Deep learning-based reconstruction for near-field MIMO radar imaging

Irfan Manisali and Figen S. Oktem
Department of Electrical Engineering, METU, Ankara, 06800 Turkey
Email: figeno@metu.edu.tr

***Abstract*—Near-field multiple-input multiple-output (MIMO) radar imaging systems are of interest in diverse fields such as medicine, through-wall imaging, and surveillance. The imaging performance of these systems highly depends on the underlying image reconstruction method. While sparsity-based methods offer better image quality than the traditional direct inversion methods, their high computational cost is undesirable in real-time applications. In this paper, we develop a novel deep learning-based reconstruction method for near-field MIMO radar imaging. The main goal is to achieve high image quality with low computational cost. The developed approach has a two-staged structure. The physics-based first stage performs adjoint operation to back project the measurements to the reconstruction space, and DNN-based second stage converts these backprojected measurements to a scene reflectivity image. As DNN, a 3D U-Net is used to jointly exploit range and cross-range correlations. We illustrate the performance of the reconstruction method using a synthetically generated dataset. The results demonstrate the effectiveness of the developed method compared to commonly used analytical approaches in terms of image quality and computation time.**

***Index Terms*—computational imaging, deep learning, radar imaging, microwave imaging, MIMO, inverse problems**

I. Introduction

Near-field radar imaging systems are of interest in diverse fields such as medicine, through-wall imaging, airport security, concealed weapon detection, and surveillance [1]–[4]. Earlier near-field radar imaging systems have operated in monostatic mode using large number of transceiver antennas.Recently, sparse multiple-input multiple-output (MIMO) arrays with spatially distributed transmit and receive antennas have gained more attention since they can offer high resolution with reduced cost, hardware complexity, and acquisition time [1]–[3], [5].

Near-field MIMO radar imaging systems aim to reconstruct the three-dimensional (3D) reflectivity distribution of the scene from the limited radar data. As a result, the imaging performance largely depends on the underlying image reconstruction method. Traditional direct inversion methods such as backprojection (and its variants) have low computational complexity but they suffer from reconstruction artifacts when observations are limited. While sparsity-based iterative reconstruction methods offer better reconstruction quality than the traditional direct inversion methods at compressive settings [4], [6]–[8], they suffer from high computational cost which is undesirable in real-time applications.

This work is supported by the Scientific and Technological Research Council of Turkey (TUBITAK) under grant 120E505.

Recently, reconstruction techniques that exploit deep learning have emerged as an alternative [9], [10]. These methods are shown to achieve high reconstruction quality with low computational cost for various imaging problems. In the context of near-field radar imaging, deep learning-based reconstruction methods have not been studied much in the literature. Most of the proposed methods are for far-field settings in SAR/ISAR or MIMO radar imaging [11]–[18]. In near-field radar imaging, there are some works for deep learning-based approaches, but most of them apply to the monostatic setting [19], [20]. For near-field and MIMO radar imaging, there are fewer works [21] and no comprehensive study. The work in [21] develops a learning-based direct inversion method that separately processes the magnitude and phase of the backprojection images with 2D-convolutional layer blocks. Because of 2D processing, the method can not exploit range correlations. Moreover, both training and testing are performed with simple point scatterers only. Hence this method is not applicable for imaging extended targets as encountered in practice.

In this work, we develop a novel deep learning-based two-stage (*Deep2S*) method to reconstruct the 3D scene reflectivity from the near-field observations of a MIMO imaging radar. The main goal is to achieve high image quality with low computational cost so that the developed Deep2S method can be used in real-time applications. The developed approach has a two-staged structure that consists of an adjoint operation followed by a 3D U-Net architecture. The adjoint stage exploits the observation model of the system and back project the measurements to the reconstruction space. The second stage employs a 3D deep neural network which is trained to convert the backprojected measurements to a magnitude-only reflectivity image. Numerical simulations are performed for a microwave imaging setting. For both training and testing, a large synthetic dataset is randomly generated to obtain 3D scenes that involves extended targets. Using this dataset, we illustrate the performance of the developed approach under different imaging scenarios and compare the performance with back-projection and sparsity-based reconstruction methods.

II. Observation model

A sample observation geometry for near-field MIMO radar imaging is illustrated in Fig. 1. The transmit and receive antennas are spatially distributed on a planar MIMO array located at $z = 0$. Each transmit antenna illuminates a scene that lies in the near-field of the array. Using Born approximation for

the scattered field, the signal captured by the receive antenna due to a single scatterer at (x, y, z) with reflectivity $s(x, y, z)$ can be expressed in the temporal frequency domain [2] as

$$r\left(x_t, y_t, x_r, y_r, k\right) = \frac{p(k)}{4\pi d_t d_r}\, s(x, y, z)\, e^{-jk(d_t + d_r)} \tag{1}$$

where $d_t = \sqrt{(x_t - x)^2 + (y_t - y)^2 + z^2}$ and $d_r = \sqrt{(x_r - x)^2 + (y_r - y)^2 + z^2}$. Here $r\left(x_t, y_t, x_r, y_r, k\right)$ denotes the temporal frequency domain measurement obtained using the transmitter at $(x_t, y_t, 0)$ and the receiver at $(x_r, y_r, 0)$, where d_t and d_r respectively denote the distances from the corresponding transmit and receive antenna elements to the scatterer, $p(k)$ is the Fourier transform of the transmitted signal with $k = 2\pi f/c$ being the frequency-wavenumber, f the temporal frequency, and c the speed of the light.

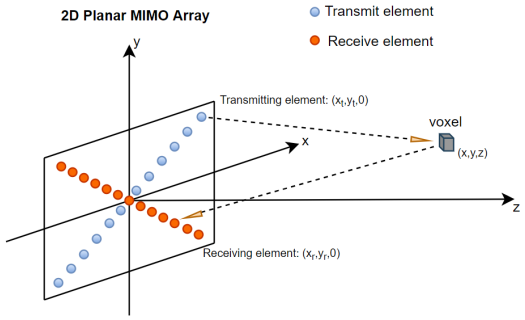

Fig. 1: Near-field MIMO Radar Imaging System.

The total received signal $\tilde{r}\left(x_t, y_t, x_r, y_r, k\right)$ due to an extended target is then given by

$$\tilde{r}\left(x_t, y_t, x_r, y_r, k\right) = \iiint \frac{p(k)}{4\pi d_t d_r}\, s(x, y, z)\, e^{-jk(d_t + d_r)}\, dx\, dy\, dz \tag{2}$$

where $s(x, y, z)$ represents the complex-valued three-dimensional reflectivity distribution of the scene.

This continuous model is converted to a discrete model by replacing the 3D reflectivity function with its discrete representation in terms of voxels. The voxel size is chosen based on the desired down-range and cross-range resolutions. We can then express the relation between the complex-valued image vector s and the measurement vector y which contains the noisy measurements obtained using different transmitter-receiver pairs and frequency steps, as follows:

$$y = As + w \tag{3}$$

where A is the observation matrix and w is the noise vector. In general, the observation matrix A is rectangular and its (m, n)th element, representing the contribution of the nth voxel to the mth measurement, can be expressed as:

$$A_{m,n} = \frac{p\left(k_m\right) e^{-jk_m d_{t_m}^{(n)}} e^{-jk_m d_{r_m}^{(n)}}}{4\pi d_{t_m}^{(n)} d_{r_m}^{(n)}} \tag{4}$$

Here the measurement index m indicates the locations of the transmitting and receiving antennas, as well as the frequency, k_m, used in this measurement. Moreover, $d_{t_m}^{(n)}$ and $d_{r_m}^{(n)}$ respectively represent the distances from the center of the nth voxel to the transmitter and receiver used.

III. DNN-based Two-Stage Reconstruction

In the inverse problem, the goal is to reconstruct the unknown reflectivity field of the scene, s, from the limited radar measurements, y. To achieve high image quality with low computational cost, we develop a physics-based learned direct reconstruction method called *Deep2S*. The method has a two-stage structure as shown in Fig. 2 where the first stage is an adjoint operation that exploits the physics-based model, and the second stage is a 3D U-Net denoiser for refinement.

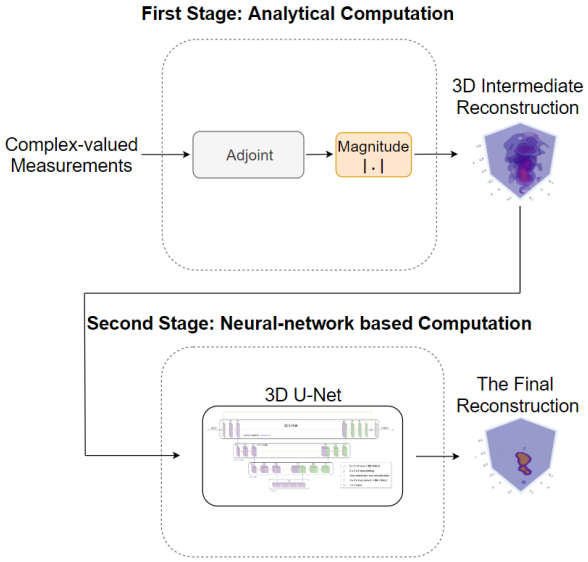

Fig. 2: DNN-based Two-Stage (Deep2S) Approach.

Since it is generally a difficult task for a network to learn the direct mapping from the measurement space to the 3D object space, we first apply the adjoint operator to the measurements to provide the network in the second stage a warm start. The adjoint operation encapsulates the physical model of the imaging system and has the benefit of fast computation due to its non-iterative nature using the hermitian of the system matrix (i.e. $A^H y$). This stage simplifies the learning process of the 3D U-Net architecture in the second stage since it back projects the radar measurements to the 3D object space.

At the second stage, a deep neural network is employed to improve this 3D intermediate result. The network is trained to convert the backprojected measurements to a magnitude-only reflectivity image. Although scene reflectivities are complex-valued, in most applications they have random phase nature, and as a result it is generally sufficient to reconstruct the magnitude of the scene reflectivity. Because of this, the magnitude of the backprojected measurements from the first stage is input to the DNN in the second stage. As DNN, a four-level 3D U-Net architecture is developed as in Fig. 3 based on the 2D U-net in [22]. This 3D U-Net architecture can jointly capture the correlations along both range and cross-range directions of a 3D extended target unlike the 2D DNN used in [21].

This architecture has an encoding and decoding path. The encoding path contains the repeated application of 3D convolutions of size $3\times3\times3$, batch normalization (BN), rectified linear unit (RELU), followed by a $2\times2\times2$ max pooling with strides of two. The decoding path consists of $3\times3\times3$ upconvolution with strides of two in each dimension, which is followed by a RELU. In the decoding path, there are also concatenations with the cropped feature maps from the encoding path. The input and output of the network is of size $25\times25\times49$ voxels.

The 3D U-Net architecture has three properties suited for our imaging problem. Firstly, due to the encoding and decoding paths, the effective receptive field of the network increases. Here, our main purpose is the refinement of the input image from the first stage. Having a large receptive field over the input image can improve the quality of the output image. Secondly, the 3D U-Net employs multichannel filters. By this way it can better extract the feature maps of its input. This increases the dimension of the latent representation of our input images, which increases the expressive power of the network. Thirdly, the 3D U-Net architecture can capture the correlation along both range and cross-range directions of the three-dimensional target, unlike the existing approaches in the literature.

Hence a feed-forward approach is obtained while incorporating the physics-based knowledge of the MIMO imaging system through the usage of the hermitian of the system matrix. The Deep2S approach has low computational complexity as desired. The computational cost of the overall approach is dominated by the 3D U-Net.

For training, the simulated radar measurements are first passed through the adjoint operation stage. The magnitude of the 3D reflectivity obtained with this adjoint operation is then input to the second stage. The DNN in the second stage is trained using these reflectivity magnitudes together with the corresponding ground-truth reflectivity magnitude of the scene, which form our training input and output respectively. For training, a synthetically generated dataset is used as explained in the next section. After training, we process our radar measurements in the test dataset with the proposed method to reconstruct the reflectivity magnitude of the unknown scene.

IV. Numerical Results

A. Simulation Setting

The sketch of the observation setting is shown in Fig. 1. As a sparse MIMO array, we consider a Mills Cross array [2], [6]. The width of the planar array is 0.3 m, which includes 12 uniformly spaced transmit antennas and 13 uniformly spaced receive antennas along its diagonals in a cross configuration. The target center is located approximately 0.5 m away from the 2D MIMO array. The frequency ranges from 4 to 16 GHz with uniformly sampled steps. In the numerical simulation, the number of frequency steps is selected as 7, 15, and 31 respectively to investigate the performance of the developed method at different compression levels.

Our goal is to infer the reflectivity image within a cube of size $0.3m\times0.3m\times0.3m$, where the voxel size is chosen as $1.25cm\times1.25cm\times0.625cm$ in x, y, and z directions, which is half of the expected theoretical resolution in each direction [2]. Then the reflectivity cube is of size $25\times25\times49$.

B. Synthetic Dataset Generation

A large experimental dataset is not available in many radar imaging applications. Because of this, the neural network is trained using a synthetic dataset that contains various randomly generated 3D targets as illustrated in Fig. 4. Firstly, the center of the object is randomly chosen from a uniform distribution between -0.05 and $0.05m$ for the x and y-axis and, 0.41 and $0.59m$ for the z-axis. Then, around the center of the object 5 virtual centers are chosen according to a Gaussian distribution with zero mean and standard deviation of 2. For every virtual center, 3 points are generated according to a Gaussian distribution with zero mean and standard deviation of 1.5. For one synthetic scene, we totally have 15 points chosen randomly within the cube. To obtain volumetric objects, these 15 points are passed through a 3D Gaussian filter with a standard deviation of 1.3. This is then passed through the sigmoid function which performs the amplitude normalization of the generated 3D targets to a maximum value of 1.

With this approach, we obtain different 3D objects that spread within the cube from a randomly chosen center. The training, validation, and test datasets contain 800, 100, and 100 images which were randomly generated in this way.

C. Training Procedure

The implementation is done using Jupyterlab 3.1.7. The training took about 8 hours using NVDIA GeForce RTX 3060. The learning rate is chosen as 10^{-3}, batch size equals as 16, and loss function is chosen as l_2 loss.

D. Results

We present the performance of the developed method by considering different compression levels in the observations. For performance comparison, the results are also obtained using the adjoint operation, backprojection (BP) and total-variation based reconstruction [6].

In Table I, the average reconstruction performance of different methods is given for 100 test images in terms of 3D PSNR and SSIM at 30 dB SNR. The number of frequency steps is increased twice each time. Hence the compression ratio is respectively %4, %8, and %16. In all cases, the Deep2S approach outperforms the other approaches in terms of PSNR and SSIM metrics. When the number of frequency steps is decreased, the reconstruction performance of all approaches decreases due to the increased ill-posedness of the problem. In the worst case when the number of frequency steps is 7, the Deep2S approach gives an average PSNR of 29.13 dB and an SSIM of 0.89. As the number of frequency steps increases, Deep2S approach passes 30 dB PSNR and 0.90 SSIM. Also, increasing the number of frequency steps from 15 to 31 does not make a significant change in the reconstruction performance. As a result, 15 frequency steps seem to be enough for this imaging scenario.

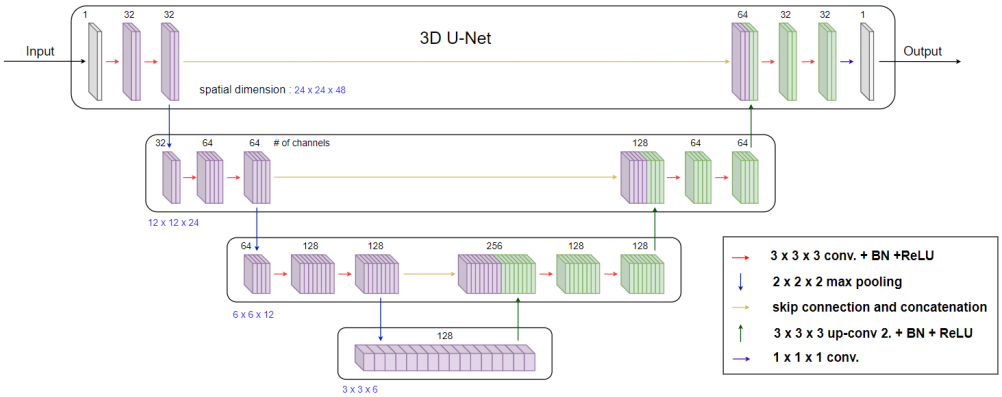

Fig. 3: Architecture of the 3D U-Net.

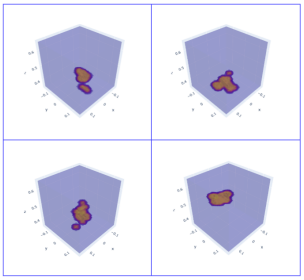

Fig. 4: Samples of Synthetically Generated Dataset (The units of x, y, and z-axis are meters).

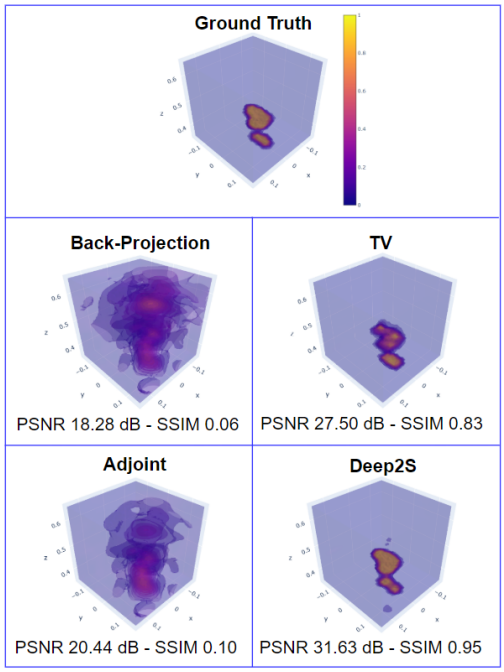

Fig. 5: Reconstructions of the different algorithms for a sample test image at 30 dB SNR (Number of Frequency Steps: 15).

Figure 5 illustrates the reconstruction performance of different methods for a sample image in the test dataset. As seen, both the TV algorithm and the Deep2S approach provide good reconstruction performance while the Deep2S approach gives a better reconstruction both visually and qualitatively. Deep2S gives nearly artifact-free result, although the adjoint operation provides a poor input to the DNN with many artifacts.

TABLE I: Average PSNR and SSIM Values for Different Number of Frequency Steps at 30 dB SNR. Best results are shown in bold.

Number of Frequency Steps	Method	PSNR (dB)	SSIM
31	Adjoint	23.0	0.35
	TV	26.3	0.83
	Deep2S	**30.4**	**0.94**
15	Adjoint	21.8	0.19
	TV	25.5	0.74
	Deep2S	**30.4**	**0.94**
7	Adjoint	20.3	0.12
	TV	22.7	0.44
	Deep2S	**29.1**	**0.89**

TABLE II: Average PSNR and SSIM Values of Adjoint Operation and BP Algorithm at 30 dB SNR.

	BP	Adjoint
PSNR	20.34	**21.77**
SSIM	0.12	**0.19**

Testing the developed DNN-based method in conditions different from the training dataset is examined in the experiment shown in Figure 6. For this purpose, we introduce a 3D target image that is not contained in the synthetically generated dataset. The 3D target image is an ellipsoid that is centered on the cube. This target image takes up more volume than the training dataset images. Also the introduced ellipsoid has a bulk large volume within the cube while the training dataset contains targets that are shapeless and randomly scattered in the volume. Figure 6 shows the results visually at 30 dB SNR (Number of Frequency Steps: 15). The best reconstruction is obtained again with the Deep2S approach both visually and qualitatively. The adjoint operation and TV reconstruction show significant volume artifacts in this

TABLE III: Average Runtime.

Adjoint	TV	Deep2S
32ms	165s	77ms

case. Moreover, the TV reconstruction fails to fill the volume inside of the ellipsoid. This result suggests that the developed approach can have robust performance for 3D targets different from the training set.

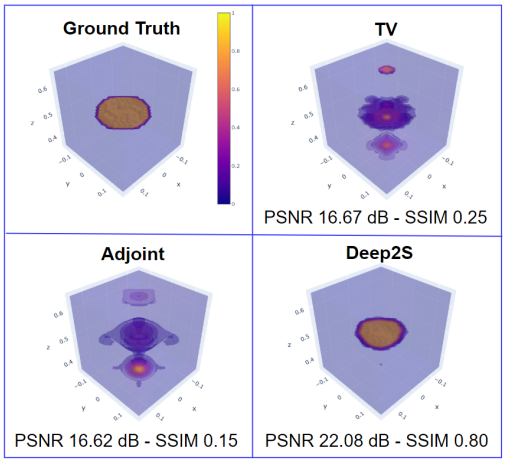

Fig. 6: Reconstructions of the different algorithms for the ellipsoid test image at 30 dB SNR (15 Frequency Steps).

In Table II, the average performance of the adjoint operation and the classical BP algorithm is given for 15 frequency steps at 30 dB SNR. The reconstruction performance of the adjoint operation is comparable (even slightly better) to the classical BP algorithm. For this reason, we use adjoint operation as the first stage of the Deep2S. The performance comparison is also carried out with the adjoint operation instead of the BP.

Table III shows the comparison of average runtime. Note that the Deep2S approach requires less than a second to reconstruct the reflectivity field of a $25 \times 25 \times 49$ scene on the CPU, while the TV reconstruction requires a runtime on the order of minutes. Hence the Deep2S approach not only surpasses the other approaches in terms of reconstruction performance but also is computationally more efficient except for the adjoint operation which gives poor performance.

V. Conclusion

In summary, we developed a novel deep learning-based approach for near-field MIMO radar imaging based on learned direct reconstruction. We demonstrate the performance of the developed method using a synthetically generated dataset and compare its performance with the commonly used analytical approaches. It has been observed that the developed method provides the best reconstruction quality while enabling fast reconstruction, which shows promise for real-time applications.

References

[1] S. S. Ahmed, A. Schiessl, F. Gumbmann, M. Tiebout, S. Methfessel, and L.-P. Schmidt, "Advanced Microwave Imaging," *IEEE Microwave Magazine*, vol. 13, no. 6, pp. 26–43, 2012.

[2] X. Zhuge and A. G. Yarovoy, "Three-Dimensional Near-Field MIMO Array Imaging Using Range Migration Techniques," *IEEE Transactions on Image Processing*, vol. 21, no. 6, pp. 3026–3033, 2012.

[3] E. Anadol, I. Seker, S. Camlica, T. O. Topbas, S. Koc, L. Alatan, F. Oktem, and O. A. Civi, "UWB 3D Near-Field Imaging with a Sparse MIMO Antenna Array for Concealed Weapon Detection," in *Radar Sensor Technology XXII*, vol. 10633. International Society for Optics and Photonics, 2018, p. 106331D.

[4] S. Li, G. Zhao, H. Li, B. Ren, W. Hu, Y. Liu, W. Yu, and H. Sun, "Near-Field Radar Imaging via Compressive Sensing," *IEEE Transactions on Antennas and Propagation*, vol. 63, no. 2, pp. 828–833, 2014.

[5] M. B. Kocamis and F. S. Oktem, "Optimal design of sparse MIMO arrays for near-field ultrawideband imaging," in *2017 25th European Signal Processing Conference (EUSIPCO)*. IEEE, 2017, pp. 1952–1956.

[6] F. S. Oktem, "Sparsity-based three-dimensional image reconstruction for near-field MIMO radar imaging," *Turkish Journal of Electrical Engineering & Computer Sciences*, vol. 27, no. 5, pp. 3282–3295, 2019.

[7] Q. Cheng, A. Alomainy, and Y. Hao, "Near-Field Millimeter-Wave Phased Array Imaging With Compressive Sensing," *IEEE Access*, vol. 5, pp. 18 975–18 986, 2017.

[8] E. A. Miran, F. S. Oktem, and S. Koc, "Sparse Reconstruction for Near-Field MIMO Radar Imaging Using Fast Multipole Method," *IEEE Access*, vol. 9, pp. 151 578–151 589, 2021.

[9] 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*, vol. 26, no. 9, pp. 4509–4522, 2017.

[10] G. Ongie, A. Jalal, C. A. M. R. G. Baraniuk, A. G. Dimakis, and R. Willett, "Deep learning techniques for inverse problems in imaging," *IEEE Journal on Selected Areas in Information Theory*, 2020.

[11] M. B. Alver, A. Saleem, and M. Çetin, "Plug-and-play synthetic aperture radar image formation using deep priors," *IEEE Transactions on Computational Imaging*, vol. 7, pp. 43–57, 2020.

[12] C. Hu, L. Wang, Z. Li, and D. Zhu, "Inverse synthetic aperture radar imaging using a fully convolutional neural network," *IEEE Geoscience and Remote Sensing Letters*, vol. 17, no. 7, pp. 1203–1207, 2019.

[13] L. Peng, X. Qiu, C. Ding, and W. Tie, "Generating 3d point clouds from a single SAR image using 3D reconstruction network," in *IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium*. IEEE, 2019, pp. 3685–3688.

[14] S. Wang, J. Guo, Y. Zhang, Y. Hu, C. Ding, and Y. Wu, "Single target SAR 3D reconstruction based on deep learning," *Sensors*, vol. 21, no. 3, p. 964, 2021.

[15] J. Chen, L. Peng, X. Qiu, C. DING, and Y. WU, "A 3D building reconstruction method for SAR images based on deep neural network," *Scientia Sinica Informationis*, vol. 49, no. 12, pp. 1606–1625, 2019.

[16] S. Zhou, Y. Li, F. Zhang, L. Chen, and X. Bu, "Automatic reconstruction of 3-D building structures for tomoSAR using neural networks," in *2019 IEEE International Conference on Signal, Information and Data Processing (ICSIDP)*. IEEE, 2019, pp. 1–5.

[17] T. Weiss, N. Peretz, S. Vedula, A. Feuer, and A. Bronstein, "Joint optimization of system design and reconstruction in MIMO radar imaging," in *2021 IEEE 31st International Workshop on Machine Learning for Signal Processing (MLSP)*. IEEE, 2021, pp. 1–6.

[18] J. Gao, B. Deng, Y. Qin, H. Wang, and X. Li, "Enhanced radar imaging using a complex-valued convolutional neural network," *IEEE Geoscience and Remote Sensing Letters*, vol. 16, no. 1, pp. 35–39, 2018.

[19] G. Sun, F. Zhang, B. Gao, Y. Zhou, Y. Xiang, and S. Pan, "Photonics-based 3D radar imaging with CNN-assisted fast and noise-resistant image construction," *Optics Express*, vol. 29, no. 13, pp. 19 352–19 361, 2021.

[20] M. Wang, S. Wei, J. Liang, X. Zeng, C. Wang, J. Shi, and X. Zhang, "RMIST-Net: Joint range migration and sparse reconstruction network for 3-D mmw imaging," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–17, 2021.

[21] Q. Cheng, A. A. Ihalage, Y. Liu, and Y. Hao, "Compressive sensing radar imaging with convolutional neural networks," *IEEE Access*, vol. 8, pp. 212 917–212 926, 2020.

[22] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in *International Conference on Medical Image Computing and Computer-assisted Intervention*. Springer, 2015, pp. 234–241.