ANED-Net: Adaptive Noise Estimation and Despeckling Network for SAR Image

Xu Wang, Yanxia Wu, Changting Shi, Ye Yuan, *Member, IEEE*, and Xue Zhang

***Abstract*—Synthetic aperture radar (SAR) images are often affected by a type of multiplicative noise known as "speckle" due to their active imaging characteristics. This property complicates the processing and interpretation of SAR images. While deep learning techniques have demonstrated success in despeckling, many models are tailored to specific noise levels. This specificity can limit a model's ability to generalize to real SAR images with varying noise levels, potentially leading to oversmoothing or overfocusing on specific details. To address these challenges, we present the Adaptive Noise Estimation and Despeckling Network (ANED-Net). This network consists of a noise-level estimation phase and a noise-level-guided nonblind denoising phase. During the nonblind denoising phase, we develop a noise-feature-guided denoising network. This network integrates a hierarchical encoder–decoder denoising module based on the Transformer block (T-unet) and a denoising enhancement control block. Together, they skillfully capture both local and global dependencies inherent in SAR images, facilitating effective noise removal. Furthermore, we also introduce a deep-attention mechanism to counteract the attentional collapse observed when the Transformer is extended in depth, enhancing the network's feature extraction capability and strengthening the model's denoising performance. Extensive tests on synthetic and real images show that ANED-Net is robust to different noise scenarios. It effectively mitigates speckle noise even at unspecified levels and outperforms many established methods.**

***Index Terms*—Deep learning, noise-level estimation, speckle noise suppression, synthetic aperture radar (SAR).**

I. Introduction

SYNTHETIC aperture radar (SAR) is an all-weather, all-day, high-resolution imaging sensor. Its unique imaging characteristics have led to its widespread use in both military and civilian applications, including military mapping, ocean monitoring, resource exploration, and oil spill detection [1], [2]. However, the coherent imaging properties of SAR introduce a type of noise known as "coherent speckle" during the imaging process. This noise severely degrades the interpretability of the images and poses significant challenges for downstream SAR imaging tasks, such as target detection and classification. Therefore, effective suppression of coherent speckle noise in SAR images is critical for their subsequent applications.

Manuscript received 2 November 2023; revised 26 December 2023; accepted 11 January 2024. Date of publication 17 January 2024; date of current version 5 February 2024. This work was supported by the Natural Science Foundation of Heilongjiang Province under Grant JJ2019LH2160. *(Corresponding author: Changting Shi.)*

The authors are with the College of Computer Science and Technology, Harbin Engineering University, Harbin 150001, China (e-mail: 2750250486@hrbeu.edu.cn; wuyanxia@hrbeu.edu.cn; shichangting@hrbeu.edu.cn; yuanye@hrbeu.edu.cn; crystal_zhang@hrbeu.edu.cn).

Digital Object Identifier 10.1109/JSTARS.2024.3355220

Since the advent of SAR, the effective mitigation of coherent speckle noise in its images has remained a topic of intense interest. Over the decades, numerous algorithms have been developed to address this challenge. Postimaging coherent speckle suppression methods for SAR images are particularly widespread and can generally be divided into two categories: 1) traditional denoising methods and 2) SAR image denoising algorithms based on deep learning.

Conventional methods for coherent speckle noise suppression can be broadly classified into three categories: 1) spatial filtering, 2) transform domain filtering, and 3) nonlocal mean filtering. Spatial domain filtering techniques, such as Lee filtering [3], extended Lee filtering [4], Kuan filtering [5], Frost filtering [6], and Gamma-MAP [7], were pioneers in SAR image denoising. They use a sliding window based on a statistical model and exploit the correlation between image pixels for noise reduction. However, because the spatial filtering denoising algorithm uses a fixed window, the spatial filtering algorithm is less adaptive and often leads to blurring of image edges and texture information during denoising because the texture-rich regions do not satisfy the statistical model.

To address the limitations of spatial filtering, researchers have introduced transform-domain filtering. Typically, this approach first transforms the image from the spatial domain to the transform domain using a specific transformation. After adjusting the coefficients in the transform domain, the denoised image is obtained through an inverse transformation. Techniques under this umbrella include methods based on the wavelet transform [8], ridgelet transform [9], curvelet transform [10], contourlet transform [11], and shearlet transform [12]. However, they often require complex shearlet transform, which can lead to excessive image smoothing, blurring of details, and degradation of image quality.

In contrast, nonlocal mean filtering exploits nonlocal similarities within the image for denoising. It identifies blocks of pixels that are similar to the current block and computes a weighted average of these blocks to reduce noise. Notable advances include algorithms, such as SAR-BM3D [13] and PPB [14]. It should be noted, however, that such methods are sensitive to the size of the search window and pixel block, and the denoising performance of NLM is degraded in the presence of severe noise.

In recent years, deep learning methods have garnered widespread attention in the field of image processing, especially convolutional neural networks (CNNs), which have shown significant success in SAR image denoising due to their powerful nonlinear feature extraction abilities. Chierchia et al. [15]

pioneered the combination of residual learning with a batch normalization strategy and introduced SAR-CNN, which outperformed traditional denoising methods. However, this method requires processing on homomorphically transformed images, which prevents end-to-end learning. To address this limitation, researchers introduced end-to-end denoising solutions, including SAR-DRN [16], SID-CNN [17], Wavelet-SRNet [18], Monet [19], and G-MONet [20], with SAR-DRN employing dilated convolutions to widen the receptive field while maintaining a lightweight structure. Monet [19] introduced a CNN with a multiobjective cost function, addressing the spatial and statistical characteristics of SAR images. G-MONet [20] utilizes a multicategory Gaussian simulator to generate training data and a multiobjective cost function, enhancing despeckling performance across various scenarios. Simultaneously, generative adversarial networks (GANs) were applied to SAR image denoising. Innovations by Wang et al. [21] with ID-GAN and Liu et al. [22] with a GAN-based speckle noise suppression network both enhanced denoising capabilities while preserving image details. Researchers have also incorporated advanced technologies such as encoder–decoder networks, dense connections, and attention networks to better extract SAR image features. Drawing inspiration from the U-Net architecture, Lattari et al. [23] developed a deep codec network using multiple skip connections to preserve image details. Gui et al. [24] tackled the vanishing gradient issue in deep CNNs with the SAR-DDCN method while Zhang et al. [25] introduced MCN-WF, using a deep architecture for more expressive feature extraction. HDRANet by Li et al. [26] improved detail preservation through spatial and channel attention modules, and Malsha et al. [27] utilized self-attention (SA) mechanisms to learn global dependencies for enhanced denoising. In addition, some researchers started to try unsupervised learning strategies to solve the difficult problem of not being able to acquire speckle-free noisy SAR images. For example, Molini et al. [28] proposed Speckle2Void, a blind CNN approach. Dalsasso et al. [29] created SAR2SAR, a network based on Noise2Noise. MERLIN [30] separates the real and imaginary parts of single-look complex SAR images as paired noisy images for Noise2Noise training to achieve significant noise reduction.

The method of using synthetic noisy images may adversely affect the despeckle results. However, the advantage of this approach is that it can easily generate speckle-noise-simulated images with clear references and can control the entire training process. In view of this, in order to maximize the effect of noise suppression, this study chooses to obtain robust despeckling results for the model with a reliable training dataset using a supervised learning approach.

Although many deep-learning-based methods have been proposed to remove speckle noise, most of the supervised denoising algorithms tend to focus on specific noise levels, and their generalization ability heavily depends on having access to extensive training data. However, there is a significant difference between the SAR noise in real scenes and the synthetic SAR noise at a specific distribution level, which leads to the unsatisfactory performance of conventional denoising methods when dealing with SAR images with unknown noise levels. To be specific, denoisers with high noise levels may cause the image to lose edge and detail information [31] while denoisers with low noise levels may leave a large amount of speckle residue after processing [31]. This implies that the use of direct denoising methods has a weak generalization ability when dealing with real SAR image noise with complex noise-level distributions. To address this issue, the study [32] used a uniform region detector to estimate the number of looks of SAR images and a pretrained FFDNet [33] model to achieve suppression of images with different noise levels. However, in SAR images, the value of the actual number of looks, L varies from pixel to pixel depending on the roughness, scale, and randomness of the scattering in each pixel. Therefore, the algorithm still causes the model to ignore the texture details of the SAR image when using a uniform number of looks.

In order to further improve the performance of the deep learning model and strike a balance between optimizing the denoising effect and preserving the image details, it becomes crucial to introduce the noise-level map, also known as the equivalent number of looks (ENL) estimation map, as the model input. We hope that the model can perform better when the noise-level map matches the real noise level of the input. Some attempts have been made for ENL estimation algorithms [34], [35], [36], but the small-sample sliding window approach based on the homogeneity assumption leads to difficulties in obtaining accurate ENL estimation results in the presence of increased image heterogeneity. With the powerful and robust feature extraction and learning capability of CNN, this article addresses the earlier problem by constructing a CNN-based noise-level estimation network. Based on this, this article proposes an adaptive noise-level despeckling framework. The framework consists of two main stages: First, noise-level estimation is performed, i.e., the ENL map of the SAR image is estimated, and then, a targeted despeckling operation is performed based on this estimation. Throughout the denoising process, the model is guided by the noise-level map so that it can more accurately analyze and utilize the noise data, thus improving the image quality. This method performs well in dealing with uncertain noise levels in SAR images and truly achieves adaptive noise removal.

Although CNNs are widely used in image restoration tasks, there are still some problems. First, using a fixed convolutional kernel may not produce optimal results in different regions of the image, because the convolutional kernel is independent of the image content [37]. Second, convolutional kernels can only capture local information and may lose important global information, especially when long-range dependencies need to be modeled. To solve these problems, methods such as adaptive convolution [38], nonlocal convolution [39], and global average pooling [40] have been proposed, but these methods still have limitations. In contrast, the SA mechanism [41] can comprehensively consider the information of all positions to capture the global features, which shows more efficient and flexible advantages. Therefore, in this article, we choose the Transformer [42] with an SA mechanism as the main network structure and integrate it into the UNet architecture to construct a network called noise-feature-guided denoising network

(NFGDN) for realizing denoising tasks in the noise-level-guided nonblind despeckling phase.

In addition, in this article, a denoising enhancement controller is designed in the NFGDN network to more accurately preserve the detail information of the original image during the denoising process. By controlling the denoising process more carefully, the fine structure of the original image can be preserved.

In the Transformer architecture, the SA mechanism plays a central role by allowing the model to focus on key regions of the input by assigning different weights to different parts of the input data. However, as the number of Transformer blocks increases, the performance of the model does not necessarily improve, but can actually degrade. This phenomenon is called "attention collapse." To address this problem, Zhou et al. [43] proposed a method called reattention, which dynamically aggregates the attention graphs of different heads by learning the transformation matrix θ to generate a new attention graph. However, the previous methods do not fundamentally change the structure of the attention mechanism or introduce more contextual information, and their change granularity is still not fine enough. Therefore, we propose a deep attention mechanism to solve this problem and achieve a better denoising effect.

Compared with the previous SAR image despeckling methods, the primary innovations presented in this article can be outlined as follows.

1) We propose the Adaptive Noise Estimation and Despeckling Network (ANED-Net) for SAR image despeckling. This network integrates noise-level estimation with nonblind denoising. During the despeckling phase, ANED-Net deeply exploits and utilizes the noise information guided under the guidance of the noise-level map to effectively counteract the varying levels of speckle noise present in SAR images.
2) We design the NFGDN within the nonblind denoising phase. It includes a hierarchical encoder–decoder denoising module anchored on the Transformer block and a denoising enhancement control (DEC) block. The overall goal of this network is to refine multiscale, local, and global representation learning. By meticulously analyzing and recognizing local and global dependencies in SAR images, it provides superior denoising results and further enhances image detail clarity.
3) We propose the implementation of deep-attention, aiming to address the issue of attentional collapse as the depth of the Transformer network is expanded, which ensures the efficient deep stacking of the network, thus strengthening the denoising effectiveness of the model.

The rest of the article is organized as follows. Section II describes the SAR image speckle noise model. Section III introduces the proposed algorithm. Section IV analyzes the experimental results. Finally, Section V concludes this article.

II. Related Work

A. Speckle Model

When radar waves hit a rough surface, inconsistencies in the phase of the received signal occur. This is due to variations in the distance between the primary scatterer and the sensor. As a result, the coherent processing of successive radar pulses leads to granular intensity variations on a pixel-by-pixel basis in the SAR image produced. This phenomenon is called coherent speckle noise. The formation mechanism of this noise is very different from the typical noise in digital image processing. The model for coherent speckle noise is as follows:

$$Y = X \cdot N \tag{1}$$

where Y represents the SAR image with coherent speckle noise. X symbolizes the ideal target feature measurement affected by the noise. N represents the coherent speckle noise generated by the SAR system during imaging. In this model, X and Y are independent. Typically, the multiplicative random noise N follows a gamma distribution with a mean of 1 and a variance of $1/L$ [44]. Therefore, the probability density function of the noise can be expressed as

$$P(N) = \frac{1}{\Gamma(L)} L^L N^{L-1} e^{-L \cdot N}, N \geq 0, L \geq 1 \tag{2}$$

where L is the ENL and $\Gamma(\cdot)$ is the gamma function.

The primary goal of SAR image denoising is to remove the coherent speckle noise N. The goal is to reconstruct the desired noiseless image X from the noisy image Y. This reconstruction is a mapping from Y to X. In particular, the intensity of the coherent speckle noise is closely related to the number of looks in the SAR image. By adjusting the number of looks, we can emulate speckle noise at different noise levels. This adjustment provides rich data for network training, which subsequently increases the robustness of the network.

B. Transformer and Attention Collapse

The Transformer [42], initially used for machine translation, fully replaces recurrent and convolutional approaches with the SA mechanism [45], [46]. The SA mechanism shows more efficiency and flexibility compared to recursion and convolution and captures global features by comprehensively considering information from all locations. This mechanism is a core part of the Transformer model, which achieves parallel processing and efficient feature learning by using multihead SA. It can be applied to a variety of visual tasks, such as image recognition [47], [48], [49], image segmentation [50], [51], and target detection [52], [53] and achieve excellent performance.

However, unlike CNNs, where performance can be improved by stacking more convolutional layers, the performance of the model does not necessarily improve as the number of Transformer blocks increases, but rather quickly saturates and even degrades. This phenomenon is called attentional collapse. The main reasons for this can be attributed to the following [43].

1) As the depth of the model increases, the similarity of the attention maps generated by each Transformer block for feature aggregation increases. This phenomenon weakens the ability of the SA mechanism to generate attentional diversity, thus limiting its ability to capture rich representations of the input data.

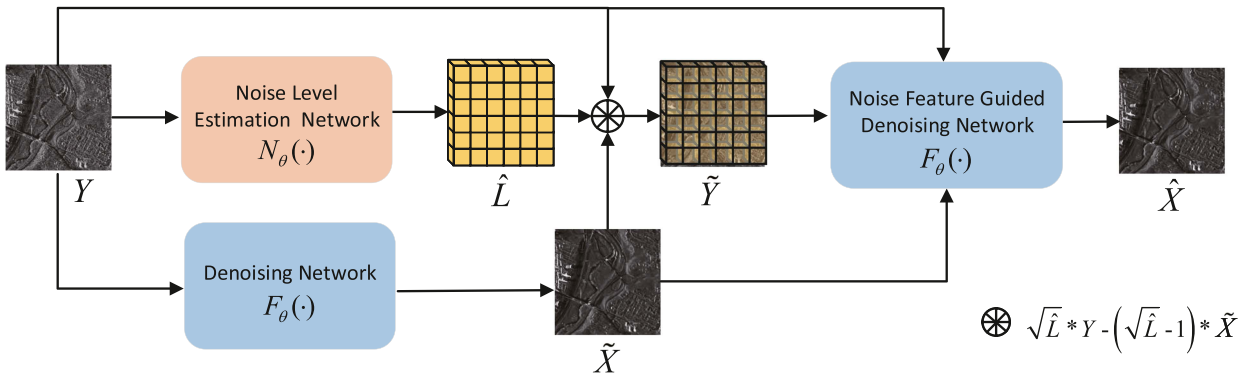

Fig. 1. Framework of the proposed ANED-Net.

2) Due to the emergence of such highly similar attention graphs, the Transformer block can degenerate into a multilayer perceptron.

This degeneration can trigger a phenomenon of model rank degradation, where the rank of the model parameter tensor is reduced by multiplying the layers. This process potentially limits the learning ability of the model and weakens its performance in complex tasks. Zhou et al. [43] proposed a method called reattention, which dynamically aggregates attention maps from different heads to generate a new attention map by learning a transformation matrix θ. The transformation matrix θ is a learnable parameter matrix that can be updated according to the backpropagation algorithm. This recomputation can increase the diversity of attention maps and solve the similarity problem of attention maps. However, this approach only solves the similarity problem of attention graphs by increasing the diversity of attention mappings. Without fundamentally changing the structure of the attention mechanism or introducing more contextual information, the granularity of the change is still not fine enough. To address this problem, we propose a new design for the SA mechanism that can extend the Transformer network to a deeper level and improve its performance.

III. Proposed Method

This section initially outlines the adaptive noise-level despeckling strategy, followed by a detailed introduction to the overall architecture of ANED-Net. Next, the two-stage network incorporated in ANED-Net is described. It then explains how to solve the attention breakdown problem in the Transformer with a deep-attention mechanism and how to preserve image detail with a denoising enhancement controller.

A. Adaptive Noise Estimation and Despeckling Network

In recent years of research, denoising algorithms such as MRDDANet [54] and SAR-CAM [55] usually use a one-time end-to-end approach to directly eliminate noise with a certain distribution level. However, in real SAR images, the noise level is closely related to the value of the number of looks L while the actual value L of each pixel is affected by the scattering characteristics, so the number of looks varies from pixel to pixel. This complex noise distribution characteristic makes the traditional denoising methods ineffective in processing real SAR images. To address this problem, we propose a new adaptive noise-level despeckling strategy that aims to more accurately remove SAR images with complex noise-level distributions. The network framework is illustrated in Fig. 1.

The strategy first estimates the noise-level map L in the noisy image Y. Then, the estimated noise-level map is fed into the nonblind denoising subnetwork together with the noisy image after a special processing, so that the denoising problem for SAR images becomes a denoising problem that can be targeted to solve the denoising problem with a known noise-level distribution. In short, the strategy is to find a noise-level map L that helps the denoising model to produce a visually clean denoised image $\hat{X}$.

Specifically, to derive the noise-level map, we use the prediction function $\hat{L} = P(Y)$, where the noisy image Y is used directly to predict the noise-level map $\hat{L}$. The parameters of the noise-level predictor are optimized by minimizing the l_1 distance

$$\theta_P = \arg\min_{\theta_P} \|L - P(Y \mid \theta_P)\|_1 \tag{3}$$

where θ_P is the parameter of the predictor P and L is the real noise-level map.

Owing to varying noise levels in SAR images, their statistical characteristics display considerable variability, making it difficult for denoising algorithms to adapt and effectively process images with different noise levels. At the same time, the stability and convergence of denoising algorithms are affected when faced with different noise levels. Especially when training deep learning models, this high variability in the data can lead to convergence difficulties or numerical stability issues during the training process. To improve the stability and convergence of the denoising algorithm and generalization and to make full use of the leading role of the noise-level map, we perform a special transformation on the original image before inputting it. Specifically, we first multiply the original image Y by $\sqrt{\hat{L}}$ to obtain a new image denoted as $\sqrt{\hat{L}}Y$, which can be further denoted as $\sqrt{\hat{L}}XN$ or $X(\sqrt{\hat{L}}N)$. As a result, the mean of the noise $\sqrt{\hat{L}}N$ becomes $\sqrt{\hat{L}}$ and the variance is kept at 1.

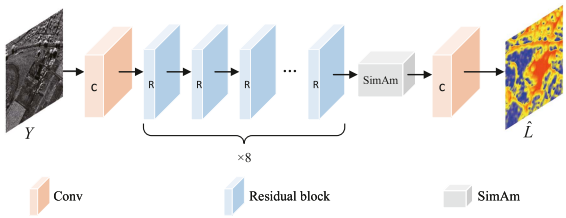

Fig. 2. Structure of NLE.

Although this transformation effectively reduces the fluctuation of the image noise, it also causes a shift in the overall image. To restore the mean of the noise to 1, we further transform the entire image to $X[\sqrt{\hat{L}}N-(\sqrt{\hat{L}}-1)]$. As a result, the noise becomes $\sqrt{\hat{L}}N-(\sqrt{\hat{L}}-1)$, whose mean is $\sqrt{\hat{L}}-(\sqrt{\hat{L}}-1)=1$ and the variance is still 1. This transformation achieves data normalization, ensuring that whatever the input value of $\hat{L}$, the data fed into the model are consistent in statistical distribution. Finally, the input image can be represented in the following form:

$$X\left[\sqrt{\hat{L}}N-(\sqrt{\hat{L}}-1)\right]=\sqrt{\hat{L}}Y-(\sqrt{\hat{L}}-1)X. \quad (4)$$

The nonblind denoising process ultimately includes two steps: First, a preliminary denoised image $\tilde{X}$ is obtained using a nonblind denoising subnetwork structure that acts as a normal denoising network, noting that there is no noise level to bootstrap this process

$$\tilde{X}=F(Y). \quad (5)$$

Then, nonblind denoising is performed using a nonblind denoiser network guided by a noise-level map

$$\hat{X}=F\left(\sqrt{\hat{L}}Y-\left(\sqrt{\hat{L}}-1\right)\tilde{X},\hat{L},\tilde{X}\right) \quad (6)$$

where $\hat{L}$ is the noise-level prediction map corresponding to Y, and $\hat{X}$ is the final denoised image guided by $\hat{L}$. Next, we refer to $\sqrt{\hat{L}}Y-(\sqrt{\hat{L}}-1)\tilde{X}$ as $\tilde{Y}$ for ease of presentation.

In this process, we attempt to eliminate the noise N from the real noisy image. Since the distribution level of SAR noise in the real image is unknown, a standard noise reduction model may not effectively determine the mapping from the noisy image to the clean image. Therefore, the estimated noise level, represented by the distribution parameter L, is input during the denoising phase. This approach allows for an accurate denoising procedure that takes into account the noise-level distribution. The following sections describe each substructure in detail.

B. Noise-Level Estimation Network

The primary goal at this point is to accurately estimate the noise level of the SAR image. Therefore, we propose a noise estimation network to estimate the noise level. The noise-level estimation block is shown in Fig. 2. The NLE module mainly consists of two basic convolutional layers, eight residual blocks, and one SimAM block [56]. Each residual block contains 2 convolutional layers, which have 64 feature channels and use a 3×3 filter with a step of 1. The padding is set to 1 to maintain consistent image dimensions. The final convolutional layer has a single output channel, representing the network output as a single-channel feature map. This can be interpreted as an estimate of the image noise level, L. Crucially, regions with different noise levels from the surrounding areas require higher weighting. SimAM attention [56] is consistent with this view. Therefore, this article introduces SimAM attention before the last convolutional layer in the noise-level predictor. Its incorporation before the last convolutional layer of the noise predictor allows for more efficient noise prediction and correction without introducing additional parameters. This network then provides effective noise-level estimation for real SAR images, thus providing important information for the subsequent nonblind noise reduction phase and ensuring a more accurate and efficient noise removal process.

C. Noise-Feature-Guided Denoising Network

The NFGDN structure proposed in this article mainly consists of a hierarchical encoder–decoder denoising module (T-unet) based on the Transformer block and a DEC module. The NFGDN is shown in Fig. 3. In the structural design, the denoising module is based on the traditional U-net structure and is extended and optimized. This module first inputs the corrected processed image along with the original noisy image, and at the same time, inspired by DeamNet [57], the image is converted from the pixel domain to the feature domain, which can extract the high-dimensional feature information in the image. Therefore, this network first maps the image from the pixel domain to the feature domain through the convolution module and fuses it to form the initial input of the model. The specific fused features are computed as follows:

$$f_0=C_f\left(C_1\left(\tilde{Y}\right)\odot C_2(Y)\right) \quad (7)$$

where f_0 represents the fused features; $C_r(r=1,2)$ indicates a convolutional layer with a convolutional kernel size of 3×3; $\odot$ symbolizes the feature cascade and refers to a convolutional layer with a convolutional kernel size of 1×1 for feature fusion. During the coding phase, the model performs a downsampling operation and detects the long-range dependencies within the image using multiple Transformer blocks. The feature dimensions are reduced by the downsampling module, preserving intermediate features for the decoding phase. Decoding begins with upsampling, then enlarges the feature sizes and combines them with the features from the coding phase. Jump connections preserve the original detail information, facilitating efficient feature extraction and transformation, resulting in a U-shaped network topology. This completes the primary noise feature processing; yet on that basis, subsequent optimization is required to ensure denoising efficiency.

The introduction of the DEC resolves this problem. It extracts key features to guide the denoising process and achieves global optimization of the features of the denoised image by multiplying the output features of the denoising module with the features of the DEC block. This approach helps preserve the intricate structure of the original image.

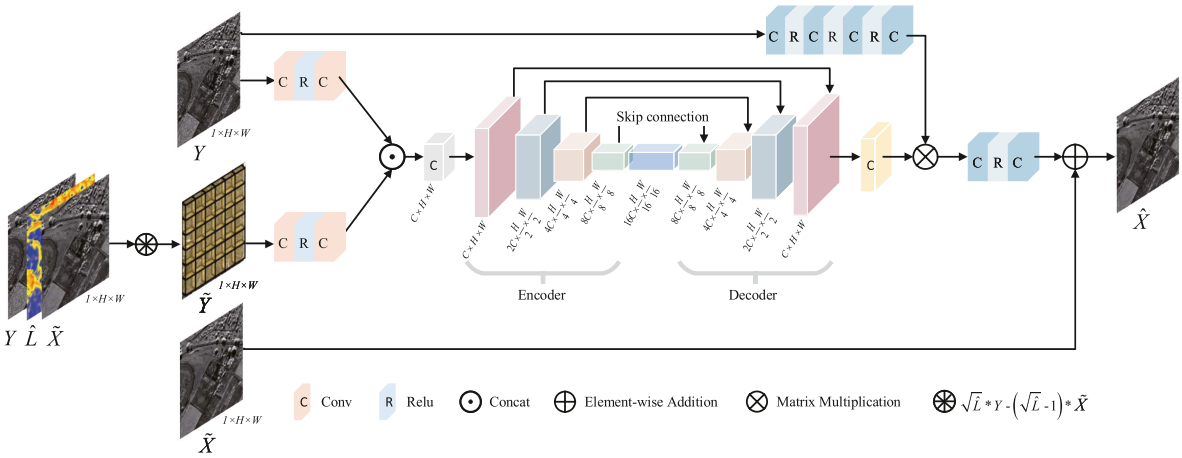

Fig. 3. Structure of NFGDN.

Specifically, this module consists of four convolutional layers tasked with feature extraction and mapping the output control signal to the [0, 1] range using a sigmoid activation function. This control signal balances the denoising and noisy image information, exerting adaptive control to the subsequent denoising process. Initially, the DEC block receives the original image features and exports a global weight matrix that is applied to the current denoising results. This allows for dynamic adjustments in the denoising process, emphasizing contextual information and maximizing detail preservation in the noisy image. The DEC block provides a refined denoising approach that significantly improves image quality and denoising effect.

Finally, the extracted features are remapped to the image domain by the Tail module and merged with the preliminary denoised image to complete the denoising process. The mathematical representation is

$$\hat{X} = C_{\text{tail}}\left(C\left(f_{T\text{-unet}} \otimes F_{\text{Dec}}\left(Y\right)\right) + \tilde{X}\right) \tag{8}$$

where $f_{(\text{T-unet})}$ represents the features identified by the denoising network, F_{Dec} symbolizes the DEC controller, which is the global optimization achieved by multiple convolutional layers and a sigmoid activation function. The symbol $\otimes$ represents the convolution operation, and C_{tail} is the final convolutional layer in the NFGDN network. C denotes a convolution layer with a convolution kernel size of 3×3.

D. Deep-Attention

In the SAR image denoising task, this study introduces a denoising module based on the Transformer block. Since the primary computational challenge in the Transformer comes from the SA layer, the computational complexity of this layer increases quadratically with the number of image patches, which limits the effectiveness of the technique for high-resolution SAR images. In order to facilitate the denoising of high-resolution SAR images, and following the work in [58], we chose to apply SA on the feature dimension rather than the spatial dimension. This approach allows the network to model the interaction of feature channels rather than direct pixel pairs, computing the cross-covariance to derive the attention map from the input features projected by the keys and queries. This mechanism of attention across feature dimensions alleviates computational constraints while preserving the core benefits of SA.

Before computing the feature covariance, this study adopts a local context mixing strategy as described in [58]. To be specific, we use a 1×1 convolution for pixel-level aggregation across channel contexts and a 3×3 deep convolution for channel-level aggregation of local contexts.

Beginning with the tensor $f_0 = C_f(C_1(\tilde{Y}) \odot C_2(Y))$, we create a query (Q), key (K), and value (V) projection abundant in the local context. This is achieved using the 1×1 convolution C_p for pixel-level crosschannel context aggregation and the 3×3 deep convolution C_d for channel-level local context aggregation

$$\begin{aligned} Q &= C_d^Q \times C_p^Q \times f_0 \\ K &= C_d^K \times C_p^K \times f_0 \\ V &= C_d^V \times C_p^V \times f_0. \end{aligned} \tag{9}$$

Here, C_p is a 1×1 pointwise convolution, and C_d is a 3×3 depthwise convolution.

It is worth noting that as the depth of the network increases, the similarity of the attentional mapping between different Transformer modules also increases, which hinders their feature extraction capabilities. This phenomenon is mainly due to the attention collapse problem. To counteract this, we propose a "deep-attention" approach that incorporates adjustable parameters to dynamically refine the query and key representations. This facilitates a more detailed attention weight calculation, which improves the performance of the multihead attention model. Specifically, the introduced adjustable parameters, α and β, for query and key improve their representations during attention weight computation. The refined attention weights are formulated as follows:

$$\begin{aligned} &\text{Deep-Attention}(Q, K, V) \\ &= \text{Softmax}\left(\frac{(Q \cdot \alpha) \cdot (K \cdot \beta)^T}{\gamma}\right) \cdot V. \end{aligned} \tag{10}$$

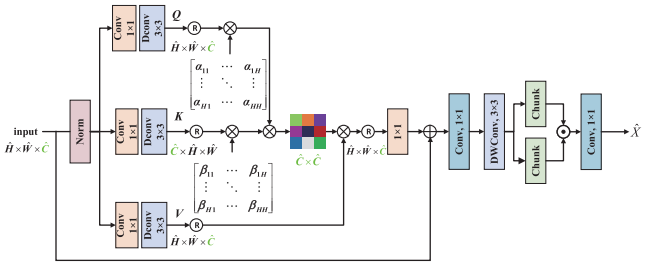

Fig. 4. Transformer block with a deep-attention mechanism.

Fig. 5. Real SAR images. (a) SAR1. (b) SAR2. (c) SAR3.

Among them, α and β are learnable parameters for optimizing Q and K. They allow the model to make fine-grained adjustments to query and key, which directly affects the generation of attentional weights. This approach provides finer control over the computation of attentional weights, thereby improving model performance. The parameters α and β are continuously learned and optimized during the training process. Here, gamma is a learning scaling parameter that controls the size of the dot product of $Q \cdot \alpha$ and $K \cdot \beta$ before applying the Softmax function.

Next, we reshape the query and key projections so that their dot products interact to generate a transposed attention map A of size $C \times C$, instead of a huge regular attention map of size $(H \times W) \times (H \times W)$.

In summary, the module aims to address the challenges in the SAR image denoising task by introducing feature-dimensional SA, local context blending, and deep-attention mechanism, the structure of which is shown in Fig. 4. Through this, the module model overcomes the limitations of the conventional SA-based approach in processing high-resolution images.

E. Loss Function

During the noise-level estimation stage, the objective is to estimate the noise level of the image with precision. In the context of the noise predictor and based on (3), the training loss l_P for noise prediction is defined as follows:

$$L_P = \left\| L - \hat{L} \right\|_1. \tag{11}$$

During the nonblind denoising stage, besides utilizing the l_1 loss between the denoised and clean images as the loss function, the total variation loss is introduced to preserve image continuity. The expression for TV loss is as follows:

$$L_{\mathrm{TV}} = \sum_{p,q} \Big[(F(Y)_{p+1,q} - F(Y)_{p,q})^2 + (F(Y)_{p,q+1} - F(Y)_{p,q})^2 \Big]^{1/2} \tag{12}$$

where $F(Y)_{(p,q)}$ denotes the pixel value of $F(Y)$ at coordinates (p, q). Therefore, the mathematical expression for the total loss function L_T is as follows:

$$L_T = \left\| X - \hat{X} \right\|_1 + \lambda L_{\mathrm{TV}} + L_P. \tag{13}$$

IV. Experimental Results and Discussion

To thoroughly evaluate the performance of our proposed ANED-Net, we performed experimental validation on both synthetic data and real datasets.

A. Data

Optical remote sensing images serve as the predominant training data for speckle noise removal models. These images are ideal references due to their high spatial resolution and minimal noise. In particular, they exhibit a reduced susceptibility to multiplicative noise interference found in SAR images, which enhances their suitability for training speckle removal models. Furthermore, given the similarities between optical remote sensing images and real SAR images, these optical images allow the model to acquire prior knowledge that is more closely aligned with SAR data. This minimizes the domain gap between training and real-world applications, ultimately improving the performance and accuracy of the despeckle model. Therefore, we used the UC Merced land use dataset [59] for our training purposes. This dataset consists of 21 different image categories, each containing 100 color images of 256 × 256 dimensions. From these, 600 were randomly selected for training, and 40 different images were selected as the test set after each training session. Since SAR images are typically grayscale, we processed the images in our training dataset into grayscale and introduced multiplicative noise within the range of L between [1, 20]. Furthermore, L is employed as labeled data to train the noise-level estimation network.

For testing, our dataset includes both simulated and real images. For the simulated images, we included five different visualization levels (L = 1, 2, 4, 8, 10) using well-known datasets, such as McMaster [60], Classic5 [61], Set12 [62], and Kodak24 [63]. The real image subset evaluated three different SAR scene images as shown in Fig. 5(a)–(c). Fig. 5(a) shows the image acquired by TerraSAR-X in StripMap mode in Barcelona, Spain, which is a single-look. Fig. 5(b) shows an image acquired by TerraSAR-X in StripMap mode of Uluru, Australia, which is a single-look, and Fig. 5(c) shows a two-look X-band amplitude image of Bedfordshire, southeast England, acquired by DRA SAR, U.K.

B. Experimental Setup

After 400 training iterations, our network model was built using the Adam optimization algorithm [64], with $\beta 1$ and $\beta 2$

values set to 0.9 and 0.999, respectively. To diversify our training samples, we employed random image enhancements such as horizontal flipping and rotations (including 90°, 180°, and 270°). The learning rate strategy employed observes a cosine decay that tapers from an initial 1e-4 to 1e-6. Within the Unet structural layout of the denoising module, the default stages for both the encoder and decoder N are fixed at 4. λ in our loss function is 0.01. Our deep learning network runs on the Pytorch 1.12.1 platform under the Linux operating system, using the NVIDIA GeForce RTX 3090 GPU for training.

C. Evaluation Index

In evaluating the performance of denoising algorithms, we selected appropriate evaluation metrics based on the characteristics of the data. For the synthetic dataset, we used the Peak Signal-to-Noise Ratio (PSNR) and structural similarity index (SSIM) as evaluation criteria [49]. PSNR mainly measures the similarity of the denoised image to the noiseless image, and a higher value of PSNR usually represents better noise suppression and image similarity. SSIM, on the other hand, is used to evaluate the performance of the denoised image in terms of preserving structure, especially edge information. SSIM is used to evaluate the performance of denoised images in preserving structure, especially the performance of edge information, and higher SSIM values indicate better recovery of image details. For real SAR images, due to the lack of noiseless reference images, we used nonreference metrics such as ENL [65], coefficient of variation (COV) [66], mean ratio (MOR) [67], edge-preservation degree based on the ratio of mean degree based on the ratio of average, EPD-ROA) [68]. ENL is calculated on a manually selected homogeneous region; if the ENL value of the region is higher, it indicates that the homogeneous region is smoother and better filtered. COV is used as a measure of the ratio of the standard deviation of pixel intensities in the homogeneous region of a SAR image to the mean value; a lower COV value indicates that the details and textures of homogeneous regions in SAR images are effectively preserved. MOR is a bias indicator. When the MOR value of the denoising algorithm is close to 1, it can be considered that the method performs excellently in preserving the radiance information and reducing the bias of the deblurred image. EPD-ROA is obtained by calculating the average value of the ratio of the edge-preserving distances of the pixels of the image and is used to evaluate the protection capability of the denoising method of the SAR image for the edge features. EPD-ROA is obtained by calculating the average value of the ratio of the edge preserving distances of the pixels of the image and is used to evaluate the ability of the SAR image denoising methods to protect the edge features. Larger EPD-ROA values indicate that the denoising method is better at protecting the edge features of SAR images.

D. Comparison Methods

In order to verify the reliability and effectiveness of the proposed algorithm, we compare it with other SAR image denoising algorithms, including the PPB [14], SAR-BM3D [13], SAR-DRN [16], MoNet [19], MRDDANet [54], SAR-CAM [55] SAR-Transformer [27], and our ANED-Net. Finally, we will analyze the performance of the proposed algorithms in terms of objective metrics and subjective vision.

E. Analysis of Synthetic Data

In this experiment, synthetic datasets with different levels of speckle noise ($L = 1, 2, 4, 8, 10$) are used, and several denoising algorithms are comprehensively evaluated on four different data sets using two performance metrics, PSNR and SSIM. The average PSNR and SSIM values obtained by each denoiser on these datasets are exhaustively listed in Tables I and II, where the best results are shown in bold. In comparison, the network structure proposed in this article performs well in terms of average PSNR/SSIM values, improving PSNR by 0.01–0.71 dB and SSIM by 0.00–0.04 compared to the next best method. To visualize the difference in the visual effect of each denoiser more intuitively, Fig. 6 shows the results of the recovery comparison at $L = 1$ noise level. From Fig. 6(c), (d), and (i), it can be seen that PPB, SAR BM3D, and SAR Transformer can effectively eliminate speckle noise, but excessive smoothing occurs, resulting in the loss of high-frequency details; from Fig. 6(e) and (f), it can be seen that SAR-DRN and MoNet produce artifact problems, which, in turn, degrade the image quality. MRDDANet SAR-CAM can effectively suppress speckle noise, but some image details are lost, as shown in Fig. 6(g) and (h). On the contrary, the denoising model proposed in this article, by adopting the encoder–decoder architecture and integrating the Transformer module to obtain the global receptive field, not only successfully preserves the structural information of the image, but also achieves a significant advantage in visual effect compared with the other seven methods through the adaptive noise-level despeckling strategy of ANED-Net, which successfully removes all the speckles and also recovers the tiny textures and edges, thus achieving a more realistic image effect.

F. Real SAR Image Analysis

In order to comprehensively evaluate the denoising capability of ANED-Net, we selected three real SAR images as samples, named SAR1 to SAR3, and made an in-depth comparison with today's mainstream denoising algorithms. Figs. 7–9 show the performance of the three images after restoration. Fig. 7 shows the restored SAR1 image. Fig. 7(b) shows that PPB can effectively remove noise, but an oversmoothing problem occurs. Meanwhile, Fig. 7(c) shows that SAR BM3D has a good noise suppression effect, but there is still a considerable amount of noise remaining in the restored image. SAR-DRN and MOnet have the problem of blurred edges and artifacts, as shown in Fig. 7(d) and (e). SAR-Transformer, SAR-CAM, and MRDDANet can suppress the noise and preserve the edge information of the image effectively, but the details remain flawed. In contrast, ANED-Net shows a better denoising effect in which it successfully preserves the details and textures of the image, achieving a good balance between denoising and preserving the image details.

Figs. 8 and 9 show the restoration results of SAR2 and SAR3, respectively, which are similar to those of SAR1. Compared

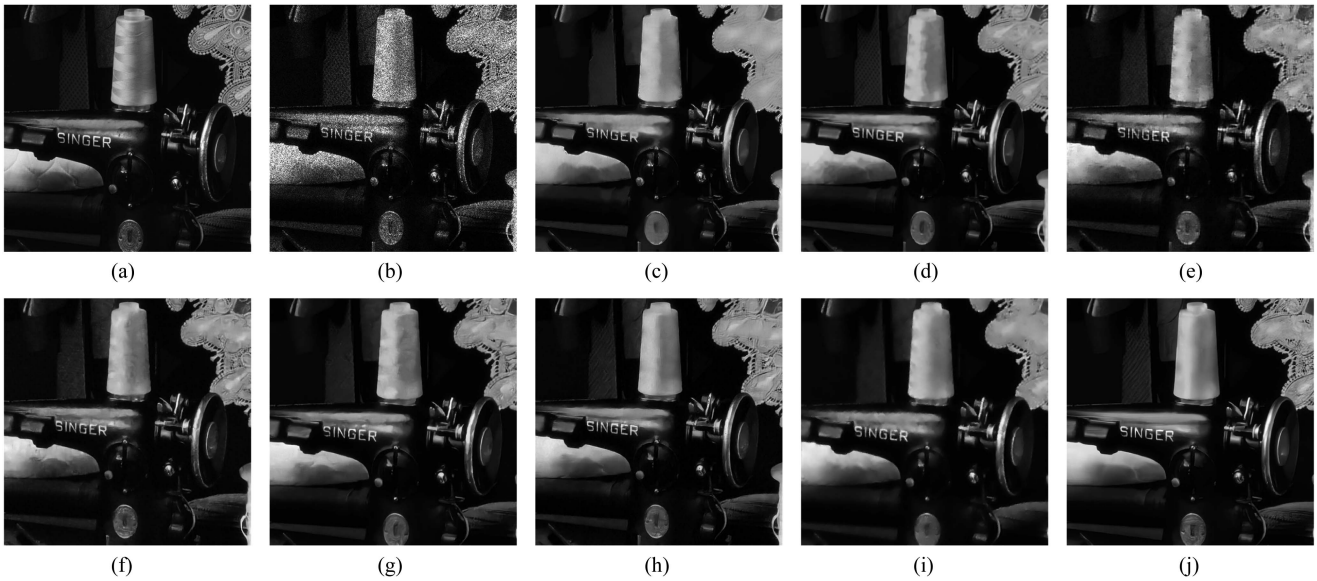

Fig. 6. Denoised results of different methods for image with $L = 1$ speckle noise. (a) Original. (b) Noise. (c) PPB. (d) SAR-BM3D. (e) SAR-DRN. (f) MoNet. (g) MRDDANet. (h) SAR-CAM. (i) SAR-Transformer. (j) Proposed.

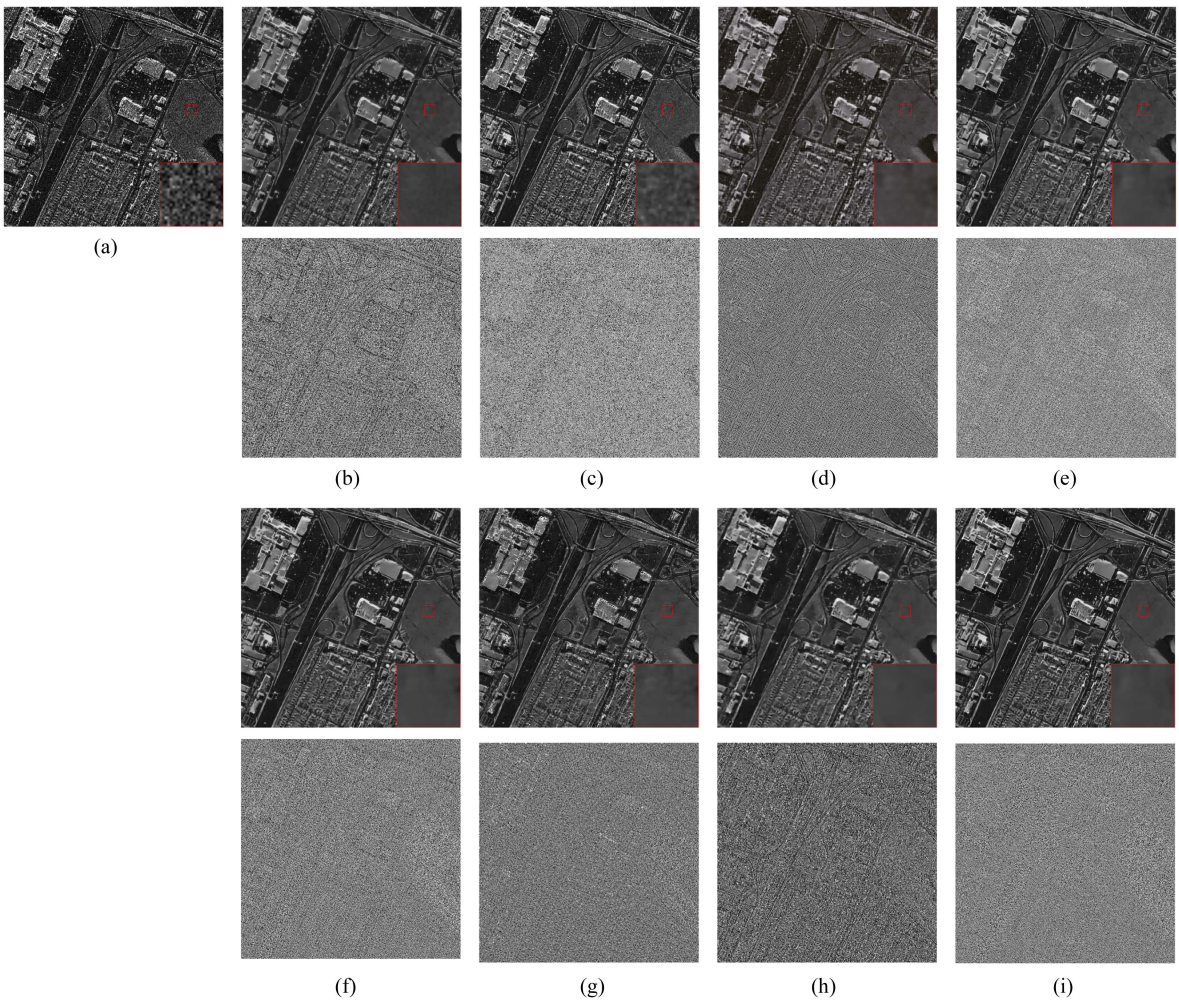

Fig. 7. Performance comparison results of different methods on the real SAR image SAR1 with ratio image. (a) Noisy image. (b) PPB. (c) SAR-BM3D. (d) SAR-DRN. (e) MONet. (f) MRDDANet. (g) SAR-CAM. (h) SAR-Transformer. (i) Proposed.

TABLE I
AVERAGE PSNR VALUE OF SEVEN DENOISING METHODS ON FOUR DATASETS

Datasets	L	PPB	SAR-BM3D	SAR-DRN	MONet	MRDDANet	SAR-CAM	SAR-Transformer	Proposed
Set12	1	24.43	23.68	24.36	25.47	25.93	25.53	24.82	**26.64**
	2	25.13	26.16	25.76	26.56	27.91	27.27	25.86	**27.98**
	4	26.87	28.20	27.22	27.85	29.17	28.95	28.10	**29.80**
	8	27.76	29.94	28.51	29.68	31.07	30.56	29.83	**31.30**
	10	28.16	30.48	29.34	30.29	31.56	31.19	30.58	**31.82**
Kodak24	1	23.95	24.21	25.50	26.37	26.48	26.30	24.92	**27.12**
	2	25.25	26.33	26.68	27.34	28.30	27.90	26.53	**29.01**
	4	27.15	28.28	28.09	27.80	29.57	29.43	28.58	**30.01**
	8	28.01	30.12	29.34	29.56	31.36	30.94	30.42	**31.76**
	10	28.59	30.16	30.28	29.89	31.93	31.45	30.56	**32.30**
McMaster	1	25.51	25.62	26.61	27.41	27.91	27.65	26.69	**28.45**
	2	26.93	27.05	27.97	28.54	29.84	29.47	28.85	**29.96**
	4	27.45	29.94	29.56	29.66	31.23	31.16	29.46	**31.73**
	8	28.14	31.74	30.86	31.73	33.21	32.84	31.86	**33.92**
	10	28.69	32.68	31.32	32.00	33.74	33.49	31.54	**34.47**
Classic5	1	23.65	24.69	25.06	25.68	26.35	25.96	25.24	**27.10**
	2	24.58	24.91	25.15	25.83	26.46	25.90	26.26	**28.53**
	4	25.99	28.97	27.59	27.80	29.60	29.11	27.69	**30.24**
	8	27.01	30.52	28.70	29.56	31.54	30.65	30.27	**31.55**
	10	27.91	31.00	29.55	29.89	31.89	31.16	30.48	**32.10**

The bold values indicate the best performance.

TABLE II
AVERAGE SSIM VALUE OF SEVEN DENOISING METHODS ON FOUR DATASETS

Datasets	L	PPB	SAR-BM3D	SAR-DRN	MONet	MRRDANet	SAR-CAM	SAR-Transformer	Proposed
Set12	1	0.60	0.72	0.67	0.72	0.74	0.75	0.71	**0.73**
	2	0.76	0.77	0.71	0.75	0.81	0.80	0.75	**0.81**
	4	0.77	0.82	0.77	0.78	0.83	0.83	0.79	**0.85**
	8	0.77	0.86	0.79	0.83	0.86	0.87	0.82	**0.89**
	10	0.80	0.87	0.83	0.85	0.88	0.88	0.82	**0.90**
Kodak24	1	0.69	0.70	0.66	0.70	0.72	0.73	0.66	**0.73**
	2	0.70	0.75	0.71	0.74	0.79	0.79	0.72	**0.76**
	4	0.70	0.81	0.77	0.77	0.82	0.82	0.76	**0.79**
	8	0.71	0.85	0.80	0.83	0.87	0.87	0.81	**0.85**
	10	0.72	0.86	0.83	0.84	0.88	0.88	0.83	**0.86**
McMaster	1	0.79	0.77	0.73	0.77	0.80	0.79	0.76	**0.78**
	2	0.80	0.82	0.77	0.80	0.85	0.84	0.81	**0.85**
	4	0.81	0.86	0.82	0.82	0.88	0.88	0.83	**0.88**
	8	0.81	0.89	0.84	0.88	0.91	0.91	0.88	**0.92**
	10	0.82	0.90	0.88	0.89	0.92	0.92	0.89	**0.92**
Classic5	1	0.71	0.67	0.62	0.66	0.71	0.69	0.65	**0.69**
	2	0.72	0.73	0.67	0.69	0.79	0.75	0.69	**0.80**
	4	0.72	0.79	0.73	0.77	0.82	0.80	0.76	**0.80**
	8	0.73	0.83	0.76	0.80	0.86	0.84	0.81	**0.86**
	10	0.74	0.85	0.80	0.81	0.87	0.85	0.83	**0.87**

The bold values indicate the best performance.

with other algorithms, the images processed by our algorithm are clearer, and the edges and textures are fully displayed.

In the evaluation of denoising algorithms, the ratio image is an important tool for evaluating denoising algorithms. It is formed by calculating the point-by-point ratio (i.e., $y/\hat{x}$) between the SAR image and the denoised image. Ideally, the ratio image should consist entirely of noise if the denoising is completely successful. However, if there is a visible texture structure in the ratio image, it may indicate that important information has been mistakenly excluded. For example, looking at Figs. 7–9, one can find visible geometric structures in the ratio images processed by algorithms such as SAR-BM3D and PPB, indicating that they oversmoothed the image edges and textures during the denoising process, which may have lost some of the image details. In contrast, our ANED-Net shows excellent performance in this aspect. It effectively suppresses speckle

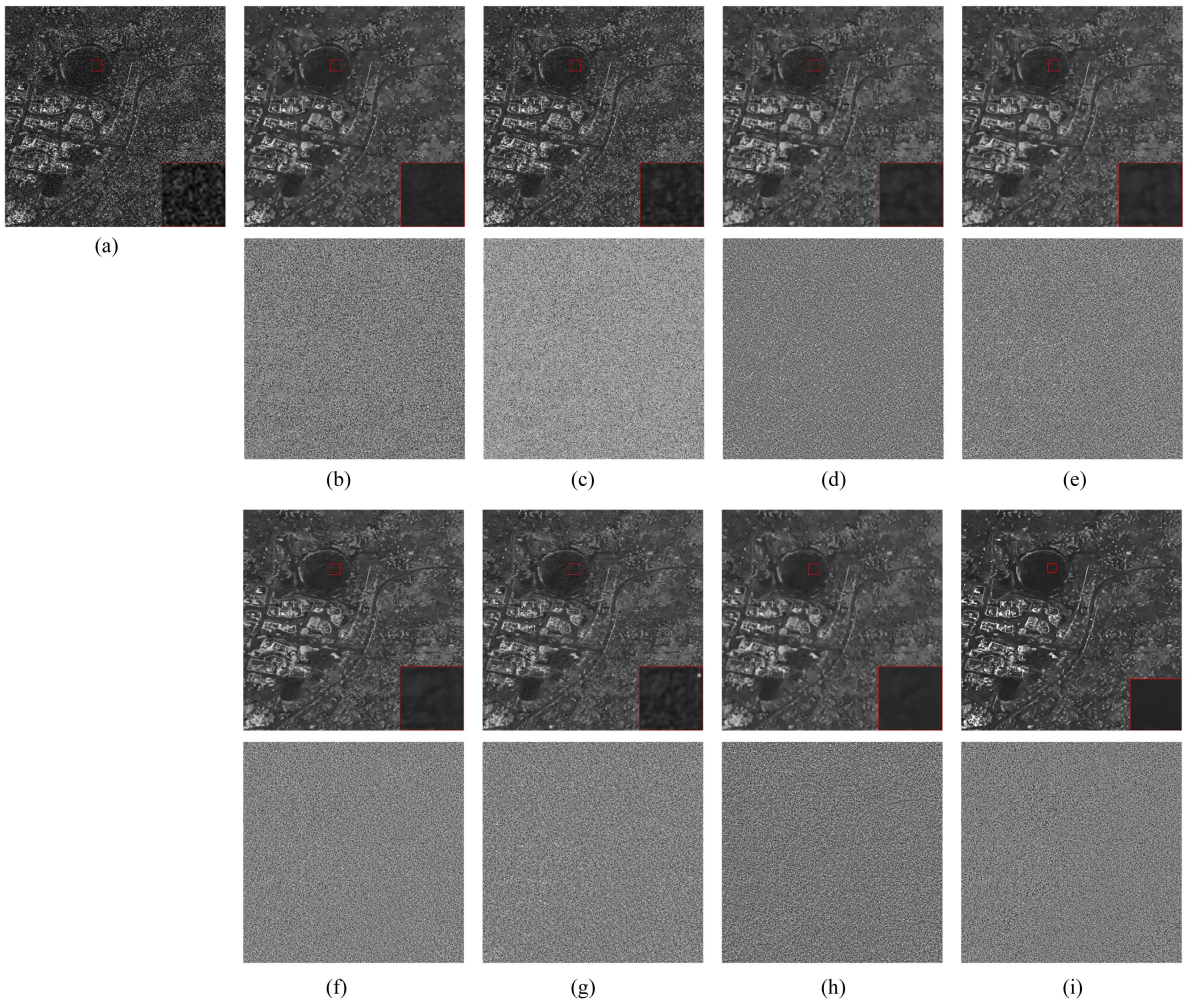

Fig. 8. Performance comparison results of different methods on the real SAR image SAR2 with ratio image. (a) Noisy image. (b) PPB. (c) SAR-BM3D. (d) SAR-DRN. (e) MONet. (f) MRDDANet. (g) SAR-CAM. (h) SAR-Transformer. (i) Proposed.

TABLE III
OBJECTIVE EVALUATION INDEX VALUES OF DIFFERENT DENOISING METHODS FOR SAR1 IMAGE

method	ENL	COV	MOR	EPD-ROA-HD	EPD-ROA-VD
PPB	250.74	0.26	0.794	0.6888	0.6898
SAR-BM3D	29.95	2.12	0.845	0.6760	0.6949
SAR-DRN	73.68	0.87	1.141	0.6243	0.6299
MoNet	139.86	0.45	1.152	0.6319	0.6363
MRRDANet	579.12	0.11	1.181	0.6466	0.6440
SAR-CAM	501.18	0.12	1.464	0.6584	0.6648
SAR-Transformer	498.90	0.12	1.110	0.6314	0.6346
Proposed	**766.33**	**0.07**	**1.050**	**0.7494**	**0.7699**

The bold values indicate the best performance.

and hardly any structure is visible in the ratio image. This observation shows that ANED-Net not only effectively removes noise but also preserves the geometric content of the underlying image.

As shown in Table III, the overall evaluation results of the five SAR image quality metrics are presented separately in the absence of real clean images. First, we used the equivalent visual number of unreferenced performance metrics for quantitative evaluation. We calculated ENL values on manually selected uniform regions (e.g., the red boxes in Figs. 7–9). The higher the ENL value, the better the filtering effect. ANED-Net achieved the highest ENL values in all selected regions, highlighting its strong speckle suppression ability. Meanwhile, the lowest COV value indicates that the details and textures of homogeneous regions in SAR images are effectively preserved. Second, to further understand the performance of these methods, we also examined the mean ratio (MOR). Ideally, the ratio image contains only speckles with a mean of 1 and a variance of $1/L$. In comparison, the MOR value of ANED-Net is closer to 1 than the other methods, demonstrating that the radiance information is better preserved in the despeckled image, introducing less bias. Finally, the EPD-ROA values of our method along the HD and VD directions are closer to 1 than the other methods, indicating that the present algorithm is better able to protect the edge features of the image. Overall, the data in Tables III–V further confirm the superiority of ANED-Net in its despeckling ability, which achieves effective noise suppression while ensuring the preservation of image detail and quality.

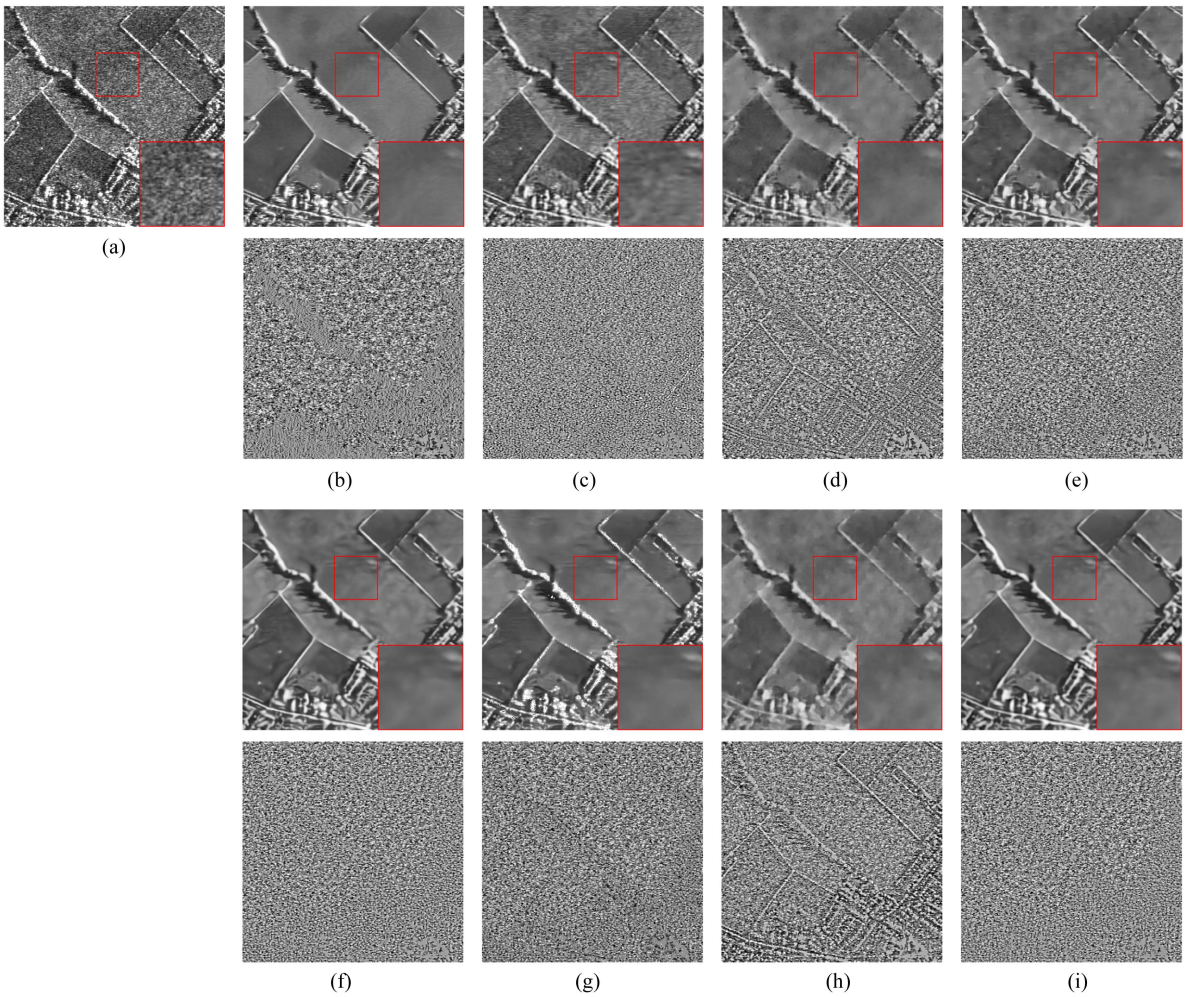

Fig. 9. Performance comparison results of different methods on the real SAR image SAR3 with ratio image. (a) Noisy image. (b) PPB. (c) SAR-BM3D. (d) SAR-DRN. (e) MONet. (f) MRDDANet. (g) SAR-CAM. (h) SAR-Transformer. (i) Proposed.

TABLE IV
OBJECTIVE EVALUATION INDEX VALUES OF DIFFERENT DENOISING METHODS FOR SAR2 IMAGE

method	ENL	COV	MOR	EPD-ROA-HD	EPD-ROA-VD
PPB	250.59	0.155	0.843	0.7335	0.7115
SAR-BM3D	19.18	1.95	0.857	0.7267	0.7039
SAR-DRN	73.87	0.51	0.850	0.6460	0.6383
MoNet	72.92	0.52	0.856	0.7301	0.6389
MRRDANet	116.13	0.32	0.860	0.6460	0.6384
SAR-CAM	38.91	0.96	0.889	0.6493	0.6415
SAR-Transformer	408.91	0.09	0.855	0.6482	0.6410
Proposed	**672.80**	**0.05**	**0.891**	**0.7351**	**0.7119**

The bold values indicate the best performance.

TABLE V
OBJECTIVE EVALUATION INDEX VALUES OF DIFFERENT DENOISING METHODS FOR SAR3 IMAGE

method	ENL	COV	MOR	EPD-ROA-HD	EPD-ROA-VD
PPB	115.40	0.88	0.947	0.9717	0.9397
SAR-BM3D	55.12	1.82	0.955	0.9791	0.9378
SAR-DRN	111.90	0.904	0.938	0.9785	0.9325
MoNet	95.54	1.05	0.950	0.9791	0.9374
MRRDANet	89.26	1.114	0.946	0.9779	0.9332
SAR-CAM	103.62	0.97	0.944	0.9819	0.9395
SAR-Transformer	117.80	0.887	0.905	0.9816	0.9373
Proposed	**121.06**	**0.84**	**0.969**	**0.9794**	**0.9443**

The bold values indicate the best performance.

G. Test for Adaptation of Noise-Level Distribution

The adaptability of despeckling algorithms to the distribution of noise levels is essential in SAR image processing for the preservation of image details and feature quality. Recent studies, such as the unsupervised denoising strategy adopted by MERLIN [30] and the diverse training data strategy constructed by G-MONet [20], have explored different approaches to achieve adaptation to diverse noise distributions. To demonstrate the effectiveness of our despeckling algorithm in adapting to noise-level distribution, this section will analyze and compare the denoising performance of the aforementioned methods. The test images were acquired in StripMap mode over Barcelona, Spain, by TerraSAR-X, with a size of 1024×1024. Fig. 10 shows that in the red-marked areas of sections (b), (c), and (d), effective denoising is achieved by all three methods. In the blue-marked regions, the results from the MERLIN and G-MONet

TABLE VI
IMPACT OF THE NOISE-LEVEL ESTIMATION MODULE ON SAR IMAGE DENOISING PERFORMANCE

Dataset	NLE	L=2 PSNR(dB)/SSIM	L=4 PSNR(dB)/SSIM	L=8 PSNR(dB)/SSIM	L=10 PSNR(dB)/SSIM
BSD500	✓	29.03/0.81	30.56/0.85	31.76/0.87	32.31/0.9
	–	28.51/0.79	30.02/0.83	31.55/0.84	32.12/0.88

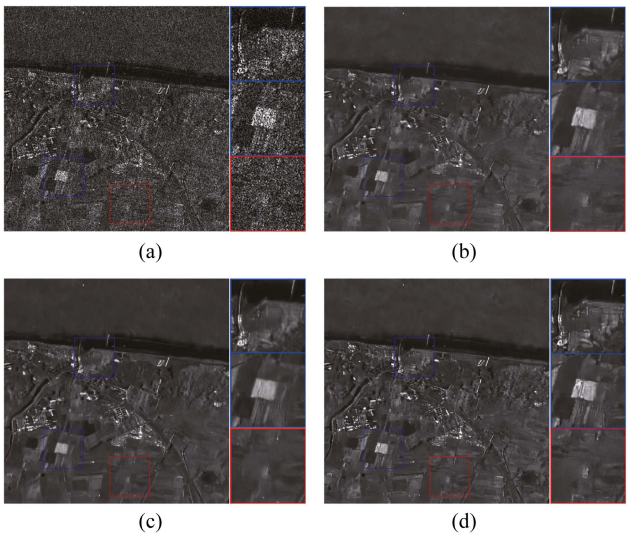

Fig. 10. TerraSAR-X image (Barcelona) and the restoration results. (a) Noise image. (b) G-MONet. (c) MERLIN. (d) Proposed.

approaches appear comparatively smoother, with some loss of texture details. In contrast, our method, as depicted in these sections, demonstrates an advantage in preserving more details and features. Owing to the noise-level estimation applied during the inference process, our algorithm excels in maintaining image details and reducing excessive smoothing, particularly in areas with greater variations in noise levels.

H. Ablation Study

To elucidate in depth the validity of the model proposed in this article, we analyze in detail the role of each component of the designed network structure.

1) Impact and Analysis of Noise Estimation: Our proposed algorithm is an adaptive noise-level denoising method for SAR images, which includes a critical noise-level estimation stage. In this section, we focus on analyzing the effect and role of noise-level estimation on SAR image denoising. To intuitively display the noise-level estimation, we output the predicted noise-level map during the model inference process and compare it with the equivalent number of looks map (ENL map) of the noisy image. Specifically, to obtain the ENL map of the original noisy image, we first applied a 3×3 sliding window to the denoised image and calculated its ENL, which is the squared ratio of the mean to the standard deviation. In the equivalent visual number plot, larger ENL values are typically labeled as cool colors such as blue while smaller ENL values are labeled as warm colors such as yellow or red. The noise-level estimation map also follows this color coding rule.

Fig. 11(a) and (f) shows test comparison images of an image acquired by an airborne system at Sandia National Laboratories and a Ku-band amplitude SAR image acquired at a racetrack in Albuquerque, New Mexico, USA, named SAR4 to SAR5. Fig. 11(b) and (g) shows the ENL maps of the original SAR images while Fig. 11(d) and (i) shows the noise-level maps estimated by the model. By comparison, it is found that the model can estimate the range value of L. Meanwhile, according to the description in [33], when the true ground noise level is unknown, it is preferable to set a higher input noise level rather than a lower one, which is more effective in removing noise to improve image quality. Our noise estimation strategy also satisfies such a phenomenon, and the predicted noise level is slightly higher than the true noise level.

Fig. 11(c), (h), (e), and (j) shows the denoising results of real SAR images with and without noise estimation, respectively, and it can be seen that the image is excessively denoised in the bare ground region and the image is excessively smooth, but after noise estimation in the bare ground region, the texture is clearer. The denoising results show that our model takes into account the differences between different textures, rather than equally using a uniform level of denoiser for all. ANED-Net can guide the model to achieve denoising and ensure texture details through the noise-level map.

Second, to confirm the important role of this noise-level estimation link in the actual image denoising process, we designed corresponding ablation experiments, which specifically analyze the scenarios with and without noise-level estimation in a comparative manner. In the experiments, we selected images from the BSD500 dataset and artificially added different levels of gamma noise (levels of 2, 4, 8, and 10, respectively), and, then, used our algorithm for the denoising process. We carefully observed and recorded the effect of the difference in the denoising effect with and without noise-level estimation. The experimental results are shown in Table VI, and it is found that the algorithm with noise-level estimation has a more significant advantage in removing noise. This further indicates that accurate noise-level estimation is crucial when dealing with real image noise. With the help of noise-level estimation, our algorithm is able to more accurately identify and remove noise from the image, which significantly improves the overall quality of the image.

2) DEC Block Effect: The DEC module plays an important role in the denoising model. It carefully controls the denoising process and helps to preserve the detailed information of the original image. To demonstrate the effectiveness of the DEC module, this article trains and evaluates the models with and without the DEC module. The comparison results are shown in Fig. 12, where Fig. 12(a) shows the synthetic noisy image with the number of looks added equal to 8, and Fig. 12(b) shows the denoised image obtained by the model without the DEC module. As illustrated by the circled area in Fig. 11(b), the denoising process in the model, without the DEC module, relies solely on the basic denoising operation, and although it can reduce the image noise, it is not effective in preserving the image detail

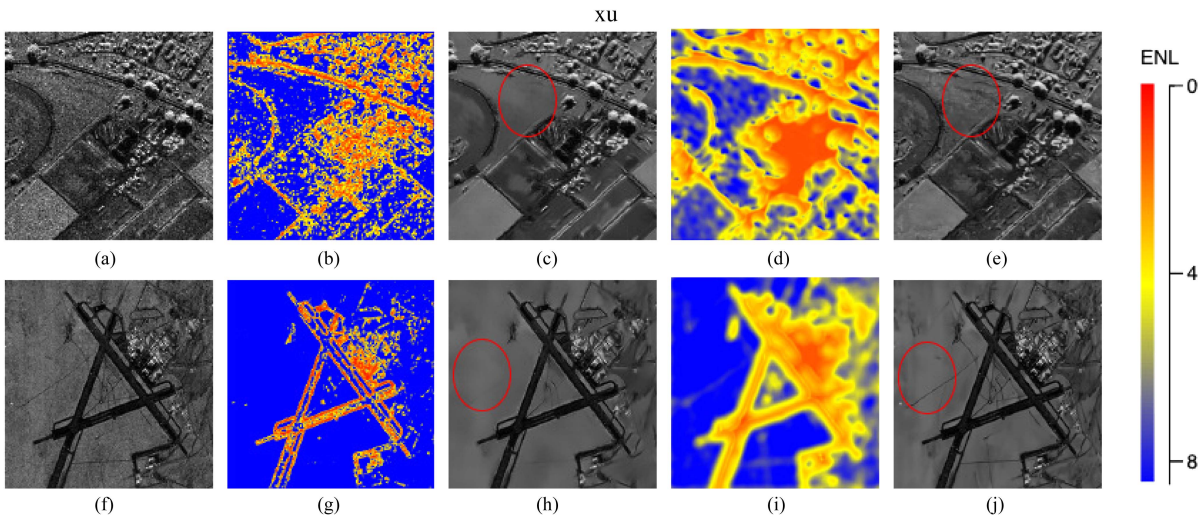

Fig. 11. Comparative analysis of SAR image noise-level estimation and its effect on denoising. (a) SAR4. (b) ENL of SAR4. (c) SAR denoising without noise estimation for SAR4. (d) Noise-level map of SAR4. (e) SAR denoising with noise estimation for SAR4. (f) SAR5. (g) ENL of SAR5. (h) SAR denoising without noise estimation for SAR5. (i) Noise level map of SAR5. (j) SAR denoising with noise estimation for SAR5.

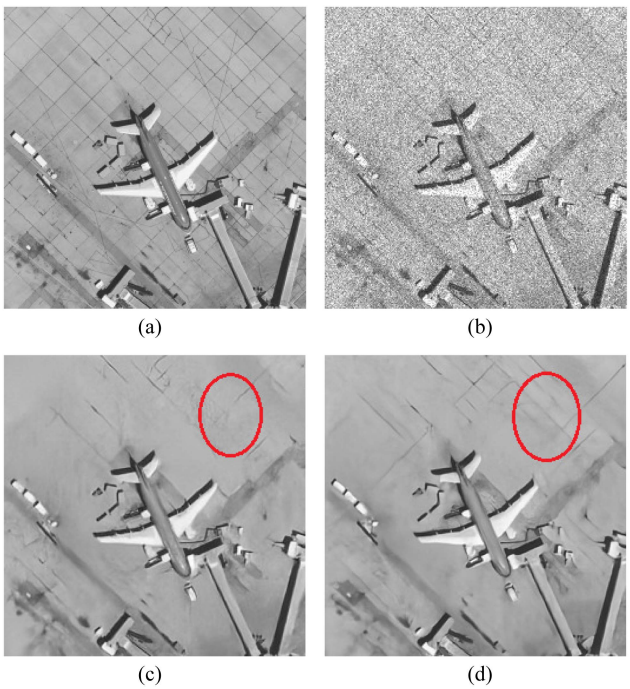

Fig. 12. Denoising effects of denoising enhanced control block. (a) Original. (b) Noise image. (c) Denoising without control block. (d) Denoising with control block.

information. However, when the DEC module is introduced, the situation changes significantly, as shown in Fig. 12(c), the texture details are better preserved with the DEC module compared to the denoised image without the DEC module. The experimental comparisons are shown in Table VII, and the results show that the DEC module significantly improves the quality and denoising effect of the image.

In summary, the DEC block takes the denoised image features from the previous stage as input and outputs a global weight matrix. Then, this weight matrix is applied to the current denoising result, thus realizing the dynamic adaptation of the denoising process. This design makes full use of contextual information in the denoising process and is able to preserve as much detail of the original image as possible during the denoising process.

TABLE VII
IMPACT OF THE DEC BLOCK ON SAR IMAGE DENOISING PERFORMANCE

Dataset	DEC	L=2 PSNR(dB)/SSIM	L=4 PSNR(dB)/SSIM	L=8 PSNR(dB)/SSIM	L=10 PSNR(dB)/SSIM
BSD500	✓	29.14/0.81	30.54/0.85	31.75/0.87	32.37/0.9
	–	28.62/0.76	30.01/0.83	31.45/0.83	32.16/0.87

TABLE VIII
INFLUENCE OF DEEP-ATTENTION ON DENOISING PERFORMANCE IN SAR IMAGES

Number of Transformer blocks	2 PSNR(dB)/SSIM	4 PSNR(dB)/SSIM	8 PSNR(dB)/SSIM	16 PSNR(dB)/SSIM
Deep-Attention	26.50/0.71	26.74/0.71	26.90/0.72	27.11/0.74
Self-Attention	26.41/0.70	26.65/0.71	26.79/0.71	26.71/0.71

3) Deep-Attention Effect: The deep-attention mechanism is designed to solve the problem of attention collapse that occurs in the Transformer in order to optimize the feature extraction capability of the model. To verify its effectiveness, we compared it with the original SA mechanism. First, we embedded a different number of Transformer blocks (e.g., 2, 6, and 8) in each layer of the structure of T-UNet. Keeping the number of Transformer blocks constant, we compared the model performance using the ANED-Net mechanism with the original SA mechanism (BSD500 dataset with noise level $L = 1$ added). Experimental results are presented in Table VIII, indicating that with an equal number of Transformer blocks, the model employing the ANED-Net mechanism demonstrates enhanced performance compared to the one utilizing the original SA mechanism. This result proves that our deep-attention mechanism can effectively improve the attention collapse phenomenon and enhance the feature extraction ability of the model. In our ANED-Net mechanism, we introduce two learnable parameters

α and β to dynamically adjust the query and key representations to optimize the calculation of attention weights. These parameters are continuously learned and optimized by the model during the training process, allowing the model to better utilize the multihead attention mechanism to capture different aspects of the input data. In conclusion, by directly comparing with the original SA mechanism, our deep-attention mechanism proves its superiority and provides an effective solution to the attention collapse problem in the Transformer.

V. Conclusion

In this article, the problem of SAR image despeckling is studied in depth, and a network called ANED-Net is proposed. This network effectively removes unknown-level speckle noise from SAR images by combining noise-level estimation and a nonblind despeckling algorithm. To improve the denoising performance, we also constructed the NFGDN, which consists of a hierarchical encoder–decoder denoising module based on Transformer blocks and a DEC block. At the same time, considering the problem of attention collapse when the Transformer network is extended in-depth, we propose the deep-attention mechanism to ensure that the model extracts features effectively even when the depth is increased. Extensive experimental results show that ANED-Net exhibits excellent denoising performance under various noise conditions compared to many other mainstream methods. In future work, we plan to conduct more in-depth research on blind denoising networks to improve the denoising effect on real SAR images with complex noise.

References

[1] A. Moreira, P. Prats-Iraola, M. Younis, G. Krieger, I. Hajnsek, and K. P. Papathanassiou, "A tutorial on synthetic aperture radar," *IEEE Geosci. Remote Sens. Mag.*, vol. 1, no. 1, pp. 6–43, Mar. 2013.

[2] F. Lattari, A. Rucci, and M. Matteucci, "A deep learning approach for change points detection in InSAR time series," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Mar. 2022, Art. no. 5223916.

[3] J.-S. Lee, "Digital image enhancement and noise filtering by use of local statistics," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-2, no. 2, pp. 165–168, Mar. 1980.

[4] J.-S. Lee, "Refined filtering of image noise using local statistics," *Comput. Graph. Image Process.*, vol. 15, no. 4, pp. 380–389, 1981.

[5] D. T. Kuan, A. A. Sawchuk, T. C. Strand, and P. Chavel, "Adaptive noise smoothing filter for images with signal-dependent noise," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-7, no. 2, pp. 165–177, Mar. 1985.

[6] Z. Sun, Z. Zhang, Y. Chen, S. Liu, and Y. Song, "Frost filtering algorithm of SAR images with adaptive windowing and adaptive tuning factor," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 6, pp. 1097–1101, Jun. 2020.

[7] A. Baraldi and F. Parmiggiani, "A refined gamma MAP SAR speckle filter with improved geometrical adaptivity," *IEEE Trans. Geosci. Remote Sens.*, vol. 33, no. 5, pp. 1245–1257, Sep. 1995.

[8] X. J. Yang and P. Chen, "SAR image denoising algorithm based on Bayes wavelet shrinkage and fast guided filter," *J. Adv. Comput. Intell. Intell. Inform.*, vol. 23, no. 1, pp. 107–113, 2019.

[9] X. Qian et al., "Ridgelet-Nets with speckle reduction regularization for SAR image scene classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 11, pp. 9290–9306, Nov. 2021.

[10] D. Devapal, N. Hashna, V. Aparna, C. Bhavyasree, J. Mathai, and K. S. Soman, "Object detection from SAR images based on curvelet despeckling," *Mater. Today: Proc.*, vol. 11, pp. 1102–1116, 2019.

[11] G. Liu, H. Kang, Q. Wang, Y. Tian, and B. Wan, "Contourlet-CNN for SAR image despeckling," *Remote Sens.*, vol. 13, no. 4, 2021, Art. no. 764.

[12] S. Liu, Q. Hu, P. Li, J. Zhao, C. Wang, and Z. Zhu, "Speckle suppression based on sparse representation with non-local priors," *Remote Sens.*, vol. 10, no. 3, 2018, Art. no. 439.

[13] S. Parrilli, M. Poderico, C. V. Angelino, and L. Verdoliva, "A nonlocal SAR image denoising algorithm based on LLMMSE wavelet shrinkage," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 2, pp. 606–616, Feb. 2012.

[14] C.-A. Deledalle, L. Denis, and F. Tupin, "Iterative weighted maximum likelihood denoising with probabilistic patch-based weights," *IEEE Trans. Image Process.*, vol. 18, no. 12, pp. 2661–2672, Dec. 2009.

[15] G. Chierchia, D. Cozzolino, G. Poggi, and L. Verdoliva, "SAR image despeckling through convolutional neural networks," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2017, pp. 5438–5441.

[16] Q. Zhang, Q. Yuan, J. Li, Z. Yang, and X. Ma, "Learning a dilated residual network for SAR image despeckling," *Remote Sens.*, vol. 10, no. 2, pp. 1–18, 2018.

[17] M. Zhang, L.-d. Yang, D.-h. Yu, and J.-b. An, "Synthetic aperture radar image despeckling with a residual learning of convolutional neural network," *Optik*, vol. 228, 2021, Art. no. 165876.

[18] R. Qin, X. Fu, J. Chang, and P. Lang, "Multilevel wavelet-SRNet for SAR target recognition," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, Jan. 2021, Art. no. 4009005.

[19] S. Vitale, G. Ferraioli, and V. Pascazio, "Multi-objective CNN-based algorithm for SAR despeckling," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 11, pp. 9336–9349, Nov. 2021.

[20] S. Vitale, G. Ferraioli, A. C. Frery, V. Pascazio, D.-X. Yue, and F. Xu, "SAR despeckling using multi-objective neural network trained with generic statistical samples," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, Sep. 2023, Art. no. 5216812.

[21] P. Wang, H. Zhang, and V. M. Patel, "Generative adversarial network-based restoration of speckled SAR images," in *Proc. IEEE 7th Int. Workshop Comput. Adv. Multi-Sensor Adaptive Process.*, 2017, pp. 1–5.

[22] Y. Sun, L. Lei, D. Guan, X. Li, and G. Kuang, "SAR image speckle reduction based on nonconvex hybrid total variation model," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 2, pp. 1231–1249, Feb. 2021.

[23] F. Lattari, B. G. Leon, F. Asaro, A. Rucci, C. Prati, and M. Matteucci, "Deep learning for SAR image despeckling," *Remote Sens.*, vol. 11, no. 13, 2019, Art. no. 1532.

[24] Y. Gui, L. Xue, and X. Li, "SAR image despeckling using a dilated densely connected network," *Remote Sens. Lett.*, vol. 9, no. 9, pp. 857–866, 2018.

[25] J. Zhang, W. Li, and Y. Li, "SAR image despeckling using multiconnection network incorporating wavelet features," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 8, pp. 1363–1367, Aug. 2020.

[26] J. Li, Y. Li, Y. Xiao, and Y. Bai, "HDRANet: Hybrid dilated residual attention network for SAR image despeckling," *Remote Sens.*, vol. 11, no. 24, 2019, Art. no. 2921.

[27] M. V. Perera, W. G. C. Bandara, J. M. J. Valanarasu, and V. M. Patel, "Transformer-based SAR image despeckling," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2022, pp. 751–754.

[28] A. B. Molini, D. Valsesia, G. Fracastoro, and E. Magli, "Speckle2Void: Deep self-supervised SAR despeckling with blind-spot convolutional neural networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Mar. 2021, Art. no. 5204017.

[29] E. Dalsasso, L. Denis, and F. Tupin, "SAR2SAR: A semi-supervised despeckling algorithm for SAR images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 4321–4329, Apr. 2021.

[30] E. Dalsasso, L. Denis, and F. Tupin, "As if by magic: Self-supervised training of deep despeckling networks with MERLIN," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Nov. 2021, Art. no. 4704713.

[31] D. Gragnaniello, G. Poggi, G. Scarpa, and L. Verdoliva, "SAR image despeckling by soft classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 6, pp. 2118–2130, Jun. 2016.

[32] X. Yang, L. Denis, F. Tupin, and W. Yang, "SAR image despeckling using pre-trained convolutional neural network models," in *Proc. Joint Urban Remote Sensing Event*, 2019, pp. 1–4.

[33] K. Zhang, W. Zuo, and L. Zhang, "FFDNet: Toward a fast and flexible solution for CNN-based image denoising," *IEEE Trans. Image Process.*, vol. 27, no. 9, pp. 4608–4622, Sep. 2018.

[34] Y. Cuii, G. Zhou, J. Yang, and Y. Yamaguchi, "Unsupervised estimation of the equivalent number of looks in SAR images," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 4, pp. 710–714, Jul. 2011.

[35] S. Foucher, J. M. Boucher, and G. B. Bénie, "Maximum likelihood estimation of the number of looks in SAR images," in *Proc. 13th Int. Conf. Microw., Radar, Wireless Commun.*, 2000, vol. 2, pp. 657–660.

[36] D. Hu et al., "Investigation of variations in the equivalent number of looks for polarimetric channels," in *Proc. 7th Int. Workshop Sci. Appl. SAR Polarimetry Polarimetric Interferometry*, 2015, pp. 1–5.

[37] Z. Liu et al., "Swin transformer: Hierarchical vision transformer using shifted windows," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2021, pp. 10012–10022.

[38] S. Niklaus, L. Mai, and F. Liu, "Video frame interpolation via adaptive convolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 670–679.
[39] X. Wang, R. Girshick, A. Gupta, and K. He, "Non-local neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 7794–7803.
[40] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, and Y. Fu, "Image super-resolution using very deep residual channel attention networks," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 286–301.
[41] A. Vaswani et al., "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, vol. 30, pp. 1–11.
[42] A. Vaswani et al., "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 5998–6008.
[43] D. Zhou et al., "DeepViT: Towards deeper vision transformer," 2021, *arXiv:2103.11886*.
[44] A. Moreira, "Improved multilook techniques applied to SAR and ScanSAR imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 29, no. 4, pp. 529–534, Jul. 1991.
[45] H. Hu, Z. Zhang, Z. Xie, and S. Lin, "Local relation networks for image recognition," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2019, pp. 3463–3472.
[46] H. Zhao, J. Jia, and V. Koltun, "Exploring self-attention for image recognition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 10076–10085.
[47] A. Dosovitskiy et al., "An image is worth 16x16 words: Transformers for image recognition at scale," in *Proc. Int. Conf. Learn. Representations*, 2020, vol. 1, pp. 1–30.
[48] H. Touvron, M. Cord, M. Douze, F. Massa, A. Sablayrolles, and H. Jégou, "Training data-efficient image transformers & distillation through attention," in *Proc. Int. Conf. Mach. Learn.*, 2021, pp. 10347–10357.
[49] L. Yuan et al., "Tokens-to-token VIT: Training vision transformers from scratch on ImageNet," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2021, pp. 558–567.
[50] W. Wang et al., "Pyramid vision transformer: A versatile backbone for dense prediction without convolutions," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2021, pp. 568–578.
[51] E. Xie, W. Wang, Z. Yu, A. Anandkumar, J. M. Alvarez, and P. Luo, "SegFormer: Simple and efficient design for semantic segmentation with transformers," in *Proc. Adv. Neural Inf. Process. Syst.*, 2021, vol. 34, pp. 12077–12090.
[52] N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov, and S. Zagoruyko, "End-to-end object detection with transformers," in *Proc. Eur. Conf. Comput. Vis.*, 2020, pp. 213–229.
[53] Z. Liu et al., "Swin transformer: Hierarchical vision transformer using shifted windows," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2021, pp. 9992–10002.
[54] S. Liu, Y. Lei, L. Zhang, B. Li, W. Hu, and Y.-D. Zhang, "MRD-DANet: A multiscale residual dense dual attention network for SAR image denoising," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Sep. 2022, Art. no. 5214213.
[55] J. Ko and S. Lee, "SAR image despeckling using continuous attention module," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 3–19, Dec. 2021.
[56] L. Yang, R.-Y. Zhang, L. Li, and X. Xie, "SimAM: A simple, parameter-free attention module for convolutional neural networks," in *Proc. Int. Conf. Mach. Learn.*, 2021, pp. 11863–11874.
[57] C. Ren, X. He, C. Wang, and Z. Zhao, "Adaptive consistency prior based deep network for image denoising," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 8596–8606.
[58] S. W. Zamir, A. Arora, S. Khan, M. Hayat, F. S. Khan, and M.-H. Yang, "Restormer: Efficient transformer for high-resolution image restoration," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2022, pp. 5728–5739.
[59] Y. Yang and S. Newsam, "Bag-of-visual-words and spatial extensions for land-use classification," in *Proc. 18th SIGSPATIAL Int. Conf. Adv. Geographic Inf. Syst.*, 2010, pp. 270–279.
[60] L. Zhang, X. Wu, A. Buades, and X. Li, "Color demosaicking by local directional interpolation and nonlocal adaptive thresholding," *J. Electron. Imag.*, vol. 20, no. 2, 2011, Art. no. 023016.
[61] A. Foi, V. Katkovnik, and K. Egiazarian, "Pointwise shape-adaptive DCT for high-quality denoising and deblocking of grayscale and color images," *IEEE Trans. Image Process.*, vol. 16, no. 5, pp. 1395–1411, May 2007.
[62] R. Zeyde, M. Elad, and M. Protter, "On single image scale-up using sparse-representations," in *Proc. 7th Int. Conf. Curves Surf.*, 2012, pp. 711–730.
[63] R. Franzen, "Kodak lossless true color image suite," *Source*, Accessed: Nov. 15, 1999. [Online]. Available: http://r0k.us/graphics/kodak/
[64] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. ICLR*, 2015, pp. 1–15.
[65] S. G. Dellepiane and E. Angiati, "Quality assessment of despeckled SAR images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 2, pp. 691–707, Feb. 2014.
[66] A. G. Mullissa, D. Marcos, D. Tuia, M. Herold, and J. Reiche, "DeSpeckNet: Generalizing deep learning-based SAR image despeckling," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Dec. 2020, Art. no. 5200315.
[67] G. D. Martino, M. Poderico, G. Poggi, D. Riccio, and L. Verdoliva, "Benchmarking framework for SAR despeckling," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 3, pp. 1596–1615, Mar. 2014.
[68] X. Ma, H. Shen, X. Zhao, and L. Zhang, "SAR image despeckling by the use of variational methods with adaptive nonlocal functionals," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 6, pp. 3421–3435, Jun. 2016.

Xu Wang received the B.S. degree in computer science and technology in 2019 from Harbin Engineering University, Harbin, China, where he is currently working toward the Ph.D. degree in software engineering with the Department of Computer Science and Technology.

His research interests include remote sensing image processing, computer vision, and deep learning.

Yanxia Wu received the B.S. degree in computer software and theory, M.S. degree in computer application technology, and Ph.D. degree in computer system architecture from Harbin Engineering University, Harbin, China, in 2002, 2005, and 2008, respectively.

She is currently a Professor with the College of Computer Science and Technology, Harbin Engineering University. Her research interests include the Internet of Things, computer architecture, intelligent computing, and computer vision.

Changting Shi received the Ph.D. degree in computer application technology from the School of Computer, Harbin Engineering University, Harbin, China, in 2010.

He is currently a Lecturer and a Master's Supervisor with Harbin Engineering University. His research interests include intelligent robots, multirobot systems, machine learning, and computer vision.

Ye Yuan (Member, IEEE) received the B.S. degree in computer science and technology and Ph.D. degree in software engineering from the College of Computer Science and Technology, Harbin Engineering University, Harbin, China, in 2017 and 2023, respectively.

His research interests include remote sensing, computer vision, and deep learning.

Xue Zhang received the B.S. degree in computer science and technology in 2019 from Northeast Electric Power University in Jilin, China. She is currently working toward the Ph.D. degree in computer science and technology with Harbin Engineering University, Harbin, China.

Her main research interests include computer vision and remote sensing image detection.