This article has been accepted for publication in IEEE Transactions on Pattern Analysis and Machine Intelligence. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TPAMI.2023.3301553

CoIR: Compressive Implicit Radar

Sean M. Farrell, Vivek Boominathan, Nathaniel Raymondi, Ashutosh Sabharwal, Ashok Veeraraghavan

Abstract—Using millimeter wave (mmWave) signals for imaging has an important advantage in that they can penetrate through poor environmental conditions such as fog, dust, and smoke that severely degrade optical-based imaging systems. However, mmWave radars, contrary to cameras and LiDARs, suffer from low angular resolution because of small physical apertures and conventional signal processing techniques. Sparse radar imaging, on the other hand, can increase the aperture size while minimizing the power consumption and read out bandwidth. This paper presents CoIR, an analysis by synthesis method that leverages the implicit neural network bias in convolutional decoders and compressed sensing to perform high accuracy sparse radar imaging. The proposed system is data set-agnostic and does not require any auxiliary sensors for training or testing. We introduce a sparse array design that allows for a $5.5\times$ reduction in the number of antenna elements needed compared to conventional MIMO array designs. We demonstrate our system's improved imaging performance over standard mmWave radars and other competitive untrained methods on both simulated and experimental mmWave radar data.

Index Terms—mmWave imaging, sparse array radar, implicit neural representations, compressed sensing

◆

1 INTRODUCTION

DEPTH imaging is a crucial component in many applications such as simultaneous localization and mapping (SLAM) [1], autonomous driving [2], and security monitoring [3]. Typically, these depth imaging applications are accomplished using a combination of visual cameras, LiDAR, and inertial sensors [4]. Visual cameras provide a high angular resolution image of the environment that can be used for near-field dense depth imaging with stereo systems or monocular depth estimation algorithms [5]. LiDARs directly output a dense point cloud of the environment with high range and angular resolutions. However, since visual cameras and LiDARs operate at optical wavelengths their depth estimation performance is significantly reduced in visually degraded environments containing low light, fog, smoke, snow, and dust [2], [5]. These natural occurrences are especially problematic for depth imaging applications that involve robot and human interactions such as disaster relief and autonomous self-driving scenarios.

Another depth sensing modality commonly used is millimeter wave (mmWave) radar. Since these radars operate at millimeter wavelengths, they can penetrate through environments with airborne particles common in fog and smoke without significant performance degradation [6]. Additionally, recent availability of low-cost and low-power single chip 77-81 GHz RF bandwidth radars make these devices favorable for integration into low form-factor and power-constrained systems [7], [8]. The main limitation of using single chip mmWave radars for depth imaging is their low angular resolution $\delta \approx \lambda/d$, which is a function of the operating wavelength λ and aperture size d. A naive approach to increasing the angular resolution is to create a larger aperture which in turn can lead to antenna coupling [9] and increased power consumption and read-out bandwidth. For example, to achieve an angular resolution of 15° for a 77 GHz linear array at Nyquist sampling would require approximately 8 antennas. If the desired angular resolution was increased to 1° this would require around 115 antennas. In practice, designing such large arrays to achieve high angular resolution mmWave imaging is often difficult or expensive.

There are several techniques used to circumvent the low angular resolution of mmWave radars: 1) synthetic aperture radars (SAR) [3] and multiple-input multiple-output (MIMO) arrays [10], [11], 2) sensor fusion [5], 3) optimization with hand-crafted priors [12], [13], [14], [15], and 4) deep learning [2], [16], [17], [18], [19], [20], [21]. These techniques either have slow acquisition times, increased hardware complexity and calibration, require large data sets, or have limited generalizability.

We propose CoIR, an analysis by synthesis method that leverages implicit neural representations (INR) and compressed sensing to perform high accuracy sparse radar imaging. We design a sparse linear array that allows for a $5.5\times$ reduction in the required receiving antennas compared to conventional MIMO linear arrays. This sub-Nyquist sampling leads to a compressed sensing inverse problem, where the objective is to estimate an image of a scene's reflectivity distribution from an under-sampled set of radar measurements. To solve this problem, our key enabling observation is that the INRs from untrained neural networks have been shown to provide an inductive bias towards natural solutions over various imaging inverse problems [22], [23], [24], [25]. In this paper, we develop an untrained neural network reconstruction method, shown in Fig.1, for sparse radar imaging and demonstrate its superior performance against other competing untrained methods on both simulated and experimental data.

Contributions: We propose CoIR, a new method for achieving high accuracy, sparse radar imaging using a single radar chip. Our code is available online[1].We make the following contributions:

- *S.M.F, V.B., N.R., A.S., A.V. are with the Department of Electrical and Computer Engineering, Rice University, Houston, TX.*
  E-mail: smf5@rice.edu

[1] https://github.com/sfarrell1/supplement-CoIR.git

This article has been accepted for publication in IEEE Transactions on Pattern Analysis and Machine Intelligence. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TPAMI.2023.3301553

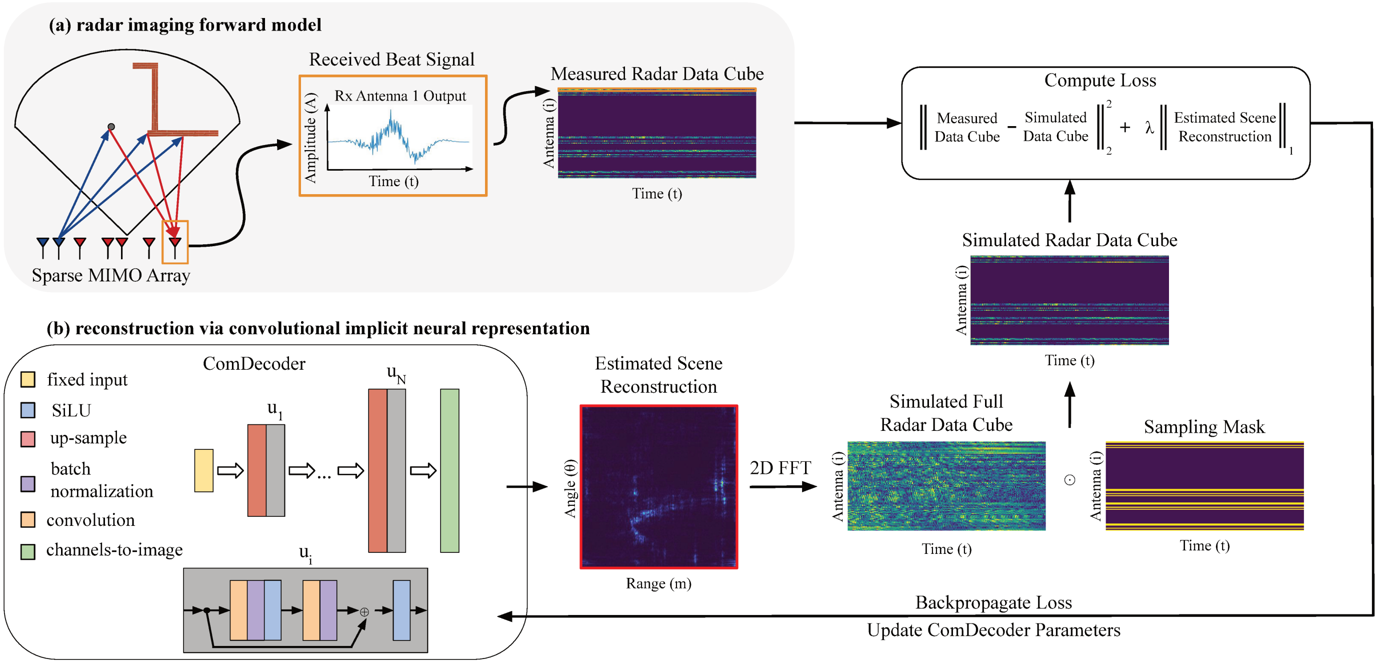

Fig. 1: **Proposed Method CoIR:** (a) In the radar imaging forward model a transmitter illuminates a scene with a mmWave FMCW pulse. Objects in the scene reflect some of the incident mmWave energy back to a sparse linear receive array. Each antenna in the receive array measures a time domain beat signal, which corresponds to a single row in the measured radar data cube. (b) ComDecoder represents the scene's reflectivity distribution (red box) using a convolutional neural network decoder. From fixed random noise ComDecoder outputs a complex valued polar image of the scenes reflectivity distribution. Using a 2D FFT, a simulated full radar data cube is synthesized. Next the rows of the simulated data cube corresponding to the sparse array antenna locations are retained. During the reconstruction, the weights of ComDecoder are updated to minimize the difference (loss) between the acquired radar data cube and the simulated radar data cube.

- We present a sparse linear array design that achieves a $5.5\times$ reduction in the number of receive antennas compared to conventional MIMO array design [26].
- We propose a fully convolutional INR architecture called ComDecoder, which is an adaptation of ConvDecoder [27] improvised for better performance.
- We compare ComDecoder against competing state-of-the-art untrained methods: Deep Image Prior (DIP), SIREN, and ℓ_1-norm gradient descent used in computer vision and radar imaging. We evaluate all methods on both simulated and experimental mmWave data from the ColoRadar dataset [4].

2 RELATED WORK

CoIR draws motivation from prior work related to mmWave imaging, compressed sensing, and implicit neural representations. We discuss relevant prior works in this section.

mmWave Imaging Systems. In recent years, there has been an increased interest in mmWave sensing applications due to the affordability of commercial sensors and increased bandwidth compared to sub-6 GHz systems [28], [29]. Prior works have used mmWaves for object tracking [30], [31], [32], human activity sensing [33], [34], [35], and material identification [36]. The main challenge of transitioning from sensing to imaging with mmWave systems is their low angular resolution, a factor of 10x below conventional optical imaging methods. To increase the angular resolution of mmWave systems and achieve imaging functionality, prior works exploit large physical arrays [37], MIMO arrays [10], [11], or SAR techniques [3].

Large physical arrays allow for real-time image capture but are expensive to build and produce large volumes of readout data. SAR techniques use spatial multiplexing to synthetically produce a large array from a small physical array. While this approach allows for low cost, single chip radars to achieve high imaging resolutions, the SAR technique typically results in slow imaging rates and bulky scanner systems. Recent work [38], [39] has investigated handheld mmWave SAR imaging techniques to circumvent the challenges of traditional SAR systems. Though handheld SAR techniques allow for increased imaging flexibility, these approaches require extensive preprocessing of radar data to compensate for motion errors and radar position ambiguity. Moreover, SAR methods still require a static imaging scene. Conventional MIMO techniques leverage multistatic radar principles to reduce the number of physical array elements but typically require a dense virtual array and thus a cascade of radar chips to achieve large apertures [7]. CoIR differs from these previous works in that we increase the aperture of our imaging system by leveraging a sparse MIMO array design that only requires a single radar chip.

More recent works have begun exploring the applications of sensor fusion [5], [40] and deep learning [2], [16], [17], [18], [19], [20], [21] to enhance the imaging

This article has been accepted for publication in IEEE Transactions on Pattern Analysis and Machine Intelligence. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TPAMI.2023.3301553

resolution and functionality of mmWave systems. Sensor fusion exploits the combination high resolution imaging modalities (optical cameras or LiDAR) with mmWave systems to increase the overall imaging resolution. However, if one sensing modality fails, it causes the performance of the whole system to degrade [5]. The closest deep learning approaches to our work are [17] and [19]. [17] proposes a sparse 2D MIMO array design for near-field mmWave imaging and exploits a trained CNN to increase the imaging resolution. [19] operates in the far-field and uses a trained U-net CNN to increase the imaging resolution of a single chip mmWave radar. The network is trained on low resolution mmWave image and high resolution LiDAR image pairs of indoor scenes. The challenge with the first approach [17] is it's limited to near-field scenes which require prior knowledge about the depth of the reflector, while latter approach [19] is tuned for indoor scenes hence limiting generalizability, and both approaches require a large training data set. CoIR, in contrast, operates in the far-field and leverages an untrained neural network to increase the resolution for a single chip mmWave sparse array radar circumventing the need for a training data set.

Sparse Radar Imaging. Our work is also related to sparse radar imaging, which we broadly categorize into two groups: sparse aperture array design and sparse reconstruction methods. Sparse aperture designs apply sub-Nyquist sampled physical arrays or MIMO virtual arrays to reduce the number of antenna elements required for high resolution imaging applications [26], [41], [42], [43], [44], [45]. These approaches are optimization-based techniques that either use convex relaxations or prior knowledge of the number of reflectors in the scene. CoIR's array design approach is inspired by [26] two step design method. However, due to our hardware integer constraint on antenna positions we cannot directly use the convex optimization approach in [26], and instead perform a grid search over feasible solutions.

Traditional sparse reconstruction methods either leverage eigenvalue super resolution methods such as MUSIC [46] and ESPRIT [47] or compressed sensing (CS) optimization techniques [12], [13], [14], [15]. Super resolution algorithms MUSIC and ESPRIT have limited practical use with radar imaging since they require signals arriving from different reflectors to be incoherent and the number of targets known. Classic CS approaches require a hand-crafted sparsity prior to limit the feasible set of solutions. For example, one approach is assuming the mmWave image is sparse in the spatial domain (i.e. the image is composed of a few reflectors) [13], [15]. Another well-established prior is total-variation norm, which enforces sparsity in the spatial domain gradients (i.e. the image is piece-wise smooth) [14]. While both of these untrained priors have shown to be reasonable in practice for sparse mmWave imaging, they can be challenging to design and are strongly scene dependent. Contrary to classic CS approaches, CoIR uses the inductive bias of untrained neural networks as a complex prior that has an affinity for solutions with visually salient features over a wide range of scenes while retaining a high impedance to noise.

Implicit Neural Representations. INRs have shown an implicit bias towards smooth natural solutions in a wide class of computer vision inverse problems. INRs can be divided broadly into convolution and MLP based architectures. Convolutional methods have shown good performance on compressed sensing problems [22], [23]; image super resolution [25]; image denoising [48]; and accelerated MRI imaging [27]. Coordinate-based MLP methods have shown successful performance on novel view synthesis [24], dynamic structure illumination [49], partial differential equation solutions [50], and image deconvolution [51]. CoIR expands on the previously mentioned INR methods and proposes a new convolutional decoder architecture tailored for sparse radar imaging: ComDecoder.

3 RADAR IMAGING BACKGROUND

In this section we formulate the MIMO radar signal and system models, including all notable modeling assumptions. We consider a MIMO radar array equipped with M_t transmit antennas and M_r receive antennas, which are arranged to form an M element uniform virtual array aperture with $d = \lambda/2$ inter-element spacing [9]. A static scene is probed by the radar system via a frequency modulated continuous waveform (FMCW) signal

$$y_{tx}(t) = e^{j2\pi(f_0 t + 0.5\frac{B}{T}t^2)} \quad 0 \leq t \leq T \tag{1}$$

where f_0 is the carrier frequency, B the chirp bandwidth, and T the pulse duration. We assume the transmitted signals are separated across the array[2] through time, code, or frequency multiplexing [52].

The scene is composed of multiple point targets, each of which reflects the impinging transmitted signal energy back to the radar array. To model the scene, we define the discrete reflectivity distribution $\bar{\mathbf{x}} \in \mathbb{C}^{K\times L}$ over a discrete polar coordinate region. Embedded in $\bar{\mathbf{x}}$ is the hardware gains, propagation losses, channel phase offsets, and complex reflectivity for each reflector. Let $n_r = 0, 1, ..., K-1$ index the range bins and $n_\theta = 0, 1, ..., L-1$ index the angle bins defining the discrete reflectivity distribution. The physical ranges and angles for each bin are contained in $\mathbf{b}_r \in \mathbb{R}^{K\times 1}$ and $\mathbf{b}_\theta \in \mathbb{R}^{L\times 1}$, respectively.

The signal received at each antenna $m = 1, \cdots, M$ in the virtual receiver aperture is dechirped [53] and sampled at the Nyquist sampling rate [54] to yield,

$$z[n,m] = \sum_{n_r=0}^{K-1}\sum_{n_\theta=0}^{L-1} \bar{x}[n_r,n_\theta] e^{j2\pi(f_0+\frac{B}{N}n)(\frac{d}{c})\sin(b_\theta[n_\theta])m} \times \\ e^{j2\pi(\frac{B}{N})(\frac{2b_r[n_r]}{c})n} + w[n,m], \quad n = 0, \cdots, N-1. \tag{2}$$

Here $z[n,m]$ is the n-th sample of the signal received at the m-th antenna and $w[n,m] \sim \mathcal{CN}\left(0,\sigma^2\right)$ is complex additive white Gaussian noise. See Supplementary Sec. 1.1 for additional details on deriving Eq. 2.

In Eq. 2 we have made use of the Born approximation such that multiple scattering events between point reflectors are negligible. Furthermore we have assumed that the

---

[2] In practical applications, it is not favorable to activate all the transmitters in the array within the same time-frequency allocation due to inter-element self-interference.

receiver array is in the far-field relative to the scene, hence the wavefront impinging the receiver array is modeled as a planar waveform. We additionally employ the narrowband assumption, which is met when $B << f_0$ [9]. Defining the angle-dependent spatial frequency

$$\psi_\theta(n_\theta) = f_0 \frac{d}{c} \sin(b_\theta[n_\theta]),$$

and the normalized range-dependent temporal frequency

$$\psi_r(n_r) = \left(\frac{B}{N}\right)\left(\frac{2b_r[n_r]}{c}\right),$$

the received signal can be rewritten

$$z[n,m] = \sum_{n_r=0}^{K-1} \sum_{n_\theta=0}^{L-1} \bar{x}[n_r, n_\theta] e^{j2\pi\psi_\theta(n_\theta)m} e^{j2\pi\psi_r(n_r)n} + w[n,m]. \tag{3}$$

Importantly, Eq. 3 is space-time separable and has a form similar to the two-dimensional Fourier transform. The forward model is compactly written across space and time as

$$\mathbf{z} = \mathcal{F}(\bar{\mathbf{x}}) + \mathbf{w}, \tag{4}$$

where $\mathcal{F}(\cdot)$ implements the 2D FFT. The advantage of this approach is it efficiently allows for the computation of a MIMO radar data cube measurement given a scene reflectivity distribution. In practice, typically the magnitude $|\bar{\mathbf{x}}|$ is used to visualize the reflectivity distribution [2], [5], [9], [17], [19], [21], [52]. The predicted $|\bar{\mathbf{x}}|$ may not theoretically match the ground truth reflectivity magnitude due to $\bar{\mathbf{x}}$ encapsulating both the true complex scene reflectivity, channel gains, and a range dependent phase offset. Nevertheless, we find that this strategy facilitates both reconstruction accuracy and computational efficiency in practice.

4 PROPOSED METHOD

We consider the problem of sparse mmWave radar imaging. An illustration of our proposed method, CoIR, is presented in Fig. 1.

Our goal is to recover an image of a scene's complex reflectivity distribution $\bar{\mathbf{x}} \in \mathbb{C}^{K \times L}$ from a single under-sampled measured radar data cube $\mathbf{z} \in \mathbb{C}^{N \times M}$. Using the MIMO radar forward model outlined in Eq. 4, the radar data cube measurements can be obtained as,

$$\mathbf{z} = \mathbf{M} \odot \mathcal{F}(\bar{\mathbf{x}}) + \mathbf{w}, \tag{5}$$

where $\mathcal{F}(\cdot)$ implements the 2D FFT, $\odot$ is the Hadamard product, $\mathbf{M} \in \{0,1\}^{N \times M}$ is a binary mask that implements under-sampling, and $\mathbf{w} \in \mathbb{C}^{N \times M}$ is complex white Gaussian noise with zero mean and variance σ^2. Recall the measurements in $\mathbf{z}$, the radar data cube, are FMCW beat signal time samples captured at each antenna. If we had a full uniform linear array with antenna spacing $d = \lambda/2$, the mask $\mathbf{M}$ would be a matrix of ones, and the radar measurements would be acquired according to Eq. 4. In this case, we can estimate the image up to the uncertainty of the additive noise as $\hat{\mathbf{x}} = \mathcal{F}^{-1}(\mathbf{z})$, where $\mathcal{F}^{-1}(\cdot)$ is the inverse 2D FFT.

Sparse radar imaging can be employed to reduce the number of antennas needed for imaging, leading to decreased power consumption and read-out bandwidth. In this work, we will focus on sparse radar imaging where the number of antennas is under-sampled, resulting in a sparse linear array. This can be modeled by multiplying $\mathbf{z}$ by a binary mask $\mathbf{M}$, which sets a subset of rows in $\mathbf{z}$ to zero. The problem then amounts to recovering an image of the scene reflectivity distribution from an under-sampled set of measurements which can be classified as a compressed sensing problem. In this compressed sensing regime, using $\mathcal{F}^{-1}(\cdot)$ to estimate the image $\hat{\mathbf{x}}$ will result in aliasing artifacts from the large sidelobes in the sparse array's point spread function (PSF), making it difficult to discriminate reflectors in the image.

We propose optimizing the weights of an untrained deep convolutional network to invert Eq. 5. Given an untrained network $G(\cdot;p) : \mathbb{R}^{c_0 \times K_0 \times L_0} \rightarrow \mathbb{R}^{c \times K \times L}$ which takes in as input $\mathbf{C} \in \mathbb{R}^{c_0 \times K_0 \times L_0}$ and is parameterized by weights $p \in \mathbb{R}^W$, producing an output $G(\mathbf{C};p) \in \mathbb{R}^{c \times K \times L}$. In this approach, the weights of the deep network are optimized such that the forward model applied to the network output matches the given radar measurements $\mathbf{z}$. We initialize $G(\mathbf{C};p)$ with a fixed $\mathbf{C}$ drawn from a uniform distribution i.i.d entries and solve an optimization problem of the form,

$$\hat{p} = \underset{p}{\operatorname{argmin}} \, ||\mathbf{z} - \mathbf{M} \odot \mathcal{F}(G(\mathbf{C};p))||_2. \tag{6}$$

In CoIR, we additionally leverage sparsity in the image domain by applying the ℓ_1-norm on the network output. This helps enforce the prior that mmWave images are sparse in the spatial domain due to the dominant specular reflections from objects at this wavelength. Thus, the final optimization process becomes,

$$\hat{p} = \underset{p}{\operatorname{argmin}} \, ||\mathbf{z} - \mathbf{M} \odot \mathcal{F}(G(\mathbf{C};p))||_2 + \lambda_L ||G(\mathbf{C};p)||_1, \tag{7}$$

where λ_L is the ℓ_1-norm hyperparameter. The estimated image of the scene is given as $\hat{\mathbf{x}} = G(\mathbf{C};\hat{p})$. This approach is very similar to traditional sparse recovery, except here we are optimizing over $\bar{\mathbf{x}}$ in the range of a deep network conditioned on the ℓ_1-norm ball. In traditional compressed sensing, the optimization is only constrained on the ℓ_1-norm ball [12], [15]. In CoIR, the deep network's convolutional structure has an implicit biased towards smooth, natural images while maintaining a high impedance to noise. We find that over-parameterizing the deep network by a factor of around 13 and adding the ℓ_1-norm prior helps to find a solution that balances fitting the salient features in the scene while suppressing noise and aliasing artifacts. See Supplementary Sec. 2.4 for analysis on over-parameterization.

CoIR addresses several key challenges related to sparse radar imaging: the design of the sparse radar aperture layout and neural network architecture. The remainder of the paper is organized as follows. First, we discuss the design choice for CoIR's sparse radar aperture in Sec. 4.1. Then we present CoIR's custom-designed CNN-based decoder architecture we call ComDecoder in Sec. 4.2. Next, Sec. 5 briefly describes competing untrained methods. Then, we present our simulated and experimental results in Sec. 6 and Sec. 7. We conclude with discussion and limitations in Sec. 8

4.1 Sparse Aperture Design

Our objective is to design a sparse MIMO virtual array that when used with ComDecoder improves radar imaging accuracy. The quality of the sparse virtual array is accessed by analyzing its PSF. The PSF's main lobe half-power-beamwidth (HPBW), maximum side lobe level (SLL), and grating lobes provide an indication of the achievable angular resolution and imaging ambiguities, respectively. We compute the sparse MIMO virutal array's PSF in the far-field by multiplying the PSFs of the physical transmitter and receiver MIMO arrays [55]. Details on computing the virtual array's PSF are in Supplemental Sec. 1.2. The design of the sparse virtual array must also satisfy hardware constraints [7] that limit the max aperture to $86\lambda/2$ with $\lambda = 3.9$ mm. Additionally, a majority of commercial single chip radar can only support four transmitters and four receivers [8]. In the design of our proposed sparse array we analyzed the virtual array's PSF and chose a design that (i) does not contain grating lobes within a $\pm 90^\circ$ field-of-view (FoV), (ii) minimizes the HPBW, (iii) minimizes the SLL, and (iv) meets our hardware constraints.

We compare our proposed sparse virtual array against three arrays constructed using conventional MIMO array design techniques [26]. We call an ideal full aperture Nyquist sampled array, *Full*, which could be constructed by neglecting the single chip hardware constraints (iv). We denote the next array as *Sub-apt* and is the largest Nyquist sampled MIMO array that can be synthesized from four transmitters and four receivers. The receivers and transmitters are positioned at $[0, 1, 2, 3]\lambda/2$ and $[0, 4, 8, 12]\lambda/2$, respectively. We call the last array *Sub-samp* since it is a sub-sampled array designed to maximize the aperture size using four transmitters and four receivers while staying within the maximum aperture limit $86\lambda/2$. The receivers and transmitters are positioned at $[0, 5, 10, 15]\lambda/2$ and $[0, 20, 40, 60]\lambda/2$, respectively. As expected, in Fig. 2 we observe that *Full* provides an ideal PSF response with a narrow HPBW, low SLL, and no grating lobes. Due to the small aperture in *Sub-apt* the HPBW is $5.2\times$ larger than in the *Full* array. In *Sub-samp* there are four grating lobes within the FoV due to the sub-Nyquist sampling of the array.

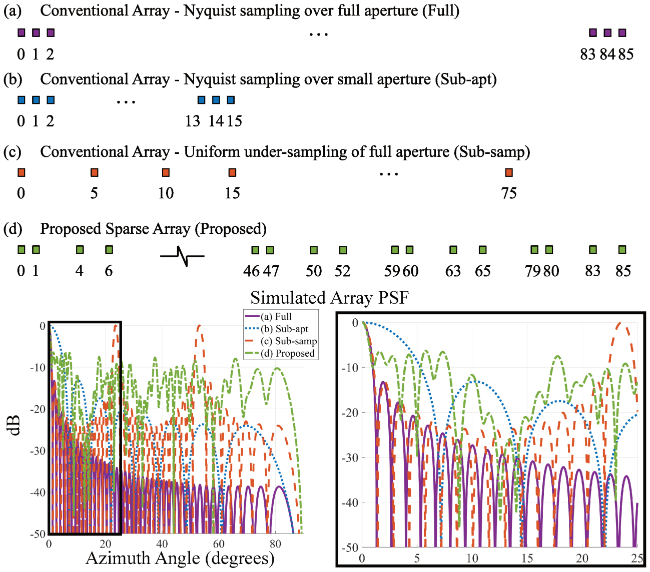

Fig. 2: **Simulated PSFs:** Comparison of different MIMO virtual array designs normalized to $\lambda/2$ increments. The proposed sparse virtual array (d) has $\approx 1^\circ$ HPBW, comparable to the full dense virtual array (a). Additionally, the proposed sparse array removes grating lobes and has reduced side lobe level (SLL) of ≈ -6 dB.

Our design methodology for CoIR's sparse array is inspired by [26], where we divide the design process into two steps. First, we choose a four element receive array to meet constraint (i). We use a four element minimum redundant array (MRA) with receivers located at $[0, 1, 4, 6](\lambda/2)$ [56]. This array design replicates the spatial frequency coverage of a seven element uniform array without introducing grating lobes within our FoV.

After establishing the receiver array design we focus on the second portion of the design process, the four element transmitter array. To minimize the HPBW (ii) we place two transmitters at $[0, 79]\lambda/2$, which maximizes the virtual array aperture while still meeting the hardware aperture constraints (iv). To determine the best position of the remaining two transmitters we implement a brute-force grid search within the set $E \in \{0, .., 79\}\lambda/2$ that minimizes the SLL of the virtual array's PSF (iii). For each of the $\approx 3 \times 10^3$ transmitter array designs evaluated in the gird search we multiply the transmitter array PSF by the fixed receiver array PSF to generate the corresponding virtual array PSF. We use the MATLAB function findpeaks [57] to determine the SLL of the virtual array's PSF. After the grid search we choose the transmitter array design that minimizes the SLL in the virtual array PSF. We found positioning transmitters at $[0, 46, 59, 79]\lambda/2$ yielded the best results with respect to constraints (i-iv). The sparse array design used in CoIR is illustrated in Fig. 2(d). The sparse array's simulated PSF has a maximum SLL of -6 dB, no grating lobes in the FoV, and $\approx 1.3^\circ$ HPBW.

4.2 Neural Network Architecture

We find a variation of the ConvDecoder [27] architecture to have significant performance improvement in both simulated and experimental data. We call this network variant ComDecoder, and like ConvDecoder both are convolutional neural networks mapping a latent variable to an image, i.e., $G : \mathbb{R}^{c_0 \times K_0 \times L_0} \rightarrow \mathbb{R}^{c \times K \times L}$, where c is the number of output channels of a K by L image with K and L being the image width and height, respectively. The ComDecoder architecture is illustrated in Fig. 1.b, where each layer except the last is composed of upsampling and then a residual block [58].

Inside the residual block a SiLU activation is used after the convolution layer instead of the traditional ReLU activation. We find using the SiLU function resulted in improved reconstructions. This could be attributed to the SiLU function being non-monotonic and helping the model be more expressive [59]. When fitting the network to a given under-sampled measurement only the parameters in the convolutional and BN layers are optimized. The final layer excludes upsampling and linearly combines the hidden channels to the output channels c using a 1x1 convolutional layer. In all cases since the image of the reflectivity distribution $\bar{\mathbf{x}}$ is complex, $c = 2$ for the real and imaginary part of the output image.

As found in [27], [48] the convolutional and upsampling layers apply strong priors on the output image. The convolutional operation captures local information among nearby pixels at varying resolutions per layer. The upsampling operation induces a fixed notion of resolution in each layer. As found in [58] the residual block structure helps propagate information captured at lower layers up to higher layers by structurally adding an identity mapping in each layer. The default architecture used in ComDecoder consists of 6 layers (including the last layer) with 128 channels per layer, and a fixed input $\mathbf{C} \in \mathbb{R}^{128\times8\times8}$ drawn from a uniform distribution. The network weights are updated by optimizing Eq. 7 and backpropagating the loss. Optimizing Eq. 7 for 2000 iterations on a single 256×256 radar reflectivity image takes under 50 seconds on a NVIDIA GeForce RTX 3080.

5 COMPETING UNTRAINED METHODS

We implement several untrained methods to perform sparse radar imaging and compare against ComDecoder. For all methods, we identify the architecture designs and regularization parameters for the network and gradient descent methods that maximize the simulated and experimental reconstruction quality. See Supplementary Sec. 2.3 for more information in hyper-parameter selection. We find the reconstruction quality for ComDecoder is maximized when a ℓ_1-norm hyperparameter of 1e-5 is used. To make the comparison fair with other network methods the same ℓ_1-norm hyperparameter is applied to all network methods. We run all iterative methods for 2000 iterations (i.e. until convergence is reached) and retain the images that correspond to the lowest loss during the optimization. Next, we describe each untrained method used for comparison.

Delay-and-Sum (DAS). The delay-and-sum (DAS) or conventional beamformer is commonly used in practice [9], [52] and implements the inverse of MIMO radar forward model Eq. 4,

$$\hat{\mathbf{x}} = \mathcal{F}^{-1}(\mathbf{z}), \tag{8}$$

where $\mathcal{F}^{-1}(\cdot)$ implements the 2D IFFT. Since the DAS reconstruction method is FFT-based it has the lowest computational complexity compared to the other untrained methods and can run in seconds on a modern CPU. In our experiments, we term this method Sparse DAS when the data cube $\mathbf{z}$ is sub-sampled. For the real-world experimental data we do not have access to the ground truth scene reflectivity, and instead approximate it using Eq. 8 with a fully sampled data cube $\mathbf{z}$; we term this method Full DAS in our results. While we consider the Full DAS as the "gold standard", our method is able to produce better reconstructions with significantly fewer aliasing artifacts.

Gradient Descent with Convex Relaxation. ComDecoder leverages the implicit bias of the structure of a convolutional decoder by optimizing the weights of the network rather than the scene reflectivity distribution. To test reconstruction performance without inductive bias we implement a gradient descent method with ℓ_1-norm regularization [13], [15] that optimizes scene reflectivity distribution directly,

$$\hat{\mathbf{x}} = \underset{\bar{\mathbf{x}}}{\operatorname{argmin}}\, ||\mathbf{z} - \mathbf{M} \odot \mathcal{F}(\bar{\mathbf{x}})||_2 + \lambda_L ||\bar{\mathbf{x}}||_1, \tag{9}$$

where λ_L is the ℓ_1-norm regularization hyperparameter. In Eq. 9 the optimization is biased towards sparse spatial domain solutions due to the ℓ_1-norm handcrafted prior. We find the reconstruction quality is maximized when λ_L = 1e-3 and 1e-2 for simulated and experimental data. In our experiments, we term this method GD+L1 Reg. The reflectivity distribution $\bar{\mathbf{x}}$ is initialized with samples from a uniform distribution.

Implicit Neural Representations. This untrained method uses an INR that typically consist of multi-layer perception layers that map an extremely low-level input to a low-level output [24]. For example, inputs could be coordinates in a scene (x, y) and the output could be a physical property of that scene like its mmWave reflectivity. These networks learn an implicit continuous representation of the input and output mapping. To implement the INR method with Eq. 7 the network G architecture is updated and the input $\mathbf{C} \in \mathbb{R}^{2\times K\times L}$ is changed to represent all the range and angle coordinates in a K by L image normalized between $[-1, 1]$. The first INR method we compare against uses ReLU activation functions with Fourier feature encoding [60]. For all experiments, we call this method INR-ReLU and we find setting $Q = 256$, $\kappa = 20$, number of layers to 7, and neurons per layer to 256 resulted in the best reconstruction quality for this method.

For additional comparisons, we implement another INR method using sinusoidal periodic activation functions [50]. This INR architecture does not require an explicit feature encoding operation but instead adjusts the initial frequency ω_0 of the network's first layer to scale its ability to learn high frequency information. For all experiments we term this method SIREN and find setting $\omega_0 = 30$, number of layers to 7, and neurons per layer to 256 resulted in the best reconstruction quality for this method.

Deep Image Prior. We implement another untrained neural network architecture based on Deep Image Prior (DIP) [25]. In [25] a U-net convolutional neural architecture is used containing an encoder, decoder, and skip connections. This approach takes a high-dimensional input with samples drawn from a uniform distribution and outputs a high-dimensional estimate, i.e. image of the scene reflectivity distribution. For all experiments, we call this method DIP. To implement the DIP method with Eq. 7 the network G architecture is updated and the input $\mathbf{C} \in \mathbb{R}^{32\times K\times L}$ is increased to match the desired output K by L image dimensions. We find using 128 channels per layer, 6 encoder and decoder layers, SiLU activation function, and nearest neighbor upsampling resulted in the best reconstruction performance for the DIP method.

DeepDecoder. The next untrained neural network method is based on DeepDecoder [48]. As the name suggests, this architecture uses a decoder architecture containing channel-wise convolution, bilinear upsampling, ReLU activation, and Batch normalization operations per layer. This method takes in a low-dimensional input and outputs a high-dimensional estimate. This architecture is unique because only the upsampling operation applies spatial pixel coupling. We call this method DeepDecoder. To implement

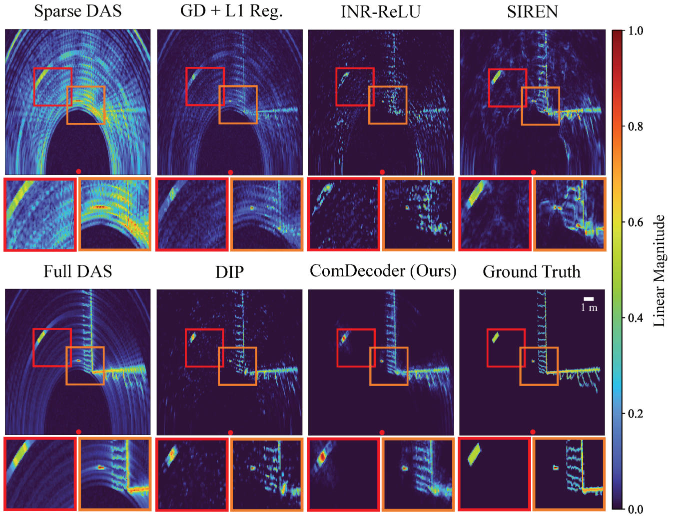

Fig. 3: **Qualitative Results for Simulated Scenes with measurement SNR at 19 dB:** Sparse radar imaging reconstruction results from a scene in our simulated data set. The bottom right image shows the ground-truth reflectivity image with two focused regions marked by red and orange squares. The red dot denotes the radar's location within the scene. The Sparse DAS reconstruction has extreme blurring from the large side lobes in the sparse array's PSF. The Full DAS has a sinc-like artifact due to the finite array aperture. The rest of the images contain the reconstructions from ComDecoder (our method) and competing methods. While all methods qualitatively improve the image quality compared to the Sparse DAS reconstruction, we observe ComDecoder has a significant improvement in fitting extending reflectors and removing sinc-like artifacts even with measurement noise.

the DeepDecoder method with Eq. 7 the network G architecture is updated and the input $\mathbf{C} \in \mathbb{R}^{512 \times K_0 \times L_0}$ matches the ComDecoder with an increased number of channels. We find using 6 layers with 512 channels per layer results in the best reconstruction performance for the DeepDecoder method.

ConvDecoder. The final untrained neural network method is based on ConvDecoder [27]. This method is a variant of DeepDecoder, with the major changes including 3 × 3 convolutions and Nearest-Neighbor upsampling per layer. Similar to DeepDecoder this method takes in a low-dimensional input and outputs a high-dimensional estimate. For all experiments, we call this method ConvDecoder. To implement our ConvDecoder method with Eq. 7 the network G architecture is updated and the input is generated following the same procedure as used for ComDecoder. We find using 6 layers with 128 channels per layer resulted in the best reconstruction performance for the ConvDecoder method.

6 SIMULATION RESULTS

This section evaluates our proposed method and all competing methods on the task of reconstructing an image of a reflectivity distribution from its simulated undersampled radar measurement. We use the peak signal-to-noise ratio (PSNR), structural similarity index (SSIM) [61], and mean absolute error (MAE) to measure the difference between the reconstructed and ground truth images. All iterative methods are optimized for 2000 iterations until convergence using the ADAM optimizer and are implemented in the PyTorch framework. The learning rates for each iterative method is chosen to provide the best performance for the given method while stabilizing the optimization. We found learning rates of 8.0e-3 worked well for CNN based methods, 1e-3 for the INR-ReLU method, and 1e-4 for SIREN. In all simulated experiments the transmit bandwidth is fixed to 1.283 GHz, carrier frequency set to 77 GHz, and use the sparse aperture design proposed in Sec. 4.1.

6.1 Generation of Simulation Data

We build a simulator modified from [2] that synthesizes mmWave radar data cubes from 2D reflectivity distributions. An example of our simulation pipeline is shown in Supplementary Sec. 2.1, Fig. 1. The first step in the simulator is the creation of realistic outdoor and indoor reflectivity distributions which requires knowledge about the placement and radar cross section of each reflector. We use LiDAR point clouds of outdoor and indoor scenes obtained from the ColoRadar data set [4] to synthesize realistic reflectivity distributions. For each point in the LiDAR point cloud, within the FoV of the radar, we use the LiDAR point's cartesian coordinates for the location of the reflector and scaling of the LiDAR point's intensity as a proxy for the reflector's radar cross section. All points are collapsed to the same height as the radar to create a 2D reflectivity distribution which is used as the ground truth reflectivity image. Next, we model the synthesized reflectivity distribution as a point reflector model in polar coordinates and use the MIMO radar forward model Eq. 4 to create a simulated radar data cube.

6.2 Sparse Radar Imaging and Sensitivity to Noise

In the first experiment, shown in Fig. 4, the performance of each sparse radar imaging method is quantified under varying noise levels added to the simulated radar data cube. In the experiment the measurement noise is swept between 35 dB and 11 dB and averaged over 25 samples from our simulated data set. In Fig. 4, we observe that our proposed method ComDecoder significantly outperforms competing methods with respect to PSNR across all noise levels. ComDecoder achieves superior performance in the SSIM metric up to 15 dB SNR where it achieves comparable performance to DeepDecoder. This aligns with computer vision literature where DeepDecoder has shown excellent performance at image denoising [48]. The MAE metric shows that ComDecoder achieves a similar level of performance as DIP. Across all metrics DIP performs second best, which is consistent with high performance on challenging inverse problems. The remaining CNN architectures ConvDecoder and DeepDecoder perform the next best followed by the MLP based architectures INR-ReLU and SIREN. All methods outperform Sparse DAS suggesting that regularization

This article has been accepted for publication in IEEE Transactions on Pattern Analysis and Machine Intelligence. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TPAMI.2023.3301553

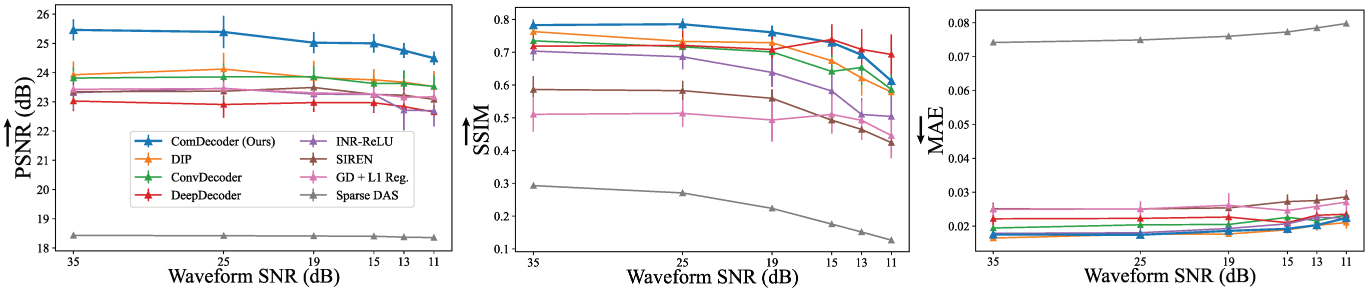

Fig. 4: **Simulated Sparse Radar Imaging Results with Additive Noise:** Quantitative reconstruction results for ComDecoder (our method) and competing methods averaged over 25 samples from our simulated data set under varying noise levels. The x-axis shows the level of i.i.d Gaussian noise added to the waveforms at each receiver. The y-axis quantifies the quality of each methods reconstruction to the ground truth image using PSNR (larger is better), SSIM (larger is better), and MAE (smaller is better) metrics. Qualitative results are presented in Supplementary Table 3 for and SNR of 19 dB.

helps reduce aliasing artifacts and capture the underlying salient features in the reflectivity distributions.

In Fig. 3 we visualize the reconstructions of two simulated scenes for all methods at an SNR of 19 dB. We expand on our choice for this SNR value in Supplementary Sec. 2.2. We observe that all methods qualitatively achieve improved reconstructions compared to Sparse DAS which uses no regularization. ComDecoder demonstrates a significant improvement in reconstructing extended reflectors and removing aliasing artifacts that appear as a radial blurring of energy in the other images. The observed increase in performance for ComDecoder over competing methods aligns with the quantitative metrics shown in Fig. 4. The MLP based methods, INR-ReLU and SIREN, struggle at differentiating and suppressing the structured aliasing artifacts from the underlying features in the scene, such as the wall corner in Fig. 3. We observe that the ℓ_1-norm regularized gradient descent method is able to localize the dominate reflectors in the scene but fails to remove aliasing artifacts to the level of neural network based methods.

We summarize each methods average reconstruction performance and speed for 25 simulated samples in Supplementary Table 3. We also performed a comparison with different CNN decoder methods – ComDecoder (ours), ConvDecoder, DeepDecoder, and the results are discussed in Supplementary Sec. 2.5.

7 EXPERIMENTAL RESULTS

This section discusses sparse radar imaging results using radar data from the ColoRadar data set [4]. The ideal reconstruction method should remove spurious artifacts caused by the high side lobes in the sparse array's PSF and recover a radar image of the scene's reflectivity distribution that is perceptually similar to the radar image formed using a dense full array, which we call Full DAS in the results. However, since the experimental radar data is noisy and is collected using a finite array the Full DAS reconstruction contains aliasing artifacts that appear as a radially smearing of energy in the radar image. In Supplementary Sec. 2.6 we investigate post-processing steps to improve the Full DAS reconstruction. We run all the methods for 2000 iterations until convergence and set the learning rate to 1e-3 for all iterative methods except SIREN which uses a learning rate of 1e-4 to stabilize the optimization. Each methods hyperparameters are set to the values used in Sec. 5

We use experimental mmWave raw ADC radar data from the ColoRadar data set [4]. The mmWave data was collected with Texas Instruments MMWCAS-RF-EVM Cascaded Imaging Radar Sensor [7] operating at 77 GHz (see Supplementary Sec. 3.4 for more details). This cascade radar sensor contains four radar chips, transmits a 1.282 GHz bandwidth FMCW pulse, and has a $86\lambda/2$ uniform linear array with an azimuth angular resolution of 1.33°. During acquisition the cascade radar sensor was mounted on a handheld rig and moved throughout several diverse indoor and outdoor environments. We specifically evaluate all methods using outdoor data collected in a courtyard scene and indoor data collected within office spaces.

7.1 Indoor and Outdoor Sparse Radar Imaging Results

Figure 5 shows sparse radar imaging reconstructions for our method, ComDecoder, and competing methods based on two samples from the ColoRadar data set [4]. Since we do not know the true reflectivity distribution for the experimental data we use Full DAS as a proxy ground truth. In Fig. 5 a LiDAR view is shown only for illustrative purposes and is not used in any method's reconstruction process.

Figure 5 (top) shows the reconstruction results for an outdoor scene with the dominate feature being a staircase. Observe that ComDecoder, SIREN, and DIP are the only methods to clearly recover the periodic reflections arising from the stair steps (red box) in the staircase. Additionally, ComDecoder significantly outperforms DIP and all other methods at removing aliasing artifacts arising from side lobes in the array's PSF (orange box). We believe that DIP struggles at achieving the same level of aliasing artifact suppression as ComDecoder because of its increased number of parameters. As shown in [25] without additional regularization such as early stopping DIP has the capacity to fit both an underlying signal and noise in a measurement.

Next to evaluate the generalizability of each methods sparse radar imaging performance we test each method on mmWave data captured in an indoor environment. The indoor reconstruction results are shown in Fig. 5 (bottom), with the dominate features in this scene being 14 poles

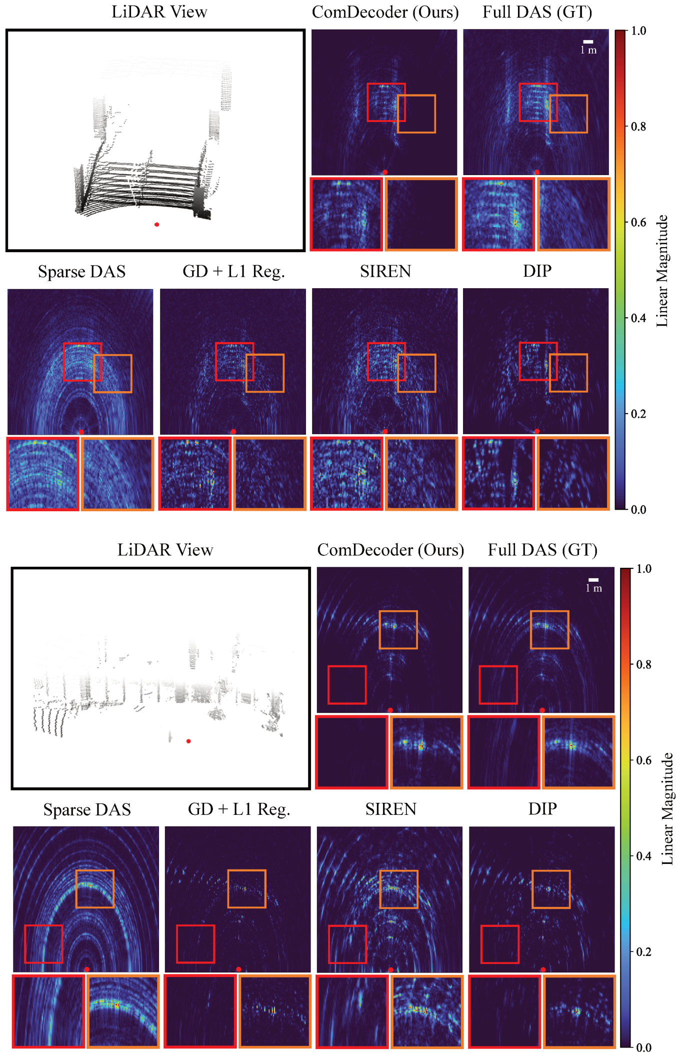

Fig. 5: **Experimental Results:** Sparse radar imaging results on experimental data from the ColoRadar dataset [4]. The 3D scene view is a LiDAR view and is only used for illustrative purposes to clearly show the contents of each scene. The red dot denotes the location of the radar module. We use the Conventional Full DAS reconstruction as the "gold standard" and compare it to the other sparse radar imaging methods. We observe in both environments that our method, ComDecoder, more accurately reconstructions the periodic features in each scene such as the stair steps and guard rails (top), and indoor wall (bottom), while removing aliasing artifacts as shown in the red and orange boxed regions respectively.

aligned in a row. All methods appear to have improved performance over the Sparse DAS reconstruction and are able to localize the reflections from the poles (orange box). In the indoor scene, Comdecoder and DIP more accurately reconstruct the salient features present in the Full DAS reconstruction with ComDecoder having superior aliasing artifact suppression (red box). We observe a significant degradation in SIREN's reconstruction quality which could be caused by an increased level of noise in the measurements from multipath effects. Since SIREN does not use convolutional layers there is not an explicit spatial local filtering bias being applied by the network that can help suppress noise.

7.2 Sparse Radar Imaging Robustness Analysis

In Fig. 6, we demonstrate reconstruction performance of ComDecoder and competing methods on several outdoor scenes with varying complexity. In each row on the left, a LiDAR view provides a perceptual view of the scene. The reconstructed image from the conventional full array is used as a benchmark to qualitatively compare against ComDecoder and the other competing methods. The orange boxed region highlights corresponding features that are captured in both the LiDAR and mmWave images. All methods show improved performance over the Sparse DAS reconstruction, but lack the performance of convolutional network methods.

Gradient descent with an ℓ_1-norm prior is able to estimate the dominate reflectors in the scene but is not able to completely suppress all aliasing artifacts. SIREN fits to both the strong reflectors in the scene and aliasing artifacts. This could be because contrary to natural images, radar images characteristically have small localized regions of high energy and a sparse distribution of weaker reflectors throughout the image. This arises from the mmWave's reflections being predominately specular instead of diffuse. Thus, even though SIREN and other MLP based architectures have shown to have an implicit biased towards natural images they struggle at differentiating reflectors from structured aliasing artifacts in the radar images.

CNN architectures have an inductive bias applied by the convolution operations, upsampling method, and skip connections. The convolution operations induce a notion of spatial locality between pixels while upsampling applies a notion of resolution per layer. The structure of the skip connections influences the flow of information between layers and can smooth out the optimization [58]. Together it is evident that these inductive biases improve the sparse radar imaging performance. Compared to DIP we observe that ComDecoder's reconstructions are qualitatively closer to the conventional full array's reconstructions with increased noise suppression. This could be attributed to ComDecoder having fewer parameters compared to DIP and uses a ResNet decoder architecture instead of a U-net. Further details on the impact that NN architecture and sparse aperture design have on ComDecoder's reconstruction quality are in Supplementary Sec. 3.1 and Sec. 3.2.

8 DISCUSSION & LIMITATIONS

While CoIR has demonstrated significant reconstruction improvements compared to other untrained techniques for sparse radar imaging, several limitations of the proposed method still provide avenues for future work. First, the radar forward model assumes static scenes and thus does not capture object motion. It is based on a single bounce point scattering model and assumes reflectors are omnidirectional scatterers. This can lead to forward model mismatch in practice and result in a decrease in performance.

This article has been accepted for publication in IEEE Transactions on Pattern Analysis and Machine Intelligence. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TPAMI.2023.3301553

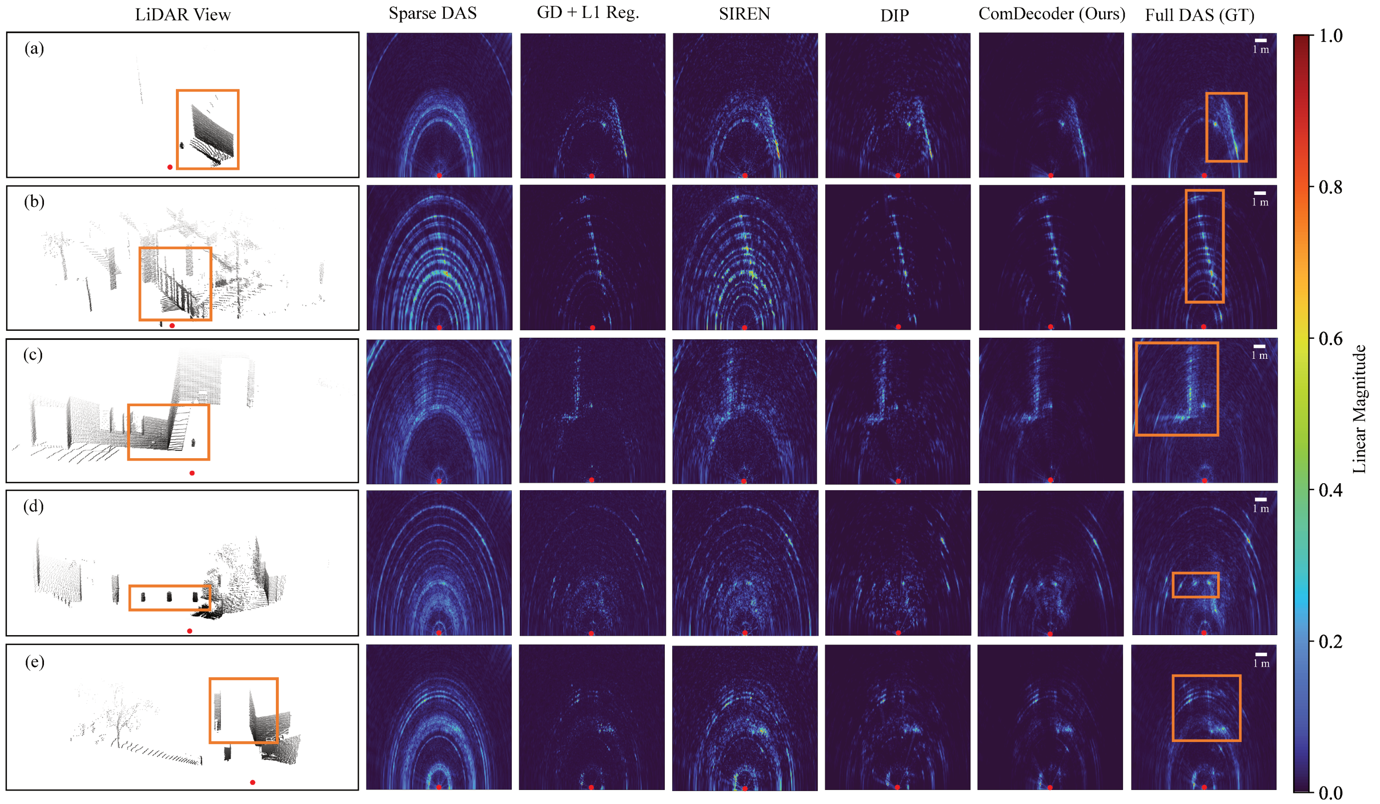

Fig. 6: **Experimental Results for Varying Outdoor Scenes:** Sparse radar imaging results on experimental data from several different scenes in the ColoRadar data set. The 3D scene view on the left is LiDAR view and is only used to provide a perceptual description of each scene. Post processing has been applied to the LiDAR point cloud to remove ground reflections to make the salient features clearer. The orange box encapsulates corresponding dominant objects visible in both the LiDAR and mmWave images. The conventional full array is used as the "gold standard" to compare against each methods sparse reconstruction. Throughout all scenes Comdecoder (our method) demonstrates an increased affinity towards natural reflectors and a reduction in aliasing artifacts, even when compared to the gold standard.

Additionally, the current reconstruction speed is on the order of tens of seconds due to the iterative nature of optimizing parameters of implicit neural networks. In future work, we plan on different strategies for initialization to speed up reconstruction time. Preliminary work on speeding up inference time can be found in Supplementary Sec. 3.3.

All the results shown in the paper are 2D depth range slices of the 3D world due to the use of linear array antenna. Reconstructing the entire 3D scene requires a 2D array of antennas, but using conventional methods will result in a quadratic increase in resource requirements such as bandwidth and processing. We foresee that our CoIR approach will be hugely beneficial in this scenario due to its use of a sparse array of antennas and this extension will be one of our future works. Additionally, CoIR opens up the opportunity to improve other array-based imaging modalities beyond mmWave radar, such as geographical remote sensing and ultrasound imaging.

9 CONCLUSION

We propose CoIR, an analysis by synthesis method that leverages the inductive bias of a CNN decoder, we call ComDecoder, to enhance the reconstruction quality in sparse radar imaging applications. Sparse radar imaging allows for a reduction in hardware resources which in turn decreases the cost and read bandwidth for a radar system. We develop a sparse MIMO aperture design for single chip radars that requires 5.5× fewer elements compared to conventional MIMO radar designs for apertures of the same size. We demonstrate ComDecoder's superior performance compared to classic and state-of-art untrained methods on both simulated and experimental mmWave radar data. Additionally, we show that CoIR can be extended to a wide variety of outdoor and indoor scenes without any scene dependent adjustments.

ACKNOWLEDGMENTS

This work was supported by the NSF Grant CNS-1956297, NSF CAREER Award no. IIS-1652633, and NSF Thermal Computational Award no. IIS-2107313.

REFERENCES

[1] M. Aladsani, A. Alkhateeb, and G. C. Trichopoulos, "Leveraging mmwave imaging and communications for simultaneous localization and mapping," in *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2019, pp. 4539–4543.

This article has been accepted for publication in IEEE Transactions on Pattern Analysis and Machine Intelligence. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TPAMI.2023.3301553

[2] J. Guan, S. Madani, S. Jog, S. Gupta, and H. Hassanieh, "Through fog high-resolution imaging using millimeter wave radar," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020.

[3] D. Sheen, D. McMakin, and T. Hall, "Three-dimensional millimeter-wave imaging for concealed weapon detection," *IEEE Transactions on Microwave Theory and Techniques*, vol. 49, no. 9, pp. 1581–1592, 2001.

[4] A. Kramer, K. Harlow, C. Williams, and C. Heckman, "Coloradar: The direct 3d millimeter wave radar dataset," *The International Journal of Robotics Research*, vol. 41, no. 4, pp. 351–360, 2022.

[5] A. Prabhakara, D. Zhang, C. Li, S. Munir, A. Sankanaryanan, A. Rowe, and S. Kumar, "A hybrid mmwave and camera system for long-range depth imaging," *arXiv preprint arXiv:2106.07856*, 2021.

[6] V. M. Patel, J. N. Mait, D. W. Prather, and A. S. Hedden, "Computational millimeter wave imaging: problems, progress, and prospects," *IEEE Signal Processing Magazine*, vol. 33, no. 5, pp. 109–118, 2016.

[7] T. Instruments. (2021) Ti mmwcas-rf-evm. [Online]. Available: https://www.ti.com/tool/MMWCAS-RF-EVM

[8] ——. (2021) Ti awr2944 evm. [Online]. Available: https://www.ti.com/tool/AWR2944EVM

[9] R. Feger, C. Wagner, S. Schuster, S. Scheiblhofer, H. Jager, and A. Stelzer, "A 77-ghz fmcw mimo radar based on an sige single-chip transceiver," *IEEE Transactions on Microwave theory and Techniques*, vol. 57, no. 5, pp. 1020–1035, 2009.

[10] M. E. Yanik and M. Torlak, "Near-field mimo-sar millimeter-wave imaging with sparsely sampled aperture data," *Ieee Access*, vol. 7, pp. 31 801–31 819, 2019.

[11] A. B. Baral and M. Torlak, "Joint doppler frequency and direction of arrival estimation for tdm mimo automotive radars," *IEEE Journal of Selected Topics in Signal Processing*, vol. 15, no. 4, pp. 980–995, 2021.

[12] R. Baraniuk and P. Steeghs, "Compressive radar imaging," in *2007 IEEE radar conference*. IEEE, 2007, pp. 128–133.

[13] R. Heckel, V. I. Morgenshtern, and M. Soltanolkotabi, "Super-resolution radar," *Information and Inference: A Journal of the IMA*, vol. 5, no. 1, pp. 22–75, 2016.

[14] D. Liu, U. S. Kamilov, and P. T. Boufounos, "Sparsity-driven distributed array imaging," in *2015 IEEE 6th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*. IEEE, 2015, pp. 441–444.

[15] L. C. Potter, E. Ertin, J. T. Parker, and M. Cetin, "Sparsity and compressed sensing in radar imaging," *Proceedings of the IEEE*, vol. 98, no. 6, pp. 1006–1020, 2010.

[16] 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.

[17] 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.

[18] F. Zhang, C. Wu, B. Wang, and K. J. R. Liu, "mmeye: Super-resolution millimeter wave imaging," *IEEE Internet of Things Journal*, vol. 8, no. 8, pp. 6995–7008, 2021.

[19] A. Prabhakara, T. Jin, A. Das, G. Bhatt, L. Kumari, E. Soltanaghaei, J. Bilmes, S. Kumar, and A. Rowe, "High resolution point clouds from mmwave radar," *arXiv preprint arXiv:2206.09273*, 2022.

[20] J. W. Smith, Y. Alimam, G. Vedula, and M. Torlak, "A vision transformer approach for efficient near-field sar super-resolution under array perturbation," in *2022 IEEE Texas Symposium on Wireless and Microwave Circuits and Systems (WMCS)*. IEEE, 2022, pp. 1–6.

[21] X. Gao, G. Xing, S. Roy, and H. Liu, "Ramp-cnn: A novel neural network for enhanced automotive radar object recognition," *IEEE Sensors Journal*, vol. 21, no. 4, pp. 5119–5132, 2020.

[22] D. Van Veen, A. Jalal, M. Soltanolkotabi, E. Price, S. Vishwanath, and A. G. Dimakis, "Compressed sensing with deep image prior and learned regularization," *arXiv preprint arXiv:1806.06438*, 2018.

[23] G. Jagatap and C. Hegde, "Phase retrieval using untrained neural network priors," in *NeurIPS 2019 Workshop on Solving Inverse Problems with Deep Networks*, 2019.

[24] B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi, and R. Ng, "Nerf: Representing scenes as neural radiance fields for view synthesis," *Communications of the ACM*, vol. 65, no. 1, pp. 99–106, 2021.

[25] U. Dmitry, A. Vedaldi, and L. Victor, "Deep image prior," *International Journal of Computer Vision*, vol. 128, no. 7, pp. 1867–1888, 2020.

[26] R. Z. Syeda, M. C. van Beurden, and A. B. Smolders, "Sparse virtual array synthesis for mimo radar imaging systems," *IET Microwaves, Antennas & Propagation*, vol. 15, no. 11, pp. 1458–1472, 2021.

[27] M. Z. Darestani and R. Heckel, "Accelerated mri with un-trained neural networks," *IEEE Transactions on Computational Imaging*, vol. 7, pp. 724–733, 2021.

[28] D. Huang, R. Nandakumar, and S. Gollakota, "Feasibility and limits of wi-fi imaging," in *Proceedings of the 12th ACM Conference on Embedded Network Sensor Systems*, ser. SenSys '14. New York, NY, USA: Association for Computing Machinery, 2014, p. 266–279. [Online]. Available: https://doi.org/10.1145/2668332.2668344

[29] M. Zhao, T. Li, M. Abu Alsheikh, Y. Tian, H. Zhao, A. Torralba, and D. Katabi, "Through-wall human pose estimation using radio signals," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 7356–7365.

[30] T. Wei and X. Zhang, "mtrack: High-precision passive tracking using millimeter wave radios," in *Proceedings of the 21st Annual International Conference on Mobile Computing and Networking*, 2015, pp. 117–129.

[31] C. Wu, F. Zhang, B. Wang, and K. R. Liu, "mmtrack: Passive multi-person localization using commodity millimeter wave radio," in *IEEE INFOCOM 2020-IEEE Conference on Computer Communications*. IEEE, 2020, pp. 2400–2409.

[32] P. Zhao, C. X. Lu, B. Wang, N. Trigoni, and A. Markham, "3d motion capture of an unmodified drone with single-chip millimeter wave radar," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE Press, 2021, p. 5186–5192. [Online]. Available: https://doi.org/10.1109/ICRA48506.2021.9561738

[33] J. Lien, N. Gillian, M. E. Karagozler, P. Amihood, C. Schwesig, E. Olson, H. Raja, and I. Poupyrev, "Soli: Ubiquitous gesture sensing with millimeter wave radar," *ACM Transactions on Graphics (TOG)*, vol. 35, no. 4, pp. 1–19, 2016.

[34] J. W. Smith, O. Furxhi, and M. Torlak, "An fcnn-based super-resolution mmwave radar framework for contactless musical instrument interface," *IEEE Transactions on Multimedia*, vol. 24, pp. 2315–2328, 2021.

[35] F. Adib, C.-Y. Hsu, H. Mao, D. Katabi, and F. Durand, "Capturing the human figure through a wall," *ACM Transactions on Graphics (TOG)*, vol. 34, no. 6, pp. 1–13, 2015.

[36] C. Wu, F. Zhang, B. Wang, and K. J. R. Liu, "msense: Towards mobile material sensing with a single millimeter-wave radio," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 4, pp. 106:1–106:20, 2020.

[37] J. Guan, A. Paidimarri, A. Valdes-Garcia, and B. Sadhu, "3d imaging using mmwave 5g signals," in *2020 IEEE Radio Frequency Integrated Circuits Symposium (RFIC)*. IEEE, 2020, pp. 147–150.

[38] H. Regmi, M. S. Saadat, S. Sur, and S. Nelakuditi, "Squigglemilli: Approximating sar imaging on mobile millimeter-wave devices," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 5, no. 3, pp. 1–26, 2021.

[39] M. S. Saadat, S. Sur, S. Nelakuditi, and P. Ramanathan, "Millicam: Hand-held millimeter-wave imaging," in *2020 29th International Conference on Computer Communications and Networks (ICCCN)*, 2020, pp. 1–9.

[40] A. Vilesov, P. Chari, A. Armouti, A. B. Harish, K. Kulkarni, A. Deoghare, L. Jalilian, and A. Kadambi, "Blending camera and 77 ghz radar sensing for equitable, robust plethysmography," *ACM Trans. Graph.(SIGGRAPH)*, vol. 41, no. 4, pp. 1–14, 2022.

[41] H. Lebret and S. Boyd, "Antenna array pattern synthesis via convex optimization," *IEEE transactions on signal processing*, vol. 45, no. 3, pp. 526–532, 1997.

[42] F. Gumbmann, P. Tran, and L.-P. Schmidt, "Sparse linear array design for a short range imaging radar," in *2009 European Radar Conference (EuRAD)*. IEEE, 2009, pp. 176–179.

[43] G. Prisco and M. D'Urso, "Maximally sparse arrays via sequential convex optimizations," *IEEE Antennas and Wireless Propagation Letters*, vol. 11, pp. 192–195, 2012.

[44] B. Mamandipoor, M. Fallahpour, G. Malysa, K. Noujeim, U. Madhow, and A. Arbabian, "Spatial-domain technique to overcome grating lobes in sparse monostatic mm-wave imaging systems," in *2016 IEEE MTT-S International Microwave Symposium (IMS)*. IEEE, 2016, pp. 1–4.

This article has been accepted for publication in IEEE Transactions on Pattern Analysis and Machine Intelligence. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TPAMI.2023.3301553

[45] S. Sun and A. P. Petropulu, "A sparse linear array approach in automotive radars using matrix completion," in *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2020, pp. 8614–8618.

[46] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE transactions on antennas and propagation*, vol. 34, no. 3, pp. 276–280, 1986.

[47] R. Roy and T. Kailath, "Esprit-estimation of signal parameters via rotational invariance techniques," *IEEE Transactions on acoustics, speech, and signal processing*, vol. 37, no. 7, pp. 984–995, 1989.

[48] R. Heckel and P. Hand, "Deep decoder: Concise image representations from untrained non-convolutional networks," *arXiv preprint arXiv:1810.03982*, 2018.

[49] R. Cao, F. L. Liu, L.-H. Yeh, and L. Waller, "Dynamic structured illumination microscopy with a neural space-time model," in *2022 IEEE International Conference on Computational Photography (ICCP)*. IEEE, 2022, pp. 1–12.

[50] V. Sitzmann, J. N. Martel, A. W. Bergman, D. B. Lindell, and G. Wetzstein, "Implicit neural representations with periodic activation functions," in *Proc. NeurIPS*, 2020.

[51] A. Reed, T. Blanford, D. C. Brown, and S. Jayasuriya, "Sinr: Deconvolving circular sas images using implicit neural representations," *IEEE Journal of Selected Topics in Signal Processing*, 2022.

[52] S. M. Patole, M. Torlak, D. Wang, and M. Ali, "Automotive radars: A review of signal processing techniques," *IEEE Signal Processing Magazine*, vol. 34, no. 2, pp. 22–35, 2017.

[53] M. I. Skolnik, *Introduction to Radar Systems*, 3rd ed., ser. McGraw-Hill International Editions: Electrical Engineering Series. Maidenhead, England: McGraw Hill Higher Education, Dec. 2000.

[54] M. Richards, *Fundamentals of radar signal processing, third edition*, 3rd ed. Columbus, OH: McGraw-Hill Education, Apr. 2022.

[55] W. L. Stutzman and G. A. Thiele, *Antenna theory and design*. John Wiley & Sons, 2012.

[56] A. Moffet, "Minimum-redundancy linear arrays," *IEEE Transactions on antennas and propagation*, vol. 16, no. 2, pp. 172–175, 1968.

[57] MATLAB, *version 9.13.0.2049777 (R2022b)*. Natick, Massachusetts: The MathWorks Inc., 2023.

[58] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[59] P. Ramachandran, B. Zoph, and Q. V. Le, "Searching for activation functions," *ArXiv*, vol. abs/1710.05941, 2018.

[60] M. Tancik, P. P. Srinivasan, B. Mildenhall, S. Fridovich-Keil, N. Raghavan, U. Singhal, R. Ramamoorthi, J. T. Barron, and R. Ng, "Fourier features let networks learn high frequency functions in low dimensional domains," *NeurIPS*, 2020.

[61] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.

Sean M. Farrell received a B.S. in engineering science from Trinity University, San Antonio, Texas, in 2019 and receiving the M.S. from the Department of Electrical and Computer Engineering, Rice University, Houston, Texas, in 2023. He is currently a Ph.D. student at Rice University, Houston, Texas. His research interests include the areas of computational imaging, signal processing, wireless communication, and radar.

Vivek Boominathan received the BTech degree in electrical engineering from the Indian Institute of Technology Hyderabad, Hyderabad, India, in 2012, and the MS and Ph.D. degrees from the Department of Electrical and Computer Engineering, Rice University, Houston, Texas, in 2016 and 2019, respectively. He is currently a postdoctoral associate with Rice University, Houston, Texas. His research interests include the areas of computer vision, signal processing, wave optics, and computational imaging.

Nathaniel Raymondi received a B.S. in electrical engineering from the University of Akron, Akron, Ohio, in 2017, and the M.S. and Ph.D. degrees from the Department of Electrical and Computer Engineering, Rice University, Houston, Texas, in 2019 and 2023, respectively. He is currently a senior professional staff at the Johns Hopkins University Applied Physics Laboratory in Laurel, Maryland. His research interests include the areas of wireless communications and radar, signal processing, and detection theory.

Ashutosh Sabharwal received the B.Tech. degree from IIT Delhi, New Delhi, India, in 1993, and the M.S. and Ph.D. degrees from The Ohio State University, Columbus, OH, USA, in 1995 and 1999, respectively. He is currently the Department Chair and an Ernest D. Butcher Professor with the Department of Electrical and Computer Engineering, Rice University, Houston, TX, USA. His research interests are in wireless theory, design, and large-scale deployed testbeds. He was a co-inventor of in-band full-duplex wireless. He is the founder of the WARP Project (warp.rice.edu), an open-source project, used with more than 125 research groups worldwide leading to more than 500 research articles. He is currently leading Rice RENEW (renew-wireless.org), to develop an open-source software-defined wireless network platform. He received the 2017 IEEE Jack Neubauer Memorial Award, the 2018 IEEE Advances in Communications Award, the 2019 ACM MobiCom Community Contribution Award, the 2019 and 2021 ACM Test-of-Time Awards, and 2023 ICC Best Paper Award. He is a Fellow of National Academy of Inventors, IEEE and ACM.

Ashok Veeraraghavan received his bachelor's degree in electrical engineering from the Indian Institute of Technology, Madras, Chennai, India, in 2002, and the MS and PhD degrees from the Department of Electrical and Computer Engineering, University of Maryland, College Park, MD,USA, in 2004 and 2008, respectively. He is currently a professor in the Electrical and Computer Engineering, with Rice University, Houston, TX, USA. Before joining Rice University, he spent three years as a research scientist at Mitsubishi Electric Research Labs, Cambridge, MA, USA. His research interests are broadly in the areas of computational imaging, computer vision, machine learning, and robotics. His thesis received the Doctoral Dissertation Award from the Department of Electrical and Computer Engineering at the University of Maryland. He is the recipient of the National Science Foundation CAREER Award in 2017. At Rice University, he directs the Computational Imaging and Vision Lab.