ICASSP 2024 - 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) | 979-8-3503-4485-1/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICASSP48485.2024.10447461

IFNET: IMAGING AND FOCUSING NETWORK FOR HANDHELD MMWAVE DEVICES

Yadong Li, Dongheng Zhang, Ruixu Geng, Jincheng Wu, Yang Hu, Qibin Sun, Yan Chen*

University of Science and Technology of China, Hefei, China

ABSTRACT

Recent advancements have showcased the potential of handheld millimeter-wave (mmWave) imaging, which applies synthetic aperture radar (SAR) principles in portable settings. However, existing studies addressing handheld motion errors either rely on costly tracking devices or employ simplified imaging models, leading to impractical deployment or limited performance. In this paper, we present IFNet, a novel deep unfolding network that combines the strengths of signal processing models and deep neural networks to achieve imaging and focusing for handheld mmWave systems. By integrating multiple priors and mapping the optimization processes into an iterative network structure, IFNet effectively compensates for phase errors and recovers high-fidelity images from severely distorted signals. Extensive experiments demonstrate that IFNet outperforms state-of-the-art methods, both qualitatively and quantitatively.

Index Terms— mmWave imaging, synthetic aperture radar, deep unfolding network

1. INTRODUCTION

Millimeter-wave (mmWave) imaging is widely applicable in various domains, including security checks [1], non-destructive testing [2], and autonomous driving [3]. Its advantages of being penetrative, light-robust, and non-ionizing radiated, make it a highly promising imaging modality compared to optical cameras and X-rays [4, 5, 6].

The spatial resolution of radio imaging is inherently constrained by the aperture size, which is determined by the number of antennas and their spacing (typically half wavelength) [7, 8]. Since building large-scale antenna arrays can be expensive and challenging, synthetic aperture radar (SAR) has been extensively employed for high-resolution radio imaging. Specifically, SAR creates large virtual antenna arrays by moving the physical antennas to transceive signals at different locations [9]. However, precise tracking of the device's trajectory is crucial to ensure a coherent combination of the received signals. Consequently, existing near-field mmWave imaging systems rely on bulky motion controllers to ensure uniform and linear motion [1, 10, 11]. Driven by the need for compact and portable imaging systems, researchers have explored mmWave SAR imaging using handheld scanning [12, 13]. However, manual scanning introduces significant motion errors that corrupt the signal phase and result in severe image distortion. While approaches like [12, 13] incorporate a motion capture system to obtain accurate handheld trajectories, it is too expensive for consumer-level applications.

Traditional autofocus techniques [14, 15], aiming at compensating for motion errors through signal processing, have been widely adopted in airborne SAR. While these methods have explicit theoretical foundations, they cannot be directly applied to near-field 3D mmWave imaging because they are based on far-field assumptions [15] or face challenges in 3D imaging [14]. To address these issues, recent studies [16, 17] proposed to utilize deep neural networks, specifically conditional generative adversarial networks (cGAN), for handheld SAR imaging. Nevertheless, they simply input the distorted amplitude image into the network while neglecting the phase information which is critical for phase error estimation and motion compensation. Consequently, these black-box networks encounter difficulties in achieving promising imaging performance and strong generalization.

To achieve robust imaging and focusing for handheld mmWave systems, this paper introduces IFNet, a deep unfolding network [18] that bridges the gap between signal processing models and deep neural networks. Specifically, we first formulate the problem of phase error compensation in handheld imaging as an optimization task with multiple priors. Next, to effectively obtain the optimal solution, we unfold the iterative optimization process into a deep neural network, harnessing the powerful non-linear mapping ability of learning-based methods. Our network takes distorted complex SAR signals as input, and performs image formation and phase error compensation with separate modules, ultimately producing reconstructed high-fidelity images. By mapping the signal processing model into a deep learning framework, our approach combines the advantages of both methodologies, resulting in improved interpretability and generalizability. Extensive evaluations demonstrate the superiority of our approach over state-of-the-art methods, both qualitatively and quantitatively, with significant average improvements of 3.3 dB in peak signal-to-noise ratio (PSNR) and 24.56% in structural similarity index measure (SSIM).

Corresponding author: Dongheng Zhang. This work was supported by NSFC under Grant 62172381, National Key R&D Programmes under Grant 2022YFC2503405 and 2022YFC0869800.

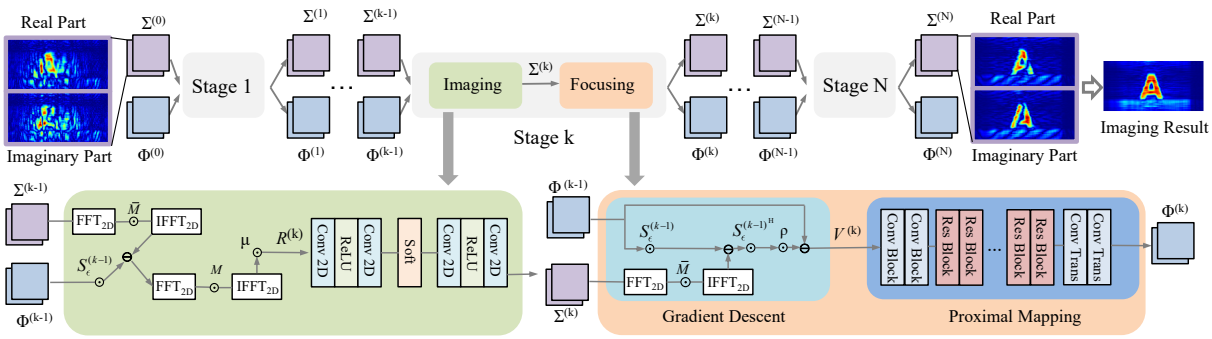

Fig. 1. The architecture of IFNet. IFNet comprises multiple stages, with each stage consisting of an imaging module and a focusing module. The design of the imaging and focusing modules are based on the optimization-based signal model.

2. METHOD

2.1. Imaging and Focusing with Sparsity and Deep Prior

The range migration algorithm (RMA) [19, 20], a widely used traditional SAR imaging method, can be defined as a pair of operators including the imaging process $\mathcal{I}(S, M)$ and signal generating process $\mathcal{G}(\Sigma, \bar{M})$ (i.e., inverse RMA), respectively [21].

$$\Sigma = IFFT_{2D}[FFT_{2D}[S] \odot M] = \mathcal{I}(S, M), \tag{1}$$

$$S = IFFT_{2D}[FFT_{2D}[\Sigma] \odot \bar{M}] = \mathcal{G}(\Sigma, \bar{M}), \tag{2}$$

where S is the received signal of a planar array, Σ is the target 2D imaging area. FFT_{2D} and $IFFT_{2D}$ denote the 2D Fast Fourier Transform (FFT) and 2D inverse FFT, respectively. M represents a specific phase term and $\bar{M}$ is the conjugate version of M. $\odot$ denotes element-wise product. In the presence of phase errors caused by handheld scanning, the signal model can be denoted as:

$$\Phi \odot S_\epsilon = \mathcal{G}(\Sigma, \bar{M}) + N, \tag{3}$$

where S_ϵ is the phase-corrupted received signal, Φ stands for the corresponding phase compensation factor, and N refers to additive white Gaussian noise. Since a majority of objects in mmWave spatial frequency domain have a high degree of sparsity, Eq. 3 can be modeled as a sparsity-driven optimization problem by adding a regularization term $\|vec(\Sigma)\|_1$. However, when the phase errors are significant, which is the case of handheld scanning, the sparsity constraint cannot guarantee satisfactory images. Since there is no prior information about the phase compensation matrix Φ, it usually results in unreliable handheld phase error estimation.

To reduce the manual scanning time and improve system efficiency, we propose to synthesize a planar array by manually moving a linear array. Hence, the handheld scanning has a certain motion pattern (a near-straight line), which means that the phase error is also subject to a specific distribution rather than completely random. Hence, prior information about Φ can be utilized to acquire stable and robust focusing results. However, handcrafted regularization is not applicable because the distribution characteristics of handheld phase error are unknown. Inspired by the success of deep neural networks, we formulate the regularization term of Φ as a deep prior, which can be learned from data both effectively and automatically. Therefore, the problem of joint handheld SAR imaging and focusing can be formulated as the problem of minimizing the following objective function:

$$\hat{\Phi}, \hat{\Sigma} = \arg\min_{\Phi,\Sigma} \frac{1}{2}\|\Phi \odot S_\epsilon - \mathcal{G}(\Sigma, \bar{M})\|_2{}^2 + \lambda\|vec(\Sigma)\|_1 + \gamma J(\Phi), \tag{4}$$

where λ is the weighting parameter indicating the strength of the regularization term $\|vec(\Sigma)\|_1$. $J(\Phi)$ is the regularization term of Φ and γ is the corresponding weighting factor. Eq. 4 can be solved by iteratively alternating between image formation (i.e., imaging part) and phase error compensation (i.e., focusing part) using a coordinate descent technique. Specifically, in the imaging part of each iteration, the cost function is minimized by updating the image σ with iterative shrinkage thresholding algorithm (ISTA):

$$R^{(k)} = \Sigma^{(k-1)} - \mu\mathcal{I}\left(\mathcal{G}(\Sigma^{(k-1)}, \bar{M}) - \Phi^{(k-1)} \odot S_\epsilon^{(k-1)}\right), \tag{5}$$

$$\Sigma^{(k)} = prox_{\lambda,\|\|_1}(R^{(k)}) = soft(R^{(k)}, \lambda), \tag{6}$$

where μ denotes the updating step size, $prox$ represents the proximal mapping operator, and $soft$ is the soft thresholding function. In the focusing part, the phase compensation factor Φ is estimated given the updated image as follows:

$$V^{(k)} = \Phi^{(k-1)} - \rho S_\epsilon^{(k-1)\mathsf{H}} \odot \left(\Phi^{(k-1)} \odot S_\epsilon^{(k-1)} - \mathcal{G}(\Sigma^{(k)}, \bar{M})\right), \tag{7}$$

$$\Phi^{(k)} = prox_J(V^{(k)}), \tag{8}$$

where ρ denotes the updating step size. Instead of deriving the proximal mapping $prox_J(V^{(k)})$ as a soft thresholding function, we formulate it as a neural network structure with improved and robust focusing ability.

2.2. Imaging and Focusing with Deep Unfolding Network

Tradition optimization method to solve Eq. 4 not only introduces a heavy computational burden due to the iterative up-

dating but also generates low-quality images because of the limited representation ability of the handcrafted regularization. Inspired by recent advances in the interpretable deep unfolding network [18, 22, 23, 24], which aims to replace the handcrafted regularization with a neural network module that has powerful learning ability, we propose IFNet by mapping the update steps in Sec. 2.1 to an end-to-end deep network structure for mmWave imaging and focusing.

The architecture of IFNet is illustrated in Fig. 1. Specifically, the network is composed of K stages, representing a total of K updating steps. Each stage consists of an imaging part and a focusing part that corresponds to image formation and phase error compensation. Specifically, since the solution to the imaging part in Sec. 2.1 is essentially an ISTA process, we adopt the design of ISTA-Net [18] and modify the specific modules according to our task. The imaging part contains the gradient descent corresponding to Eq. 5 and two convolutional blocks with a soft-thresholding function corresponding to Eq. 6. Each convolutional block consists of a convolutional layer of 8 kernels with a size of 3, a ReLU activation function, and a convolutional layer of 2 kernels with a size of 3. The focusing part is composed of a gradient descent step corresponding to Eq. 7 and an encoder-decoder module corresponding to Eq. 8. The encoder-decoder module consists of two convolutional blocks with the first one having 32 convolution kernels with a size of 7 and the second one having 64 convolution kernels with a size of 3. Each convolution layer is followed by an instance normalization layer and a ReLU layer. Then, we employ 5 ResBlock [25] to extract more abstract features and 2 transpose convolution blocks to reconstruct the images. The derived solution from Eq. 4 to Eq. 8 includes several hyperparameters, i.e., λ, μ, ρ, which need to be adjusted manually in the traditional optimization process. Similar to the existing work [18], we set all of them to be learnable parameters that can be automatically optimized by the network.

The input of IFNet is the SAR image distorted by the manual scanning phase errors. Different from prior work that only utilized the amplitude of the complex images, we preserve the phase information by separating the real part and imaginary part and concatenating them as a 2-channel image. This helps the network to better capture the correlation between the images and phase errors. To facilitate the training process, we perform min-max normalization on the real and imaginary parts, respectively. The network is optimized with a loss function of mean square error, which measures the difference between the distorted images and well-focused images.

3. EXPERIMENTS

3.1. Implementation Details

Device Configuration. To collect SAR data, we employ the TI MMWCAS-RF-EVM radar consisting of 12 transmitting antennas and 16 receiving antennas. The radar operates at 76-81 GHz and can formulate an 86-element linear array with MIMO technology. We set the radar parameters as start frequency, 77 GHz; chirp slope, 38.5 MHz/μs; chirp duration, 40 μs; the number of ADC samples, 256; ADC sampling rate, 8 Msps.

Network Training. The dimension of the input and output SAR image of the network is $256 \times 128 \times 2$. The initial values of hyperparameters are set as $\lambda = 0.01, \mu = \rho = 0.5$. We implement the network with PyTorch and train it for 30 epochs with a learning rate of 0.0001 and a batch size of 64. To demonstrate the superiority of IFNet, we have implemented the cGAN architecture proposed in [16] as a baseline and made a comparison.

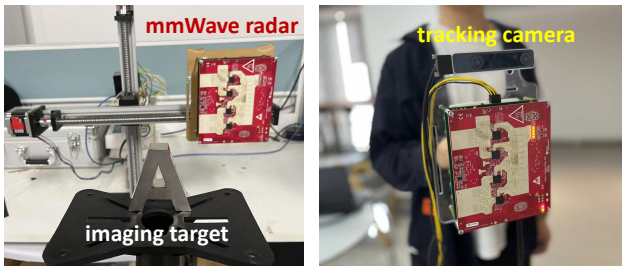

Fig. 2. Collection of ground truths and handheld scanning.

3.2. Dataset Collection

Large-scale training datasets are crucial for ensuring the robustness of data-driven methods. However, there is currently no publicly available handheld SAR imaging dataset, and creating such a dataset can be highly time-consuming. To address this challenge, we adopt a hybrid approach to generate sufficient training data. Firstly, as depicted in Fig. 2, we utilize a motion controller to collect SAR signals without phase errors, which serve as the ground truth. Next, we record 200 scanning trajectories performed by four volunteers using the Intel RealSense T265 tracking camera, as shown in Fig. 2. To diversify the dataset, we randomly select five trajectories and calculate their mean trajectory. This allows us to synthesize a large number of different handheld trajectories that can produce different phase errors. Finally, we compute the phase errors by measuring the differences between the handheld trajectories and the motion controller trajectories. These phase errors are then applied to the ground-truth images, generating distorted images that reflect the motion errors.

The dimension of the synthetic planar array is 200×86. The distance between the target and the array plane is 0.3 m, and the size of the target is 6 cm$\times$ 6 cm. In total, our dataset comprises 10,400 images from 26 different letters, with each letter distorted by 400 different handheld trajectories. To evaluate the performance of our model on new objects, we use 8,000 images from 20 letters for training and reserve 2,400 images from the remaining 6 letters for testing.

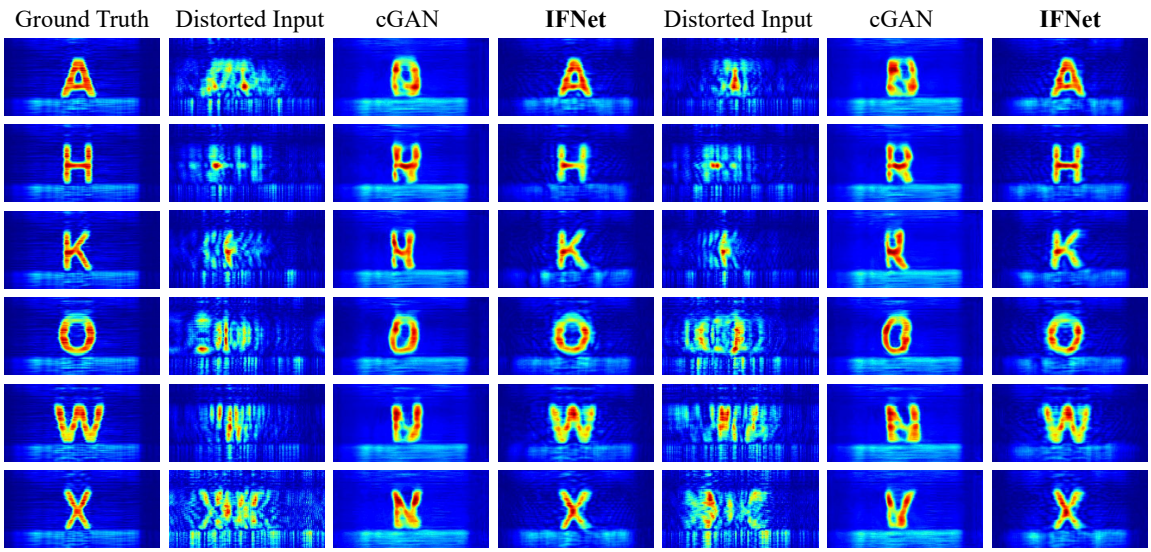

Fig. 3. Qualitative comparison. Even for severely distorted images, IFNet can successfully reconstruct the shape of the target. In contrast, although cGAN [16] effectively reduces the defocus effect, it fails to accurately restore the shape for new objects.

3.3. Qualitative Results

The qualitative comparison between IFNet and cGAN is presented in Fig. 3. The results clearly demonstrate that the image quality generated by IFNet surpasses that of cGAN. This improvement can be attributed to the fact that IFNet is constructed based on an optimization-based signal model, which enhances its imaging and focusing capabilities for new objects. In contrast, while cGAN is able to mitigate the defocus effect, it struggles to generate accurate target shapes due to limited generalization.

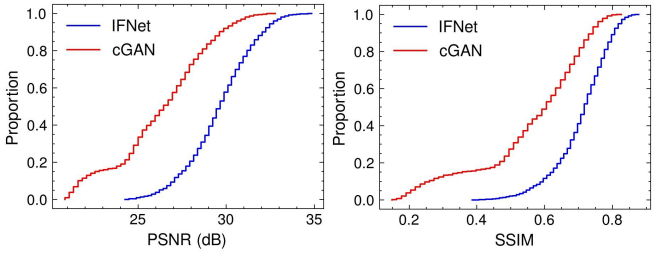

Fig. 4. The ECDF of PSNR and SSIM.

3.4. Quantitative Evaluation

Fig. 4 depicts the empirical cumulative distribution function (ECDF) of the PSNR and SSIM for the images generated by IFNet and cGAN. It is evident that IFNet outperforms cGAN in both metrics. The average PSNR and SSIM values for images produced by IFNet are 29.69 dB and 0.71, respectively, while for cGAN, the PSNR and SSIM values are 26.39 dB and 0.57, respectively. This indicates that IFNet excels in reconstructing target shapes and producing more accurate images.

3.5. Parameter Size

In addition to its superior performance, IFNet also exhibits a significantly smaller model size compared to cGAN, as illustrated in Table 1. The parameter size of IFNet is merely one-fifth of that of cGAN. This discrepancy can be attributed to the incorporation of phase information in IFNet, which greatly assists the network in accurately estimating phase errors. In contrast, cGAN solely relies on the amplitude image, and as a result, it faces challenges when reconstructing images for new or unfamiliar objects.

Table 1. Comparison of the number of network parameters

Models	cGAN	IFNet
Parameters	17.23 M	3.43 M

4. CONCLUSION

In this paper, we introduced IFNet, a deep unfolding network that combines signal processing models and deep neural networks to address the challenges of handheld mmWave imaging. By integrating multiple priors and employing an iterative network structure, IFNet effectively compensated for phase errors and recovered high-fidelity images from severely distorted signals. Extensive experiments demonstrated the superior performance of IFNet compared to state-of-the-art methods both quantitatively and qualitatively. The proposed IFNet has the potential to significantly advance the field of handheld mmWave imaging and open new avenues for practical applications.

5. REFERENCES

[1] Shichao Li and Shiyou Wu, "Low-cost millimeter wave frequency scanning based synthesis aperture imaging system for concealed weapon detection," *IEEE Transactions on Microwave Theory and Techniques*, vol. 70, no. 7, pp. 3688–3699, 2022.

[2] Mohamed A. Abou-Khousa, Mohammed Saif Ur Rahman, Kristen M. Donnell, and Mohammad Tayeb Al Qaseer, "Detection of surface cracks in metals using microwave and millimeter-wave nondestructive testing techniques—a review," *IEEE Transactions on Instrumentation and Measurement*, vol. 72, pp. 1–18, 2023.

[3] Sujeet Milind Patole, Murat Torlak, Dan Wang, and Murtaza Ali, "Automotive radars: A review of signal processing techniques," *IEEE Signal Processing Magazine*, vol. 34, no. 2, pp. 22–35, 2017.

[4] Jinbo Chen, Dongheng Zhang, Zhi Wu, Fang Zhou, Qibin Sun, and Yan Chen, "Contactless electrocardiogram monitoring with millimeter wave radar," *IEEE Transactions on Mobile Computing*, pp. 1–17, 2022.

[5] Wenxuan Li, Dongheng Zhang, Yadong Li, Zhi Wu, Jinbo Chen, Dong Zhang, Yang Hu, Qibin Sun, and Yan Chen, "Real-time fall detection using mmwave radar," in *ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2022, pp. 16–20.

[6] Bin-Bin Zhang, Dongheng Zhang, Yadong Li, Yang Hu, and Yan Chen, "Unsupervised domain adaptation for rf-based gesture recognition," *IEEE Internet of Things Journal*, pp. 1–1, 2023.

[7] Ali Raza Barket, Weidong Hu, Bing Wang, Waseem Shahzad, and Jabir Shabbir Malik, "Selection criteria of image reconstruction algorithms for terahertz short-range imaging applications," *Opt. Express*, vol. 30, no. 13, pp. 23398–23416, Jun 2022.

[8] Yadong Li, Dongheng Zhang, Jinbo Chen, Jinwei Wan, Dong Zhang, Yang Hu, Qibin Sun, and Yan Chen, "Towards domain-independent and real-time gesture recognition using mmwave signal," *IEEE Transactions on Mobile Computing*, pp. 1–15, 2022.

[9] Kun Qian, Zhaoyuan He, and Xinyu Zhang, "3d point cloud generation with millimeter-wave radar," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 4, no. 4, dec 2020.

[10] Amir Mirbeik, Robin Ashinoff, Tannya Jong, Allison Aued, and Negar Tavassolian, "Real-time high-resolution millimeter-wave imaging for in-vivo skin cancer diagnosis," *Scientific Reports*, vol. 12, no. 1, pp. 4971, 2022.

[11] Ying He, Dongheng Zhang, and Yan Chen, "3d radio imaging under low-rank constraint," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 33, no. 8, pp. 3833–3847, 2023.

[12] Guillermo Álvarez Narciandi, Jaime Laviada, and Fernando Las-Heras, "Freehand mm-wave imaging with a compact mimo radar," *IEEE Transactions on Antennas and Propagation*, vol. 69, no. 2, pp. 1224–1229, 2021.

[13] Guillermo Álvarez Narciandi, Jaime Laviada, Yuri Álvarez López, Guillaume Ducournau, Cyril Luxey, Cybelle Belem-Goncalves, Frederic Gianesello, Nour Nachabe, Carlos Del Rio, and Fernando Las-Heras, "Freehand system for antenna diagnosis based on amplitude-only data," *IEEE Transactions on Antennas and Propagation*, vol. 69, no. 8, pp. 4988–4998, 2021.

[14] J. Wang and X. Liu, "Sar minimum-entropy autofocus using an adaptive-order polynomial model," *IEEE Geoscience and Remote Sensing Letters*, vol. 3, no. 4, pp. 512–516, 2006.

[15] D.E. Wahl, P.H. Eichel, D.C. Ghiglia, and C.V. Jakowatz, "Phase gradient autofocus-a robust tool for high resolution sar phase correction," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 30, no. 3, pp. 827–835, 1994.

[16] Jaime Laviada, Guillermo Álvarez Narciandi, and Fernando Las-Heras, "Artifact mitigation for high-resolution near-field sar images by means of conditional generative adversarial networks," *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1–11, 2022.

[17] Christos Vasileiou, Josiah Smith, Shiva Thiagarajan, Matthew Nigh, Yiorgos Makris, and Murat Torlak, "Efficient cnn-based super resolution algorithms for mmwave mobile radar imaging," in *2022 IEEE International Conference on Image Processing (ICIP)*, 2022, pp. 3803–3807.

[18] Jian Zhang and Bernard Ghanem, "Ista-net: Interpretable optimization-inspired deep network for image compressive sensing," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.

[19] Xiaodong Zhuge and Alexander 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.

[20] Muhammet Emin Yanik, Dan Wang, and Murat Torlak, "Development and demonstration of mimo-sar mmwave imaging testbeds," *IEEE Access*, vol. 8, pp. 126019–126038, 2020.

[21] Mou Wang, Shunjun Wei, Zichen Zhou, Jun Shi, Xiaoling Zhang, and Yongxin Guo, "3-d sar autofocusing with learned sparsity," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–18, 2022.

[22] Kai Zhang, Luc Van Gool, and Radu Timofte, "Deep unfolding network for image super-resolution," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020.

[23] Wenhui Wu, Jian Weng, Pingping Zhang, Xu Wang, Wenhan Yang, and Jianmin Jiang, "Uretinex-net: Retinex-based deep unfolding network for low-light image enhancement," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2022, pp. 5901–5910.

[24] Yan Yang, Jian Sun, Huibin Li, and Zongben Xu, "Deep admm-net for compressive sensing mri," in *Advances in Neural Information Processing Systems*, D. Lee, M. Sugiyama, U. Luxburg, I. Guyon, and R. Garnett, Eds. 2016, vol. 29, Curran Associates, Inc.

[25] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016.