EFFICIENT THREE-DIMENSIONAL NEAR-FIELD IMAGING WITH PHYSICS-INFORMED DEEP LEARNING FOR MIMO RADAR

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

OKYANUS ORAL

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
ELECTRICAL AND ELECTRONICS ENGINEERING

JUNE 2024

Approval of the thesis:

EFFICIENT THREE-DIMENSIONAL NEAR-FIELD IMAGING WITH PHYSICS-INFORMED DEEP LEARNING FOR MIMO RADAR

submitted by **OKYANUS ORAL** in partial fulfillment of the requirements for the degree of **Master of Science in Electrical and Electronics Engineering Department, Middle East Technical University** by,

Prof. Dr. Naci Emre Altun
Dean, Graduate School of **Natural and Applied Sciences** _____

Prof. Dr. İlkay Ulusoy
Head of Department, **Electrical and Electronics Engineering** _____

Assoc. Prof. Dr. S. Figen Öktem Seven
Supervisor, **Electrical and Electronics Engineering, METU** _____

Examining Committee Members:

Prof. Dr. Tolga Çiloğlu
Electrical and Electronics Engineering, METU _____

Assoc. Prof. Dr. S. Figen Öktem Seven
Electrical and Electronics Engineering, METU _____

Prof. Dr. T. Engin Tuncer
Electrical and Electronics Engineering, METU _____

Prof. Dr. A. Aydın Alatan
Electrical and Electronics Engineering, METU _____

Prof. Dr. Orhan Arıkan
Electrical and Electronics Engineering, Bilkent University _____

Date:10.06.2024

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Surname: Okyanus Oral

Signature :

ABSTRACT

EFFICIENT THREE-DIMENSIONAL NEAR-FIELD IMAGING WITH PHYSICS-INFORMED DEEP LEARNING FOR MIMO RADAR

Oral, Okyanus
M.S., Department of Electrical and Electronics Engineering
Supervisor: Assoc. Prof. Dr. S. Figen Öktem Seven

June 2024, 90 pages

Near-field radar imaging systems are used in a wide range of applications, such as medical diagnosis, through-wall imaging, concealed weapon detection, and nondestructive evaluation. In this thesis, we consider the inverse problem of reconstructing the three-dimensional (3D) complex-valued reflectivity distribution of the near-field scene from the sparse multiple-input multiple-output (MIMO) array measurements. Motivated by recent advances, we develop physics-informed deep learning techniques for the image reconstruction and array optimization tasks encountered in near-field MIMO radar imaging. Firstly, we develop a novel plug-and-play (PnP) reconstruction method that exploits deep priors and regularization on the magnitude. Our approach provides a unified general framework to effectively handle arbitrary regularization on the magnitude of a complex-valued unknown and is equally applicable to other radar image formation problems including SAR. Secondly, we focus on existing learned direct inversion methods that enable real-time imaging and perform modifications to improve these methods. We demonstrate the effectiveness of all developed approaches under various compressive and noisy observation scenarios using both simulated and experimental data. We also analyze the resolution achieved at compressive settings

with sparse MIMO arrays. The developed methods enable not only state-of-the-art performance for 3D real-world targets but also fast computation. Lastly, we develop a novel method for joint optimization of MIMO arrays and reconstruction methods. We illustrate the performance of the jointly optimized imaging system by utilizing various reconstruction methods and different observation settings, and compare the performance with the commonly used MIMO arrays.

Keywords: 3D Near-Field MIMO Radar Imaging, Complex-Valued Reconstruction, Plug-and-Play Methods, Deep Learning, Joint Optimization

ÖZ

FİZİK TABANLI DERİN ÖĞRENME TEKNİKLERİ İLE ÜÇ BOYUTLU YAKIN ALAN MIMO RADAR GÖRÜNTÜLEME

Oral, Okyanus
Yüksek Lisans, Elektrik ve Elektronik Mühendisliği Bölümü
Tez Yöneticisi: Doç. Dr. S. Figen Öktem Seven

Haziran 2024 , 90 sayfa

Yakın alan radar görüntüleme sistemleri tıbbi teşhis ve gizli silah tespiti gibi çeşitli uygulamalarda kullanılmaktadır. Bu tezde, seyrek çok-girişli çok-çıkışlı (MIMO) radar anten dizisi ölçümlerinden yakın alan sahnesinin üç boyutlu karmaşık değerli yansıtırlık dağılımının geri çatılması ters problemine odaklanılmaktadır. Yakın alan MIMO radar görüntülemesinde karşılaşılan imgenin geri çatılması ve anten dizilimi eniyilemesi problemleri için, son gelişmelerden harekete geçerek, fizik tabanlı ve derin öğrenmeye dayalı teknikler geliştiriyoruz. İlk olarak, imgenin geri çatılması problemi için derin öğrenmeye dayalı önsel bilgilerden ve büyüklük üzerindeki düzenlileştirmelerden yararlanan yeni bir tak-çalıştır yöntemi geliştiriyoruz. Bu yaklaşımımız, karmaşık değerli bilinmeyenin büyüklüğü üzerinde herhangi bir düzenlileştirme uygulayabilmek için genel bir çerçeve sağlamakta ve sentetik açıklıklı radar dahil radar görüntülemede karşılaşılan diğer ters problemlere benzer bir şekilde uygulanabilmektedir. İkinci olarak, mevcut imge oluşturma yöntemlerinden gerçek zamanlı görüntülemeyi mümkün kılan öğrenilmiş direkt evirme yöntemlerine odaklanıyoruz ve bu yöntemleri iyileştiriyoruz. Geliştirilen tüm yaklaşımların başarımını çeşitli ölçüm senaryoları

altında hem simüle edilmiş hem de deneysel verileri kullanarak gösteriyoruz. Ayrıca seyrek MIMO dizileriyle sıkıştırmalı örnekleme için elde edilen çözünürlüğü de analiz ediyoruz. Geliştirilen yöntemler, yalnızca üç boyutlu gerçek hedefler için en iyi başarımı sağlamakla kalmayıp aynı zamanda hızlı hesaplanabilmektedir. Son olarak, MIMO dizilimlerinin ve imge oluşturma yöntemlerinin aynı anda eniyilenmesi için yeni bir yöntem geliştiriyoruz. Çeşitli imge oluşturma yöntemleri ve farklı ölçüm senaryoları kullanarak eniyilenmiş görüntüleme sistemlerinin başarımını gösteriyor ve bu başarımı yaygın kullanılan MIMO dizileriyle karşılaştırıyoruz.

Anahtar Kelimeler: 3-Boyutlu Yakın Alan MIMO Radar Görüntüleme, Karmaşık Değerli Geriçatım, Tak-Çalıştır Yöntemleri, Derin Öğrenme, Bileşik Eniyileme

to my family

ACKNOWLEDGMENTS

First and foremost, I would like to express my sincere gratitude to my supervisor, Assoc. Prof. Dr. Figen S. Öktem, for her invaluable expertise, guidance, and support, which have greatly contributed to my personal and academic growth.

I am grateful to the members of my thesis committee, Prof. Dr. Tolga Çiloğlu, Prof. Dr. T. Engin Tuncer, Prof. Dr. A. Aydın Alatan, and Prof. Dr. Orhan Arıkan, for showing interest in my studies, reviewing and commenting on my thesis.

This work is supported by the Scientific and Technological Research Council of Turkey (TUBITAK) under Grant 120E505 (1001 Research Program).

I would like to express my deepest gratitude to my family, who are the most fun, uplifting, and inspiring people I know.

TABLE OF CONTENTS

LIST OF TABLES

TABLES

LIST OF FIGURES

FIGURES

LIST OF ALGORITHMS

ALGORITHMS

LIST OF ABBREVIATIONS

ABBREVIATIONS	
1D	1 Dimensional
2D	2 Dimensional
3D	3 Dimensional
Adam	Adaptive Moment Estimation
ADMM	Alternating Direction Method of Multipliers
BM3D	Block-Matching and 3D Filtering
BP	Back-Projection
CG	Conjugate-Gradient
CL	Compression Level
CNN	Convolutional Neural Network
CUDA	Compute Unified Device Architecture
CV-Deep2S	Complex-Valued Deep Two-Stage
C-SALSA	Constrained Split Augmented Lagrangian Shrinkage Algorithm
DeepDI	Deep Direct Inversion
Deep2S	Deep Two-Stage
Deep2S+	Deep Two-Stage Plus
DNN	Deep Neural Network
E2E	End-to-End
FMM	Fast Multipole Method
GPU	Graphics Processing Unit
ISAR	Inverse Synthetic-Aperture Radar
KM	Kirchhoff Migration

MCA	Mill's Cross Array
MIMO	Multiple-Input Multiple-Output
MRI	Magnetic Resonance Imaging
MSE	Mean Squared Error
PnP	Plug-and-Play
PSNR	Peak Signal to Noise Ratio
ReLU	Rectified Linear Unit
RSA	Ring Spiral Array
RTX	Ray Tracing Texel eXtreme
RV-CNN	Real-Valued Convolutional Neural Network
SAR	Synthetic-Aperture Radar
SNR	Signal to Noise Ratio
Ti	Titan
TV	Total-Variation
URA	Uniform Rectangular Array
U-ℓ_1	Unrolled Iterative Shrinkage Thresholding Algorithm

CHAPTER 1

INTRODUCTION

1.1 Motivation and Problem Definition

Near-field radar imaging systems are used in a wide range of applications such as medical diagnosis, through-wall imaging, concealed weapon detection, and nondestructive evaluation [1–4]. Compared to the classical monostatic planar arrays with colocated transmitter and receiver antennas, multiple-input multiple-output (MIMO) arrays offer reduced hardware complexity and cost for high-resolution imaging applications. As a result, there has been a growing interest in using MIMO arrays for near-field radar imaging [2–8].

Reconstructing the three-dimensional (3D) complex-valued scene reflectivity from the sparse MIMO measurements is a highly ill-posed problem. Therefore, the reconstruction quality greatly depends on the utilized imaging algorithm and the array topology. As a result, many studies in the literature focus on improving the reconstruction performance through the development of efficient reconstruction algorithms [9–18] and antenna array designs [5, 8, 19–24].

Traditional imaging algorithms, such as direct inversion schemes, offer low computational complexity, but their performance substantially degrades in ill-posed settings with limited and noisy data [9–12]. On the other hand, motivated by the compressed sensing theory, the regularized reconstruction methods with sparsity priors offer promising imaging performance in compressive settings [14–18]. However, a common drawback of these methods is their iterative nature, resulting in longer computation time than the direct inversion approaches. More importantly, hand-crafted regularization functionals correspond to using simple image priors, which do not ac-

curately represent the distribution of target images.

With the recent developments in deep learning, learned reconstruction methods emerged as powerful alternatives to both regularized and direct reconstruction methods [25–33]. These methods employ deep neural networks (DNNs) in the reconstruction process to form images either directly from measurements or use DNNs within model-based imaging algorithms. To reduce the information load on the utilized deep neural networks and exploit physics-based information, the state-of-the-art learned reconstruction methods incorporate the forward model of the imaging system into the reconstruction process. This alleviates the need to capture the physics of the imaging system solely from statistical regularities in the data. Despite the recent success of these methods, most of these approaches are developed for 2D and real-valued image reconstruction problems [28, 32, 34]. In particular, there is very limited research for physics-based deep learning methods, for 3D near-field MIMO radar imaging.

Moreover, for many applications in radar imaging, the phase of the reflectivity at a particular point can be modeled as random and uncorrelated with the phase at other points [35, 36]. This is due to the phase shifts that can occur when imaging rough surfaces and also at the air/target interface due to the electrical properties of materials [35]. Although this is the case, most of the existing reconstruction methods, including deep-learning based ones, do not take into account the random phase nature of the scene reflectivity values [17, 18, 37, 38]. There is also no unified framework in the literature to account for the random phase when enforcing regularization [15, 33, 36, 39].

Since array topology plays an important role in the imaging performance, there has also been a significant effort in the literature to design and optimize antenna arrays to yield better reconstruction quality [5, 8, 19–24, 40, 41]. In spite of this effort, there are only a few works that systematically optimize the arrays to directly improve the reconstruction performance. Instead, most of the existing works design the MIMO arrays using heuristics and ad-hoc criteria [5, 19, 21, 24, 41]. Recent advancements in computational imaging have shown that imaging performance can be improved significantly by jointly optimizing the reconstruction algorithm and the parameters of the observation system [20, 23, 40, 42–48]. Although there are some works in near-field radar imaging for joint optimization [20, 23, 40], there is no comprehensive and

generalizable framework demonstrated for end-to-end optimization of MIMO arrays for 3D near-field imaging.

1.2 Relevant Work

In this thesis, we consider the inverse problem of reconstructing the 3D complex-valued reflectivity distribution of the near-field scene from the sparse radar measurements. Motivated by recent advances, we develop physics-informed deep learning techniques for the image reconstruction and array optimization tasks encountered in this imaging problem.

Despite the recent success of plug-and-play (PnP) algorithms that exploit powerful denoisers within regularized inversion frameworks, there is no study on PnP methods that exploit deep priors for near-field radar imaging where we encounter a 3D complex-valued image reconstruction problem. Here, we develop a novel PnP method to reconstruct the 3D complex-valued reflectivity distribution of the near-field scene from sparse MIMO radar measurements. Due to the random phase nature of scene reflectivity values in various applications, our method exploits regularization on the magnitude. By providing a general expression for the proximal mapping associated with such regularization functionals, we show that arbitrary regularization on magnitude can be efficiently enforced within our PnP framework using simple and computationally efficient update steps.

We also focus on existing learned direct inversion methods that enable real-time imaging and perform modifications to improve these methods. Existing learned direct inversion methods in the literature either perform complex-valued processing or magnitude-only processing. To the best of our knowledge, there is no work that compared the performance of complex-valued refinement with magnitude-only refinement to reconstruct reflectivities involving random phase, and showed the superiority of one of them to the other. For this reason, we also perform this comparative evaluation that has been missing in the earlier works.

We demonstrate the effectiveness of the developed approaches under various compressive and noisy observation scenarios using both simulated and experimental data.

We also analyze the resolution offered by sparse MIMO arrays in such compressive imaging settings.

Lastly, we develop a general framework for the joint optimization of MIMO arrays and reconstruction methods in an end-to-end fashion. We demonstrate the developed framework by using various reconstruction methods under different observation settings. We also compare the performance of the optimized imaging systems with the commonly used MIMO arrays.

1.3 Novelties and Contributions

The contributions and novelties of this thesis are as follows:

- Providing a unified PnP framework to effectively handle arbitrary regularization on the magnitude of a complex-valued unknown involving random phase,
- Development of a novel deep learning-based PnP reconstruction method for 3D complex-valued imaging with application to near-field MIMO radar imaging,
- End-to-end joint optimization of MIMO arrays and reconstruction methods,
- Improvements and comparative performance analysis on learned direct reconstruction methods,
- Comprehensive experiments on synthetic 3D scenes with quantitative and qualitative analysis by considering various compressive and noisy observation scenarios,
- Performance evaluation with experimental measurements to demonstrate reconstruction of 3D real-world targets, and comparison with the commonly used direct inversion and regularized reconstruction methods.

1.4 The Outline of the Thesis

The outline of the thesis is as follows. In Chapter 2, we present the observation model relating the near-field scene to the MIMO array measurements. This chapter also dis-

cusses the inverse problem of reconstructing the scene reflectivity values from the MIMO radar data and reviews the existing imaging algorithms. We present the developed plug-and-play reconstruction method for regularization on the magnitude in Chapter 3. Here, we also provide the simulation setting and the experimental data that are utilized throughout the thesis. Chapter 4 presents our analysis and improvements for the existing learned direct inversion methods in the literature. In Chapter 5 the developed joint optimization method is presented. Lastly, we conclude the thesis in Chapter 6 by providing the final remarks and discussing the possible research directions for future studies.

CHAPTER 2

THREE-DIMENSIONAL NEAR-FIELD MIMO RADAR IMAGING

This chapter is organized as follows. In Section 2.1 we describe the working principle of a near-field MIMO radar imaging system by providing the observation model that relates the near-field MIMO array measurements to the reflectivity distribution of the scene. In Section 2.2 we define the inverse problem for three-dimensional image reconstruction and provide a summary on the existing image reconstruction approaches in the literature. Here, our goal is to provide an overview of the strengths and the drawbacks of existing imaging algorithms, where the analyses motivate the methods developed in this thesis.

2.1 Observation Model

Consider the general MIMO imaging setting illustrated in Fig. 2.1 with spatially distributed transmit and receive antennas on the antenna array located at $z = 0$. In order to infer the 3D reflectivity distribution of the scene, each transmit antenna, located at $\mathbf{r}_T = [x_T, y_T, 0]^T$, illuminates the scene with a pulse signal and the scattered field from the scene is measured by a receive antenna, located at $\mathbf{r}_R = [x_R, y_R, 0]^T$.

Under Born approximation, time-domain response of a single point-scatterer with reflectivity $s(\mathbf{r})$ and located at $\mathbf{r} = [x, y, z]^T$ can be expressed as follows [9]:

$$\tilde{y}(\mathbf{r}_T, \mathbf{r}_R, t) = \frac{p(t - \frac{d(\mathbf{r}_T, \mathbf{r})}{c} - \frac{d(\mathbf{r}_R, \mathbf{r})}{c})}{4\pi\ d(\mathbf{r}_T, \mathbf{r})\ d(\mathbf{r}_R, \mathbf{r})} s(\mathbf{r}). \tag{2.1}$$

Here $\tilde{y}(\mathbf{r}_T, \mathbf{r}_R, t)$ denotes the time-domain measurement acquired using the transmit and receive antenna pair located respectively at $\mathbf{r}_T$ and $\mathbf{r}_R$ due to a single scatterer. The transmitted pulse is denoted by $p(t)$, and c denotes the speed of light. The distances

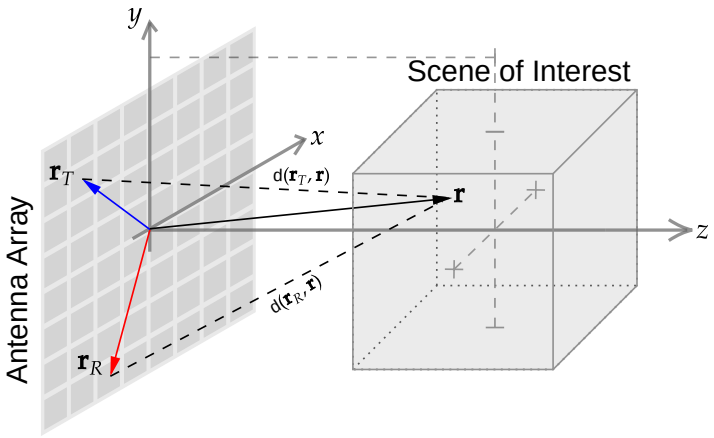


Figure 2.1: Schematic view of a near-field MIMO radar imaging system.

of the scatterer to the corresponding transmitter and receiver are given by $d(\mathbf{r}_T, \mathbf{r}) = \|\mathbf{r}_T - \mathbf{r}\|_2$ and $d(\mathbf{r}_R, \mathbf{r}) = \|\mathbf{r}_R - \mathbf{r}\|_2$ respectively.

By taking 1D Fourier transform over t, the received signal due to a single scatterer can be expressed in the temporal frequency domain as follows:

$$\tilde{y}(\mathbf{r}_T, \mathbf{r}_R, k) = h(\mathbf{r}_T, \mathbf{r}_R, k, \mathbf{r}) s(\mathbf{r}), \tag{2.2}$$

where

$$h(\mathbf{r}_T, \mathbf{r}_R, k, \mathbf{r}) = \tilde{p}(k) \frac{e^{-jk\left(d(\mathbf{r}_T, \mathbf{r}) + d(\mathbf{r}_R, \mathbf{r})\right)}}{4\pi\, d(\mathbf{r}_T, \mathbf{r})\, d(\mathbf{r}_R, \mathbf{r})}, \tag{2.3}$$

and $\tilde{p}(k)$ is the temporal Fourier transform of the transmitted pulse with $k = \frac{2\pi}{c} f$ denoting the frequency-wavenumber whereas f denoting the temporal frequency. Using (2.2), the measurement, $y(\mathbf{r}_T, \mathbf{r}_R, k)$, due to an extended target can be expressed as the superposition of these responses from point-scatterers:

$$y(\mathbf{r}_T, \mathbf{r}_R, k) = \int_x \int_y \int_z h(\mathbf{r}_T, \mathbf{r}_R, k, \mathbf{r}) s(\mathbf{r})\, d\mathbf{r}. \tag{2.4}$$

Since the measurements are discrete, and the image reconstruction algorithm will be run on a computer, a discrete forward model is needed. For this, the coordinate variables are discretized based on the expected range and cross-range resolutions of the used MIMO imaging system [9]. Then the discretized scene reflectivity values can be related to the discrete measurements obtained using different transmitter-receiver

pairs and frequency steps as

$$y(\mathbf{r}_{T_m}, \mathbf{r}_{R_m}, k_m) = \sum_n h(\mathbf{r}_{T_m}, \mathbf{r}_{R_m}, k_m, \mathbf{r}_n) s(\mathbf{r}_n). \tag{2.5}$$

Here the subscript m indicates the location of the transmitting and receiving antennas as well as the frequency used in the mth measurement. Moreover, the subscript n indicates the voxel number in the discretized 3D scene.

By using lexicographical ordering, the measurements and the reflectivity values of the image voxels are put into the following vectors:

$$\mathbf{y} = [y(\mathbf{r}_{T_1}, \mathbf{r}_{R_1}, k_1), \ldots, y(\mathbf{r}_{T_M}, \mathbf{r}_{R_M}, k_M)]^T \in \mathbb{C}^M, \tag{2.6}$$

$$\mathbf{s} = [s(\mathbf{r}_1), \ldots, s(\mathbf{r}_N)]^T \in \mathbb{C}^N, \tag{2.7}$$

where M and N respectively represent the number of measurements and voxels. Then using (2.5) we can express the noisy measurements in matrix-vector form as follows:

$$\mathbf{y} = \mathbf{A}\mathbf{s} + \mathbf{w}. \tag{2.8}$$

The matrix $\mathbf{A} \in \mathbb{C}^{M \times N}$ is the observation matrix whose (m, n)th element is given by

$$\mathbf{A}_{m,n} = h(\mathbf{r}_{T_m}, \mathbf{r}_{R_m}, k_m, \mathbf{r}_n), \tag{2.9}$$

which represents the contribution of the nth voxel at location $\mathbf{r}_n$ to the mth measurement taken using the transmitter at $\mathbf{r}_{T_m}$, receiver at $\mathbf{r}_{R_m}$, and frequency $\frac{c}{2\pi} k_m$. Also $\mathbf{w} \in \mathbb{C}^M$ represents the additive noise vector. We assume white Gaussian noise since it commonly holds in practical applications of interest. Hence each noise component is uncorrelated over different voxels and has variance σ_w^2.

2.2 Inverse Problem

In the inverse problem, the goal is to estimate the 3D complex-valued reflectivity field, $\mathbf{s}$, from the acquired radar measurements, $\mathbf{y}$. This corresponds to solving an underdetermined problem with sparse MIMO measurements $M \ll N$. As a result, the reconstruction quality greatly depends on the utilized computational imaging method.

The existing reconstruction methods in the literature can be grouped into three main categories. These are traditional direct inversion, regularized reconstruction, and deep learning based reconstruction methods.

2.2.1 Traditional Direct Inversion Methods

Traditional direct inversion methods do not utilize any prior information and are solely derived to obtain a direct solution based on the forward (observation) model expression. These methods generally involve back-projecting measurements to the image domain and then employ a filter-like operation [9–12].

Commonly used traditional direct inversion methods for near-field radar imaging include range migration [9], back-projection [4], and Kirchhoff migration [11]. Based on the encountered observation model, direct inversion methods can be expressed in the following general form:

$$\hat{s}(\mathbf{r}_n) = \sum_m y(\mathbf{r}_{T_m}, \mathbf{r}_{R_m}, k_m)\, \xi(\mathbf{r}_n, \mathbf{r}_{R_m}, \mathbf{r}_{T_m}, k_m)\, e^{jk\left(d(\mathbf{r}_{T_m}, \mathbf{r}_n) + d(\mathbf{r}_{R_m}, \mathbf{r}_n)\right)} \tag{2.10}$$

Here $\xi(\mathbf{r}_n, \mathbf{r}_{R_m}, \mathbf{r}_{T_m}, k_m)$ is the filtering kernel expressed in the frequency wave-number domain, which has different expressions for different methods. For example, in the case of Kirchoff migration (KM), the filtering kernel becomes [4]

$$\xi(\mathbf{r}_n, \mathbf{r}_{R_m}, \mathbf{r}_{T_m}, k_m) = 4\frac{z_n - z_{R_m}}{d^2(\mathbf{r}_{R_m}, \mathbf{r}_n)}(1 + jk\, d(\mathbf{r}_{R_m}, \mathbf{r}_n))\frac{z_n - z_{T_m}}{d^2(\mathbf{r}_{T_m}, \mathbf{r}_n)}(1 + jk\, d(\mathbf{r}_{T_m}, \mathbf{r}_n)). \tag{2.11}$$

KM algorithm is an analytical approximate method that accounts for scattering from rough surfaces by using the Helmholtz–Kirchhoff integral theorem [11]. On the other hand, in the back-projection method, measured waves at the antenna plane are propagated back to the object plane by only using round-trip delays, which means there is no filtering kernel (i.e., $\xi = 1$) [4]. The adjoint of the forward operator is also in the form given in (2.10) where the filtering kernel is $\xi = \frac{1}{4\pi d(\mathbf{r}_{R_m}, \mathbf{r}_n) d(\mathbf{r}_{T_m}, \mathbf{r}_n)}$. Hence, the adjoint operation also back-propagates the measurements to the object domain with a scale inversely proportional to the radial distance of the antenna elements [12].

Although these methods offer low computational complexity, their reconstruction performance substantially degrades in ill-posed settings with limited and/or noisy data.

2.2.2 Regularized Reconstruction Methods

Regularization-based methods can yield more successful reconstructions than the traditional methods by incorporating prior information about the unknown 3D image

cube into the reconstruction process [13–18].

A systematic approach to regularization is to incorporate the prior knowledge about unknown solution in a deterministic or stochastic setting, and leads to a minimization with a regularization functionall, $\mathcal{R}(\cdot)$, penalizing the solutions that do not comply with the assumed prior information [13,17]. Under the white Gaussian noise assumption, we can formulate the inverse problem by using a regularization functional as follows

$$\hat{\mathbf{s}} = \arg\min_{\mathbf{s}} \mathcal{R}(\mathbf{s}) \text{ subject to } \|\mathbf{y} - \mathbf{A}\mathbf{s}\|_2 \leq \epsilon \tag{2.12}$$

where ϵ is a parameter that should be chosen based on the noise variance (i.e., $\sqrt{M \cdot \sigma_w^2}$).

A commonly used regularization functional is isotropic total-variation (TV) [49], which promotes sparsity on the gradient of the image vector, and is computed by,

$$\mathcal{R}_{TV}(\mathbf{s}) \triangleq \|\sqrt{|\mathbf{D}_x\mathbf{s}|^2 + |\mathbf{D}_y\mathbf{s}|^2 + |\mathbf{D}_z\mathbf{s}|^2}\|_1, \tag{2.13}$$

where $\mathbf{D}_x$,$\mathbf{D}_y$, and $\mathbf{D}_z$ are the matrices corresponding to discrete difference operators along x,y and z directions. Another widely used regularization functional is ℓ_1 norm ($\mathcal{R}_{\ell_1}(\mathbf{s}) \triangleq \|\mathbf{s}\|_1$) which promotes sparsity directly in the image domain. These commonly used sparsity priors are motivated by the compressed sensing theory [50] and are shown to offer promising imaging performance at various compressive imaging settings including radar imaging [15–18,36,38,39].

Although these methods offer promising imaging performance, utilizing handcrafted regularization functionals corresponds to exploiting simple image priors, where information about the target images is not learned. Hence, these priors are not adaptive to the distribution of data and do not accurately represent the distribution of target images. Moreover, sparsity-based reconstruction methods generally have iterative nature obtained using variable splitting methods such as alternating direction method of multipliers (ADMM) [51] and proximal gradient descent [49]. As a result, these methods often have longer computation time than the direct inversion approaches, which is undesirable in real-time applications.

For near-field radar imaging, existing regularized reconstruction methods generally enforce smoothness or sparsity on the complex-valued reflectivity distribution [17, 18,37,38]. These methods are therefore built on the assumption that the scene reflec-

tivity has locally correlated phase and magnitude. However, for many applications, the phase of the reflectivity at a particular point can be more accurately modeled as random and uncorrelated with the phase at other points [35,36,39]. This is because phase shift can occur when imaging rough surfaces and also at the air/target interface due to the electrical properties of materials [35]. It has been observed in various related SAR works that enforcing regularization only on the magnitude improves the performance compared to enforcing it directly on the complex-valued reflectivity [15, 33, 36, 39].

2.2.3 Deep Learning Based Reconstruction Methods

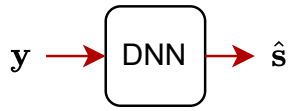
Deep learning based reconstruction methods can be grouped into two categories, these are learned direct inversion and learning-based regularized reconstruction approaches.

2.2.3.1 Learned Direct Inversion

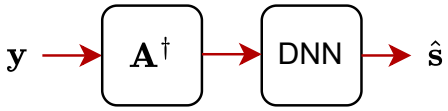
Recently, reconstruction techniques that exploit deep learning have emerged as alternatives to the traditional direct inversion methods [25]. These methods are shown to simultaneously achieve high reconstruction quality and low computational cost for various imaging problems [25–28].

Learning-based direct inversion methods are aimed to perform the reconstruction directly from the measurements using a deep neural network. Hence the neural network is trained to learn the direct mapping from the observations to the desired image solely using training data. Although these methods yield the state-of-the-art performance for simpler inverse problems like denoising [27], they can not provide successful results whenever the observation model is complex, the unknown image does not look alike observations, or there is not a lot of training data available. For this reason, commonly an efficient approximate inverse of the forward model is first applied analytically and these initial images are provided to the network as a warm start. Subsequently a deep neural network is employed to refine this initial reconstruction [25–28]. A representative block diagram of learned direct inversion methods are shown in Fig.2.2. An important advantage of learned direct inversion methods is their low computational complexity due to their feed-forward (non-iterative) nature, which is ideal for real-time

imaging applications.



(a) Learning-based direct inversion.



(b) An example of physics-based learned direct inversion.

Figure 2.2: A representative block diagram for learned direct inversion methods.

Most of the learned direct inversion methods in the radar imaging literature are for far-field settings in SAR/ISAR or MIMO radar imaging [20, 52–58]. In the near-field and MIMO radar imaging context, there are only a few works for learned direct reconstruction [59, 60] from which only the work in [60] is developed for imaging 3D extended targets with random phase.

2.2.4 Learning-Based Regularized Reconstruction

In recent years, plug-and-play (PnP) regularization became one of the most widely used frameworks for solving computational imaging problems [28, 30, 32, 33]. The key idea behind plug-and-play regularization is to substitute an off-the-shelf denoiser in place of the proximal operator associated with regularization in an iterative reconstruction framework such as ADMM. This enables the utilization of powerful priors without explicitly specifying regularization functionals, where solving the inverse problem becomes equivalent to finding a consensus equilibrium for data-fidelity and image-prior updates [30, 61]. A representative block diagram for the PnP methods is shown in Fig. 2.3.

Initial works on PnP regularization methods relied on analytical priors by the use of handcrafted denoisers such as BM3D [34, 37, 62]. However, with the rise of deep learning and the observation that DNNs provide superior denoising performance [28], there has been a growing interest in using learned PnP regularization approaches that exploit deep priors [30, 32, 33] which provide state of the art performance.

To the best of our knowledge, there are no deep pior-based PnP approaches developed

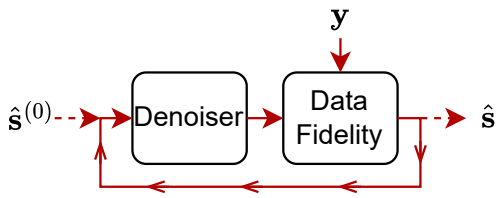


Figure 2.3: A representative block diagram of plug-and-play regularization. Within an iterative optimization framework, the plugged-in denoiser replaces the proximal mapping associated with regularization, whereas the data-fidelity block enforces the fidelity of the reconstructions to the measurements.

for 3D near-field radar imaging. A related PnP work in SAR imaging [37] utilizes 2D analytical (but not deep) denoising priors to reconstruct 3D extended targets. This method also does not take into account the random phase nature of the reflectivity values and considers regularization on the complex-valued reflectivity. On the other hand, another work from the SAR literature develops a deep prior-based PnP method for magnitude regularization [33]. Nonetheless, this method requires an inefficient computation to update the phase of the solutions and is developed for 2D far-field imaging.

Another commonly utilized framework for learning-based regularized reconstruction is algorithm unrolling, where a model-based iterative method is unrolled into an end-to-end trainable network [28, 29, 31, 63]. Algorithm unrolling can be used for both regularization with analytical priors, such as ℓ_1 regularization [64], or with regularization schemes that exploit deep priors, such as PnP regularization with deep denoiser prior [31]. Although both learned PnP- and unrolling-based approaches can yield state-of-the-art reconstruction quality, PnP methods have the advantage of adaptability to different imaging settings and significantly less training time.

Despite the recent success of the methods that exploit DNNs within a model-based framework, most of these approaches have been developed for 2D or real-valued image reconstruction problems [28, 30, 32–34]. Furthermore, there is no study on such methods for near-field radar imaging where we encounter a 3D complex-valued image reconstruction problem.

CHAPTER 3

PLUG-AND-PLAY REGULARIZATION ON MAGNITUDE WITH DEEP PRIORS

3.1 Introduction

In this chapter, we develop a novel and efficient PnP method for reconstructing the 3D complex-valued reflectivity distribution of the near-field scene from sparse MIMO measurements. Due to the random phase nature of the scene reflectivities in various applications, we formulate the image formation problem by exploiting regularization on the magnitude of the reflectivity function. We provide a general expression for the proximal mapping associated with such regularization functionals operating on the magnitude. By utilizing this expression, we develop a computationally efficient PnP reconstruction method that consists of simple update steps. To utilize within the developed PnP framework, we also train a 3D deep denoiser that can jointly exploit range and cross-range correlations. The source codes of this developed approach are available at https://github.com/METU-SPACE-Lab/PnP-Regularization-on-Magnitude.

Our approach provides a unified PnP framework to effectively handle arbitrary regularization on the magnitude of a complex-valued unknown, which appears to be missing in the previous related radar imaging works [15, 33]. The effectiveness of the developed learning-based PnP approach is illustrated in microwave imaging under various compressive and noisy observation scenarios using both simulated data and experimental measurements. We also compare the performance with the commonly used traditional methods (back-projection and Kirchhoff migration), and with the sparsity-based approaches involving ℓ_1 and TV regularization.

Some parts of this chapter have been presented in [65] and published in [66].

Compared to the earlier works in near-field MIMO radar imaging, the developed technique not only provides state-of-the-art reconstruction performance for 3D real-world targets, but also enables fast computation. In particular, compared to the traditional direct inversion methods and sparsity-based approaches, the developed reconstruction technique achieves the best reconstruction quality at compressive settings with both simulated and experimental data.

Despite the recent success of the PnP that exploit deep priors, most of these approaches have been developed for 2D or real-valued image reconstruction problems [28, 30, 32–34]. Furthermore, there is no study on such methods for near-field radar imaging where we encounter a 3D complex-valued image reconstruction problem. To the best of our knowledge, our approach is the first deep prior-based PnP approach developed for near-field radar imaging. A related PnP work in SAR imaging [37] utilizes 2D analytical (but not deep) denoising priors to reconstruct 3D extended targets. This approach also considers regularization on the complex-valued reflectivity. Differently, our approach exploits regularization on the magnitude of the reflectivity due to its random phase nature in various applications. There is also a related learned PnP approach with magnitude regularization which has been developed for 2D (far-field) SAR imaging [33]. However, this method requires an inefficient iterative computation to update the phase. In contrast, our approach does not have a phase update step and all the other update steps are simple and efficient to compute thanks to the closed-form expression used for the proximal mapping. The presented closed-form expression for the proximal mapping associated with arbitrary regularization on the magnitude also provides a generalization of the proximal mappings associated with TV and ℓ_1 regularization on magnitude [15]. Hence our PnP framework provides a generalizable and powerful means for effectively enforcing arbitrary regularization on magnitude, and is equally applicable to other radar image formation problems (including SAR).

Different than the related learning-based works in near-field MIMO radar imaging [59, 60,64,67], our approach is a deep prior-based PnP approach developed for imaging 3D extended targets. In particular, the works in [59, 60, 67] present deep learning-based non-iterative reconstruction methods by refining an initial analytical reconstruction using DNNs. Other learning-based work in [64] develops an unrolling-based method. But unlike our approach, this method is not DNN-based (i.e. not deep prior-based) and

only learns the hyperparameters (such as soft threshold and regularization parameters) of the unrolled ℓ_1 regularization-based reconstruction algorithm.

The main contributions of this chapter can be summarized as follows:

- Providing a unified PnP framework to effectively handle arbitrary regularization on the magnitude of a complex-valued unknown (involving random phase),
- Development of a novel deep learning-based plug-and-play reconstruction method for 3D complex-valued imaging with application to near-field MIMO radar imaging,
- Comprehensive experiments on synthetic 3D scenes with quantitative and qualitative analysis by considering various compressive and noisy observation scenarios,
- Performance evaluation with experimental measurements to demonstrate reconstruction of 3D real-world targets, and comparison with the commonly used direct inversion and regularized reconstruction methods.

The chapter is organized as follows. In Section 3.2 we formulate the inverse problem by enforcing regularization on the magnitude and then develop our plug-and-play approach. The architecture of the deep denoiser utilized for learned PnP reconstruction is also presented here. Section 3.3 presents the imaging results for various compressive and noisy observation scenarios. The details of the simulated and experimental settings considered, and the training procedure are also presented here. We conclude the chapter by providing final remarks in Section 3.4.

3.2 Developed Plug-and-Play Reconstruction Approach

In this section, we first formulate the inverse problem by enforcing regularization on the magnitude and then develop our plug-and-play approach using the ADMM framework. The architecture of the 3D deep denoiser utilized for learned PnP reconstruction is also presented here.

3.2.1 Inverse Problem

Due to the random phase nature of the unknown scene reflectivities, we formulate the inverse problem of reconstructing the scene reflectivity values, $\mathbf{s}$, from the sparse MIMO measurements, $\mathbf{y}$, using a regularization functional, $\mathcal{R}(|\cdot|)$, that only operates on the magnitude:

$$\min_{s} \mathcal{R}(|\mathbf{s}|) \text{ subject to } \|\mathbf{y} - \mathbf{A}\mathbf{s}\|_2 \leq \epsilon \tag{3.1}$$

where ϵ is a parameter that should be chosen based on the noise variance (i.e. $\sqrt{M \cdot \sigma_w^2}$), and $|\mathbf{s}|$ denotes the magnitude of the reflectivity vector $\mathbf{s}$.

3.2.2 Variable Splitting and ADMM

To solve this regularized inverse problem, we first convert the constrained problem in (3.1) to an unconstrained one using the penalty function, $\iota_{\|\mathbf{y}-\mathbf{v}_1\|_2 \leq \epsilon}(.)$, and then apply variable splitting as follows:

$$\begin{aligned} &\min_{s,\mathbf{v}_1,\mathbf{v}_2} \left(\iota_{\|\mathbf{y}-\mathbf{v}_1\|_2 \leq \epsilon}(\mathbf{v}_1) + \mathcal{R}(|\mathbf{v}_2|)\right) \\ &\text{subject to } \mathbf{A}\mathbf{s} - \mathbf{v}_1 = 0\ ,\ \mathbf{s} - \mathbf{v}_2 = 0 \end{aligned} \tag{3.2}$$

Here the indicator function $\iota_{\|\mathbf{y}-\mathbf{v}_1\|_2 \leq \epsilon}(\mathbf{v}_1)$ takes value 0 if the constraint in (3.1) is satisfied and $+\infty$ otherwise, whereas $\mathbf{v}_1$, $\mathbf{v}_2$ are the auxiliary variables.

We solve the optimization problem in (3.2) with the C-SALSA approach [68]. In the corresponding ADMM framework [51], we first obtain the associated augmented Lagrangian form given by

$$\begin{aligned} &\mathcal{L}_{\rho_1,\rho_2}(\mathbf{s}, \mathbf{v}_1, \mathbf{v}_2, \mathbf{d}_1, \mathbf{d}_2) \\ &\quad = \iota_{\|\mathbf{y}-\mathbf{v}_1\|_2 \leq \epsilon}(\mathbf{v}_1) + \frac{\rho_1}{2}\|\mathbf{A}\mathbf{s} - \mathbf{v}_1 - \mathbf{d}_1\|_2^2 - \frac{\rho_1}{2}\|\mathbf{d}_1\|_2^2 \\ &\qquad + \mathcal{R}(|\mathbf{v}_2|) + \frac{\rho_2}{2}\|\mathbf{s} - \mathbf{v}_2 - \mathbf{d}_2\|_2^2 - \frac{\rho_2}{2}\|\mathbf{d}_2\|_2^2 \end{aligned} \tag{3.3}$$

Here $\mathbf{d}_1$, $\mathbf{d}_2$ denote the dual variables for $\mathbf{A}\mathbf{s}$ and $\mathbf{s}$, and $\rho_1, \rho_2 \in \mathbb{R}^+$ are the penalty parameters for the auxiliary variables $\mathbf{v}_1$ and $\mathbf{v}_2$. We then alternatively minimize this augmented Lagrangian function over $\mathbf{s}$, $\mathbf{v}_1$, and $\mathbf{v}_2$ to obtain the update steps for these variables.

Firstly, the minimization over $\mathbf{s}$ corresponds to solving a least-squares problem with the following normal equation:

$$(\mathbf{A}^H\mathbf{A} + \kappa\mathbf{I})\mathbf{s}^{l+1} = \mathbf{A}^H(\mathbf{v}_1^l + \mathbf{d}_1^l) + \kappa(\mathbf{v}_2^l + \mathbf{d}_2^l) \tag{3.4}$$

where the superscript l is the iteration count, and $\kappa \triangleq \frac{\rho_1}{\rho_2}$ is a hyper-parameter that needs to be adjusted. Since solving this normal equation using matrix inversion is impractical due to the large size, we instead use few conjugate-gradient (CG) iterations to update the scene reflectivity $\mathbf{s}$.

Secondly, the minimization over $\mathbf{v}_1$ corresponds to the proximal operator of the penalty function $\iota_{\|\mathbf{y}-\mathbf{v}_1\|_2 \leq \epsilon}(\cdot)$, which can be computed as the projection of $\mathbf{A}\mathbf{s}^{l+1} - \mathbf{d}_1^l$ onto ϵ-radius hyper-sphere with center $\mathbf{y}$ as follows:

$$\mathbf{v}_1^{l+1} = \mathbf{y} + \begin{cases} \epsilon \frac{\mathbf{A}\mathbf{s}^{l+1} - \mathbf{d}_1^l - \mathbf{y}}{\|\mathbf{A}\mathbf{s}^{l+1} - \mathbf{d}_1^l - \mathbf{y}\|_2}, & \text{if } \|\mathbf{A}\mathbf{s}^{l+1} - \mathbf{d}_1^l - \mathbf{y}\|_2 > \epsilon \\ \mathbf{A}\mathbf{s}^{l+1} - \mathbf{d}_1^l - \mathbf{y}, & \text{if } \|\mathbf{A}\mathbf{s}^{l+1} - \mathbf{d}_1^l - \mathbf{y}\|_2 \leq \epsilon \end{cases} \tag{3.5}$$

Lastly, the minimization over $\mathbf{v}_2$ corresponds to the proximal operator for the regularization function, $\mathcal{R}(|\cdot|)$, that operates on the magnitude of the complex-valued vector $\mathbf{v}_2$:

$$\mathbf{v}_2^{l+1} = \mathbf{\Psi}_{\alpha\mathcal{R}(|\cdot|)}(\mathbf{s}^{l+1} - \mathbf{d}_2^l) \tag{3.6}$$

where $\mathbf{\Psi}_{\alpha\mathcal{R}(|\cdot|)}$ is the respective proximal operator given by

$$\mathbf{\Psi}_{\alpha\mathcal{R}(|\cdot|)}(\mathbf{p}) \triangleq \arg\min_{\mathbf{v}} \left(\alpha\mathcal{R}(|\mathbf{v}|) + \frac{1}{2}\|\mathbf{v} - \mathbf{p}\|_2^2 \right) \tag{3.7}$$

for a complex-valued vector $\mathbf{p}$, with $\alpha \triangleq \frac{1}{\rho_2}$ determining the amount of regularization. This update step corresponds to solving a denoising problem for a complex-valued unknown, $\mathbf{v}$, with regularization enforced on its magnitude and noisy observation given as $\mathbf{p}$. To develop a computationally efficient PnP reconstruction method that consists of simple update steps, we provide a general expression for the solution of this denoising problem (equivalently, for the proximal operator in (3.7)). This will enable us to effectively handle arbitrary regularization on the magnitude, which appears to be missing in the previous radar imaging works.

3.2.3 Denoising with Regularization on Magnitude

In this section, we provide a general expression for the solution of the complex-valued denoising problem in (3.7) which involves regularization on the magnitude. For this, we first express each complex-valued vector as a product of a diagonal phase matrix and a magnitude vector as follows:

$$\mathbf{v} = \mathbf{\Phi}_{\mathbf{v}}|\mathbf{v}|, \quad \mathbf{p} = \mathbf{\Phi}_{\mathbf{p}}|\mathbf{p}|, \tag{3.8}$$

where $\mathbf{\Phi}_{\mathbf{v}} = \text{diag}(e^{j\angle\mathbf{v}})$ and $\mathbf{\Phi}_{\mathbf{p}} = \text{diag}(e^{j\angle\mathbf{p}})$ are complex-valued unitary matrices that contain the phase of the vectors $\mathbf{v}$ and $\mathbf{p}$ on their diagonals, respectively, whereas $|\mathbf{v}|$ and $|\mathbf{p}|$ represent real-valued and non-negative vectors that contain the respective magnitudes. By using these expressions, the optimization problem in (3.7) can be viewed as a joint minimization over the magnitude and phase of $\mathbf{v}$:

$$\min_{|\mathbf{v}|,\,\angle\mathbf{v}} \left(\alpha\mathcal{R}(|\mathbf{v}|) + \frac{1}{2}\|\mathbf{\Phi}_{\mathbf{v}}|\mathbf{v}| - \mathbf{\Phi}_{\mathbf{p}}|\mathbf{p}|\|_2^2 \right) \tag{3.9}$$

This joint minimization problem is equivalent to

$$\min_{|\mathbf{v}|} \left(\min_{\angle\mathbf{v}} \left(\alpha\mathcal{R}(|\mathbf{v}|) + \frac{1}{2}\|\mathbf{\Phi}_{\mathbf{v}}|\mathbf{v}| - \mathbf{\Phi}_{\mathbf{p}}|\mathbf{p}|\|_2^2 \right) \right) \tag{3.10a}$$

$$\equiv \min_{|\mathbf{v}|} \left(\alpha\mathcal{R}(|\mathbf{v}|) + \min_{\angle\mathbf{v}} \left(\frac{1}{2}\|\mathbf{\Phi}_{\mathbf{v}}|\mathbf{v}| - \mathbf{\Phi}_{\mathbf{p}}|\mathbf{p}|\|_2^2 \right) \right) \tag{3.10b}$$

Hence to solve this complex-valued denoising problem, our strategy is to first solve the minimization over the phase, $\angle\mathbf{v}$, in closed-form, and then by substituting the optimal phase solution, $\angle\hat{\mathbf{v}}$, to the above cost function, to solve the remaining minimization over the magnitude, $|\mathbf{v}|$.

For minimization over the phase, we have

$$\angle\hat{\mathbf{v}} = \arg\min_{\angle\mathbf{v}} \left(\frac{1}{2}\|\mathbf{\Phi}_{\mathbf{v}}|\mathbf{v}| - \mathbf{\Phi}_{\mathbf{p}}|\mathbf{p}|\|_2^2 \right) \tag{3.11}$$

$$= \arg\min_{\angle\mathbf{v}} \left(\frac{1}{2}\|\mathbf{\Phi}_{\mathbf{p}}^H\mathbf{\Phi}_{\mathbf{v}}|\mathbf{v}| - |\mathbf{p}|\|_2^2 \right) \tag{3.12}$$

where the last expression follows from the unitary property of the phase matrices, i.e. $\mathbf{\Phi}_{\mathbf{p}}^H = \mathbf{\Phi}_{\mathbf{p}}^{-1}$. After expanding the ℓ_2 norm expression and simplifying it using the unitary property of phase matrices and omitting the terms that do not depend on the phase $\angle\mathbf{v}$, we obtain

$$\angle\hat{\mathbf{v}} = \arg\max_{\angle\mathbf{v}} \left(\frac{1}{2} \left(|\mathbf{v}|^T\mathbf{\Phi}_{\mathbf{p}}\mathbf{\Phi}_{\mathbf{v}}^*|\mathbf{p}| + |\mathbf{p}|^T\mathbf{\Phi}_{\mathbf{p}}^*\mathbf{\Phi}_{\mathbf{v}}|\mathbf{v}| \right) \right) \tag{3.13}$$

Here we also use the fact that $|\mathbf{v}|$ and $|\mathbf{p}|$ are real-valued and hence their Hermitian transpose is simply equal their transpose, and since phase matrices are diagonal, their Hermitian transpose is simply equal their conjugation. Using the diagonality of the phase matrices, this further simplifies to

$$\angle\hat{\mathbf{v}} = \arg\max_{\angle\mathbf{v}} \left(|\mathbf{p}|^T \mathfrak{Re}\{\mathbf{\Phi}_\mathbf{p}^* \mathbf{\Phi}_\mathbf{v}\} |\mathbf{v}| \right). \tag{3.14}$$

Hence to find the optimal phase, we need to maximize $\sum_{n=1}^{N} |p_n||v_n| \cos(\angle v_n - \angle p_n)$ over all elements $\angle v_n$ of the vector $\angle\mathbf{v}$. Since each term in this summation contains only one element of $\angle\mathbf{v}$, maximization can be decoupled for each element, which yields $\angle p_n$ as the optimal value of $\angle v_n$. This shows that the optimal phase, $\angle\hat{\mathbf{v}}$, for the denoising problem in (3.9) is equal to the phase of the given noisy observation $\mathbf{p}$:

$$\angle\hat{\mathbf{v}} = \angle\mathbf{p}. \tag{3.15}$$

That is, the proximal mapping of a function that operates on the magnitude of a complex-valued vector must directly pass the phase values of the proximal point.

After solving the minimization over the phase in closed-form, we now substitute the optimal phase solution, $\angle\hat{\mathbf{v}}$, to the cost function in (3.10b) and consider the remaining minimization over the magnitude, $|\mathbf{v}|$:

$$|\hat{\mathbf{v}}| = \arg\min_{|\mathbf{v}|} \left(\alpha\mathcal{R}(|\mathbf{v}|) + \frac{1}{2} \||\mathbf{v}| - |\mathbf{p}|\|_2^2 \right) \tag{3.16}$$

where we use the unitary property of the phase matrix $\mathbf{\Phi}_\mathbf{p}$ as before. Note that this expression is equivalent to the Moreau proximal mapping, $\mathbf{\Psi}_{\alpha\mathcal{R}(\cdot)}$, associated with the regularization function $\mathcal{R}(\cdot)$ and applied on the magnitude $|\mathbf{p}|$. Hence the optimal magnitude $|\hat{\mathbf{v}}|$ for the denoising problem in (3.9) corresponds to denoising of the magnitude of the noisy observation $\mathbf{p}$ with noise variance α:

$$|\hat{\mathbf{v}}| = \mathbf{\Psi}_{\alpha\mathcal{R}(\cdot)}(|\mathbf{p}|). \tag{3.17}$$

For the scalar-valued case, a similar derivation is encountered in [69].

Therefore, the solution of the complex-valued denoising problem in (3.7) with magnitude regularization can be computed as

$$\mathbf{\Psi}_{\alpha\mathcal{R}(|\cdot|)}(\mathbf{p}) = e^{j\angle\mathbf{p}} \odot \mathbf{\Psi}_{\alpha\mathcal{R}(\cdot)}(|\mathbf{p}|), \tag{3.18}$$

where $\odot$ denotes element-wise multiplication. This corresponds to denoising the magnitude of $\mathbf{p}$ using the proximal (denoising) operator $\mathbf{\Psi}_{\alpha\mathcal{R}(\cdot)}$ and merging the denoised magnitude with the unprocessed phase of $\mathbf{p}$. Since (3.18) decouples the magnitude and phase solutions, it enables us to use real-valued denoisers (proximal operators) $\mathbf{\Psi}_{\alpha\mathcal{R}(\cdot)}$ for the solution of the complex-valued denoising problem (in (3.7)).

3.2.4 Algorithmic Summary of the Developed Method

The steps of the developed PnP method are summarized in Algorithm 1 and illustrated in Fig. 3.1. Each iteration of the algorithm mainly consists of four computationally efficient update steps. The first step is the update of the image $\mathbf{s}$ as given in line 4 and carried out using few CG iterations. The second step is the update of the auxiliary variable $\mathbf{v}_1$ by computing the projection given in line 6 and efficiently computed using scaling operations. The third step is the complex-valued denoising step given in line 7 to update the auxiliary variable $\mathbf{v}_2$. As shown, this denoising is equivalent to directly passing the phase but denoising the magnitude of $\mathbf{s}^{l+1} - \mathbf{d}_2^l$ using the proximal operator $\mathbf{\Psi}_{\alpha\mathcal{R}(\cdot)}$. To exploit data-driven deep priors, we use a trained denoiser as proximal operator, as explained in the next section. The last steps are the dual-updates given in lines 8 and 9.

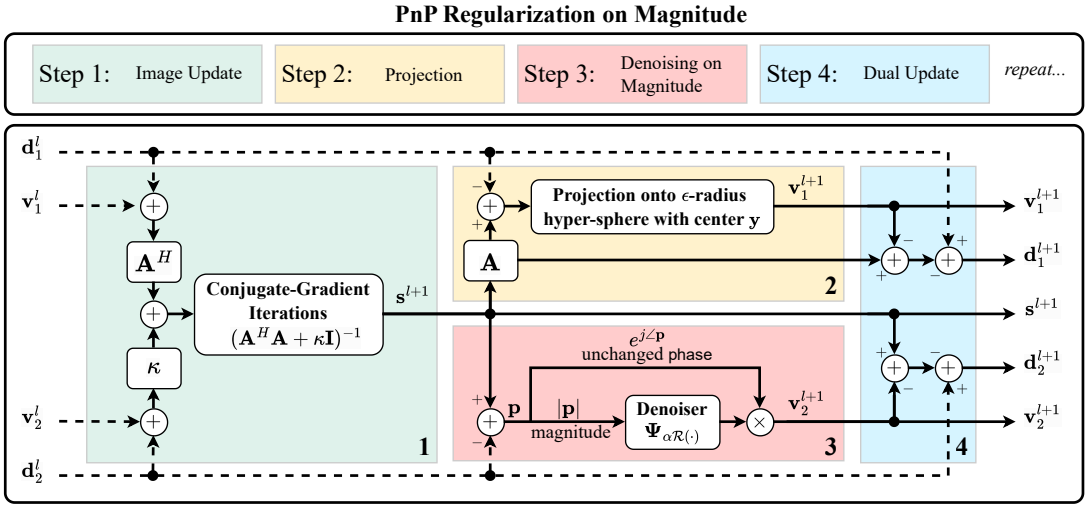


Figure 3.1: Developed PnP Method for Complex-valued Reconstruction with Regularization on Magnitude.

Algorithm 1: PnP Regularization on Magnitude for Complex-Valued Reconstruction

inputs: $\mathbf{\Psi}_{\alpha\mathcal{R}(\cdot)}$, $\mathbf{y}$, $\mathbf{A}$, $\mathbf{s}^0, \mathbf{v}_2^0, \mathbf{v}_1^0$, $\epsilon > 0$, $\kappa > 0$, $\alpha > 0$

$\mathbf{d}_1^0, \mathbf{d}_2^0 \leftarrow \mathbf{0}$, $l \leftarrow 0$

repeat

$$\mathbf{s}^{l+1} = (\mathbf{A}^H\mathbf{A} + \kappa\mathbf{I})^{-1}(\mathbf{A}^H(\mathbf{v}_1^l + \mathbf{d}_1^l) + \kappa(\mathbf{v}_2^l + \mathbf{d}_2^l))$$

$$\mathbf{u}^l = \mathbf{A}\mathbf{s}^{l+1} - \mathbf{d}_1^l$$

$$\mathbf{v}_1^{l+1} = \mathbf{y} + \begin{cases} \epsilon\frac{\mathbf{u}^l - \mathbf{y}}{\|\mathbf{u}^l - \mathbf{y}\|_2}, & \text{if } \|\mathbf{u}^l - \mathbf{y}\|_2 > \epsilon \\ \mathbf{u}^l - \mathbf{y}, & \text{if } \|\mathbf{u}^l - \mathbf{y}\|_2 \leq \epsilon \end{cases}$$

$$\mathbf{v}_2^{l+1} = e^{j\angle(\mathbf{s}^{l+1} - \mathbf{d}_2^l)} \odot \mathbf{\Psi}_{\alpha\mathcal{R}(\cdot)}(|\mathbf{s}^{l+1} - \mathbf{d}_2^l|)$$

$$\mathbf{d}_1^{l+1} = \mathbf{d}_1^l - (\mathbf{A}\mathbf{s}^{l+1} - \mathbf{v}_1^{l+1})$$

$$\mathbf{d}_2^{l+1} = \mathbf{d}_2^l - (\mathbf{s}^{l+1} - \mathbf{v}_2^{l+1})$$

$l \leftarrow l + 1$

until *some stopping criterion is satisfied*;

output: $\mathbf{s}^l$

Note that our development is implicit about the choice of the regularizer ($\mathcal{R}(|\cdot|)$) and the related proximal operator ($\mathbf{\Psi}_{\alpha\mathcal{R}(\cdot)}$). Therefore, we can efficiently adopt plug-and-play framework, which enables the utilization of powerful priors, such as deep denoisers, in place of the proximal operator, without explicitly specifying the regularizer.

Moreover, our PnP approach provides a generalizable and powerful means for efficiently handling arbitrary regularization on the magnitude of a complex-valued unknown. Our approach is applicable with any forward model matrix $\mathbf{A}$, and hence can be used for other complex-valued image formation problems including SAR reconstruction.

3.2.5 3D Deep Denoiser for Learned PnP Regularization

Following the success of convolutional neural networks (CNN) on denoising [28, 32, 70], we train and deploy a deep CNN-based denoiser for the third step of our PnP

approach. Our denoiser is a 3D U-net developed based on the 2D U-net architecture in [71] and is shown in Fig. 3.2. To be able to effectively handle a wide range of noise levels, our denoiser is designed for non-blind Gaussian denoising similar to [32], and hence takes as input also the noise level. This non-blind denoiser replaces the proximal operator $\mathbf{\Psi}_{\alpha\mathcal{R}(\cdot)}$ in line 7 of the Algorithm 1, which is used to denoise the input magnitudes.

The proposed denoiser is a 3-level encoder-decoder architecture with repeated 3D convolutional blocks (C) followed by batch normalization (B) and ReLU (R). Due to 3D processing, the denoiser can jointly exploit range and cross-range correlations. On each level, max pooling (Max. Pool.) is used to reduce the spatial size of the input tensor by a factor of 2 in each dimension and transposed convolution blocks (T.Conv.) are used to increase by 2. At each decoding level, the output of the transposed convolution block is concatenated with the encoder outputs. The concatenated outputs are then fed to the respective decoding blocks. A single-channel 3D convolution block follows the last decoding block. The number of output channels of all convolutional blocks is indicated inside parentheses in Figure 3.2.

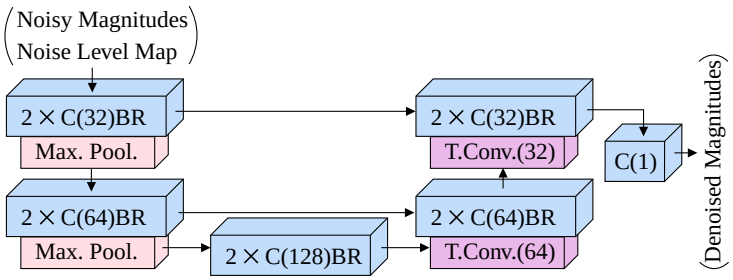


Figure 3.2: Network architecture of the proposed 3D deep denoiser. "C", "B", "R", "Max. Pool." and "T.Conv." represent 3D convolution, batch normalization, ReLU activation, max-pooling operation, and transposed convolution, respectively. The number of output channels is denoted inside parentheses. Total number of parameters is 1,357,505.

The input of our U-net is the 3D reflectivity magnitude that will be denoised and the 3D noise level map. The noise level map enables to adjust the amount of denoising in our non-blind denoiser network and its values are set to the constant $\sqrt{\alpha}$ in (3.16). The output of the U-net is the 3D denoised reflectivity magnitude.

3.3 Experiments and Results

We now demonstrate the effectiveness of the developed learning-based PnP approach under various compressive and noisy observation scenarios in microwave imaging. For this, we first train the implemented denoiser using a synthetically generated large dataset consisting of 3D extended targets. We then perform comprehensive experiments on synthetic 3D scenes, and comparatively evaluate the performance with the widely used back-projection (BP) and Kirchhoff migration (KM) algorithms, as well as using sparsity-based regularization in the form of isotropic total-variation (TV) and ℓ_1. Lastly, we illustrate the performance with experimental measurements to demonstrate the successful reconstruction of 3D real-world targets.

3.3.1 Training of the 3D Deep Denoiser

Because a large experimental dataset is not available for microwave imaging, we use a synthetic dataset [60] to train our denoiser network. The utilized synthetic dataset consists of randomly generated complex-valued image cubes of size $25 \times 25 \times 49$. We use 800 image cubes for training, 100 image cubes for testing, and another 100 image cubes for validation. Each synthetic image cube is obtained by randomly generating 15 points within the cube and then applying a 3D Gaussian filter to convert these points to a volumetric object. The magnitudes are normalized (via sigmoid function) to 1, while adding a random phase to each image voxel from a uniform distribution between 0 and 2π.

The denoiser network replaces the proximal operator $\mathbf{\Psi}_{\alpha\mathcal{R}(\cdot)}$ in line 7 of the Algorithm 1 with the goal of denoising the reflectivity magnitudes. We accordingly train our deep denoiser by minimizing the mean squared error between the 3D ground truth magnitudes and Gaussian noise added magnitudes on 800 training scenes. At each iteration of training, a new Gaussian noise realization is added to each ground truth magnitude by randomly and uniformly choosing the noise standard deviation, σ_v, from the interval $[0, 0.2]$. In addition, the constant noise level map is formed using this value for noise standard deviation, i.e. $\sqrt{\alpha} = \sigma_v$, and concatenated to the 3D noisy magnitude. As a result, the network learns to denoise the reflectivity magnitudes in a non-blind

manner.

For training, we use a batch size of 16 with the maximum number of epochs set as 2000. We utilize Adam optimizer [72] with an initial learning rate of 10^{-3}, and drop the learning rate by a factor of 10 if the validation loss does not improve for 25 epochs. We stop the training when the validation loss does not improve for 50 epochs. At the end of training, we use the network weights that provide the minimum validation loss. Training converges approximately 15 minutes on NVIDIA GeForce RTX 3080 Ti GPU using PyTorch 1.12.0 with CUDA Toolkit 11.6.0 in Python 3.10.6.

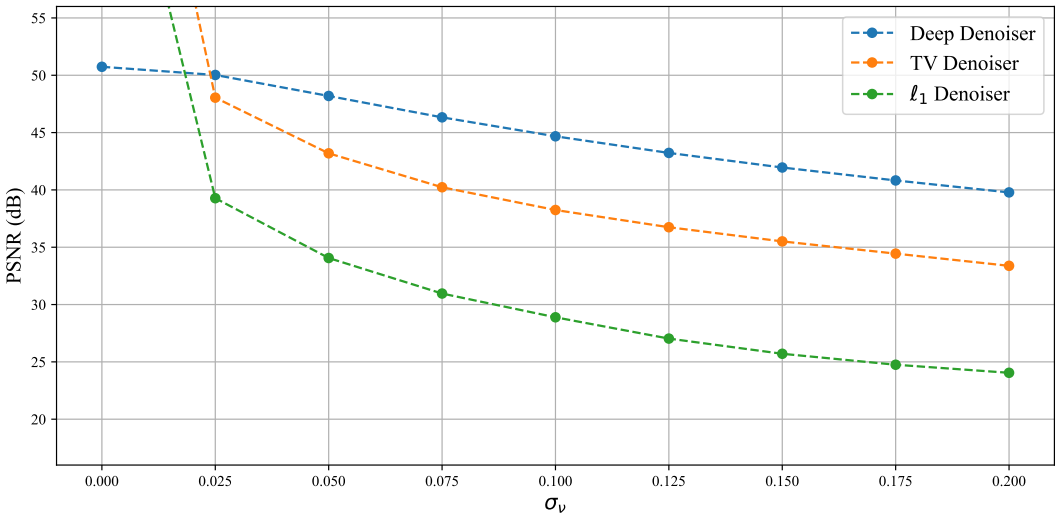
To analyze the performance of our learning-based PnP approach, we use the same trained denoiser without any modification for both simulated and experimental data.

We now present the denoising performance of the trained DNN in comparison with the other denoising approaches (ℓ_1 and TV regularization). The average 3D peak signal-to-noise ratio (PSNR) is computed using 100 test images for different values of noise standard deviation σ_ν and provided in Fig. 3.3a. Sample denoised magnitudes are also shown in Fig. 3.3b-f. As seen, the deep denoiser significantly outperforms other methods.

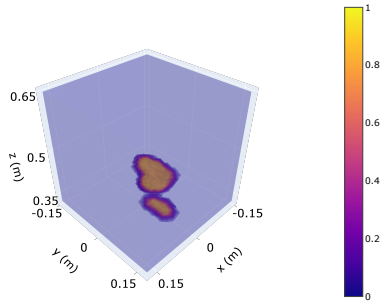
Data-driven DNN-based denoisers are currently the best choice for plug-and-play regularization because DNNs provide state-of-the-art performance for the denoising problem as demonstrated in various works in the literature [28]. In contrast to the existing analytical (hand-crafted) denoisers such as those based on l_1 and TV regularization, DNN-based denoisers are data-adaptive denoisers that learn how to remove the noise for the data of interest. Since the parameters of the deep denoiser are optimized based on the training data, prior information about the target images is learned. On the other hand, TV and ℓ_1 regularization functions are hand-crafted and correspond to much simpler priors.

3.3.2 Performance Analysis with Simulated Data

We first analyze the performance of the developed imaging technique at various noise and compression levels using the synthetic scenes in the test dataset. For this, we consider a microwave imaging setting similar to Fig. 2.1. The scene of interest has

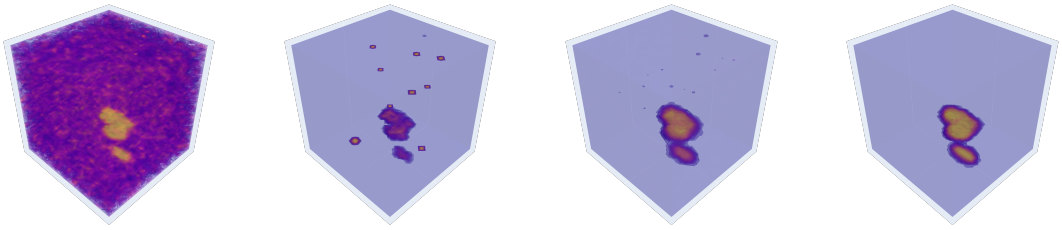


(a) Average Denoising PSNR



(b) Ground Truth

ℓ_1 Denoising TV Denoising Deep Denoiser



(c) Input Magnitudes (d) 23.66 dB (e) 33.45 dB (f) 39.85dB

Figure 3.3: Denoising performance of different methods; (a) average test PSNR with respect to noise standard deviation σ_ν, (b) ground truth magnitudes of the sample test image, (c) noisy input magnitudes at $\sigma_\nu = 0.2$, (d)-(f) denoised outputs corresponding to ℓ_1, TV and deep-prior based denoising and the respective PSNRs (dB).

physical dimension of 30 cm × 30 cm × 30 cm, and its center is located 50 cm away from the antenna array.

As MIMO array topology, commonly used Mill's Cross array [9] is utilized. The used planar array has a width of 0.3 m, and contains 12 transmit and 13 receive antennas, which are uniformly spaced on the diagonals in a cross configuration as shown in Fig. 3.4. The frequency, f, is swept between 4 GHz and 16 GHz with uniform steps.

In an uncompressed observation setting, the expected theoretical resolution [9] is 2.5 cm in the cross-range directions, x and y, and 1.25 cm in the down-range direction, z. With the goal of achieving these resolutions in the sparse case, we choose the image voxel size as 1.25 cm along x, y directions, and 0.625 cm along z direction (i.e. half of these resolutions). For the scene of interest, this results in an image cube of $25\times25\times49$ voxels, which is same as the size of the synthetic scenes generated. Using these synthetic image cubes with the forward model in (2.8), we simulate measurements at various signal-to-noise ratios (SNR $= 10\log_{10}(\frac{\|\mathbf{As}\|_2^2}{M\cdot\sigma_w^2})$) and compression levels (CL $= 1 - \frac{M}{N}$) for our analysis.

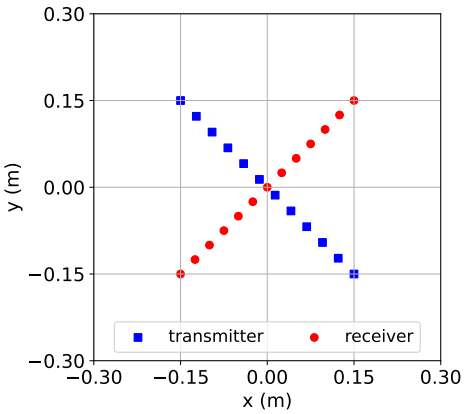


Figure 3.4: Mill's Cross Array.

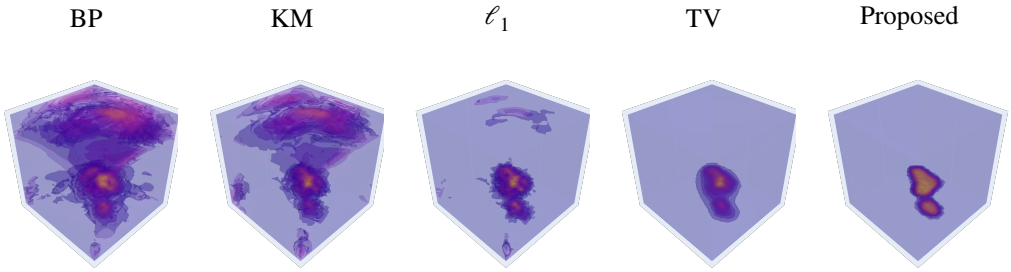
Before discussing the results, we provide the implementation details of the developed learning-based PnP approach, as well as the approaches used for comparison. For all regularization-based approaches, we enforce regularization on the reflectivity magnitudes and utilize the developed PnP approach in Algorithm 1 with different denoising (proximal update) steps. In particular, as the proximal operator, $\mathbf{\Psi}_{\alpha\mathcal{R}(\cdot)}$, we utilize soft-thresholding in the case of ℓ_1 regularization and 5 iterations of Chambolle algo-

rithm [73,74] in the case of TV regularization. Although there are methods in the literature to decide on the value ρ_2 (or equivalently α) adaptively, these methods introduce additional internal parameters to tune and can even negatively affect the convergence properties of the ADMM algorithm [51]. Here we choose the regularization parameter α in (3.18) by searching for its optimal value in the validation dataset between 10^{-5} and 10^{-1} in a coarse to fine fashion. For this, we first use logarithmic steps with scale factors of 10. Once this coarse search is completed, we perform a linear search between the two best-performing parameters to fine-tune the performance. In practice, since the parameter α determines the amount of regularization, the selection of its value is mostly affected by the SNR value for a spesific observation system. Hence, in a real-world setting, its value can be determined through a prior calibration step. Our results also showed that the performance of the tested methods is not highly sensitive to the value of this regularization parameter and stays relatively constant within a logarithmic scale.

We initialize each iterative algorithm with $\mathbf{s}^0 = \frac{\mathbf{A}^H\mathbf{y}}{\max(|\mathbf{A}^H\mathbf{y}|)}$, and in each $\mathbf{s}$-update-step, the conjugate gradient algorithm is run for 5 iterations. TV and ℓ_1-based approaches converge to a solution for a sufficiently large κ in (3.4). Accordingly, we choose $\kappa = 5 \cdot 10^4$ and run the iterations until the stopping criterion is satisfied, which is when the relative change $\frac{\||\mathbf{s}|^{l+1} - |\mathbf{s}|^l\|_2}{\||\mathbf{s}|^l\|_2}$ drops below $5 \cdot 10^{-4}$. Because the convergence of learned PnP is an ongoing area of research and is not always guaranteed [75,76], we limit the maximum number of iterations in the developed learning-based approach to 30. For the choice of κ, we search the optimal value using the validation dataset and set it as $\kappa = 5 \cdot 10^2$. Moreover, for real-world applications, automatic parameter selection strategies similar to [77] would be highly useful during the calibration process.

To comparatively evaluate the performance of the developed approach, we first consider the case with a medium SNR of 30 dB and a high compression level of 90%. This corresponds to using 20 frequency steps between 4 and 16 GHz and is equivalent to reconstructing the reflectivity cube with only 10% data. For the sample test image in Fig.3.3, the reconstructions obtained with different approaches are illustrated in Fig. 3.5 using the same colormap. To quantitatively evaluate the performance, we also provide 3D PSNR between the normalized reconstructed magnitudes, $\frac{|\hat{\mathbf{s}}|}{\max|\hat{\mathbf{s}}|}$, and the ground truth magnitudes, $|\mathbf{s}|$, which is calculated as $\text{PSNR} = 10\log_{10}\left(\frac{1}{\text{MSE}}\right)$ where

$\text{MSE} = \frac{1}{N}\| \, |\mathbf{s}| - \frac{|\hat{\mathbf{s}}|}{\max |\hat{\mathbf{s}}|}\|_2^2$ is the mean squared error. Although all algorithms reconstruct a complex-valued reflectivity distribution, the reconstructed phase is not used in this evaluation since it is random and does not contain any useful information. As seen in Fig. 3.5, the developed learning-based approach provides the best image quality with a reconstruction closely resembling the ground truth and achieving a PSNR of 30.12 dB. On the other hand, TV reconstruction suffers from over-smoothing, whereas ℓ_1 based reconstruction contains speckle-like artifacts and an artifact cluster at the top. The visual quality of KM and BP reconstructions are even worse with many more reconstruction artifacts due to noisy and compressed data, where KM performs slightly better than BP.



(a) 20.76 dB (b) 22.08 dB (c) 25.42 dB (d) 26.13 dB (e) 30.12 dB

Figure 3.5: Sample reconstructions with $\frac{M}{N} = 10\%$ data (i.e. 90% compression level) and 30 dB measurement SNR. Reconstructions obtained using different methods with their PSNR (dB) indicated underneath each figure. (Maximum projections along each dimension and 3D rotating views are available for all reconstructions at https://github.com/METU-SPACE-Lab/PnP-Regularization-on-Magnitude as video.)

To compare the reconstruction speed, average run-time of each method is computed over 100 test scenes as given in Table 3.1. As seen, the developed approach is capable of providing the best reconstruction quality with an average runtime of few seconds and is the fastest method after the direct inversion-based approaches (which largely fail). Moreover, TV and ℓ_1 regularized solutions take much longer time to compute.

Table 3.1: Average Run-Time on 100 Test Scenes at 30 dB SNR and with 10% data.

	BP	KM	ℓ_1	TV	Proposed
Δt	13.4 ms	13.4 ms	29.6 s	21.2 s	3.66 s

3.3.2.1 Compression Level Analysis

We now analyze the effect of the compression level on the performance for the 30 dB SNR case. We consider compression levels of 97.5%, 95%, 92.5%, 90%, 85% and 80%, which respectively correspond to using 5, 10, 15, 20, 30, and 40 frequency steps between 4 and 16 GHz, and are equivalent to reconstructing the reflectivity cube with 2.5%, 5%, 7.5%, 10%, 15% and 20% available data. Here the compression level of 97.5% is provided to show the breaking point of the proposed approach. For each case, the average PSNR is computed for the 100 test scenes reconstructed and is given in Table 3.2.

Table 3.2: Average PSNR on 100 Test Scenes for Different Amounts of Available Data at 30 dB Measurement SNR.

$\frac{M}{N}$	2.5%	5%	7.5%	10%	15%	20%
Back-Projection	16.41	19.75	21.71	23.49	24.56	24.60
Kirchhoff Migration	18.43	21.18	22.95	24.51	25.41	25.42
ℓ_1 Regularization	22.76	24.08	24.90	25.70	25.85	25.85
TV Regularization	19.20	22.26	24.18	26.26	26.45	26.46
Proposed Method	21.53	27.27	29.82	30.40	30.65	30.75

As seen from the table, the developed learning-based approach significantly outperforms the other approaches for all compression levels other than 97.5% (i.e. 2.5% data). In particular, the average PSNR exceeds 30 dB when we perform a reconstruction with 10% or higher data. It is also interesting to observe that the performance of the developed method at the 95% compression level (i.e. 5% data) with 27.27 dB PSNR is even better than the performance of all compared methods at the lowest compression level (i.e. 20% data). At the 97.5% compression level with only 2.5%

available data all methods fail to provide faithful reconstructions with PSNRs less than 23 dB, which suggests that the information provided by this amount of data is insufficient. As expected, all regularization-based approaches outperform the direct inversion methods (BP and KM), especially at highly compressive settings. Moreover, data-adaptive deep priors enable superior performance compared to hand-crafted analytical priors, TV, and ℓ_1. From these analytical priors, ℓ_1 starts to yield better performance than TV at the compression levels higher than 90% (i.e. with less than 10% data availability). From the direct inversion-based methods, KM consistently performs better than BP and approaches the performance of ℓ_1 regularization at the increased data availability rates. Because of this, from this point forward, we will omit the BP from the visual comparisons and only present the results of KM. In general, the performance of each method starts to increase slowly with the increased data availability rates beyond 15%. This suggests that the bottleneck on the measurement diversity becomes the sparse MIMO array topology when the number of frequency steps exceeds 30.

For visual comparison, sample reconstructions obtained with 2.5%, 5%, 10%, and 20% data are also given in Fig. 3.6. As seen, for all approaches, the reconstruction quality improves with the increasing amount of data. Moreover, we can observe that KM is the most severely affected method by the amount of available data, and at high compression levels its reconstruction suffers from large grating lobes. At compression levels corresponding to 5% and 2.5% data, TV reconstruction also contains large artifacts in addition to the over-smoothing effect. On the other hand, although the ℓ_1-based method suffers from speckle-like artifacts, its performance does not change much up until the highest compression level (corresponding to 2.5% data). For the highest compression level, we see that ℓ_1 reconstruction is point-like and not an extended target. On the other hand, although the PSNR of the developed method is less, it outputs an extended target that resembles the shape of the ground truth. After this breaking point for the compression level, the proposed learning-based PnP method yields almost artifact-free reconstructions for all other compression levels.

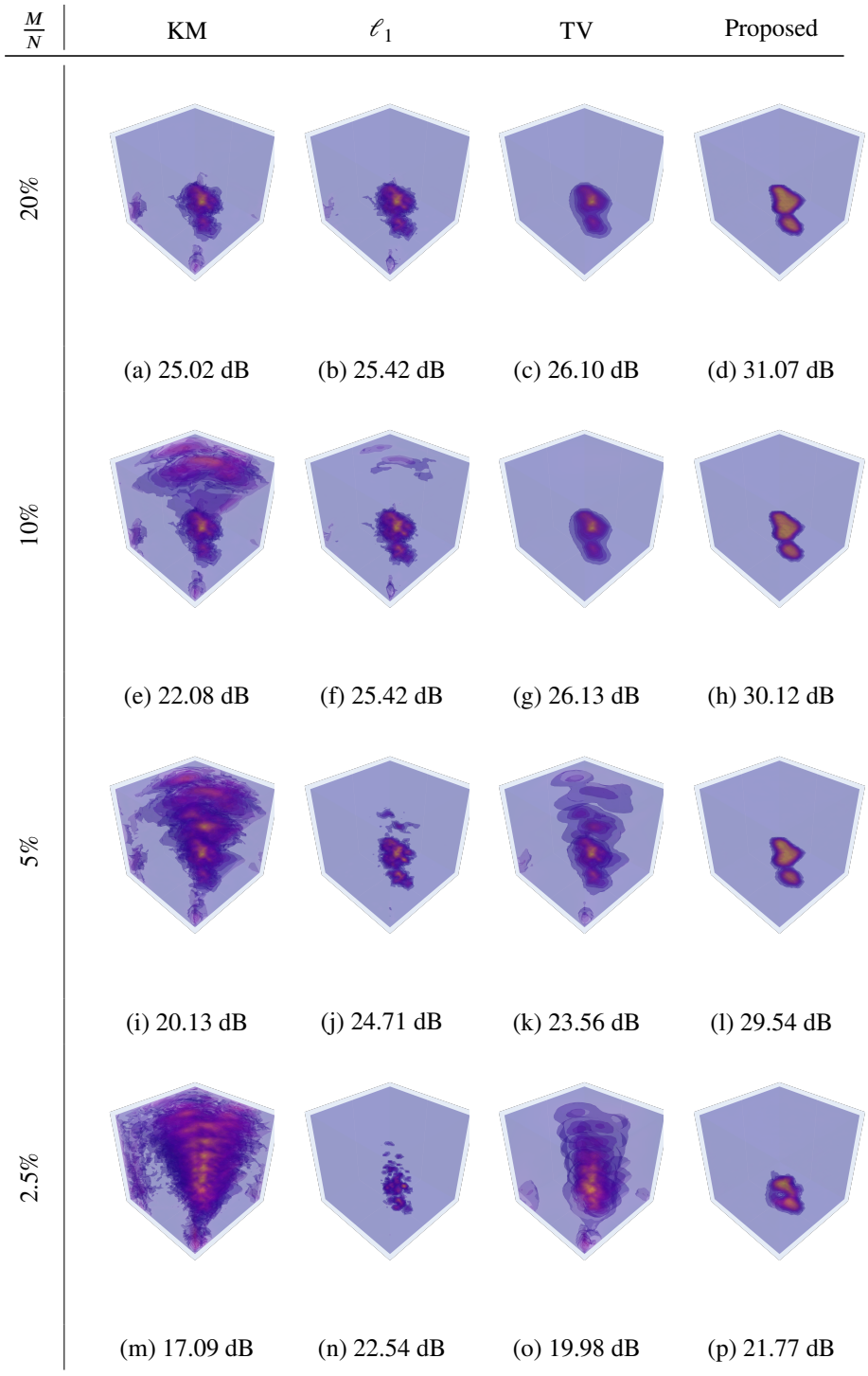


Figure 3.6: Sample reconstructions obtained for different amounts of available data, $\frac{M}{N} = 2.5\%, 5\%, 10\%, 20\%$, at 30 dB measurement SNR. PSNR (dB) of each reconstruction is indicated underneath.

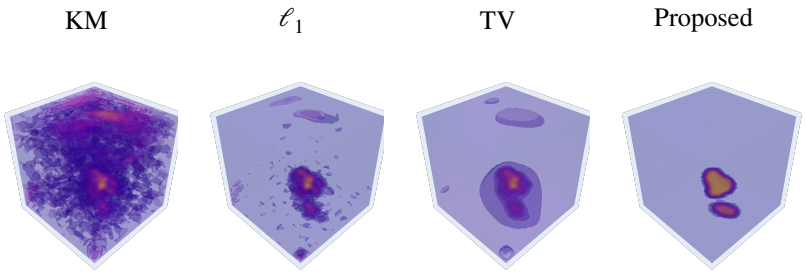
3.3.2.2 Noise Level Analysis

We now fix the available data to 10% and analyze the effect of SNR on the quality of reconstructions. For this, we gradually drop the SNR from 30 dB to 0 dB with steps of 10 dB. The average PSNR of each method is given in Table 3.3 at different SNRs. As seen, the developed learning-based approach outperforms the other methods also for all noise levels. In particular, the performance of the developed method even at the lowest SNR case (i.e. 0 dB) with 28.31 dB PSNR is better than the performance of all compared methods at the highest SNR case (i.e. 30 dB). Similar to the results in the compression level analysis, all regularization-based approaches outperform the direct inversion methods, and in the most ill-posed case with 0 dB SNR, ℓ_1 prior yields better reconstruction than TV.

Table 3.3: Average PSNR on 100 Test Scenes for Different Measurement SNRs using 10% Data.

SNR	0 dB	10 dB	20 dB	30 dB
Back-Projection	22.20	23.35	23.48	23.49
Kirchhoff Migration	22.37	24.26	24.49	24.51
ℓ_1 Regularization	25.40	25.69	25.70	25.70
TV Regularization	23.87	26.02	26.26	26.26
Proposed Method	28.31	29.28	30.12	30.40

In Fig. 3.7 sample reconstructions for 0 dB SNR case are given. Compared to the reconstructions given in Fig. 3.5 for 30 dB SNR case, KM result is severely degraded at this low SNR due to high noise amplification. On the other hand, ℓ_1 and TV-based reconstructions still show some fidelity to the original image, but with more artifacts. More importantly, even for this highly noisy and compressive observation setting, the proposed learning-based PnP method is capable of providing a clean reconstruction that maintains high fidelity to the ground truth.



(a) 20.58 dB (b) 24.89 dB (c) 22.93 dB (d) 28.88 dB

Figure 3.7: Sample reconstructions with $\frac{M}{N} = 10\%$ data at 0 dB measurement SNR. PSNR values are indicated underneath the figures.

3.3.3 Performance Analysis with Experimental Data

We now demonstrate the performance of the developed approach on real-world scenes using experimental measurements available online [78,79]. These experimental measurements were acquired for a scene that contains a toy revolver approximately 50 cm away from a sparse MIMO array [79]. The used MIMO array has 16 transmit and 9 receive Vivaldi antennas that are distributed in a spiral configuration on the antenna plane as shown in Fig. 3.8. The experimental measurements were recorded at 251 uniformly sampled frequencies from 1 to 26 GHz. We aim to infer the reflectivity distribution within a 30 cm $\times$ 30 cm $\times$ 30 cm image cube that contains the revolver.Similar to [79], we choose the sampling interval as 0.5 cm along all three dimensions. This results in an unknown image cube of $61 \times 61 \times 61$ voxels.

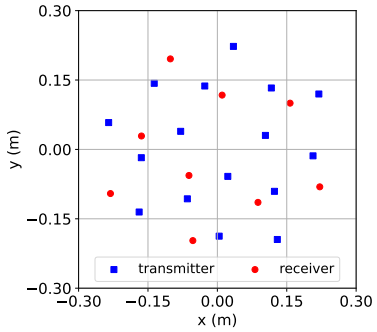


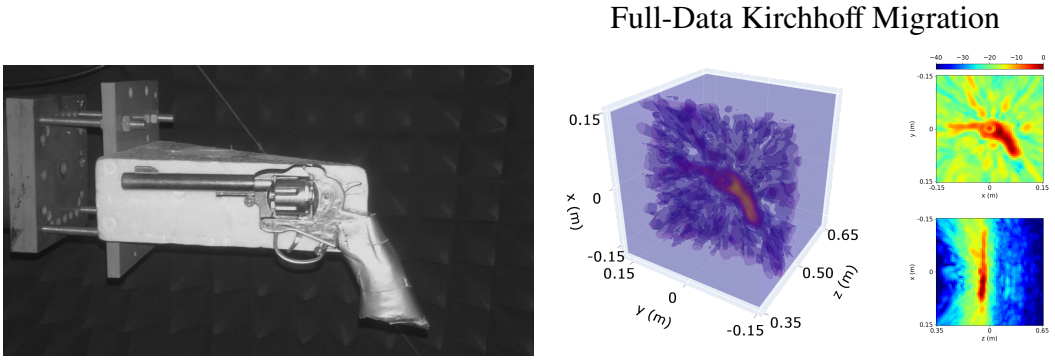
Figure 3.8: Spiral MIMO Array [79].

Since our focus is on compressive imaging, we consider sparse frequency measurements from the band of 4–16 GHz (similar to the simulated setting). In particular, from the available data, we use 7 and 11 uniformly sampled frequencies between 4 and 16 GHz, which respectively correspond to compression ratios of 99.56% and 99.31%. These are equivalent to reconstructing the reflectivity cube with only 0.44% and 0.69% data, yielding to extremely compressive settings.

To reconstruct this real scene using the developed approach with deep prior as well as with TV and ℓ_1 priors, we use the same κ parameters determined in the previous simulated setting. For the choice of the regularization parameter α, we again perform a search for the optimal value to obtain the best reconstruction quality. Moreover, the parameter ϵ in (3.5) is empirically set to $\frac{1}{\sqrt{10}}\|y\|_2$, which approximately corresponds to measurement at 10 dB SNR. Additionally, since the maximum value of the reflectivity magnitudes in the real scene can be different from the synthetic scenes used in training, the reflectivity magnitude at each iteration is scaled with its maximum value prior to entering to the denoiser (in order to fall into the range $[0, 1]$). Then the denoised magnitude at the output of the denoiser is scaled back.

A photograph of the imaged toy revolver is shown in Fig. 3.9a. Note that although this photograph provides a visual reference for comparisons, it does not represent the ground truth reflectivity magnitudes. As an additional reference for comparisons, we also obtain the KM reconstruction of the scene using the full frequency data available (i.e. 251 frequency steps in the band 1-26 GHz), which corresponds to a highly overdetermined setting with $\frac{M}{N} = 361.44\%$ data availability. This *full-data* KM reconstruction is given in Fig. 3.9b to reveal the general shape of the scene reflectivity. But despite using all of the available data, it still contains widespread artifacts, especially over the cross-range dimensions. This is the expected behavior of direct inversion methods with sparse arrays due to the resulting aliasing [79].

Reconstructions obtained for two different compressive settings with 0.44% and 0.69% data are shown in Fig. 3.10. It is seen that the developed approach with deep prior provides the best results with the least amount of artifacts. In particular, KM reconstructions suffer from significant grating lobes and aliasing on range direction (which appears in the form of replication) resulting due to the sparsely sampled frequencies.



(a) Photograph of the Toy Revolver [78, 79]

(b) ($^{M}/_{N}$ =361.46%)

Figure 3.9: Imaged revolver and its reconstruction with *full-data* Kirchhoff Migration [78, 79] provided as visual reference. For the KM reconstruction, the left image provides the 3D view of the image cube in linear scale, whereas the two images on the right show front/side views in dB scale obtained by maximum projection of the image cube onto the $x-y$ and $y-z$ planes, respectively.

Although not as prominent, similar replication artifacts on range direction are also present in the results of hand-crafted regularization approaches. Most notably, TV reconstructions fail to resolve aliasing and contain replicated silhouettes of the revolver. While TV reconstructions perform visually better than KM, they perform poorly compared to ℓ_1 regularization at these highly compressive settings (as similar with the observations in the earlier analysis). In ℓ_1 regularized reconstructions, there are less artifacts along the cross-range directions compared to TV, but the revolver appears as eroded, and there are distributed speckle artifacts, which are more common along the range direction (aligned with the locations of the aliasing artifacts in KM- and TV-based solutions).

On the other hand, the proposed PnP approach with deep prior is capable of providing a near-perfect reconstruction with only 0.69% data. Few aliasing artifacts occur over the range direction at the higher compressed setting with 0.44% data. Nevertheless, in both cases, the edges of the object are sharply reconstructed, and the frame, cylinder, trigger guard, and muzzle of the revolver are all clearly visible. Hence the proposed approach is much less prone to sparse sampling and aliasing, thanks to the power of learned deep priors.

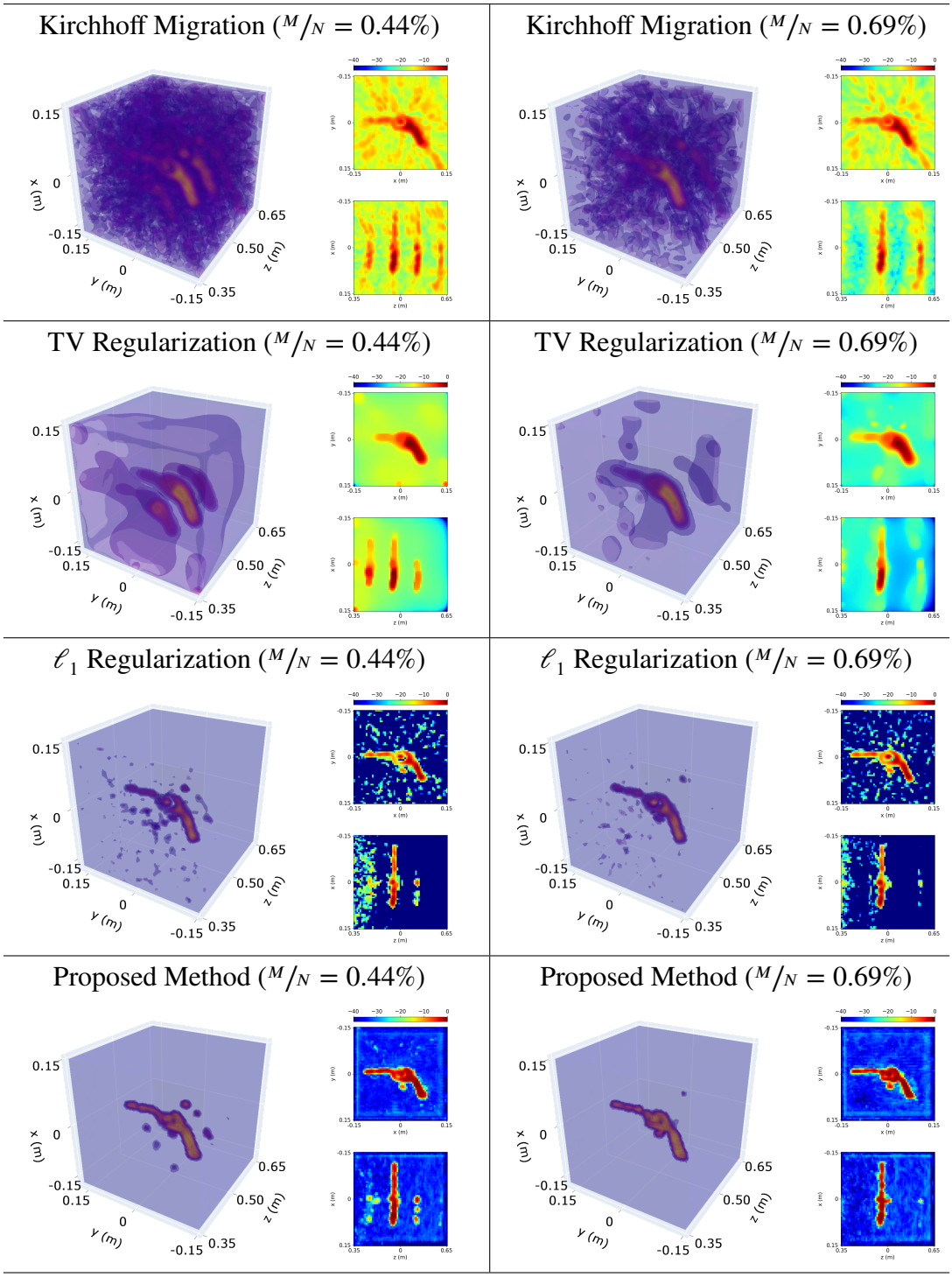


Figure 3.10: Reconstructions of the toy revolver, obtained with different methods at two compressive settings using 7 ($\frac{M}{N}$ =0.44%) and 11 ($\frac{M}{N}$ =0.69%) frequency steps. For each reconstruction, the left image provides the 3D view of the image cube in linear scale, whereas the two images on the right show front/side views in dB scale obtained by maximum projection of the image cube onto the $x-y$ and $y-z$ planes, respectively.

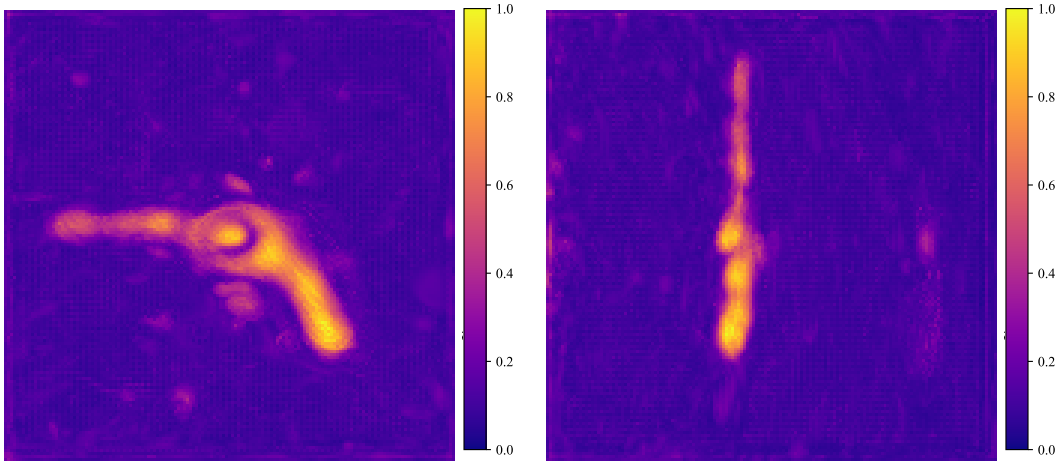
The proposed method not only provides the highest reconstruction quality but also takes only 6 seconds (for the case with 0.44% data). Hence it is again the second fastest method after KM which performs poorly. On the other hand, TV and ℓ_1 regularized solutions suffer from significantly longer computation time, which are approximately 150 seconds.

Overall these real scene experiments demonstrate that the utilization of deep priors in a plug-and-play algorithm enables state-of-the-art reconstruction quality even at highly compressive experimental settings, while also yielding significantly reduced run-time compared to hand-crafted analytical priors. Note that the learned prior is also capable of representing unseen real-world objects, although the training has been performed with synthetic and randomly generated much simpler extended targets. Moreover, even though this experimental observation setting (including antenna array type, number of measurements taken, etc.) differs from the previously analyzed simulated setting, our learning-based method can be directly used without re-training since it is based on PnP framework (and not unrolling). Hence the proposed learning-based PnP method is highly adaptable to experimental data and different observation settings.

3.3.3.1 Reconstruction at a Finer Spatial Resolution

We also test our approach's performance at a finer resolution, for a datacube of size $151 \times 151 \times 151$, within the same physical space (corresponding to 2mm resolution) using 11 frequency steps. Since 3D rendering becomes difficult at this grid size, we only provide the maximum projections of the obtained reconstruction in Fig. 3.11. As seen in this figure, we do not observe any splitting behavior with increased spatial resolution.

We expect the approach to provide similar performance at finer resolutions as long as compression level (data availability) kept similar and finer resolution used also for the training dataset. As the spatial resolution of the test object digresses away from the training dataset's resolution, the performance can inevitably be affected. However, we still observe good performance for both with a grid size of $61 \times 61 \times 61$ and $151 \times 151 \times 151$, in spite of the fact that a grid size of $25 \times 25 \times 49$ has been used for the training dataset.



(a) Projection onto the $x - y$ plane.

(b) Projection onto the $y - z$ plane.

Figure 3.11: Reconstructed image of revolver of size $151 \times 151 \times 151$ (with a 2mm resolution) using 11 frequency steps .

3.4 Discussion

We have developed a novel and efficient plug-and-play approach that enables the reconstruction of 3D complex-valued images involving random phase by exploiting both analytic and deep priors. Our approach provides a unified general framework to effectively handle arbitrary regularization on the magnitude of a complex-valued unknown and is applicable to various complex-valued image formation problems including SAR and MIMO radar imaging with far- or near-field settings. Our development is based on a general closed-form expression provided for the solution of a complex-valued denoising problem with regularization on the magnitude. By utilizing this expression in an ADMM framework, a computationally efficient PnP reconstruction method that consists of simple update steps is obtained.

In this chapter, we applied the developed PnP method to near-field compressive MIMO imaging for reconstruction of the 3D complex-valued scene reflectivities with random phase nature. Within our PnP framework, we utilized a 3D deep denoiser to take advantage of data-adaptive deep priors. To the best of our knowledge, our approach is the first deep prior-based PnP approach demonstrated for near-field radar imaging.

The effectiveness of our approach is illustrated under various compressive and noisy observation scenarios in microwave imaging using both simulated and experimental data. The results show that the developed PnP approach with learned deep prior achieves the state-of-the-art reconstruction quality at highly compressive settings with a generalizability capability for unseen real-world objects and high adaptability to experimental data. The approach also has the advantage of reduced run-time and applicability to different observation settings without re-training due to its PnP nature. Compared to approaches with analytical priors, it is also more robust to sparse data and noise. We observe both with simulated and experimental data that frequency steps as few as 10 provide sufficient measurement diversity for reconstruction of scenes with average complexity. This is an important observation since earlier works generally use hundreds of frequency steps for similar tasks. As expected the bandwidth is more critical than the number of frequency samples taken within this band.

Lastly we note that although the developed PnP method is quite fast with a runtime on the order of seconds, further acceleration and reduction in memory use can be achieved by more efficiently computing the forward and adjoint operators, using methods like fast multipole method (FMM) [16]. Moreover, exploring the performance of the developed method with different 3D denoiser architectures, and joint optimization of the denoiser and MIMO array configuration may improve the reconstruction quality. Enriching our training dataset can also help to improve the performance. Likewise, utilizing a training dataset synthesized for a specific imaging task, such as a dataset consisting of 3D models of concealed weapons, can allow the deep architecture to better learn the task-oriented prior information and can improve the performance.

CHAPTER 4

EFFICIENT PHYSICS-BASED LEARNED RECONSTRUCTION FOR REAL-TIME IMAGING

4.1 Introduction

Existing traditional direct inversion methods in near-field MIMO radar imaging offer fast computation but suffer from low image quality at compressive settings. On the other hand, while regularized inversion methods can offer better reconstruction quality, they suffer from high computational cost which is undesirable in real-time applications. Recently, reconstruction techniques that exploit deep learning have emerged as alternatives to these analytical approaches [25]. These methods are shown to simultaneously achieve high reconstruction quality and low computational cost for various imaging problems [25–28].

To achieve high image quality with low computational cost, the deep two-stage (Deep2S) and deep direct inversion (DeepDI) approaches were developed in [60] as novel deep learning-based direct inversion methods that can be used in real-time applications. These methods consist of two stages. In the first stage, the measurements are back-projected to the image domain to obtain 3D intermediate reconstructions. For this, the Deep2S approach utilizes the adjoint operation and is a physics-based approach. On the other hand, the DeepDI approach replaces this physics-based first stage with a fully connected layer and is a purely learning-based method. In the second stage of both approaches, the intermediate reconstructions from the first stage are refined by a DNN to obtain a magnitude-only reflectivity image. Different than the related learned direct inversion methods in near-field MIMO radar that utilize 2D convolutional ker-

Some parts of this chapter have been published in [80].

nels to process intermediate reconstructions [59], the DeepDI and Deep2S approaches utilize CNNs that exploit 3D convolutions. This allows these approaches to exploit range correlations more effectively, which is particularly important for reconstructing extended 3D targets.

Although the performance of the Deep2S and DeepDI methods were tested for various simulation settings in [60], their performance was not illustrated within an experimental setting. Here, we illustrate the performance with experimental data to demonstrate applicability to real-world targets and measurements. Moreover, we comparatively analyze the effect of complex-valued processing of intermediate reconstructions, as is the case for some of the earlier works in the literature [55,57,58,81], with magnitude-only processing, as used in the Deep2S approach. Additionally, we develop an extension of the Deep2S approach and improve its performance by casting the adjoint operation associated with the MIMO array to a fully trainable complex-valued layer, thus creating a novel hybrid approach between DeepDI and Deep2S. Furthermore, we investigate the resolution achieved with the Deep2S approach at compressive MIMO imaging settings and compare this to the expected theoretical resolution for the conventional (non-compressive) settings. The source codes and dataset used in this chapter are made available at https://github.com/METU-SPACE-Lab/Efficient-Learned-3D-Near-Field-MIMO-Imaging.

The main contributions of this chapter can be summarized as follows:

- Comparative performance evaluation by changing magnitude-only processing with complex-valued processing,
- Development of a learned reconstruction method by casting the adjoint operation associated with the MIMO array to a fully trainable complex-valued layer,
- Resolution analysis with sparse MIMO arrays,
- Performance analysis with experimental data.

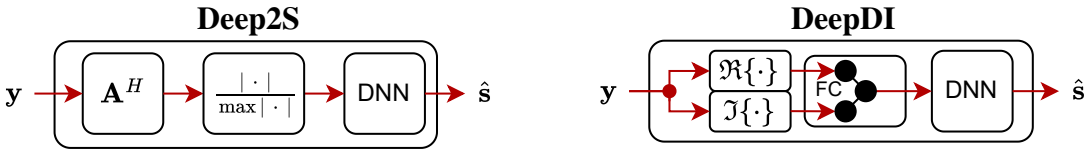
This chapter is organized as follows. First, the Deep2S and DeepDI approaches are explained in Section 4.2. Section 4.3 provides comparisons of different network architectures, where we also develop a novel hybrid approach between Deep2S and DeepDI.

In Section 4.4 we illustrate the performance of the methods with both simulated and experimental measurements. This section also presents resolution analysis. In Section 4.5, conclusions, and final remarks are provided.

4.2 Existing Learned Direct Inversion Methods

In this section, the Deep2S and DeepDI approaches [60] are reviewed. The main goal of these approaches is to achieve high image quality with low computational cost so that the methods can be used in real-time applications. Learning-based direct inversion methods enable such capabilities. For this reason, the Deep2S and the DeepDI approaches are based on learned direct reconstruction [26–28, 82, 83].

The Deep2S approach is a physics-based learned reconstruction method with a two-stage structure as shown in Fig. 4.1a. Here the first stage is an adjoint operation that exploits the physics-based model to provide an initial reconstruction, and the second stage is a 3D U-Net denoiser for refinement. On the other hand, the DeepDI approach, shown in Fig. 4.1b, replaces the physics-based first stage with a fully connected layer and processes the real and imaginary parts of the measurements directly. Hence, this approach aims to perform the reconstruction using only deep neural networks and does not use the physics-based model.



(a) Block Diagram of the Deep2S Approach.

(b) Block Diagram of the DeepDI Approach. "FC" denotes a single fully connected layer.

Figure 4.1: Block diagrams of the Deep2S, DeepDI approaches [60, 80].

The deep neural network utilized in [60] is illustrated in Fig. 4.2, which is similar to the 3D deep denoiser utilized in Chapter 3. The network architecture has encoding and decoding paths which respectively contain the repeated application of 3D convolution, and 3D upconvolution blocks. The number of channels is indicated above each block in the figure where the channels involve filters of size $3 \times 3 \times 3$. These con-

volutional layers are followed by batch normalization (BN) and rectified linear unit (ReLU). While the encoding path contains $2 \times 2 \times 2$ max-pooling layers with strides of two to decrease the spatial size, the decoding path involves upconvolutions with strides of 2 in each dimension to increase the spatial size. Because the input of the network is chosen to be of size $25 \times 25 \times 49$ voxels in x, y, and z directions, cropping is first performed to arrive at the size $24 \times 24 \times 48$ so that spatial size reduction can be performed through $2 \times 2 \times 2$ max-pooling operations. In the decoding path, there are also concatenations with the cropped feature maps from the encoding path as shown with the padding arrow in Fig. 4.2. Hence both the input and output of the network are of size $25 \times 25 \times 49$. Lastly, the output of the network is followed by a sigmoid function to normalize the magnitudes between 0 and 1. The total number of parameters in this 3D U-Net architecture is 2,873,153, which is more than twice the number of parameters of the network in Fig. 3.2.

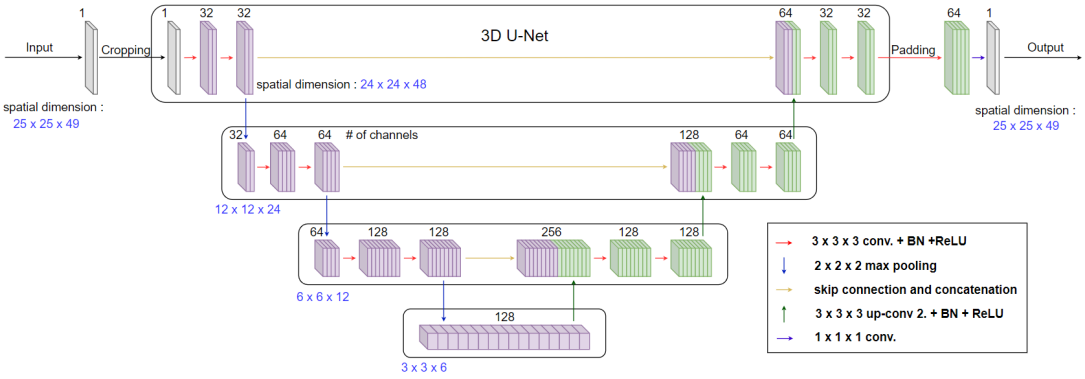


Figure 4.2: Block diagram of the 3D U-net used in [60, 80].

In this chapter, for a fair comparison with the results in [60], we use this original 3D U-Net architecture.

4.3 Improvements and Modifications

In this section, we modify the Deep2S and DeepDI approaches with the goal of analyzing the effect of different architectures on the reconstruction performance.

4.3.1 Improvement on the First Stage

As the first modification, we replace the adjoint operation in the first stage of the Deep2S approach with a fully connected trainable layer to perform multiplication with a learned matrix instead of adjoint matrix. Here, our goal is to analyze whether multiplication with a learned matrix provides better performance than multiplication with the adjoint matrix ($\mathbf{A}^H$) to form the intermediate reconstruction.

Utilizing this trainable layer followed by magnitude operation as the first stage, we obtain a hybrid approach between DeepDI and Deep2S, and call it Deep2S+, where its block diagram is shown in Fig. 4.3. In this setting, we exploit the physics-based information to initialize the weights of the fully trainable model, which reduces the need for substantial training data.

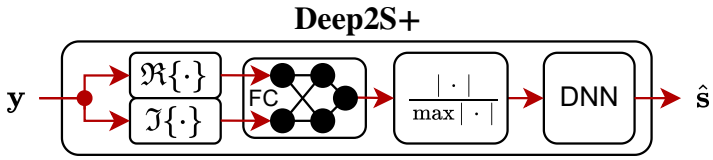


Figure 4.3: Block Diagram of the Deep2S+. "FC" denotes the trainable fully connected layer, which replaces multiplication with the complex-valued adjoint matrix.

The first stage of Deep2S+ serves to project the complex-valued measurements to the image space by performing a linear operation as follows:

$$\begin{bmatrix} \mathbf{v}_R \\ \mathbf{v}_I \end{bmatrix} = \begin{bmatrix} \mathbf{P}_R & -\mathbf{P}_I \\ \mathbf{P}_I & \mathbf{P}_R \end{bmatrix} \begin{bmatrix} \mathbf{y}_R \\ \mathbf{y}_I \end{bmatrix} \tag{4.1}$$

Here $\mathbf{y}_R = \Re\{\mathbf{y}\}, \mathbf{y}_I = \Im\{\mathbf{y}\} \in \mathbb{R}^M$ are real and imaginary components of the measurements, $\mathbf{P}_R, \mathbf{P}_I \in \mathbb{R}^{N\times M}$ represent the real and imaginary components of the applied matrix, and similarly $\mathbf{v}_R$, $\mathbf{v}_I$ are real and imaginary components of the result. This operation simply corresponds to the multiplication of the complex-valued measurements with a complex-valued matrix, and is similar to the operations used for the construction of complex-valued convolutional neural networks [55,81]. The trainable parameters of this layer are $\mathbf{P}_R$ and $\mathbf{P}_I$. Here to exploit physics-based information, we initialize the projection layer in the first stage using $\mathbf{A}^H$, i.e. by setting $\mathbf{P}_R = \Re\{\mathbf{A}^H\}$ and $\mathbf{P}_I = \Im\{\mathbf{A}^H\}$. Note that this initialization provides the adjoint result as output.

Therefore, at the beginning of the training of Deep2S+, we exactly have the Deep2S, and the goal is to improve both the first and second stages of the Deep2S through training (hence the name Deep2S+).

Note that both DeepDI and Deep2S+ are purely DNN-based approaches. But while the first stage of DeepDI independently processes the real and imaginary components of the measurements to form an intermediate reconstruction for the image magnitude, Deep2S+ jointly processes these components (similar to Deep2S) to form a complex-valued intermediate reconstruction (whose magnitude is then input to the second stage). The total number of parameters in Deep2S+ approach is equal to 146,198,153 (hence 2 fold increase compared to DeepDI which has 76,013,778 number of trainalbe parameters).

4.3.2 Modification on the Second Stage

As the second modification, we provide the complex-valued intermediate reconstructions as input to the DNN instead of providing only their magnitudes. The goal here is to investigate whether processing the complex-valued intermediate reconstructions, as done in some of the earlier works [55, 57, 58, 81], increases the imaging quality over only processing the magnitude. Compared to the earlier works that refine the intermediate reconstructions in complex-valued form by either processing real and imaginary parts as two independent channels or using complex-valued CNNs (with a larger number of trainable parameters), the Deep2S approach simply processes only the magnitudes of intermediate reconstructions by taking into account the random phase nature of the reflectivities in various applications. To the best of our knowledge, there is no work that compared the performance of complex-valued refinement with magnitude-only refinement to reconstruct reflectivities involving random phase, and showed the superiority of one of them to the other. For this reason, we perform this comparative evaluation that has been missing in the earlier works.

To process complex-valued intermediate reconstructions, we remove the magnitude operation from the first stage of Deep2S. Instead, we provide the complex-valued representation as input to the 3D U-Net after normalizing it with its magnitude. The first layer of the 3D U-Net is hence modified to accept two-channel input, which con-

sists of the real and imaginary components of the complex-valued intermediate reconstruction. The resulting U-Net architecture contains 2,874,017 parameters (which is slightly larger than before). This type of complex-valued refinement is often referred as RV-CNN in the earlier radar imaging works. Here we call this modified version of Deep2S as CV-Deep2S (Complex-Valued Deep2S).

4.4 Analyses and Comparative Evaluation

For our analyses and simulations, we use the same simulation setting in Chapter 3, where the scene of interest is cube-shaped with a sidelength of 0.3 m. As the sparse MIMO array topology, we again use the same Mills Cross array in Fig. 3.4. The target center is located approximately 0.5 m away from the 2D MIMO array in its near-field. The frequency, f, is swept between 4 GHz to 16 GHz with uniform steps. In the numerical simulations, the number of frequency steps is selected as 15. Considering the theoretical resolution for the uncompressed setting we choose the voxel size as $1.25\,\text{cm} \times 1.25\,\text{cm} \times 0.625\,\text{cm}$ in x, y, and z dimensions as in the previous chapter. Therefore, the reflectivity image that we want to infer contains $25 \times 25 \times 49$ voxels in x, y, and z directions, respectively.

4.4.1 Training of the Deep Architectures

The training is performed using the synthetically generated dataset used in Chapter 3 with ground truth / noisy measurement pairs. We simulate noisy measurements associated with the random-phase added scenes at 30dB SNR and 15 frequency steps using the forward model in (2.8). The implementation and training are performed with TensorFlow [84]- Keras [85]. As the loss function, we use the mean square error between the 3D ground truth magnitudes and the 3D reconstructed magnitudes. The default Adam optimizer [72] is used with an initial learning rate of 10^{-3}. The batch size is set to 16, and the maximum number of epochs is set to 100. The early stopping criterion is selected as 15 consecutive epochs with no drop in validation loss.

We apply transfer learning from the original pre-trained Deep2S [60] for Deep2S+ and initialize the 3D U-Net architecture in the second stage with the pre-trained model of

the same setting. Since the second stage of CV-Deep2S and the Deep2S architectures differ, we train the CV-Deep2S architecture from the start.

4.4.2 Performance Analysis with Simulated Data

The average performance of DeepDI, CV-Deep2S, Deep2S, and Deep2S+ are shown in Table 4.1. Here, in addition to the PSNR, we also compute the structural similarity index (SSIM) [86] of the reconstructions. SSIM is a perceptual quality metric, and similar to PSNR, a higher SSIM value indicates a better reconstruction or higher similarity between the reconstructed and the ground truth images. Since the reconstructed images are 3D, we compute the SSIM for each two-dimensional slice over the range direction and then take its average value.

Table 4.1: Average PSNR and SSIM Values for 15 Frequency Steps at 30dB SNR for Different Network Architectures

DeepDI	CV-Deep2S	Deep2S	Deep2S+
23.40 / 0.81	29.80 / 0.80	29.20 / 0.93	29.40 / 0.95

As seen, the performance increase with Deep2S+ compared to DeepDI is substantial, which illustrates the significance of utilizing physics-based information to give a head-start to a fully trainable architecture. Deep2S+ also provides improvement over Deep2S with a higher PSNR and SSIM, which shows that the training improved the first and second stages of the Deep2S. On the other hand, although the average PSNR of CV-Deep2S exceeds Deep2S and Deep2S+, its average SSIM is substantially lower and even less than DeepDI. This suggests that processing the intermediate reconstructions in complex-valued form (instead of magnitude) does not yield an improvement on the performance.

Compared to the average PSNR achived with the PnP method developed in Chapter 3, the average reconstruction PSNR of Deep2S+ is 0.4dB lower for the same simulation setting with 15 frequency steps and 30 dB SNR. On the other hand, this method offers fast computation with a runtime in the order of milliseconds, whereas the developed

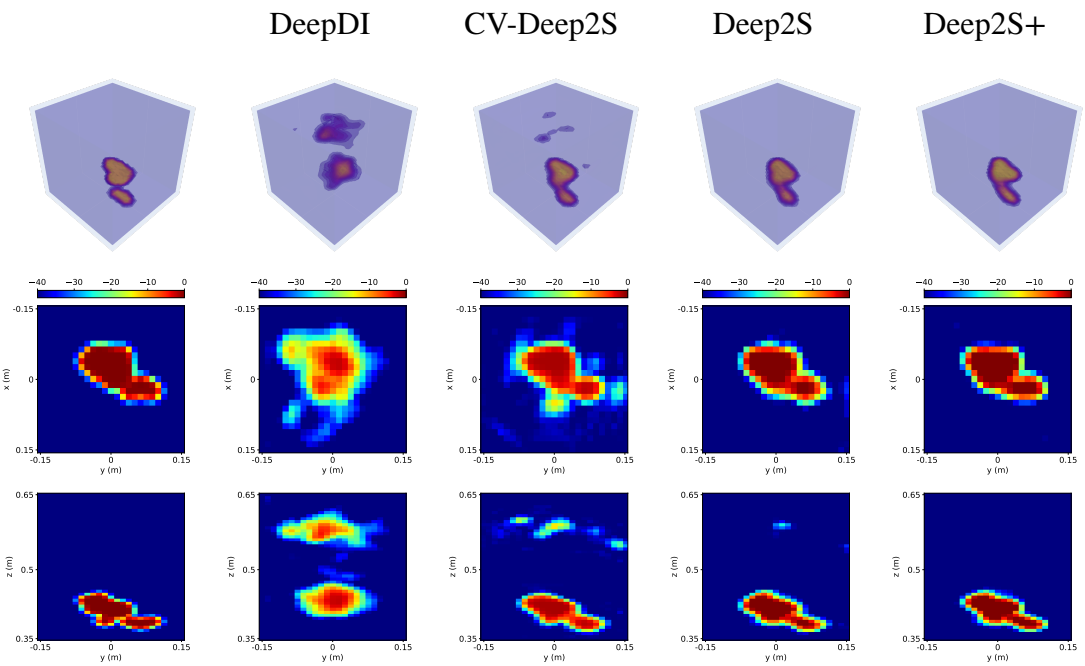
PnP method has a runtime in the order of seconds, which can be critical for real-time imaging. Moreover, while PnP method requires hyper-parameter selection, Deep2S+ do not require this. On the other hand, the developed PnP method can be used for different imaging systems without requiring re-training while Deep2S+ requires re-training. Hence slight performance degradation for Deep2S+ may be tolerated depending on the application of interest and its requirements.

To also visually evaluate the performance, DeepDI, CV-Deep2S, Deep2S, and Deep2S+ reconstructions are illustrated in Fig. 4.4 for a sample image cube from the test dataset. As seen, CV-Deep2S reconstruction contains artifact clusters, especially around $z = 0.57$ plane, while Deep2S and Deep2S+ reconstructions are mostly artifact-free. While the highest PSNR is achieved by CV-Deep2S for this sample image, SSIM is substantially lower compared to the other two (similar to the average performance in Table 4.1). Compared to Deep2S and CV-Deep2S, Deep2S+ better reduces the artifacts along the z-axis arising due to under-sampling (as seen in the maximum projection onto the $y - z$ plane), but slightly suffers from over-smoothing (which might be due to being fully adapted to the training data). These results, along with Table 4.1, suggest that processing the complex-valued intermediate reconstructions does not yield increased image quality as compared to magnitude-only processing.

4.4.3 Performance Analysis with Experimental Data

After our analysis using simulated data, we now demonstrate the performance of the Deep2S approach with experimental measurements of the toy revolver [78, 79] used in Chapter 3. Note that the MIMO array used for these experimental measurements, shown in Fig. 3.8, differs from the Mills Cross array used for the simulated data. Although this difference was not an issue for the developed PnP approach in terms of imaging performance, a different imaging geometry from the simulations requires re-training of the Deep2S approach for operation without a model mismatch. This is because the DNN used in the Deep2S approach acts as an artifact remover whose input-output relation is dependent on the MIMO array geometry.

For this experimental setting, we choose the voxel size the same as the simulated setting (i.e. 1.25cm×1.25cm×0.625cm). Then the reflectivity image that we want



(a) Ground Truth (b) 22.58 / 0.66 (c) 30.90 / 0.79 (d) 29.50 / 0.94 (e) 28.88 / 0.94

Figure 4.4: Reconstructions of the first test image of the synthetically generated dataset at 30 dB SNR (Number of Frequency Steps: 15, $^{M}/_{N} = 7.5\%$) using different architectures. The PSNR (dB) and SSIM of reconstructions are indicated underneath each figure. The first row provides the 3D view of the image cube in linear scale whereas the second and third rows show front/side views in dB scale obtained by maximum projection of the image cube onto the $x-y$ and $y-z$ planes, respectively. (3D rotating views of these reconstructions can be found at https://github.com/METU-SPACE-Lab/Efficient-Learned-3D-Near-Field-MIMO-Imaging as video.)

to infer consists of $25 \times 25 \times 49$ voxels in x, y, and z directions, respectively. For measurements, 15 uniform frequency steps between 4 GHz and 16 GHz are used as before. As a result, the goal is to reconstruct the unknown reflectivity image with $\sim 7\%$ data (M/N) in this compressive MIMO imaging case.

To reconstruct from experimental data, we re-train CV-Deep2S by utilizing the 800 random phase added training images and simulating measurements with the forward model for the used spiral array at 30dB SNR. We also re-train the Deep2S and Deep2S+ models for this measurement setting using transfer learning from the models trained for the cross-array setting. To additionally investigate the performance of the Deep2S with a model mismatch, we use the trained Deep2S model for the cross-array setting without re-training for the spiral array case and call this Deep2S*.

The experimental results are shown in Fig. 4.5, where we also provide the adjoint operation outputs to visually assess the artifacts removed from the intermediate first stage reconstruction. As seen, the reconstruction with adjoint contain widespread volume artifacts, similar to the direct inversion results in Chapter 3. These large artifacts are especially distributed over the cross-range dimensions, which are mainly due to under-sampling with the sparse MIMO array and is related to spatial aliasing [79]. The adjoint result also contains significant artifacts along the range direction (especially around the $z = 0.65$m and $z = 0.55$m planes), which is mainly related to the sparsely sampled 4-16 GHz frequency band. On the other hand, all of the deep learning based approaches substantially reduce these widespread artifacts present in the adjoint reconstruction.

In particular, with Deep2S, the reconstructed revolver is clearly separated from the scene background as desired. It provides a clean and sharp image with a very high dynamic range and contrast, and hence yields the highest imaging quality among the tested methods. Compared to the adjoint operation and Deep2S* involving model mismatch, both the side-lobe artifacts and range artifacts are substantially reduced through re-training for the spiral array setting with simulated data. This illustrates that Deep2S has some flexibility to be used in experimental settings that are different from the trained one; but when the amount of artifacts in the adjoint result significantly changes compared to the trained setting, it benefits from re-training for the new obser-

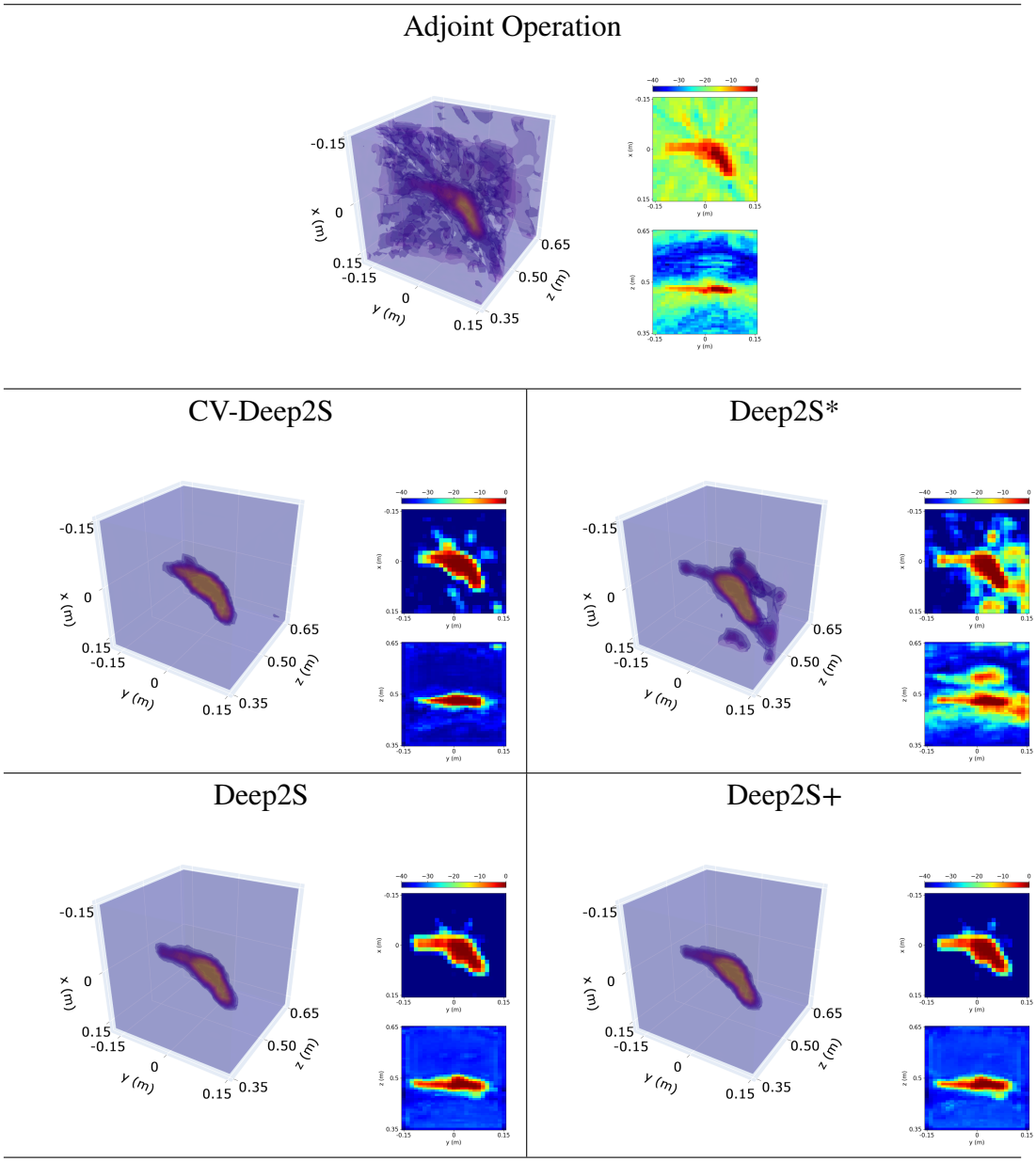


Figure 4.5: Reconstructed reflectivity magnitudes using 15 frequency steps. Deep2S* shows the reconstruction with model mismatch, i.e. the Deep2S model that utilizes the 3D U-Net trained on the Mill's Cross array. For each reconstruction, the left image provides the 3D view of the image cube in linear scale, whereas the two images on the right show front/side views in dB scale obtained by maximum projection of the image cube onto the $x-y$ and $y-z$ planes, respectively.

vation setting, as expected. Moreover, compared to CV-Deep2S, Deep2S appears to better preserve the shape of the revolver and also yields less artifacts. Hence similar to the simulated case, we again observe with experimental data that directly processing the complex-valued intermediate reconstructions does not provide improvement over only processing the magnitudes, and even seems to reduce the generalizability of the approach. Lastly, although Deep2S+ reconstruction is nearly identical to that of Deep2S, similar to the simulated setting, we also provide its reconstruction for completeness.

4.4.4 Resolution Analysis

The resolution of a near-field MIMO imaging system is characterized in a conventional observation setting for a non-sparse antenna array and non-sparse frequency sampling [9]. However, as well-known, the reconstruction quality and resolution degrade in compressive settings with limited data. This is the case here since the data is acquired with a sparse MIMO array at few frequency steps. Here we investigate the resolution achieved with the Deep2S approach at compressive settings and compare this to the expected theoretical resolution for the conventional (non-compressive) setting. For this, firstly, we perform conditioning-based resolution analysis (similar to [87]) using only the observation matrix, which is independent of the reconstruction method. Secondly, we consider multiple separated point targets as sample scenes and demonstrate the resolving capability of the Deep2S approach by analyzing the reconstructed images.

In the first analysis, we consider the conditioning of the inverse problem where the goal is to estimate the values of multiple separated point targets whose locations are known when 15 frequency steps are used. Note that if this problem cannot be properly solved (owing to high condition number), the original reconstruction task of estimating both the value and location of the point targets will also not be possible to solve. Based on this fact, we investigate the condition number of the mentioned inverse problem, which is characterized by the condition number of the submatrix obtained from the observation matrix $\mathbf{A}$ by only keeping the columns associated with the locations of the considered point targets. For this, we consider scenes containing 2 or 4 point

targets with different separation distances. The targets lie on the middle slice of the image cube along the range direction, and are placed with same distance to the center of the $x-y$ plane (as an example, see Fig. 4.7a). The horizontal and vertical separation of the targets are changed from 1cm to 20cm with 1cm steps. Figure 4.6 shows the condition number of the submatrices of $\mathbf{A}$ for these cases. As expected, conditioning gets worse as the number of point targets increases, or their separation distance decreases. More importantly, condition number appears to saturate when the separation distance is increased to 5-6 cm, which suggests that achievable resolution is close to these values. This corresponds to almost 2-fold worse resolution in our compressive setting with only $\sim 7.5\%$ data compared to the expected theoretical resolution of the conventional non-compressive setting (i.e. 2.5 cm).

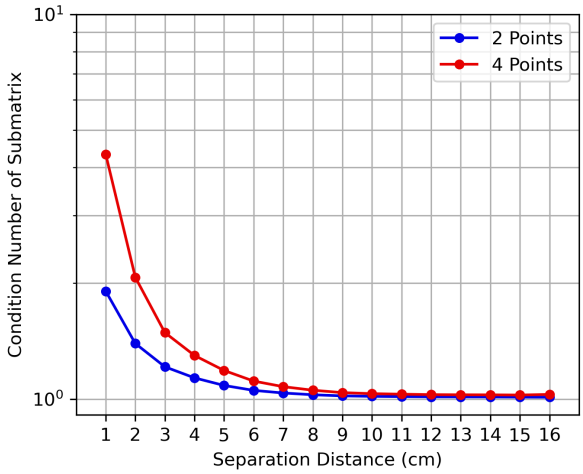
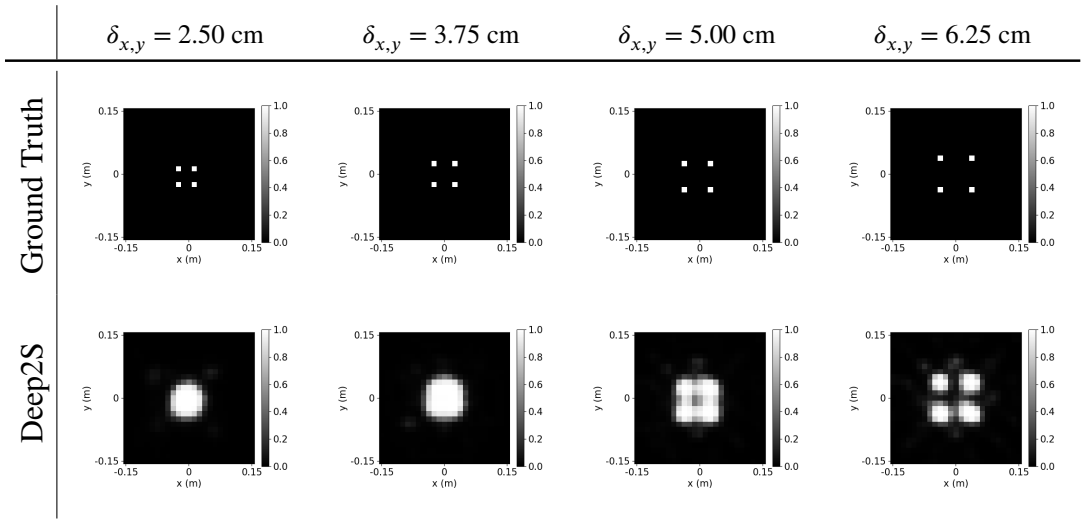
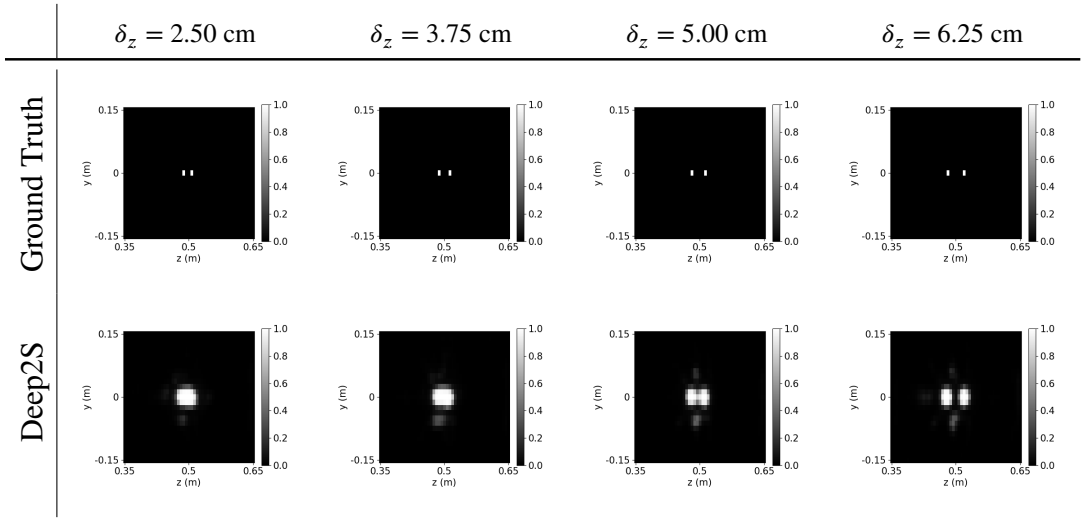


Figure 4.6: Conditioning of the relevant submatrices of $\mathbf{A}$ for different number of point sources and separation distances.

In the second analysis, we investigate the resolving capability of our approach by analyzing the reconstructed images for scenes containing multiple point targets. For this, we place 4 point targets onto the middle $z = 50$cm plane, and 2 point targets along the z axis, with different separation distances as shown in Fig. 4.7. The reconstructions obtained for the observation setting with 30 dB SNR and 15 frequency steps are also shown in the same figure together with the maximum value projected 2D plots of these scenes. As seen, our approach manages to resolve targets separated by a distance of nearly 5cm along the cross range directions, and 2.5cm along the



(a) Cross-Range Resolution Analysis at $z = 50$cm plane



(b) Range Resolution Analysis at $x, y = 0$cm line.

Figure 4.7: Demonstration of resolution using point targets at SNR = 30 dB (Number of Frequency Steps: 15). Separation distance along x/y directions ($\delta_{x,y}$) and along z direction (δ_z) are indicated on the left for each row. Front/side views are shown, which are obtained by maximum projection of the image cube onto the x-y and y-z planes, respectively.

range direction. Hence, similar to the conditioning-based analysis, there is almost 2-fold worse resolution in this compressive setting with only $\sim 7.5\%$ data compared to the expected theoretical resolution of the conventional non-compressive and noise-free setting. However, note that these results are obtained purely based on unrealistic scenes containing point targets, which do not exist in our training dataset that only contains extended 3D targets to mimic the real-world objects. We feel the results of this analysis may change without much change in the achievable resolution for real-world objects by enriching the training dataset with scenes containing point targets.

4.5 Discussion

From the analyzed approaches, the Deep2S method exploits the physics-based knowledge by utilizing the adjoint of the forward model in its first stage. The adjoint operation has the benefit of fast computation due to its non-iterative nature. For comparison, the DeepDI approach replaces the physics-based first stage with a fully connected neural network. In this two-stage structure, the reconstruction is performed directly from the radar measurements using only neural networks, and the observation model is not used. To investigate whether multiplication with a learned matrix provides better performance than multiplication with the adjoint matrix to form the intermediate reconstructions, we developed a hybrid approach, Deep2S+. This method replaces the adjoint operation in the first stage with a fully connected layer, which is learned through transfer learning from the adjoint matrix. All adjustable parameters of the developed approaches are learned end-to-end, which avoids the difficulties of parameter tuning that exist in sparsity-driven compressed sensing recovery methods.

Our comparative results demonstrate that processing the intermediate reconstructions in complex-valued form (instead of magnitude) does not yield an improvement on the performance. It has been observed that both Deep2S and Deep2S+ methods are capable of providing high-quality reconstructions even with highly under-sampled data while enabling fast reconstruction, which show promise for real-time compressive imaging applications.

CHAPTER 5

JOINT OPTIMIZATION OF MIMO ARRAYS AND RECONSTRUCTION METHODS

5.1 Introduction

In near-field radar imaging, reconstruction of the 3D complex-valued scene reflectivity is commonly performed using the data acquired with sparse arrays. Consequently, the design of the array topology greatly affects the reconstruction quality. Accordingly, there has been a significant effort in the radar imaging literature to design and optimize antenna arrays to yield better reconstruction quality [5, 8, 19–24].

The studies on improving or developing radar arrays for near-field imaging can be grouped into two categories. These are heuristic [5, 19, 21], and metric-based designs [8, 19, 20, 22–24, 40]. The difference between the former and the latter comes from the design process, where the heuristic procedure evaluates the performance only after the design is completed.

The heuristically designed arrays are commonly evaluated by their focusing performance or by the distribution of their virtual antennas (such as element shadowing over rotational angles) [5, 19, 21]. The optimality of these heuristic designs with respect to the tested performance criteria is commonly not demonstrated [20], and as a result open to further analysis. Consequently, as metric-based alternatives, these performance criteria are used as ad-hoc metrics within optimization frameworks to yield optimized designs. Most works in this category optimize parametrized antenna positions by relying on black-box optimization methods such as evolutionary algorithms [24, 41] or grid-search [22, 40]. Another systematic approach for designing MIMO arrays is optimizing the array parameters to improve the reconstruction quality

of a particular reconstruction algorithm evaluated on a dataset [8, 20, 23]. The strategies used for this purpose include greedy search for the antenna locations [8], and gradient-based optimization strategies to obtain a binary sampling pattern for antenna positions [20, 23] or directly optimizing a set of antenna coordinates as continuous-valued parameters [23].

Recent advancements in computational imaging have shown that imaging performance can be improved significantly by jointly optimizing the reconstruction algorithm and the parameters of the observation system in an end-to-end (E2E) fashion [20, 23, 40, 42–48]. Some of the jointly optimized systems in the literature include diffractive lenses [42], coded apertures [44] or programmable masks [45] for computational optics, subsampling patterns for undersampled MRI [46], illumination patterns for computational microscopy [47], and seismic acquisition geometry for geoscience applications [48]. Similarly, there are some works in near-field radar imaging for E2E optimization, notably [20, 23, 40].

The key challenge of E2E optimization of MIMO arrays for 3D imaging is the computational and memory requirements to store the observation model and to simulate the measurements. Given the sheer number of voxels of the discretized 3D scene and the number of measurements by the virtual antenna elements, the effective use of discrete optimization methods for the antenna locations by the use of coded masks [23, 40] is not feasible for this problem. On the other hand, optimizing a set of antenna positions as continuous-valued parameters can provide a more viable alternative [20].

In this chapter, we develop a general efficient E2E optimization framework to jointly optimize MIMO arrays and reconstruction methods for 3D near-field imaging. For this, we treat the transmit and receive antenna positions as continuous-valued parameters and enforce them to lie within the antenna aperture. We demonstrate the generality of the developed E2E optimization framework by jointly optimizing the antenna arrays for different reconstruction methods, such as Kirchhoff migration [11], adjoint operation [12, 60], unrolled ℓ_1 regularization [29, 63], and Deep2S [60], which can prove useful for different applications. Our results suggest that the optimized arrays not only offer better reconstruction quality compared to the commonly used MIMO arrays but also are more versatile to be used with different imaging algorithms.

Compared to the work in [20], which develops a similar continuous valued joint optimization method for 2D near-field MIMO imaging, our formulation takes the aperture boundaries into account and is developed for 3D imaging. While the referenced method concentrates on joint optimization of the receiver antenna positions and the weights of a deep learning-based direct inversion method, in our work, we perform joint optimization for both transmit and receive antennas and demonstrate the results using four different imaging algorithms. Additionally, instead of utilizing oversampled beamforming outputs as the target images; we utilize a synthetic dataset consisting of 3D extended targets. This allows us to fine-tune the system and algorithm parameters against ground truth images, potentially achieving performance improvements beyond the oversampled direct reconstruction, as demonstrated in Chapter 3.

The main contributions of this chapter can be summarized as follows:

- Development of a novel generic E2E optimization method for 3D near-field radar imaging,

- New MIMO array configurations optimized specifically for traditional direct reconstruction, regulized reconstruction, and deep-learning based direct inversion,

- Performance comparisons of the commonly used MIMO arrays with different reconstruction algorithms.

This chapter is organized as follows. In Section 5.2 we formalize the joint optimization as a constrained minimization problem and develop our solution approach. We also explain the reconstruction algorithms utilized within our joint optimization framework here. Section 5.3 presents the implementation details and includes analyses of the utilization of different numbers of transmitter and receiver antennas and the reconstruction performance with different optimized arrays. Lastly, we conclude this chapter with Section 5.4 by providing the final remarks.

5.2 Developed Method

This section demonstrates the developed framework for jointly optimizing the parameters of the MIMO imaging system and the parameters of imaging algorithms.

5.2.1 Joint Optimization Problem

In the joint optimization problem, our goal is to concurrently optimize both the parameters of the imaging algorithm, denoted by θ, and the parameters of the MIMO array, denoted by μ.

We formulate this joint optimization as a constrained minimization problem as follows,

$$\hat{\theta}, \hat{\mu} = \arg\min_{\theta,\mu} \sum_{\mathbf{s}\in\chi} \frac{1}{N_\chi} \mathcal{L}(\mathbf{s}, \hat{\mathbf{s}}) \tag{5.1}$$

$$\text{subject to } \hat{\mathbf{s}} = \mathcal{D}_{\theta,\mu}(\mathbf{A}_\mu \mathbf{s} + \mathbf{w}) \ , \ \mu \in \Omega \tag{5.2}$$

where χ denotes the training dataset consisting of N_χ number of random phase added 3D extended targets and $\mathcal{L}(\cdot,\cdot)$ is the cost function computing the fidelity of the reconstructed reflectivity values $\hat{\mathbf{s}}$ to the ground truth reflectivity values, $\mathbf{s}$. Moreover, $\mathcal{D}_{\theta,\mu}$ and $\mathbf{A}_\mu$ respectively denote the utilized reconstruction algorithm and the forward model matrix, where their dependency on the algorithm and system parameters are explicitly indicated with subscripts, θ and μ. Lastly, Ω denotes the set of feasible solutions for the parameters of the observation system.

The number of trainable parameters of learned reconstruction methods is, in general, on the order of millions. This limits the efficient use of most second-order optimization approaches for this large-scale problem. Accordingly, we use a first-order gradient-based method to solve this optimization problem.

To solve the minimization problem in (5.1) we use projected gradient descent [88]

whose iterations are given by,

$$\theta^{l+1} = \theta^{l} - \eta_{\theta} \frac{\partial}{\partial \theta^{l}} \left(\frac{1}{N_{\chi}} \sum_{s \in \chi} \mathcal{L}(\mathbf{s}, \hat{\mathbf{s}}) \right) \tag{5.3}$$

$$\mu^{l+1} = \mathrm{Proj}_{\Omega} \left(\mu^{l} - \eta_{\mu} \frac{\partial}{\partial \mu^{l}} \left(\frac{1}{N_{\chi}} \sum_{s \in \chi} \mathcal{L}(\mathbf{s}, \hat{\mathbf{s}}) \right) \right) \tag{5.4}$$

where η_{μ} and η_{θ} are the step-size parameters for the observation system and algorithm parameters, respectively. $\mathrm{Proj}_{\Omega}(\cdot)$ denotes the projection operator given by

$$\mathrm{Proj}_{\Omega}(\bar{\mu}) \triangleq \arg\min_{\mu} \|\bar{\mu} - \mu\|_2 \text{ s.t. } \mu \in \Omega. \tag{5.5}$$

and is equivalent to the proximal mapping of an indicator function that enforces the system constraints.

In this work, we only optimize the transmit and receive element positions. As a result, Ω denotes the area that lies within the antenna aperture, and the projection operator corresponds to correcting the positions of antennas if they lie outside the aperture by assigning the closest point to their position inside the aperture. For instance, for an antenna aperture at the $z = 0$ plane, if the $x - y$ coordinates of the antennas, x_a and y_a, take values in the intervals $x_a \in [-\frac{h}{2}, +\frac{h}{2}]$, $y_a \in [-\frac{h}{2}, +\frac{h}{2}]$, then the projected coordinates would be computed as follows,

$$x_a^{proj.} = \begin{cases} +\frac{h}{2}, \text{ if } \frac{h}{2} < x_a \\ x_a, \text{ if } -\frac{h}{2} \leq x_a \leq \frac{h}{2} \\ -\frac{h}{2}, \text{ if } x_a < -\frac{h}{2} \end{cases} \quad \text{and} \quad y_a^{proj.} = \begin{cases} +\frac{h}{2}, \text{ if } \frac{h}{2} < y_a \\ y_a, \text{ if } -\frac{h}{2} \leq y_a \leq \frac{h}{2} \\ -\frac{h}{2}, \text{ if } y_a < -\frac{h}{2} \end{cases} . \tag{5.6}$$

5.2.2 Utilized Reconstruction Methods

We perform joint optimization when using the commonly used analytic direct inversion methods such as Kirchhoff migration [11] and adjoint operation [12, 60], ℓ_1 regularization in the form of unrolled iterative shrinkage thresholding algorithm (U-ℓ_1) [29, 63], and physics-based learned direct reconstruction, namely, the Deep2S approach [60].

The learnable parameters of the reconstruction algorithm, θ, differ for each of the utilized algorithms. As traditional direct inversion methods, the KM algorithm and

the adjoint operation do not require optimizaiton. On the other hand, as an iterative regularization-based method, U-ℓ_1 requires setting the step size and regularization parameters. Similarly, as a physics-based learned direct inversion method, the Deep2S approach requires the weights of the utilized DNN architecture to be learned.

For U-ℓ_1, to improve the imaging performance, we utilize L independent learnable step-size parameters for each unrolled iteration [63]. Similarly, as the optimum regularization parameter depends on the imaging setting, we jointly learn L independent soft threshold parameters for each unrolling step.

For the second stage of the Deep2S approach, we utilize a 3D U-net architecture, which is the modified version of the U-net shown in Fig. 3.2. This modified architecture inputs only the magnitudes of the 3D intermediate reconstructions. Hence, the total number of parameters of this network is 1,356,641, which is slightly less than the 3D U-net used in Chapter 3 with 1,357,505 parameters.

5.3 Experiments and Results

In our simulations, we consider the same microwave imaging setting investigated in Chapter 3 and Chapter 4, where the scene of interest has a physical dimension of 30 cm $\times$ 30 cm $\times$ 30 cm, and its center is located 50 cm away from the antenna array. The simulated MIMO array is planar and square-shaped with a side length of 30 cm. The operating frequency, f, is swept between 4 GHz and 16 GHz with 15 uniform steps. Accordingly, we again choose the image voxel size as 1.25 cm along x, y directions, and 0.625 cm along z direction with the goal of achieving the expected theoretical resolution [9]. For the scene of interest, this results in an image cube of $25 \times 25 \times 49$ voxels.

5.3.1 Implementation Details

To implement the descent steps in (5.3) and (5.4), we exploit the PyTorch environment to create a computation graph for auto-differentiation. The block diagram of the generated computation graph is shown in Fig. 5.1.

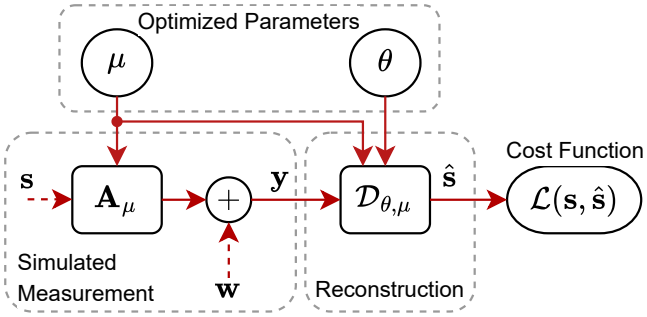


Figure 5.1: Block diagram of the computation graph for E2E optimization where solid lines denote the differentiated path for back-propagation.

We perform joint optimization by utilizing the same synthetically generated dataset used in Chapter 3 and Chapter 4, where the training and validation sets contain 800 and 100 random-phase added image cubes, respectively. We set the cost function in (5.1) to the mean square error computed over the normalized reflectivity magnitudes, (i.e., $\mathcal{L}(\mathbf{s},\hat{\mathbf{s}}) = \frac{1}{N}\| |\mathbf{s}| - \frac{|\hat{\mathbf{s}}|}{\max|\hat{\mathbf{s}}|}\|_2^2$) due to the random-phase nature of scene reflectivity values in various applications.

To accelerate the optimization, we substitute the gradient descent steps with a single iteration of the Adam [72] optimization algorithm. Moreover, to reduce the memory requirements and the optimization duration, we use a batch size of 16 and set the maximum number of epochs to 2000. We stop the training when the validation loss does not improve for 50 epochs. At the end of optimization, we use the parameters that provide the minimum validation loss.

At each iteration of joint optimization, noisy measurements are simulated using (2.8) by generating a new Gaussian noise realization. For each image in the training batch, measurement SNR is randomly selected from a uniform distribution between 0 to 30 dB. Then, white Gaussian noise with standard deviation, $\sigma_w = \frac{\|\mathbf{As}\|_2}{\sqrt{M}} 10^{-\frac{SNR}{20}}$, corresponding to this SNR realization is added to the simulated noiseless measurement of this image.

For joint optimization with the Deep2S, we initialize the weights of the U-net architecture randomly and do not use any transfer learning. On the other hand, for joint

optimization with U-ℓ_1 we use $L = 3$ unrolling iterations and initialize the step size and soft threshold parameters of the iterative shrinkage algorithm as 10^{-6} and 10^{-2} respectively. As classical direct inversion methods, the KM algorithm and adjoint operation do not have any parameters to optimize.

We use an initial learning rate of 10^{-2} for the parameters of the observation system, μ. For the Deep2S approach, the learning rate for the DNN weights is initialized as 10^{-3}. For the U-ℓ_1, due to the scale difference between the step size and soft threshold parameters of the iterative shrinkage algorithm, different learning rates are used, 10^{-6} and 10^{-4}, respectively. Lastly, we use a scheduler to drop the learning rates by a factor of 10 if the validation loss does not improve for 25 epochs.

Joint optimization takes approximately 2 hours on a NVIDIA GeForce RTX 3080 Ti GPU using PyTorch 1.12.0 with CUDA Toolkit 11.6.0 in Python 3.10.6.

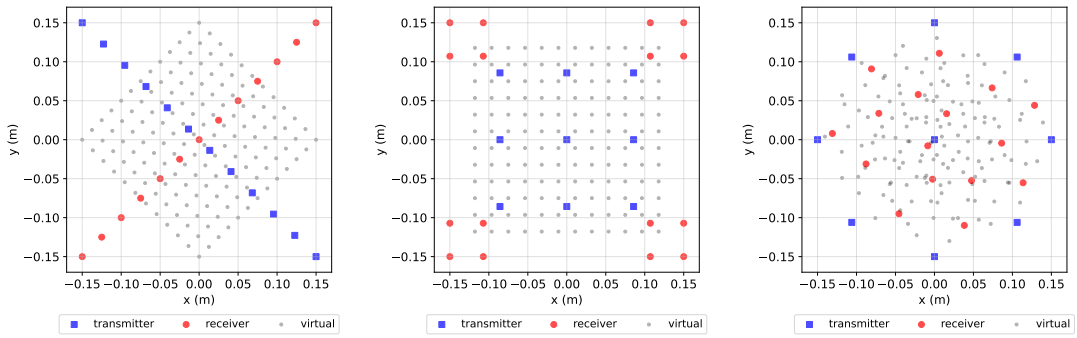
5.3.2 Analysis with Different Number of Transmit and Receive Antennas

We now jointly optimize MIMO arrays consisting of different numbers of transmit and receive antennas with the goal of understanding the optimum number of transmit and receive elements when using KM, adjoint, U-ℓ_1, and Deep2S reconstruction methods.

For our simulations, we fix the total number of antennas to $N_{Tx} + N_{Rx} = 25$ where N_{Tx} and N_{Rx} denote the number of transmit and receive elements, respectively. We perform joint optimization for each possible configuration of $(N_{Tx}, N_{Rx} = 25 - N_{Tx})$. Realize from (2.3) that simultaneously switching all transmit elements to receive elements and all receive elements to transmit elements does not change the forward model matrix. For instance, for the spiral array in Figure 3.8, if we were to have 16 receive elements and 9 transmit elements at the exact locations of the original 16 transmit and 9 receive elements, we would obtain the same forward model matrix. However, one should keep in mind that although these arrays yield the same forward model, the total acquisition time for the measurements is lower for the array with fewer transmitter antennas, hence this array topology is more desirable. Here, by varying the number of transmit elements from $N_{Tx} = 1$ to $N_{Rx} = 12$ and by using $N_{Rx} = 25 - N_{Tx}$ receive elements, we analyze the performance of the optimized array configurations for all

combinations of $(N_{Tx}, N_{Rx} = 25 - N_{Tx})$ antenna pairs.

In our subsequent analyses for comparisons, we also provide the reconstruction performance obtained with the commonly used MIMO arrays [5], namely, the Mill's Cross (MCA), Uniform Rectangular (URA), and Ring Spiral (RSA) Arrays which are shown in Fig. 5.2.



(a) MCA (b) URA (c) RSA

Figure 5.2: Commonly used MIMO arrays for near-field radar imaging.

The average reconstruction PSNRs computed over the test dataset for 0, 10, 20, and 30 dB measurement SNR when using 15 frequency steps are shown in Table 5.1 for the configurations $(N_{Tx}, N_{Rx}) = (9, 16)$ and $(N_{Tx}, N_{Rx}) = (12, 13)$. Here, we demonstrate the reconstruction performance when joint optimization is performed for the same number of transmit and receive elements as the commonly utilized arrays in Fig. 5.2.

Table 5.1: Test PSNR (dB) of the Commonly Used and the Best Performing Optimized Arrays when using 9 and 12 Transmitter Antennas.

	9 Transmitter Antennas			12 Transmitter Antennas	
	URA	RSA	Optimized	MCA	Optimized
Adjoint	22.68	22.54	22.78	21.94	22.93
KM	22.21	22.08	23.50	22.37	23.69
U-ℓ_1	24.90	25.06	25.60	24.72	25.64
Deep2S	30.00	29.77	30.85	29.57	30.85

As seen from the table, for both 9 and 12 transmitter antennas, the reconstruction performances of the jointly optimized systems exceed the reconstruction performances obtained with the commonly used arrays. Moreover, for all reconstruction algorithms, the optimized systems with 12 transmitter antennas have the highest average PSNR. Interestingly, the performance achieved with the jointly optimized arrays for the Deep2S approach is nearly the same for both 9 and 12 transmitter antennas. From the optimized systems, we achieve the highest performance gain for the KM method, with more than a 1.3dB difference (compared to MCA). This result indicates that the developed joint optimization method can successfully improve the antenna positions as the KM method has no algorithm parameters to optimize. On the other hand, the second highest performance gain is obtained for the Deep2S approach with a 0.85 dB difference (compared to URA). Since the DNN used in the Deep2S approach has the highest number of algorithm parameters, this result indicates that we can successfully optimize the algorithm parameters during joint optimization.

We now analyze the joint optimization results for all the remaining configurations to understand the effect of the number of transmitter and receiver antennas on the reconstruction performance. The average reconstruction PSNRs of the optimized MIMO imaging systems are shown in Fig. 5.3, where the graph illustrates the average PSNR on the test dataset for 0, 10, 20, and 30 dB measurement SNR. We omit the adjoint results from this graph since the performance of the KM method exceeds the adjoint operation, and both joint optimization results show a similar trend.

As seen from the figure, for various configurations, the performances achieved with the jointly optimized arrays consistently exceed the performance offered by the baseline MIMO arrays, illustrating the success of the developed E2E optimization framework. Moreover, the average reconstruction performance is the highest for all reconstruction algorithms when the optimization is performed for MIMO arrays consisting of 12 transmit and 13 receiver antennas. This result was expected since the number of independent measurements that can be acquired with a MIMO array is proportional to the number of virtual antennas (i.e., $N_{Tx} \times N_{Rx}$), which takes its highest value when $N_{Tx} \approx N_{Rx}$.

Another observation from Fig. 5.3 is that the average PSNR of the U-ℓ_1 saturates ear-

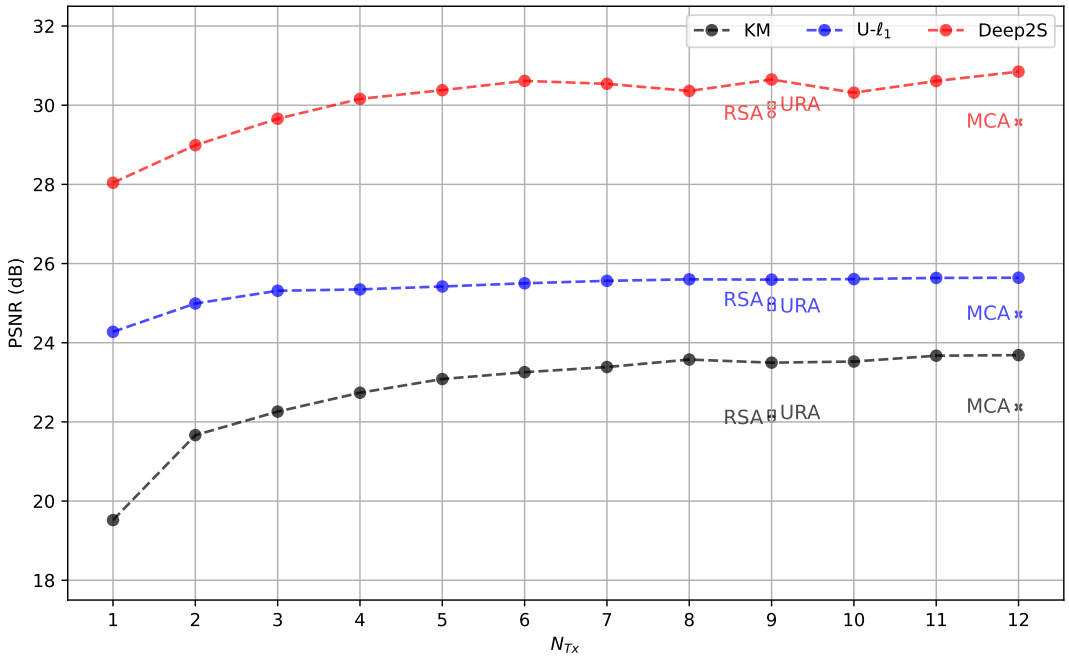
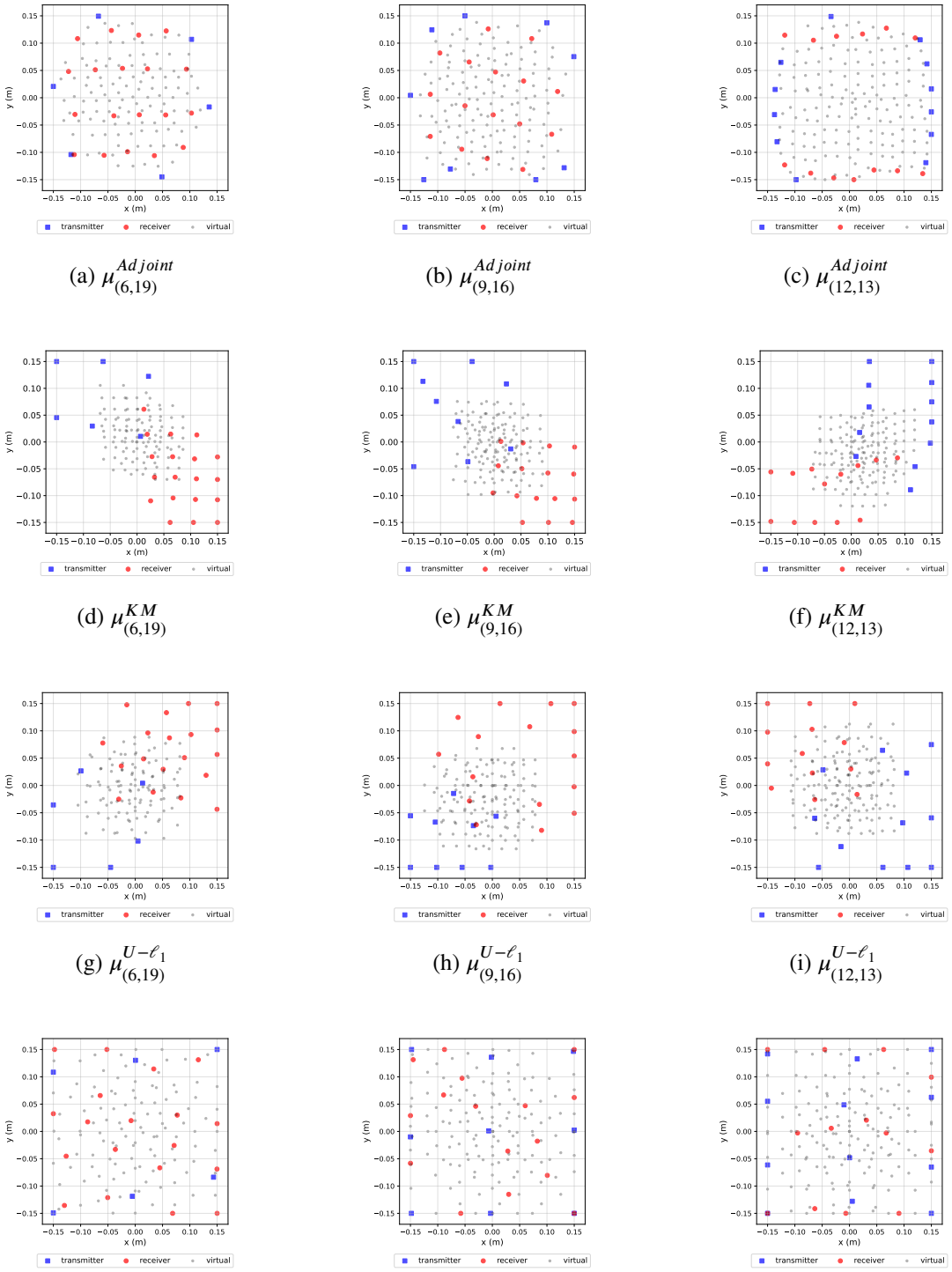


Figure 5.3: Average test performance over different measurement SNRs with respect to the number of transmitter antennas (the number of receiver antennas is set to $N_{Rx} = 25 - N_{Tx}$). Algorithm performances with the commonly used MIMO arrays, namely, Mill's Cross, Uniform Rectangular, and Ring Spiral Arrays are respectively donated with "×","□" and "○".

lier (after 3 transmitter antennas) than other methods. Moreover, the performance difference between its worst- and best-performing joint optimization results is the smallest and has a difference of around 2 dB. This may be attributed to using an image prior, which decreases the dependency of the imaging performance on the array topology. In a similar sense, the joint optimization results with the KM method seem to be the most affected by the choice of a number of transmitter antennas. Its worst- and best-performing joint optimization results have around a 4dB difference. This can be attributed to KM being a traditional direct inversion method where the observation setting is the sole factor affecting its performance. Lastly, as a learning-based direct reconstruction method, the Deep2S achieves the highest image quality (as expected) with more than 28dB PSNR. We believe that as the imaging performance increases, additional performance improvements are harder to achieve. Regardless, the joint optimization results achieved with the Deep2S approach are higher than 30 dB PSNR (which corresponds to the performance of URA) even after using 4 transmitter antennas.

Some examples of the jointly optimized MIMO arrays are shown in Fig. 5.4 where the notation, $\mu^{\mathcal{D}_{\theta,\mu}}_{(N_{Tx},N_{Rx})}$, denotes that the parameters of the antenna array are jointly optimized with the reconstruction algorithm $\mathcal{D}_{\theta,\mu}$ for the case of (N_{Tx}, N_{Rx}) transmitter and receiver antennas.

As seen, the optimized antenna positions have self-emerged interpretable patterns. For instance, when using 6 and 9 transmitter antennas, the jointly optimized MIMO arrays with the adjoint operation place transmitter antennas around the receive elements. We note that similar (heuristically designed) MIMO arrays are presented in the works [22, 40]. Similarly, when using 12 transmitter antennas, the distribution of the antenna positions becomes similar to a commonly used rectangular array [19]. Also note that on all of the joint optimization results, the antennas are placed at the boundaries. Intuitively, distributing the antenna elements to the boundary to yield the largest aperture makes sense as the theoretical cross-range resolution [5] is finer for a larger aperture. Additionally, we observe that the jointly optimized MIMO arrays with the Deep2S have the largest virtual aperture. This may be due to the DNN used at its second stage, enabling more efficient use of measurement diversity in comparison to the other approaches. Furthermore, through the placement of both transmitter and



(a) $\mu^{Adjoint}_{(6,19)}$ (b) $\mu^{Adjoint}_{(9,16)}$ (c) $\mu^{Adjoint}_{(12,13)}$

(d) $\mu^{KM}_{(6,19)}$ (e) $\mu^{KM}_{(9,16)}$ (f) $\mu^{KM}_{(12,13)}$

(g) $\mu^{U-\ell_1}_{(6,19)}$ (h) $\mu^{U-\ell_1}_{(9,16)}$ (i) $\mu^{U-\ell_1}_{(12,13)}$

(j) $\mu^{Deep2S}_{(6,19)}$ (k) $\mu^{Deep2S}_{(9,16)}$ (l) $\mu^{Deep2S}_{(12,13)}$

Figure 5.4: Jointly optimized MIMO arrays. The jointly optimized algorithms are indicated with a subscript and the number of transmitter/receiver antennas (N_{Tx}, N_{Rx}) are indicated with a subscript, $\mu^{D_{\theta,\mu}}_{(N_{Tx},N_{Rx})}$, beneath each figure.

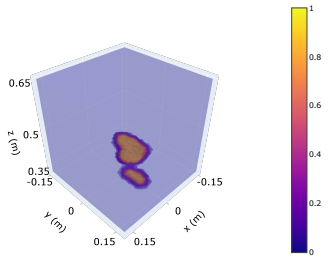
receiver antennas to the corners and the boundaries, its virtual antennas sample nearly all the aperture with a nearly uniform density. Again, we note that ad-hoc metrics that evaluate similar concepts as sampling density are also used in the literature to design MIMO arrays [19]. All these results indicate that the developed E2E optimization method outputs sensible designs.

Sample reconstructions at 30dB SNR with 15 frequency steps for the best-performing optimized arrays and commonly used MIMO arrays are provided in Fig.5.5. As seen, the PSNR of the reconstructions with the optimized arrays are consistently higher than those of the commonly used MIMO arrays. For both KM and U-ℓ_1, the artifacts present on the reconstructions with the commonly used MIMO arrays are substantially reduced with the optimized systems, especially along the range direction. Parallel to our observations from Fig. 5.3, compared to the joint optimization results of the other imaging algorithms, the performance gain for the KM method is the most notable and the highest, with a 2.3dB difference. Similarly, when we compare the reconstructions with the Deep2S approach, the PSNR achieved with the optimized system is more than 1.3 dB higher than the second-best performing array, URA. Moreover, with the optimized array, although a point-like artifact is present near the object, we can see that the shape of the smaller bottom part of the target is better preserved and separated from the main body of the target.

5.3.3 Analysis with Different Optimized Arrays

Although the joint optimization is performed for a specific reconstruction algorithm, the optimized arrays may still offer better image quality than the commonly used MIMO arrays when tested with the imaging algorithms in which the arrays are not jointly optimized. To analyze this, we use the best performing jointly optimized arrays, denoted by $\mu^{Adjoint}_{(12,13)}$, $\mu^{KM}_{(12,13)}$, $\mu^{U-\ell_1}_{(12,13)}$, and $\mu^{Deep2S}_{(12,13)}$ for adjoint, KM, U-ℓ_1 and Deep2S respectively and freeze the parameters of the observation system but train the parameters of the algorithms, θ.

The test performances of the MIMO arrays are presented in Table 5.2. Parallel to our observations before, we see that when the optimized array matches with its jointly optimized algorithm, the system achieves the highest performance. We also see that



(a) Ground Truth

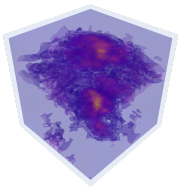
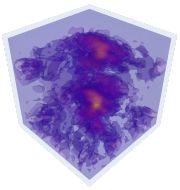
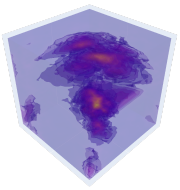
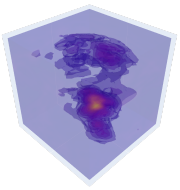
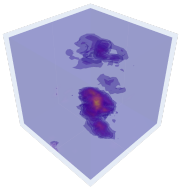
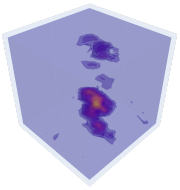
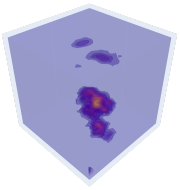
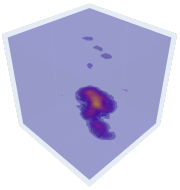
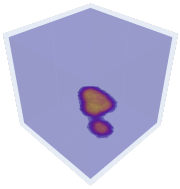
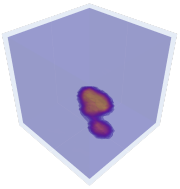
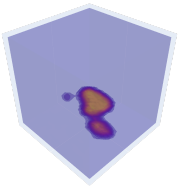
$\mathcal{D}_{\theta,\mu}$	URA	RSA	MCA	$\mu^{\mathcal{D}_{\theta,\mu}}_{(12,13)}$
KM				
	(b) 21.16 dB	(c) 21.30 dB	(d) 21.71 dB	(e) 24.01 dB
U-ℓ_1				
	(f) 25.01 dB	(g) 25.38 dB	(h) 25.09 dB	(i) 25.57 dB
Deep2S				
	(j) 30.96 dB	(k) 29.79 dB	(l) 30.94 dB	(m) 32.32 dB

Figure 5.5: Reconstructions with the commonly used and the optimized arrays at 30 dB Measurement SNR using 15 frequency steps. PSNR values are indicated underneath the figures.

$\mu_{(12,13)}^{Deep2S}$, i.e., the MIMO array jointly optimized with Deep2S, achieves a greater reconstruction quality than the commonly used MIMO arrays for all reconstruction algorithms except for the adjoint operation. This further supports our argument that the jointly optimized Deep2S systems better exploit measurement diversity.

Table 5.2: Test PSNR (dB) of the Commonly Used and the Optimized Arrays with Different Reconstruction Algorithms Computed over 0, 10, 20, and 30 dB measurement SNR using 15 frequency steps.

	Commonly Used Arrays			Jointly Optimized Arrays			
Algorithm	URA	RSA	MCA	$\mu_{(12,13)}^{Adjoint}$	$\mu_{(12,13)}^{KM}$	$\mu_{(12,13)}^{U-\ell_1}$	$\mu_{(12,13)}^{Deep2S}$
Adjoint	22.68	22.54	21.94	**22.93**	21.98	22.26	22.35
KM	22.21	22.08	22.37	22.21	**23.69**	23.42	22.72
U-ℓ_1	24.90	25.06	24.72	24.76	25.55	**25.64**	25.34
Deep2S	30.00	29.77	29.57	30.33	28.79	29.01	**30.85**
Rank	*5th*	*6th*	*7th*	*3rd*	*4th*	*2nd*	*1st*

To systematically evaluate the reconstruction performance of the arrays when using different reconstruction algorithms, we compute their standard scores [89] and rank the results by the average standard scores at the end of the table. As seen, the optimized arrays not only perform the best for the corresponding jointly optimized reconstruction algorithm but also are more versatile compared to URA, RSA, and MCA. From these results, it can be seen that the array $\mu_{(12,13)}^{KM}$ is the most specialized among the optimized arrays, as it achieves the greatest performance gain over the best-performing commonly used array but ranks the lowest in terms of versatility. On the other hand, $\mu_{(12,13)}^{Deep2S}$ shows the highest and $\mu_{(12,13)}^{U-\ell_1}$ the second highest versatility. These results further indicate that, through joint optimization, the developed MIMO arrays can provide more informative measurements.

5.4 Discussion

We have developed a novel joint optimization method for 3D near-field MIMO imaging. The developed joint optimization method can be used with any reconstruction algorithm that can be cast to a differentiable layer. Although we only tested our end-to-end optimization framework for optimizing the antenna positions, the joint optimization can be easily extended to the other design parameters such as operating frequencies.

We demonstrated the generality of our end-to-end optimization framework for imaging methods such as traditional direct inversion approaches, regularized reconstruction, and learned direct inversion. The optimization results yielded sensible array designs with self-emerging and interpretable patterns. Moreover, the developed systems consistently outperform those commonly found in existing literature while also offering versatility to be used with other imaging algorithms.

We believe that while our simulations provide valuable insights, empirical validation through real-world hardware implementations is an essential part of the design process, which is a topic of further study. In such a case, the cost function may be modified to penalize the arrays that have high manufacturing costs or to enforce manufacturing or other practical constraints for the validity of the forward model. Moreover, although we have performed optimization for a planar MIMO array, considering a different topology can also improve the imaging quality [21]. Similarly, the polarization of the antenna elements can provide additional measurement diversity [90], and hence the polarization of the antennas can be included both during formulation of the forward model and joint optimization.

We also believe that imaging performance can also be qualitatively improved by utilizing cost functions with perceptual quality [91]. Furthermore, if the optimized systems are intended to be used for a classification application, end-to-end optimization can be performed by using discriminative cost functions to directly improve the classification accuracy [42]. Utilizing more comprehensive datasets can also boost the imaging performance or yield more robust MIMO array designs. While not explored in this study, integrating our joint optimization method with an unrolled PnP reconstruction

may also improve the performance.

CHAPTER 6

CONCLUSION

In this thesis, we developed efficient physics-informed deep learning techniques for the image reconstruction and array optimization tasks encountered in 3D near-field MIMO radar imaging. Accordingly, we started our discussions in Chapter 2 by explaining the observation model and summarizing the imaging algorithms.

In Chapter 3, we developed a novel PnP approach that enables generalizable means of efficiently enforcing regularization on the magnitude. The developed approach involves simple update steps and can be effectively used with both analytical and deep priors. To the best of our knowledge, our approach is the first deep prior-based PnP approach developed for near-field radar imaging where we encounter a 3D complex-valued image reconstruction problem. Moreover, it is also equally applicable to other similar complex-valued image formation problems, including SAR reconstruction. We demonstrated the effectiveness of our approach through experimental data and extensive simulations involving extended targets. We have illustrated that the developed learning-based method can be directly used without re-training for different observation scenarios since it is based on the PnP framework. The results show that the developed technique not only provides state-of-the-art reconstruction performance but also enables fast computation.

In Chapter 4, we focused on the Deep2S and DeepDI approaches in the literature which enable high reconstruction quality for real-time applications. To improve their performance, we developed a hybrid method called Deep2S+, which lies between Deep2S and DeepDI, enabling physics-based initialization of a fully trainable network. We showed that the performance of the approaches can be improved without requiring substantial training data via physics-based initialization. Here, we also ana-

lyzed the choice of network architecture for imaging quality. Our analyses showed that the complex-valued processing of intermediate reconstructions does not provide any advantage over magnitude-only processing. Moreover, we demonstrated the applicability of these methods to real-world extended targets by using experimental measurements. Although the methods require re-training when the observation model changes, the results demonstrate that Deep2S and Deep2S+ (with a runtime on the order of milliseconds) provide clean and sharp images with a very high dynamic range and contrast.

In Chapter 4, we also analyzed the resolution offered by the Mill's Cross array in compressive settings. Here we considered the conditioning of the inverse problem of estimating only the values of multiple separated point targets when we know their exact locations. Our results suggested that achievable cross-range resolution in the analyzed compressive setting is close to 5-6 cm rather than the expected theoretical resolution of the conventional non-compressive setting (i.e. 2.5 cm).

Lastly, in Chapter 5, we developed a novel end-to-end joint optimization method based on projected gradient descent. We demonstrated the success of the developed framework through simulations by utilizing various reconstruction methods and different observation settings. The developed systems consistently outperformed those commonly used in the literature while offering versatility to be used with other reconstruction algorithms. Moreover, the optimization results yielded sensible array designs with self-emerging and interpretable patterns.

We now provide some general remarks for future studies. We believe that for all of the developed approaches, utilizing a dataset consisting of more realistic or task-oriented training data or utilizing different cost functions could improve the training and performance. Algorithmically, the most straightforward extension of the work presented in this thesis can be unrolling the developed deep-prior-based PnP method. Also, the presented joint optimization method can be extended to frequency selection, allowing the end-to-end optimized systems to achieve higher reconstruction quality, especially at higher compression levels. We also believe that empirical validation through real-world hardware implementations is an essential part of the design process and a topic of further study. More efficient computation of the forward and adjoint operators via

methods like the fast multipole method would also be highly beneficial in terms of accelerating the execution of the algorithms both during training and testing. Furthermore, implementing such an FMM module as a differentiable layer and contributing to open-source libraries such as PyTorch and TensorFlow would be valuable for research in this field.

REFERENCES

[1] F. Fioranelli, S. Salous, and X. Raimundo, "Frequency-modulated interrupted continuous wave as wall removal technique in through-the-wall imaging," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 10, pp. 6272–6283, 2014.

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

[3] X. Zhuge and A. G. Yarovoy, "A sparse aperture MIMO-SAR-based UWB imaging system for concealed weapon detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 1, pp. 509–518, 2010.

[4] E. Anadol, I. Seker, S. Camlica, T. O. Topbas, S. Koc, L. Alatan, F. Oktem, and O. A. Civi, "UWB 3D near-field imaging with a sparse MIMO antenna array for concealed weapon detection," in *Radar Sensor Technology XXII*, vol. 10633, pp. 458–472, SPIE, 2018.

[5] X. Zhuge and G. Yarovoy, Alexander, "Study on two-dimensional sparse MIMO UWB arrays for high resolution near-field imaging," *IEEE Trans. Antennas Propag.*, vol. 60, no. 9, pp. 4173–4182, 2012.

[6] S. S. Ahmed, A. Schiess, and L.-P. Schmidt, "Near field mm-wave imaging with multistatic sparse 2D-arrays," in *2009 European Radar Conference (EuRAD)*, pp. 180–183, IEEE, 2009.

[7] M. E. Yanik and M. Torlak, "Near-field MIMO-SAR millimeter-wave imaging with sparsely sampled aperture data," *IEEE Access*, vol. 7, pp. 31801–31819, 2019.

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

[9] X. Zhuge and A. G. Yarovoy, "Three-dimensional near-field MIMO array imaging using range migration techniques," *IEEE Trans. Image Process.*, vol. 21, no. 6, pp. 3026–3033, 2012.

[10] Y. Álvarez, Y. Rodriguez-Vaqueiro, B. Gonzalez-Valdes, F. Las-Heras, and A. García-Pino, "Fourier-based imaging for subsampled multistatic arrays," *IEEE Trans. Antennas Propag.*, vol. 64, no. 6, pp. 2557–2562, 2016.

[11] X. Zhuge, A. G. Yarovoy, T. Savelyev, and L. Ligthart, "Modified Kirchhoff migration for UWB MIMO array-based radar imaging," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 6, pp. 2692–2703, 2010.

[12] D. L. Marks, O. Yurduseven, and D. R. Smith, "Fourier accelerated multistatic imaging: A fast reconstruction algorithm for multiple-input-multiple-output radar imaging," *IEEE Access*, vol. 5, pp. 1796–1809, 2017.

[13] P. C. Hansen, *Discrete inverse problems: insight and algorithms*, vol. 7. SIAM, 2010.

[14] F. S. Oktem, L. Gao, and F. Kamalabadi, "Computational spectral and ultrafast imaging via convex optimization," in *Handbook of Convex Optimization Methods in Imaging Science*, pp. 105–127, Springer, 2017.

[15] H. E. Güven, A. Güngör, and M. Cetin, "An augmented Lagrangian method for complex-valued compressed SAR imaging," *IEEE Trans. Comput. Imag.*, vol. 2, no. 3, pp. 235–250, 2016.

[16] E. A. Miran, F. S. Oktem, and S. Koc, "Sparse reconstruction for near-field MIMO radar imaging using fast multipole method," *IEEE Access*, vol. 9, pp. 151578–151589, 2021.

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

[18] Z. Yang and Y. R. Zheng, "Near-field 3-D synthetic aperture radar imaging via compressed sensing," in *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 2513–2516, 2012.

[19] Q. Cheng, Y. Liu, H. Zhang, and Y. Hao, "A generic spiral MIMO array design method for short-range UWB imaging," *IEEE Antennas and Wireless Propagation Letters*, vol. 19, no. 5, pp. 851–855, 2020.

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

[21] H. Cetinkaya, S. Kueppers, R. Herschel, and N. Pohl, "Millimeter-wave MIMO array based on semi-circular topology," *IEEE Sensors Journal*, vol. 20, no. 14, pp. 7740–7749, 2020.

[22] K. Tan, S. Wu, Y. Wang, S. Ye, J. Chen, and G. Fang, "A novel two-dimensional sparse MIMO array topology for UWB short-range imaging," *IEEE Antennas and Wireless Propagation Letters*, vol. 15, pp. 702–705, 2016.

[23] S. Wang, S. Li, B. Ren, K. Miao, G. Zhao, and H. Sun, "Convex optimization-based design of sparse arrays for 3-D near-field imaging," *IEEE Sensors Journal*, vol. 23, no. 9, pp. 9640–9648, 2023.

[24] Q. An, A. Hoorfar, H. Lv, and J. Wang, "Task-specific sparse MIMO array design for TWRI using multi-objective CMA-ES," in *2021 XXXIVth General Assembly and Scientific Symposium of the International Union of Radio Science (URSI GASS)*, pp. 1–4, 2021.

[25] S. López-Tapia, R. Molina, and A. K. Katsaggelos, "Deep learning approaches to inverse problems in imaging: Past, present and future," *Digital Signal Processing*, vol. 119, p. 103285, 2021.

[26] K. H. Jin, M. T. McCann, E. Froustey, and M. Unser, "Deep convolutional neural network for inverse problems in imaging," *IEEE Trans. Image Process.*, vol. 26, no. 9, pp. 4509–4522, 2017.

[27] A. Lucas, M. Iliadis, R. Molina, and A. K. Katsaggelos, "Using deep neural networks for inverse problems in imaging: beyond analytical methods," *IEEE Signal Processing Magazine*, vol. 35, no. 1, pp. 20–36, 2018.

[28] G. Ongie, A. Jalal, C. A. M. R. G. Baraniuk, A. G. Dimakis, and R. Willett, "Deep learning techniques for inverse problems in imaging," *IEEE Journal on Selected Areas in Information Theory*, vol. 1, no. 1, pp. 39–56, 2020.

[29] V. Monga, Y. Li, and Y. C. Eldar, "Algorithm unrolling: Interpretable, efficient deep learning for signal and image processing," 2020.

[30] U. S. Kamilov, C. A. Bouman, G. T. Buzzard, and B. Wohlberg, "Plug-and-play methods for integrating physical and learned models in computational imaging: Theory, algorithms, and applications," *IEEE Signal Process. Mag.*, vol. 40, no. 1, pp. 85–97, 2023.

[31] H. K. Aggarwal, M. P. Mani, and M. Jacob, "MoDL: Model-based deep learning architecture for inverse problems," *IEEE Trans. Med. Imag.*, vol. 38, no. 2, pp. 394–405, 2018.

[32] K. Zhang, Y. Li, W. Zuo, L. Zhang, L. Van Gool, and R. Timofte, "Plug-and-play image restoration with deep denoiser prior," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 10, pp. 6360–6376, 2021.

[33] M. B. Alver, A. Saleem, and M. Çetin, "Plug-and-play synthetic aperture radar image formation using deep priors," *IEEE Trans. Comput. Imag.*, vol. 7, pp. 43–57, 2021.

[34] S. V. Venkatakrishnan, C. A. Bouman, and B. Wohlberg, "Plug-and-play priors for model based reconstruction," in *2013 IEEE Global Conference on Signal and Information Processing*, pp. 945–948, 2013.

[35] D. Munson and J. Sanz, "Image reconstruction from frequency-offset Fourier data," *Proceedings of the IEEE*, vol. 72, no. 6, pp. 661–669, 1984.

[36] M. Çetin and W. C. Karl, "Feature-enhanced synthetic aperture radar image formation based on nonquadratic regularization," *IEEE Trans. Image Process.*, vol. 10, no. 4, pp. 623–631, 2001.

[37] Y. Wang, Z. He, X. Zhan, Q. Zeng, and Y. Hu, "A 3-D sparse SAR imaging method based on plug-and-play," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–14, 2022.

[38] S. Li, G. Zhao, H. Li, B. Ren, W. Hu, Y. Liu, W. Yu, and H. Sun, "Near-field radar imaging via compressive sensing," *IEEE Trans. Antennas Propag.*, vol. 63, no. 2, pp. 828–833, 2015.

[39] 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.

[40] K. Tan, S. Wu, Y. Wang, S. Ye, J. Chen, X. Liu, G. Fang, and S. Yan, "On sparse MIMO planar array topology optimization for UWB near-field high-resolution imaging," *IEEE Transactions on Antennas and Propagation*, vol. 65, no. 2, pp. 989–994, 2017.

[41] A. Di Serio, P. Hügler, F. Roos, and C. Waldschmidt, "2-D MIMO radar: A method for array performance assessment and design of a planar antenna array," *IEEE Transactions on Antennas and Propagation*, vol. 68, no. 6, pp. 4604–4616, 2020.

[42] J. Bacca, E. Martinez, and H. Arguello, "Computational spectral imaging: a contemporary overview," *J. Opt. Soc. Am. A*, vol. 40, pp. C115–C125, Apr 2023.

[43] D. Mengu, M. S. S. Rahman, Y. Luo, J. Li, O. Kulce, and A. Ozcan, "At the intersection of optics and deep learning: statistical inference, computing, and inverse design," *Adv. Opt. Photon.*, vol. 14, pp. 209–290, Jun 2022.

[44] L. Wang, T. Zhang, Y. Fu, and H. Huang, "HyperReconNet: Joint coded aperture optimization and image reconstruction for compressive hyperspectral imaging," *IEEE Transactions on Image Processing*, vol. 28, no. 5, pp. 2257–2270, 2019.

[45] J. Lopez, E. Vargas, and H. Arguello, "Depth estimation from a single optical encoded image using a learned colored-coded aperture," 2023.

[46] C. D. Bahadir, A. Q. Wang, A. V. Dalca, and M. R. Sabuncu, "Deep-learning-based optimization of the under-sampling pattern in MRI," *IEEE Transactions on Computational Imaging*, vol. 6, pp. 1139–1152, 2020.

[47] M. R. Kellman, E. Bostan, N. Repina, and L. Waller, "Physics-based learned design: Optimized coded-illumination for quantitative phase imaging," 2019.

[48] A. Hernandez-Rojas and H. Arguello, "3D geometry design via end-to-end optimization for land seismic acquisition," in *2022 IEEE International Conference on Image Processing (ICIP)*, pp. 4053–4057, 2022.

[49] A. Beck and M. Teboulle, "Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems," *IEEE Trans. Image Process.*, vol. 18, no. 11, pp. 2419–2434, 2009.

[50] E. J. Candes and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 21–30, 2008.

[51] S. Boyd, N. Parikh, E. Chu, B. Peleato, J. Eckstein, *et al.*, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends® in Machine learning*, vol. 3, no. 1, pp. 1–122, 2011.

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

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

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

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

[56] G. Sun, F. Zhang, B. Gao, Y. Zhou, Y. Xiang, and S. Pan, "Photonics-based 3D radar imaging with CNN-assisted fast and noise-resistant image construction," *Optics Express*, vol. 29, no. 13, pp. 19352–19361, 2021.

[57] T. Yang, H. Shi, M. Lang, and J. Guo, "ISAR imaging enhancement: Exploiting deep convolutional neural network for signal reconstruction," *International Journal of Remote Sensing*, vol. 41, no. 24, pp. 9447–9468, 2020.

[58] H. Mu, Y. Zhang, C. Ding, Y. Jiang, M. H. Er, and A. C. Kot, "DeepImaging: A ground moving target imaging based on CNN for SAR-GMTI system," *IEEE Geoscience and Remote Sensing Letters*, vol. 18, no. 1, pp. 117–121, 2020.

[59] Q. Cheng, A. A. Ihalage, Y. Liu, and Y. Hao, "Compressive sensing radar imaging with convolutional neural networks," *IEEE Access*, vol. 8, pp. 212917–212926, 2020.

[60] I. Manisalı, "Deep learning-based reconstruction methods for near-field MIMO radar imaging," Master's thesis, Middle East Technical University, 2022.

[61] G. T. Buzzard, S. H. Chan, S. Sreehari, and C. A. Bouman, "Plug-and-play unplugged: Optimization-free reconstruction using consensus equilibrium," *SIAM Journal on Imaging Sciences*, vol. 11, no. 3, pp. 2001–2020, 2018.

[62] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "BM3D image denoising with shape-adaptive principal component analysis," in *SPARS'09-Signal Processing with Adaptive Sparse Structured Representations*, 2009.

[63] J. Zhang and B. Ghanem, "ISTA-Net: Interpretable optimization-inspired deep network for image compressive sensing," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 1828–1837, 2018.

[64] S. Wei, Z. Zhou, M. Wang, H. Zhang, J. Shi, X. Zhang, and L. Fan, "Learning-based split unfolding framework for 3-D mmW radar sparse imaging," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–17, 2022.

[65] O. Oral and F. S. Oktem, "Plug-and-play reconstruction with 3D deep prior for complex-valued near-field MIMO imaging," in *2023 31st European Signal Processing Conference (EUSIPCO)*, pp. 496–500, 2023.

[66] O. Oral and F. S. Oktem, "Plug-and-play regularization on magnitude with deep priors for 3D Near-Field MIMO imaging," *IEEE Transactions on Computational Imaging*, vol. 10, pp. 762–773, 2024.

[67] 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)*, pp. 1–6, IEEE, 2022.

[68] M. V. Afonso, J. M. Bioucas-Dias, and M. A. Figueiredo, "An augmented Lagrangian approach to the constrained optimization formulation of imaging inverse problems," *IEEE Trans. Image Process.*, vol. 20, no. 3, pp. 681–695, 2010.

[69] J. Fessler, "EECS 598-006, optimization methods for signal and image processing and machine learning, chapter 5 proximal methods," vol. 20, 2021.

[70] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising," *IEEE Trans. Image Process.*, vol. 26, no. 7, pp. 3142–3155, 2017.

[71] O. Ronneberger, P. Fischer, and T. Brox, "U-net: convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*, pp. 234–241, Springer, 2015.

[72] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[73] A. Chambolle and T. Pock, "A first-order primal-dual algorithm for convex problems with applications to imaging," *Journal of Mathematical Imaging and Vision*, vol. 40, pp. 120–145, 2011.

[74] E. Y. Sidky, J. H. Jørgensen, and X. Pan, "Convex optimization problem prototyping for image reconstruction in computed tomography with the Chambolle–Pock algorithm," *Physics in Medicine Biology*, vol. 57, no. 10, p. 3065, 2012.

[75] E. Ryu, J. Liu, S. Wang, X. Chen, Z. Wang, and W. Yin, "Plug-and-play methods provably converge with properly trained denoisers," in *Proceedings of the 36th International Conference on Machine Learning* (K. Chaudhuri and R. Salakhutdinov, eds.), vol. 97 of *Proceedings of Machine Learning Research*, pp. 5546–5557, PMLR, 09–15 Jun 2019.

[76] S. H. Chan, X. Wang, and O. A. Elgendy, "Plug-and-play ADMM for image restoration: Fixed-point convergence and applications," *IEEE Trans. Comput. Imag.*, vol. 3, no. 1, pp. 84–98, 2017.

[77] K. Özaslan, "Fast and robust solution techniques for large scale linear least squares problems," Master's thesis, Bilkent University, 2020.

[78] J. Wang, "EM data acquired with irregular planar MIMO arrays," 2020.

[79] J. Wang, P. Aubry, and A. Yarovoy, "3-D short-range imaging with irregular MIMO arrays using NUFFT-based range migration algorithm," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 7, pp. 4730–4742, 2020.

[80] I. Manisali, O. Oral, and F. S. Oktem, "Efficient physics-based learned reconstruction methods for real-time 3D near-field MIMO radar imaging," *Digital Signal Processing*, vol. 144, p. 104274, 2024.

[81] H. Jing, S. Li, K. Miao, S. Wang, X. Cui, G. Zhao, and H. Sun, "Enhanced millimeter-wave 3-D imaging via complex-valued fully convolutional neural network," *Electronics*, vol. 11, no. 1, p. 147, 2022.

[82] K. Kulkarni, S. Lohit, P. Turaga, R. Kerviche, and A. Ashok, "Reconnet: Non-iterative reconstruction of images from compressively sensed measurements," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 449–458, 2016.

[83] B. Zhu, J. Z. Liu, S. F. Cauley, B. R. Rosen, and M. S. Rosen, "Image reconstruction by domain-transform manifold learning," *Nature*, vol. 555, no. 7697, pp. 487–492, 2018.

[84] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng, "TensorFlow: Large-scale machine learning on heterogeneous systems," 2015. Software available from tensorflow.org.

[85] F. Chollet *et al.*, "Keras." `https://keras.io`, 2015.

[86] 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.

[87] F. S. Oktem, O. F. Kar, C. D. Bezek, and F. Kamalabadi, "High-resolution multispectral imaging with diffractive lenses and learned reconstruction," *IEEE Transactions on Computational Imaging*, vol. 7, pp. 489–504, 2021.

[88] S. Boyd, L. Xiao, and A. Mutapcic, "Subgradient methods," *lecture notes of EE392o, Stanford University, Autumn Quarter*, vol. 2004, no. 01, 2003.

[89] M. Saisana, *Standard Scores*, pp. 6321–6322. Dordrecht: Springer Netherlands, 2014.

[90] A. Dalkilic, F. S. Oktem, and L. Alatan, "Enhanced near-field microwave imaging system with polarization diversity," *IEEE Antennas and Wireless Propagation Letters*, vol. 23, no. 3, pp. 1085–1089, 2024.

[91] Q. Yang, P. Yan, Y. Zhang, H. Yu, Y. Shi, X. Mou, M. K. Kalra, Y. Zhang, L. Sun, and G. Wang, "Low-dose CT image denoising using a generative adversarial network with Wasserstein distance and perceptual loss," *IEEE Transactions on Medical Imaging*, vol. 37, no. 6, pp. 1348–1357, 2018.