*remote sensing*

MDPI

*Article*

# Autoencoder Neural Network-Based STAP Algorithm for Airborne Radar with Inadequate Training Samples

**Jing Liu [1,*], Guisheng Liao [1], Jingwei Xu [1], Shengqi Zhu [1], Filbert H. Juwono [2] and Cao Zeng [1]**

[1] National Laboratory of Radar Signal Processing, Xidian University, Xi'an 710071, China
[2] Computer Science Program, University of Southampton Malaysia, Nusajaya 79100, Malaysia
* Correspondence: ngleliujing@xidian.edu.cn

**Abstract:** Clutter suppression is a key problem for airborne radar, and space-time adaptive processing (STAP) is a core technology for clutter suppression and moving target detection. However, in practical applications, the non-uniform time-varying environments including clutter range dependence for non-side-looking radar lead to the training samples being unable to satisfy the sample requirements of STAP that they should be independent identical distributed (IID) and that their number should be greater than twice the system's degree of freedom (DOF). The lack of sufficient IID training samples causes difficulty in the convergence of STAP and further results in a serious degeneration of performance. To overcome this problem, this paper proposes a novel autoencoder neural network for clutter suppression with a unique matrix designed to be decoded and encoded. The main challenges are improving the accuracy of the estimation of the clutter-plus-noise covariance matrix (CNCM) for STAP convergence, designing the form of the data input to the network, and making the network successfully explored to the improvement of CNCM. For these challenges, the main proposed solutions include designing a unique matrix with a certain dimension and a series of covariance data selections and matrix transformations. Consequently, the proposed method compresses and retains the characteristics of the covariances, and abandons the deviations caused by the non-uniformity and the deficiency of training samples. Specifically, the proposed method firstly develops a unique matrix whose dimension is less than half of the DOF, meanwhile, it is based on a processing of the selected clutter-plus-noise covariances. Then, an autoencoder neural network with $l_2$ regularization and the sparsity regularization is proposed for the unique matrix to be decoded and encoded. The training of the proposed autoencoder can be achieved by reducing the total loss function with the gradient descent iterations. Finally, an inverted processing for the autoencoder output is designed for the reconstruct ion of the clutter-plus-noise covariances. Simulation results are used to verify the effectiveness and advantages of the proposed method. It performs obviously superior clutter suppression for both side-looking and non-side-looking radars with strong clutter, and can deal with the insufficient and the non-uniform training samples. For these conditions, the proposed method provides the relatively narrowest and deepest IF notch. Furthermore, on average it improves the improvement factor (IF) by 10 dB more than the ADC, DW, JDL, and original STAP methods.

**Keywords:** clutter suppression; space-time adaptive processing; autoencoder neural network; inadequate training samples

check for updates

**Citation:** Liu, J.; Liao, G.; Xu, J.; Zhu, S.; Juwono, F.H.; Zeng, C. Autoencoder Neural Network-Based STAP Algorithm for Airborne Radar with Inadequate Training Samples. *Remote Sens.* **2022**, *14*, 6021. https://doi.org/10.3390/rs14236021

Academic Editor: Benoit Vozel

Received: 14 October 2022
Accepted: 23 November 2022
Published: 28 November 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

## 1. Introduction

Airborne radar has huge advantages over ground-based radar in far sight and flexibility. When looking down, radar on a moving platform suffers from the core problem of clutter suppression because the ground clutter is strong and its Doppler spectrum is seriously widened due to the platform movement, which further causes the Doppler spectrum of the target to be drowned in the clutter spectrum [1–4]. For this, the space-time adaptive processing (STAP) of airborne radar has become the core technology for clutter suppression and target detection. With the space-time two-dimensional spectral domain, it obtains a better performance than the traditional one-dimensional processing [4–6].

However, estimating the clutter-plus-noise covariance matrix (CNCM) accurately is essential for STAP, which requires that the number of the independent and identically distributed (IID) samples should be greater than twice the system's degrees of freedom (DOF) to achieve an output signal-to-clutter-plus-noise ratio (SCNR) loss within 3 dB [7]. Unfortunately, due to the array configurations and the complex clutter environment, sufficient training samples are difficult to obtain in non-uniform time-varying environments [8,9]. In particular, as a phenomenon of clutter non-uniformities, the clutter range dependence occurs when the configuration of the array antenna is non-side-looking, which directly results in the training samples failing to meet the IID condition [10]. Therefore, in applications, the difficulty in obtaining enough IID training samples makes the performance of STAP degrade significantly.

Reducing the dimension or rank is often used to solve the problems mentioned above, by decreasing the system's DOF and selecting parts of the training sample data to estimate the CNCM of the cell to be detected [11–14]. For example, in STAP dimension reduction methods with fixed structure, the auxiliary channel processing (ACP) method [15] has been proposed by means of setting the two-dimensional beam near the clutter ridge as the auxiliary beam. It reduces the DOF from MN to M+N-1. Utilizing the two-dimensional Fourier transform and simultaneously dealing with beamforming and Doppler frequency processing, a joint domain localized (JDL) [16] method that selects a rectangle joint processing local area is proposed. It reduces the dimension of adaptive processing to the product of length and width of the selected processing local area. Besides, for limited IID training samples, STAP based on the sparse representation (SR) theory called SR-STAP are proposed [17–19]. In addition, sparse Bayesian learning (SBL) [20]-based STAP has also received attention because of its robustness. However, their performance relies on the match between the sparse representation of the clutter and the setting of the dictionary matrix; meanwhile, at present most of STAP methods for clutter suppression are based on the side-looking array whose samples are theoretically IID.

Although the above-mentioned methods reduce the consequences of the insufficiency and non-uniformity of the training samples, STAP with low degrees of freedom or reduced samples leads to performance degradation.

For the clutter non-uniformity caused by the clutter range dependence, it can also be solved by range distance compensation and scale transition [21–24]. Clutter range compensation mainly compensates the clutter spectra of adjacent training cells so that the space-time distribution characteristics of the clutter after compensation are as consistent as possible to those to be detected. Thus, the performance of the clutter suppression of the follow-up STAP processing can be improved. For example, Borsari proposed the Doppler warping (DW) algorithm [25,26] based on the non-side-looking array. The Doppler frequencies of the range cells to be detected and compensated are obtained by using the inertial navigation parameters of the radar system. The two-dimensional clutter spectra of the range cells to be compensated for are shifted into one dimension so that in the Doppler frequency direction, they coincide with the clutter spectrum of the range cell to be detected. Then, the correlation of the range distance of the clutter is eliminated. Furthermore, Klemm proposes an angle-Doppler compensation (ADC) method [27] which obtains the Doppler frequency of each range cell through adaptive estimation of the echo data, rather than through the inertial navigation parameters of the system. Additionally, Pearson proposed the high-order Doppler warping (HODW) method [28], which compensates the adjacent range cells from multiple spatial angle directions. However, the existing methods are hardly able to take both the insufficiency of training samples and the non-uniformity, such as the clutter range dependence caused by the non-side-looking array, into consideration at the same time.

In this paper, we propose an autoencoder neural network-based STAP algorithm for airborne radar with inadequate training samples. The aim of this paper is to overcome the lack of sufficient IID training samples, and even to overcome the non-uniform samples that cannon obey the IID condition. The novel method firstly based on a processing of the selected clutter-plus-noise covariances designs a unique matrix whose dimension is

less than half of the DOF. Next, an autoencoder neural network with $l_2$ regularization and sparsity regularization is proposed for the unique matrix to be decoded and encoded. Then, the training of neural network of the proposed autoencoder is achieved by reducing the total loss function with the gradient descent iterations. Finally, the invert processing for the output of the autoencoder is designed for the reconstruction of the clutter-plus-noise covariances. Compared to the conventional methods, it performs obviously superior clutter suppression for both side-looking and non-side-looking radars with strong clutter and improves the improvement factor (IF) greatly. The proposed method can adapt better to non-uniform clutter environments and non-ideal sample conditions. In this paper, the non-uniformity of samples mainly refers to those caused by clutter range dependence in non-side-looking arrays.

Notation: $(\cdot)^{\mathrm{H}}$, $(\cdot)^{-1}$ and $(\cdot)^*$ denote conjugate-transpose, inverse, and conjugate operators, respectively. $\otimes$ and $\odot$ denote the Kronecker and the Khatri–Rao product operators, respectively. $\frac{\partial(\cdot)}{\partial(\cdot)}$ denotes the derivative. $\mathbb{C}^{K\times 1}$ denotes the $K \times 1$ dimension.

## 2. Signal Model

Consider an airborne monostatic radar system equipped witha uniform linear array (ULA) whose number of the receiving array elements is $N$. The elements are half-wavelength $d = \lambda/2$ inner spaced, and $K$ pulses are transmitted at a constant pulse repetition frequency (PRF) $f_{PRF}$ during a coherent processing interval (CPI). A range cell is composed of all the clutter scattering cells which hold the same slant distance as the antenna array, and the components from different scatterers in the clutter echo are independent of each other. For the antenna array and the clutter scatterers, it is assumed that the platform is moving at a uniform speed $v_a$, and the angle (non-side-looking angle) between the ULA and the moving direction of the platform is $\vartheta$. Moreover, the angle between the clutter scatterers and the antenna array is $\alpha$, and the angle between the clutter scatterers and the moving direction of the platform is $\beta$. Furthermore, relative to the speed vector of the platform, the elevation angle and the azimuth angle of the clutter scatterers are $\theta$ and $\varphi$, respectively. Then, each column of $N \times K$ dimensional echo data matrix collected from $K$ pulses in a CPI is stacked to obtain a vector $\mathbf{x}_{cn,l}$, as follows [29,30]

$$\mathbf{x}_{cn,l} = \sum_{i=1}^{N_c} \xi_{i,l} \mathbf{a}_{st}(f_{s,i}, f_{d,i}) + \mathbf{n}_l \tag{1}$$

where $N_c$ is the total number of the clutter blocks in the clutter cell, and $\xi_{i,l}$ is the complex amplitude of the $i$th clutter block in the clutter cell. The $l$ stands for the $l$th range cell. Meanwhile, $\mathbf{a}_{st}(f_{s,i}, f_{d,i})$ is the space-time steering vector of the $i$th clutter block, and $\mathbf{n}_l$ is the zero-mean noise vector. In Equation (1), the space-time two-dimensional steering vector $\mathbf{a}_{st}(f_{s,i}, f_{d,i})$ can be expressed in detail as

$$\mathbf{a}_{st}(f_{s,i}, f_{d,i}) = \mathbf{a}_t(f_{d,i}) \otimes \mathbf{a}_s(f_{s,i}) \tag{2}$$

in which the temporal steering vector $\mathbf{a}_t(f_{d,i})$ is expressed as

$$\mathbf{a}_t(f_{d,i}) = [1, e^{j2\pi \frac{f_{d,i}}{f_{PRF}}}, \ldots, e^{j2\pi(K-1)\frac{f_{d,i}}{f_{PRF}}}] \in \mathbb{C}^{K\times 1} \tag{3}$$

and the spatial steering vector $\mathbf{a}_s(f_{s,i})$ is

$$\mathbf{a}_s(f_{s,i}) = [1, e^{j2\pi f_{s,i}}, \ldots, e^{j2\pi(N-1)f_{s,i}}] \in \mathbb{C}^{N\times 1} \tag{4}$$

where $\frac{f_{d,i}}{f_{PRF}}$ and $f_{s,i}$ are the normalized spatial frequency and Doppler frequency, respectively. According to the spatial geometric relationship, the Doppler frequency $f_{d,i}$ of the clutter scatterer can be written as

$$f_{d,i} = \frac{2v_a}{\lambda}\cos(\theta_i)\cos(\varphi_i) = \frac{2v_a}{\lambda}\cos(\beta_i) \tag{5}$$

where $\beta_i$ is the spatial cone angle. In addition, the spatial frequency of the scatterer is

$$f_{s,i} = \frac{d}{\lambda}\cos(\alpha_i) = \frac{d}{\lambda}\cos(\theta_i)\cos(\varphi_i - \vartheta_i) \tag{6}$$

The all-space-time adaptive processing means that all the space and the time degrees of freedom are used in the process of clutter suppression. The purpose of STAP is to suppress clutter as much as possible and keep the target gain of the output unchanged. Thus, according to the linear constraint minimum variance (LCMV) criterion, the adaptive optimal weight vector $\mathbf{w}_{opt}$ can be obtained from the following formula

$$\min_{\mathbf{w}} \mathbf{w}^{\mathrm{H}}\mathbf{R}_X\mathbf{w}, \quad \text{s.t.} \quad \mathbf{w}^{\mathrm{H}}\mathbf{a}_{st}(f_{s0}, f_{d0}) = 1 \tag{7}$$

where $\mathbf{a}_{st}(f_{s0}, f_{d0})$ is the space-time steering vector of the target and $f_{d0}$ and $f_{s0}$ are its temporal frequency and spatial frequency. According to the Lagrange multiplier method, the optimal weight vector $\mathbf{w}_{opt}$ can be obtained from

$$\mathbf{w}_{opt} = \frac{\mathbf{R}_X^{-1}\mathbf{a}_{st}(f_{s0}, f_{d0})}{\mathbf{a}_{st}^{\mathrm{H}}(f_{s0}, f_{d0})\mathbf{R}_X^{-1}\mathbf{a}_{st}(f_{s0}, f_{d0})} \tag{8}$$

where $\mathbf{w}_{opt}$ is the weight vector of STAP and $\mathbf{a}_{st}(f_{s0}, f_{d0})$ is the space-time vector with $NK \times 1$ dimension. $\mathbf{R}_X$ is the covariance matrix, and in practical applications, $\mathbf{R}_X$ is unknown. Therefore, maximum likelihood estimation is usually used to estimate the covariance matrix, as follows [31]:

$$\mathbf{R}_X \approx \hat{\mathbf{R}}_X = \frac{1}{L}\sum_{l=1}^{L}\mathbf{x}(l)\mathbf{x}^{\mathrm{H}}(l) \tag{9}$$

where $\mathbf{x}(l)$ for $l = 1, 2, \ldots, L$ are the training samples near the cell to be detected and they exclude samples with a strong clutter scattering point or suspected target pollution. $L$ is the total number of the training samples used to estimate $\hat{\mathbf{R}}_X$. By solving Equation (7) with $\hat{\mathbf{R}}_X$, the optimal space-time weight vector $\mathbf{w}_{opt}$ can be obtained.

It should be mentioned that according to the formula transformation of the sum-difference angle, for the clutter scatterer, the relationship between its spatial frequency and Doppler frequency is

$$[\frac{f_d}{f_{dm}} - \cos(\alpha)\cos(\vartheta)]^2 + \cos^2(\alpha)\sin^2(\vartheta) = \cos^2(\theta)\sin^2(\vartheta) \tag{10}$$

When the antenna is a positive side-looking structure with $\vartheta = 0$, which means that the array direction is consistent with the moving direction of platform, the clutter spectrum is linearly distributed on the space-time two-dimensional plane and has range stability. When the antenna is a forward-looking structure, that is $\vartheta = 90$, the trajectory of the clutter spectrum in the space-time two-dimensional plane is a cluster of ellipses, and the radii of the ellipses change with distance. When the antenna is a non-side-looking structure with $0 < \vartheta < 90$, the clutter spectrum in the space-time two-dimensional plane presents an oblique elliptical distribution, and the radii of the ellipses change with distance. Therefore, the corresponding training samples do not meet the IID condition.

## 3. The Proposed Algorithm

For the covariance matrix $\hat{\mathbf{R}}_X^{cn}(i,j)$ of clutter-plus-noise, its element can by calculated by

$$\begin{aligned}\hat{\mathbf{R}}_X^{cn}(i,j) &= \mathbf{X}_{cn}(i,:)\mathbf{X}_{cn}^{\mathrm{H}}(j,:) \\ &= \frac{1}{L}\sum_{l=1}^{L}\mathbf{X}_{cn}(i,l)\mathbf{X}_{cn}^{*}(j,l) \quad (1 \le i,j \le NK)\end{aligned} \tag{11}$$

where $\bar{\mathbf{T}}(k_1,k_2)$ generally stands for the element of $\bar{\mathbf{T}}$ located at the $k_1$th row and the $k_2$th column, and $\bar{\mathbf{T}}(k_1,:)$ is the $k_1$th row of the matrix $\bar{\mathbf{T}}$. $\mathbf{X}_{cn}$ is composed of $\mathbf{X}_{cn} = [\mathbf{x}_{cn,1},\ldots,\mathbf{x}_{cn,L}]$. The proposed method designs a unique sub-matrix to be decoded and encoded based on the following construction

$$\hat{\mathbf{R}}_X^{cn} = \left[\begin{array}{c|c} \hat{\mathbf{R}}_{(1)}^{cn} \in \mathbb{C}^{p\times(NK-p)} & \hat{\mathbf{R}}_{(2)}^{cn} \in \mathbb{C}^{p\times p} \\ \hline \hat{\mathbf{R}}_{(3)}^{cn} \in \mathbb{C}^{(NK-p)\times(NK-p)} & \hat{\mathbf{R}}_{(4)}^{cn} \in \mathbb{C}^{(NK-p)\times p} \end{array}\right], \quad p < \frac{NK}{2} \tag{12}$$

where the size of the dimension $p$ is designed as $p < NK/2$ in the proposed method, which means that $p$ is a positive integer that is not too large and less than half of the system's degree of freedom. In detail, the expression of $\hat{\mathbf{R}}_{(2)}^{cn}$ is

$$\hat{\mathbf{R}}_{(2)}^{cn} = \frac{1}{L}\sum_{l=1}^{L}\begin{bmatrix} \mathbf{X}_{cn}(1,l)\mathbf{X}_{cn}^{*}(NK-p+1,l) & \mathbf{X}_{cn}(1,l)\mathbf{X}_{cn}^{*}(NK-p+2,l) & \cdots & \mathbf{X}_{cn}(1,l)\mathbf{X}_{cn}^{*}(NK,l) \\ \mathbf{X}_{cn}(2,l)\mathbf{X}_{cn}^{*}(NK-p+1,l) & \mathbf{X}_{cn}(2,l)\mathbf{X}_{cn}^{*}(NK-p+2,l) & \cdots & \mathbf{X}_{cn}(2,l)\mathbf{X}_{cn}^{*}(NK,l) \\ \vdots & & \ddots & \vdots \\ \mathbf{X}_{cn}(p,l)\mathbf{X}_{cn}^{*}(NK-p+1,l) & \mathbf{X}_{cn}(p,l)\mathbf{X}_{cn}^{*}(NK-p+2,l) & \cdots & \mathbf{X}_{cn}(p,l)\mathbf{X}_{cn}^{*}(NK,l) \end{bmatrix} \tag{13}$$

where $\hat{\mathbf{R}}_{(2)}^{cn}$ contains parts of the clutter-plus-noise covariances. Then, for the reconstruction of the CNCM, a unique matrix is proposed as follows:

$$\begin{aligned}\mathbf{C}_{re} &= [\hat{\mathbf{R}}_{(2)}^{cn} - \bar{\mathbf{M}}] \odot [\sqrt{\frac{p}{\sum_{i=1}^{p}(\hat{\mathbf{R}}_{(2)}^{cn}(i,1) - \bar{\mathbf{M}}(i,1))^2}}\mathbf{I}_{one},\ldots,\sqrt{\frac{p}{\sum_{i=1}^{p}(\hat{\mathbf{R}}_{(2)}^{cn}(i,p) - \bar{\mathbf{M}}(i,p))^2}}\mathbf{I}_{one}] \\ &= [\hat{\mathbf{R}}_{(2)}^{cn} - \bar{\mathbf{M}}] \odot \mathbf{M}_{va}\end{aligned} \tag{14}$$

where $\mathbf{I}_{one} = [1;1;\ldots;1] \in \mathbb{C}^{p\times 1}$, and the $p \times p$ dimensional matrix $\bar{\mathbf{M}}$ is

$$\bar{\mathbf{M}} = \frac{1}{p}\begin{bmatrix} \sum_{i=1}^{p}\hat{\mathbf{R}}_{(2)}^{cn}(i,1), & \sum_{i=1}^{p}\hat{\mathbf{R}}_{(2)}^{cn}(i,2) & \cdots & \sum_{i=1}^{p}\hat{\mathbf{R}}_{(2)}^{cn}(i,p) \\ \sum_{i=1}^{p}\hat{\mathbf{R}}_{(2)}^{cn}(i,1), & \sum_{i=1}^{p}\hat{\mathbf{R}}_{(2)}^{cn}(i,2) & \cdots & \sum_{i=1}^{p}\hat{\mathbf{R}}_{(2)}^{cn}(i,p) \\ \vdots & & \ddots & \vdots \\ \sum_{i=1}^{p}\hat{\mathbf{R}}_{(2)}^{cn}(i,1), & \sum_{i=1}^{p}\hat{\mathbf{R}}_{(2)}^{cn}(i,2) & \cdots & \sum_{i=1}^{p}\hat{\mathbf{R}}_{(2)}^{cn}(i,p) \end{bmatrix} \tag{15}$$

Therefore, the numerical magnitude and variance of the elements $\mathbf{C}_{re}(p_1,p_2)$ for $1 \le p_1,p_2 \le p$ are standardized, by which the covariances are processed and transformed to better adapt to the network in the proposed method. Then, a unique matrix to be decoded and encoded for covariance reconstruction and clutter suppression is proposed as

$$\mathbf{C}_{re}^{r} = \mathrm{real}\{\mathbf{C}_{re}\} \tag{16}$$

where real$\{\cdot\}$ is the real part of $\{\cdot\}$, and $\mathbf{C}_{re}$ is obtained from Equations (12) and (14). With $\mathbf{C}_{re}^{r}$, the proposed method further designs the autoencoder network, which firstly defines

$$\mathbf{H}(:, p_2) = \hbar(\mathbf{C}_{re}^{r}(:, p_2)) = \digamma(\mathbf{W}\mathbf{C}_{re}^{r}(:, p_2) + \mathbf{b}) \tag{17}$$

where $\mathbf{H}$ denotes the output of the encoder, and $\digamma(\cdot)$ is its activation function. Moreover, $\mathbf{W}$ and $\mathbf{b}$ are the weight matrix and the bias coefficient vector of the encoding layer, respectively. It is worth noting that at the same time, the output $\mathbf{H}$ is the input of the decoder, which is expressed as

$$\hat{\mathbf{C}}_{re}^{r}(:, p_2) = g(\mathbf{H}(:, p_2)) = \pounds(\mathbf{U}\mathbf{H}(:, p_2) + \mathbf{f}) \tag{18}$$

where $\hat{\mathbf{C}}_{re}^{r}$ is the output of the decoder, and $\pounds(\cdot)$ is its activation function, which is different from $\digamma(\cdot)$. $\mathbf{U}$ and $\mathbf{f}$ are the weight matrix and the bias coefficient vector of the decoding layer, respectively. For the activation function, based on the positive saturating transfer, it is chosen as

$$\mathbf{H}(i, p_2) = \begin{cases} 0, & \text{if} \quad [\mathbf{W}\mathbf{C}_{re}^{r}(:, p_2) + \mathbf{b}](i) \leq 0 \\ \mathbf{W}(i, :)\mathbf{C}_{re}^{r}(:, p_2) + \mathbf{b}(i), & \text{if} \quad 0 < [\mathbf{W}\mathbf{C}_{re}^{r}(:, p_2) + \mathbf{b}](i) \leq 1 \\ 1, & \text{if} \quad [\mathbf{W}\mathbf{C}_{re}^{r}(:, p_2) + \mathbf{b}](i) > 1 \end{cases} \tag{19}$$

and based on the nonlinear logistic sigmoid function, the decoder is designed as

$$\hat{\mathbf{C}}_{re}^{r}(:, p_2) = \frac{1}{1 + e^{-[\mathbf{U}\mathbf{H}(:, p_2) + \mathbf{f}]}} = \frac{1}{1 + e^{-[\mathbf{U}\digamma(\mathbf{W}\mathbf{C}_{re}^{r}(:, p_2) + \mathbf{b}) + \mathbf{f}]}} \tag{20}$$

where the expression of $\hat{\mathbf{C}}_{re}^{r}$ represents the estimation of the originated input $\mathbf{C}_{re}^{r}$, because the proposed autoencoder neural network makes its covariance estimation output approximate its input $\mathbf{C}_{re}^{r}$ as close as possible by means of the encoding and decoding. As a consequence, the criterion of the loss function of autoencoder network can be written as [32]

$$\Theta_{loss}(\mathbf{C}_{re}^{r}, \hat{\mathbf{C}}_{re}^{r}) = \Delta(\mathbf{C}_{re}^{r}, \hat{\mathbf{C}}_{re}^{r}) \approx 0 \tag{21}$$

which is represented as the following problem [33,34]:

$$(\hat{\mathbf{W}}, \hat{\mathbf{b}}, \hat{\mathbf{U}}, \hat{\mathbf{f}}) = \underset{\mathbf{W},\mathbf{b},\mathbf{U},\mathbf{f}}{\arg\min} \langle [\Delta(\mathbf{C}_{re}^{r}, \hat{\mathbf{C}}_{re}^{r})] \rangle \tag{22}$$

To further solve the problem of Equation (22), the loss function is deduced and analyzed in detail. The total loss function can be divided into three parts, and the criterion of Equation (22) directly results in the first part as follows:

$$E_0 = \frac{1}{p} \sum_{p_2=1}^{p} E_0^{p_2} = \frac{1}{2p} \sum_{p_1=1}^{p} \sum_{p_2=1}^{p} |\mathbf{C}_{re}^{r}(p_1, p_2) - \hat{\mathbf{C}}_{re}^{r}(p_1, p_2)|^2 \tag{23}$$

where $1 \leq p_1, p_2 \leq p$ with $p$ defined in Equation (12). In order to avoid the phenomenon of overfitting and improve the generalization ability of the network, the $l_2$ weight regularization term is brought in as follows:

$$E_1 = \frac{1}{2}\{\sum_{i=1}^{h_n} \sum_{j=1}^{p} [\mathbf{W}(i, j)]^2 + \sum_{i=1}^{p} \sum_{j=1}^{h_n} [\mathbf{U}(i, j)]^2\} \tag{24}$$

where $h_n$ is the chosen number of neurons in the hidden layer. Then, in order to enforce a constraint on the sparsity of the output from the hidden layer and encourage a better compression that benefits for the characteristic extraction, a sparsity regularization term is introduced as

$$\begin{aligned} E_2 &= \sum_{i=1}^{h_n} \mathrm{KL}(\beta \| \hat{\beta}_i) \\ &= \sum_{i=1}^{h_n} [\beta \log(\frac{\beta}{\hat{\beta}_i}) + (1-\beta)\log(\frac{1-\beta}{1-\hat{\beta}_i})] \end{aligned} \tag{25}$$

where $\mathrm{KL}(\cdot)$ is the function of Kullback–Leibler divergence measured how different two distributions are. It takes the value zero when $\hat{\beta}_i = \beta$; otherwise, it becomes larger. Furthermore, $\hat{\beta}$ is the average activation value of $\hbar[\mathbf{C}^r_{re}(:, p_2)]$ in Equation (17) for $p_2 = 1, 2, \ldots, p$, and $\hat{\beta}_i$ is the $i$th element of $\hat{\beta}$. Meanwhile, $\beta$ is the desired value of $\hat{\beta}$. As a result, with $E_0$, $E_1$ and $E_2$, the total loss function is

$$\begin{aligned} L_{oss} &= E_0 + \eta E_1 + \lambda E_2 \\ &= \frac{1}{2p} \sum_{p_1=1}^{p} \sum_{p_2=1}^{p} |\mathbf{C}^r_{re}(p_1, p_2) - \hat{\mathbf{C}}^r_{re}(p_1, p_2)|^2 + \frac{\eta}{2} \{ \sum_{i=1}^{h_n} \sum_{j=1}^{p} [\mathbf{W}(i,j)]^2 \\ &+ \sum_{i=1}^{p} \sum_{j=1}^{h_n} [\mathbf{U}(i,j)]^2 \} + \lambda \sum_{i=1}^{h_n} [\beta \log(\frac{\beta}{\hat{\beta}_i}) + (1-\beta)\log(\frac{1-\beta}{1-\hat{\beta}_i})] \end{aligned} \tag{26}$$

where $\eta$ and $\lambda$ are the coefficients for the $l_2$ weight regularization and the sparsity regularization, respectively. To further train the autoencoder and obtain the desired network coefficients for reconstructing $\mathbf{C}^r_{re}(p_1, p_2)$ making use of the gradient descent method, each iteration that updates $\mathbf{W}(i,j)$ and $\mathbf{b}(i)$ is as follows [35]

$$\mathbf{W}(i,j) = \mathbf{W}(i,j) - \epsilon \frac{\partial L_{oss}}{\partial \mathbf{W}(i,j)} \tag{27}$$

for the weight matrix of encoder and

$$\mathbf{b}(i) = \mathbf{b}(i) - \epsilon \frac{\partial L_{oss}}{\partial \mathbf{b}(i)} \tag{28}$$

for the bias matrix of encoder, and $\frac{\partial(\cdot)}{\partial(\cdot)}$ and $\epsilon$ are the derivative and learning rate, respectively. For the decoder, the weight matrix $\mathbf{U}$ and the bias matrix $\mathbf{f}$ are updated the same as $\mathbf{W}$ and $\mathbf{b}$. Next, $\frac{\partial L_{oss}}{\partial \mathbf{W}(i,j)}$ is derived in detail, as follows:

$$\begin{aligned} \frac{\partial L_{oss}}{\partial \mathbf{W}(i,j)} &= \frac{\partial E_0^{p_2}}{\partial \mathbf{W}(i,j)} + \eta \frac{\partial E_1}{\partial \mathbf{W}(i,j)} + \lambda \frac{\partial E_2}{\partial \mathbf{W}(i,j)} \\ &= \frac{1}{p} \sum_{p_2=1}^{p} \frac{\partial E_0}{\partial [\sum_{p_1=1}^{p} \mathbf{W}(i,p_1)\mathbf{C}^r_{re}(p_1,p_2) + \mathbf{b}(i)]} \frac{\partial [\sum_{p_1=1}^{p} \mathbf{W}(i,p_1)\mathbf{C}^r_{re}(p_1,p_2) + \mathbf{b}(i)]}{\partial \mathbf{W}(i,j)} \\ &+ \eta \frac{\partial [\frac{1}{2}\{\sum_{i=1}^{h_n}\sum_{j=1}^{p}[\mathbf{W}(i,j)]^2 + \sum_{i=1}^{p}\sum_{j=1}^{h_n}[\mathbf{U}(i,j)]^2\}]}{\partial \mathbf{W}(i,j)} + \lambda \frac{\partial \{\sum_{i=1}^{h_n}[\beta \log(\frac{\beta}{\hat{\beta}_i}) + (1-\beta)\log(\frac{1-\beta}{1-\hat{\beta}_i})]\}}{\partial \mathbf{W}(i,j)} \\ &= \frac{1}{p} \sum_{p_2=1}^{p} \mathbf{C}^r_{re}(j,p_2)\xi_{i,p_2} + \eta \mathbf{W}(i,j) + \lambda \frac{\partial}{\partial \mathbf{W}(i,j)} \{\beta(\log\beta - \log\hat{\beta}_i) + (1-\beta)[\log(1-\beta) - \log(1-\hat{\beta}_i)]\} \\ &= \frac{1}{p} \sum_{p_2=1}^{p} \mathbf{C}^r_{re}(j,p_2)\xi_{i,p_2} + \eta \mathbf{W}(i,j) + \lambda(-\frac{\beta}{\hat{\beta}_i} + \frac{1-\beta}{1-\hat{\beta}_i})\hat{\gamma}_i \end{aligned} \tag{29}$$

where the brief representation $\xi_{i,p_2}$ and $\hat{\gamma}_i$ are $\xi_{i,p_2} = \frac{\partial E_0^{p_2}}{\partial[\sum_{p_1=1}^{p} \mathbf{W}(i,p_1)\mathbf{C}_{re}^r(p_1,p_2)+\mathbf{b}(i)]} = \frac{\partial E_0^{p_2}}{\partial \mathbf{Z}(i,p_2)}$ and $\hat{\gamma}_i = \frac{\partial \hat{\beta}_i}{\partial \mathbf{W}(i,j)}$. Then, the proposed method further deduces $\xi_{i,p_2}$ step by step, as follows

$$\begin{aligned}
\xi_{i,p_2} =& \frac{\partial[\frac{1}{2p}\sum_{p_1=1}^{p}|\mathbf{C}_{re}^r(p_1,p_2)-\hat{\mathbf{C}}_{re}^r(p_1,p_2)|^2]}{\partial[\sum_{p_1=1}^{p}\mathbf{W}(i,p_1)\mathbf{C}_{re}^r(p_1,p_2)+\mathbf{b}(i)]} \\
=& \frac{1}{2}\frac{\partial}{\partial[\mathbf{Z}(i,,p_2)]}\{\sum_{j=1}^{p}[\mathbf{C}_{re}^r(j,p_2)-\pounds(\tilde{\mathbf{Z}}(j,p_2))]^2\} \\
=& \sum_{j=1}^{p}\{-[\mathbf{C}_{re}^r(j,p_2)-\pounds(\tilde{\mathbf{Z}}(j,p_2))]\frac{\partial[\pounds(\tilde{\mathbf{Z}}(j,p_2))]}{\partial\tilde{\mathbf{Z}}(j,p_2)}\}\frac{\partial\tilde{\mathbf{Z}}(j,p_2)}{\partial\mathbf{Z}(i,p_2)} \\
=& \sum_{j=1}^{p}\tilde{\xi}_{i,p_2}^u\frac{\partial[\sum_{k=1}^{h_n}\mathbf{U}(j,k)F(\mathbf{Z}(k,p_2))+\mathbf{f}(j)]}{\partial\mathbf{Z}(i,p_2)} \\
=& \sum_{j=1}^{p}\tilde{\xi}_{i,p_2}^u\mathbf{U}(j,i)F'(\mathbf{Z}(i,p_2))
\end{aligned} \tag{30}$$

where $\tilde{\mathbf{Z}}(j,p_2) = \sum_{k=1}^{h_n}\mathbf{U}(j,k)F(\mathbf{Z}(k,p_2)) + \mathbf{f}(j)$, and $\mathbf{Z}(i,p_2) = \sum_{p_1=1}^{p}\mathbf{W}(i,p_1)\mathbf{C}_{re}^r(p_1,p_2) + \mathbf{b}(i)$. Additionally, $\tilde{\xi}_{i,p_2}^u = -[\mathbf{C}_{re}^r(j,p_2)-\pounds(\tilde{\mathbf{Z}}(j,p_2))]\pounds'(\tilde{\mathbf{Z}}(j,p_2))$. $F'(\mathbf{Z}(i,p_2))$ and $\pounds'(\tilde{\mathbf{Z}}(j,p_2))$ are the derivatives with regard to $\mathbf{Z}(i,p_2)$ and $\tilde{\mathbf{Z}}(j,p_2)$, respectively. In Equation (29), with the average activation value $\hat{\beta}$ obtained from $\hbar[\mathbf{C}_{re}^r(:,p_2)]$, that is

$$\hat{\beta}_i = \frac{1}{p}\sum_{p_2=1}^{p}\hbar_i[\mathbf{C}_{re}^r(:,p_2)] = \frac{1}{p}\sum_{p_2=1}^{p}F[\mathbf{Z}(i,p_2)] = \frac{1}{p}\sum_{p_2=1}^{p}F[\sum_{p_1=1}^{p}\mathbf{W}(i,p_1)\mathbf{C}_{re}^r(p_1,p_2)+\mathbf{b}(i)] \tag{31}$$

the detailed derivation of $\hat{\gamma}_i = \frac{\partial\hat{\beta}_i}{\partial\mathbf{W}(i,j)}$ in Equation (29) is

$$\hat{\gamma}_i = \frac{1}{p}\sum_{p_2=1}^{p}F'[\mathbf{Z}(i,p_2)]\frac{\partial[\sum_{p_1=1}^{p}\mathbf{W}(i,p_1)\mathbf{C}_{re}^r(p_1,p_2)+\mathbf{b}(i)]}{\partial\mathbf{W}(i,j)} = \frac{1}{p}\sum_{p_2=1}^{p}F'[\mathbf{Z}(i,p_2)]\mathbf{C}_{re}^r(j,p_2) \tag{32}$$

On the other hand, for Equation (29), the derivative for the bias of the network, is deduced as follows:

$$\begin{aligned}
\frac{\partial L_{oss}}{\partial\mathbf{b}(i)} =& \frac{\partial E_0}{\partial\mathbf{b}(i)} + \eta\frac{\partial E_1}{\partial\mathbf{b}(i)} + \lambda\frac{\partial E_2}{\partial\mathbf{b}(i)} \\
=& \frac{1}{p}\sum_{p_2=1}^{p}\frac{\partial E_0^{p_2}}{\partial[\sum_{p_1=1}^{p}\mathbf{W}(i,p_1)\mathbf{C}_{re}^r(p_1,p_2)+\mathbf{b}(i)]}\frac{\partial[\sum_{p_1=1}^{p}\mathbf{W}(i,p_1)\mathbf{C}_{re}^r(p_1,p_2)+\mathbf{b}(i)]}{\partial\mathbf{b}(i)} \\
&+ \eta\frac{\partial[\frac{1}{2}\{\sum_{i=1}^{h_n}\sum_{j=1}^{p}[\mathbf{W}(i,j)]^2+\sum_{i=1}^{p}\sum_{j=1}^{h_n}[\mathbf{U}(i,j)]^2\}]}{\partial\mathbf{b}(i)} + \lambda\frac{\partial\{\sum_{i=1}^{h_n}[\beta\log(\frac{\beta}{\hat{\beta}_i})+(1-\beta)\log(\frac{1-\beta}{1-\hat{\beta}_i})]\}}{\partial\mathbf{b}(i)} \\
=& \xi_{i,p_2} + (-\frac{\beta}{\hat{\beta}_i}+\frac{1-\beta}{1-\hat{\beta}_i})\tilde{\gamma}_i
\end{aligned} \tag{33}$$

where $\xi_i$ is achieved from Equation (30). Meanwhile different from $\hat{\gamma}_i$ in Equation (29), $\tilde{\gamma}_i$ is

$$\tilde{\gamma}_i = \frac{\partial\hat{\beta}_i}{\partial\mathbf{b}(i)} = \frac{1}{p}\sum_{p_2=1}^{p}F'[\mathbf{Z}(i,p_2)]\frac{\partial[\sum_{p_1=1}^{p}\mathbf{W}(i,p_1)\mathbf{C}_{re}^r(p_1,p_2)+\mathbf{b}(i)]}{\partial\mathbf{b}(i)} = \frac{1}{p}\sum_{p_2=1}^{p}F'[\mathbf{Z}(i,p_2)] \tag{34}$$

At this point, utilizing Equations (27)–(34), $\mathbf{W}$ and $\mathbf{b}$ can be updated. Hence, by repeating the iterative steps of the gradient descent method and reducing the total loss function gradually, the neural network of the proposed autoencoder with the designed

$\mathbf{C}_{re}^{r}$ from Equations (11)–(16), can be acquired. Therefore, after the encoding and decoding, the reconstructed $p < \frac{NK}{2}$ dimensional matrix $\mathbf{C}_{re}^{r}$ based on the processing of the selected covariances is proposed as

$$\tilde{\hat{\mathbf{C}}}_{re}^{r} = \hat{\mathbf{C}}_{re}^{r} \odot [\sqrt{\frac{\sum_{i=1}^{p}(\hat{\mathbf{R}}_{(2)}^{cn}(i,1) - \bar{\mathbf{M}}(i,1))^2}{p}}\mathbf{I}_{one}, \ldots, \sqrt{\frac{\sum_{i=1}^{p}(\hat{\mathbf{R}}_{(2)}^{cn}(i,p) - \bar{\mathbf{M}}(i,p))^2}{p}}\mathbf{I}_{one}] + \bar{\mathbf{M}} \tag{35}$$

where $\bar{\mathbf{M}}$ is calculated by Equation (15). Then based on Equation (16), the final design of the covariance reconstruction $\hat{\mathbf{R}}_{(2)}^{cn}$ is

$$\tilde{\hat{\mathbf{R}}}_{(2)}^{cn} = \tilde{\hat{\mathbf{C}}}_{re}^{r} + \text{imag}\{\mathbf{C}_{re}\} \tag{36}$$

where $\mathbf{C}_{re}$ is the same with that in Equation (14), and imag$\{\cdot\}$ represents its imaginary part. With the reconstructed and corrected $\tilde{\hat{\mathbf{R}}}_{(2)}^{cn}$ and the structure $\tilde{\hat{\mathbf{R}}}_{X}^{cn} = [\hat{\mathbf{R}}_{(1)}^{cn}, \tilde{\hat{\mathbf{R}}}_{(2)}^{cn}; \hat{\mathbf{R}}_{(3)}^{cn}, \hat{\mathbf{R}}_{(4)}^{cn}]$ in Equation (12), the weight vector of STAP as Equation (8) can be exploited to suppress the clutter and obtain better performance. The proposed method is summarized in Table 1.

**Table 1.** Summary of the proposed method.

1. By matching filter processing and pulse stacking, the received data are obtained as Equation (1).
2. According to the space-time covariance, a unique matrix is proposed.
   (a) The dimension of the unique matrix is designed as less than half of the DOF as Equations (11)–(13);
   (b) A series of covariance data selections and matrix transformations are proposed, as Equations (14)–(16).
3. Designing the unique matrix as the input, the proposed method further constructs the autoencoder network for it with $l_2$ regularization and sparsity regularization.
   (a) The activation functions are obtained from Equations (17)–(20);
   (b) The loss function is obtained from Equations (21)–(26).
4. Training the proposed network by means of reducing the total loss function with gradient descent iterations as Equations (27)–(34).
5. Taking advantage of the designed structure of the unique matrix and the output of the proposed network, the inverse matrix transformation is brought in and covariances are reconstructed, as Equations (35)–(36).
6. Performing STAP with the reconstructed clutter-plus-noise covariances, clutter suppression can be achieved by Equations (7) and (8).

## 4. Simulation Results

In this section, some simulations performed to verify the effectiveness and advantages of the proposed method are discussed. Consider an airborne radar system that has a number of receiving array elements, number of pulses in a CPI, height of the airborne platform, pulse repetition frequency, distance between elements, and wavelength of $N = 8$, $K = 9$, $H = 6000$ m, $f_{PRF} = 9$, $d = 0.3$ m, and $\lambda = 0.15$ m, respectively. In addition, for the proposed method, the transfer functions of the encoder and the decoder are set as Equations (22) and (23), and the coefficient for the $l_2$ weight regularizer is $\eta = 0.01$. Meanwhile, the sparsity regularization is set as $\lambda = 15$, and the maximum number of epochs is 300. The noise power is fixed, and the scattering coefficients of clutter patches are subject to the complex Gaussian distribution with the amplitude determined by the clutter-to-noise ratio (CNR). When comparing the clutter suppression performance of the proposed method with that of different representative methods, the improvement factor (IF) given by IF $= \frac{|\mathbf{w}^{\mathrm{H}}\mathbf{s}|^2}{\mathbf{w}^{\mathrm{H}}\mathbf{R}\mathbf{w}} \frac{\text{trace}(\mathbf{R})}{\mathbf{s}^{\mathrm{H}}\mathbf{s}}$ is exploited. Other parameters used to measure the comprehensive performance will be introduced in detail in each of the following simulations.

### *4.1. Space-Time Distributions of Clutter and Two Dimensional Frequency Response*

Figure 1 depicts the space-time distributions of the clutter spectra for the original STAP method and the proposed method, where a side-looking radar with the angle between its antenna axis and the velocity of the airborne platform being $\vartheta = 0^\circ$ is considered, and its velocity is $v_a = 70$ m/s. Additionally, the number of the training range cells is $L = 40$, and the clutter-to-noise ratio is CNR $= 50$ dB. For the proposed method, the number of neurons in the hidden layer is set as $h_n = 10$, and the sparsity proportion is $sp = 0.01$. This simulation verifies that the proposed method effectively obtains a high-accuracy estimation of the clutter spectrum and overcomes the relatively low accuracy with the broadened clutter ridge provided by the original data of STAP method.

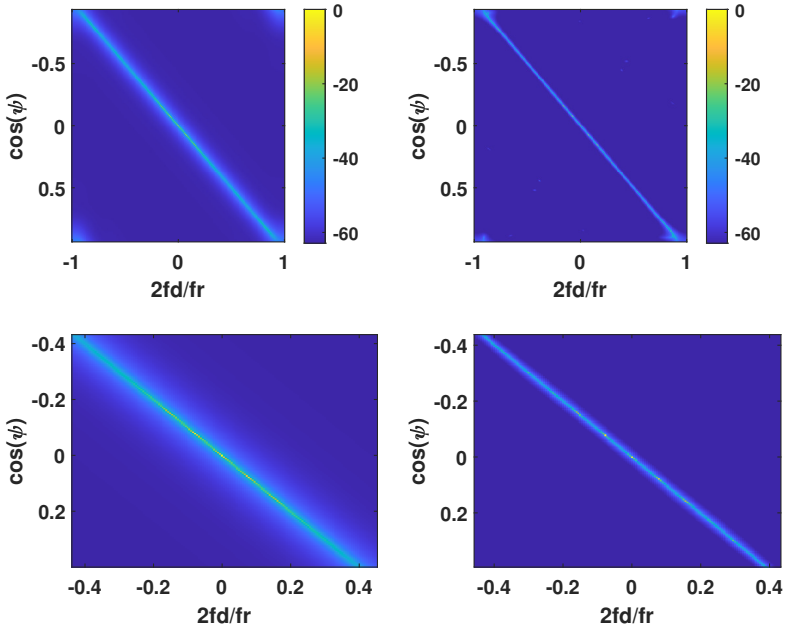
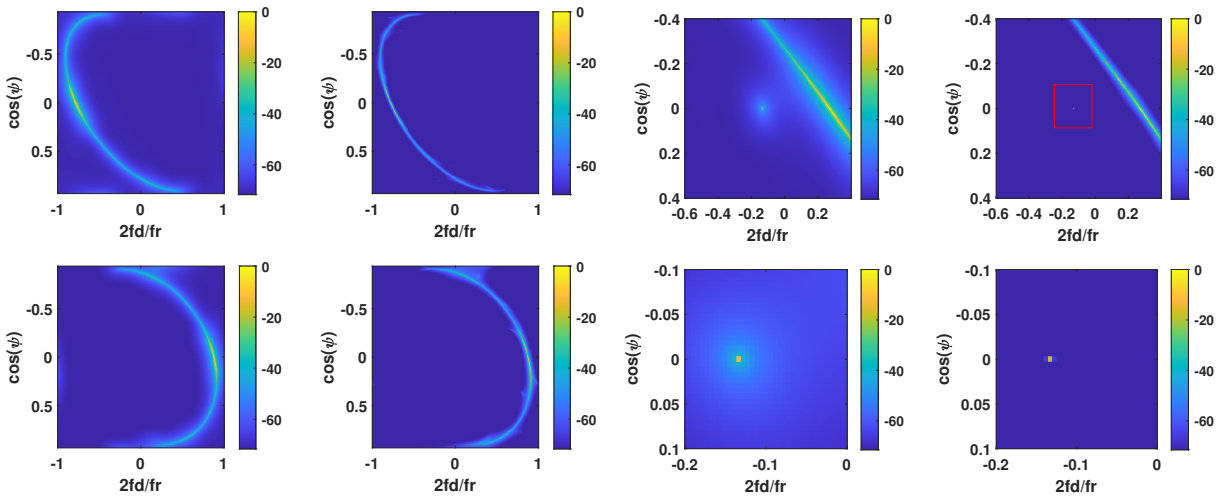



**Figure 1.** Space-time distributions of clutter spectra with data of original STAP (**left**) and proposed method (**right**).

Figure 2 demonstrates the space-time distributions of non-uniform clutter caused by non-side-looking array with data of the original STAP (left columns in Figure 2a,b) and the proposed method (right columns in Figure 2a,b). In Figure 2a, non-side-looking radars with non-side-looking angles $\vartheta = 60^\circ$ (upper row) and $\vartheta = -75^\circ$ (lower row) are considered, respectively. Meanwhile, the velocity of the airborne platform is $v_a = 70$ m/s, and the number of the training samples is $L = 40$. In addition, CNR $= 60$ dB. To further measure the effectiveness of the proposed method for different non-uniform clutter caused by different non-side-looking arrays and conditions, in Figure 2b, $\vartheta = -15^\circ$, $v_a = 80$ m/s, $L = 60$ are considered. Furthermore, the echo of a target with speed $v = 120$ m/s is contained to show its effect. It can be directly observed from Figure 2a,b that, different from Figure 1 with a small number of IID samples, when considering non-uniform clutter caused by different non-side-looking arrays, the proposed method is still effective. With this kind of non-uniform clutter, the clutter ridge provided by the proposed method becomes narrower and clearer. This indicates that the proposed method improves the estimation accuracy of CNCM, which can be used to bring in better clutter suppression. At the same time, Figure 2b emphasizes that at this condition when the spectrum contains the target, the proposed algorithm simultaneously overcomes the broadening phenomena of the clutter ridge and the target to contribute to better clutter suppression.




(**a**) Without Target (**b**) With Target

**Figure 2.** (**a**) Space-time distributions of non-uniform clutter caused by non-side-looking array with data of original STAP and proposed method. (**b**) Space-time distributions of target and non-uniform clutter.

Figure 3 shows the frequency response of the space-time two-dimensions for the original STAP method and the proposed method. In Figure 3, the subfigures on the top left and on the bottom left are the frequency responses of the original STAP method with CNR = 50 dB and CNR = 60 dB, respectively. Furthermore, for comparison, the subfigures on the top right and on the bottom right are those of the proposed method with CNR = 50 dB and CNR = 60 dB, respectively. The other conditions remain the same as Figure 1. It can be seen that in the corresponding clutter region, a deep notch of the space-time frequency response is achieved by the proposed method along the clutter oblique band. However, the performance of the original STAP method degrades severely because of the insufficiency of the training samples. This means that the proposed method can filter out the clutter effectively and improve the performance of processor greatly.

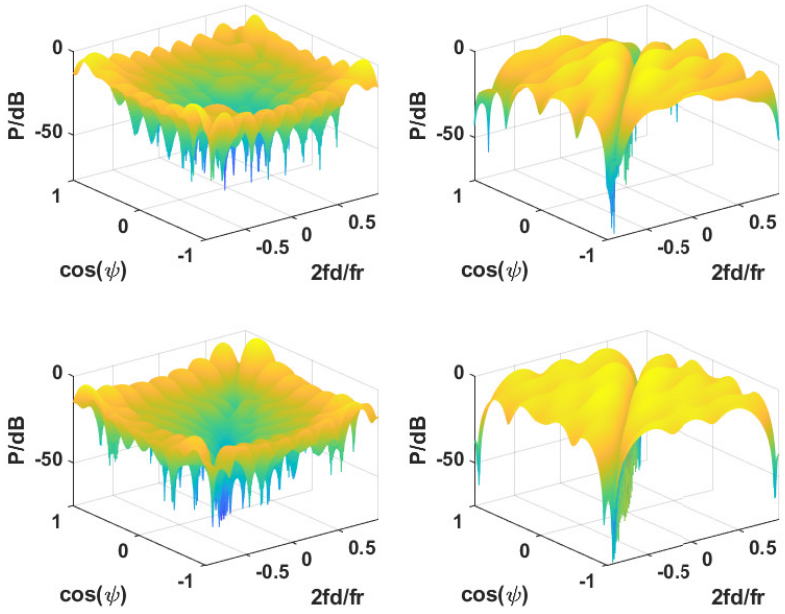



**Figure 3.** Frequency response of space-time two dimensions for original STAP method (**left**) and proposed method (**right**).

### *4.2. Improvement Factor Results*

Figure 4 demonstrates the IF versus normalized Doppler frequency for different methods with a non-side-looking radar, whose angle between the antenna axis and the platform velocity is $\vartheta = 60^\circ$. Other simulation conditions are the same as those in Figure 1. Under the condition of non-side-looking that faces the clutter range dependence, due to the different space-time distributions of clutter spectra of the range cells, averaging the samples from the adjacent range cells leads to the serious broadening of the clutter spectrum. This indicates serious degradation of the clutter suppression performance of STAP method, especially the clutter suppression performance of the main lobe. After the range compensations by the DW method and ADC method, the non-uniformity level of clutter is improved. On average, the IF curve obtained by the proposed algorithm is higher and narrower than the other IF curves of STAP method, JDL method, DW method, and ADC method. Then, the proposed algorithm achieves better clutter suppression performance than the other methods.

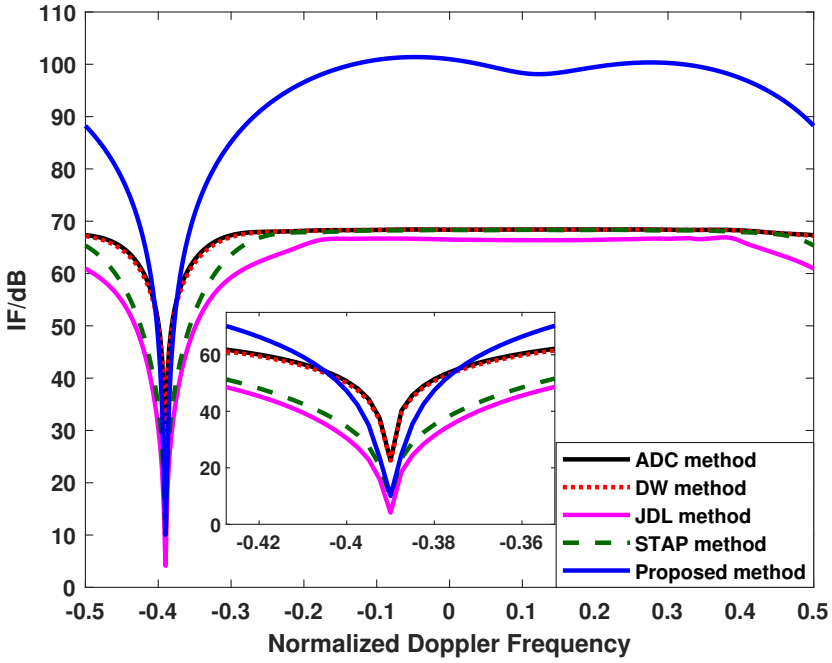



**Figure 4.** IF versus normalized Doppler frequency for non-side-looking radar.

To compare the influence of different platform parameters, Figure 5 shows the IF versus normalized Doppler frequency for different methods under the same condition with Figure 4 except for the velocity of the airborne platform $v_t = 170$ m/s. In Figure 5, the range dependence and the deficiency of training samples cause the relatively shallow notches in the IF curves of DW method and ADC method. In contrast, the IF notch of the proposed method remains simultaneously deep and narrow. Furthermore, the proposed method improves the IF more than 10 dB than the others, on average. Figures 4 and 5 indicate that in the presence of the non-uniformity and the deficiency of training samples, the proposed method has a better ability to suppress the strong clutters.

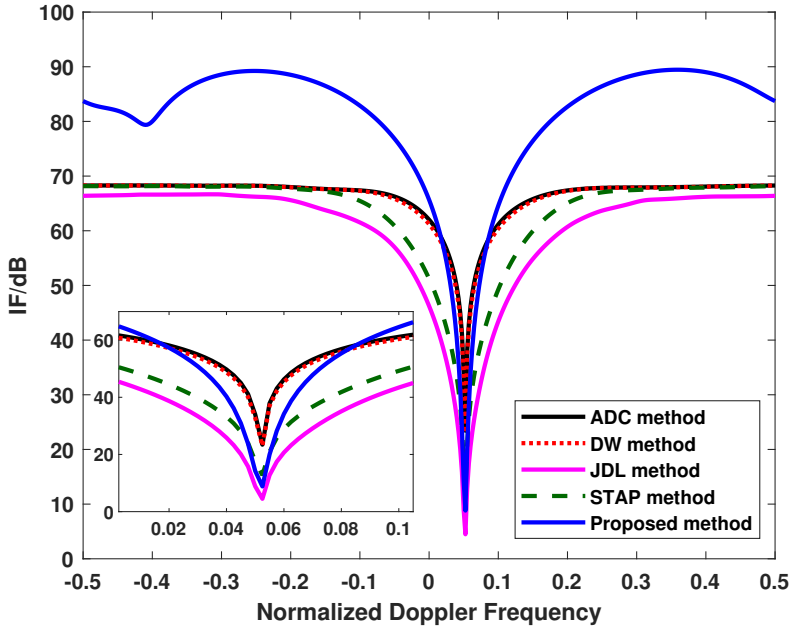



**Figure 5.** IF versus normalized Doppler frequency for different platform parameters.

Figure 6 verifies the IF versus normalized Doppler frequency for different methods, where a side-looking radar with $\vartheta = 0°$, and different clutter-to-noise ratios CNR $= 40$ dB and CNR $= 60$ dB are considered. As shown in Figure 6, for all the methods, an increase in CNR corresponds to an improvement in IF. For different CNRs, compared with the other methods, the proposed method provides the narrowest and deepest IF notch. The obvious superiority of the proposed method results from the designed $p$-dimensional matrix $\tilde{\hat{\mathbf{C}}}_{re}^{r}$ based on the processed and selected covariances; meanwhile, it results from its automatic encoded and decoded reconstruction that compresses the characteristics of CNCM and abandons the deviations caused by the non-uniformity and the deficiency of training samples.

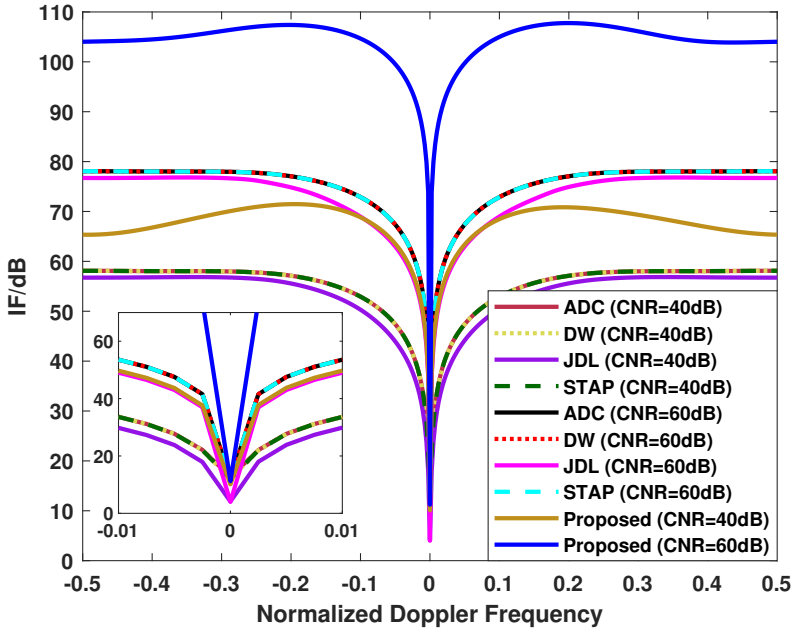
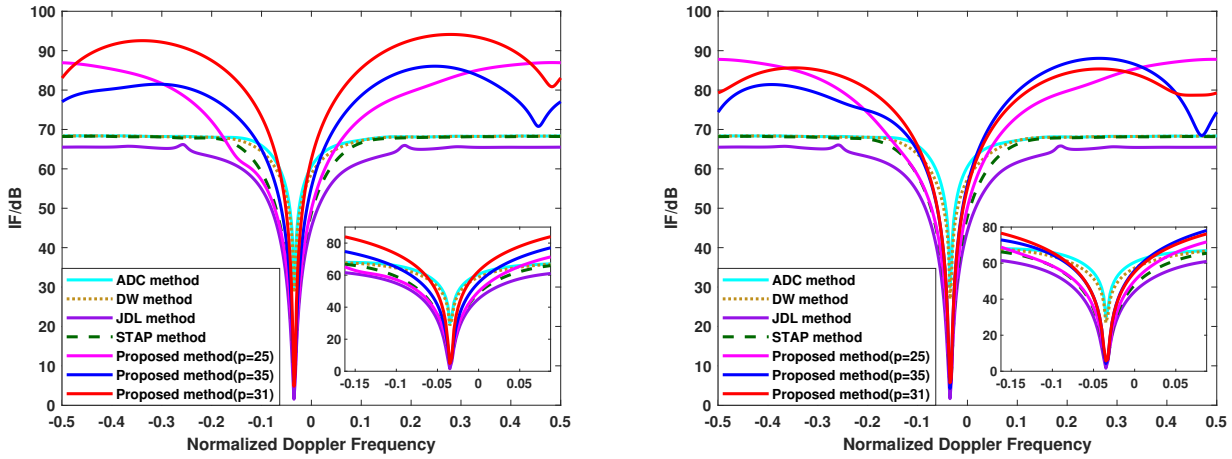



**Figure 6.** IF versus normalized Doppler frequency for side-looking radar.

Figure 7a,b show the IFs of different methods versus normalized Doppler frequency, where for a forward-looking radar with $-90°$ and $v_a = 150$ m/s, CNR is fixed at CNR = 50 dB.

For the proposed method, to compare the influence of different dimensions of the designed matrix $\mathbf{C}_{re}^{r} \in \mathbb{C}^{p \times p}$ to be decoded and encoded, three contrasts $p = 25$, $p = 31$ and $p = 35$ are considered. In regard to the number of the training samples obtained from the corresponding range cells, $L = 50$ is used in Figure 7a, while $L = 60$ is used in Figure 7b. Figure 7 verifies for different numbers of the training samples, the proposed method maintains superior performance. In addition, Figure 7a,b demonstrates that the dimension of $\mathbf{C}_{re}^{r}$ has an effect on the clutter suppression performance. A small dimension of $\mathbf{C}_{re}^{r}$ results in the degraded performance of clutter suppression because of the inadequate reconstruction of CNCM, whereas the overlarge dimension of $\mathbf{C}_{re}^{r}$ is also unsuitable on account of the incompetence of the reconstruction with excessive characteristics and deviations.




(**a**) $L = 50$ (**b**) $L = 60$

**Figure 7.** IF versus normalized Doppler frequency for CNR = 50 dB. (**a**) The number of training samples is $L = 50$. (**b**) The number of training samples is $L = 60$.

### *4.3. Convergence Results of the Proposed Algorithm*

Figure 8 shows the root mean square error (RMSE) between the output of the proposed autoencoder neural network and its input versus the number of epochs. Specifically, the input is the designed $p \times p$ dimensional $\mathbf{C}_{re}^{r}$ based on the processed and selected covariances, and the output is its reconstructed data $\hat{\mathbf{C}}_{re}^{r}$ after the encoding and decoding by the proposed method. Large RMSE means relatively large differences between $\mathbf{C}_{re}^{r}$ and $\hat{\mathbf{C}}_{re}^{r}$. It can be seen that the RMSE drops dramatically at about $n_{ep} = 100$ to $n_{ep} = 250$ epochs, and then, with the increased number of epochs, the RMSE remains unchanged. For different CNRs, the convergence of network training can be obtained by the proposed method. Once converged, more epochs are not helpful in improving the accuracy of the reconstruction.

Figure 9 illustrates the RMSE of the proposed network versus the number of epochs, where different numbers of hidden neurons and sparsity proportions are analyzed for comparison. Under different parameter settings, the proposed method can converge with a certain number of the hidden neurons. It can be observed in Figure 9 that the setting of $sp = 0.0001$ provides the lowest RMSE and the fastest convergence, which can be achieved when the number of epochs reaches just $n_{ep} = 5$ to $n_{ep} = 30$. Additionally, with the same $sp$, the number of the hidden neurons $Hiddensize = 30$ leads to a higher RMSE than $Hiddensize = 10$. For these reasons, a lower sparsity proportion encourages the higher degree of sparsity. Furthermore, within the proper limits, higher sparsity with lower $sp$ and fewer hidden neurons force the proposed network to capture only the more important characteristics and be less sensitive to the deviations of the inaccurate samples. This is beneficial to a more accurate reconstruction of $\mathbf{C}_{re}^{r}$. With regard to the proper limits of the sparsity at this condition, as can be seen from Figure 9, when the proportion $sp$ drops down to $sp = 0.000001$, higher sparsity with lower $sp$ cannot result in a better RMSE between the output of the proposed autoencoder network and its input. However, lower

sparsity with $sp = 0.01$ is improper because of the corresponding large RMSEs. Therefore, $0.01 \leq sp \leq 0.000001$ can be used to be the sparsity proportion for the proposed network.

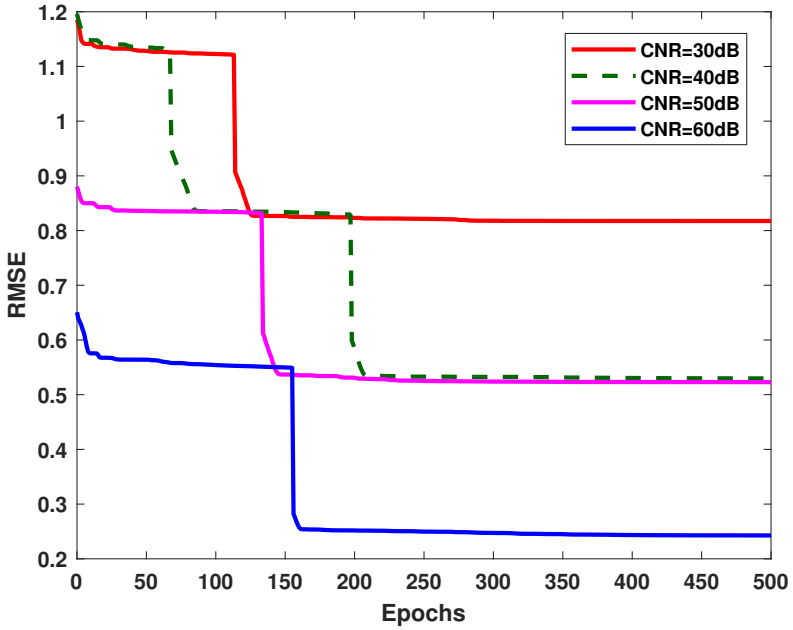



**Figure 8.** RMSE between the output and the original input versus the number of epochs.

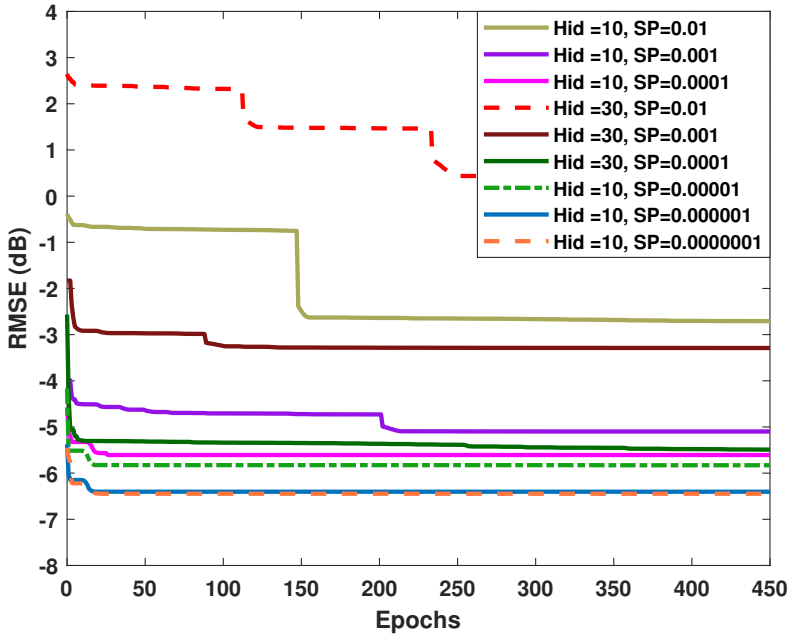



**Figure 9.** RMSE of the proposed network versus the number of epochs for different hidden sizes and sparsity proportions.

### *4.4. Computation Time of the Proposed Algorithm*

Figure 10 shows the computation time of the proposed method versus the total number of epochs $n_{ep}$, with different parameter settings of $p$, *Hiddensize*, and $sp$. Moreover, the computation time of the other analyzed methods versus the number of the training samples $L$ is also shown in Figure 10. Different from the other analyzed methods whose computation complexity is mainly affected by $L$, the computational burden of the proposed method mainly comes from the total number of epochs. In Figure 10, with the reduction of the epochs $n_{ep}$, the speed is improved gradually and obviously. The setting of $sp = 0.0001$ acquires obvious

advantages when the epochs $n_{ep} > 50$. Additionally, at the same $sp$, fewer dimensions of the designed $\mathbf{C}_{re}^{r}$ and fewer hidden neurons can contribute to lowering the computation time of the proposed method. However, as can also be seen from Figure 10, compared with the DW, ADC, JDL, and STAP methods, the proposed method is obviously slower because of the network training for the reconstruction of clutter-plus-noise covariances.

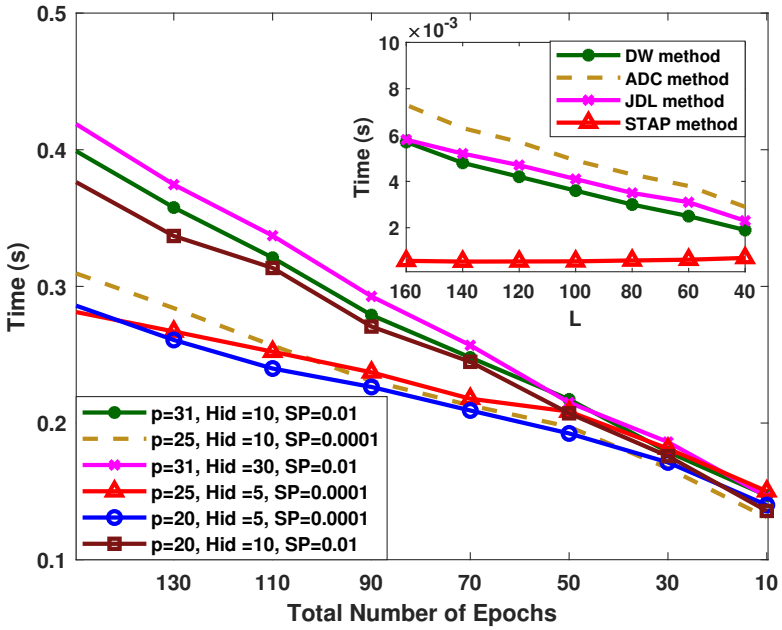



**Figure 10.** Computation time versus the total number of epochs.

## 5. Discussion

With the designed unique matrix, the proposed method compresses and retains the characteristics of the covariances and abandons the deviations caused by the deficiency of IID training samples. The design of the dimension $p$ of the unique matrix in the proposed method is less than half of the DOF, that is $p < NK/2$. This is because, according to the properties of clutter-plus-noise covariances and the structure of CNCM, $p < NK/2$ can ensure that the clutter-plus-noise power $\{\hat{\mathbf{R}}_X^{cn}(i,i)\}_{i=1,2,\ldots,NK}$ located at the diagonal position is excluded from the selected covariances. This protects the most important power characteristics of CNCM and avoids the characteristic losses caused by the reconstruction process. Therefore, as the input of the network, the unique matrix with a certain dimension and matrix transformations allows the nonlinearity of the network be successfully explored and contributed to the improvement of CNCM.

Traditional dimension or rank reduction methods and distance compensation methods try to adapt to the inadequate IID training samples at the cost of reducing the degree of freedom of the system, giving up part of the sample data or relying on the inertial navigation system parameters. These disadvantages make the performance of clutter suppression and target detection not ideal and make the level of self-adaptability low. In contrast, the proposed method overcomes the above deficiencies, and it improves self-adaptability without reducing the degree of freedom. For different conditions of inadequate IID training samples, including the non-uniformity caused by the clutter range dependence of non-side-looking radar and the insufficient quantity, the proposed method provides obviously better clutter suppression performance. However, the proposed method has relatively high computational complexity, as shown in the simulations. The main cost of computing time is network training containing gradient descent process with iterations. As a result, DW, ADC, JDL, and STAP algorithms, which are without iterations, are faster than the proposed method. For further improvement of computing time, faster network training methods can be used to replace the gradient descent method. Moreover, controlling the total number of iterations can also make a contribution.

## 6. Conclusions

In this paper, we propose an autoencoder neural network-based STAP algorithm for airborne radar with inadequate training samples. In the proposed method, a matrix to be decoded and encoded whose dimension is less than half of the DOF is designed by processing of the selected clutter-plus-noise covariances. Then, by training the proposed autoencoder neural network and the invert processing for its output, the reconstruction of the clutter-plus-noise covariances is achieved for STAP. The simulation results have verified that the proposed method performs obviously superior clutter suppression for both side-looking and non-side-looking radars with strong clutter. Moreover, the proposed method can better adapt to the non-ideal sample conditions including the non-uniform clutter environments and inadequate training samples.

**Author Contributions:** Conceptualization, J.L. and G.L.; methodology, J.L.; software, J.L. and F.H.J.; resources, J.L., G.L., J.X. and C.Z.; writing—original draft preparation, J.L.; writing—review and editing, J.L. and J.X.; supervision, S.Z. and F.H.J.; funding acquisition, G.L., J.X., S.Z., J.L. and C.Z. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by the National Natural Science Foundation of China (NSFC) under Grants 61901340, 61931016, and 62071344, the Young Talent Starlet in Science and Technology in Shaanxi under Grant No. 2022KJXX-38, the Fundamental Research Funds for the Central Universities under Grant JB210212, the Science and Technology Innovation Team of Shaanxi Province under Grant 2022TD-38, and the stabilization support of National Radar Signal Processing Laboratory under Grant KGJ202X0X.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Kang, M.S.; Kim, K.T. Automatic SAR Image Registration via Tsallis Entropy and Iterative Search Process. *IEEE Sens. J.* **2020**, *20*, 7711–7720. [CrossRef]
2. Gong, M.G.; Cao, Y.; Wu, Q.D. A Neighborhood-Based Ratio Approach for Change Detection in SAR Images. *IEEE Geosci. Remote Sens. Lett.* **2012**, *9*, 307–311. [CrossRef]
3. Hakim, W.L.; Achmad, A.R.; Eom, J.; Lee, C.W. Land Subsidence Measurement of Jakarta Coastal Area Using Time Series Interferometry with Sentinel-1 SAR Data. *J. Coast. Res.* **2020**, *102*, 75–81. [CrossRef]
4. Ward, J. *Space-Time Adaptive Processing for Airborne Radar*; Technical Report; MIT Lincoln Laboratory: Lexington, KY, USA, 1998.
5. Klemm, R. *Principles of Space-Time Adaptive Processing*; The Institution of Electrical Engineers: London, UK, 2002.
6. Guerci, J.R. *Space-Time Adaptive Processing for Radar*; Artech House: Norwood, MA, USA, 2003.
7. Reed, I.S.; Mallett, J.D.; Brennan, L.E. Rapid convergence rate in adaptive arrays. *IEEE Trans. Aerosp. Electron. Syst.* **1974**, *AES-10*, 853–863. [CrossRef]
8. Zhang, Q.; Mikhael, W.B. Estimation of the clutter rank in the case of subarraying for space-time adaptive processing. *Electron. Lett.* **1997**, *35*, 419–420. [CrossRef]
9. Melvin, W.L. Space-time adaptive radar performance in heterogeneous clutter. *IEEE Trans. Aerosp. Electron. Syst.* **2000**, *36*, 621–633. [CrossRef]
10. Lapierre, F.D.; Ries, P.; Verly, J.G. Foundation for mitigating range dependence in radar space-time adaptive processing. *IET Radar Sonar Navig.* **2009**, *3*, 18–29. [CrossRef]
11. Guerci, J.R.; Goldstein, J.S; Reed, I.S. Optimal and adaptive reduced-rank STAP. *IEEE Trans. Aerosp. Electron. Syst.* **2000**, *36*, 647–663. [CrossRef]
12. Liao, G.S.; Bao, Z.; Xu, Z.Y. A framework of rank-reduced space-time adaptive processing for airborne radar and its applications. *Sci. China Ser. E Technol. Sci.* **1997**, *40*, 505–512. [CrossRef]
13. Zhang, L.; Bao, Z.; Liao, G.S. A comparative study of eigenspace based rank reduced STAP methods. *Acta Electron. Sin.* **2000**, *28*, 27–30.
14. Goldstein, J.S. Reduced rank adaptive filtering. *IEEE Trans. Signal Process.* **1997**, *45*, 492–496. [CrossRef]
15. Wang, W.L.; Liao, G.S.; Zhang, G.B. Improvement on the performance of the auxiliary channel STAP in the non-homogeneous environment. *J. Xidian Univ.* **2004**, *20*, 426–429.
16. Wang, Y.; Peng, Y. Space-time joint processing method for simultaneous clutter and jamming rejection in airborne radar. *Electron. Lett.* **1996**, *32*, 258. [CrossRef]
17. Yang, Z.C.; Li, X.; Wang, H.Q.; Jiang, W.D. On clutter sparsity analysis in space-time adaptive processing airborne radar. *IEEE Geosci. Remote Sens. Lett.* **2013**, *10*, 1214–1218. [CrossRef]
18. Sun, K.; Meng, H.; Wang, Y.; Wang, X. Direct data domain STAP using sparse representation of clutter spectrum. *Signal Process.* **2011**, *91*, 2222–2236. [CrossRef]

19. Yang, Z.C.; Lamare, R.; Liu, W. Sparsity-based STAP using alternating direction method with gain/phase errors. *IEEE Trans. Aerosp. Electron. Syst.* **2017**, *53*, 2756–2768.
20. Ji, S.; Xue, Y.; Carin, L. Bayesian compressive sensing. *IEEE Trans. Signal Process.* **2008**, *56*, 2346–2356. [CrossRef]
21. Lim, C.H.; Mulgrew, B. Filter banks based JDL with angle and separate Doppler compensation for airborne bistatic radar. In Proceedings of the International Radar Symposium India, Bangalore, India, 18–22 December 2005; pp. 583–587.
22. Lapierre, F.; Droogenbroeck, M.V.; Verly, J.G. New methods for handling the dependence of the clutter spectrum in non-sidelooking monostatic STAP radars. In Proceedings of the IEEE Acoustics, Speech, and Signal Processing Conference, Hong Kong, China, 6–10 April 2003; pp. 73–76.
23. Lapierre, F.; Verly, J.G. Registration-based range dependence compensation for bistatic STAP radars. *EURASIP J. Appl. Signal Process.* **2005**, *1*, 85–98. [CrossRef]
24. Lapierre, F.; Verly, J.G. Computationally-efficient range dependence compensation method for bistatic radar STAP. In Proceedings of the IEEE International Radar Conference, Arlington, VA, USA, 9–12 May 2005; pp. 714–719.
25. Borsari, G.K. Mitigating effects on STAP processing caused by an inclined array. In Proceedings of the 1998 IEEE Radar Conference, Dallas, TX, USA, 11–14 May 1998; pp. 135–140.
26. Kreyenkamp, O.; Klemm, R. Doppler compensation in forward-looking STAP radar. *IEE Proc. Radar Sonar Navig.* **2001**, *148*, 253–258. [CrossRef]
27. Himed, B.; Zhang, Y.; Hajjari, A. STAP with angle-Doppler compensation for bistatic airborne radars. In Proceedings of the IEEE Radar Conference, Long Beach, CA, USA, 25 April 2002; pp. 311–317.
28. Pearson, F.; Borsari, G. Simulation and analysis of adaptive interference suppression for bistatic surveillance radars. In Proceedings of the Adaptive Sensor Array Process, Lexington, MA, USA, 5–6 June 2007; LincoIn Laboratory: VIrkshop, MA, USA, 2007.
29. Liu, K.; Wang, T.; Wu, J.; Liu, C.; Cui, W. On the Efficient Implementation of Sparse Bayesian Learning-Based STAP Algorithms. *Remote Sens.* **2022**, *14*, 3931. [CrossRef]
30. Xu, J.W.; Liao, G.S.; Huang, L.; Zhu, S.Q. Joint magnitude and phase constrained STAP approach. *Digit. Signal Process.* **2015**, *46*, 32–40. [CrossRef]
31. Ottersten, B.; Stoica, P.; Roy R. Covariance matching estimation techniques for array signal processing applications. *Digit. Signal Process.* **1998**, *8*, 185–210. [CrossRef]
32. Bengio, Y.; Courville, A.; Vincent, P. Representation learning: A review and new perspectives. *IEEE Trans. Pattern Anal. Mach. Intell.* **2013**, *35*, 1798–1828. [CrossRef] [PubMed]
33. Goodfellow, I.; Bengio, Y.; Courville, A. *Deep Learning*; MIT Press: Cambridge, UK, 2016; Volume 1, Chapter 14, pp. 499–507.
34. Gregor, K.; Lecun, Y. Learning fast approximations of sparse coding. In Proceedings of the 27th International Conference on International Conference on Machine Learning, Haifa, Israel, 21–24 June, 2010; pp. 399–406.
35. Zhang, S.X.; Choromanska, A.; Yann, L.C. Deep learning with Elastic Averaging SGD. *Adv. Neural Inf. Process. Syst.* **2015**, *28*, 1–24.