



Article

Pseudo- $\mathcal{L}_0$ -Norm Fast Iterative Shrinkage Algorithm Network: Agile Synthetic Aperture Radar Imaging via Deep Unfolding Network

Wenjiao Chen ¹ , Jiwen Geng ^{2,*}, Fanjie Meng ¹ and Li Zhang ³

¹ The Department of Space Control and Communication, Space Engineering University, Beijing 102249, China; chenwenjiao008@163.com (W.C.); beilimeng1992@163.com (F.M.)

² The School of Information Science and Engineering, Southeast University, Nanjing 214135, China

³ The 15th Research Institute of China Electronics Technology Corporation, Beijing 100083, China; zhangli35@cetc.com.cn

* Correspondence: gengjiwen@buaa.edu.cn

Abstract: A novel compressive sensing (CS) synthetic-aperture radar (SAR) called AgileSAR has been proposed to increase swath width for sparse scenes while preserving azimuthal resolution. AgileSAR overcomes the limitation of the Nyquist sampling theorem so that it has a small amount of data and low system complexity. However, traditional CS optimization-based algorithms suffer from manual tuning and pre-definition of optimization parameters, and they generally involve high time and computational complexity for AgileSAR imaging. To address these issues, a pseudo- $\mathcal{L}_0$ -norm fast iterative shrinkage algorithm network (pseudo- $\mathcal{L}_0$ -norm FISTA-net) is proposed for AgileSAR imaging via the deep unfolding network in this paper. Firstly, a pseudo- $\mathcal{L}_0$ -norm regularization model is built by taking an approximately fair penalization rule based on Bayesian estimation. Then, we unfold the operation process of FISTA into a data-driven deep network to solve the pseudo- $\mathcal{L}_0$ -norm regularization model. The network's parameters are automatically learned, and the learned network significantly increases imaging speed, so that it can improve the accuracy and efficiency of AgileSAR imaging. In addition, the nonlinearly sparsifying transform can learn more target details than the traditional sparsifying transform. Finally, the simulated and data experiments demonstrate the superiority and efficiency of the pseudo- $\mathcal{L}_0$ -norm FISTA-net for AgileSAR imaging.

Keywords: synthetic-aperture radar (SAR); AgileSAR imaging; pseudo- $\mathcal{L}_0$ -norm fast iterative shrinkage algorithm network (pseudo- $\mathcal{L}_0$ -norm FISTA-net)



Citation: Chen, W.; Geng, J.; Meng, F.; Zhang, L. Pseudo- $\mathcal{L}_0$ -Norm Fast Iterative Shrinkage Algorithm Network: Agile Synthetic Aperture Radar Imaging via Deep Unfolding Network. *Remote Sens.* **2024**, *16*, 671. <https://doi.org/10.3390/rs16040671>

Academic Editor: Piotr Samczynski

Received: 7 December 2023

Revised: 3 February 2024

Accepted: 9 February 2024

Published: 13 February 2024



Copyright: © 2024 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

High-resolution wide-swath (HRWS) synthetic-aperture radars (SARs) are an important research topic due to their highly efficient observation [1,2]. Although azimuthal multi-channel SARs [3] and multi-input multi-output (MIMO) SARs [4] can achieve high azimuthal resolution and wide-range swath, they have a large amount of data and a highly complex antenna. As the compressive sensing (CS) theorem develops [5–7], an innovative system concept for sparse scenes, called AgileSAR, is unconstrained by the Nyquist theorem so that the amount of data is less than that of the mentioned HRWS system [8]. Simultaneously, the CS algorithm guarantees azimuthal resolution, and time-space sampling persuades wide-swath coverage mosaicked by several range sub-swaths to achieve 5 m resolution and a 300 km swath with a single channel. AgileSAR adopts sub-Nyquist sampling along the azimuthal dimension, and the sparse scene is recovered by AgileSAR imaging [8].

In spite of the fact that AgileSAR can achieve the above merits, AgileSAR imaging still has the following limitations: Under the assumption of satisfying the restricted isometry property (RIP) [9], CS algorithms are vital to AgileSAR imaging. At present, they

mainly include traditional optimization-based CS methods and recent network-based CS methods. The optimization-based CS methods include greedy algorithms [10,11], $\mathcal{L}_1$ -norm optimization algorithms [12,13], and Bayesian-based methods [14–16], where the $\mathcal{L}_1$ -norm optimization algorithm has better performance in terms of the recovered error evaluated by mean squared error (MSE) [17–19]. Although optimization-based CS methods have achieved better recovered performance, these methods still suffer from some problems, i.e., manual tuning and pre-definition of optimization parameters (e.g., the regularization parameter and the thresholding parameter), high computational complexity, low signal-to-noise ratio (SNR) resistance, and the prior assumption of observed scenes. With the development of deep learning techniques, some network-based CS methods are proposed to address the manual tuning issue and improve the performance and speed of signal reconstruction [20–29]. Specifically, deep unfolding networks relate optimization-based CS methods with deep neural networks so that they can provide better interpretability, and these methods have been successfully applied in SAR imaging for remote sensing.

However, the current CS methods do not take full advantage of prior scene information we may hold, and the sparse property is imposed uniformly and independently on each variable. Consequently, some low signal-to-noise ratio (SNR) targets in the sparse scene cannot be accurately recovered, often yielding false targets by the current CS methods. Although some algorithms have already been proposed [30,31], there is still no idea of how and why to select the penalized rule in AgileSAR imaging to further mitigate the impact of empirical parameter setting on reconstructed quality and computational complexity. In addition, when the observed scenes are non-sparse, CS methods with traditional fixed sparsifying transforms (e.g., DCT, wavelet [32,33]) usually cannot result in good reconstruction performance.

To address these challenges in AgileSAR imaging, a pseudo- $\mathcal{L}_0$ -norm fast iterative shrinkage algorithm network (pseudo- $\mathcal{L}_0$ -norm FISTA-net) is proposed in this paper. Firstly, a pseudo- $\mathcal{L}_0$ -norm regularization model is built based on Bayesian estimation. This model approximately fairly penalizes the regularization item with prior scene information, i.e., the reciprocal of its previous solution, to nearly acquire the number of nonzero values. Due to the fact that this model approximates a $\mathcal{L}_0$ -norm regularization equation, we name it the pseudo- $\mathcal{L}_0$ -norm regularization model. Then, we unfold the iterative updating process of FISTA into an interpretable deep network to solve the regularization model. The network and the network's parameters are automatically learnable through training in a data-driven manner, overcoming the difficulties of manually adjustable parameters and high computation complexity. Additionally, the learnable nonlinear transform sparsifies the non-sparse scenes instead of the traditional sparsifying transform to improve reconstruction performance. Finally, the quantitative analysis and numerical results for simulation are detailed in the following sections to demonstrate the evident advantage of the proposed algorithm.

The rest of this paper is organized as follows: In Section 2, the observation model for AgileSAR is established, and we propose a pseudo- $\mathcal{L}_0$ -norm FISTA-net for AgileSAR imaging and present the details of the AgileSAR imaging network. In Section 3, simulation experiments and data experiments of real TerraSAR-X images confirm the effectiveness and efficiency of our proposed method. In Section 4, a discussion is given to show the performance and advantages of our proposed algorithm. Section 5 concludes this paper.

2. Materials and Methods

In this section, the observation model of AgileSAR is described. To further improve the performance and efficiency of AgileSAR imaging, a pseudo- $\mathcal{L}_0$ -norm FISTA-net is proposed.

2.1. AgileSAR Signal Model

In the traditional HRWS system, e.g., the azimuthal multi-channel SAR [3] and MIMO SAR [4], the equivalent sampling still satisfies the Nyquist theorem. To lower the amount of

data and relieve the contradiction between high resolution and wide swath, AgileSAR with sub-Nyquist sampling frequency has been proposed and can achieve these requirements [8]. As usual, the SAR system is deemed to be stationary after transmitting one pulse and receiving an echo before the next transmission [5]. The imaging geometry of SAR is shown in Figure 1.

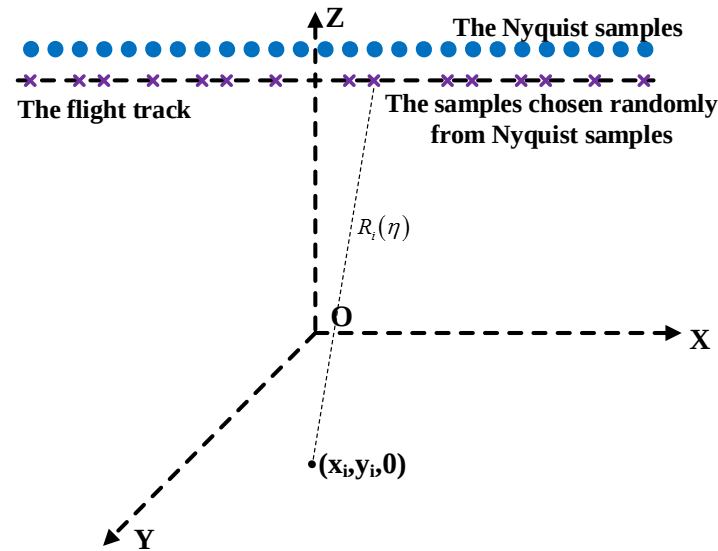


Figure 1. The imaging geometry. η is the slow time along the azimuth and $R_i(\eta)$ represents the range between the radar and the point target located at the coordinate $(x_i, y_i, 0)$ at the azimuth time η . x_i and y_i denote the azimuth and range coordinates, respectively. Because a two-dimensional image, i.e., azimuth and range, is considered, the coordinate $(x_i, y_i, 0)$ is simplified as (x_i, y_i) . • denotes the Nyquist samples. × demonstrates the real azimuthal samples chosen randomly from Nyquist samples in the AgileSAR system.

Similar to the traditional SAR system, the raw data in AgileSAR for many point targets can be written as

$$s_c(\tau, \eta) = \sum_{x_i, y_i} \sigma_i W_i(\tau, \eta) \exp\left\{j\pi K_r [\tau - 2R_i(\eta)/c]^2\right\} \cdot \exp(-j4\pi R_i(\eta)/\lambda) + n(\tau, \eta) \quad (1)$$

where τ is the fast time along the range and σ_i and $W_i(\tau, \eta)$ are the backscattering coefficient and the weighting pattern corresponding to the i -th target at (x_i, y_i) . K_r denotes the chirp rate of a linear frequency-modulated (LFM) signal, c is the light speed, λ is the wavelength, and $n(\tau, \eta)$ is the system noise.

To relieve the inherent contradiction in the HRWS system, sub-Nyquist sampling in AgileSAR is implied along the azimuthal dimension, and the sampling method is as shown in Figure 1. The AgileSAR system only adopts the traditional matched filtering (MF) method to not recover the scene exactly [5]. Accordingly, AgileSAR imaging includes three steps: (1) range compression based on MF; (2) range cell migration correction (RCMC) (interpolate by zero-padding every azimuthal signal after range compression and then sum up all the range-compressed signals of grids on the same range cell to the pre-defined grid); and (3) azimuth reconstruction with sub-Nyquist samples based on the CS algorithm. After implementing range compression and RCMC in Equation (1), the signal at a certain range cell is represented by

$$s_c(\tau_0, \eta) = \sum_{x_i, y_i} \sigma_i W_i(\tau_0, \eta) T_r \sin c\{K_r T_r (\tau_0 - 2R_i(\eta - \eta_{ci})/c)\} \times \exp\{-j4\pi R_i(\eta - \eta_{ci})/\lambda\} + n(\tau_0, \eta) \quad (2)$$

where η_{ci} is the beam center crossing time for target (x_i, y_i) . T_r is the pulse width, and $\sin c(\cdot)$ denotes the $\sin c$ function.

Let $\sigma = [\sigma_1, \sigma_2, \dots, \sigma_M]^T$ be the vectorized backscattering cross-sections of targets on the same range cell, and $\mathbf{s}_{N \times 1} = [s_c(\tau_0, \eta_1), s_c(\tau_0, \eta_2), \dots, s_c(\tau_0, \eta_N)]^T$ be the vectored signal after range compression and RCMC. This can be expressed as

$$\mathbf{s}_{N \times 1} = \mathbf{D}_{N \times M} \sigma_{M \times 1} + \mathbf{n}_{N \times 1} \quad (3)$$

where N is the sampling number on the azimuthal dimension, and M is the number of resolution cells at a certain range cell in the observed scene. $\mathbf{D}_{N \times M} = \{D_i(\tau_0, \eta_n)\}_{n=1, i=1}^{N, M}$ denotes the mapping relation between the received signal and the scene, and $D_i(\tau_0, \eta_n) = W_i(\tau_0, \eta_n) \cdot T_r \cdot \exp\{-j4\pi R_i(\eta_n - \eta_{ci})/\lambda\}$. $\mathbf{n}_{N \times 1} = [n(\tau_0, \eta_1), n(\tau_0, \eta_2), \dots, n(\tau_0, \eta_N)]^T$ is the noise.

2.2. Pseudo- $\mathcal{L}_0$ -Norm FISTA-Net for AgileSAR Imaging

In this subsection, the CS theorem is briefly introduced. Then, a pseudo- $\mathcal{L}_0$ -norm regularization model is built based on Bayesian estimation. Finally, the pseudo- $\mathcal{L}_0$ -norm FISTA-net unfolds the iterative updating process of FISTA into the deep network to solve the pseudo- $\mathcal{L}_0$ -norm regularization model for AgileSAR imaging.

2.2.1. CS Theorem

In CS, the $\mathcal{L}_1$ -norm regularization model can effectively recover nearly sparse signals with/without measurement noise from remarkably few measurements, and its application is so wide that it broadly could be considered the modern least-squares method [12,13]. The $\mathcal{L}_1$ -norm regularization model has good recovery performance in terms of the recovered error evaluated by MSE [17]. Its equation is to solve underdetermined problem (3) without subscript

$$\hat{\sigma} = \underset{\sigma}{\operatorname{argmin}} \left\{ \|\mathbf{s} - \mathbf{D}\sigma\|_2^2 + \alpha \|\sigma\|_1 \right\} \quad (4)$$

where α is the regularization parameter. The first item $\|\mathbf{s} - \mathbf{D}\sigma\|_2^2$ ensures fidelity, and the second item $\|\sigma\|_1$ guarantees the sparsity of the recovered scene. The parameter γ_1 balances between the recovered error and the sparsity and is empirically chosen by minimizing the recovered error of the whole scene.

Even though the $\mathcal{L}_1$ -norm regularization model has achieved good recovery performance, some low SNR targets cannot be accurately recovered, often yielding false targets. The further optimization of the regularization model contributes to improving the reconstruction performance. Is there an alternative to the $\mathcal{L}_1$ -norm regularization model to achieve better recovered quality? To solve this problem, a reweighted $\mathcal{L}_1$ -based algorithm has already been proposed to improve the recovered performance [30,31]. However, the selection of an approximately fair penalization rule does not have a theoretical analysis.

2.2.2. Pseudo- $\mathcal{L}_0$ -Norm Regularization Model

For the sparse undetermined equation, the $\mathcal{L}_0$ -norm regularization model has the best recovered performance, but this model is a non-polynomial hard (NP-hard) problem [12]. Inspired by this, we make full use of prior scene information to penalize the regularization item to be close to $\mathcal{L}_0$ -norm to achieve the better performance of the $\mathcal{L}_0$ -norm regularization model and make the optimization equation solvable. Due to the fact that it is closer to the $\mathcal{L}_0$ -norm regularization model in some sense, we name it the pseudo- $\mathcal{L}_0$ -norm regularization model. Subsequently, we adopt Bayesian estimation to analyze and deduce an approximately fair penalization rule. The rule is deduced as follows: Usually, the noise $\mathbf{n}_{N \times 1}$ is assumed to be a Gaussian distribution with a zero mean and variance σ_n^2 [12]. For simplicity, all the following matrixes/vectors omit subscripts.

$$p_n(n) = p_{\mathbf{s}/\sigma}(\mathbf{s}/\sigma) = \left(\frac{1}{\sqrt{2\pi}\sigma_n} \right)^N \exp \left\{ -\frac{\|\mathbf{s} - \mathbf{D}\sigma\|_2^2}{2\sigma_n^2} \right\} \quad (5)$$

The Laplace distribution promotes most coefficients to be small so that it can describe the sparse scene [32]. Assuming $\Gamma(\cdot)$ nonlinearly sparsifies the observed scene σ ,

$$p(\sigma_i) = \left(\frac{\xi_i}{2}\right) \exp\{-\xi_i \|\Gamma(\sigma_i)\|_1\} \quad (6)$$

where ξ_i is the scale parameter of the Laplace distribution, and $\xi_i > 0$. Then the probability distribution of the vector $\sigma_{M \times 1}$ is

$$p_{\sigma}(\sigma) = \prod_{i=1}^M \left[\left(\frac{\xi_i}{2}\right) \exp(-\xi_i \|\Gamma(\sigma_i)\|_1) \right] = \prod_{i=1}^M \left[\left(\frac{\xi_i}{2}\right) \right] \exp\left(-\sum_{i=1}^M \xi_i \|\Gamma(\sigma_i)\|_1\right) \quad (7)$$

Based on the Bayesian rule in information theory, the maximum a posteriori (MAP) probability of the vector $\sigma_{M \times 1}$ is

$$\hat{\sigma} = \operatorname{argmax}_{\sigma} [p_{\sigma/s}(\sigma/s)] = \operatorname{argmax}_{\sigma} [p_{s/\sigma}(s/\sigma) \cdot p_{\sigma}(\sigma)] \quad (8)$$

Taking the logarithm of the above Formula (8),

$$\hat{\sigma} = \operatorname{argmax}_{\sigma} [\log[p_{s/\sigma}(s/\sigma)] + \log[p_{\sigma}(\sigma)]] = \operatorname{argmax}_{\sigma} \left[-\frac{1}{2\sigma_n^2} \|s - D\sigma\|_2^2 - \|\xi \cdot \Gamma(\sigma)\|_1 \right] = \operatorname{argmin}_{\sigma} [\|s - D\sigma\|_2^2 + \beta \|\xi \cdot \Gamma(\sigma)\|_1] \quad (9)$$

where the reweighting matrix ξ is the diagonal matrix with $\{\xi_1, \xi_2, \dots, \xi_M\}$ on the diagonal and zeros elsewhere. $\Gamma(\cdot)$ denotes the vector–matrix form of the sparse transform $\Gamma(\cdot)$. Similar to α in Equation (4), β is the regularization parameter. After the above deduction, optimization Equation (9) seems to be a modification of the $\mathcal{L}_1$ -norm regularization model.

To solve Equation (9), the matrix ξ is first calculated. The logarithm function of the vector $\sigma_{M \times 1}$ is

$$\log[p_{\sigma}(\sigma)] = \log \left[\prod_{i=1}^M \left[\left(\frac{\xi_i}{2}\right) \right] \exp\left(-\sum_{i=1}^M \xi_i \|\Gamma(\sigma_i)\|_1\right) \right] = \log \left[\prod_{i=1}^M \left(\frac{\xi_i}{2}\right) \right] + \log \left[\exp\left(-\sum_{i=1}^M \xi_i \|\Gamma(\sigma_i)\|_1\right) \right] = \sum_{i=1}^M \log\left(\frac{\xi_i}{2}\right) - \sum_{i=1}^M \xi_i \|\Gamma(\sigma_i)\|_1 \quad (10)$$

Let the partial derivative function of Formula (10) with respect to ξ_i be equal to zero, and the estimation of the scale parameter ξ_i be

$$\frac{\partial \log[p_{\sigma}(\sigma)]}{\partial \xi_i} = \frac{1}{\xi_i} - \|\Gamma(\sigma_i)\|_1 = 0 \quad (11)$$

$$\xi_i = 1/(\|\Gamma(\sigma_i)\|_1) \quad (12)$$

If $\|\Gamma(\sigma_i)\|_1 = 0$, Formula (12) makes no sense. Formula (12) is modified as

$$\xi_i = 1/(\|\Gamma(\sigma_i)\|_1 + \iota) \quad (13)$$

where $\iota > 0$ is some very small positive constant.

The above deduction confirms the penalization rule by the prior distribution, i.e., the Laplace distribution, and this model makes full use of prior data information to achieve good performance. Certainly, different probability distributions confirm different penalization rules. In reality, without knowing the variables themselves, using the iteratively updated method to establish the approximately fair penalization rule allows successively better estimation of nonzero variables. According to optimization Equations (9) and (13), we know that the regularized item is iteratively penalized by itself reciprocally to nearly acquire the number of nonzero values, and the large coefficients are more heavily penalized to discourage their effects compared to those with small coefficients and are more likely to be identified as nonzero. Once the nonzero locations are identified, their influence attenuates in order to allow more sensitivity to identify the remaining small but nonzero elements. This means that this model can recover more target details more accurately.

2.2.3. AgileSAR Imaging Based on Pseudo- $\mathcal{L}_0$ -Norm FISTA-Net

To solve the pseudo- $\mathcal{L}_0$ -norm regularization model for AgileSAR imaging, this paper proposes a pseudo- $\mathcal{L}_0$ -norm FISTA-net by taking full advantage of the merits of optimization-based and network-based methods. The basic idea of the pseudo- $\mathcal{L}_0$ -norm FISTA-net is to map the iterative steps of FISTA to a fixed number of phases of a deep network framework, in which self-adapted parameters and sparse priors integrated with a CNN module are learnable using appropriate training data. Therefore, better reconstructed performance and higher efficiency can also be achieved due to the self-adapted optimal parameters of the deep network.

A. Network Mapping of Pseudo- $\mathcal{L}_0$ -Norm FISTA-net

To solve the regularization problem (9), FISTA accelerates convergence and consumes less time by adding momentum [34]. The pseudo- $\mathcal{L}_0$ -norm FISTA-net takes advantage of the merits of optimization-based and network-based CS methods. The basic idea of the pseudo- $\mathcal{L}_0$ -norm FISTA-net is to map the FISTA iterative steps to a deep network framework, and it is formulated as

$$\mathbf{v}^{(g)} = \boldsymbol{\sigma}^{(g-1)} - \mu^{(g)} \mathbf{D}^H (\mathbf{D} \boldsymbol{\sigma}^{(g-1)} - \mathbf{s}) \quad (14)$$

$$\boldsymbol{\xi}^{(g)} = \text{diag} \left(1 / \left(|\Gamma(v_1^{(g)})| + \iota^{(g)} \right), 1 / \left(|\Gamma(v_2^{(g)})| + \iota^{(g)} \right), \dots, 1 / \left(|\Gamma(v_M^{(g)})| + \iota^{(g)} \right) \right) \quad (15)$$

$$\boldsymbol{\omega}^{(g)} = \Gamma^{-1} \left(\left(\boldsymbol{\xi}^{(g)} \right)^{-1} \cdot \Theta \left(\boldsymbol{\xi}^{(g)} \cdot \Gamma(\mathbf{v}^{(g)}), \boldsymbol{\delta}^{(g)} \right) \right) \quad (16)$$

$$\boldsymbol{\sigma}^{(g)} = \boldsymbol{\omega}^{(g)} + \gamma^{(g)} \left(\boldsymbol{\omega}^{(g)} - \boldsymbol{\omega}^{(g-1)} \right) \quad (17)$$

where $\Theta(\sigma_i, \delta) = \text{sign}(\sigma_i) \max\{|\sigma_i - \delta|, 0\}$ is the soft threshold function. $\mu^{(g)}$, $\iota^{(g)}$, $\Gamma(\cdot)$, $\boldsymbol{\delta}^{(g)}$, and $\gamma^{(g)}$ are learnable variables of the deep unfolding network. Pseudo- $\mathcal{L}_0$ -norm FISTA-net alternates among the $\mathbf{v}^{(g)}$ module from one-step gradient descent, the $\boldsymbol{\xi}^{(g)}$ module, the $\boldsymbol{\omega}^{(g)}$ module, and the $\boldsymbol{\sigma}^{(g)}$ module. It generalizes the four types of operations to have learnable parameters as network layers. Given the undersampled data $\mathbf{s}$, they flow over all modules and finally generate a reconstructed image $\boldsymbol{\sigma}$. Figure 2 illustrates the overall framework of the pseudo- $\mathcal{L}_0$ -norm FISTA-net, and more details are provided in the following paragraphs.

- (1) $\mathbf{v}^{(g)}$ Module: This layer achieves an estimation by solving the standard convex quadratic program $\|\mathbf{s} - \mathbf{D}\boldsymbol{\sigma}\|_2^2$ and updates the reconstructed image based on the gradient descent operation of the closed-form solution Equation (14), using the output of the previous layer $\boldsymbol{\sigma}^{(g-1)}$ as the input for the current layer. Additionally, the weight $\mu^{(g)}$ is learned through end-to-end training rather than being fixed.
- (2) $\boldsymbol{\xi}^{(g)}$ Module: This layer takes advantage of prior scene information $\mathbf{v}^{(g)}$ held in the previous layer. Bayesian estimation analyzes and deduces an approximately fair penalization rule $\xi_i = 1 / (|\Gamma(v_i)| + \iota)$, which explains the method of applying prior scene information. This reweighting method penalizes the regularization item to be close to $\mathcal{L}_0$ -norm to achieve good recovered performance of the $\mathcal{L}_0$ -based algorithm; thus, it can be beneficial for the recovery of low SNR targets.
- (3) $\boldsymbol{\omega}^{(g)}$ Module: The proximal operator aims to remove the noise and false targets of the previous layer $\mathbf{v}^{(g)}$ through thresholding in the sparse transform domain. Complex image details are captured by fine-tuning existing sparse transformations. The pseudo- $\mathcal{L}_0$ -norm FISTA-net aims to learn a more flexible representation $\Gamma(\cdot)$ and threshold $\boldsymbol{\delta}^{(g)}$ from the training data.

Inspired by the powerful representation power of CNNs and their universal approximation property, we design a general nonlinear sparse transform function $\Gamma(\cdot)$. In the pseudo- $\mathcal{L}_0$ -norm FISTA-net, the CNN network module is adopted to replace $\Gamma(\cdot)$, and it is a combination of three linear convolutional operators (without bias terms) separated by

a Rectified Linear Unit (ReLU), as shown in the dashed box in Figure 2. The first convolutional operator and the last convolutional operator both correspond to 1 filter (each of size 3×3), and the second convolutional operator corresponds to 32 filters (each of size 3×3). The first convolution layer corresponds to one filter in order to adapt single-channel data. Mathematically, the sparse transform $\Gamma(\cdot)$ is invertible, i.e., $\Gamma(\cdot) \cdot \Gamma^{-1}(\cdot) = \mathbf{I}$, where $\mathbf{I}$ is the identity operator. A symmetric network $\Gamma^{-1}(\cdot)$ is denoted by a mirror-symmetrical framework. Therefore, $\omega^{(g)}$ can be efficiently computed by (16). The parameters of $\Gamma(\cdot)$ are shared throughout different iterations. As the processing of the iterations continues, $\delta^{(g)}$, a learnable shrinkage thresholding parameter, varies at each iteration. The advantage of this setting is that it can maintain the flexibility to self-adapt to the noise at each iteration.

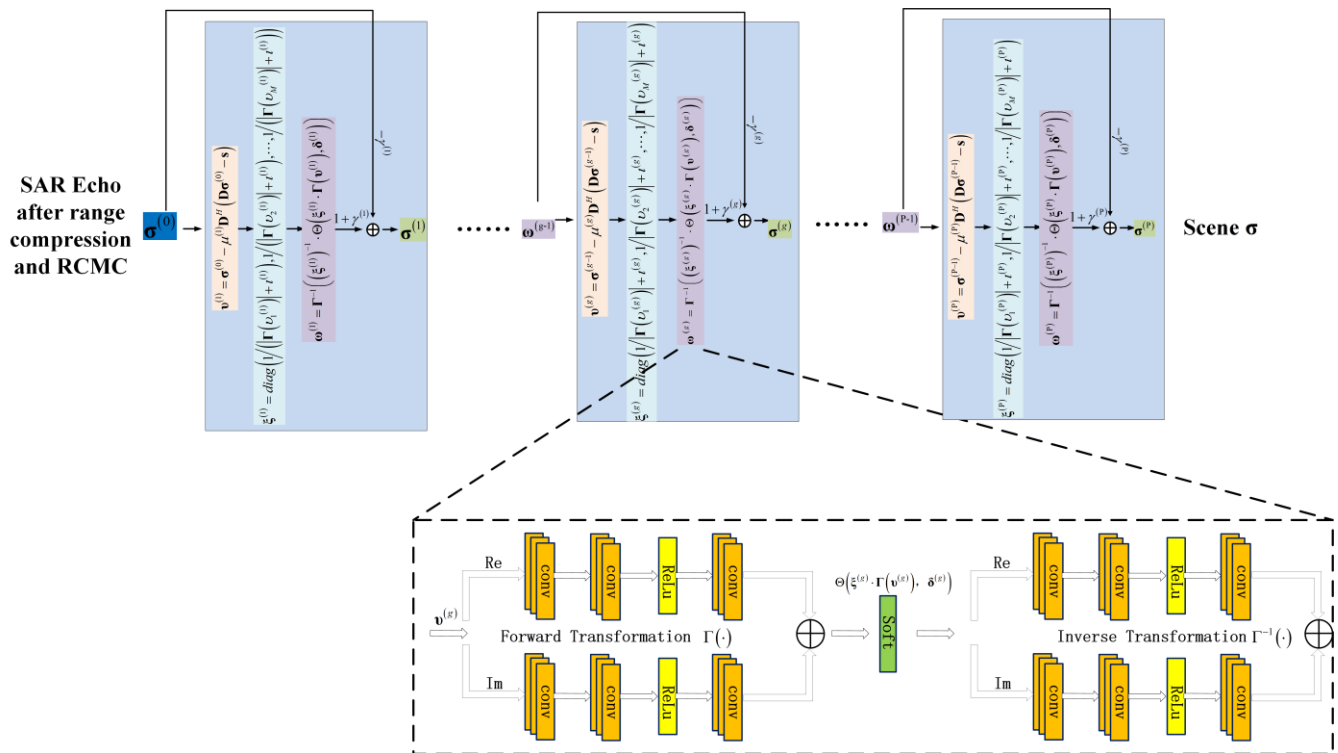


Figure 2. The overall framework of the proposed pseudo- $\mathcal{L}_0$ -norm FISTA-net. Specifically, the pseudo- $\mathcal{L}_0$ -norm FISTA-net consists of four main modules, i.e., gradient descending, reweight updating, proximal mapping, and momentum updating.

Additionally, the scene reconstructed from the echo is a complex-valued recovery, so we need to extend the pseudo- $\mathcal{L}_0$ -norm FISTA-net to the complex-valued domain. The real-imaginary individual model takes the basic operators for complex-valued data into consideration, especially for multiplication operators; they can be converted into matrix operators based on the real and imaginary parts, respectively, i.e., for complex multiplication $\mathbf{s} = \mathbf{D}\boldsymbol{\sigma} + \mathbf{n}$. We establish the following data structures for echo $\mathbf{s}$ and imaging results $\boldsymbol{\sigma}$, which simply doubles the data's dimensional size by separating the real and imaginary parts of the original data, as follows:

$$\mathbf{s}_n = \begin{bmatrix} (\mathbf{s}_n)^{real} \\ (\mathbf{s}_n)^{imag} \end{bmatrix}, \boldsymbol{\sigma}_m = \begin{bmatrix} (\boldsymbol{\sigma}_m)^{real} \\ (\boldsymbol{\sigma}_m)^{imag} \end{bmatrix} \quad (18)$$

As for the measurement matrix $\mathbf{D}$, we adopt the modification as

$$\mathbf{D}_{nm} = \begin{bmatrix} (\mathbf{D}_{nm})^{real} & -(\mathbf{D}_{nm})^{imag} \\ (\mathbf{D}_{nm})^{imag} & (\mathbf{D}_{nm})^{real} \end{bmatrix} \quad (19)$$

where $\mathbf{D}_{nm}$ is an element of $\mathbf{D}$. With the modifications above, we can effectively convert complex-valued problems into a real-imaginary individual model that can be expressed as

$$\begin{bmatrix} \mathbf{s}^{real} \\ \mathbf{s}^{imag} \end{bmatrix} = \begin{bmatrix} \mathbf{D}^{real} & -\mathbf{D}^{imag} \\ \mathbf{D}^{imag} & \mathbf{D}^{real} \end{bmatrix} \begin{bmatrix} \boldsymbol{\sigma}^{real} \\ \boldsymbol{\sigma}^{imag} \end{bmatrix} + \begin{bmatrix} \mathbf{n}^{real} \\ \mathbf{n}^{imag} \end{bmatrix} \quad (20)$$

Additionally, the cross-item between the real and imaginary components of the complex-valued signal was considered in (20). With the modifications above, we can effectively convert complex-valued problems into real-valued ones. Thus, the pseudo- $\mathcal{L}_0$ -norm FISTA-net can be applied to microwave imaging problems.

- (4) $\boldsymbol{\sigma}^{(g)}$ Module: In this layer, a momentum term, which chooses the update weights of two previous results, is introduced to accelerate the convergence rate. Simultaneously, this advantage is further optimized by an autonomously learnable parameter $\gamma^{(g)}$.

B. Network-Based Parameter Constraint

Although $\mu^{(g)}$, $\delta^{(g)}$, and $\gamma^{(g)}$ are learnable and there are no manual parameters in a pseudo- $\mathcal{L}_0$ -norm FISTA-net, extra constraints are introduced to ensure convergence properly. However, non-positive step size and thresholding values contradict the definition of these variables during iteration in an ISTA-net. Therefore, $\mu^{(g)}$, $\delta^{(g)}$, and $\gamma^{(g)}$ are constrained to be positive. Additionally, the gradient step $\mu^{(g)}$ decays smoothly with iterations. The thresholding value $\delta^{(g)}$ also iteratively decreases because noise variance is suppressed progressively with iterations. The two-step update weight $\gamma^{(g)}$ should increase monotonously, corresponding to the two-step update weight in the FISTA. Therefore, we define

$$\mu^{(g)} = \ln(1 + \exp(a_1 g + b_1)), a_1 < 0 \quad (21)$$

$$\theta^{(g)} = \ln(1 + \exp(a_2 g + b_2)), a_2 < 0 \quad (22)$$

$$\gamma^{(g)} = \frac{\ln(1 + \exp(a_3 g + b_3)) - \ln(1 + \exp(a_3 + b_3))}{\ln(1 + \exp(a_3 g + b_3))}, a_3 > 0 \quad (23)$$

One benefit of $\ln(1 + \exp(\cdot))$ is its simple derivative function.

The loss function is an index to quantify the effects of the network, and its choice depends on many factors, e.g., the choice of objective function and the efficiency of running gradient descent. In this paper, reconstruction error and sparsity constraint were considered in the loss function, and adjustable parameters κ were adopted to control the tradeoff between two constrained terms. With the training data $\{\boldsymbol{\sigma}, \mathbf{s}\}$ and the learnable parameters $\{\mu^{(g)}, \iota^{(g)}, \Gamma(\cdot), \delta^{(g)}, \gamma^{(g)}\}_{g=1}^{N_g}$, the global loss function

$$Loss\{\mu^{(g)}, \iota^{(g)}, \Gamma(\cdot), \delta^{(g)}, \gamma^{(g)}\} = \sum_{g=1}^{N_{data}} \left(\|\hat{\boldsymbol{\sigma}} - \boldsymbol{\sigma}\|_2^2 + \kappa \sum_g^{N_g} \left\| \Gamma^{-1}\left(\Gamma(\mathbf{v}^{(g)})\right) - \mathbf{v}^{(g)} \right\|_2^2 \right) \quad (24)$$

where $\|\cdot\|_2$ represents the $\mathcal{L}_2$ -norm; N_{data} is the number of training samples; N_g , a learnable parameter, is the total number of network phases; κ is the weight parameter for two different constrained factors; and $\hat{\boldsymbol{\sigma}}$ and $\boldsymbol{\sigma}$ represent the SAR image recovered with undersampled and Nyquist-sampled echoes, respectively. Once the loss function has been determined, the parameters can be learned automatically.

C. Initialization

Initialization near the minimum of the objective function can accelerate convergence. For $\boldsymbol{\sigma}^{(g)}$, $\hat{\boldsymbol{\sigma}}^0 = \mathbf{D}^H \mathbf{s}$. The parameters $\{a_1, b_1, a_2, b_2, a_3, b_3\}$ and the weight parameter κ are initialized as $\{-0.5, -0.2, 1, -2, -1, 0\}$ and 0.1, respectively.

D. Implementation Details

The unknown parameters of the pseudo- $\mathcal{L}_0$ -norm FISTA-net are $\mu^{(g)}$, $\iota^{(g)}$, $\Gamma(\cdot)$, $\delta^{(g)}$, and $\gamma^{(g)}$. The layer number of the network is fixed while conducting end-to-end training. Since the networks are shared throughout and $\{a_1, b_1, a_2, b_2, a_3, b_3\}$ are decoupled with iteration, we may choose a different number of iterations for reconstruction. The networks were implemented in Pycharm with the Pytorch 2.0.0 library, and the training of the pseudo- $\mathcal{L}_0$ -norm FISTA-net was performed on a workstation with NVIDIA RTX 3090Ti GPUs.

Once the network has been built, the training dataset can be used to learn and train the network. The input, output, and tagged data for the network are the undersampled raw data after RCMC, the SAR image, and the SAR image reconstructed with Nyquist-sampled echoes, respectively. Given a training dataset and a certain number of network layers, the network can automatically adjust the learnable parameters through training to obtain the reconstructed scene.

3. Experiment

In this section, we validate the performance of the pseudo- $\mathcal{L}_0$ -norm FISTA-net for AgileSAR imaging through simulation experiments and data experiments.

3.1. Data Description

For SAR-observed scenes, both training and testing datasets are obtained with SAR images from simulations and the TerraSAR-X satellite. A total of 500 SAR images σ , each with a size of 256×256 , are used to simulate echo data s and selected for training. Additionally, three SAR images are used for testing. Their sizes are 256×256 , 1601×4141 , and 1024×2048 . During testing, the three test images are first straightened and preprocessed to the corresponding size for the training model. The training and testing datasets consist of simulated sparse scenes, real sparse SAR scenes, and real complex SAR scenes.

In order to verify the validity and effectiveness of AgileSAR imaging based on the pseudo- $\mathcal{L}_0$ -norm FISTA-net, the reflectivity functions of the images were used to simulate raw data according to the simulated parameters, and then the raw data were randomly received, as shown in Figure 1. The results obtained with the Nyquist-sampled echo after chirp scaling algorithm (CSA) were used as labels to train the proposed network.

3.2. Evaluation Index

To quantify the performance of the pseudo- $\mathcal{L}_0$ -norm FISTA-net, the recovered performance can be evaluated by normalized MSE (NMSE), structural similarity index (SSIM), and peak signal-to-noise ratio (PSNR) in this paper. NMSE is defined as the ratio of the error energy to the true signal energy

$$\text{NMSE} = \frac{\sum \|\hat{\sigma} - \sigma\|_2^2}{\sum \|\sigma\|_2^2} \quad (25)$$

where $\hat{\sigma}$ is the recovered scene. σ is achieved by sampling on the uniform grid for reference. After performing the imaging processing procedure, NMSE can also be acquired. SSIM compares local patterns of pixel intensities that have been normalized for luminance and contrast [35]. In SSIM, the product of three components of similarity between the reference scene σ and the recovered scene $\hat{\sigma}$ is computed to estimate the image's local quality, as follows:

$$\text{SSIM}(\hat{\sigma}, \sigma) = [l(\hat{\sigma}, \sigma)]^{s_1} \cdot [c(\hat{\sigma}, \sigma)]^{s_2} \cdot [s(\hat{\sigma}, \sigma)]^{s_3} \quad (26)$$

where $l(\hat{\sigma}, \sigma)$, $c(\hat{\sigma}, \sigma)$, and $s(\hat{\sigma}, \sigma)$ are luminance, contrast, and structural similarity. s_1 , s_2 , and s_3 are positive constants used to adjust the relative importance of these three components. A simple setting $s_1 = s_2 = s_3 = 1$ is adopted in SSIM.

The PSNR is a performance index of the quality of a reconstructed image relative to the label image, which is described as

$$\text{PSNR} = 10 * \log_{10} \frac{\max(\hat{\sigma}^2)}{\|\hat{\sigma} - \sigma\|_2^2} \quad (27)$$

A higher PSNR value indicates that the reconstructed image is of higher quality and closer to the label image. The lower the MMSE, the larger the SSIM, and the higher the PSNR are, the better the image quality is.

3.3. Experimental Results

3.3.1. AgileSAR Imaging under Different Optimization Algorithms

To verify the effectiveness and superiority of the pseudo- $\mathcal{L}_0$ -norm FISTA-net for AgileSAR imaging, we conduct experiments in one simulated SAR scene and two real SAR scenes, and the proposed method is compared with the CSA, an optimization-based algorithm. The simulated parameters are shown in Table 1. The reconstruction results are illustrated in Figures 3–5, and the quantitative index analysis is in Table 2. In each reconstructed result, (a) is the result reconstructed by CSA with Nyquist sampling raw data, (b) is the result reconstructed by the orthogonal matching pursuit (OMP) algorithm with sub-Nyquist sampling raw data in AgileSAR, (c) is the result reconstructed by the $\mathcal{L}_1$ -norm optimization algorithm with sub-Nyquist sampling raw data in AgileSAR, (d) is the result reconstructed by the Bayesian-based algorithm with sub-Nyquist sampling raw data in AgileSAR, and (e) is the result reconstructed by the pseudo- $\mathcal{L}_0$ -norm FISTA-net with sub-Nyquist sampling raw data in AgileSAR. The recovered algorithms for results (b), (c), and (d) belong to the optimization-based algorithm. Additionally, (a) is used as the label.

Table 1. Simulated parameters.

Parameter	Data
Average PRF (Hz) in the AgileSAR	381
PRF (Hz) in the traditional SAR	1907
Range sampling frequency (MHz)	40
The referred slant range (km)	888
The pulse width (us)	50
The doppler bandwidth (Hz)	1401
Wavelength (mm)	5.55
Velocity (m/s)	7513
Height (km)	693
The squint angle (°)	0

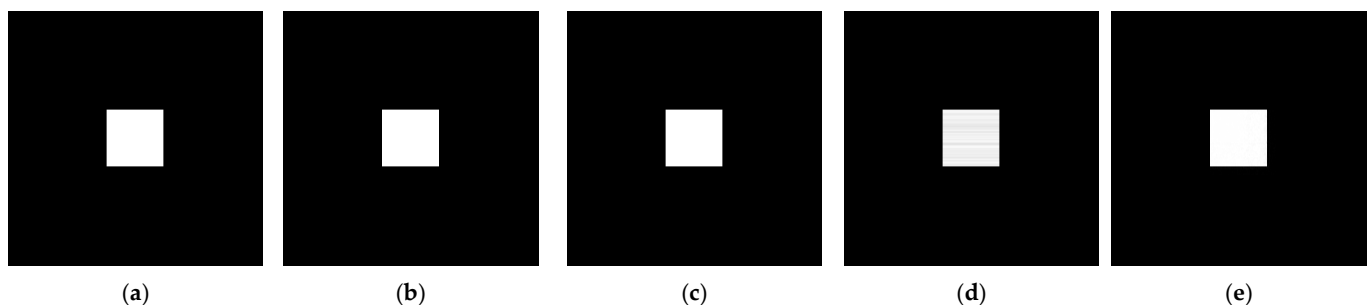


Figure 3. The reconstructed results: (a) the result reconstructed by RDA with Nyquist sampling raw data; (b) the result reconstructed by the OMP algorithm with 20% sub-Nyquist sampling raw data in AgileSAR; (c) the result reconstructed by the $\mathcal{L}_1$ -norm optimization algorithm with 20% sub-Nyquist sampling raw data in AgileSAR; (d) the result reconstructed by the Bayesian-based algorithm with 20% sub-Nyquist sampling raw data in AgileSAR; (e) the result reconstructed by the pseudo- $\mathcal{L}_0$ -norm FISTA-net with 20% sub-Nyquist sampling raw data in AgileSAR.

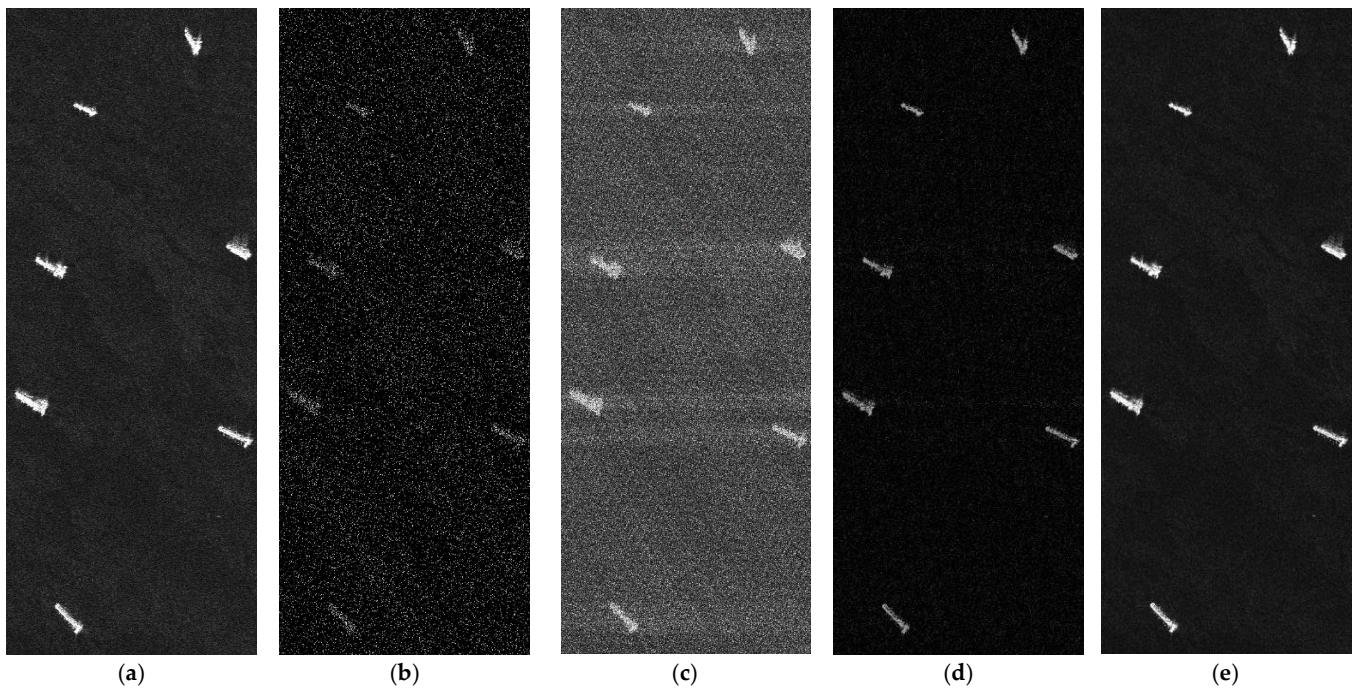


Figure 4. The reconstructed results: (a) the result reconstructed by RDA with Nyquist sampling raw data; (b) the result reconstructed by the OMP algorithm with 20% sub-Nyquist sampling raw data in AgileSAR; (c) the result reconstructed by the $\mathcal{L}_1$ -norm optimization algorithm with 20% sub-Nyquist sampling raw data in AgileSAR; (d) the result reconstructed by the Bayesian-based algorithm with 20% sub-Nyquist sampling raw data in AgileSAR; (e) the result reconstructed by the pseudo- $\mathcal{L}_0$ -norm FISTA-net with 20% sub-Nyquist sampling raw data in AgileSAR.

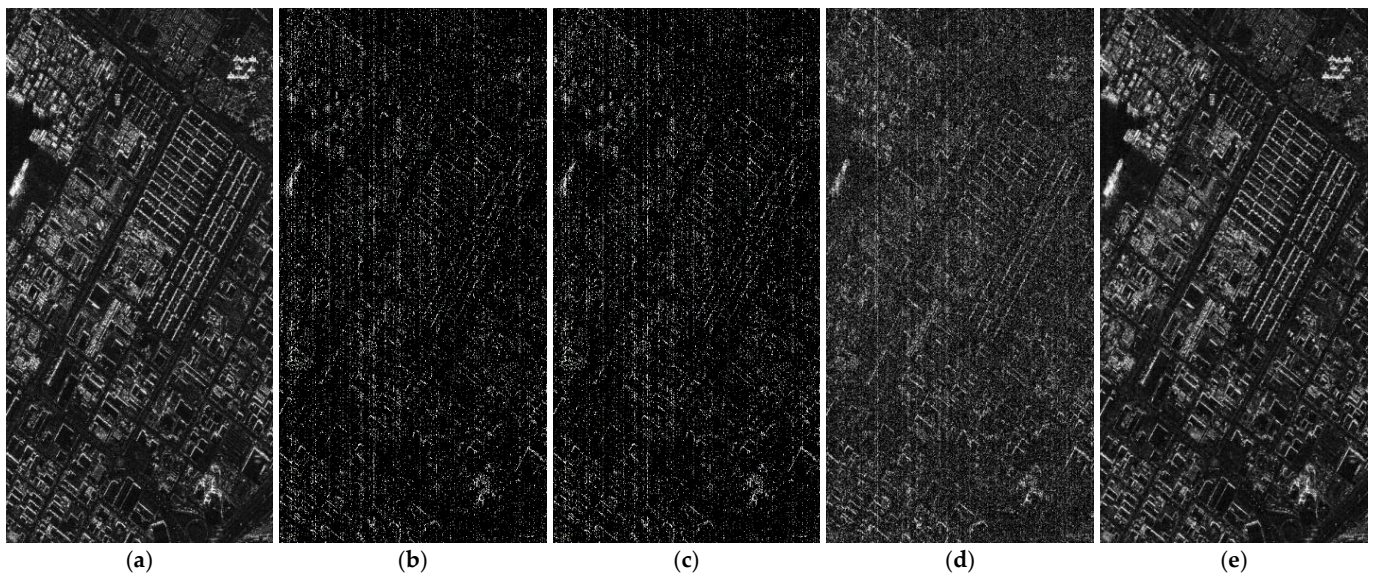


Figure 5. The reconstructed results: (a) the result reconstructed by RDA with Nyquist sampling raw data; (b) the result reconstructed by the OMP algorithm with 20% sub-Nyquist sampling raw data in AgileSAR; (c) the result reconstructed by the $\mathcal{L}_1$ -norm optimization algorithm with 20% sub-Nyquist sampling raw data in AgileSAR; (d) the result reconstructed by the Bayesian-based algorithm with 20% sub-Nyquist sampling raw data in AgileSAR; (e) the result reconstructed by the pseudo- $\mathcal{L}_0$ -norm FISTA-net with 20% sub-Nyquist sampling raw data in AgileSAR.

Table 2. Performance indexes under different algorithms.

	Different Algorithms	NMSE	SSIM	PSNR (dB)	Time (s)
Figure 3	Optimization-based algorithms	OMP	0.0042	0.9850	36.86
		$\mathcal{L}_1$ -norm optimization	0.0042	0.9850	36.86
		Bayesian-based	0.0041	0.9850	36.91
	Pseudo- $\mathcal{L}_0$ -norm FISTA-net algorithm		0.0001	0.9957	52.88
Figure 4	Optimization-based algorithms	OMP	1.0046	0.0685	17.23
		$\mathcal{L}_1$ -norm optimization	0.7913	0.1998	18.27
		Bayesian-based	0.7249	0.1065	18.65
	Pseudo- $\mathcal{L}_0$ -norm FISTA-net algorithm		0.0216	0.8980	33.91
Figure 5	Optimization-based algorithms	OMP	0.8674	0.0263	11.39
		$\mathcal{L}_1$ -norm optimization	0.8129	0.0529	11.67
		Bayesian-based	0.7539	0.0842	11.99
	Pseudo- $\mathcal{L}_0$ -norm FISTA-net algorithm		0.0615	0.7770	22.88

For the sparse scenes in Figures 3 and 4, both the optimization-based algorithm and the pseudo- $\mathcal{L}_0$ -norm FISTA-net reconstruct the scene exactly. For the complex scene in Figure 5, the pseudo- $\mathcal{L}_0$ -norm FISTA-net can reconstruct the whole scene, while the optimization-based algorithm only focuses on the outline.

In the comparison of optimization-based algorithms, for the sparse scenes, i.e., the first scene and the second scene, the $\mathcal{L}_1$ -norm optimization algorithm has better performance indexes but a high computation time. For the complex scene, i.e., the third scene, the slight difference in indexes in Table 2 does not demonstrate the recovered performance of different optimization-based algorithms because the three optimization-based algorithms do not exactly recover the whole scene.

In the comparison of optimization-based algorithms and the Bayesian-based algorithm, the proposed method exhibits lower NMSE, higher SSIM, and PSNR values for the three SAR images in Table 2. This indicates that the pseudo- $\mathcal{L}_0$ -norm FISTA-net can reconstruct the scene more accurately. Additionally, the imaging time after training decreases, and it significantly improves the efficiency of AgileSAR imaging. The imaging results illustrate that, under appropriate training scenes, the pseudo- $\mathcal{L}_0$ -norm FISTA-net presents a good ability to preserve target details. The fidelity of the imaging results is evident as target details are maintained. This suggests the potential of the proposed imaging method for preserving details, which offers promising prospects for high-resolution wide-swath imaging applications.

3.3.2. AgileSAR Imaging under Different Undersampling Ratios

To evaluate the performance under different undersampling ratios, we train our pseudo- $\mathcal{L}_0$ -norm FISTA-net under undersampling ratios {15%, 20%, 30%} and then verify the performance for three testing SAR images. The performance indexes under different ratios are illustrated in Table 3. The performance indexes demonstrate that the recovered results perform better as the undersampling ratios increase. The undersampling ratio of the data used in the experiment in Figures 3–5 is 20%, as illustrated by the parameters in Table 1. The average PRF in the AgileSAR is 381 Hz, and the PRF in the traditional SAR is

1907 Hz, so the undersampling ratio is 20%. Considering the limited space, we only display the recovered results with a 20% undersampling ratio in Figures 3–5.

Table 3. Performance indexes under different undersampling ratios.

	Algorithm	Undersampling Ratios	NMSE	SSIM	PSNR (dB)
The first scene	Pseudo- $\mathcal{L}_0$ -norm FISTA-net algorithm	30%	0.0001	0.9967	53.89
		20%	0.0001	0.9957	52.88
		15%	0.0135	0.9654	31.76
The second scene	Pseudo- $\mathcal{L}_0$ -norm FISTA-net algorithm	30%	0.0126	0.9411	36.23
		20%	0.0216	0.8980	33.91
		15%	0.0980	0.5457	27.34
The third scene	Pseudo- $\mathcal{L}_0$ -norm FISTA-net algorithm	30%	0.0446	0.8308	24.27
		20%	0.0615	0.7770	22.88
		15%	0.2297	0.4286	17.16

3.3.3. AgileSAR Imaging under Different Phase Numbers

To evaluate the effect of the number of layers, we train the network with different phase numbers ranging from 5 to 15 and then verify the performance on the testing set. Figure 6 curves illustrate the quantitative evaluation results of the pseudo- $\mathcal{L}_0$ -norm FISTA-net on the testing set: Figure 6a shows the curves of the performance indexes for the first scene with respect to different phase numbers, Figure 6b shows the curves of the performance indexes for the second scene with respect to different phase numbers, and Figure 6c shows the curves of the performance indexes for the third scene with respect to different phase numbers. The curves demonstrate that the performance indexes improve when the number of layers is six and fluctuate when the number of layers is larger than six. GPU computational cost increases as the number of layers increases. Thus, considering the tradeoff between GPU computation cost and recovered performance, the six-layer configuration is a preferable setting for our experiments.

3.3.4. AgileSAR Imaging under Different Epoch Numbers

The iterative number of epochs is set to 200 as the basis for the iterative stopping in our manuscript. To evaluate the effect of the number of epochs, we train the network with different epochs ranging from 40 to 200 and then verify the performance on the testing set. Figure 7 curves illustrate the quantitative evaluation results of the pseudo- $\mathcal{L}_0$ -norm FISTA-net on the testing set: Figure 7a shows the curves of the performance indexes for the first scene with respect to different epoch numbers, Figure 7b shows the curves of the performance indexes for the second scene with respect to different epoch numbers, and Figure 7c shows the curves of the performance indexes for the third scene with respect to different epoch numbers. The curves demonstrate that the performance indexes improve when the epoch number is 100 and slightly fluctuate when the epoch number is larger than 100. The network can achieve better recovered performance when trained for a larger number of epochs. Thus, considering the recovered performance, the 200-epoch configuration is a preferable setting for our experiments.

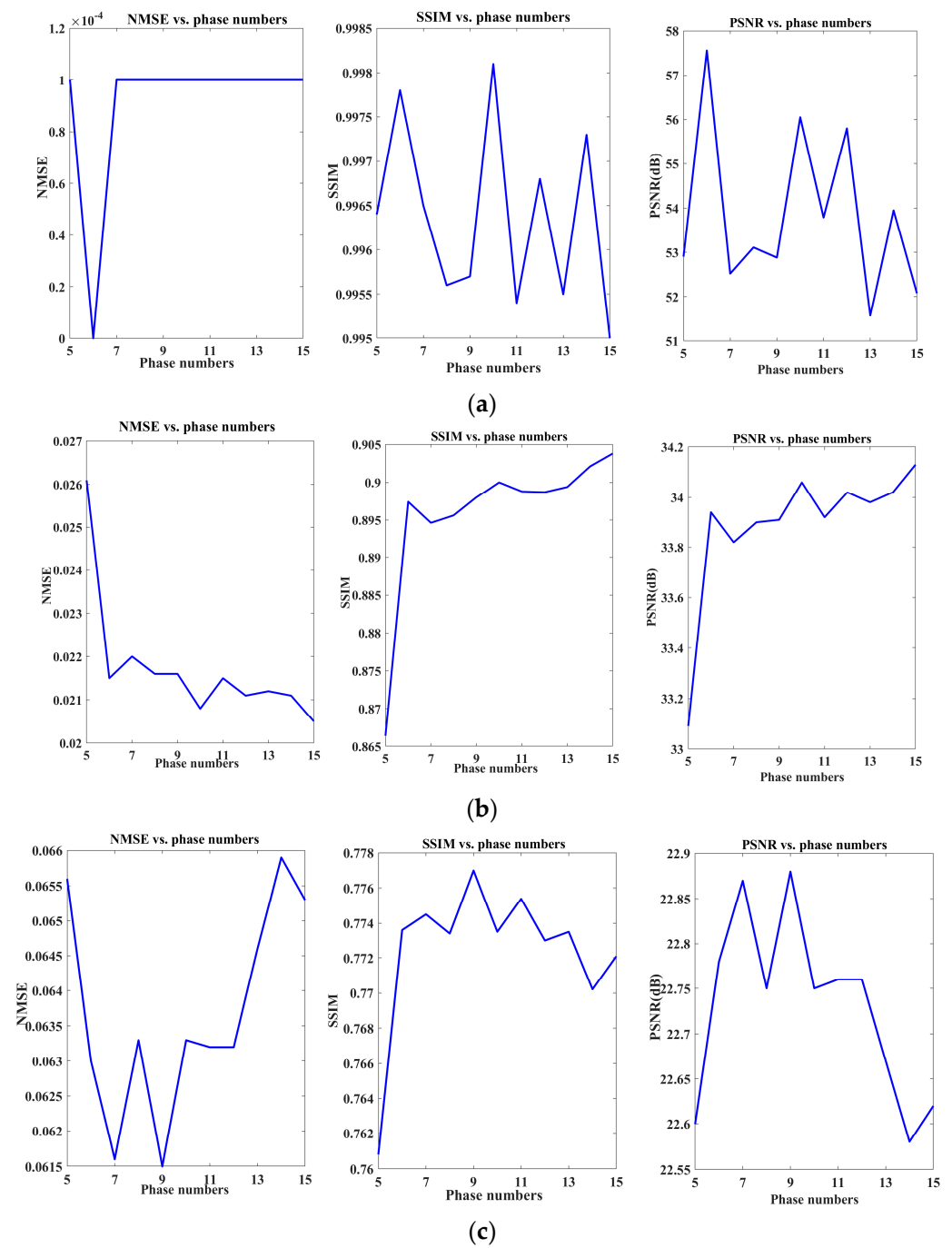


Figure 6. The curves (a), (b) and (c) of performance indexes for the first, second and third testing scenes with respect to different phase numbers, respectively.

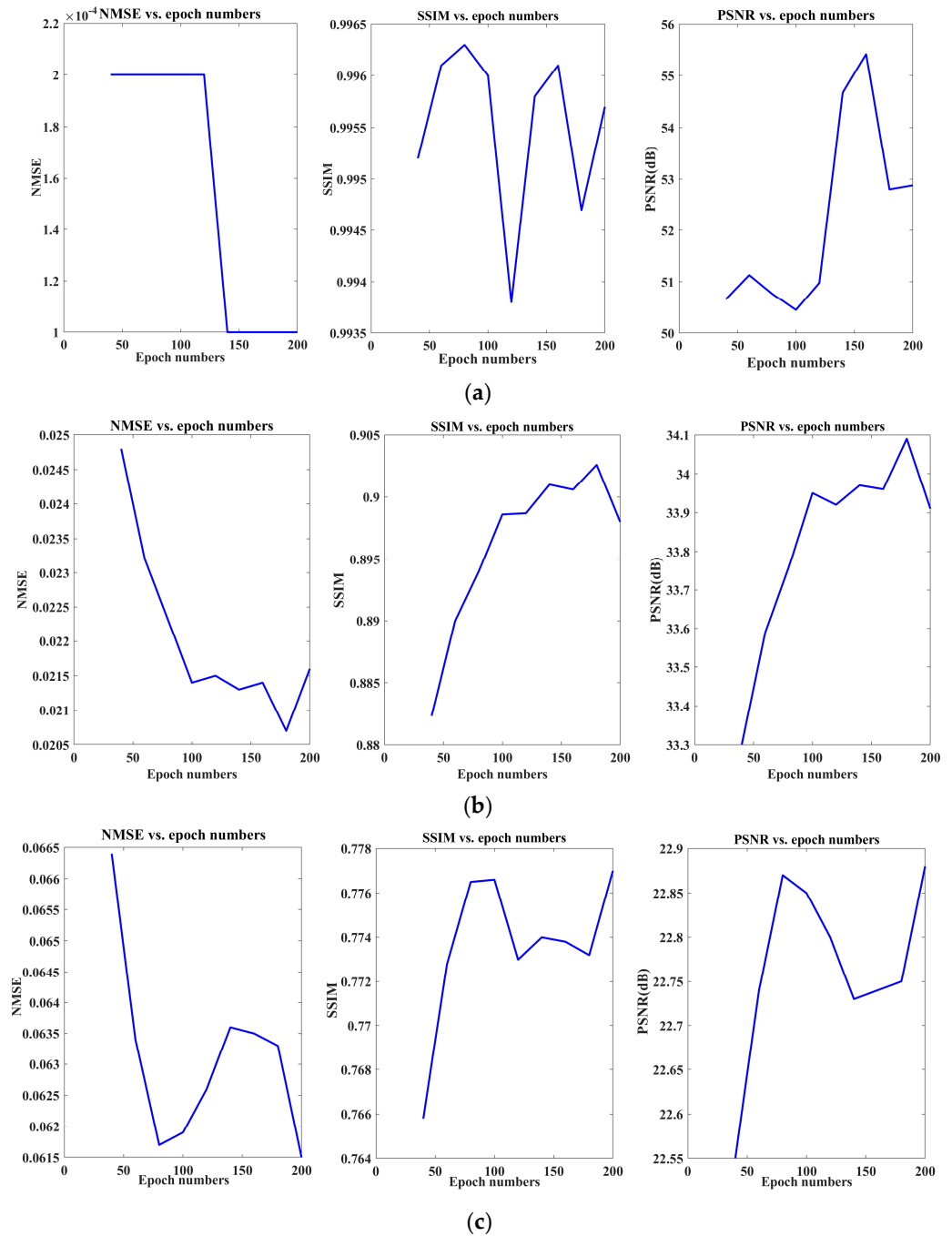


Figure 7. The curves (a), (b) and (c) performance indexes for the first, the second and the third testing scenes with respect to different epoch numbers, respectively.

4. Discussion

HRWS SAR imaging has always been the goal of spaceborne SAR systems in remote sensing applications [2]. Because high resolution and wide swath are inherently conflicting requirements, the HRWS SAR system brings new difficulties and challenges. These requirements are simultaneously satisfied by advanced observing modes, e.g., the azimuthal multi-channel SAR [3] and MIMO SAR [4], with both of these observing modes having the characteristics of large amounts of data and a long antenna. As the CS theorem develops, a novel imaging mode named AgileSAR has been proposed without large amounts of data and a long antenna for the sparse scene, and AgileSAR imaging based on CS algorithms can recover the sparse scene [8].

Although the $\mathcal{L}_1$ -norm regularization model in CS algorithms has achieved quite good performance, low SNR targets are not accurately recovered, and the model may yield false targets. To further improve the recovered performance of AgileSAR imaging, we propose a pseudo- $\mathcal{L}_0$ -norm FISTA-net. Firstly, this algorithm presents a pseudo- $\mathcal{L}_0$ -norm regularization model, which penalizes the regularization item with the reciprocal of its previous solution to nearly acquire the number of nonzero values. This penalized rule is approximately fair and makes full use of prior data information. Then, it unfolds the iterative process of FISTA into the deep network to solve the pseudo- $\mathcal{L}_0$ -norm regularization model so that it improves the efficiency and accuracy of AgileSAR imaging. Additionally, the learnable sparsifying transform in this network can reconstruct more target details instead of the traditional sparsifying transform. For the recovered error, we deduce the expression of MSE as follows: Assume that $\mathbf{F}_{N \times S}$ is the submatrix constructed by taking S columns from the recovered matrix $\mathbf{D}_{N \times M}$, which are specified by index vector $\mathbf{\Lambda}$, and each element of $\mathbf{\Lambda}$ satisfies the following:

$$\sigma_{\Lambda_l} \neq 0, (l = 1, 2, \dots, S) \quad (28)$$

The estimated covariance matrix for the nonzero components σ_l is

$$\mathbf{C} = \left(\mathbf{F}_{N \times S}^H \mathbf{F}_{N \times S} + \gamma_2 \Sigma(\sigma_{\Lambda_l}) \right)^{-1} \mathbf{F}_{N \times S}^H \mathbf{F}_{N \times S} \left(\mathbf{F}_{N \times S}^H \mathbf{F}_{N \times S} + \gamma_2 \Sigma(\sigma_{\Lambda_l}) \right)^{-1} \quad (29)$$

where $\Sigma(\sigma_{\Lambda_l}) = \text{diag}(\xi_{\Lambda_1}/\sigma_{\Lambda_1}, \xi_{\Lambda_2}/\sigma_{\Lambda_2}, \dots, \xi_{\Lambda_S}/\sigma_{\Lambda_S})$ is a diagonal matrix. The estimated error is the trace of the covariance matrix $\mathbf{C}$

$$\begin{aligned} \|\hat{\sigma}_{\Lambda} - \sigma_{\Lambda}\|_2^2 &= \sigma_n^2 \text{trace}(\mathbf{C}) \\ &= \sigma_n^2 \text{trace} \left(\left(\mathbf{F}_{N \times S}^H \mathbf{F}_{N \times S} + \gamma_2 \Sigma(\sigma_{\Lambda_l}) \right)^{-1} \mathbf{F}_{N \times S}^H \mathbf{F}_{N \times S} \left(\mathbf{F}_{N \times S}^H \mathbf{F}_{N \times S} + \gamma_2 \Sigma(\sigma_{\Lambda_l}) \right)^{-1} \right) \\ &= \sigma_n^2 \text{trace} \left(\left(\left(\mathbf{F}_{N \times S}^H \mathbf{F}_{N \times S} + \gamma_2 \Sigma(\sigma_{\Lambda_l}) \right)^{-1} \right)^2 \mathbf{F}_{N \times S}^H \mathbf{F}_{N \times S} \right) \\ &\geq \sigma_n^2 \left[2 \text{trace} \left(\left(\mathbf{F}_{N \times S}^H \mathbf{F}_{N \times S} + \gamma_2 \Sigma(\sigma_{\Lambda_l}) \right)^{-1} \right) - \text{trace} \left(\left(\mathbf{F}_{N \times S}^H \mathbf{F}_{N \times S} \right)^{-1} \right) \right] \end{aligned} \quad (30)$$

Assuming that $\mathbf{A} = \mathbf{F}_{N \times S}^H \mathbf{F}_{N \times S}$ and $\mathbf{B} = \mathbf{F}_{N \times S}^H \mathbf{F}_{N \times S} + \gamma_2 \Sigma(\sigma_{\Lambda_l})$, their eigenvalues are $\lambda_1, \lambda_2, \dots, \lambda_S$ and $\alpha_1, \alpha_2, \dots, \alpha_S$, respectively. Based on the structure form of the recovered matrix $\mathbf{D}_{N \times M}$, the $\mathcal{L}_2$ -norm of each column is nearly equal, assuming $\|\mathbf{D}_l\|_2^2 \approx L$, where L is the number of samples in one aperture time [8]. Formula (30) is simplified as

$$\|\hat{\sigma}_{\Lambda} - \sigma_{\Lambda}\|_2^2 \geq \sigma_n^2 \left[2 \text{trace}(\mathbf{B}^{-1}) - \text{trace}(\mathbf{A}^{-1}) \right] \quad (31)$$

Gail's circle theorem [36] indicates that

$$\begin{aligned} |\lambda_l - L| &\leq (S-1) \cdot u \cdot L \\ (1 - (S-1) \cdot u) \cdot L &\leq \lambda_l \leq (1 + (S-1) \cdot u) \cdot L \\ \frac{1}{(1+(S-1) \cdot u) \cdot L} &\leq \lambda_l^{-1} \leq \frac{1}{(1-(S-1) \cdot u) \cdot L} \\ \frac{S}{(1+(S-1) \cdot u) \cdot L} &\leq \sum_{l=1}^S \lambda_l^{-1} \leq \frac{S}{(1-(S-1) \cdot u) \cdot L} \\ \frac{S}{(1+(S-1) \cdot u) \cdot L} &\leq \text{trace}(\mathbf{A}^{-1}) \leq \frac{S}{(1-(S-1) \cdot u) \cdot L} \end{aligned} \quad (32)$$

$$\begin{aligned}
& \left| \alpha_l - \left(L + \frac{\gamma_2 \cdot \xi_{\Lambda_l}}{\sigma_{\Lambda_l}} \right) \right| \leq (S-1) \cdot u \cdot L \\
& \frac{\sigma_{\Lambda_l} \cdot [1 - (S-1) \cdot u] \cdot L + \gamma_2 \cdot \xi_{\Lambda_l}}{\sigma_{\Lambda_l}} \leq \alpha_l \leq \frac{\sigma_{\Lambda_l} \cdot [1 + (S-1) \cdot u] \cdot L + \gamma_2 \cdot \xi_{\Lambda_l}}{\sigma_{\Lambda_l}} \\
& \frac{\sigma_{\Lambda_l}}{\sigma_{\Lambda_l} \cdot [1 + (S-1) \cdot u] \cdot L + \gamma_2 \cdot \xi_{\Lambda_l}} \leq \alpha_l^{-1} \leq \frac{\sigma_{\Lambda_l}}{\sigma_{\Lambda_l} \cdot [1 - (S-1) \cdot u] \cdot L + \gamma_2 \cdot \xi_{\Lambda_l}} \\
& \frac{S}{[1 + (S-1) \cdot u] \cdot L + \gamma_2 \cdot \max\left(\frac{\xi_{\Lambda_l}}{\sigma_{\Lambda_l}}\right)} \leq \sum_{l=1}^S \alpha_l^{-1} \leq \frac{S}{[1 - (S-1) \cdot u] \cdot L + \gamma_2 \cdot \min\left(\frac{\xi_{\Lambda_l}}{\sigma_{\Lambda_l}}\right)} \\
& \frac{S}{[1 + (S-1) \cdot u] \cdot L + \gamma_2 \cdot \max\left(\frac{\xi_{\Lambda_l}}{\sigma_{\Lambda_l}}\right)} \leq \text{trace}(\mathbf{B}^{-1}) \leq \frac{S}{[1 - (S-1) \cdot u] \cdot L + \gamma_2 \cdot \min\left(\frac{\xi_{\Lambda_l}}{\sigma_{\Lambda_l}}\right)}
\end{aligned} \tag{33}$$

where $u = \max_{1 \leq m_1 \neq m_2 \leq M} |\langle \mathbf{D}_{m_1}, \mathbf{D}_{m_2} \rangle| / (\|\mathbf{D}_{m_1}\|_2 \cdot \|\mathbf{D}_{m_2}\|_2)$ is the mutual coherence coefficient, which reflects the maximum similarity between any two different columns m_1, m_2 in the recovered matrix

$\mathbf{D}_{N \times M}$. Combine Formulas (32) and (33), as follows:

$$\begin{aligned}
& \frac{2S}{[1 + (S-1) \cdot u] \cdot L + \gamma_2 \cdot \max\left(\frac{\xi_{\Lambda_l}}{\sigma_{\Lambda_l}}\right)} - \frac{S}{(1 - (S-1) \cdot u) \cdot L} \leq 2\text{trace}(\mathbf{B}^{-1}) - \text{trace}(\mathbf{A}^{-1}) \\
& \leq \frac{2S}{[1 - (S-1) \cdot u] \cdot L + \gamma_2 \cdot \min\left(\frac{\xi_{\Lambda_l}}{\sigma_{\Lambda_l}}\right)} - \frac{S}{(1 + (S-1) \cdot u) \cdot L}
\end{aligned} \tag{34}$$

So, we have the MSE of the pseudo- $\mathcal{L}_0$ -norm regularization model

$$\|\hat{\boldsymbol{\sigma}}_{\Lambda} - \boldsymbol{\sigma}_{\Lambda}\|_2^2 \geq \sigma_n^2 \left[2\text{trace}(\mathbf{B}^{-1}) - \text{trace}(\mathbf{A}^{-1}) \right] \geq \sigma_n^2 \left[\frac{2S}{[1 + (S-1) \cdot u] \cdot L + \gamma_2 \cdot \max\left(\frac{\xi_{\Lambda_l}}{\sigma_{\Lambda_l}}\right)} - \frac{S}{(1 - (S-1) \cdot u) \cdot L} \right] \tag{35}$$

Similarly, the MSE of the $\mathcal{L}_1$ -norm regularization model

$$\|\hat{\boldsymbol{\sigma}}_{l_1} - \boldsymbol{\sigma}_{\Lambda}\|_2^2 \geq \sigma_n^2 \left[2\text{trace}(\mathbf{B}^{-1}) - \text{trace}(\mathbf{A}^{-1}) \right] \geq \sigma_n^2 \left[\frac{2S}{[1 + (S-1) \cdot u] \cdot L + \gamma_1 \cdot \max\left(\frac{1}{\sigma_{\Lambda_l}}\right)} - \frac{S}{[1 - (S-1) \cdot u] \cdot L} \right] \tag{36}$$

From the above deduction, we can know the optimal performance achieved by the two regularization models above. By comparing (35) and (36), it can be observed that the difference is in the dashed block. The pseudo- $\mathcal{L}_0$ -norm optimization algorithm is more sensitive to small coefficients. For low SNR targets, i.e., $\xi_{\Lambda_l} = 1/(\Gamma(\sigma_{\Lambda_l}) + \iota) > 1$, our proposed algorithm performs better and can recover low SNR targets more accurately.

AgileSAR imaging includes three steps: range compression, range cell migration correction (RCMC), and azimuth compression. After range compression and RCMC, a pseudo- $\mathcal{L}_0$ -norm FISTA-net is employed to achieve azimuth compression. The simulation experiments and data experiments in Figures 3–5 demonstrate that the algorithm can achieve smaller NMSE, higher SSIM, and PSNR compared to the optimization-based algorithm. Additionally, the proposed method significantly reduces the imaging time after training the model, which improves the efficiency of AgileSAR imaging.

The potential applications of the pseudo- $\mathcal{L}_0$ -norm FISTA-net can recover more target details and reduce computation complexity, so that the efficiency and accuracy of AgileSAR imaging are further improved. This algorithm can also be applied to the traditional HRWS system, e.g., the azimuthal multi-channel SAR and MIMO SAR, to lower the amount of data to relieve the pressure on data storage.

5. Conclusions

In this paper, we propose a signal processing algorithm for an innovative single-channel HRWS system called AgileSAR for sparse scenes. To further improve AgileSAR imaging performance and efficiency, a pseudo- $\mathcal{L}_0$ -norm FISTA-net is proposed. Firstly,

the pseudo- $\mathcal{L}_0$ -norm regularization model takes an approximately fair penalization rule with prior scene information to nearly acquire the number of nonzero variables based on Bayesian estimation, so that it can more accurately recover more target details compared with the $\mathcal{L}_1$ -norm regularization model. Then, the proposed algorithm unfolds the iterative process into the deep network to solve the pseudo- $\mathcal{L}_0$ -norm regularization model to improve the accuracy and efficiency of AgileSAR imaging. In addition, the learnable sparsifying transform in this network can recover more details than the traditional sparsifying transform. Finally, the quantitative analysis and simulated experiments demonstrate the effectiveness and efficiency of the proposed algorithm.

Author Contributions: Methodology, W.C., J.G., F.M. and L.Z.; Validation, W.C., J.G. and F.M.; Writing—original draft, W.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Restrictions apply to the availability of these data. Data were bought from Deutsches Zentrum für Luft- und Raumfahrt (DLR) and are available from the author Wenjiao Chen with the permission of DLR.

Acknowledgments: The authors would like to thank the editors and the reviewers for their help and suggestions.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Krieger, G.; Moreira, A.; Fiedler, H.; Hajnsek, I.; Werner, M.; Younis, M.; Zink, M. TanDEM-X: A satellite formation for high-resolution SAR interferometry. *IEEE Trans. Geosci. Remote Sens.* **2007**, *45*, 3317–3341. [\[CrossRef\]](#)
2. Krieger, G.; Younis, M.; Gebert, N.; Huber, S.; Moreira, A. Advanced concepts for high-resolution wide-swath SAR imaging. In Proceedings of the 8th European Conference on Synthetic Aperture Radar, Aachen, Germany, 7–10 June 2010; pp. 1–4.
3. Sikaneta, I.; Gierull, C.H.; Cerutti-Maori, D. Optimum signal processing for multichannel SAR: With application to high-resolution wideswath imaging. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 6095–6109. [\[CrossRef\]](#)
4. Krieger, G. MIMO-SAR: Opportunities and pitfalls. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 2628–2645. [\[CrossRef\]](#)
5. Candes, E.J.; Tao, T. Decoding by linear programming. *IEEE Trans. Inf. Theory* **2005**, *51*, 4203–4215. [\[CrossRef\]](#)
6. Donoho, D.L.; Elad, M.; Temlyakov, V.N. Stable recovery of sparse overcomplete representations in the presence of noise. *IEEE Trans. Inf. Theory* **2006**, *52*, 6–18. [\[CrossRef\]](#)
7. Baraniuk, R. Compressive sensing. *IEEE Signal Process Mag.* **2007**, *24*, 118–121. [\[CrossRef\]](#)
8. Yu, Z.; Chen, W.; Xiao, P.; Li, C. AgileSAR: Achieving Wide-Swath Spaceborne SAR Based on Time-Space Sampling. *IEEE Access* **2019**, *7*, 674–686. [\[CrossRef\]](#)
9. Tillmann, A.M.; Pfetsch, M.E. The computational complexity of the restricted isometry property, the nullspace property, and related concepts in compressed sensing. *IEEE Trans. Inf. Theory* **2014**, *60*, 1248–1259. [\[CrossRef\]](#)
10. Tropp, J.A.; Gilbert, A.C. Signal recovery from partial information via orthogonal matching pursuit. *IEEE Trans. Inf. Theory* **2007**, *53*, 4655–4666. [\[CrossRef\]](#)
11. Cerrone, C.; Cerull, R.; Golden, B. Carousel greedy: A generalized greedy algorithm with applications in optimization. *Comput. Oper. Res.* **2017**, *85*, 97–112. [\[CrossRef\]](#)
12. Candes, E.; Tao, T. The Dantzig selector: Statistical estimation when p is much larger than n . *Ann. Stat.* **2007**, *35*, 2313–2351.
13. Tibshirani, R. Regression shrinkage and selection via the Lasso. *J. R. Stat. Soc. B* **1996**, *58*, 267–288. [\[CrossRef\]](#)
14. Ji, S.; Xue, Y.; Carin, L. Bayesian Compressive Sensing. *IEEE Trans. Signal Process* **2008**, *56*, 2346–2356. [\[CrossRef\]](#)
15. Tipping, M.E.; Faul, A.C. Fast marginal likelihood maximization for sparse Bayesian models. In Proceedings of the Ninth International Workshop on Artificial Intelligence and Statistics, Key West, FL, USA, 3–6 January 2003; pp. 3–6.
16. Babacan, S.D.; Molina, R.; Katsaggelos, A.K. Bayesian compressive sensing using laplace priors. *IEEE Trans. Image Process.* **2010**, *19*, 53–63. [\[CrossRef\]](#) [\[PubMed\]](#)
17. Arjouni, Y.; Kaabouch, N.; Ghazi, H.E.; Tamtaoui, A. Compressive sensing: Performance comparison of sparse recovery algorithms. In Proceedings of the Annual Computing and Communication Workshop and Conference (CCWC) 2017, Las Vegas, NV, USA, 9–11 January 2017; pp. 1–7.

18. Joshi, S.; Siddamal, K.V.; Saroja, V.S. Performance analysis of compressive sensing reconstruction. In Proceedings of the 2015 2nd International Conference on Electronics and Communication Systems (ICECS), Coimbatore, India, 26–27 February 2015; pp. 724–729.
19. Celik, S.; Basaran, M.; Erkucuk, S.; Cirpan, H. Comparison of compressed sensing based algorithms for sparse signal reconstruction. In Proceedings of the 2016 24th Signal Processing and Communication Application Conference (SIU), Zonguldak, Turkey, 16–19 May 2016.
20. Xiong, K.; Zhao, G.; Wang, Y.; Shi, G. SPB-Net: A Deep Network for SAR Imaging and Despeckling with DownSampled Data. *IEEE Trans. Geosci. Remote Sens.* **2021**, *59*, 9238–9256. [\[CrossRef\]](#)
21. Xiong, K.; Zhao, G.; Wang, Y.; Shi, G.; Chen, S. Lq-SPB-Net: A Real-Time Deep Network for SAR Imaging and Despeckling. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 5209721. [\[CrossRef\]](#)
22. Xiong, K.; Zhao, G.; Wang, Y.; Shi, G. SAR Imaging and Despeckling Based on Sparse, Low-Rank, and Deep CNN Priors. *IEEE Geosci. Remote Sens. Lett.* **2021**, *19*, 4501205. [\[CrossRef\]](#)
23. Hershey, J.R.; Roux, J.L.; Weninger, F. Deep Unfolding: Model-Based Inspiration of Novel Deep Architectures. *arXiv* **2014**. [\[CrossRef\]](#)
24. Zhang, J.; Ghanem, B. ISTA-Net: Interpretable Optimization-Inspired Deep Network for Image Compressive Sensing. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–23 June 2018; pp. 1828–1837.
25. You, D.; Xie, J.; Zhang, J. ISTA-Net++: Flexible Deep Unfolding Network for Compressive Sensing. In Proceedings of the 2021 IEEE International Conference on Multimedia and Expo (ICME), Shenzhen, China, 5–9 July 2021.
26. Zhang, K.; Gool, L.V.; Timofte, R. Deep Unfolding Network for Image Super-Resolution. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020; pp. 3214–3223.
27. Xiang, J.; Dong, Y.; Yang, Y. FISTA-Net: Learning A Fast Iterative Shrinkage Thresholding Network for Inverse Problems in Imaging. *IEEE Trans. Med. Imaging* **2021**, *40*, 1329–1339. [\[CrossRef\]](#)
28. Zhou, G.; Xu, Z.; Fan, Y.; Zhang, Z.; Qiu, X.; Zhang, B.; Fu, K.; Wu, Y. HPHR-SAR-Net: Hyperpixel High-Resolution SAR Imaging Network Based on Nonlocal Total Variation. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2023**, *16*, 8595–8608. [\[CrossRef\]](#)
29. Wang, M.; Zhang, Z.; Qiu, X.; Gao, S.; Wang, Y. ATASI-Net: An Efficient Sparse Reconstruction Network for Tomographic SAR Imaging with Adaptive Threshold. *IEEE Trans. Geosci. Remote Sensing* **2023**, *61*, 4701918. [\[CrossRef\]](#)
30. Zou, H. The adaptive lasso and its oracle properties. *J. Am. Stat. Assoc.* **2006**, *101*, 1418–1429. [\[CrossRef\]](#)
31. Candès, E.J.; Wakin, M.B.; Boyd, S.P. Enhancing Sparsity by Reweighted L1 Minimization. *J. Fourier Anal. Appl.* **2007**, *14*, 877–905. [\[CrossRef\]](#)
32. Seeger, M.W.; Nickisch, H. Compressed sensing and Bayesian experimental design. In Proceedings of the 25th international conference on Machine Learning (ICML), 5–9 July 2008; pp. 912–919.
33. Rousset, F.; Ducros, N.; Farina, A.; Valentini, G.; Andrea, C.; Peyrin, F. Adaptive basis scan by wavelet prediction for single-pixel imaging. *IEEE Trans. Comput. Imaging* **2017**, *3*, 36–46. [\[CrossRef\]](#)
34. Beck, A.; Teboulle, M. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM J. Imag. Sci.* **2009**, *2*, 183–202. [\[CrossRef\]](#)
35. Wang, Z.; Bovik, A.C.; Sheikh, H.R.; Simoncelli, E.P. Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Process.* **2004**, *13*, 600–612. [\[CrossRef\]](#)
36. Golub, G.; Loan, C.F.V. *Matrix Computations*, 3rd ed.; Johns Hopkins University Press: Baltimore, MD, USA, 1996.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.