN [18], and SAR2SAR [19] can deliver high-quality restoration. However, because of their purely data-driven nature, it can be challenging to interpret their intermediate representations or ensure adherence to classical statistical models. Comprehensive reviews of recent developments in deep learning for SAR despeckling are provided in [4]. Strategies for applying deep learning to SAR despeckling, including transfer from pre-trained natural-image denoisers and end-to-end training on SAR data, are analyzed in [20]. Therefore, instead of directly competing with these heavily parameterized architectures purely on quantitative metrics, the present research focuses on an alternative direction: enhancing a classical, theoretically grounded filter by making it differentiable and spatially adaptive. Consequently, the primary baseline for evaluation is the classical Frost filter with an optimally selected static coefficient, isolating the specific performance gain achieved by introducing spatial adaptation via the attention mechanism, rather than comparing against purely data-driven CNNs.

Transfer learning enables effective reuse of features learned from large-scale datasets such as ImageNet [21]. Architectures such as VGG16 (Visual Geometry Group) [22] provide hierarchical texture representations that transfer across natural-image tasks, while convolutional models developed for SAR despeckling [23] apply the same inductive bias to radar intensity imagery; SAR imaging fundamentals and speckle phenomenology are surveyed in [24]. The transferability of learned features across domains was systematically investigated in [25]. The theoretical foundations and practical applications of transfer learning in computer vision are comprehensively reviewed in [26].

Attention mechanisms, including the convolutional block attention module (CBAM) [27], SENet [28], and non-local networks [29], allow models to adaptively weight informative regions. This property is useful for despeckling, where object boundaries and homogeneous areas should receive different degrees of smoothing.

Based on the critical analysis of the current scientific literature, the following unresolved scientific and applied problem was identified: the lack of end-to-end hybrid methods capable of autonomously controlling spatial adaptation by integrating a differentiable analytical model of local statistics (the Frost filter) directly into the neural network computation graph. Consequently, *the object of research* is defined as the process of suppressing multiplicative speckle noise in synthetic aperture radar (SAR) images, and *the aim of this research* is to automate the process of adaptive SAR image filtering by developing the hybrid VGG16-FrostNet method.

To achieve this aim, the following research tasks were formulated.

1. To develop a differentiable mathematical model of the classical Frost filter to enable its integration into a neural network architecture.
2. To develop the architecture based on transfer learning (pretrained VGG16 model) combined with the CBAM and the differentiable Frost filter, and to formulate a comprehensive loss function for its end-to-end training.
3. To perform an experimental evaluation of the developed method and conduct a comparative analysis with the baseline filter using full-reference and edge-preservation metrics.

In practice, the proposed approach supports automated SAR data processing pipelines without manual tuning of filtering parameters.

2. Materials and Methods

2.1. Research methods

The following research methods were employed: mathematical modeling of multiplicative noise and the Frost filter for formulating a differentiable filtering framework; deep learning with convolutional neural networks and attention mechanisms for adaptive parameter prediction; transfer learning for reusing pretrained feature representations; and comparative statistical analysis using full-reference, perceptual, and edge-preservation image quality metrics for performance evaluation.

2.2. Source data

Experiments were conducted on synthetic data designed to emulate Sentinel-1 IW GRD (Interferometric Wide swath, Ground Range Detected) intensity products with VV polarization. The use of synthetic data is motivated by the need for reference images for supervised training and quantitative evaluation. To construct the synthetic dataset, histogram matching between Sentinel-2 optical images and reference Sentinel-1 SAR images was applied [30], enabling the generation of photo-realistic training samples with a known reference.

The dataset comprised 1528 single-channel 1024 × 1024 intensity images (uint16) for training and 382 images for testing. For each clean image, a paired sample (noisy, clean) was generated by applying the multiplicative noise model, defined below in equations (9), (10), with ENL randomly sampled from a uniform distribution over [3.0, 6.0] to cover variability across SAR systems and products. During training, data augmentation was applied by randomly extracting a 128 × 128 image crop from each 1024 × 1024 image, followed by random horizontal and vertical flips. Validation and final evaluation were performed on full 1024 × 1024 images.

2.3. Baselines and quality assessment metrics

To assess the proposed method, it was compared with the classical Frost filter using an optimally selected (static) damping coefficient. For each test image, the damping coefficient A was swept over [0.2, 2.6] with a step of 0.2, and the best value was selected by maximizing PSNR (Frost best-A baseline).

Filtering quality was evaluated using a set of complementary full-reference, perceptual, and edge-preservation metrics (1)–(8).

PSNR (peak signal-to-noise ratio) [31]:

$$\mathrm{PSNR}=20\log_{10}\left(\frac{\mathrm{MAX}_I}{\sqrt{\mathrm{MSE}}}\right),$$

$$\mathrm{MSE}=\frac{1}{HW}\sum_{i,j}\left(I_{filtered}(i,j)-I_{clean}(i,j)\right)^2, \tag{1}$$

where $MAX_I = 1.0$ for normalized images, and H and W denote image height and width, respectively.

SSIM (Structural Similarity Index) [31]

$$\mathrm{SSIM}(x,y)=\frac{\left(2\mu_x\mu_y+C_1\right)\left(2\sigma_{xy}+C_2\right)}{\left(\mu_x^2+\mu_y^2+C_1\right)\left(\sigma_x^2+\sigma_y^2+C_2\right)}, \tag{2}$$

where μ_x, μ_y – the mean intensities, σ_x^2, σ_y^2 – the variances, σ_{xy} – the covariance, and $C_1 = 0.01^2$, $C_2 = 0.03^2$ – the stabilization constants.

MS-SSIM (multi-scale structural similarity index) [32]

$$\text{MS-SSIM}(x,y)=\left[l_M(x,y)\right]^{\alpha_M}\prod_{j=1}^{M}\left[c_j(x,y)\right]^{\beta_j}\left[s_j(x,y)\right]^{\gamma_j}, \tag{3}$$

where $l_M(x, y)$ – the luminance comparison at the coarsest scale, $c_j(x, y)$ and $s_j(x, y)$ – the contrast and structure comparisons at scale j, and the standard five-scale weights are ($w_1 = 0.0448$, $w_2 = 0.2856$, $w_3 = 0.3001$, $w_4 = 0.2363$, $w_5 = 0.1333$). In the standard MS-SSIM setting, the luminance term is used only at the coarsest scale.

FSIM (feature similarity index measure) [33]

$$\mathrm{FSIM}=\frac{\sum_{p\in\Omega}S_{PC}\left(p\right)S_{G}\left(p\right)PC_{m}\left(p\right)}{\sum_{p\in\Omega}PC_{m}\left(p\right)}, \tag{4}$$

where $S_{PC}(p)$ – the phase-congruency similarity, $S_G(p)$ – the gradient-magnitude similarity, $PC_m(p) = \max(PC_1(p), PC_2(p))$ – the phase-congruency weight, and $\boldsymbol{\Omega}$ – the set of all image pixels.

HaarPSI (Haar wavelet-based Perceptual Similarity Index) [34]

$$\mathrm{HaarPSI}_{f_1,f_2}=l_{\alpha}^{-1}\left(\frac{\sum_{x}\sum_{k=1}^{2}HS_{f_1,f_2}^{(k)}\left[x\right]\cdot W_{f_1,f_2}^{(k)}\left[x\right]}{\sum_{x}\sum_{k=1}^{2}W_{f_1,f_2}^{(k)}\left[x\right]}\right)^{2}, \tag{5}$$

where $HS^{(k)}{}_{f1,f2}[x]$ denotes the local similarity map obtained from Haar-wavelet responses, $W^{(k)}{}_{f1,f2}[x]$ – the corresponding weight map, $k \in \{1, 2\}$ selects horizontal and vertical features, and l_α – the logistic mapping used in the original definition.

MDSI (mean deviation similarity index) [35]

$$\mathrm{MDSI}=\left(\frac{1}{\left|\Omega\right|}\sum_{p\in\Omega}\left|D\left(p\right)-\mu_{D}\right|^{\rho}\right)^{1/\rho}, \tag{6}$$

where $D(p)$ – the combined local distortion map, μ_D – its mean value over $\boldsymbol{\Omega}$, and $\boldsymbol{\rho}$ – the pooling exponent. Lower MDSI values correspond to better image quality.

EPI (Edge Preservation Index) [36]

$$\mathrm{EPI}=\frac{\sum_{i}\left(E_{f}\left(i\right)-\mu_{E_{f}}\right)\left(E_{c}\left(i\right)-\mu_{E_{c}}\right)}{\sqrt{\sum_{i}(E_{f}\left(i\right)-\mu_{E_{f}})^{2}\sum_{i}(E_{c}\left(i\right)-\mu_{E_{c}})^{2}}}, \tag{7}$$

where E_f, E_c – edge maps extracted using a 3 × 3 Sobel operator from the filtered and reference images, respectively, and μ_{Ef}, μ_{Ec} – the mean values of the corresponding edge maps.

FOM (Pratt's Figure of Merit) [37]

$$\mathrm{FOM}=\frac{1}{\max(N_{detected},N_{reference})}\sum_{i=1}^{N_{detected}}\frac{1}{1+\alpha d_{i}^{2}}, \tag{8}$$

where $N_{detected}$, $N_{reference}$ – the numbers of detected and reference edge pixels, d_i – the distance from the i-th detected edge pixel to the nearest reference edge, and $\alpha = 1/9$ – a weighting parameter.

3. Results and Discussion

3.1. Mathematical modeling and problem statement

Speckle noise in SAR intensity images is commonly modeled as multiplicative noise, following the local-statistics formulation in [6] and the radar imaging model and adaptive filter in [5]:

$$I\left(x,y\right)=X\left(x,y\right)\cdot S\left(x,y\right), \tag{9}$$

$$S\sim\Gamma\left(L,1/L\right), \tag{10}$$

where $X(x, y)$ denotes the underlying scene intensity, $I(x, y)$ – the observed intensity, $S(x,y)$ – the speckle multiplier with a gamma distribution, and L – the equivalent number of looks (ENL). The selected parameterization ensures $E[S] = 1$ and $Var[S] = 1/L$, corresponding to the standard multiplicative speckle model for SAR imagery, which has been validated across various SAR systems and imaging conditions [2].

Based on the multiplicative noise model (9), (10), the formal problem statement is formulated as follows: given a noisy observation $I(x, y)$, find a nonlinear mapping F_θ parameterized by weights θ that produces an estimate $\hat{X}(x, y) = F_\theta(I(x, y))$ approximating the true scene intensity $X(x, y)$. The optimal parameters θ^* are determined by minimizing the expected composite loss

$$\theta^{*}=\mathrm{argmin}_{\theta}\mathbb{E}\left[\mathcal{L}\left(F_{\theta}\left(I\right),X\right)\right]. \tag{11}$$

A central component of the mapping F_θ is the integration of classical local-statistics logic into the neural computation graph. To achieve this, a differentiable mathematical model of the Frost filter was developed. The classical Frost filter weight equation employs the squared coefficient of variation C^2_I in the exponent. Direct integration of this formulation into a gradient-based optimization pipeline presents a fundamental difficulty: in heterogeneous regions and near object boundaries, C_I assumes values greater than unity, and consequently C^2_I exceeds C_I by a large margin, causing the exponent to become large in absolute value. In finite-precision (float32) arithmetic, the exponential function underflows to zero for sufficiently negative arguments, rendering the gradients of the filter weights with respect to the damping coefficient numerically negligible and effectively blocking gradient flow through the filter layer during backpropagation.

To overcome this limitation, the mathematical model was modified by replacing C^2_I with its first-power counterpart $C_I = \sigma_I/\mu_I$. Since for $C_I > 1$ the inequality $C_I < C^2_I$ holds, this substitution compresses the dynamic range of the local-statistics term in the exponent, keeping the exponential function and its derivatives numerically representable. This modification does not reduce the filter expressiveness, because the spatially varying damping coefficient A_{map} – predicted by the neural network for each pixel – implicitly compensates for the reduced magnitude of C_I relative to C^2_I. The expanded range of A_{map} (detailed in Section 3.2) provides sufficient capacity for this adjustment. The resulting differentiable filter weights are computed using a 7 × 7 window as

$$w_{ij}=\exp\left(-A_{map}\left(x,y\right)\cdot C_{I}\left(x,y\right)\cdot d_{ij}\right), \tag{12}$$

where $C_I = \sigma_I/\mu_I$ – the local coefficient of variation estimated from local mean and variance, and d_{ij} – the Euclidean distance from the window center. The filtered intensity is computed as

$$I_{frost}\left(x,y\right)=\frac{\sum_{i,j}w_{ij}\cdot I\left(x+i,y+j\right)}{\sum_{i,j}w_{ij}}. \tag{13}$$

An auxiliary sub-task is to predict the spatially varying damping map $A_{map}(x, y)$ that directly controls this differentiable mathematical model, ensuring that smoothing adapts to local image content. The proposed solution to these formulated problems is the hybrid VGG16-FrostNet method developed in Section 3.2, which optimizes the mapping F_θ end-to-end according to the objective (11).

3.2. The VGG16-FrostNet architecture

The development of the proposed architecture requires constructing a specialized neural network architecture capable of integrating the mathematical model (12), (13). The key design principle consists in predicting the spatially varying damping map (16) to act as an internal parameter for continuous tensor operations, making the hybrid mechanism trainable end-to-end. The overall architecture of the developed method is shown in Fig. 1.

The architecture consists of five main components.

VGG16 backbone: The first two blocks of a pretrained VGG16 network (layers 1–9 from the feature extractor) are used to extract low-level texture features. To accommodate single-channel SAR images, the input convolution is modified from (3, 64, 3 × 3) to (1, 64, 3 × 3) by averaging RGB-channel weights: $W_{gray} = (1/3)\Sigma W_i$. The backbone is trained jointly with the remaining components (freeze_backbone = False) to adapt ImageNet-pretrained features to the SAR domain.

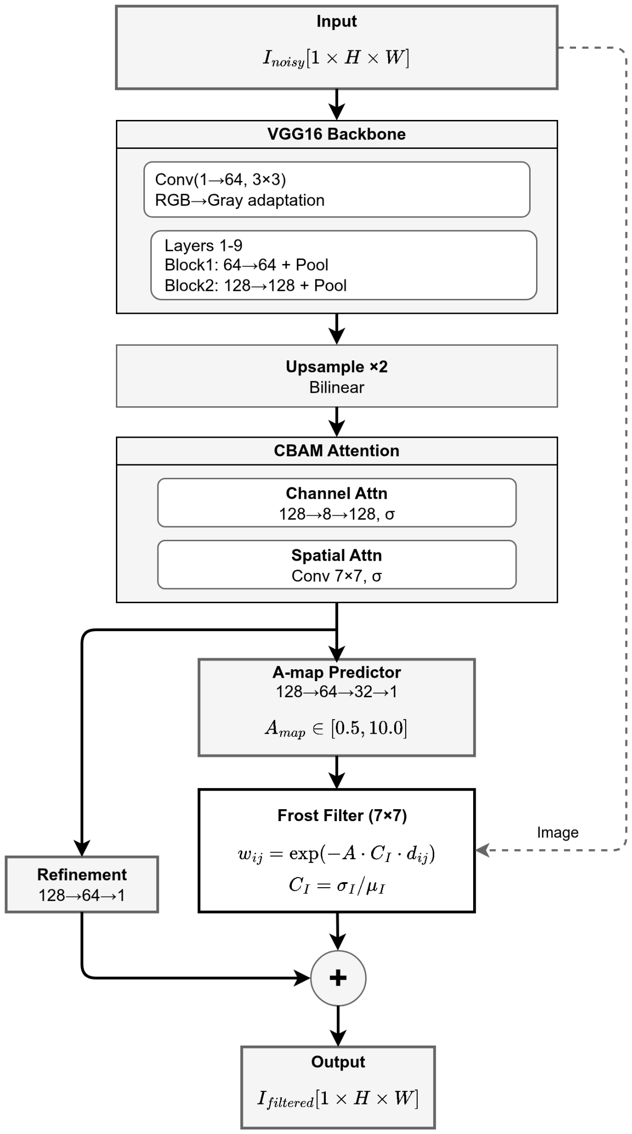


Fig. 1. VGG16-FrostNet architecture: a noisy input image is processed by a pretrained VGG16 backbone, followed by the CBAM attention module, an A_map coefficient predictor, a differentiable Frost filter, and a residual refinement branch

Rationale for selecting VGG16: the VGG16 architecture [22] employs small 3 ×3 convolution kernels, enabling effective extraction of low-level texture features (e. g., edges and micro-textures) that are critical for separating boundaries from homogeneous regions in SAR images. The early VGG16 blocks are known to learn transferable representations across domains [25].

CBAM [27]: Channel attention and spatial attention are applied sequentially to modulate backbone features, as formalized in:

$$M_c(F) = \sigma\Big(\mathrm{MLP}\big(\mathrm{AvgPool}(F)\big) + \mathrm{MLP}\big(\mathrm{MaxPool}(F)\big)\Big), \quad (14)$$

$$M_s(F) = \sigma\Big(\mathrm{Conv}_{7\times7}\big(\big[\mathrm{AvgPool}(F);\mathrm{MaxPool}(F)\big]\big)\Big), \quad (15)$$

where F denotes the input feature map, M_c – the channel-attention map computed using an MLP (reduction = 16), and M_s – the spatial-attention map obtained via concatenation of channel-wise average and max pooling followed by a 7 × 7 convolution. Modulation is applied sequentially: channel attention yields $F_c = F \otimes M_c(F)$, then spatial attention is applied to obtain $F' = F_c \otimes (1 + M_s(F_c))$, which can be written compactly as $F' = F \otimes M_c(F) \otimes (1 + M_s(F \otimes M_c(F)))$.

Coefficient-map predictor: A sequence of convolutional layers with channel widths 128 –> 64 –> 32 –> 1 produces an adaptive damping coefficient map A_{map} constrained to the range [0.5, 10.0]

$$A_{map} = 0.5 + 9.5 \cdot \sigma\left(A_{raw}\right), \quad (16)$$

where A_{raw} – the raw network output and σ denotes the sigmoid function. The expanded range (compared with the typical A from 1 to 3 used in the classical Frost filter) enables adaptation to different region types: higher values (5.0 to 10.0) near boundaries for edge preservation and lower values (0.5 to 2.0) in homogeneous areas for strong noise suppression. This predicted tensor is then directly passed to the differentiable Frost filter formulated in (12), (13).

Residual refinement branch: A parallel branch with channel widths 128 –> 64 –> 1 restores high-frequency details that may be weakened by filtering. The final output (17) is computed as

$$I_{filtered} = I_{frost} + I_{refinement}. \quad (17)$$

In this way, the residual path compensates for detail loss, while the main Frost branch continues to suppress noise in homogeneous regions.

Training relies on a composite loss function:

$$\mathcal{L}_{total} = \mathcal{L}_{L1} + \lambda_e \mathcal{L}_{edge} + \lambda_s \mathcal{L}_{ssim} + \lambda_a \mathcal{L}_{attn} + \lambda_r \mathcal{L}_{refine}, \quad (18)$$

$$\mathcal{L}_{L1} = \frac{1}{HW}\sum_{i,j}\left|I_{filtered}(i,j) - I_{clean}(i,j)\right|, \quad (19)$$

$$\mathcal{L}_{edge} = \frac{1}{HW}\sum_{i,j}\left|E_{pred}(i,j) - E_{target}(i,j)\right|, \quad (20)$$

where E – the edge map, $E = \sqrt{(g_x^2 + g_y^2 + \varepsilon)}$, and $g_x = G_x \cdot I$, $g_y = G_y \cdot I$ – the horizontal and vertical gradients obtained with 3 × 3 Sobel filters

$$\mathcal{L}_{ssim} = 1 - \overline{\mathrm{SSIM}}, \quad (21)$$

where $SSIM$ – the structural similarity index defined in (2)

$$\mathcal{L}_{attn} = \frac{1}{HW}\sum_{i,j}\left(A_{map}(i,j) - E_{norm}(i,j)\right)^2, \quad (22)$$

where E_{norm} – the min-max normalized edge map of the reference image, $E_{norm} = (E_{target} - \min(E_{target}))/(\max(E_{target}) - \min(E_{target}) + \varepsilon)$, with min and max taken over all pixels

$$\mathcal{L}_{refine} = \frac{1}{HW}\sum_{i,j}\left|I_{refinement}(i,j)\right|. \quad (23)$$

The objective contains five terms. L1 loss (L_{L1}) measures the main pixelwise difference between the prediction and the reference; see (19). Edge loss (L_{edge}, $\lambda_e = 0.1$) compares 3 × 3 Sobel gradients g_x, g_y of the prediction and the reference via L1, as in (20) [38]. SSIM loss (L_{ssim}, $\lambda_s = 0.1$) is 1 – $SSIM$ averaged over a Gaussian window (size 11, $\sigma = 1.5$), (21) [31]. Attention regularization (L_{attn}, $\lambda_a = 0.05$) is the MSE between the spatial-attention map and min-max normalized edges of the reference, (22). Refinement regularization (L_{refine}, $\lambda_r = 0.01$) is an L1 sparsity penalty on the refinement output, (23).

These coefficients were chosen from validation experiments. In the final setting, the edge and SSIM terms use $\lambda_e = \lambda_s = 0.1$, attention regularization uses $\lambda_a = 0.05$, and refinement regularization uses $\lambda_r = 0.01$. L1 remains the dominant part of the objective, while the additional terms introduce boundary, structural, and attention-related constraints. These coefficients, as well as the overall architectural design, were chosen based on preliminary validation experiments and established best practices in image restoration [4]. A comprehensive ablation study to isolate the exact marginal contribution of each loss term and structural component (e. g., CBAM versus other attention modules) falls outside the scope of this proof-of-concept study, which aims primarily to demonstrate the viability and interpretability of the end-to-end differentiable Frost filter.

The hybrid VGG16-FrostNet method is defined as an integrated end-to-end trainable processing pipeline that unifies three components developed above: the differentiable mathematical model of the Frost filter (12), (13), the hybrid architecture combining the pretrained VGG16 backbone with the CBAM, and the composite loss function (18)–(23). The method prescribes a specific algorithmic procedure:

a) input normalization and multiplicative noise modeling according to (9), (10);

b) a forward pass through the architecture, during which the CBAM predicts the spatially varying damping map A_{map} (16) that directly controls the differentiable Frost filter;

c) computation of the composite loss;

d) backpropagation through all components – including the differentiable filter layer – with simultaneous update of all trainable parameters.

The key challenge in developing this method was ensuring compatibility between two fundamentally different paradigms: classical statistical filtering (Frost filter) and neural network gradient-based optimization. This was achieved through the differentiable reformulation of the filter (described above) and through linking the output of the CBAM attention module to the input damping parameter A_{map} of the filter. This connection transforms the filtering parameter from an externally tuned constant into an internal state variable of the network, predicted autonomously during each forward pass. The attention regularization loss (22) further enforces physically meaningful behavior of A_{map}: higher damping values near structural boundaries (for edge preservation) and lower values in homogeneous regions (for strong noise suppression).

The model was trained using a transfer-learning strategy based on the framework established in [26].

The effectiveness of fine-tuning pretrained features for domain adaptation was demonstrated in [25], where the pretrained VGG16 backbone was fine-tuned together with the remaining modules for SAR imagery. This setup reused low-level representations learned on ImageNet [21] and adapted them to the SAR despeckling task.

Training was carried out with the Adam optimizer and a learning rate of 0.0005. The batch size was 8 image crops of size 128 × 128 pixels, randomly extracted from the original 1024 × 1024 images. The model was trained for 60 epochs, and the differentiable Frost filter used a 7 × 7 window. The pretrained VGG16 backbone (blocks 1–2), CBAM attention modules, the coefficient predictor, and the refinement branch were optimized jointly. All experiments were performed on an NVIDIA RTX 4090 GPU.

3.3. Experimental evaluation and comparison with baselines

Fig. 2 shows training curves for VGG16-FrostNet, illustrating the evolution of the loss function and quality metrics during training.

As shown in Fig. 2, the model converged steadily over 60 training epochs. The training loss is presented in normalized form only to make visual comparison with the validation curve easier. Because the two curves remained close throughout training, there was no clear sign of overfitting. On the validation set, PSNR and SSIM rose gradually and reached about 34.2 dB and 0.97, respectively. Taken together, these trajectories showed that the ImageNet-initialized VGG16 backbone could be adapted to SAR imagery without optimization instability.

Fig. 3 shows the distribution of adaptive coefficients A_{map} over the validation set.

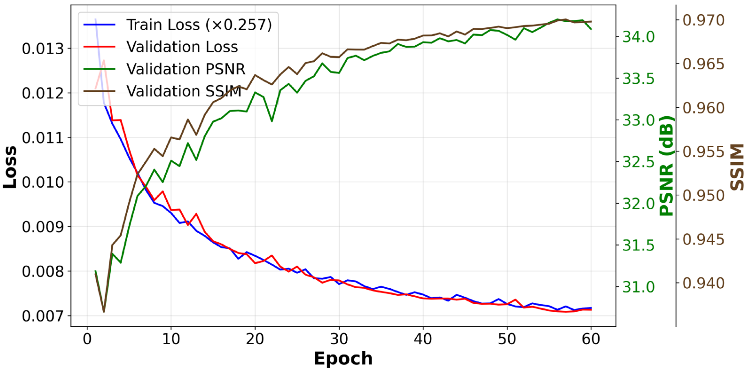


Fig. 2. Training curves for VGG16-FrostNet: loss on the training and validation sets (left axis) and PSNR/SSIM on the validation set (right axes)

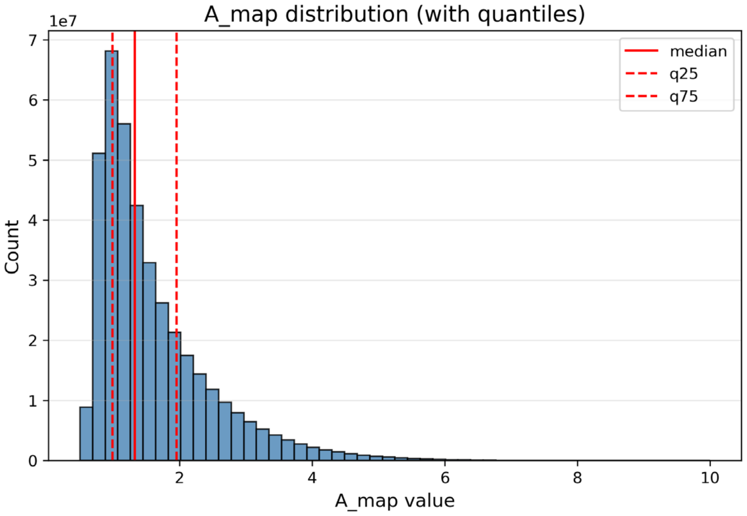


Fig. 3. Distribution of adaptive coefficients A_map over the validation set. Vertical lines indicate the median (solid) and quartiles Q25 and Q75 (dashed)

The histogram of A_{map} values over the validation set of 382 images (Fig. 3) exhibited a skewed distribution with a median of 1.32 and a mean of 1.60. The quartiles were Q25 = 0.99 and Q75 = 1.95, indicating the predominance of low-to-moderate damping coefficients. The observed range spanned 0.50 to 9.27, consistent with the imposed constraint A_{map} in [0.5, 10.0]. The concentration of values in the low-coefficient region (0.5 to 2.0) aligned with weaker smoothing in structurally salient areas and stronger smoothing in homogeneous regions.

Quantitative results on the test set of 382 images (1024 × 1024 pixels) are summarized in Table 1.

Table 1

Comparison of filtering quality across methods

Method	Noisy image	Frost (best-A)	VGG16-FrostNet (proposed)
PSNR (dB)	24.73 ± 1.45	30.82 ± 1.44	34.18 ± 1.90
SSIM	0.8014 ± 0.0574	0.9361 ± 0.0249	0.9700 ± 0.0152
MS-SSIM	0.9385 ± 0.0209	0.9720 ± 0.0106	0.9871 ± 0.0064
FSIM	0.9352 ± 0.0204	0.9529 ± 0.0122	0.9782 ± 0.0060
HaarPSI	0.4888 ± 0.0648	0.6279 ± 0.0392	0.7870 ± 0.0436
MDSI	0.3067 ± 0.0212	0.2867 ± 0.0181	0.2401 ± 0.0168
EPI	0.5236 ± 0.1026	0.7078 ± 0.0605	0.8903 ± 0.0338
FOM	0.7137 ± 0.1393	0.4380 ± 0.0951	0.8340 ± 0.0457

VGG16-FrostNet achieved a mean PSNR of 34.18 ± 1.90 dB and SSIM of 0.9700 ± 0.0152 on the 382-image test set (values are reported as mean ± standard deviation). Relative to the classical Frost filter with an optimally selected static damping coefficient (Frost best-A), the proposed method improved PSNR by 3.36 dB and SSIM by 0.0340. Relative to the noisy input, the PSNR gain was 9.45 dB. VGG16-FrostNet also outperformed the baseline on extended metrics: MS-SSIM 0.9871 ± 0.0064 versus 0.9720 ± 0.0106, FSIM 0.9782 ± 0.0060 versus 0.9529 ± 0.0122, and HaarPSI 0.7870 ± 0.0436 versus 0.6279 ± 0.0392. The MDSI score (lower is better) was 0.2401 ± 0.0168 for the proposed method versus 0.2867 ± 0.0181 for Frost best-A, indicating improved structural preservation. Boundary-oriented metrics further highlighted the advantage of spatial adaptation: EPI was 0.8903 ± 0.0338 versus 0.7078 ± 0.0605 (increase of 0.1825), and FOM was 0.8340 ± 0.0457 versus 0.4380 ± 0.0951 (increase of 0.3960). These improvements were attributed to per-pixel, spatially varying selection of damping coefficients, which eliminated per-image parameter sweeps and supported higher-throughput processing.

To analyze the dependence of filtering performance on the initial noise level, Fig. 4 presents scatter plots relating the metrics of noisy images to those of filtered images for four measures: PSNR, SSIM, EPI, and HaarPSI. The diagonal line $y = x$ corresponds to no improvement; points above the diagonal indicate metric gains after filtering. Linear regression revealed a strong positive correlation between noisy and filtered metrics, supporting the stability of the proposed method across noise levels.

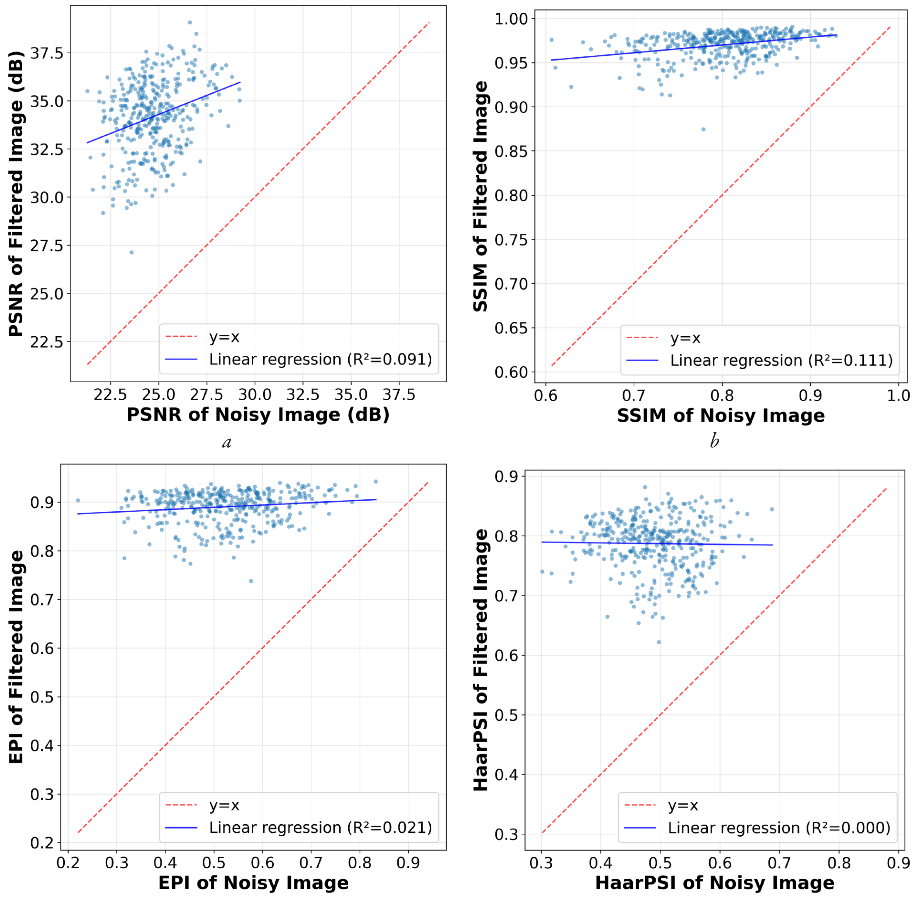
To assess robustness, filtering performance was analyzed as a function of noise level characterized by the equivalent number of looks (ENL). The 382-image test set was partitioned into three ENL ranges: [3.0, 4.0), [4.0, 5.0), and [5.0, 6.0], corresponding to different speckle levels in SAR imagery. For each range, the mean metric values of filtered images and the gain relative to the corresponding noisy images are reported in Table 2.

Each table entry reports the mean metric value for filtered images within the corresponding ENL range, along with the change relative to the mean metric value of the noisy images (shown in parentheses). Positive changes for PSNR, SSIM, MS-SSIM, FSIM, HaarPSI, EPI, and FOM indicate improvement after filtering, with larger values corresponding to stronger improvement. For MDSI (where lower is better), a negative change (i. e., a decrease) likewise indicates improvement.

The robustness analysis demonstrated stable performance across the full ENL range [3.0, 6.0]. As ENL decreased (stronger noise), the PSNR gain increased from 8.93 dB for [5.0, 6.0] to 10.10 dB for [3.0, 4.0), indicating effective adaptation to varying noise levels. Similar trends were observed for SSIM, MS-SSIM, FSIM, HaarPSI, EPI, and FOM, where larger gains occurred under heavier speckle. MDSI consistently decreased across all ENL bins, with the largest improvements under stronger noise. Overall, all metrics remained high across the tested ENL range, supporting the robustness of the proposed approach.

Fig. 5 presents a visual comparison of despeckling results and associated diagnostic maps for a representative scene.

Visual inspection indicates that VGG16-FrostNet suppressed speckle in homogeneous areas while preserving edges and fine structural details. The spatial attention map highlights object boundaries and structurally salient regions, supporting adaptive filtering behavior. The adaptive coefficient map A_{map} reflects spatially varying damping in the differentiable Frost filter, exhibiting higher values near boundaries and lower values in homogeneous regions.



a *b*

c *d*

Fig. 4. Relationship between filtered-image and noisy-image metrics:
a – PSNR; *b* – SSIM; *c* – EPI; *d* – HaarPSI. The dashed diagonal $y = x$ indicates no improvement

Table 2

Robustness analysis versus noise level (ENL)

Metric	ENL [3.0, 4.0]	ENL [4.0, 5.0]	ENL [5.0, 6.0]
PSNR (dB)	33.78 (+10.10)	34.21 (+9.40)	34.52 (+8.93)
SSIM	0.9668 (+0.2061)	0.9702 (+0.1641)	0.9727 (+0.1396)
MS-SSIM	0.9853 (+0.0616)	0.9873 (+0.0468)	0.9887 (+0.0387)
FSIM	0.9748 (+0.0536)	0.9786 (+0.0418)	0.9809 (+0.0348)
HaarPSI	0.7717 (+0.3241)	0.7881 (+0.2975)	0.7995 (+0.2759)
MDSI	0.2483 (–0.0717)	0.2393 (–0.0667)	0.2335 (–0.0619)
EPI	0.8827 (+0.4096)	0.8919 (+0.3595)	0.8957 (+0.3354)
FOM	0.8308 (+0.1380)	0.8361 (+0.1218)	0.8348 (+0.1030)

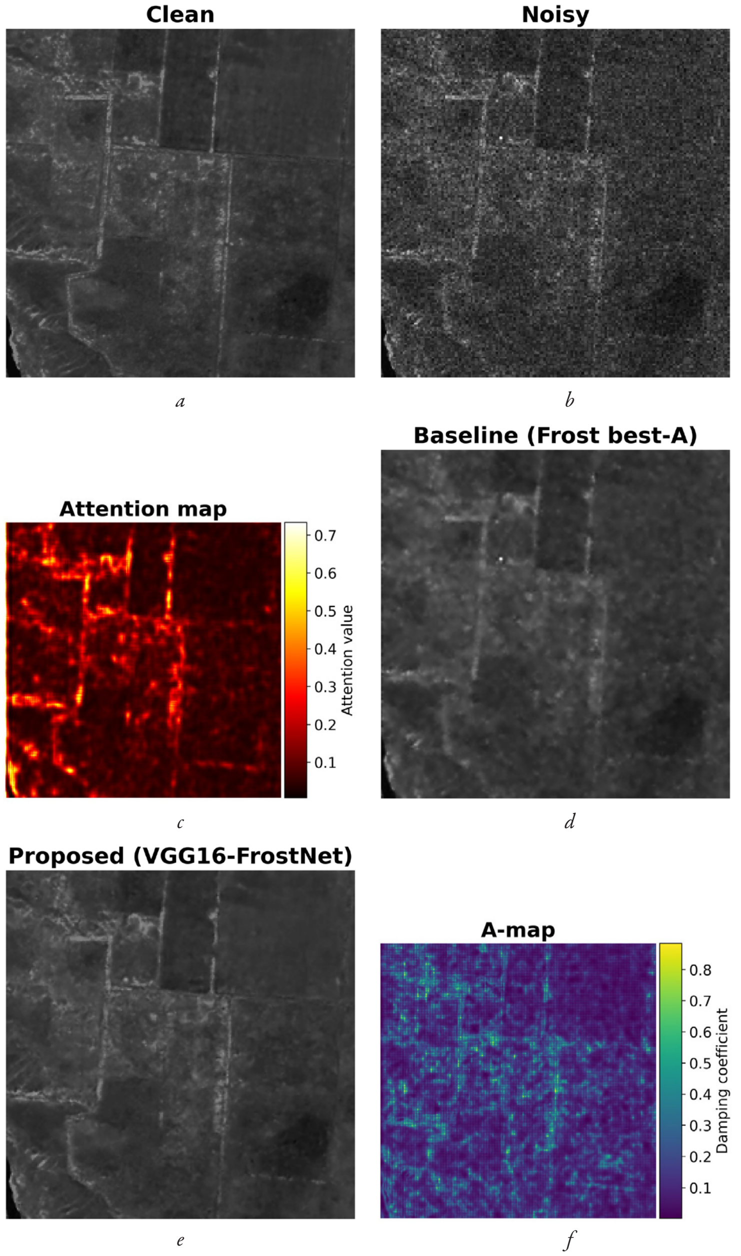


Fig. 5. Visual comparison and diagnostic maps:
a – clean image; *b* – noisy image; *c* – attention map; *d* – baseline (Frost best-A);
e – proposed (VGG16-FrostNet); *f* – A-map

The processing time per 1024 × 1024 image was approximately 40.66 ms on an NVIDIA RTX 4090 GPU.

3.4. Discussion

The quantitative performance of VGG16-FrostNet, including a 3.36 dB gain in PSNR over the classical Frost filter with an optimal static coefficient, is explained by the successful integration of a fully differentiable mathematical model of the Frost filter with the CBAM attention mechanism. Unlike the traditional approach, which applies a single fixed damping factor, per-pixel prediction of damping coefficients ensures simultaneous weak smoothing near edges and strong noise suppression in homogeneous regions. This mechanism directly translates theoretical local-statistics modeling into an end-to-end optimized neural network computation graph.

Compared to existing deep-learning approaches that rely entirely on data-driven feature extraction (such as standard CNNs or U-Nets), VGG16-FrostNet actively retains the analytical interpretability of classical models. While previous studies focused on predicting expected filtering efficiency or mapping optimal parameters for separate filters using external meta-data, the proposed method embeds the filtering equation directly into the network. This allows the model to surpass classical baseline methods by eliminating manual tuning while circumventing the "black-box" nature of purely data-driven architectures.

The principal distinctive feature and scientific novelty of this research is the analytical fusion of transfer learning (VGG16) and spatial-channel attention (CBAM) with the exact mathematical formulation of the Frost filter. Unlike prior meta-learning techniques, CBAM dynamically predicts the explicit damping coefficient map A_{map} as an internal state tensor variable. This specific feature directly solves the stated research problem: it automates spatial adaptation autonomously during a forward pass and enables the model to be optimized end-to-end via a composite loss function, delivering a highly controllable, edge-preserving filtering response.

In practice, VGG16-FrostNet is recommended for deployment in automated SAR data processing pipelines where manual filter tuning per scene is unfeasible. The interpretable spatial attention maps can be utilized by practitioners to automatically inspect geometric distortions or verify object boundaries directly during processing. The largest practical gains are observed when processing large-scale imagery containing both extensive agricultural or water surfaces (requiring strong smoothing) and detailed urban infrastructure (requiring boundary preservation).

The current research faces several objective limitations. Since the primary focus is the theoretical formulation of the differentiable Frost filter, training and full-reference evaluations were isolated to synthetic data with gamma-distributed speckle (ENL 3.0–6.0). While this ensures perfect methodological reproducibility, validation on real SAR data presents complex geometric challenges (layover, shadow, and foreshortening) and requires no-reference evaluation, which is outside the scope of this initial theoretical validation. Moreover, this research was constrained to VV polarization using raw digital numbers rather than radiometrically calibrated backscatter products. For practitioners utilizing these results in their own operating conditions, it is crucial to account for these nuances and apply additional fine-tuning if dealing with multi-polarimetric or severely distorted data.

Future research will focus on validating the architecture using real Sentinel-1 imagery where no-reference metrics are required.

Promising extensions include adapting the methodology to multi-polarimetric SAR data (VH/VV combinations), incorporating the structural filtering block into multi-scale U-Net paradigms, and implementing self-supervised training procedures. Adaptation of the inference algorithms for edge-computing environments also remains a highly relevant engineering task.

4. Conclusions

1. The classical local-statistics equations of the Frost filter were reformulated into continuous, backpropagatable 2D convolution tensor operations (as detailed in Section 3.2). This specific mathematical transformation permitted the traditional static damping parameter to be substituted with a dynamic, per-pixel tensor variable that scales local spatial smoothing gradients during gradient descent. By establishing this hybrid computational mechanism – comprising the damping map prediction (14)–(16) and the differentiable Frost filter (12), (13) – a differentiable Frost layer was successfully developed and proven viable as a direct analytical processing component within an end-to-end optimized neural network architecture.

2. A targeted hybrid neural network architecture was designed by exploiting the pretrained VGG16 backbone (blocks 1–2) for robust feature extraction and fusing it with the CBAM spatial-channel attention module. The CBAM module functioned effectively as an autonomous spatial controller, successfully predicting the spatially varying damping coefficient map A_{map} within the physical boundaries of 0.5 to 10.0. Interpretability analysis confirmed that this hybrid design mathematically governs the adaptive filtering behavior: the model applies strong edge preservation (A_{map} up to 10.0) along structurally salient object boundaries, while enforcing strong smoothing (A_{map} close to 0.5–2.0) across widespread homogeneous regions.

3. A comparative experimental evaluation on test datasets simulating SAR conditions (equivalent number of looks 3.0–6.0) validated the superiority of the hybrid approach over traditional baselines. The developed VGG16-FrostNet method achieved a mean PSNR of 34.18 dB and SSIM of 0.97, securing a significant 3.36 dB gain in PSNR compared exclusively to the classical Frost filter utilizing an optimal static coefficient. Physical edge-preservation indicators, specifically EPI (0.8903) and FOM (0.8340), quantitatively established that the proposed architecture suppresses multiplicative speckle while reliably preserving high-frequency image structure. Theoretically, these findings verify that classical analytical filters can be embedded into modern deep-learning computation graphs without sacrificing physical interpretability. Practically, the VGG16-FrostNet method is highly beneficial for deployment in automated SAR data pipelines within Earth remote sensing systems, where it eliminates manual parameter tuning and ensures reliable, spatially adaptive image analysis.

Acknowledgments

Sentinel-1 data were obtained from the Copernicus open-access data platform (dataspace.copernicus.eu). The computations used resources from the National Aerospace University "Kharkiv Aviation Institute".

Conflict of interest

The authors declare that they have no conflict of interest regarding this research, including financial, personal, authorship, or other, that could influence the research and its results presented in this article.

Financing

The research was performed without financial support.

Data availability

Data will be made available on reasonable request.

Use of artificial intelligence

The authors confirm that no artificial intelligence technologies were used in the preparation of this manuscript.

Authors' contributions

Raed Al-Senaikh: Conceptualization, Methodology, Software, Formal analysis, Investigation, Data curation, Writing – original draft, Visualization; ***Oleksii Rubel***: Conceptualization, Methodology, Supervision, Writing – review and editing, Project administration.

References

1. Oliver, C., Quegan, S. (2004). *Understanding synthetic aperture radar images.* SciTech Publishing, 479. Available at: https://books.google.com/books?id=0Ev3IoGbVSIC
2. Argenti, F., Lapini, A., Bianchi, T., Alparone, L. (2013). A Tutorial on Speckle Reduction in Synthetic Aperture Radar Images. *IEEE Geoscience and Remote Sensing Magazine,* 1 (*3*), 6–35. https://doi.org/10.1109/mgrs.2013.2277512
3. Touzi, R. (2002). A review of speckle filtering in the context of estimation theory. *IEEE Transactions on Geoscience and Remote Sensing, 40 (11),* 2392–2404. https://doi.org/10.1109/tgrs.2002.803727
4. Fracastoro, G., Magli, E., Poggi, G., Scarpa, G., Valsesia, D., Verdoliva, L. (2021). Deep Learning Methods For Synthetic Aperture Radar Image Despeckling: An Overview Of Trends And Perspectives. *IEEE Geoscience and Remote Sensing Magazine, 9 (2),* 29–51. https://doi.org/10.1109/mgrs.2021.3070956
5. Frost, V. S., Stiles, J. A., Shanmugan, K. S., Holtzman, J. C. (1982). A Model for Radar Images and Its Application to Adaptive Digital Filtering of Multiplicative Noise. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 4 (2),* 157–166. https://doi.org/10.1109/tpami.1982.4767223
6. Lee, J.-S. (1980). Digital Image Enhancement and Noise Filtering by Use of Local Statistics. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 2 (2),* 165–168. https://doi.org/10.1109/tpami.1980.4766994
7. Kuan, D. T., Sawchuk, A. A., Strand, T. C., Chavel, P. (1985). Adaptive Noise Smoothing Filter for Images with Signal-Dependent Noise. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 7 (2),* 165–177. https://doi.org/10.1109/tpami.1985.4767641
8. Lopes, A., Touzi, R., Nezry, E. (1990). Adaptive speckle filters and scene heterogeneity. *IEEE Transactions on Geoscience and Remote Sensing, 28 (6),* 992–1000. https://doi.org/10.1109/36.62623
9. Khudov, H., Makoveichuk, O., Tokarev, S., Andriushchenko, A., Pukhovyi, O., Rohulia, O. et al. (2026). Improving a method for filtering images acquired from a space-based radar observation system based on the Kuan algorithm. *Eastern-European Journal of Enterprise Technologies, 1 (9 (139)),* 40–46. https://doi.org/10.15587/1729-4061.2026.352347
10. Filipponi, F. (2019). Sentinel-1 GRD Preprocessing Workflow. *3rd International Electronic Conference on Remote Sensing,* 11. https://doi.org/10.3390/ecrs-3-06201
11. Abramov, S., Krivenko, S., Roenko, A., Lukin, V., Djurovic, I., Chobanu, M. (2013). Prediction of filtering efficiency for DCT-based image denoising. *2013 2nd Mediterranean Conference on Embedded Computing* (*MECO*). Budva: IEEE, 97–100. https://doi.org/10.1109/meco.2013.6601327
12. Lukin, V. V., Abramov, S. K., Rubel, A., Krivenko, S. S., Naumenko, A., Vozel, B. et al. (2014). An approach to prediction of signal-dependent noise removal efficiency by dct-based filter. *Telecommunications and Radio Engineering, 73 (18),* 1645–1659. https://doi.org/10.1615/telecomradeng.v73.i18.40
13. Rubel, O. S., Lukin, V. V., De Medeiros, F. S. (2015). Prediction of Despeckling Efficiency of DCT-Based Filters Applied to SAR Images. *2015 International Conference on Distributed Computing in Sensor Systems.* Fortaleza: IEEE, 159–168. https://doi.org/10.1109/dcoss.2015.16
14. Rubel, O., Lukin, V., Rubel, A., Egiazarian, K. (2019). NN-Based Prediction of Sentinel-1 SAR Image Filtering Efficiency. *Geosciences, 9* (*7*), 290. https://doi.org/10.3390/geosciences9070290
15. Rubel, O., Lukin, V., Rubel, A., Egiazarian, K. (2021). Selection of Lee Filter Window Size Based on Despeckling Efficiency Prediction for Sentinel SAR Images. *Remote Sensing, 13 (10),* 1887. https://doi.org/10.3390/rs13101887
16. Rubel, O. S., Rubel, A. S., Lukin, V., Egiazarian, K. (2022). Optimal parameters selection of the Frost filter based on despeckling efficiency prediction for Sentinel SAR images. *Electronic Imaging, 34 (10),* 193-1-193–196. https://doi.org/10.2352/ei.2022.34.10.ipas-193

17. Chierchia, G., Cozzolino, D., Poggi, G., Verdoliva, L. (2017). SAR image despeckling through convolutional neural networks. *2017 IEEE International Geoscience and Remote Sensing Symposium* (*IGARSS*). Fort Worth: IEEE, 5438–5441. https://doi.org/10.1109/igarss.2017.8128234

18. Zhang, K., Zuo, W., Chen, Y., Meng, D., Zhang, L. (2017). Beyond a Gaussian Denoiser: Residual Learning of Deep CNN for Image Denoising. *IEEE Transactions on Image Processing, 26 (7),* 3142–3155. https://doi.org/10.1109/tip.2017.2662206

19. Dalsasso, E., Denis, L., Tupin, F. (2021). SAR2SAR: A Semi-Supervised Despeckling Algorithm for SAR Images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 14,* 4321–4329. https://doi.org/10.1109/jstars.2021.3071864

20. Dalsasso, E., Yang, X., Denis, L., Tupin, F., Yang, W. (2020). SAR Image Despeckling by Deep Neural Networks: from a Pre-Trained Model to an End-to-End Training Strategy. *Remote Sensing, 12 (16),* 2636. https://doi.org/10.3390/rs12162636

21. Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S. et al. (2015). ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision, 115 (3),* 211–252. https://doi.org/10.1007/s11263-015-0816-y

22. Simonyan, K., Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. *Proceedings of the International Conference on Learning Representations* (*ICLR*). arXiv. Available at: https://arxiv.org/abs/1409.1556

23. Vitale, S., Ferraioli, G., Pascazio, V. (2021). Multi-Objective CNN-Based Algorithm for SAR Despeckling. *IEEE Transactions on Geoscience and Remote Sensing, 59 (11),* 9336–9349. https://doi.org/10.1109/tgrs.2020.3034852

24. Moreira, A., Prats-Iraola, P., Younis, M., Krieger, G., Hajnsek, I., Papathanassiou, K. P. (2013). A tutorial on synthetic aperture radar. *IEEE Geoscience and Remote Sensing Magazine, 1 (1),* 6–43. https://doi.org/10.1109/mgrs.2013.2248301

25. Yosinski, J., Clune, J., Bengio, Y., Lipson, H. (2014). How transferable are features in deep neural networks? *Advances in Neural Information Processing Systems* (*NeurIPS*), 27. Available at: https://papers.nips.cc/paper/5347-how-transferable-are-features-in-deep-neural-networks

26. Pan, S. J., Yang, Q. (2010). A Survey on Transfer Learning. *IEEE Transactions on Knowledge and Data Engineering, 22 (10),* 1345–1359. https://doi.org/10.1109/tkde.2009.191

27. Woo, S., Park, J., Lee, J.-Y., Kweon, I. S. (2018). CBAM: Convolutional Block Attention Module. *Computer Vision – ECCV 2018,* 3–19. https://doi.org/10.1007/978-3-030-01234-2_1

28. Hu, J., Shen, L., Sun, G. (2018). Squeeze-and-Excitation Networks. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition.* Salt Lake City: IEEE, 7132–7141. https://doi.org/10.1109/cvpr.2018.00745

29. Wang, X., Girshick, R., Gupta, A., He, K. (2018). Non-local Neural Networks. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition.* Salt Lake City: IEEE, 7794–7803. https://doi.org/10.1109/cvpr.2018.00813

30. Al-Senaikh, R., Rubel, O. (2025). Predicting Filtered Image Quality Using Transfer Learning on Sentinel-1 Speckle Noise with DenseNet-121. *Ukrainian Journal of Remote Sensing, 12 (4),* 4–15. https://doi.org/10.36023/ujrs.2025.12.4.293

31. Wang, Z., Bovik, A. C., Sheikh, H. R., Simoncelli, E. P. (2004). Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing, 13 (4),* 600–612. https://doi.org/10.1109/tip.2003.819861

32. Wang, Z., Simoncelli, E. P., Bovik, A. C. (2003). Multiscale structural similarity for image quality assessment. *The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers, 2003.* Pacific Grove: IEEE, 1398–1402. https://doi.org/10.1109/acssc.2003.1292216

33. Zhang, L., Zhang, L., Mou, X., Zhang, D. (2011). FSIM: A Feature Similarity Index for Image Quality Assessment. *IEEE Transactions on Image Processing, 20 (8),* 2378–2386. https://doi.org/10.1109/tip.2011.2109730

34. Reisenhofer, R., Bosse, S., Kutyniok, G., Wiegand, T. (2018). A Haar wavelet-based perceptual similarity index for image quality assessment. *Signal Processing: Image Communication, 61,* 33–43. https://doi.org/10.1016/j.image.2017.11.001

35. Nafchi, H. Z., Shahkolaei, A., Hedjam, R., Cheriet, M. (2016). Mean Deviation Similarity Index: Efficient and Reliable Full-Reference Image Quality Evaluator. *IEEE Access, 4,* 5579–5590. https://doi.org/10.1109/access.2016.2604042

36. Sattar, F., Floreby, L., Salomonsson, G., Lovstrom, B. (1997). Image enhancement based on a nonlinear multiscale method. *IEEE Transactions on Image Processing, 6* (*6*), 888–895. https://doi.org/10.1109/83.585239

37. Abdou, I. E., Pratt, W. K. (1979). Quantitative design and evaluation of enhancement/thresholding edge detectors. *Proceedings of the IEEE, 67* (*5*), 753–763. https://doi.org/10.1109/proc.1979.11325

38. Mathieu, M., Couprie, C., LeCun, Y. (2016). Deep multi-scale video prediction beyond mean square error. *Proceedings of the International Conference on Learning Representations* (*ICLR*). arXiv. https://doi.org/10.48550/arXiv.1511.05440

✉***Raed Al-Senaikh,*** *PhD Student, Department of Information and Communication Technologies named after O. O. Zelensky, National Aerospace University "Kharkiv Aviation Institute", Kharkiv, Ukraine, e-mail: r.z.alsenaikh@khai.edu, ORCID: https://orcid.org/0000-0002-8059-4237*

Oleksii Rubel, *PhD, Associate Professor, Department of Information and Communication Technologies named after O. O. Zelensky, National Aerospace University "Kharkiv Aviation Institute", Kharkiv, Ukraine, ORCID: https://orcid.org/0000-0001-6206-3988*

✉*Corresponding author*