PAPER • OPEN ACCESS

Synthetic aperture radar images denoising based on multi-scale attention cascade convolutional neural network

To cite this article: Huilin Shan *et al* 2023 *Meas. Sci. Technol.* **34** 085403

View the article online for updates and enhancements.

Synthetic aperture radar images denoising based on multi-scale attention cascade convolutional neural network

Huilin Shan[1,2]**, Xiangwei Fu**[2,]*** , Zongkui Lv**[2]**, Xingchen Xu**[2]**, Xingtao Wang**[2] **and Yinsheng Zhang**[1,2]

[1] The School of Electronics and Information Engineering, Wuxi University, Wuxi 214105, People's Republic of China
[2] The School of Electronics and Information Engineering, Nanjing University of Information Science and Technology, Nanjing 210044, People's Republic of China

E-mail: 20211249607@nuist.edu.cn

CrossMark

Abstract

Synthetic aperture radar (SAR) images are often affected by speckle noise, which can hinder accurate interpretation and subsequent use of the images in applications such as target detection and segmentation. To address this issue, we propose a denoising algorithm based on a multi-scale attention cascade convolutional neural network (MSAC-Net). Our algorithm employs multi-scale asymmetric convolution to extract image features and an attention mechanism to integrate these features. Additionally, we designed a multi-layer deep cascade convolutional network to enhance the generalization ability of the model features. Experimental results show that our proposed MSAD-Net model significantly outperforms state-of-the-art SAR image denoising algorithms. Specifically, it achieves a significant improvement in peak signal-to-noise ratio, with an increase of about 0.81–13.97 dB, and structural similarity index measure, with an increase of about 0.01–0.14. Overall, our study presents a novel denoising algorithm for SAR images that greatly improves the accuracy of subsequent image applications.

Keywords: synthetic aperture radar, image denoising, attention mechanism, convolutional neural network

(Some figures may appear in colour only in the online journal)

1. Introduction

Synthetic aperture radar (SAR) finds applications in Earth surface observation, resource management, military, and security fields [1]. SAR produce high-resolution images by receiving and processing radar echo signals. Unlike traditional optical remote sensing satellites, SAR can acquire images despite weather disturbances such as clouds, fog, rain and snow, thus providing stronger all-weather observation capabilities, higher resolution, broader coverage, more accurate measurement capabilities, and more flexible data processing methods. However, SAR images may display speckle noise, which appears as small dots or speckles and is usually a mixture of low and high-frequency components. Speckle noise is characterized by a granular appearance with a black and white texture. This is due to the fact that SAR images are formed through coherent processing of the radar echoes from consecutive pulses, resulting in a pixel-by-pixel variation in the echo intensity [2]. The unique imaging principle of SAR leads to

---

* Author to whom any correspondence should be addressed.

an excessive multiplicative component in SAR images, which gives rise to speckle noise that is more difficult to remove compared to additive noise. Speckle noise and additive noise can be represented as follows:

$$F_{\text{speckle\_img}} = f_{\text{g\_img}} \bullet n_{\text{mult}}$$
$$F_{\text{additive\_img}} = f_{\text{g\_img}} + n_{\text{add}} \tag{1}$$

where $F_{\text{speckle\_img}}$ is the noisy SAR image, $f_{\text{g\_img}}$ is the noise-free SAR image, n_{mult} is the multiplicative noise due to coherent interference, and n_{add} is the additive noise .

Speckle noise can impact image quality and accuracy, particularly in applications requiring precise target identification and measurement. Therefore, restoring a clean image from SAR images with noise is an urgent problem that needs to be addressed.

At present, SAR image denoising is mainly divided into two approaches: image filtering processing [3] and non-coherent multi-look processing [4]. Image filtering processing aims to filter noise by using filters. Previous research by Donoho [5] and Johnstone [6] proposed a wavelet value shrinkage method for inherent speckle noise in SAR images. The method involves performing wavelet transform on both the noise and signal. In the wavelet domain, the wavelet transform itself has a strong ability to concentrate signal energy. This enables it to effectively distinguish the signal from the noise in the frequency domain, and then filter it using a set threshold to achieve the denoising effect. Later, this theory became a fundamental part of traditional denoising algorithms. Another mainstream method in spatial filtering denoising is non-local mean filtering [7]. This algorithm differs from threshold filtering as it is a spatial global algorithm. It converts similar gray information in the entire image into the current pixel gray. Similarity is expressed by Euclidean distance, and effective denoising is achieved by the gray-weighted average of the entire image. The SAR-BM3D algorithm represents this approach [8]. After that, Murali Mohan Babu *et al* [9] further advanced the work of literature [8] by utilizing bilateral sampling to sample image blocks in BM3D, followed by discrete wavelet transform. They also introduced the concept of wavelet thresholding to improve the retention of image details in the denoised image. Later, the sparse representation model became popular for SAR image denoising due to its excellent performance in removing speckle noise [10]. However, its large algorithm size consumes high computing resources, making it less practical for widespread use. Rudin *et al* [11] proposed a total variation (TV) model for denoising, which achieved good results in speckle noise removal despite the staircasing effect it produced. In recent years, literature [12–14] has made significant improvements and optimizations to the TV denoising algorithm. By minimizing the staircasing effect while preserving image edge information, the improved algorithm model maintains its adaptability.

The WNNM algorithm, proposed by Gu *et al* [15], is based on low-rank matrix decomposition and nuclear norm minimization. This algorithm takes advantage of the fact that the low rank of the image and the high rank of the image noise are fundamentally different. By solving the optimization problem of the low rank matrix using the nuclear norm minimization method, this algorithm utilizes the non-local similarity of the image for image denoising. Compared to the BM3D algorithm, the WNNM algorithm shows a significant qualitative improvement in denoising effect. However, despite its good performance in removing SAR speckle noise, this algorithm has significant limitations in practical applications. Another SAR image denoising technique is incoherent multi-look denoising. This method decomposes the image into multiple Doppler bandwidths, using different synthetic apertures for each bandwidth, and then superimposes the decomposed images. While this method effectively removes speckle noise, it can cause a decrease in resolution of the decomposed image. This decrease in resolution may have significant impacts on other tasks such as change detection.

In recent years, deep learning has rapidly developed and is increasingly being used in various industries to process signal tasks. This is because deep learning can effectively extract the distribution and features of signals. For instance, in the field of neuroscience, studies [16, 17] have utilized deep learning to process electromagnetic wave signals in the human brain. In the medical field, a study [18] has detected changes in the thyroid gland in the human body using CT scans to aid doctors in diagnosing patients' conditions. Other studies [19–21] have innovatively changed the algorithm models for detection and recognition, and achieved excellent performance in detection and recognition accuracy. In the field of image denoising, deep learning has also been widely applied. However, the performance of deep learning models may vary depending on the specific application. For natural images, Zhang *et al* [22] proposed the classic denoising convolutional neural network (DnCNN), an end-to-end denoising mechanism that integrates residual network learning and the batch normalization (BN) method to accelerate the network's convergence. The network exhibits good robustness in feature extraction and achieves excellent denoising results. Subsequently, DANet Yue *et al* [23] used the generative adversarial network (GAN) framework to train their model. However, the instability of training based on the GAN model becomes the most significant limitation in adversarial networks, so it takes longer time to converge. Nevertheless, their algorithm produces good results in ordinary image denoising. Another model, FFDNet, was proposed by Zhang *et al* [24], which improves the shortcomings of DnCNN and introduces the noise level mapping M as the input of the model. FFDNet performs well in denoising effect. Since the SAR image noise itself is multiplicative noise, Chierchia *et al* [25] proposed SAR-CNN for SAR image denoising. However, due to the particularity of noise, the algorithm has certain requirements for its training samples. SAR-CNN only selects non-coherent multi-view images, which have inherent disadvantages in noise sample training, and its denoising effect is general. The average peak signal-to-noise ratio (PSNR) is only 25.95 dB. Liu *et al* [26] proposed an effective method for suppressing speckle noise in SAR images using GAN. The network consists of a generator and a discriminator, and uses the TV loss function as the overall optimization

scheme. Compared to traditional neural network algorithms, this method does not require logarithmic transformation and directly performs end-to-end network training, achieving good denoising performance. Dalsasso *et al* [27] proposed a semi-supervised denoising model (SAR2SAR) that mainly uses time series to obtain residual information from the image, which is essentially different from the neural network model. The method achieved denoising performance comparable to that of neural networks. In 2022, Thakur and Maji [28] proposed AGSDNet, which combines attention mechanism and gradient to denoise. While achieving better denoising results, the model itself is heavy-weighted, requires a large number of data sets as support, and the network convergence rate is general.

In summary, traditional denoising algorithms for SAR images typically use global denoising techniques. The principle involves processing and judging the image using global, similar information. However, in cases of high-resolution images, the algorithm requires extensive preprocessing, such as smoothing and pixel discrimination through neighborhood processing of each image block. This algorithm consumes a large amount of computing resources, and its practical application is limited by time and space constraints, making it inefficient for denoising tasks. While these algorithms improve denoising effects, they increase time costs. Deep learning algorithms have shown promise in image denoising. However, current methods such as those described above still have limitations, such as slow network convergence speeds, heavy model weights, and reduced accuracy. To address these issues, this paper proposes a denoising algorithm based on a multi-scale attention cascaded layers network (MSAC-Net). The network employs the idea of multi-scale asymmetric convolution (MAC) and attention. Compared with single convolution kernels, MAC kernels have a good image receptive field, meaning they can extract image information from different scales to capture more detailed image details. Subsequently, the convolution kernels of different scales are integrated into the network, and the attention mechanism is introduced to the stitched feature map to divide the attention of the features, enhancing the main features of the image. The dense cascade block (DCB) is used to further strengthen the features in the middle of the network. Finally, the image restoration and reconstruction are achieved through the subtraction calculation method.

2. Model structure

2.1. Multi-scale asymmetric convolution module

Most CNNs currently use regular convolution structures, such as 3×3, 4×4, 5×5 and $n \times n$. While regular convolution kernels have shown good results in feature extraction, they still have deficiencies in extracting certain types of image edge and fuzzy information. This paper proposes the use of an asymmetric convolution structure, which is an algorithm structure first proposed by Chunwei Tian in 2021, called asymmetric convolution kernel (ACNet) [29]. The network is highly effective in recovering noisy low-resolution images, and asymmetric convolutions can achieve model compression and acceleration
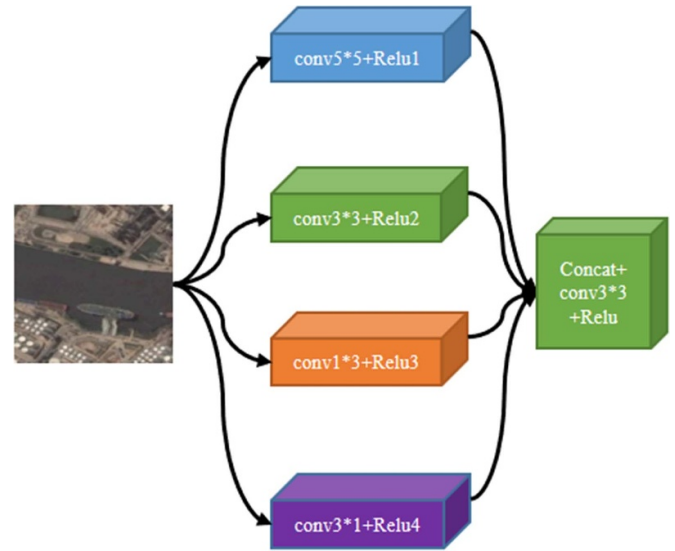

**Figure 1.** Multi-scale asymmetric convolution module.

compared to existing square convolutions. Previous studies have shown that standard convolutions can be decomposed into and convolutions to reduce the number of parameters. The theory behind this is relatively simple: if the rank of the two-dimensional convolution kernel is 1, the operation can be equivalently converted into a series of one-dimensional convolutions. However, since the eigenvalues learned by the convolution kernel in deep networks follow a distribution, and their internal rank is higher than in practice, directly applying the feature image to the convolution kernel can result in significant information loss. Jin *et al* [30] used structural constraints to separate two-dimensional convolutions, and the operation time was nearly twice as fast while still achieving good accuracy. Denton *et al* [31] proposed an infinite approximation low-rank matrix by utilizing the singular value decomposition technique, and subsequently refined the upper network to enhance recovery performance. Jaderberg *et al* [32] successfully applied horizontal and vertical kernels to train their network by minimizing reconstruction errors. Asymmetric convolution is also widely utilized for network structure design. For instance, in Inception-v3, convolution is replaced by a combination of convolution and asymmetric convolution. Additionally, asymmetric convolution is a widely adopted technique for designing network architectures. For example, in Inception-v3, conv 7×7 is replaced by conv 1×7 and conv 7×1. MAC module is used conv 3×3, conv 5×5, conv 1×3, conv 3×1, as shown in figure 1. These four convolution kernels are used to extract the initial features of the image, and finally Concat (Cat) them. The front-end part of the MAC can be expressed as:

$$F(x) = \text{Cat}\left[\text{ReLU}_i\left(\text{conv}\,5 \times 5\,(x)\right), \text{ReLU}_i\left(\text{conv}\,3 \times 1\,(x)\right),\right.$$
$$\left.\text{ReLU}_i\left(\text{conv}\,3 \times 3\,(x)\right), \text{ReLU}_i\left(\text{conv}\,1 \times 3\,(x)\right)\right]. \quad (2)$$

In the expression (1), each convolution kernel needs to pass through the ReLU_i ($i = 1, 2, 3,$ and 4) activation function to filter the image value to the range $[0, \infty]$, so as to prevent the
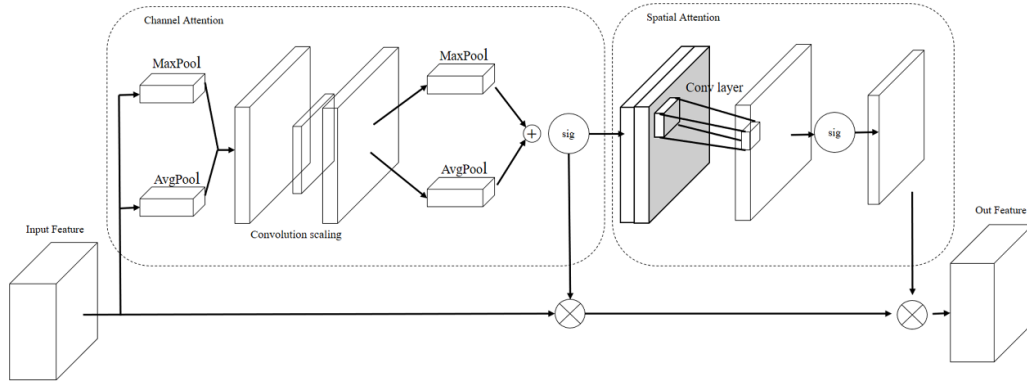
Figure 2. The architecture of CBAM. '⊕' represents feature fusion addition, '⊗' represents feature value multiplication.
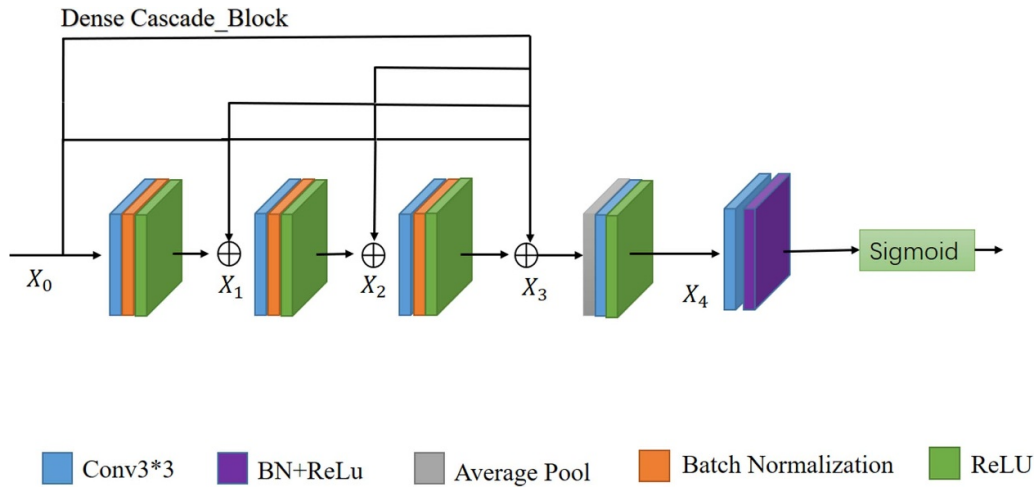
Figure 3. Dense cascade block. '⊕' represents the addition of feature data; '——' connection represents residual connection, '→' represents the direction of data transmission.

gradient explosion and disappearance of the image during the back propagation. $F(x)$ denote output of the front-end model. The expression of the final output of the MAC as:

$$\mathrm{MAC}(O) = \mathrm{ReLU}\left[\mathrm{Conv}\,3 \times 3\left(F(x)\right)\right]. \qquad (3)$$

In the expression (2), $\mathrm{MAC}(O)$ denotes the output of the whole module.

2.2. Attention mechanism

The attention mechanism is a process that autonomously learns a set of weight coefficients throughout the network. It emphasizes the regions of interest while suppressing irrelevant background regions in a dynamic weighting manner. Currently, there are three main attention mechanisms: channel attention, spatial attention, and self-attention. Channel attention constructs weight coefficients for different channels of the current model, and learns to capture the best model channel, and independently reduces the weight system of the remaining channels to strengthen important features and suppress non-important ones. The representative of channel attention is mainly SEnet [33]. Spatial attention improves the feature representation of key regions. The spatial information in the

original image is transformed into another space through the spatial transformation module, and the key information is retained. A weighted mask is generated for each location, and the output is weighted, thereby enhancing the specific target region of interest while weakening the irrelevant background region. This paper introduces the current mainstream attention mechanism, CBAM [34]. Its structure is shown in figure 2, which centralizes the dual attention mechanism of space and channel, and its effect is better.

2.3. Dense cascade block layers

Cascaded is a neural network that has the ability to automatically train and add hidden units. It offers several advantages such as a high learning rate, customizable network neurons and depth, and effective backpropagation. Drawing inspiration from Dense Net [35], this paper proposes a novel dense cascade block (DCB) network structure, as illustrated in figure 3. Each node in the network can be represented as:

$$X_i = \sum_{n=0}^{i-1} X_n + \mathrm{ReLU}\left(\mathrm{BN}\left(\mathrm{Conv}\left(X_{i-1}\right)\right)\right) \qquad (i = 1, 2, 3) \quad (4)$$

$$X_4 = \mathrm{ReLU}\left(\mathrm{Conv}\left(\mathrm{AvgPool}\left(X_3\right)\right)\right). \qquad (5)$$

The structure of the proposed approach comprises a series of convolution modules. During forward propagation, the output of each module is used as input for the subsequent module, resulting in dense connections that enable the parameters to interact and iterate among the convolution layers. Ultimately, the global receptive field of the image is obtained through the final generalized average pooling layer, allowing the network to learn image information more effectively. The expression for the DCB module is as follows:

$$\text{DCB}(o) = \sigma\left(\text{BN}\left(\text{Conv}\left(X_4\right)\right)\right). \quad (6)$$

The output of dense cascade block layers is represented by DCB(o), the σ denote sigmoid function, the sigmoid function maps any real number to a value in the range (0,1), and is defined as follows:

$$\text{Sigmoid}(x) = \frac{1}{1 + \exp(-x)}. \quad (7)$$

3. Denoising mode

3.1. Denoising method

In this section, we introduce MSAC-Net, a denoising network designed by us, as shown in figure 4. MSAC-Net is mainly composed of a MAC module, an attention module, and a DCB. In this paper, the network is designed as an end-to-end structure, where the input is a noisy image and the output is a clean image. The denoising mechanism of this paper is as follows: firstly, the initial ACNet and attention module are used to obtain the basic features of the image. Secondly, the DCB independently learns the residual information of the image. Finally, the output features of the last layer are fed back, and the image is reconstructed after denoising. The network parameters of MSAC-Net are shown in table 1.

3.2. Loss function

Due to the particularity of SAR image, the noise image is mostly multiplicative noise, so the data set trained by this model needs to be preprocessed by adding noise. The image is denoised according to $Y(a,b) = n(\sigma,\delta) \times X(a,b)$, where $Y(a,b)$ represents the input noise image, $X(a,b)$ represents the unnoised image, $n(\sigma,\delta)$ and represents the noise with a variance of σ. After the input noise image passes through the network MSAC-Net, a residual image (Resim) is output, expressed as:

$$\text{MSAC - Net}\left[Y(a,b)\right] = \text{Resim}(a,b) = Y(a,b) - n(\sigma,\delta). \quad (8)$$

Subsequently, a given mean square error is used to train the denoising network, and its loss function is expressed as:

$$\text{loss}(w,b) = \frac{1}{N}\left\|\text{MSAC} - \text{Net}\left[Y(a,b)\right] - n(\sigma,\delta)\right\|^2 \quad (9)$$

where w and b are the weights and bias learned in the network, and $\text{loss}(w,b)$ are the set of all parameters learned in the network.

3.3. Evaluating indicator

In this paper, we conduct qualitative and quantitative experiments to evaluate and demonstrate the performance of our proposed MSAC-Net model in image denoising. We visually assess the clarity and completeness of the denoised image, and use objective evaluation metrics such as PSNR and structural similarity index measure (SSIM). If the size of the original clean image is $N \times M$ and the denoised image is y then PSNR can be expressed as:

$$\text{PSNR} = 10\log_2\left\{\frac{N \times Mf^2 \max}{\sum\limits_{i=1}^{N}\sum\limits_{j=1}^{M}\left[x(a,b) - y(a,b)\right]^2}\right\}. \quad (10)$$

Among them, $f$max represents the maximum intensity of the input image. For some 8 bit grayscale images, there are 256 possible grayscale values, so $f$max = 256. To measure the denoising effect of the model, we employ PSNR. However, it is also important to consider other indicators that can evaluate the difference between the original image and the denoised image. SSIM is one such quality assessment measure that compares the two images. For calculating SSIM, we use two non-negative image signals, x and y

$$\text{SSIM}(x,y) = \frac{\left(2\mu_x\mu_y + c_1\right)\left(2\sigma_{xy} + c_2\right)}{\left(\mu_x^2 + \mu_y^2 + c_1\right)\left(\sigma_x^2 + \sigma_y^2 + c_2\right)} \quad (11)$$

where μ_x and μ_y represent the average strength of X and Y, respectively, and σ_x and σ_y represent the standard deviation of X and Y, respectively. σ_{xy} is the covariance of images x and y, c_1 and c_2 are constant values. Local parameters μ_x, μ_y, σ_x, σ_y and σ_{xy} are in a 8×8 square window and are calculated in pixel pans.

3.4. Comparison algorithms

Comparing algorithms is an essential step in data analysis and machine learning research, as each algorithm has its unique advantages and limitations. Moreover, comparing different algorithm models can help us understand their performance on various datasets and evaluation metrics. In this study, we evaluate the performance of MSAC-Net by comparing it with state-of-the-art algorithm models, including WNNM [14], SAR-BM3D [8], SAR-CNN [23], GAN [21], SAR2SAR [25], and AGSDNet [28]. To ensure a fair comparison, we utilize the default settings provided by the authors in their respective algorithm literature and calculate PSNR and SSIM as error metrics to comprehensively evaluate their practical application value.

4. Experimental results and analysis

4.1. Experimental platform

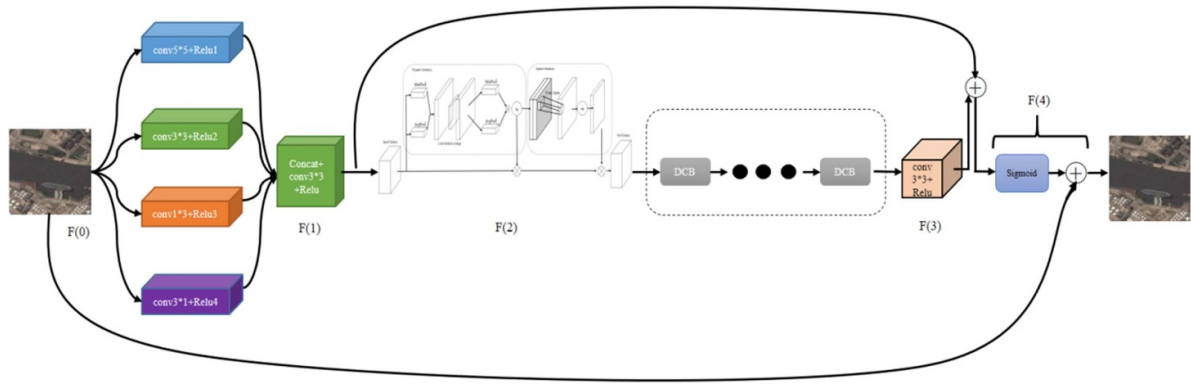The experimental platform of this paper is built, as shown in table 2.

Figure 4. The architecture of MSAC-Net.

Table 1. MSAC-Net network parameters. '—' means no; (1,0) means to fill in rows, (0,1) means to fill in columns; the 'label'–'label' represents the middle position of the two labels; '*n*' in [label–label]$_n$ represents the number from top to bottom.

Local	Kernel size	Stride	Padding	Channel	Pooling (size)
$[\boldsymbol{F}(0) - \boldsymbol{F}(1)]_1$	5×5	1	2	64	—
$[\boldsymbol{F}(0) - \boldsymbol{F}(1)]_2$	3×3	1	1	64	—
$[\boldsymbol{F}(0) - \boldsymbol{F}(1)]_3$	1×3	1	(0,1)	128	—
$[\boldsymbol{F}(0) - \boldsymbol{F}(1)]_4$	3×1	1	(1,0)	128	—
$F(1)$	3×3	1	1	64	—
$F(2)$	—	—	—	32	—
$\boldsymbol{x}0$	3×3	1	1	32	—
$\boldsymbol{X}_1 - \boldsymbol{X}_2$	3×3	1	1	64	—
$\boldsymbol{X}_2 - \boldsymbol{X}_3$	3×3	1	1	32	—
$\boldsymbol{X}_3 - \boldsymbol{X}_4$	3×3	1	1	64	Average (2×2)
$F(3)$	3×3	1	1	1	—
$F(4)$	3×3	1	1	1	—

Table 2. Experiment platform.

Hardware	Detail
GPU	NVIDIA GeForce RTX 3090 GPU
CPU	12th Gen intel(R) Core(TM) i9-12900KF @ 3.19 GHz
Language framework	Anconda + Python3.8.13 + Pytorch1.11.0
GPU memory	24 G
CPU memory	32 G
System	Windows 10

4.2. Datasets

The MSAC-Net network proposed in this paper requires noisy images and corresponding noise-free images as training sets. However, obtaining image datasets such as SAR can be difficult. Therefore, we selected the NWPU-RESISC45 dataset of Northwestern Polytechnical University [36] for our study and manually synthesized the required dataset. To do this, we first selected 1000 images from the NWPU-RESISC45 dataset, each with 256 × 256 pixels. These images included 100 images of each type, such as airport, coastline, rainforest, terrace, bridge, ship, highway, land and mountain. We used a total of 1000 images as training sets, selecting 20 images of each type for a total of 200 images as verification sets. We also collected a total of 1200 clean, noise-free images. To create the noisy image dataset, we added Gaussian noise to the clean images in a multiplicative manner, using a mean value of 0 and a variance. This completed the dataset of noise images. We conducted comparative experiments in this paper, which included four groups of images with different variance levels ($\sigma = 20, 30, 40, 50$), as shown in figure 5. We verified that the noise adding method used in the dataset was consistent with the method used in the training dataset.

4.3. Experimentation

In this paper, we conducted an experiment using 1000 clean SAR images. At the start of training, each image, which had a resolution of 256 × 256 pixels, was divided into 466 pane images with a resolution of 40 × 40 pixels using a sliding window approach with a step size of 10 pixels. We applied various data augmentation techniques, such as flipping, translating, scaling, rotating and mirroring, to enhance the dataset. The resulting images were used to train the MSAC-Net network.
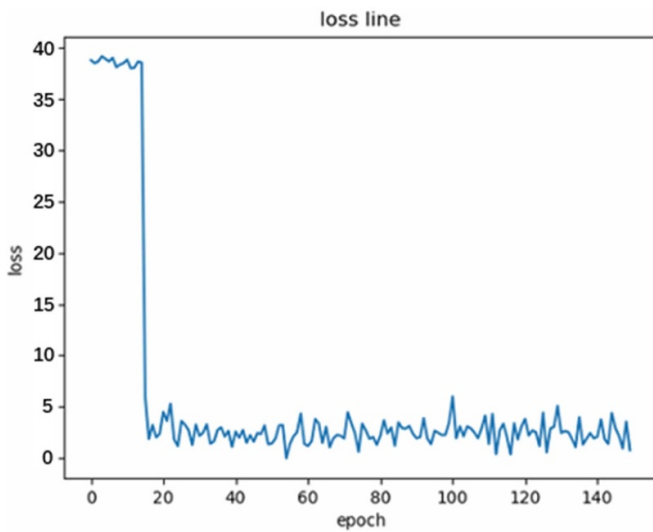
We set the epoch of the network training to 150, with a BN setting momentum of 0.95 and a batch size of 64. We used the Adam [37] optimizer to optimize the network's weight coefficients, and introduced dynamic adjustment learning rate (lr). We initially set the learning rate to 0.01, and then reduced it by a factor of 0.3 at epochs 15, 55, and 115, respectively. The other settings were kept at their default values. The model was trained for 150 rounds, with a total of 1092 188 data iterations. The loss curve is shown in figure 6. During testing, we found that the MSAC-Net could effectively denoise images of any size. In this paper, we implemented Gaussian noise additive processing on images using Python, and used the Pytorch framework to implement the model algorithms.

4.4. Ablation experiment

The purpose of this experiment is to verify the importance and role of each module in the MSAC-Net model. To achieve this, we prepared a small SAR image dataset of 100 images
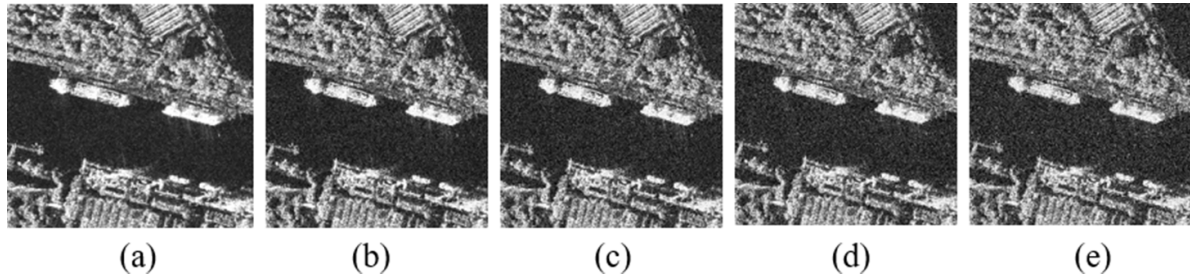
Figure 5. Images at different noise levels. (a) Ground image. (b) Noisy image with $\sigma = 20$. (c) Noisy image with $\sigma = 30$. (d) Noisy image with $\sigma = 40$. (e) Noisy image with $\sigma = 50$.
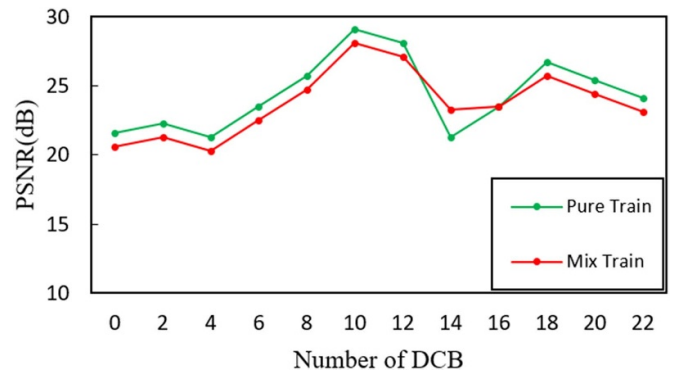

Figure 6. The train model of loss line.

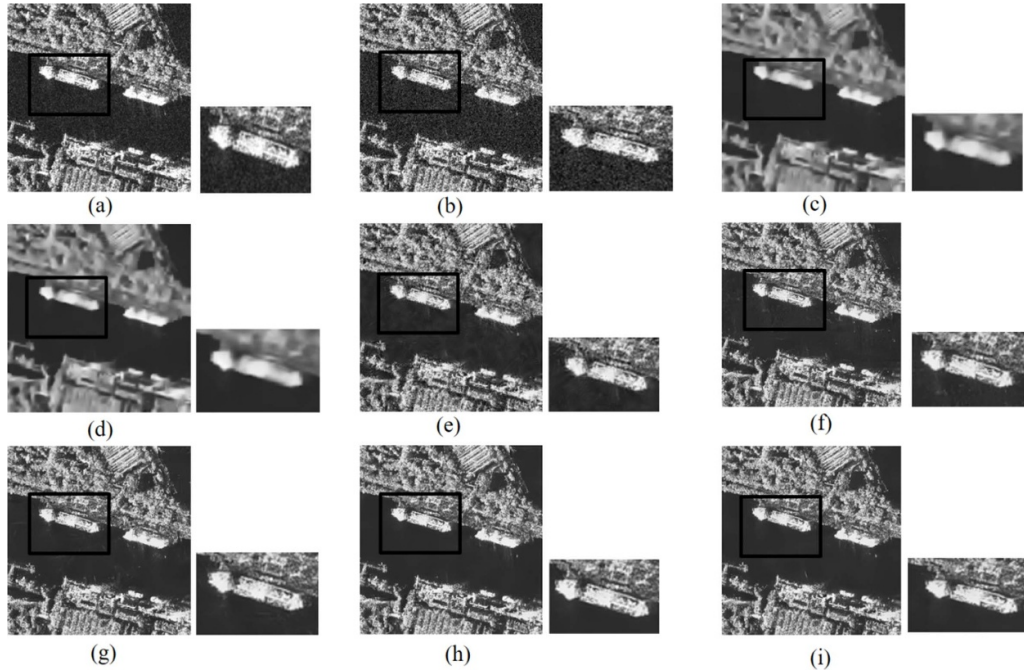**Figure 7.** PSNR values of different DCB layers in different modes.

a moderate impact. In contrast, CBAM has a relatively small effect on the overall model. Overall, our experiment validates the importance and role of different components in MSAC-Net, and the above conclusions can be drawn.

Table 3. Ablation learning.

	MAC	CBAM	DCB	PSNR (dB)
Model A	✓	✗	✓	27.68
Model B	✗	✓	✓	26.34
Model C	✓	✓	SG	23.75
Model D	✓	✓	✓	29.81

4.5. Experimental analysis under different background images

To achieve the best denoising effect of MSAC-Net, this study performed ablation experiments on DCB layers of varying depths. The study compared the performance of different DCB layers using a single background type image for pure train mode and different DCB layers using a multi-background type image for mix train mode. Each layer underwent 150 rounds of training, with a noise level of 40 ($\sigma = 40$). The ship image was selected for pure training mode while ten types of images were used for mix training mode. The ship image was verified ten times in both pure and mix training modes, and the average PSNR value of these ten trials was calculated as shown in figure 7. From the figure, it is evident that the highest PSNR values were achieved when the DCB layer depth was 12 layers, reaching 29.3 dB and 28.5 dB for pure and mix training modes, respectively. Despite the pure training mode having a 0.8 dB higher value than the mix training mode, the 12 layers were still identified as the optimal DCB layers for the network proposed in this study.

In summary, as shown in figure 7, the PSNR value of the network in a pure training mode (using a single type of image) can reach 29.3 dB, while in a mixed training mode (using multiple types of images), its PSNR value is 28.5 dB. Therefore,

with a noise level of 35. We built three different modules for training and testing and compared their denoising performance. The three modules are the MAC module, the attention module (CBAM), and the dense cascade layer module (DCB). By reasonably deleting and adding these three modules, we created four different combinations: Model A (containing only MAC and DCB), Model B (containing only CBAM and DCB), Model C (containing a single DCB and MAC combined with CBAM), and Model D (containing multiple layers of DCB and MAC combined with CBAM). We tested the four combinations on the same image, and we used PSNR to quantify their denoising performance, as shown in table 3.

In table 3, 'SG' means single-layer DCB, '✓' means inclusion relationship, and '✗' means removal relationship. After conducting experiments, we discovered that DCB plays a significant role in the performance of our model, while MAC has

Figure 8. Denoising effect comparison; (a) original image; (b) noisy image; (c) image denoised with WNNM; (d) image denoised with SAR-BM3D; (e) image denoised with SAR-CNN; (f) image denoised with GAN; (g) image denoised with SAR2SAR; (h) image denoised with AGSDNet; (i) image denoised with MSAC-Net.

we can conclude that training the network with only a single type of image can effectively improve its denoising performance. However, due to the complexity and diversity of SAR noise images in practical applications, the pure training mode cannot guarantee that the entire algorithm will have better generalization ability and robustness. Consequently, multi-type image training is more suitable for practical applications, and it can still perform well in terms of denoising performance.

4.6. Comparative analysis

In this paper, we compare seven denoising algorithms using verification images of ships, mountains and coasts. Figure 8 shows the denoising effect on the ship image, while figures 9 and 10 demonstrate the denoising effects on the coast and mountain images, respectively. These figures visually compare the denoising results of the different algorithms at a noise level of 30 ($\sigma = 30$).

This study presents an analysis of the denoising effects of seven algorithms on three images. In figure 8, it can be observed that the denoising results of WNNM, SAR-CNN, and SAR-BM3D exhibit many artifacts and severe texture loss, indicating incomplete removal of noise and blurred visual effects. Although GAN preserved some details, the ship appears blurry, and most of the edge information between the ship and the shore has been erased. While SAR2SAR and AGSDNet removed most of the noise, some scattered speckle noise around the ship's edge was not effectively suppressed, resulting in numerous artifacts around the ship's body and the water surface. In contrast, the proposed algorithm achieved excellent restoration of ship and shore

information, and the water surface around the ship was processed cleanly compared to the original image. In figure 9, both WNNM and SAR-BM3D algorithms significantly blur the wave edge information in the coastal image. In the denoising result images of SAR-CNN, GAN and SAR2SAR, many noise points remain, and some corners are distorted or lost. Although AGSDNet removes most of the noise around the waves, the overall appearance is far from the original image. The proposed method produces cleaner results in the wave edge processing, making the image appear clearer than the original. In figure 10, the four algorithms, GAN, WNNM, SAR-BM3D and SAR-CNN, exhibited mediocre denoising performance, as some regions remain relatively blurred. SAR2SAR fails to recover the sharp edges of mountainous terrain, and AGSDNet exhibits some noise in the corners of gullies. The proposed method removes the majority of the noise, resulting in extremely clear gully lines and superior overall denoising performance compared to the first six algorithms.

The main purpose of the experiment is to test the performance of the proposed MSAC-Net on datasets. Table 4 is the PSNR results of different methods on different images. Table 5 is the SSIM results of different methods on different images. In table 4 and table 5, bold numbers represent the highest in each row. The rows in tables 4 and 5 are the PSNR and SSIM values generated by different algorithms when the noise parameter is $\sigma = 20, 30, 40, 45, 50$, and the columns are the PSNR and SSIM values of the same method at different noise levels. The methods included are: SAR-BM3D model with denoising parameter σ, SAR-CNN model with denoising parameter σ, WNNM model with denoising parameter σ, GAN model
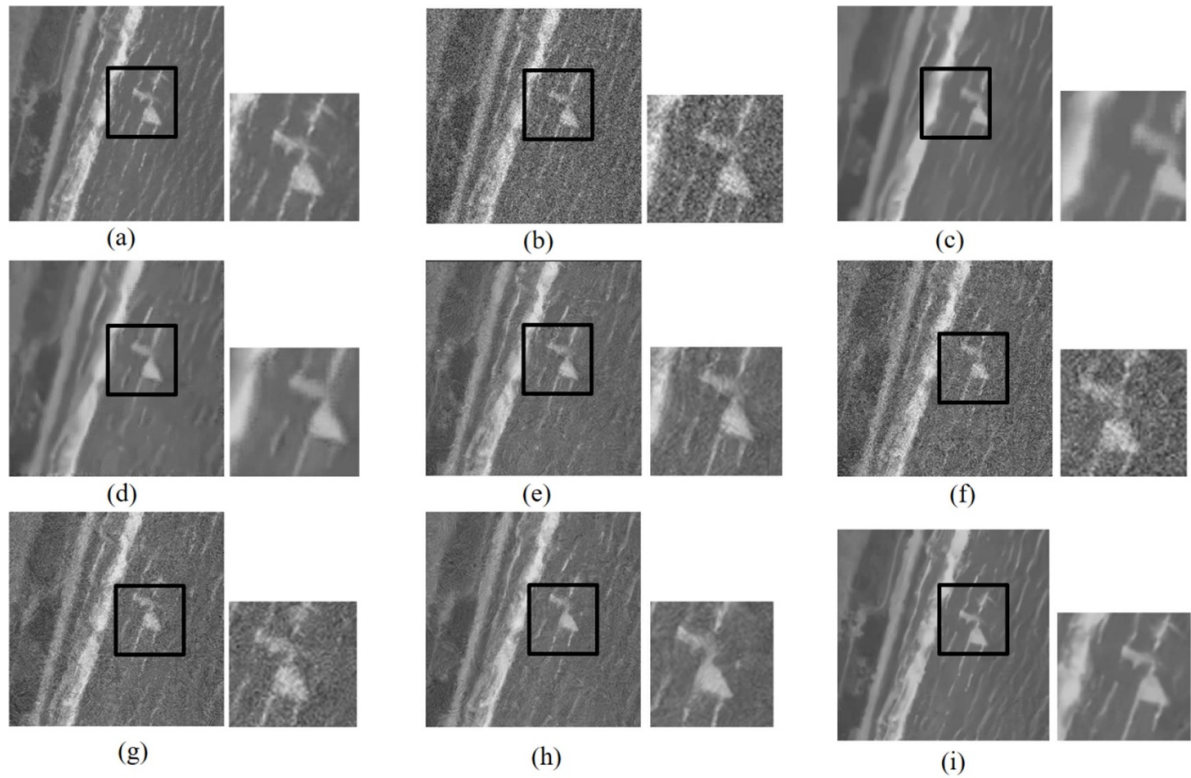
Figure 9. Coastal denoising renderings. (a) Original image; (b) noisy image; (c) image denoised with WNNM; (d) image denoised with SAR-BM3D; (e) image denoised with SAR-CNN; (f) image denoised with GAN; (g) image denoised with SAR2SAR; (h) image denoised with AGSDNet; (i) image denoised with MSAC-Net.
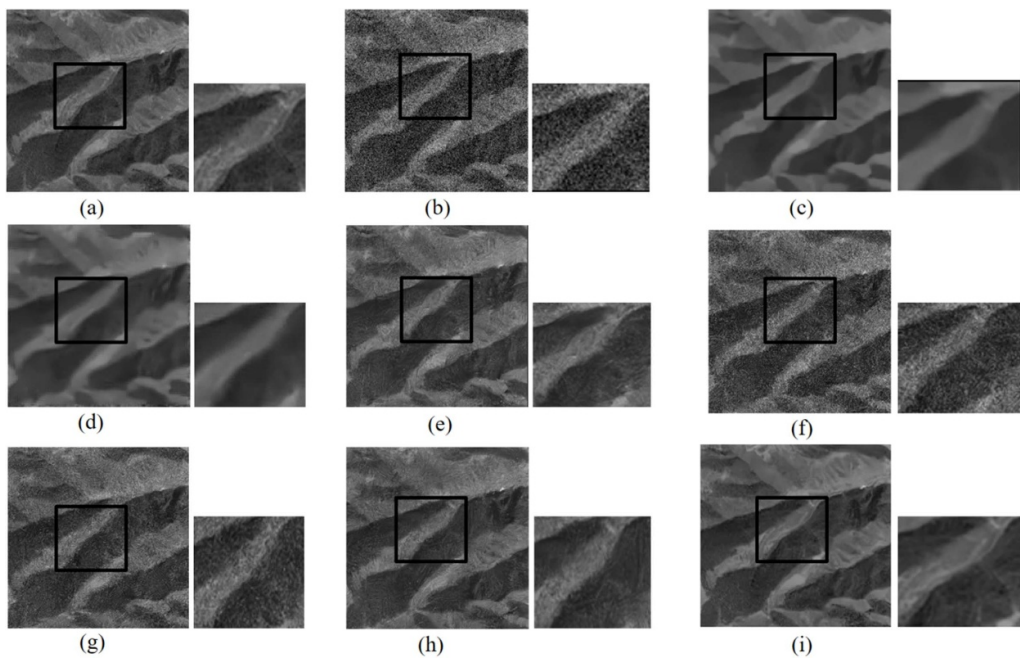
Figure 10. Mountain denoising renderings. (a) Original image; (b) noisy image; (c) image denoised with WNNM; (d) image denoised with SAR-BM3D; (e) image denoised with SAR-CNN; (f) image denoised with GAN; (g) image denoised with SAR2SAR; (h) image denoised with AGSDNET; (i) image denoised with MSAC-Net.

Table 4. The denoising level of each algorithm for various types SAR images under different noise levels (PSNR).

Image	σ	PSNR (dB)						
		WNNM	SAR-BM3D	SAR-CNN	GAN	SAR2SAR	AGSDNet	MSAC-Net
Ship	20	28.87	29.28	31.85	30.54	30.61	31.01	**32.31**
	30	26.62	27.13	30.14	29.36	28.17	30.47	**31.01**
	40	26.35	26.52	29.86	28.94	28.23	28.55	**29.97**
	45	25.28	25.93	27.77	27.62	27.92	**29.61**	28.11
	50	25.17	24.01	27.61	25.07	26.83	27.51	**27.62**
Mountain	20	28.15	29.13	31.10	29.34	29.57	29.56	**33.76**
	30	26.53	27.65	**30.03**	27.13	26.47	27.11	29.88
	40	25.72	26.04	28.53	25.60	25.39	25.41	**29.59**
	45	25.14	25.99	27.17	25.17	24.99	25.08	**28.58**
	50	24.05	25.92	**27.31**	24.64	24.37	23.61	26.21
Coastal	20	28.22	30.21	31.01	29.37	29.92	29.14	**31.83**
	30	26.63	28.52	29.43	27.13	26.47	27.66	**30.58**
	40	26.55	27.73	28.59	25.89	25.27	26.41	**29.98**
	45	25.39	26.24	26.87	25.79	24.99	25.53	**28.93**
	50	24.37	25.55	**25.02**	24.46	24.13	**25.08**	25.07

Table 5. The denoising level of each algorithm for various types SAR images under different noise levels (SSIM).

Image	σ	SSIM						
		WNNM	SAR-BM3D	SAR-CNN	GAN	SAR2SAR	AGSDNet	MSAC-Net
Ship	20	0.7929	0.8324	0.8923	0.8721	0.8814	0.9062	**0.9107**
	30	0.7538	0.7826	0.8621	0.8817	0.8637	0.8897	**0.8969**
	40	0.7063	0.7505	0.8669	0.8429	0.8553	**0.8815**	0.8763
	45	0.6818	0.6867	0.7926	0.8621	0.8432	**0.8762**	**0.8618**
	50	0.6821	0.6027	0.7818	0.8411	0.8291	**0.8438**	**0.7881**
Mountain	20	0.7867	0.8354	0.9012	0.8617	0.8803	0.9142	**0.9255**
	30	0.7128	0.7891	0.9010	0.8608	0.8637	0.8907	**0.8764**
	40	0.6957	0.7497	0.8725	0.8421	0.8553	0.8681	**0.8683**
	45	0.6708	0.7134	0.7971	0.8519	0.8432	**0.8583**	**0.8287**
	50	0.6684	0.6808	0.8211	0.8107	0.8291	**0.8432**	0.7878
Coastal	20	0.7911	0.8561	0.9036	0.8797	0.8704	0.9011	**0.9052**
	30	0.7187	0.7899	0.8947	**0.8703**	0.8612	0.8633	0.8623
	40	0.7064	0.7587	0.8809	0.8412	0.8551	**0.8815**	0.8524
	45	0.6817	0.7308	0.7437	0.8134	0.8432	**0.8762**	0.8194
	50	0.6512	0.6915	**0.7018**	0.7938	**0.8311**	0.8276	0.7547

with denoising parameter σ, SAR2SAR model with denoising parameter σ, AGSDNet model with denoising parameter σ and MSAC-Net in this paper. From table 4, it can be seen that the PSNR value of the proposed MSAC-Net is about 2.51 dB higher than that of SAR-BM3D, about 0.74 dB higher than that of SAR-CNN, about 3.36 dB higher than that of WNNM, about 2.49 dB higher than that of GAN, about 2.67 dB higher than that of SAR2SAR, about 2.11 dB higher than that of AGSDNet. Furthermore, except for the noise level in $\sigma = 50$ or $\sigma = 40$, MSAC-Net is 1.5 dB less than AGSDNet, and the PSNR values obtained by the rest of the model at each noise level are higher than the other six values compared. Especially when the noise parameter is 20, the method in this paper is 3.44 dB higher than the WNNM algorithm. In terms of structural similarity, it can be seen that the structural similarity

of MSAC-Net is mostly the highest value in the comparison method. Therefore, combining PSNR and SSIM, the proposed methods is superior to the state-of-the-art methods in denoising performance.

5. Conclusion

In this paper, we propose a new denoising model called MSAC-Net to solve the problem of inherent noise generated by SAR images. Our model uses an end-to-end architecture that does not require a separate subnet or manual intervention. The proposed method consists of three modules: the MAC module, the feature extraction module with attention mechanism, and the feature enhancement module with

dense cascade network (DCB). Our model achieves convergence without requiring a large amount of dataset, and reaches convergence for image data after 150 epochs of training. It exhibits outstanding training efficiency and good portability. Additionally, the denoising results show that compared with state-of-the-art algorithms, MSAC-Net can not only produce a better denoising effect but also has good robustness. The network achieves a good balance between controlling noise reduction and detail trade-off.

However, while the model achieves better denoising performance, it cannot be very lightweight, and powerful computational devices are required for practical applications. In this section, we propose potential measures to improve the model. Firstly, we can introduce sparse matrices, pruning, and other methods. Secondly, we can simplify the network structure by reducing the number of layers and convolutional kernel size as much as possible. Finally, we can utilize a series of network compression methods, such as low-rank decomposition, to make the model lightweight.

In conclusion, the proposed algorithm in this paper demonstrates that MSAC-Net can provide perceptually satisfying denoising results and outperforms other state-of-the-art algorithms in terms of PSNR and SSIM. The flexibility, efficiency, and effectiveness of MSAC-Net offer a new solution for SAR image denoising.

Data availability statement

All data that support the findings of this study are included within the article (and any supplementary files).

Author statement

I hereby declare that there are no conflicts of interest involved in the submission of this manuscript, and that all authors have granted their approval for its publication. Additionally, I would like to state on behalf of my co-authors that the work presented in this manuscript is entirely original research, and has not been previously published, nor is it currently being considered for publication, either in its entirety or in part, elsewhere. It is confirmed that all authors listed have reviewed and endorsed the enclosed manuscript.

Conflict of interest

The authors declare no conflict of interest.

ORCID iD

Xiangwei Fu https://orcid.org/0000-0002-3845-9242

References

[1] Bao Z, Sun C Y and Xing M D 2000 Principles and algorithms for inverse synthetic aperture radar imaging of manoeuvring targets *IEEE 2000 Int. Radar Conf.* (*Alexandria, VA, USA*) pp 316–21

[2] Najafi A, Hasanlou M and Akbari V 2017 Land cover changes detection in polarimetric SAR data using algebra, similarity and distance based methods *ISPRS-Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.* **42** 195–200

[3] Nair R R, David E and Rajagopal S 2019 A robust anisotropic diffusion filter with low arithmetic complexity for image *Eurasip J. Image Video Process.* **2019** 48

[4] Chen X X, Pu B X, Yu J, Wang R S and Zhong R F 2022 Research on building roof segmentation algorithm from multiple view stereo point clouds *Geo-Spatial Inf.* **20** 33–37

[5] Donoho D L 1995 De-noising by soft-thresholding *IEEE Trans. Inf. Theory* **41** 613–27

[6] Donoho D L and Johnstone I M 1995 Adapting to unknown smoothness via wavelet shrinkage *J. Am. Stat. Assoc.* **90** 1200–24

[7] Buades A, Coll B and Morel J 2005 A non-local algorithm for image denoising *2005 IEEE Computer Society Conf. on Computer Vision and Pattern Recognition* (*CVPR'05*) vol 2 pp 60–65

[8] Parrilli S, Poderico M, Angelino C V and Verdoliva L 2012 A nonlocal SAR image denoising algorithm based on LLMMSE wavelet shrinkage *IEEE Trans. Geosci. Remote Sens.* **50** 606–16

[9] Murali Mohan Babu Y, Subramanyam M V and Giri Prasad. M N 2015 A modified BM3D algorithm for SAR image despeckling *Proc. Comput. Sci.* **70** 69–75

[10] Foucher S 2008 SAR image filtering via learned dictionaries and sparse representations *Proc. IEEE Int. Geoscience and Remote Sensing Symp.* (Boston, MA: IEEE Press) pp 229–32

[11] Rudin L I, Osher S and Fatemi E 1992 Nonlinear total variation based noise removal algorithms *Physica* D **60** 259–68

[12] Ding H B, Wang Z Z and Liu D 2021 Comparison of denoise methods of LiDAR signal *Acta Opt. Sin.* **41** 9–18

[13] Liu J, Huang X H, Wu J H, Su N G and Yu L X 2020 Adaptive total variation speckle denoising method based on sine-cosine decomposition *China J. Lasers* **47** 172–9

[14] Li S, Fu B, Wei J, Lv Y, Wang Q and Tu J 2021 Ultrasonic logging image denoising based on CNN and feature attention *IEEE Access* **9** 116845–56

[15] Gu S, Zhang L, Zuo W and Feng X 2014 Weighted nuclear norm minimization with application to image denoising *Proc. IEEE Conf. on Computer Vision and Pattern Recognition* (*Columbus, OH, USA*) pp 2862–9

[16] Singh A, Raj K, Kumar T, Verma S and Roy A M 2023 Deep learning-based cost-effective and responsive robot for autism treatment *Drones* **7** 81

[17] Lobbia A, Hankonen N, Presseau J, Rodrigues A and Sniehotta F F 2018 The efficacy of cognitive stimulation therapy (CST) for people with mild-to-moderate dementia *Eur. Psychol.* **24** 7–25

[18] Wu C, Li S, Liu X, Jiang F and Shi B 2022 DMs-MAFM+ EfficientNet: a hybrid model for predicting dysthyroid optic neuropathy *Med. Biol. Eng. Comput.* **60** 3217–30

[19] Khan W, Raj K, Kumar T, Roy A M and Luo B 2022 Introducing Urdu digits dataset with demonstration of an efficient and robust noisy decoder-based pseudo example generator *Symmetry* **14** 1976

[20] Singh A *et al* 2023 Understanding EEG signals for subject-wise definition of Armoni activities (arXiv:2301.00948)

[21] Jin Z, Hu G and Zhao K 2022 Mannose-anchored quaternized chitosan/thiolated carboxymethyl chitosan composite NPs as mucoadhesive carrier for drug delivery *Carbohydrate Polym.* **283** 119174

[22] Zhang K, Zuo W, Chen Y, Meng D and Zhang L 2017 Beyond a Gaussian denoiser: residual learning of deep CNN for image denoising *IEEE Trans. Image Process.* **26** 3142–55

[23] Yue Z, Zhao Q, Zhang L and Meng D 2020 Dual adversarial network: toward real-world noise removal and noise generation *Computer Vision—ECCV* ed V Andrea, H Bischof, T Brox and J-M Frahm (Cham: Springer) pp 41–58

[24] Zhang K, Zuo W and Zhang L 2018 FFDNet: toward a fast and flexible solution for CNN-based image denoising *IEEE Trans. Image Process.* **27** 4608–22

[25] Chierchia G, Cozzolino D, Poggi G and Verdoliva L 2017 SAR image despeckling through convolutional neural networks *2017 IEEE Int. Geoscience and Remote Sensing Symp. (IGARSS)* (*Fort Worth, TX, USA*) pp 5438–41

[26] Liu R, Li Y and Jiao L 2020 SAR image specle reduction based on a generative adversarial network *2020 Int. Joint Conf. on Neural Networks (IJCNN)* (*Glasgow, UK*) pp 1–6

[27] Dalsasso E, Denis L and Tupin F 2021 SAR2SAR: a semi-supervised despeckling algorithm for SAR images *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **14** 4321–9

[28] Thakur R K and Maji S K 2022 AGSDNet: attention and gradient-based SAR denoising network *IEEE Geosci. Remote Sens. Lett.* **19** 4506805

[29] Tian C, Xu Y, Zuo W, Lin C-W and Zhang D 2022 Asymmetric CNN for image superresolution *IEEE Trans. Syst. Man Cybern.: Syst.* **52** 3718–30

[30] Jin J, Dundar A and Culurciello E 2014 Flattened convolutional neural networks for feedforward acceleration (arXiv:1412.5474)

[31] Denton M, Zaremba W, Bruna J, LeCun Y and Fergus R 2014 Exploiting linear structure within convolutional networks for efficient evaluation (arXiv:1404.0736)

[32] Jaderberg M, Vedaldi A and Zisserman A 2014 Speeding up convolutional neural networks with low rank expansions (arXiv:1405.3866)

[33] Hu J, Shen L and Sun G 2018 Squeeze-and-excitation networks *2018 IEEE/CVF Conf. on Computer Vision and Pattern Recognition* pp 7132–41

[34] Woo S, Park J, Lee J-Y and Kweon I-S 2018 CBAM: convolutional block attention module (arXiv:1807.06521)

[35] Huang G, Liu Z, Van Der Maaten L and Weinberger K Q 2017 Densely connected convolutional networks *2017 IEEE Conf. on Computer Vision and Pattern Recognition* pp 2261–9

[36] Cheng G, Han J and Lu X 2017 Remote sensing image scene classification: benchmark and state of the art *Proc. IEEE* 1865–83

[37] Kingma D P and Ba J 2015 Adam: a method for stochastic optimization (arXiv:1412.6980)