A Novel Inverse Synthetic Aperture Radar Imaging Method Using Convolutional Neural Networks

Changyu Hu[1], Ling Wang[1], *Member, IEEE*, Ze Li[1], Otmar Loffeld[2], *Senior Member, IEEE*
1. Key Laboratory of Radar Imaging and Microwave Photonics of the Ministry of Education
Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China
2. Center for Sensor Systems, University of Siegen, Siegen 57076, Germany

***Abstract*—The compressive sensing theory provides a novel type of image reconstruction method for radar imaging. A good image can be obtained using much less data as compared to the conventional imaging methods. However, there are two issues of CS based ISAR imaging methods. One is the time-consuming iterative optimization based reconstruction, which always leads to slow computational speed as compared to classical ISAR imaging methods. The other is the sparse representation of the target scene. The assumption of spatial sparse or transforming the target scene into assumed sparse domains may not well solve the problem of sparse representation. This two issues limit the application and reconstruction performance of CS based imaging methods. We propose a new ISAR imaging method that uses the convolutional neural network (CNN) to perform the image formation using under-sampled ISAR data. An effective CNN architecture for ISAR imaging is presented, which can exploit the sparse feature of the target image extremely well by multi-layer nonlinear processing. The imaging results of real ISAR data show that the proposed CNN based ISAR imaging method can obtain higher quality images with less data than the state-of-the-art CS reconstruction algorithms and improve the imaging efficiency obviously.**

***Index Terms*—Radar; imaging; Inverse synthetic aperture radar (ISAR); convolutional neural network (CNN); deep learning**

I. Introduction

The conventional ISAR imaging uses the range-Doppler (RD) type of methods. In the past decade, compressive sensing (CS) based imaging has gained much interest. A variety of image reconstruction algorithms developed using CS theory have been published [1]–[7]. The CS based imaging methods can obtain good imaging results with under-sampled data or non-complete data on the condition that the scene is approximately sparse or sparse in a transformation domain [8].

In CS based ISAR imaging, the data after motion compensation can be modeled as follows [6]:

$$\vec{\mathbf{G}}_s = \mathbf{\Psi}\vec{\boldsymbol{\sigma}} + \vec{\mathbf{n}}_s \tag{1}$$

where $\vec{\mathbf{G}}_s \in \mathbb{C}^m$ is a randomly under-sampled measurement vector, $\mathbf{\Psi} \in \mathbb{C}^{m\times n}$ with $m<n$, $n = N_r N_a$ is the measurement matrix, which is a partial Fourier matrix. $\vec{\boldsymbol{\sigma}} \in \mathbb{C}^n$ denotes the ISAR target scene to be imaged and $\vec{\mathbf{n}}_s \in \mathbb{C}^m$ is the noise in the measurements.

Corresponding author : Prof. Ling Wang/ Email: wanglrpizess@163.com

Under the assumption that the target scene $\vec{\boldsymbol{\sigma}}$ is sparse, the imaging problem is formulated as an constrained l_1 minimization problem [5] as follows:

$$\hat{\vec{\boldsymbol{\sigma}}} = \arg\min_{\vec{\boldsymbol{\sigma}}} \|\vec{\boldsymbol{\sigma}}\|_1 \quad \text{s.t.} \quad \|\vec{\mathbf{G}}_s - \mathbf{\Psi}\vec{\boldsymbol{\sigma}}\|_2^2 < \epsilon \tag{2}$$

The sparse representation of $\vec{\boldsymbol{\sigma}}$ can be used to enhance the reconstruction of certain features [9]. Let $\mathbf{D} \in \mathbb{C}^{n\times n}$ be a sparse transform or dictionary. The sparse representation of the target scene is given by $\vec{\mathbf{w}} = \mathbf{D}\vec{\boldsymbol{\sigma}}$ where the vector $\vec{\mathbf{w}} \in \mathbb{C}^n$ denotes the sparse coefficients in the domain expanded by $\mathbf{D}$. Thus, the image reconstruction is re-formulated as follows:

$$\hat{\vec{\mathbf{w}}} = \arg\min_{\vec{\mathbf{w}}} \|\vec{\mathbf{w}}\|_1 \quad \text{s.t.} \quad \|\vec{\mathbf{G}}_s - \mathbf{\Psi}\mathbf{D}\vec{\mathbf{w}}\|_2^2 < \epsilon \tag{3}$$

Once $\hat{\vec{\mathbf{w}}}$ is obtained, the target image is formed by $\hat{\vec{\boldsymbol{\sigma}}} = \mathbf{D}\hat{\vec{\mathbf{w}}}$.

There are a variety of reconstruction algorithms [10] for solving (2) or (3). However, the algorithms are always of an iterative structure. A large amount of iterations are usually needed for the algorithms to converge to a reasonable solution, which makes them time-consuming and are not comparable to the classic range-Doppler based radar imaging methods. In addition, the unknowns $\hat{\vec{\boldsymbol{\sigma}}}$ to be reconstructed are not exactly sparse in the domain expressed by a pre-determined or a fixed dictionary. The resultant sparse representation of the target image is not good enough and not adaptive to the target to be imaged. Some researchers have attempted to obtain better imaging results by applying the dictionary learning to better sparsely represent the target image [11], [12].

At present, a new type of signal recovery approach for under-sampled data has appeared and gained much interest, which employs the deep learning technology [13]–[15]. In view of this approach, we develop a novel ISAR imaging method for randomly under-sampled data using the convolutional neural network (CNN).

In the CNN framework, the imaging problem can be regarded as finding a best representation of the input data in an "image space". In our work, we first generate an initial image by processing the randomly under-sampled data simply using the traditional range-Doppler method. The quality of the initial image is very poor as expected since the pulses cannot be coherently integrated. We use this poor initial image as the input to the CNN imaging network. Then, we train the network using training data sets that we specially construct using existing ISAR real data. Note that we use data augmentation to generate sufficient training data sets for learning. We compare

the CNN based imaging method with several CS imaging methods. The imaging results of real ISAR data show that the performance of our CNN based imaging method is superior to the CS imaging methods.

The rest of our paper is organized as follow: Section II presents the CNN based ISAR imaging method including the CNN architecture, training strategy and generation of the input image to the imaging network. Section III presents processing results of real ISAR data and the analysis of the results. Section V draws the conclusion.

II. CNN-Based Imaging method

A. An effective CNN architecture for ISAR imaging

CNN is a typical architectures in deep learning (DL) technology [16], which is comprised of multiple convolutional layers, activation function layers and sometimes several full connection layers. The unique characteristics of CNN are sparse connections and weights sharing mechanism, which can reduce the number of parameters in CNN and speed up the network learning. The main advantage of CNN is its excellent capability of feature extraction, especially deep-layer abstract features. A number of CNN architecture based DL methods have been applied to image classification [17], object detection [18] and image reconstruction [19], etc.

We develop a CNN architecture for ISAR imaging, as shown in Fig. 1. The proposed imaging network consists of five convolutional layers. Each layer generates a new representation of the output of the last layer. In the first three layers, there is an activation function layer after each layer where the rectified linear unit (ReLU) is used as the non-linear activation function. As shown in Fig. 1, the light blue planes denotes the feature maps and the orange planes denotes the nonlinear activation processing results of the feature maps.

The convolutional kernel size used by each convolutional layer is 11×11, 3×3, 7×7, 11×11 and 1×1, respectively. The number of the feature maps generated by each convolutional layer is 1, 64, 32, 1 and 1, respectively. The adoption of multi-channel feature extraction in middle layers is to ensure effective and complete feature capture leading to a good mapping between the input data and output image.

The size of the feature maps generated by the each convolutional layers is 100×100, which keeps the same with size of the input data, i.e., the low-quality initial image of the target. The output of the imaging network is the well focused target image that we expect.

B. Training of the imaging network

The training data is composed of the training sets and validation sets. The training sets and validation sets are composed of 500 training examples and 100 validation examples, respectively. Each training example and validation example contain an initial input image and a label image.

The label images and initial images (input images) are generated using available ISAR real data. The motion compensated ISAR data is well processed using RD algorithm and the resulting good images are defined as the label images. The corresponding randomly under-sampled Fourier-domain ISAR data is performed simply using two-dimensional inverse Fourier transform to form the initial image used as the input image.

TABLE I: Setting of hyperparameters for network training

Parameters	**base_lr**	**l_strategy**	**v_interval**	**v_iteration**
Values	10e−5	fixed	250	50

In order to make the network parameters converge quickly and accurately, we set the hyperparameters as listed in Table I. The base_lr denotes the initial learning rate. The l_strategy denotes the changing strategy of learning rate during the network training. The v_interval denotes the training times between the current validation and the previous validation. The v_iteration denotes the iteration times in the validation stage. The optimal algorithm used for CNN training is stochastic gradient descent (SGD) in our work

The loss function used in the training stage is defined as the averaged reconstruction errors over all training samples. Let $L(\mathbf{W})$ denote the loss function. We have

$$L(\mathbf{W}) = \frac{1}{K}\sum_{i=1}^{K} \|f(\tilde{\boldsymbol{\sigma}}_i, \mathbf{W}) - \hat{\boldsymbol{\sigma}}_i\|^2 \tag{4}$$

where $\mathbf{W}$ denotes the weights and biases of the imaging network, $f(\tilde{\boldsymbol{\sigma}}_i, \mathbf{W})$ denotes the network output for the i^{th} training example with initial image $\tilde{\boldsymbol{\sigma}}_i$ and label image $\hat{\boldsymbol{\sigma}}_i$ and K denotes the size of the training set, i.e., the number of training samples. The loss function is minimized by adjusting the weights and biases in each layers of the imaging network using backpropagation and SGD algorithms.

C. Generation of the input image to the CNN

We assume there are M_r and M_a samples in the range and cross-range dimensions after the random down-sampling, respectively. Note that the data is usually in the range-frequency and cross-range-time domain. For the radar data using dechirp processing in the range, the domain in which the under-sampling is performed is the range-time and cross-range-time domain.

There are two methods to obtain the input image. One is to use the weighted least square (WLS) method to solve a underdetermined system of linear equations to get an initial image with size of $N_r \times N_a$. Note that the processing is performed by arranging the two-dimensional data into a one-dimensional vector and rearranging the solution back into a two-dimensional matrix.

The other is to expand the $M_r \times M_a$ matrix to a $N_r \times N_a$ one by inserting zero at the positions that are not sampled or just randomly setting zero if the range and slow-time indices of the samples are not known. The resulting zero padded matrix is much like the non-complete data with certain samples being missing. Then, we perform the two-dimensional Fourier transform to obtain the input image.

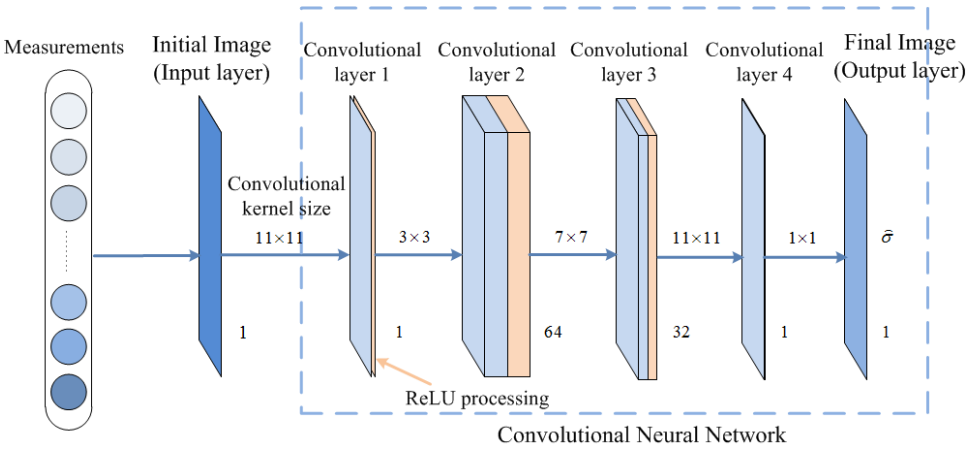


Fig. 1: CNN architecture for ISAR imaging.

III. Demonstrations

A. Data description

In this section, the performance of the proposed CNN based imaging method is demonstrated using ISAR real data of a plane. Note that the ISAR real data considered here is the same type of data as the training data.

The ISAR data were collected by a ground-based ISAR operating at C band. The bandwidth of the transmitted waveform is 400 MHz. A de-chirp processing was used in the received signal processing. The motion compensation has been performed using the minimum entropy based global range alignment algorithm [20] and the improved phase gradient algorithm (PGA) [21].

The size of the data is 100×100, i.e., 100 range cells and 100 slow time samples, corresponding to 10^4 measurements. After the motion compensation, the data is randomly under-sampled in both range and cross-range dimension with under-sampled ratio of 25% leading to 2500 measurements.

B. Metrics for image quality evaluation

To provide a quantitative evaluation of the imaging performance of the proposed imaging method, we use two types of metrics. One type of metrics is the "true-value" based metrics and the other is the conventional metrics for evaluating the image quality.

The "true-value" based evaluation is based on the comparison of the reconstructed image with a reference image, which represents the "true-value" image. Since we do not have ground-truth images of the noncooperative plane target, we use a high-quality image reconstructed by conventional RD method using full data as the reference image. Thus, the metrics evaluate the performance of the proposed imaging method as compared to the RD method.

The metrics used in "true-value" based evaluation are as follows [22]: false alarm (FA), missed detection (MD) and relative root mean square error (RRMSE). FA denotes the number of scatterers that are reconstructed in the image but are not present in the reference image. MD denotes the number of scatterers that are not reconstructed in the newly generated image but are reconstructed in the reference image. RRMSE measures the reconstruction error of the amplitude of the scatterers.

The conventional metrics for evaluating the image quality are target-to-clutter Ratio (TCR), entropy of the image (ENT) and image contrast (IC). The TCR that we use in our work is defined as follows:

$$\mathrm{TCR} = 10\log_{10}\left(\frac{\sum_{(i,j)\in\Omega_\tau}|\sigma(i,j)|^2}{\sum_{(i,j)\in\Omega_C}|\sigma(i,j)|^2}\right) \tag{5}$$

where (i,j) denotes the pixel index, $\sigma(i,j)$ denotes the reconstructed value at (i,j) of the normalized image $\hat{\sigma}$ and Ω_τ, Ω_C denote the target region and clutter region in σ. We determine Ω_τ and Ω_C by performing a binarization processing on the image reconstructed by the conventional RD method with full data. The pixels whose values are greater than a specified threshold are classified in to Ω_τ, otherwise into Ω_C.

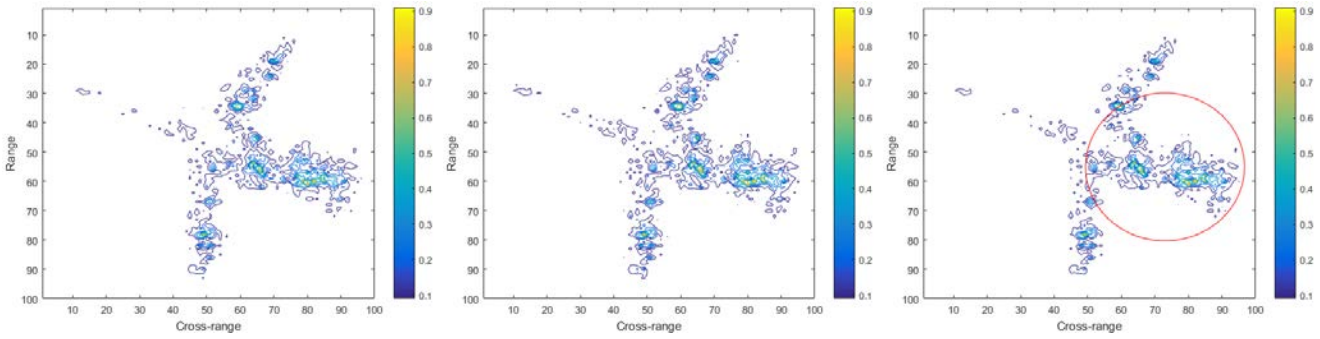
C. Imaging results and analysis

We performed the image formation using the proposed CNN based imaging method. We trained the imaging network with three different times. Fig. 2 presents the reconstructed images.

For performance comparison, we present the image obtained by performing the conventional RD method on full data, as shown in Fig. 3(a) as well as the images reconstructed by orthogonal matching pursuit (OMP) [23] method, null-space l_1 norm minimization method [5], greedy kalman filtering(GKF) method [6] and dictionary learning based methods [11] using 25% measurements, as shown in Fig. 3(a) - Fig. 3(f). Note that all images are displayed with the same contour level.

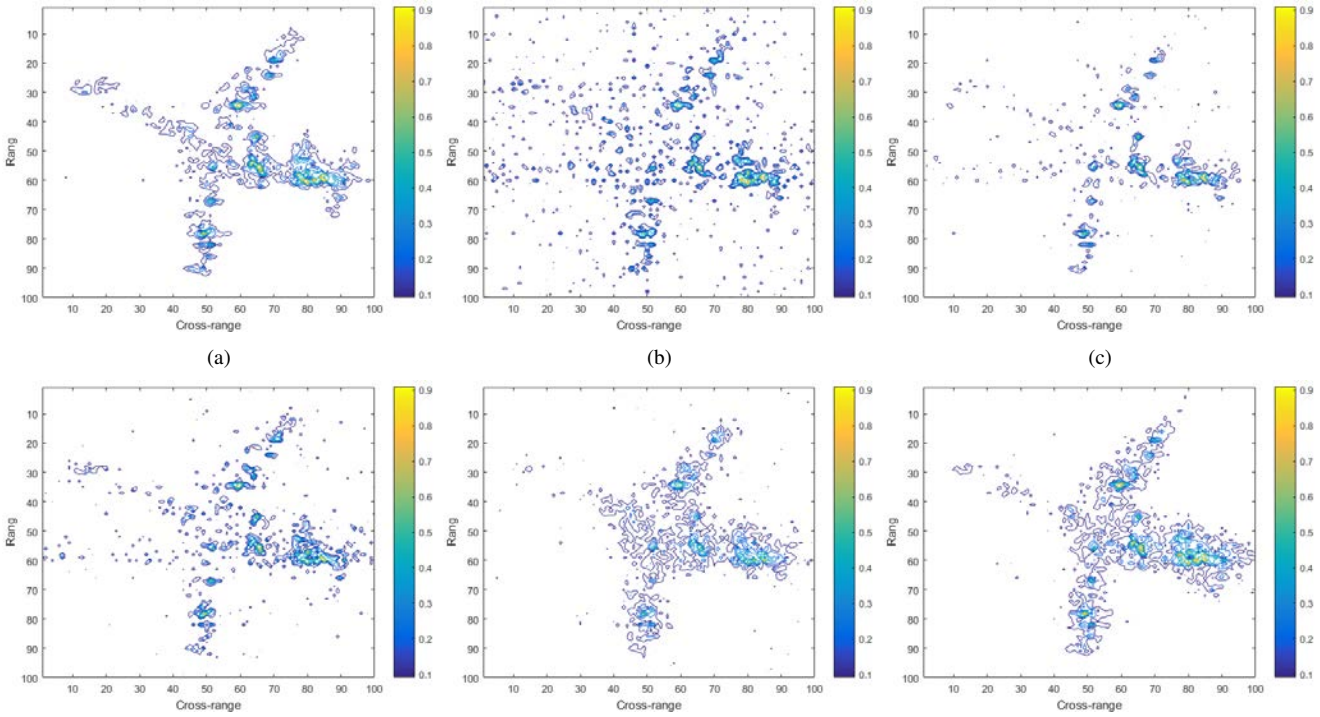
From Fig. 2, we see that the plane is well reconstructed by the proposed method. There is little visual difference in the results obtained with different training times. However, the image reconstructed with the network trained by more times presents the shape of the target clearer than those obtained with less trained network, as indicated by the circled region.

Comparing Fig. 2 with Fig. 3(a)-Fig. 3(f), we see that with same number of under-sampled data, the CNN based imaging method can provide much better reconstructions than the existing CS imaging methods considered here. There are



(a) (b) (c)

Fig. 2: The images reconstructed by the proposed CNN based imaging method using 25% measurement where the imaging networks are trained by (a) 5000 times (CNN1), (b) 15000 times (CNN2) and (c) 25000 times (CNN3).



(a) (b) (c)

(d) (e) (f)

Fig. 3: The images reconstructed by (a) RD method using full data and by the (a) OMP method, (b) null-space l_1 minimization based method, (c) GKF method, (d) on-line dictionary learning based method and (e) off-line dictionary learning based method using 25% measurements.

smaller artifacts in the images obtained by the proposed method. The shape of the target is much well reconstructed, as shown in Fig. 2. With one fourth of the raw data, we can obtain the nearly same image as the RD image.

We evaluate the image quality using the metrics mentioned above. The results are listed in Table II. We refer to the CNN trained with three different times as CNN1, CNN2 and CNN3, respectively.

The FA list in the first column in Table II indicates that the CNN(CNN1-CNN3) based imaging methods has small artifacts.

The values of the MD also indicate the superiority of our method in reconstructing the scatterers. The values of MD of dictionary learning based methods are close to those of our method, but the target shape is not reconstructed as well as our method, as shown in the images Fig. 3(e) and Fig. 3(f).

The null-space l_1 norm minimization method has the smallest FA, but it has the largest MD, which may be due to its nature to find the sparsest reconstruction. In this regard, we see the CNN based imaging method has small FA and MD simultaneously. This is an attractive characteristic.

From the RRMSE listed in the fourth column of Table II, we see that our method has the smallest RRMSE, which indicates that the amplitude reconstruction error of our method relative to the RD imaging is smallest as compared to other imaging methods considered.

TABLE II: Evaluation of image quality for different imaging methods

Methods	FA	MD	RRMSE	TCR(dB)	ENT	IC	Time(s)
CNN1	81	186	0.2908	24.0248	5.4757	8.6113	0.1011
CNN2	89	170	0.2398	24.7090	5.4734	8.5561	0.1205
CNN3	72	184	0.2681	24.9916	5.4339	8.7784	0.1025
OMP	449	266	0.4492	16.0869	6.0284	6.3179	60.2342
GKF	222	233	0.3315	19.7872	5.6509	7.7449	1.3408e3
On-line dictionary learning	244	160	0.3251	20.3453	5.7985	7.2270	5.9377
Off-line dictionary learning	192	189	0.3203	20.5431	5.7080	7.4934	2.3620
Null-space l_1 norm minimization	50	304	0.2902	25.8064	4.9401	11.0844	321.1142

As the training times of CNN increase, the TCR and IC of the reconstructed image become larger and the ENT of the image becomes smaller, as shown in the fifth column of Table II. The proposed CNN based imaging method leads to higher TCR and IC and smaller ENT than other imaging methods.

The last column of II shows the computational time of each imaging method. Once the imaging network is trained well, the imaging time of the CNN reaches the order of magnitude $0.1\mathrm{s}$. The imaging efficiency is much higher than other iterative optimization based imaging methods.

IV. Conclusion

We proposed a novel CNN based ISAR imaging method. An explicit CNN architecture for performing ISAR imaging is presented. The experimental results show that the new imaging method can obtain much better results than the state-of-the-art CS reconstruction algorithms. Furthermore, the CNN based imaging method has very high computational efficiency, which meets the requirements of real-time processing well.

References

[1] D. L. Donoho. Compressed sensing. *IEEE Transactions on Information Theory*, 52(4):1289–1306, April 2006.

[2] R. Baraniuk and P. Steeghs. Compressive radar imaging. In *2007 IEEE Radar Conference*, pages 128–133, April 2007.

[3] Joachim H. G. Ender. On compressive sensing applied to radar. *Signal Processing*, 90(5):1402–1414, 2010.

[4] J. H. G. Ender. Autofocusing isar images via sparse representation. In *European Conference on Synthetic Aperture Radar*, pages 203–206, 2012.

[5] L. Wang and O. Loffeld. ISAR imaging using a null space l_1 minimizing kalman filter approach. In *2016 4th International Workshop on Compressed Sensing Theory and its Applications to Radar, Sonar and Remote Sensing (CoSeRa)*, pages 232–236, Sept 2016.

[6] Ling Wang, Otmar Loffeld, Kaili Ma, and Yulei Qian. Sparse ISAR imaging using a greedy kalman filtering. *Signal Processing*, 138:1(10), 2017-09-01.

[7] S. Tomei, A. Bacci, E. Giusti, M. Martorella, and F. Berizzi. Compressive sensing-based inverse synthetic radar imaging imaging from incomplete data. *IET Radar, Sonar Navigation*, 10(2):386–397, 2016.

[8] M. Cetin and W. C. Karl. Feature-enhanced synthetic aperture radar image formation based on nonquadratic regularization. *IEEE Transactions on Image Processing*, 10(4):623–631, Apr 2001.

[9] S. Samadi, M. Cetin, and M. A. Masnadi-Shirazi. Sparse representation-based synthetic aperture radar imaging. *IET Radar, Sonar Navigation*, 5(2):182–193, Feb 2011.

[10] Simon Foucart and Holger Rauhut. *A Mathematical Introduction to Compressive Sensing*. Birkhuser Basel, 2013.

[11] Ling Wang Changyu Hu and Otmar Loffeld. *Inverse Synthetic Aperture Radar Imaging Exploiting Dictionary Learning*. IEEE Radar Conference, 2018.

[12] A. Soanlui and M. Cetin. Dictionary learning for sparsity-driven sar image reconstruction. In *2014 IEEE International Conference on Image Processing (ICIP)*, pages 1693–1697, Oct 2014.

[13] Ali Mousavi, Ankit B. Patel, and Richard G. Baraniuk. A deep learning approach to structured signal recovery. *CoRR*, abs/1508.04065, 2015.

[14] K. Kulkarni, S. Lohit, P. Turaga, R. Kerviche, and A. Ashok. Reconnet: Non-iterative reconstruction of images from compressively sensed measurements. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 00, pages 449–458, June 2016.

[15] Ali Mousavi and Richard G. Baraniuk. Learning to invert: Signal recovery via deep convolutional networks. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 2272–2276, 2017.

[16] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016.

[17] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 25*, pages 1097–1105. Curran Associates, Inc., 2012.

[18] R. Girshick, J. Donahue, T. Darrell, and J. Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. In *2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 00, pages 580–587, June 2014.

[19] K. H. Jin, M. T. Mccann, E Froustey, and M Unser. Deep convolutional neural network for inverse problems in imaging. *IEEE Transactions on Image Processing A Publication of the IEEE Signal Processing Society*, 26(9):4509–4522, 2017.

[20] D. Zhu, L. Wang, Y. Yu, Q. Tao, and Z. Zhu. Robust isar range alignment via minimizing the entropy of the average range profile. *IEEE Geoscience and Remote Sensing Letters*, 6(2):204–208, April 2009.

[21] Wang Ling; Zhu Dai yin; Zhu Zhao-da. Study on ship imaging using sar real data. *Journal Of Electronics and Information Technology*, 29(2):401, 2007.

[22] A. Bacci, E. Giusti, D. Cataldo, S. Tomei, and M. Martorella. ISAR resolution enhancement via compressive sensing: A comparison with state of the art SR techniques. In *2016 4th International Workshop on Compressed Sensing Theory and its Applications to Radar, Sonar and Remote Sensing (CoSeRa)*, pages 227–231, Sept 2016.

[23] R. Venkataramani and Y. Bresler. Further results on spectrum blind sampling of 2d signals. In *Proceedings 1998 International Conference on Image Processing. ICIP98 (Cat. No.98CB36269)*, volume 2, pages 752–756 vol.2, Oct 1998.