MITSUBISHI ELECTRIC RESEARCH LABORATORIES
https://www.merl.com

Deep-Unfolded Autofocus Imaging for Distributed MIMO Radar

Terada, Tsubasa; Mansour, Hassan; Boufounos, Petros T.; Takahashi, Ryuhei

TR2026-128 September 17, 2026

Abstract

We propose a deep-unfolded, multiple-input multiple-output (MIMO) extended autofocus imaging method for distributed radar that efficiently enhances imaging performance. The proposed method unfolds an iterative imaging algorithm using deep denoisers with fewer unrolled iterations, enabling faster and more stable inference. Moreover, to fully exploit the information available across transmit and receive antennas, extending distributed radar to a MIMO configuration improves the imaging signal-to-noise ratio. At the same time, the direct-path signal between the transmit and receive antennas is explicitly incorporated into the data-fidelity term of the shift-kernel estimation problem, thereby mitigating autofocus degradation for long-range targets. Numerical simulations demonstrate consistently high peak signal-to-noise ratio and structural similarity index measure, as well as clear reconstruction of target shapes and salient features composed of multiple scattering points.

IEEE International Conference on Image Processing (ICIP) 2026

Mitsubishi Electric Research Laboratories, Inc.
201 Broadway, Cambridge, Massachusetts 02139

DEEP-UNFOLDED AUTOFOCUS IMAGING FOR DISTRIBUTED MIMO RADAR

Tsubasa Terada[⋆†], *Hassan Mansour*[⋆], *Petros T. Boufounos*[⋆], *Ryuhei Takahashi*[†]

[⋆]Mitsubishi Electric Research Laboratories,
201 Broadway, Cambridge, MA 02139, USA
{terada, mansour, petrosb}@merl.com

[†]Information Technology R&D Center, Mitsubishi Electric Corporation
Ofuna 5-1-1, Kanagawa, Japan
{Terada.Tsubasa@cb, Takahashi.Ryuhei@ab}.MitsubishiElectric.co.jp

ABSTRACT

We propose a deep-unfolded, multiple-input multiple-output (MIMO) extended autofocus imaging method for distributed radar that efficiently enhances imaging performance. The proposed method unfolds an iterative imaging algorithm using deep denoisers with fewer unrolled iterations, enabling faster and more stable inference. Moreover, to fully exploit the information available across transmit and receive antennas, extending distributed radar to a MIMO configuration improves the imaging signal-to-noise ratio. At the same time, the direct-path signal between the transmit and receive antennas is explicitly incorporated into the data-fidelity term of the shift-kernel estimation problem, thereby mitigating autofocus degradation for long-range targets. Numerical simulations demonstrate consistently high peak signal-to-noise ratio and structural similarity index measure, as well as clear reconstruction of target shapes and salient features composed of multiple scattering points.

***Index Terms*—** Radar autofocus, blind deconvoluation, sparse image reconstruction, deep denoisers, deep unfolding

1. INTRODUCTION

Imaging radar enables the characterization of target shapes and features beyond point-target representations, and its imaging resolution is closely related to the aperture length. Synthetic aperture radar (SAR) and inverse synthetic aperture radar (ISAR) achieve high-resolution imaging by enlarging the synthetic aperture through the spatial motion of the radar or the target [1, 2]. These methods are effective when the relative motion between the radar and the target is known or predictable, while requiring an observation time proportional to the synthetic aperture length. In contrast, distributed radar systems obtain a large aperture length by spatially distributing multiple radars, enabling instantaneous observation while being well suited for targets with unknown motion [3–5]. Furthermore, incorporating cross-path (bistatic) signals into the coherent integration improves the overall signal-to-noise ratio (SNR) compared to processing only monostatic returns.

High-resolution distributed radar imaging requires accurate knowledge of the antenna phase center positions of each radar. However, when precise optical surveying becomes problematic, such as when radars are mounted on mobile platforms, position estimation relies on the global navigation satellite system (GNSS) and the inertial navigation system (INS). At best, the accuracy of these systems is on the order of centimeters [6], which exceeds the wavelength of the radar pulse and thus limits high-resolution imaging.

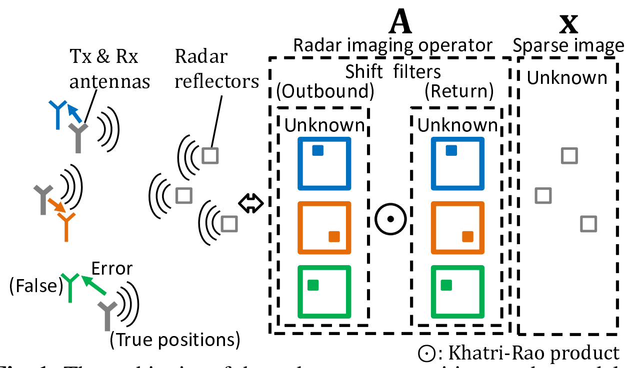

Fig. 1. The ambiguity of the radar antenna position can be modeled via 2D shift kernels $\mathbf{h}$, resulting in measurements of the form $\mathbf{y} = \mathbf{A}(\mathbf{h})\mathbf{x}$ for all transmit-receive antenna combinations.

As a result, multiple techniques have been proposed to compensate for this error by modeling the effect of the perturbation as unknown gain and phase errors in the acquired measurements, or as unknown time-shifts when considering far-field imaging assumptions [7–15]. However, in many real-world applications, the idealized conditions assumed in those techniques do not hold, making the models imprecise [16]. More recently, deep unfolding [17, 18] has emerged in SAR autofocus and inverse imaging applications as an effective blind deconvolution technique that achieves computation and iteration efficiency [19–21].

In this work, we fully exploit the advantages of distributed radar systems by introducing a multiple-input multiple-output (MIMO) configuration and using direct-path signals between transmitting antennas, receiving antennas, and the scene targets. We also reformulate the estimation algorithm in the imaging stage via deep unfolding to enhance performance by learning complex target structures using deep denoisers. Moreover, by learning algorithm parameters such as step sizes, deep unfolding reduces the number of required iterations and accelerates inference. In addition, a deep denoiser, proven effective in radar autofocus imaging [22], is integrated into each unfolding stage in a step-adaptive manner, increasing its representation capacity and further improving performance.

To utilize direct-path signals, we explicitly incorporate the signals between radars into the data fidelity term of the shift-kernel estimation problem. As the distance to the imaging region increases, the phase fronts of waves reflected from the region exhibit higher spatial correlation, which degrades the noise robustness of antenna position error estimation. However, the direct-path signals between distributed radars arranged close to the cross-range direction enhance the orthogonality of the phase fronts, enabling robust antenna position error estimation performance even for distant imaging regions.

The remainder of this paper is organized as follows. Section 2 formulates distributed radar autofocus imaging extended to a MIMO configuration. Section 3 presents an algorithm that reconstructs radar image estimation via deep unfolding and shift-kernel estimation exploiting MIMO and direct-path signals using a block coordinate descent method. Section 4 evaluates the target shape estimation performance through numerical simulations. Finally, Section 5 concludes the paper.

2. PROBLEM FORMULATION FOR DISTRIBUTED MIMO RADAR

2.1. Signal Model

We consider measurements acquired at discrete angular frequencies $\{\omega_f\}_{f=1}^{F}$, where $\omega_f \in \mathbb{R}$ denotes the fth angular frequency. For each ω_f, the corresponding measurement vector is $\mathbf{y}(\omega_f) \in \mathbb{C}^{M^2}$, where M is the number of transmitting and receiving array antennas sharing a common phase center. We stack these measurements as $\mathbf{y} = \left(\mathbf{y}^T(\omega_1), \ldots, \mathbf{y}^T(\omega_F)\right)^T$, so that $\mathbf{y} \in \mathbb{C}^{M^2F}$. Note that, while we assume for simplicity that all antennas perform both transmission and reception in this study, the proposed framework can be readily extended to cases with different numbers of transmitting and receiving antennas or different phase centers within each radar.

The unknown targets located within a spatial region of interest that is discretized on a grid $\Omega \subset \mathbb{R}^2$ is represented by $\mathbf{x} \in \mathbb{C}^N$, where $|\Omega| = N$ and $N = N_\mathrm{x} \times N_\mathrm{y}$ is the number of spatial grid points. The imaging region is treated as two-dimensional, assuming that elevation beams are formed for each subarray and imaging is performed for each elevation angle; however, extension to three dimensions is straightforward in the same manner. Let $\mathbf{l}_n \in \Omega$, $n \in \{1, \ldots, N\}$ denote the nth grid location. The vector $\mathbf{x}$ is $\mathbf{x} = \left(X(\mathbf{l}_1), \ldots, X(\mathbf{l}_N)\right)^T$, where $X(\mathbf{l}_n) \in \mathbb{C}$ denotes the reflectivity value. In general, target reflections exhibit angle-dependent characteristics due to speckle resulting from the coherent summation of individual scattering points. In this paper, since a distributed radar system with very high spatial resolution is employed, we assume that individual scattering points can be resolved and separately observed.

The global radar imaging operator $\tilde{\mathbf{A}} \in \mathbb{C}^{M^2F\times N}$ is formed from per-frequency matrices $\tilde{\mathbf{A}}(\omega_f) \in \mathbb{C}^{M^2\times N}$ via

$$\tilde{\mathbf{A}} = \left(\tilde{\mathbf{A}}^T(\omega_1), \ldots, \tilde{\mathbf{A}}^T(\omega_F)\right)^T, \tag{1}$$

yielding the receiving signal model

$$\mathbf{y} = \tilde{\mathbf{A}}\,\mathbf{x} + \mathbf{n}, \tag{2}$$

where $\mathbf{n} \in \mathbb{C}^{M^2F}$ is the noise signal vector.

For each ω_f, the matrix $\tilde{\mathbf{A}}(\omega_f)$ can be modeled as a Khatri–Rao structure,

$$\tilde{\mathbf{A}}(\omega_f) = P(\omega_f)\left(\tilde{\mathbf{U}}(\omega_f) \odot \tilde{\mathbf{U}}(\omega_f)\right), \tag{3}$$

where $\odot$ denotes the Khatri–Rao product that corresponds to the columnwise Kronecker product and $P(\omega_f) \in \mathbb{C}$ is a frequency spectrum that corresponds to a time-domain pulse $p(t)$ emitted to space from transmitters, and $\tilde{\mathbf{U}}$ denotes the field propagation matrix.

Let $\Gamma \subset \mathbb{R}^2, |\Gamma| = M$ be the set of all the spatial locations of the antennas. The field matrix $\tilde{\mathbf{U}}(\omega_f) \in \mathbb{C}^{M\times N}$ is given by $\tilde{\mathbf{U}}(\omega_f) = \left(\mathbf{u}(\omega_f, \tilde{\mathbf{r}}_1), \ldots, \mathbf{u}(\omega_f, \tilde{\mathbf{r}}_M)\right)^T$, where $\tilde{\mathbf{r}}_m \in \Gamma$ denotes the mth sensor location. Each field vector is $\mathbf{u}(\omega_f, \tilde{\mathbf{r}}_m) = \left(U(\omega_f, \tilde{\mathbf{r}}_m, \mathbf{l}_1), \ldots, U(\omega_f, \tilde{\mathbf{r}}_m, \mathbf{l}_N)\right)^T$.

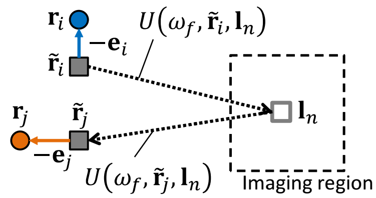

Fig. 2. The signal propagates from transmit antenna i to receive antenna j ($i, j \in \{1, \ldots, M\}$) via reflection by object n ($n \in \{1, \ldots, N\}$) in the imaging region. The antenna position ambiguity can be compensated by shifting the position by vector $\mathbf{e}_{(\cdot)}$.

The scalar field value is modeled as

$$U(\omega_f, \tilde{\mathbf{r}}_m, \mathbf{l}_n) = a(\tilde{\mathbf{r}}_m, \mathbf{l}_n)\,\exp\left(-j\,\frac{\omega_f}{c}\|\tilde{\mathbf{r}}_m - \mathbf{l}_n\|_2\right), \tag{4}$$

where $a(\tilde{\mathbf{r}}_m, \mathbf{l}_n) \in \mathbb{R}$ is a propagation loss, and c is the speed of light.

In real-world environments, antenna positioning errors arise due to manufacturing tolerances, installation inaccuracies, or GNSS receiver positioning errors. These errors not only make it difficult to obtain a clear image but also degrade detection performance, which can be critical for reliability-sensitive applications such as autonomous driving. Let $\mathbf{e}_m$ denote the error vector that corrects the assumed position $\mathbf{r}_m \in \Gamma$ to the true position $\tilde{\mathbf{r}}_m$, such that $\tilde{\mathbf{r}}_m = \mathbf{r}_m + \mathbf{e}_m$. Fig. 2 illustrates a case in which both the transmitting and receiving antennas have position errors.

To account for this spatial deviation, we exploit convolution in the spatial domain [16]. We model the impact of $\mathbf{e}_m$ by introducing a spatial shift kernel $\mathbf{h}_m \in \mathbb{R}_+^{N_\mathrm{e}}$, where $N_\mathrm{e} = N_{\mathrm{e_x}} \times N_{\mathrm{e_y}}$ denotes the dimension of the kernel. By treating the propagation mismatch as a translation of the steering vectors, $\mathbf{h}_m$ is represented as a Dirac delta function $\delta(\mathbf{l} - \mathbf{e}_m)$ of a target position $\mathbf{l}$. This kernel effectively shifts the columns of the field matrix $\mathbf{U}$ by $+\mathbf{e}_m$, which enables us to cast the estimation as a sparse recovery problem. Collecting the the shift kernels into $\mathbf{h}_{1:M} = (\mathbf{h}_1, \ldots, \mathbf{h}_M)$, we define the radar imaging operator $\mathbf{A}(\omega_f, \mathbf{h}_{1:M})$ as a function of the spatial offsets, obtained by replacing each field vector $\mathbf{u}(\omega_f, \mathbf{r}_m)$ with its discrete convolution with $\mathbf{h}_m$, and forming its Khatri–Rao product:

$$\begin{aligned}\tilde{\mathbf{A}}(\omega_f) &= \mathbf{A}(\omega_f, \mathbf{h}_{1:M}) \\ &= P(\omega_f)\left(\mathbf{U}(\omega_f, \mathbf{h}_{1:M}) \odot \mathbf{U}(\omega_f, \mathbf{h}_{1:M})\right),\end{aligned} \tag{5}$$

where,

$$\mathbf{U}(\omega_f, \mathbf{h}_{1:M}) = \begin{pmatrix} (\mathbf{u}(\omega_f, \mathbf{r}_1)) * \mathbf{h}_1)^T \\ \vdots \\ (\mathbf{u}(\omega_f, \mathbf{r}_M)) * \mathbf{h}_M)^T \end{pmatrix}. \tag{6}$$

Here, $*$ denotes the two-dimensional convolution operator. Then, the stacked operator is

$$\mathbf{A}(\mathbf{h}_{1:M}) = \left(\mathbf{A}^T(\omega_1, \mathbf{h}_{1:M}), \ldots, \mathbf{A}^T(\omega_F, \mathbf{h}_{1:M})\right)^T. \tag{7}$$

Accordingly, the received-signal model can be rewritten as

$$\mathbf{y} = \mathbf{A}(\mathbf{h}_{1:M})\,\mathbf{x} + \mathbf{n}, \tag{8}$$

Here, we reorder $\mathbf{y}$ receiver-wise. Specifically, letting the stacked vector as $\mathbf{y} = (\mathbf{y}_1^T, \ldots, \mathbf{y}_M^T)^T$, each receiver measurement $\mathbf{y}_m$ is given by $\mathbf{y}_m = \mathbf{A}_m(\mathbf{h}_{1:M})\mathbf{x} + \mathbf{n}_m$, where

$$\mathbf{A}_m(\mathbf{h}_{1:M}) = \left(\mathbf{A}_m^T(\omega_1, \mathbf{h}_{1:M}), \ldots, \mathbf{A}_m^T(\omega_F, \mathbf{h}_{1:M})\right)^T,$$

and

$$\mathbf{A}_m(\omega_f, \mathbf{h}_{1:M}) = P(\omega_f)((\mathbf{u}(\omega_f, \mathbf{r}_m) * \mathbf{h}_m)^T \odot \mathbf{U}(\omega_f, \mathbf{h}_{1:M})).$$

Following [16], the joint estimation of $\mathbf{x}$ and $\mathbf{h}_{1:M}$ is formulated as

$$\begin{aligned}
&\underset{\substack{\mathbf{x}\in\mathbb{C}^N \\ \mathbf{h}_{1:M}\in\mathbb{R}_+^{N_e}}}{\arg\min} && \frac{1}{M}\sum_{m=1}^{M} \|\mathbf{y}_m - \mathbf{A}_m(\mathbf{h}_{1:M})\mathbf{x}\|_2^2 \\
&\text{subject to} && R(\mathbf{x}) \leq \tau, \\
& && \|\mathbf{h}_m\|_0 = 1,\ \mathbf{1}^T\mathbf{h}_m = 1,\ \forall m,
\end{aligned} \tag{9}$$

where $\mathbf{1}$ is the all-ones vector, $R(\mathbf{x})$ is a spatial regularizer that promotes predefined structure in $\mathbf{x}$, and τ is a tunable constraint parameter. One example of a suitable regularizer $R(\mathbf{x})$ is the fused-Lasso function [16]. Alternatively, we show below that the structure promoted by $R(\mathbf{x})$ can be more effectively learned in a data-driven fashion through a deep denoiser embedded in an unfolded network architecture.

3. RECONSTRUCTION ALGORITHM

The search for a stationary point of the non-convex minimization problem in (9) is carried out using a block coordinate descent method, which alternates between updates $\mathbf{x}$ and $\mathbf{h}_{1:M}$.

3.1. Estimating the Radar Image $\mathbf{x}$

Given fixed antenna shift estimates $\mathbf{h}_{1:M}$, the update for $\mathbf{x}$ is formulated as the minimization of a data fidelity term constrained to the image of a deep denoiser:

$$\begin{aligned}
\hat{\mathbf{x}} \quad = \quad & \underset{\mathbf{x}}{\arg\min} \textstyle\sum_{m=1}^{M} \frac{1}{2}\|\mathbf{y}_m - \mathbf{A}_m(\mathbf{h}_{1:M})\mathbf{x}\|_2^2 \\
& \text{subject to } \mathbf{x} = \mathcal{D}_{\boldsymbol{\theta}}(\mathbf{x}),
\end{aligned} \tag{10}$$

where $\mathcal{D}_{\boldsymbol{\theta}}(\mathbf{x})$ is a feed forward neural network with parameters $\boldsymbol{\theta}$ that is trained to project noisy intermediate images onto the manifold of focused radar images.

The solution to (10) can be achieved by variable splitting; where an auxiliary variable $\mathbf{z}$ is introduced and updated by a gradient step with respect to the data fidelity term around an intermediate estimate of $\mathbf{x}$. Then $\mathbf{x}$ is updated by projecting $\mathbf{z}$ onto the image of the denoiser $\mathcal{D}_{\boldsymbol{\theta}}$. More specifically, we have in every iteration k:

$$\begin{aligned}
&\mathbf{z}_k = \left(\mathbf{I} - \alpha_k \textstyle\sum_{m=1}^{M} \mathbf{A}_m^H \mathbf{A}_m\right)\mathbf{x}_k + \alpha_k \textstyle\sum_{m=1}^{M} \mathbf{A}_m^H \mathbf{y}_m \\
&\mathbf{x}_{k+1} = \mathcal{D}_{\boldsymbol{\theta}_k}(\mathbf{z}_k),
\end{aligned} \tag{11}$$

where α_k is a descent step size at iteration k.

3.2. Training the deep denoiser $\mathcal{D}_{\boldsymbol{\theta}}$

We adopt an iteration unfolding strategy in this paper, where K iterations of (11) are unfolded as layers of a deep neural network with $\boldsymbol{\theta}$ as the learnable parameters. In this case, a trained denoiser $\mathcal{D}_{\boldsymbol{\theta}}$ is computed by solving:

$$\Theta^*, \mathcal{S}^* = \arg \min_{\{\boldsymbol{\theta}_k\},\{\alpha_k\}} \mathbb{E}_{\mathbf{x}\in\mathcal{X}} \|\mathbf{x}_K(\boldsymbol{\theta}, \boldsymbol{\alpha}) - \mathbf{x}\|^2. \tag{12}$$

where $\boldsymbol{\theta}$ and $\boldsymbol{\alpha}$ include all iteration-specific learnable parameters $\{\boldsymbol{\theta}_k\}$ and $\{\alpha_k\}$, respectively, for $k \in \{1 \ldots K\}$. We observed that providing iteration specific step size and denoising parameters achieves better reconstruction performance with fewer iterations K.

3.3. Estimating the shift kernels $\mathbf{h}_{1:M}$

Let $[M] = \{1, \ldots, M\}$ and $[F] = \{1, \ldots, F\}$ be the sets of indices corresponding to the Khatri-Rao components and the frequency dimension, respectively. We assume the rows of the matrix $\mathbf{A}(\mathbf{h}_{1:M})$ and the vector $\mathbf{y}$ are indexed by the triplet $(i, j, f) \in [M] \times [M] \times [F]$. We define the index set $\mathcal{I}_m$, which collects the row indices associated with the mth component of the underlying Khatri-Rao structure, as follows:

$$\mathcal{I}_m = \{(i, j, f) \in [M] \times [M] \times [F] \mid (i = m \vee j = m)\}. \tag{13}$$

The cardinality of this set is $|\mathcal{I}_m| = F(2M - 1)$. Further, for a fixed frequency index $f \in [F]$, we define the per-frequency subset $\mathcal{I}_{m,f} := \mathcal{I}_m \cap ([M] \times [M] \times \{f\})$. The submatrix $\mathbf{A}_{\mathcal{I}_m}(\mathbf{h}_{1:M})$ and subvector $\mathbf{y}_{\mathcal{I}_m}$ are constructed by extracting the rows indexed by $\mathcal{I}_m$.

The data fidelity cost function for estimating $\mathbf{h}_m$ is given by

$$D_{\mathrm{h}}^{\mathrm{MIMO}}(\mathbf{h}_m) := \frac{1}{2}\|\mathbf{y}_{\mathcal{I}_m} - \mathcal{A}_{\mathcal{I}_m}\mathbf{h}_m\|_2^2. \tag{14}$$

At alternating iteration t, given the image $\mathbf{x}^t$ and the partial parameter vector $\mathbf{h}_{\setminus m}^{(t,m-1)}$, the operator $\mathcal{A}_{\mathcal{I}_m}$ is defined as

$$\mathcal{A}_{\mathcal{I}_m}(\mathbf{x}^t, \mathbf{h}_{\setminus m}^{(t,m-1)}) = \begin{pmatrix} \mathcal{A}_{\mathcal{I}_{m,1}}(\mathbf{x}^t, \mathbf{h}_{\setminus m}^{(t,m-1)}) \\ \vdots \\ \mathcal{A}_{\mathcal{I}_{m,F}}(\mathbf{x}^t, \mathbf{h}_{\setminus m}^{(t,m-1)}) \end{pmatrix} \mathbf{P}, \tag{15}$$

$$\begin{aligned}
&\mathcal{A}_{\mathcal{I}_{m,f}}(\mathbf{x}^t, \mathbf{h}_{\setminus m}^{(t,m-1)}) \\
&\quad := P(\omega_f) \begin{pmatrix} \mathbf{x}^{t\,T}\mathbf{F}_2^H \mathbf{D}_{\mathbf{v}_{f,m}} \mathbf{F}_2 \\ \mathbf{1}_2 \otimes \left(\bar{\mathbf{U}}_{f,m}^t \mathbf{D}_{\mathbf{x}^t} \mathbf{F}_2^H \mathbf{D}_{\mathbf{w}_{f,m}} \mathbf{F}_2\right) \end{pmatrix},
\end{aligned} \tag{16}$$

where

$$\mathbf{v}_{f,m} = \mathbf{F}_2(\mathbf{u}(\omega_f, \mathbf{r}_m) \circ \mathbf{u}(\omega_f, \mathbf{r}_m)), \tag{17}$$

$$\mathbf{w}_{f,m} = \mathbf{F}_2 \mathbf{u}(\omega_f, \mathbf{r}_m), \tag{18}$$

$$\bar{\mathbf{U}}_{f,m}^t = \mathbf{U}_{[M]\setminus\{m\}}(\omega_f, \mathbf{h}_{\setminus m}^{(t,m-1)}). \tag{19}$$

Here, $\mathbf{P}$ is a binary selection matrix corresponding to N_e grid points within the range of the shift kernel. $\mathbf{D}_{\mathbf{z}}$ denotes the diagonal matrix formed by the vector $\mathbf{z}$, $\mathbf{U}_{[M]\setminus m}(\cdot)$ denotes the submatrix of $\mathbf{U}(\cdot)$ formed by selecting the rows corresponding to all antennas except the mth one. The vector $\mathbf{h}_{\setminus m}^{(t,m-1)}$ consists of the updated components for indices $j < m$ and the previous estimates for $j > m$: $\mathbf{h}_{\setminus m}^{(t,m-1)} := (\mathbf{h}_{1:m-1}^t, \mathbf{h}_{m+1:M}^{t-1})$. The symbol $\otimes$ represents the Kronecker product, $\circ$ denotes the element-wise product, and $\mathbf{F}_2$ refers to the two-dimensional Fourier transform matrix.

The first row of (16) represents the case where both the transmitter and receiver are antenna m. The subsequent row covers the signals where only one of the antennas is m. Note that swapping the transmitter and receiver roles yields two measurements for each pair of antennas sharing the same path.

3.4. Direct-Path Signal Utilization

The estimation accuracy of antenna position errors in the cross-range direction deteriorates when targets are at long ranges relative to the aperture size, unless the imaging targets are spatially well-distributed. To address this limitation, we incorporate direct-path signals between distributed Tx–Rx pairs into the estimation of the shift kernel.

$$D_{\mathrm{h}}^{\mathrm{joint}}(\mathbf{h}_m) = \frac{1}{2}\left\| \begin{pmatrix} \mathbf{y}_{\mathcal{I}_m} \\ \mathbf{y}^{\mathrm{Dir}}_{\mathcal{I}_m^{\neq}} \end{pmatrix} - \begin{pmatrix} \mathcal{A}_{\mathcal{I}_m} \\ \mathcal{B}_{\mathcal{I}_m^{\neq}} \end{pmatrix} \mathbf{h}_m \right\|_2^2. \tag{20}$$

In (20), we suppress the dependence of $\mathcal{B}_{\mathcal{I}_m^{\neq}}$ on $\mathbf{h}_{\setminus m}^{(t,m-1)}$ for notational simplicity, and write it simply as $\mathcal{B}_{\mathcal{I}_m^{\neq}}$.

We first define the index sets used for the mth subproblem as $\mathcal{I}_m^{\neq} := \{(i,j,f) \in \mathcal{I}_m \mid i \neq j\}$, and let $V := |\mathcal{I}_m^{\neq}|$. We enumerate $\mathcal{I}_m^{\neq}$ as

$$\mathcal{I}_m^{\neq}(\nu) = (i_\nu, j_\nu, f_\nu), \qquad \nu = 1, \ldots, V, \tag{21}$$

$$\kappa_\nu := i_\nu + j_\nu - m. \tag{22}$$

Since $(i_\nu, j_\nu, f_\nu) \in \mathcal{I}_m^{\neq}$, exactly one of $\{i_\nu, j_\nu\}$ equals m; hence κ_ν denotes the other antenna index.

The direct-path operator is formed as

$$\mathcal{B}_{\mathcal{I}_m^{\neq}} = \begin{pmatrix} \mathcal{B}^T_{\mathcal{I}_m^{\neq}(1)} & \cdots & \mathcal{B}^T_{\mathcal{I}_m^{\neq}(V)} \end{pmatrix}^T, \tag{23}$$

where

$$\mathcal{B}_{\mathcal{I}_m^{\neq}(\nu)} = \mathbf{x}_{\mathrm{cnt}}^T \mathbf{F}_2^H \mathbf{D}_{\mathbf{b}_{\mathcal{I}_m^{\neq}(\nu)}} \mathbf{F}_2, \tag{24}$$

$$\mathbf{b}_{\mathcal{I}_m^{\neq}(\nu)} = \mathbf{F}_2 \left(B_{\mathcal{I}_m^{\neq}(\nu)}(1), \ldots, B_{\mathcal{I}_m^{\neq}(\nu)}(N_e) \right)^T. \tag{25}$$

Let $\mathbf{x}_{\mathrm{cnt}} = [\delta_{i,\mathrm{cnt}}]_{i=1}^{N_e}$, where $\delta_{i,\mathrm{cnt}}$ is the Kronecker delta centered at the grid point corresponding to $\mathbf{r}_m$ This is because, unlike reflected waves, direct waves do not involve the summation of grid points within the region of interest.

Finally, the direct-path steering entry is modeled by

$$B_{\mathcal{I}_m^{\neq}(\nu)}(n) = P(\omega_{f_\nu})\, a(\mathbf{r}_m + \mathbf{e}_{m,n},\, \mathbf{r}_{\kappa_\nu}) \exp\Big(-j\,\frac{\omega_{f_\nu}}{c}\|(\mathbf{r}_m + \mathbf{e}_{m,n}) - \mathbf{r}_{\kappa_\nu}\|_2\Big), \tag{26}$$

where the antenna shifts of $\mathbf{r}_{\kappa_\nu}$ are fixed to the previously estimated values $\mathbf{e}_{\kappa_\nu}^{(t,m-1)}$, and $\mathbf{e}_{m,n}$ denote all possible position perturbations of antenna m that are covered by the shift kernel $\mathbf{h}_m$.

3.5. Reconstruction algorithm

We adopt an alternating minimization procedure that switches between updating the radar image $\mathbf{x}$ while fixing the shift kernels $\{\mathbf{h}_m\}$, followed by fixing the radar image and updating the kernels. The complete procedure is summarized in Algorithm 1.

For updating the image $\mathbf{x}$ given current antenna shift estimates $\hat{\mathbf{h}}_{1:M}^t$, we use the deep unfolding approach described in section 3.1. Given the optimized parameters $\{\Theta^*, \mathcal{S}^*\}$ obtained through the supervised training regime in (12), the image update at iteration t is defined as the solution $\hat{\mathbf{x}}_K^t$ to (11) after K unfolded steps. Next, we update the antenna shift kernels $\mathbf{h}_m$ for each $m \in [M]$. Using the updated image $\mathbf{x}^{t+1}$, we solve the following constrained optimization problem:

$$\hat{\mathbf{h}}_m^{t+1} = \underset{\mathbf{h}_m \in \mathbb{R}_+^{N_e}}{\arg\min}\, D_{\mathrm{h}}^{\mathrm{joint}}(\mathbf{h}_m) \quad \text{subject to} \quad \|\mathbf{h}_m\|_0 = 1, \quad \mathbf{1}^T\mathbf{h}_m = 1. \tag{27}$$

The solution to (27) above can be obtained with a brute force search over all N_e one-hot kernels. Note that we adopt a greedy approach for updating the kernels that loops over the antenna shifts only once. While this approach is suboptimal, we demonstrate in section 4 that it sufficiently corrects for the antenna position ambiguities.

Algorithm 1 Block Coordinate Descent for Deep-Unfolded Distributed MIMO Radar Autofocus Imaging

Input: Measurements $\{\mathbf{y}_m\}_{m=1}^M$, Trained network parameters $\{\Theta^*, \mathcal{S}^*\}$.
Initialize: Set initial antenna kernels $\hat{\mathbf{h}}_{1:M}^{(0)}$.
1: **for** $t \leftarrow 0$ to $T-1$ **do**
2: Obtain $\hat{\mathbf{x}}_K^{t+1}(\Theta^*, \mathcal{S}^*)$ by solving (11) for K steps.
3: **for** $m \leftarrow 1$ to M **do**
4: Obtain $\hat{\mathbf{h}}_m^{t+1}$ by solving (27)
5: **end for**
6: **end for**
Return: $\hat{\mathbf{x}}^T$, $\hat{\mathbf{h}}_{1:M}^T$.

4. PERFORMANCE EVALUATION

We conducted numerical simulations to validate the effectiveness of the proposed method. In the observation setup, we use ten radars, each equipped with an array antenna. The radar transmits a signal with a fractional bandwidth of 1.33, sampled at 12 uniformly spaced frequency points. Let λ_c denote the wavelength corresponding to the center frequency of the transmit band. The radars are deployed along a 120° circular arc with a diameter of 90,000 λ_c, with approximately uniform spacing of 10,000 λ_c, and the azimuth boresight of each radar is directed toward the center of the circle. Each radar had an azimuthal field of view of $\pm 60^\circ$, and the direct-path signals from other radars within this field of view were utilized.

The imaging region was located at a range of 50,000 λ_c from the midpoint of the radar array, in the direction normal to the circular arc defined by the radar array geometry. The grid spacing was set to $\lambda_c/12$, defining an imaging region of 64×64 grids and an antenna position error search area of 25×25 grids. For the evaluation of imaging performance on generic target shapes, the dataset comprised 6,000 images generated from 12 geometric shapes (circle, semicircle, oval, triangle, square, rectangle, parallelogram, rhombus, trapezoid, kite, star, and pentagon) with randomized sizes, positions, and rotations (500 samples each). To better reflect realistic radar target characteristics, the line-of-sight components and partial reflections from the upper and lower boundaries were randomly subsampled, and the amplitude of each scattering point was assigned randomly. We then add Gaussian random noise to the radar measurement with the target peak signal-to-noise ratio (PSNR). To reduce the amount of training data, no radar operator mismatch was introduced during training, except for the convolutional neural network (CNN)-based method described later; such errors were added only during performance evaluation. These were split into 80% for training, 10% for validation, and 10% for testing. The proposed network utilized five unfolding layers with a maximum of 30 outer loop iterations and was optimized using Adam with a learning rate of 1×10^{-3} and a weight decay of 10^{-8}. We used DnCNN [23] as the denoiser $\mathcal{D}_{\boldsymbol{\theta}_k}$. Simulations were performed under noise conditions of 20 dB PSNR, defined after coherent integration of all MIMO reflection paths.

Figure 3 presents a qualitative comparison of images reconstructed from radar received signals contaminated with noise and antenna position errors. With respect to the ground-truth image, we compare the back-projection image without antenna position error compensation, the CNN reconstruction [23] trained using the back-projection image as input, the deep prior autofocus image similar to [22] but with training data restrictions, and the deep unfolding

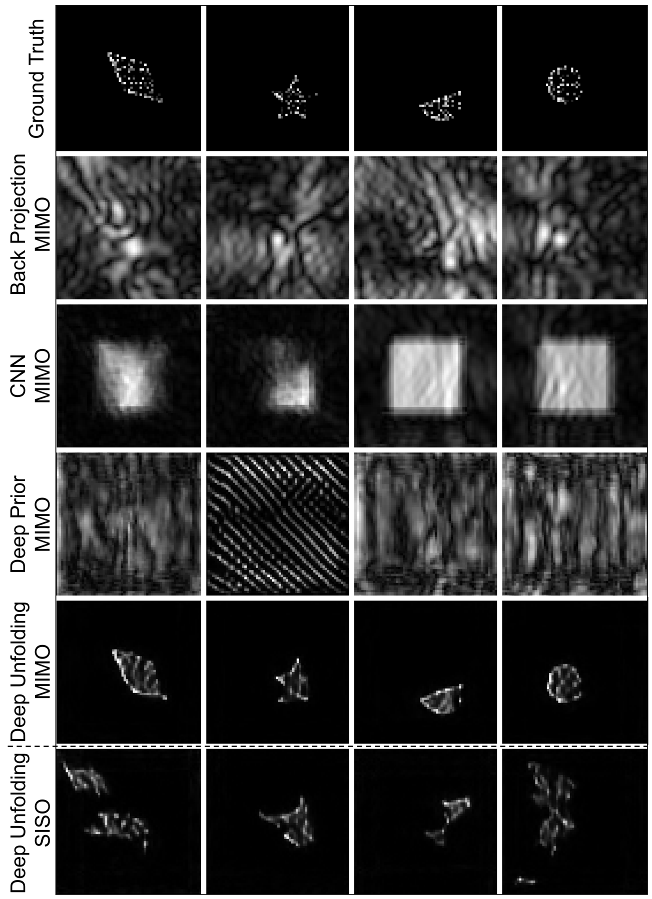

Fig. 3. Reconstruction examples for four radar scenes at PSNR = 20 dB. From top to bottom, the rows show the ground-truth image, the back-projection image without antenna position error compensation, the CNN reconstruction trained using the back-projection image as input, the deep prior autofocus image, the deep unfolding autofocus images with MIMO and SISO configurations.

autofocus imaging approach described above. For the CNN method, training incorporated antenna position errors, as the CNN performs direct imaging without explicit error estimation. Only the deep unfolding method presents results for both MIMO and single-input single-output (SISO) configurations, whereas all other estimation results are based on a MIMO signal model. In addition, the deep prior and deep unfolding methods, which explicitly estimate antenna position errors, utilize the direct-path signal.

The back projection, CNN, and deep prior methods all failed to achieve accurate imaging because of the antenna position error and sidelobe artifacts. Although the CNN was trained on diverse antenna position error patterns, it failed to adequately compensate for these errors, resulting in unclear target reconstruction. The deep prior method similarly produced indiscernible target shapes, due to training data free from antenna position errors, a limited number of iterations to match the deep unfolding setup, and insufficient model expressiveness caused by the small number of learnable parameters. In contrast, the proposed deep unfolding method achieves accurate imaging performance, allowing characteristic shapes such as the rhombus and star structures to be clearly identified.

Table 1 presents the PSNR and structural similarity index measure (SSIM) evaluations of the reconstructed images relative to the

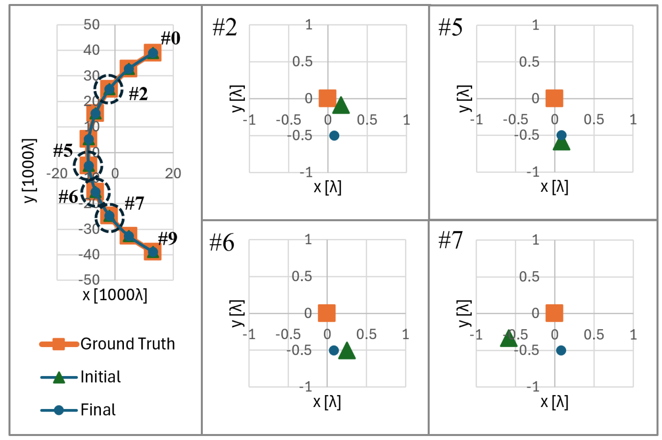

Fig. 4. An example of estimated antenna position errors. The left panel displays the overall array configuration. In the right panel, we plotted the positions of four selected antennas relative to the ground truth. The final estimates demonstrate that, while a fixed offset exists, the relative positions are successfully estimated.

Table 1. PSNR [dB] and SSIM comparison, where deep unfolding combined with MIMO and direct-path signal utilization attains the best performance in both metrics.

Method and Configuration		Evaluation Results	
$\hat{\mathbf{x}}$	$\hat{\mathbf{h}}_{1:M}$	PSNR	SSIM
Fused Lasso	MIMO	24.1	0.514
Deep Prior	MIMO	19.8	0.677
Deep Unfolding	*MIMO*	*25.8*	*0.809*
Deep Unfolding	SISO	24.9	0.755

*Note: Global shift correction is performed for evaluations.

ground truth images averaged over the test dataset. We compare our proposed deep unfolding method utilizing MIMO and the direct wave, with the deep prior [22], fused lasso [16], and the deep unfolding result under a conventional SISO configuration. It can be seen that the proposed method achieves the highest performance among all compared methods in terms of both metrics. The performance metrics are computed after correcting for the global shift ambiguity that is inherent to this blind deconvolution problem by shifting the reconstructed images to maximize the correlation with the ground truth images. Note that the global shift ambiguity is also evident in the computation of the corrected antenna positions. Fig. 4 illustrates that while the relative positions of the estimated antenna locations match the true antenna positions, there remains a global shift ambiguity in the final estimation.

5. CONCLUSION

We proposed a deep-unfolded, MIMO-extended autofocus imaging method for distributed radar that effectively enhances imaging performance. The proposed method unfolds an iterative imaging algorithm based on deep denoisers by running only a small number of iterations. Moreover, to fully exploit the advantages of distributed radar systems, a MIMO configuration and the use of direct-path signals are introduced. Numerical simulations validated the effectiveness of the proposed approach, demonstrating improved PSNR and SSIM compared with state of the art methods, as well as clear reconstruction of target shapes and salient features composed of multiple scattering points.

6. REFERENCES

[1] John C. Curlander and Robert N. McDonough, *Synthetic Aperture Radar: Systems and Signal Processing*, Wiley Series in Remote Sensing and Image Processing. Wiley, New York, 1991.

[2] Victor C. Chen and Marco Martorella, *Inverse Synthetic Aperture Radar Imaging: Principles, Algorithms and Applications*, Institution of Engineering and Technology (IET) / SciTech Publishing, 2014.

[3] M. A. Herman and T. Strohmer, "High-resolution radar via compressed sensing," *IEEE transactions on signal processing*, vol. 57, no. 6, pp. 2275–2284, 2009.

[4] Y. Yu, A. P. Petropulu, and H. V. Poor, "Mimo radar using compressive sampling," *IEEE Journal of Selected Topics in Signal Processing*, vol. 4, no. 1, pp. 146–163, 2010.

[5] C. R. Berger and J. M. F. Moura, "Noncoherent compressive sensing with application to distributed radar," in *Information Sciences and Systems (CISS), 2011 45th Annual Conference on*. IEEE, 2011, pp. 1–6.

[6] Quasi-Zenith Satellite System Services Inc. (QSS), "Quasi-zenith satellite system service performance report for 1sth fy2025: Centimeter level augmentation service (clas)," Service performance report, Quasi-Zenith Satellite System Services Inc. (QSS), Jan. 2026.

[7] T Derham, S Doughty, C Baker, and K Woodbridge, "Ambiguity functions for spatially coherent and incoherent multistatic radar," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 46, no. 1, 2010.

[8] W. Ye, T. S. Yeo, and Z. Bao, "Weighted least-squares estimation of phase errors for sar/isar autofocus," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 37, no. 5, pp. 2487–2494, 1999.

[9] Kuang-Hung Liu, Ami Wiesel, and David C Munson, "Synthetic aperture radar autofocus via semidefinite relaxation," *IEEE transactions on image processing*, vol. 22, no. 6, pp. 2317–2326, 2013.

[10] Kuang-Hung Liu, Ami Wiesel, and David C Munson, "Synthetic aperture radar autofocus based on a bilinear model," *IEEE Transactions on image Processing*, vol. 21, no. 5, pp. 2735–2746, 2012.

[11] 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.

[12] N. O. Onhon and M. Cetin, "A sparsity-driven approach for joint sar imaging and phase error correction," *IEEE Transactions on Image Processing*, vol. 21, no. 4, pp. 2075–2088, 2012.

[13] Ali Ahmed, Benjamin Recht, and Justin Romberg, "Blind deconvolution using convex programming," *IEEE Transactions on Information Theory*, vol. 60, no. 3, pp. 1711–1732, 2014.

[14] Yanjun Li, Kiryung Lee, and Yoram Bresler, "Optimal sample complexity for blind gain and phase calibration.," *IEEE Trans. Signal Processing*, vol. 64, no. 21, pp. 5549–5556, 2016.

[15] X. Du, C. Duan, and W. Hu, "Sparse representation based autofocusing technique for isar images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 51, no. 3, pp. 1826–1835, 2013.

[16] Hassan Mansour, Dehong Liu, Ulugbek Kamilov, and Petros T. Boufounos, "Sparse Blind Deconvolution for Distributed Radar Autofocus Imaging," *IEEE Transactions on Computational Imaging*, vol. 4, no. 4, pp. 537–551, Dec. 2018.

[17] Karol Gregor and Yann LeCun, "Learning fast approximations of sparse coding," in *Proceedings of the 27th International Conference on International Conference on Machine Learning*, Madison, WI, USA, 2010, ICML'10, p. 399–406, Omnipress.

[18] J.R. Hershey, J. Le Roux, and F. Weninger, "Deep Unfolding: Model-Based Inspiration of Novel Deep Architectures," *arXiv*, Aug. 2014.

[19] Yuelong Li, Mohammad Tofighi, Junyi Geng, Vishal Monga, and Yonina C. Eldar, "Efficient and interpretable deep blind image deblurring via algorithm unrolling," *IEEE Transactions on Computational Imaging*, vol. 6, pp. 666–681, 2020.

[20] Zhi Liu, Shuyuan Yang, Quanwei Gao, Zhixi Feng, Min Wang, and Licheng Jiao, "Afnet and pafnet: Fast and accurate sar autofocus based on deep learning," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–13, 2022.

[21] Xiaoyu Qin, Bin Deng, and Hongqiang Wang, "A high-resolution learning imaging method for thz-sar moving targets based on af-rpca-net," *Signal Processing*, vol. 243, pp. 110499, 2026.

[22] Hassan Mansour, Suhas Lohit, and Petros T. Boufounos, "Distributed Radar Autofocus Imaging Using Deep Priors," in *IEEE International Conference on Image Processing (ICIP)*, Oct. 2022, pp. 2511–2515.

[23] Kai Zhang, Wangmeng Zuo, Yunjin Chen, Deyu Meng, and Lei Zhang, "Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising," *IEEE Transactions on Image Processing*, vol. 26, no. 7, pp. 3142–3155, 2017.