Application of improved CNN in SAR image noise reduction

Li Lu1,a, Ganchun Zhang2,b, Ying Nie3,c, Jiayi Liu4,d, Yibiao Fang5,e, Guoling Zhang6,f, Yahui Wu7,g

1Weapon Science and Technology, Air Force Engineering University, Xian Shanxi
2Military Tactics, Air Force Engineering University, Xian Shanxi
392861 Army, Air Force Engineering University, Sanyuan Shanxi
4Information and Communication Engineering, Air Force Engineering University, Xian Shanxi
5Cyberspace Security, Air Force Engineering University, Xian Shanxi
6Cyberspace Security, Air Force Engineering University, Xian Shanxi
7Vehicle Engineering, Air Force Early Warning Academy, Whhan hubei

a764603926@qq.com, b985529465@qq.com, c1290738043@qq.com, d sixandone1@163.com, e fangyibiao@foxmail.com, 
 f823721228@qq.com, g178990092@qq.com

ABSTRACT: Aiming at the problem of coherent speckle noise in SAR image processing, this paper improves the network based on the convolutional neural network method, adjusts the hierarchical relationship in the network hierarchy, and introduces the principles of batch normalization and residual network to eliminate the gradient disappearance when the network level is too deep. The loss function is improved by using multiple convolution kernels of different sizes which are used to extract a variety of different levels of feature information. Finally, based on the peak signal-to-noise ratio (PSNR) and edge index two parameters, the experimental comparison with PPB, SAR-BM3D and NL-SAR three types of denoising methods, The experimental results show that the SAR image generated by the improved CNN network can better achieve despeckle and denoising, and the edges of the image are kept intact.

1. Introduction

Synthetic Aperture Radar (SAR) can produce high-resolution images of terrain targets all-weather and all-day. It has excellent mobility at night and can overcome the limitations of optical and infrared systems. The many excellent characteristics of SAR images make it widely used in military and civilian fields. However, in the process of processing SAR image data, the SAR image is often affected by the coherent speckle noise carried by itself [1]. These coherent speckles are generated by the random interference of coherent echoes from a large number of scattered points, making SAR images suitable for ground objects. The inaccurate reflection of the scattering characteristics of the image seriously affects the image quality and reduces the ability to extract target information in the image. Therefore, the suppression of coherent speckle noise is a prerequisite for processing SAR images and is a key difficulty for in-depth research on SAR images.
In recent years, many researchers have tried to remove coherent speckle noise from the image and save more image information at the same time. Literature [2] proposed multi-view averaging processing filtering, using SAR images on different bands for averaging operation, so that coherent speckle noise can be smoothly filtered out, but this method is not clean; literature [3-5] The spatial domain filtering method is adopted, but the processed image becomes blurred and the edge information details are missing; the literature [6-9] uses wavelet theory to decompose the SAR image into a combination of multiple scales, and reorganize the wavelet part containing noise However, there is no suitable theoretical value for the selection of the wavelet scale, and it is easy to delete the picture information by mistake. Literature [10] uses the SAR-BM3D calculation method, but the algorithm is complex and the calculation is too large. These traditional denoising methods have certain advantages, but it is difficult to efficiently obtain high-quality SAR images.

In recent years, extensive research has shown that deep learning technology can learn deep-level information, and it can automatically learn image features, which reduces a lot of tedious work compared to traditional methods. Research scholars have applied deep learning to image denoising and have achieved good results. Literature [11] uses neural network denoising, and through layer-by-layer training, it achieves faster and more accurate convergence; literature [12] verifies sufficient training data can make the denoising effect of the multilayer perceptron reach the best. For the SAR images in this article, literature [13-15] also provides a new idea for SAR image denoising. For example, literature [15] combines image denoising and image coloring in the network structure to obtain high-quality SAR image.

Aiming at the denoising of SAR images, this paper combines the idea of CNN network and residual learning to build a new target loss function, introduces a smoother Mish activation function, and builds a denoising model that meets the characteristics of SAR images. In the process of network training, convolutional separation is performed to obtain noisy sub-images; residual learning and batch normalization theory are introduced to avoid the disappearance of gradients when the network level is too deep, and to achieve faster convergence; in feature extraction, convolution kernels of different sizes are used to extract features, and feature information of multiple scales is obtained. Experimental comparison shows that the CNN network designed in this paper can remove the coherent speckle noise of SAR images more effectively, and the edge information of the image can be kept intact.

2. Related knowledge

2.1. Residual network

In the CNN training process, as the network level deepens, deeper details can be obtained. However, too deep network levels are prone to gradient disappearance, which leads to network performance degradation.

Residual learning is an important method used to solve performance degradation and gradient disappearance. Its principle structure is shown in Figure 1 below.

To put it simply, the two-layer network is cross-connected every certain number of layers. This connection is equivalent to performing the same mapping, without generating additional parameters, and without increasing computational complexity. The entire network can still be trained through end-to-end backpropagation. Its network model formula is derived as follows, and the formula of a residual block is expressed as:

\[ x_{l+1} = x_l + F(x_l, W_l) \]  \hspace{1cm} (1)

After recursive calculation, the characteristics of the deep network can be obtained as:

\[ x_L = x_0 + \sum_{i=1}^{L-1} F(x_i, W_i) \]  \hspace{1cm} (2)

Equation (2) expresses that the network structure in the deep layer can be composed of the sum of the shallow layer features and the residual function as \( \sum_{i=1}^{L-1} F(x_i, W_i) \), indicating that the shallow layer
and deep layer features are interrelated. Then in the process of backpropagation, assuming the loss function is $E$, then the chain rule of backpropagation is:

$$\frac{\partial E}{\partial x_l} = \frac{\partial E}{\partial x_L} \cdot \frac{\partial x_l}{\partial x_L} = \frac{\partial E}{\partial x_L} \cdot \left[1 + \sum_{i=1}^{L-1} \frac{\partial}{\partial x_L} F(x_i, w_i)\right]$$

This formula can ensure that the gradient will not disappear and will not be close to 0, because no matter how many terms are multiplied, the value will not be -1. Using the structure of the residual network can simplify the learning process, enhance gradient propagation, and avoid the phenomenon of gradient disappearance. At the same time, the generalization ability of the network is greatly enhanced, breaking the asymmetry of the network [16].

### 2.2 Batch normalization

Batch Normalization (BN), in simple terms, is to add a normalization layer to normalize the input of the network [17].

BN has two advantages. One is that it does not need to perform complex settings on various parameters (such as learning rate, weight attenuation coefficient, drop out ratio, etc.) like the stochastic gradient descent method; the other is that it can solve the middle layer data in the training process. The situation where the distribution has changed [18].

The BN method can be carried out as follows. First, it can be assumed that there is an $m$-dimensional input, as shown in the following formula (4)

$$x = (x^{(1)}, \cdots, x^{(m)})$$

The hidden layer $l$ in the first layer is $z^{(1)}$, and the calculation of its mean and variance are:

$$\mu = \frac{1}{m} \sum_{i=1}^{m} z^{(i)}$$

$$\sigma^2 = \frac{1}{m} \sum_{i=1}^{m} (z^{(i)} - \mu)^2$$

Then normalize each dimension, as in formula (7)

$$z^{(i)} = \frac{z^{(i)} - \mu}{\sqrt{\sigma^2 + \epsilon}}$$

$\epsilon$ as a constant, the purpose is to prevent the variance from being zero. However, in order to adapt to the requirements of different input changes and different mean variances, the learning reconstruction parameters $\alpha$ and $\beta$ are introduced here, as shown in equation (8), so that they can be normalized to any size of mean or variance.

$$z^{(i)} = \alpha z^{(i)}_{\text{norm}} + \beta$$

---

**Figure 1** Principle diagram of residual learning

![Principle diagram of residual learning](image)
On the whole, the function of the BN layer is to control the mean and standard deviation of each layer's output through parameters, and can choose any normalization size. It can reduce the connection between the front and rear layers, enhance the ability of each layer to learn independently, and ultimately improve the performance of the entire network.

2.3. Generation of coherent speckle noise

Synthetic Aperture Radar is a coherent imaging system. When emitting coherent electromagnetic waves, if it encounters a rough surface, the reflected signal will exist in multiple scatterers inside the resolution unit, and their distance and position from the sensor are different. Therefore, when receiving the echo signal, it can be found that the signal is coherent in frequency, but falls on a different phase. When the echo phases are the same, it is a strong signal, otherwise it is a weak signal. Therefore, the scattering echo coefficient will have large random fluctuations, resulting in many grain-like spots on the image, which is the cause of coherent speckle noise [19].

The noise generated by this signal superposition is called multiplicative noise, and its model is (9)

\[ f(x, y) = \mu(x, y) \ast \eta(x, y), (x, y) \in \Omega \]

where \( f \) is a noisy image, \( x, y \) is the horizontal and vertical components, and \( \mu(x, y) \) is a point in the image, where \( \mu \in \Omega \), \( \eta \) is the multiplicative noise to be removed, and it obeys the Gamma distribution (10)

\[ P(\eta) = \frac{L^\eta \Gamma(L) \eta^{L-1}}{\Gamma(L)} e^{-\frac{L}{\eta}} \]

Where \( \Gamma(\cdot) \) represents the Gamma function, and, \( \eta \geq 0 \), \( L \geq 1 \).

2.4. Denoising evaluation index

Regarding the degree of denoising effect, the main considerations are the following three quantitative indicators:

(1) Evaluation of polarization information retention
(2) Evaluation of coherent speckle suppression
(3) Edge retention evaluation

Based on two quantitative considerations, this paper selects peak signal-to-noise ratio (PSNR) and edge index to evaluate the denoising performance.

(1) PSNR

PSNR reflects the degree of distortion of the image, and its expression is equation (11)

\[ PSNR = 10 \log \left\{ \frac{\text{MAX}_i^2}{\frac{1}{M \times N} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} [R(i, j) - I(i, j)]^2} \right\} \]

Among them, \( I \) and \( R \) represent the image before and after denoising, the aspect ratio is
$M \times N$, $\text{MAX}_i^j$ represents the maximum possible pixel value of the image, usually 255. This parameter reflects the loss degree of the image after denoising. The larger the value, the more complete the image information is retained and the better the denoising effect.

(2) Edge index

The edge $\gamma$ index can well reflect whether the edge of the image is intact. This is an important evaluation index for SAR image denoising. Its expression is as follows:

$$
\gamma = \frac{\Gamma(\Delta S - \overline{\Delta S}, \Delta L - \overline{\Delta L})}{\sqrt{\Gamma(\Delta S - \Delta S, \Delta S - \Delta S)\Gamma(\Delta L - \overline{\Delta L}, \overline{\Delta L} - \Delta L)}}
$$

(12)

Where $\gamma$ represents the edge index value, where, $\Delta S$, $\Delta L$ are the order $i \times j$ matrix, $\overline{\Delta S}$ and $\overline{\Delta L}$ are the mean values before and after denoising, in addition:

$$
\Gamma(X,Y) = \sum_{i=1}^{n} \sum_{j=1}^{n} X(i,j)Y(i,j)
$$

(13)

$X,Y$ is the $i \times j$ order matrix, $\Gamma(X,Y)$ is the result of the summation. The $\gamma$ closer to 1, the effect of edge retention is the best.

3. Network model

This paper constructs a network model based on the synthetic aperture radar image denoising method of convolutional neural network, which can realize end-to-end denoising training, and obtain noise-free SAR image directly from the noise image denoising. Only the convolutional layer is used in the network. At the same time, the Mish activation function is used in the network. Compared with the ReLU activation function, its curve is smoother, which can penetrate more information into the neural network to improve accuracy and generalization ability, The graph is shown in figure 3. Next, introduce the designed network.

![Figure 3 Mish and ReLU activation functions](image_url)

The structure level of the network model is shown below. There are a total of 12 layers of convolutional networks. The network structure is shown in Figure 2 below, and the information of each layer is shown in Table 1. The first layer is the convolutional layer and the Mish activation function, using 64 convolution $1 \times 1$ kernels; the second to the 11th layers include convolutional layers, batch normalization and Mish activation functions, $32 \ 1 \times 1$ and $32 \ 3 \times 3$; in the 12th The layer is a convolutional layer, using $64 \ 3 \times 3$ convolution kernels, as shown in Table 1. At the same time, a cross connection is made every two layers, so that a more efficient network structure can be obtained with a smaller number of channels and a lower computational cost. Using different sizes of convolution kernels can obtain features of different scales, increasing the width of the network, and the acquired features are often better than a single convolution kernel, and using a small convolution kernel first can reduce the dimensionality and reduce the amount of calculation, as shown in the following figure 4 shown.
The network denoising process is as follows: Build a convolutional neural network based on parameter-learnable activation functions, use the ordinary convolutional layer plus Mish activation function in the first layer, and connect with 2 to 11 layers of multiple convolutional layers, batch normalization and Mish activation function in series. The last layer uses only the convolutional layer for final output. Perform batch normalization and jump connection in the network, and perform appropriate zero padding to ensure that the output of each layer shares the same image input size. In the convolutional network of 2 to 11 layers, the convolution kernel is used to reduce the dimensionality, and then the convolution kernel is used for feature extraction. Finally, the extracted noisy sub-image is divided by the input image to obtain the desired denoised image. In the loss function (14) of the network, according to the model and training parameters, the loss functions of mean squared error (MSE), total variation TV (Total Variation, TV) minimization and edge detection are used, and use RAdam optimizer to optimize gradient descent. The MSE loss function can predict the difference between the output image and the original image; the TV loss function is equivalent to a regularization term to maintain the smoothness of the image and avoid amplifying noise during the restoration algorithm; the edge loss function is used to judge the image The edges after denoising remain intact.

$$L = L_{MSE} + \lambda_{TV} \cdot L_{TV} + L_{Edge} \quad (14)$$

$$L_{TV} = \sum_{i,j} \sqrt{(x_{i,j-1} - x_{i,j})^2 + (x_{i+1,j} - x_{i,j})^2} \quad (15)$$

$$L_{Edge} = \left(\frac{\partial X}{\partial \mu} - \frac{\partial \hat{X}}{\partial \mu}\right)^2 + \left(\frac{\partial X}{\partial \nu} - \frac{\partial \hat{X}}{\partial \nu}\right)^2 \quad (16)$$

$$L_{MSE} = \frac{1}{n} \sum_{i=1}^{n} \left[ R(\Delta) - (x - y) \right]^2 \quad (17)$$

Where $x$ is the denoised image, $n$ is the number of samples, $y$ is the noisy image, $R(\Delta)$ is the learned noisy sub-image, $X$ and $\hat{X}$ is the noise-free image with horizontal and vertical gradients, and $Edge$ is the image gradient along the horizontal and vertical directions A measure of the gradient difference in direction. $\lambda_{TV} = 2 \times 10^{-4}$ is the initial weight of the TV loss function, $i,j$ is the image row and column number, respectively, $x_{i,j-1}$ and represents the element of the $j-1$ row. The loss function can learn and train the network model in an end-to-end manner, and the use of TV loss can make the image smoother and make the edges of the image clearer.
4. Experiment and analysis

4.1. Experimental parameters
In this experiment, the experimental data adopts the spotlight MSTAR data set of ground military vehicles, and the pixel size is 128×128. There are a total of 6761 images in jpg format. The image set has 4 types of ground objects (BRDM-2, SLICY, ZSU_23_4, 2S1), including 4 types of objects with different elevation angles of 15 degrees and 17 degrees. The number of each data set is shown in Table 2 below. The loss function of the training process uses the Radam (Rectified Adam) optimizer [20], which is robust to different learning rates, while still able to converge quickly and obtain higher accuracy. The learning rate is set to 0.0002, and the software system Using the tensorflow experimental framework in deep learning, based on the GPU1080TI hardware environment. The experimental results are compared with the three methods using PPB [21], SAR-BM3D [22], and NL-SAR [23].

| data set  | BRDM-2 | SLICY | ZSU_23_4 | 2S1 |
|-----------|--------|-------|----------|-----|
| Quantity  | 1415   | 2539  | 1401     | 1401|

4.2. Experimental results
In order to verify the denoising ability of the above four methods, the selected evaluation indicators are peak signal-to-noise ratio and edge index respectively. The convergence curve of the loss function is shown in Figure 5. It can be seen that after 25 rounds, it is close to convergence. The denoising data with the other three methods are shown in Tables 3 and 4.

Figure 6 is a denoising comparison diagram of the four types of selected 2S1 slice images. From left to right, they are the original image, PPB, SAR-BM3D, NL-SAR, and the improved CNN method. It can be seen from the above two tables that the PPB, SAR-BM3D and NL-SAR methods have good performance. However, in terms of the overall results, the edge index value of the improved CNN method in this paper is close to 1, and the edge is preserved more complete; higher PSNR value, better visual effect after denoising. From the denoising image generated, the coherent speckles in the image generated by the improved CNN are obviously reduced, and the completeness, smoothness and clarity are improved compared with other methods, which proves the feasibility of the method in this paper.

![Figure 5](image)

![Figure 6](image)

| Types    | PPB  | SAR-BM3D | NL-SAR | CNN  |
|----------|------|----------|--------|------|
| BRDM-2   | 21.07| 23.15    | 22.77  | 25.37|
| SLICY    | 24.03| 25.12    | 25.53  | 27.41|
Table 4 Values of the four methods

| Types     | PPB  | SAR-BM3D | NL-SAR | CNN  |
|-----------|------|----------|--------|------|
| BRDM-2    | 0.328| 0.412    | 0.378  | 0.556|
| SLICY     | 0.601| 0.564    | 0.598  | 0.669|
| ZSU_23_4  | 0.423| 0.492    | 0.376  | 0.514|
| 2S12      | 0.352| 0.421    | 0.501  | 0.529|

5. Conclusion

This article is based on the improvement of the convolutional neural network, making it a new type of convolutional neural network suitable for SAR image denoising. Considering that the gradient disappears when the network level is too deep, the BN layer and residual network are added between each network level, so that the training can proceed smoothly. Based on the general CNN denoising method, this paper uses a variety of different sizes of convolution kernels for feature extraction to obtain multi-scale feature information, which effectively reduces the training time. After comparing the three types of denoising methods, PPB, SAR-BM3D and NL-SAR, we can see the advantages of improved CNN in SAR image denoising.

REFERENCE

[1] F. Argenti, A. Lapini, L. Alparone, and T. Bianchi, “A tutorial on speckle reduction in synthetic aperture radar images,” IEEE Geosci. Remote Sens. Mag., vol. 1, pp. 6–35, 2013.
[2] Ni Jie. Research on SAR image filtering algorithm and new evaluation model [D]. Xidian University, 2018.
[3] J.-S. Lee, “Speckle analysis and smoothing of synthetic aperture radar images,” Computer graphics and image processing, vol. 17, no. 1, pp.24–32, 1981.
[4] Ma X, Shen H, Zhao X, Zhang L. SAR Image Despeckling by the Use of Variational Methods With Adaptive Nonlocal Functionals[J].IEEE Transactions on Geoscience and Remote Sensing,2016:1-15.
[5] Mahdavi S, Salehi B, Moloney C, et al. Speckle Filtering of Synthetic Aperture Radar images Using Filters with Object-Size-Adapted Windows[J]. International Journal of Digital Earth,2017.
[6] Mastriani M. New wavelet-based superresolution algorithm for speckle reduction in SAR images[J]. 2016.
[7] MuraliMohanBabu. Y,Subramanyam. M.V,GiriPrasad. M.N. A Modified BM3D Algorithm for SAR Image Despeckling[J]. Procedia Computer Science,2015,70.
[8] Isin Erer,Nur Huseyin Kaplan. Fast local SAR image despeckling by edge-avoiding wavelets[J]. Signal, Image and Video Processing,2019,13(6).
[9] Liu Zihao, Miao Xingang, Tang Boyan. Based on wavelet transform SAR image speckle noise suppression algorithm[J]. Robotics and Applications, 2019(04): 43-45.
[10] Lu Zili, Jia Xin, Zeng Chuangzhan. Fast SAR-BM3D speckle suppression algorithm based on feature extraction [J]. Electronic Measurement Technology, 2018, 41(01): 123-128.
[11] V. Jain, H. Seung. Natural image denoising with convolutional networks [C]. Advances in Neural Information Processing Systems 21-Proceedings of the 2008 Conference, 2009: 769–776.

[12] H.C. Burger et al. Image denoising: can plain neural networks compete with BM3D? [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2012: 2392–2399.

[13] P. Wang, H. Zhang, and V. M. Patel, “Sar image despeckling using a convolutional neural network,” IEEE Signal Processing Letters, vol. 24, no. 12, pp. 1763–1767, Dec 2017.

[14] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, “Beyond a Gaussian Denoiser: Residual Learning of Deep CNN for Image Denoising,” arXiv:1608.03981v1, 2016.

[15] Wang P , Patel V M . Generating High Quality Visible Images from SAR Images Using CNNs[J]. 2018.

[16] He K , Zhang X , Ren S , et al. Deep Residual Learning for Image Recognition[C]// 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE Computer Society, 2016.

[17] S. Ioffe and C. Szegedy, “Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift,” arXiv:1502.03167v3, 2015.

[18] Ioffe S , Szegedy C . Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift[J]. 2015.

[19] Pi Yiming, Yang Jianyu, Fu Yusheng, et al. Principle of synthetic aperture radar imaging [M]. 2007.

[20] Liu L , Jiang H , He P , et al. On the Variance of the Adaptive Learning Rate and Beyond[J]. 2019.

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

[22] 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.

[23] C.-A. Deledalle, L. Denis, F. Tupin, A. Reigber, and M. Jäger, “NL-SAR: A unified nonlocal framework for resolution-preserving (Pol)(In)SAR denoising,” IEEE Trans. Geosci. Remote Sens., vol. 53, no. 4, pp. 2021–2038, 2015.