# NOVEL HYBRID-LEARNING ALGORITHMS FOR IMPROVED

# MILLIMETER-WAVE IMAGING SYSTEMS

by

Josiah W. Smith

APPROVED BY SUPERVISORY COMMITTEE:

_____

Murat Torlak, Chair

_____

Naofal Al-Dhahir

_____

Carlos A. Busso-Recabarren

_____

Randall E. Lehmann

*Dedicated to my wife, Morgan.*

NOVEL HYBRID-LEARNING ALGORITHMS FOR IMPROVED

MILLIMETER-WAVE IMAGING SYSTEMS

by

JOSIAH W. SMITH, BSEE

DISSERTATION

Presented to the Faculty of

The University of Texas at Dallas

in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY IN

ELECTRICAL ENGINEERING

THE UNIVERSITY OF TEXAS AT DALLAS

May 2022

# ACKNOWLEDGMENTS

Finally, I am eternally indebted to my wife, my adventure partner, and my love for her endless love for me and support through the most challenging times. She is my inspiration and passion – providing me with strength to overcome challenges and love to extend to others.

March 2022

# NOVEL HYBRID-LEARNING ALGORITHMS FOR IMPROVED

# MILLIMETER-WAVE IMAGING SYSTEMS

Josiah W. Smith, PhD
The University of Texas at Dallas, 2022

Supervising Professor: Murat Torlak, Chair

Increasing attention is being paid to millimeter-wave (mmWave), 30 GHz to 300 GHz, and terahertz (THz), 300 GHz to 10 THz, sensing applications including security sensing, industrial packaging, medical imaging, and non-destructive testing. Traditional methods for perception and imaging are challenged by novel data-driven algorithms that offer improved resolution, localization, and detection rates. Over the past decade, deep learning technology has garnered substantial popularity, particularly in perception and computer vision applications. Whereas conventional signal processing techniques are more easily generalized to various applications, hybrid approaches where signal processing and learning-based algorithms are interleaved pose a promising compromise between performance and generalizability. Furthermore, such hybrid algorithms improve model training by leveraging the known characteristics of radio frequency (RF) waveforms, thus yielding more efficiently trained deep learning algorithms and offering higher performance than conventional methods.

This dissertation introduces novel hybrid-learning algorithms for improved mmWave imaging systems applicable to a host of problems in perception and sensing. Various problem spaces are explored, including static and dynamic gesture classification; precise hand localization for human computer interaction; high-resolution near-field mmWave imaging using forward synthetic aperture radar (SAR); SAR under irregular scanning geometries; mmWave image

super-resolution using deep neural network (DNN) and Vision Transformer (ViT) architectures; and data-level multiband radar fusion using a novel hybrid-learning architecture. Furthermore, we introduce several novel approaches for deep learning model training and dataset synthesis. Depending on the application, a varying balance of classical signal processing techniques and deep learning is applied to optimally leverage the advantages of each technique. To verify the proposed algorithms, we employ virtual prototyping via simulation and develop custom-built imaging testbeds for empirical testing. Our custom tools for algorithm development, dataset generation, system-level design, and deployment are made public to promote further innovation in this arena. The simulation and experimental results demonstrate the wide application space of hybrid-learning algorithms and the efficacy of joint signal processing data-driven algorithms for radar sensing, perception, and imaging.

TABLE OF CONTENTS

# LIST OF FIGURES

# LIST OF TABLES

# CHAPTER 1

# INTRODUCTION

## 1.1 Background and Motivation

Low-cost electromagnetic (EM) imaging systems have gained attention over the past decade as commercially available radar platforms have become increasingly affordable. Millimeter-wave (mmWave) radar has attracted exceptional interest for applications such as gesture recognition [1, 2, 3, 4, 5, 6, 7], concealed threat detection [8, 9, 10, 11], and medical imaging [12, 13, 14, 15, 16, 17], owing to its semi-penetrating non-ionizing nature and low power consumption. Although extensive studies have been conducted on deep learning for image processing and computer vision [18, 19, 20, 21] and conventional signal processing of radar signals [8, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31], we examine an emerging field by leveraging the advantages of both approaches to develop novel hybrid-learning algorithms. The degree to which data-driven and conventional algorithms are employed varies by application, constraints, and requirements. As we approach several distinct applications, we apply the hybrid-learning approach of optimally leveraging the strengths of each domain to produce effective and efficient systems.

As machine learning algorithms are gaining significant attention for large-scale to edge applications, machine learning on radar data is gaining momentum particularly for classification and perception. Radar hand gesture recognition is of notable interest as increasing importance is placed on privacy and non-invasive sensing methods are generally preferred. In addition, sensing using optical or depth cameras often requires ideal lighting and temperature conditions [32, 33]. mmWave radar devices have recently emerged as a promising alternative offering low-cost system-on-chip sensors whose output signals contain precise spatial information even under non-ideal imaging conditions [7]. However, proper handling of radar signals is essential for high-fidelity gesture sensing systems and can result in considerably varied performance.

In addition to gesture sensing, data-driven perception on synthetic aperture radar (SAR) images is used in applications such as smartphone imaging [34], UAV SAR [35], and automotive imaging [36]. However, high-resolution image reconstruction for irregular SAR scanning geometries using existing algorithms requires computationally prohibitive techniques. Efficient algorithms for common SAR patterns (linear [37, 38], rectilinear/planar [8, 39, 40, 41], circular [42, 43, 44, 45], and cylindrical [46, 47, 48, 49, 50]) have been investigated in the literature, but computationally tractable algorithms remain unexplored for irregular scanning geometries.

On the other hand, while many studies on high-resolution near-field mmWave imaging have been conducted on the signal processing front [8, 39, 46, 51, 52, 53], incorporating data-driven techniques into the imaging pipeline has received limited attention [54, 55]. Optical image super-resolution techniques have gained significant attention in recent years from the machine learning community [18, 56]; however, SAR image super-resolution has largely been relegated to the far-field imaging regime [57, 58, 59, 60, 61, 62, 63]. Besides, near-field SAR super-resolution presents several unique challenges, particularly the availability of large, meaningful datasets for training convolutional neural network (CNN)-based algorithms. Moreover, prior works assume simplistic targets consisting of only randomly placed point targets [43, 64, 65, 66]. To address these shortcomings, we propose a novel simulation platform for generating large meaningful datasets containing radar data from sophisticated objects and 3-D models. With the flexibility of this toolbox, we address emerging problems in SAR image enhancement for irregular scanning geometries using data-driven techniques.

Uniting our efforts towards preprocessing algorithms and improved data-driven imaging algorithms, we investigate a contactless gesture control framework that employs deep learning-aided super-resolution techniques in combination with classical radar signal processing methods to improve performance. Existing studies on contactless gesture control, such as gesture radar, commonly apply machine learning and deep learning techniques. Gurbuz et al. employ a multi-frequency radio-frequency (RF) sensor to recognize American sign

language patterns with high accuracy using several machine learning techniques such as support vector machines (SVM), random forest, linear discriminant analysis, and k-nearest neighbors [67]. Gesture recognition algorithms have been examined for mmWave radar using SVM classifiers [3], CNN techniques [2, 4, 68, 69], LSTM networks [70], etc. [1]. However, these methods separately apply data-driven and signal processing techniques. By employing hybrid-learning algorithms, our proposed method combines the advantages of signal processing and deep learning to yield a performance gain over previous work [71].

Finally, interleaved hybrid-learning algorithms, particularly for near-field imaging, could significantly improve perception, as innate signal characteristics can be leveraged across spatial and spectral domains. Specifically, the dual relationships between RF signals and their spectral representations can be leveraged for superior resolution and image focusing. Towards this end, we investigate a novel hybrid-learning approach for multiband signal fusion to achieve 3-D SAR super-resolution. Compared to traditional signal processing approaches [72, 73, 74, 75, 76, 77, 78, 79, 80], our method enables technologies such as concealed weapon detection and occluded item classification as intricate targets can be recovered with high-resolution. Using this approach, we achieve 21 GHz bandwidth from two 4 GHz bandwidth radars operating at 60–64 GHz and 77–81 GHz. Extensive simulation and empirical experiments are provided to validate the robustness and generalizability of the proposed complex-valued CNN architecture [66]. Hybrid-learning algorithms require careful consideration of the mechanics of the problems but can offer considerable performance gains compared to signal processing or deep learning alone.

## 1.2 Research Objectives and Previous Work

The main objective of this dissertation is to present a framework through which to approach mmWave imaging problems by leveraging the advantages and trade-offs between conventional radar signal processing methods and modern data-driven algorithms. The proposed

technique is denoted as hybrid-learning as a hybrid approach that employs expertise and intuition in both the conventional radar signal processing domain and machine learning arena. To achieve this goal, we focus our efforts on several perception and imaging problems and develop novel data-driven techniques for improved classification, localization, and imaging.

Towards machine learning classification of radar signals, we investigate various front-end signal processing techniques and their impact on perception systems. We explore data preprocessing techniques for static (stationary) and dynamic (moving) hand gestures using mmWave radar, considering both fidelity and computational load. By developing a thorough understanding of the challenges and opportunities in gesture recognition, we demonstrate a novel training technique by employing "sterile" data during the model training process. Additionally, we develop a model to decompose irregular SAR scanning geometries in the near-field to develop an efficient image reconstruction technique that overcomes the excessive computational burden required in previous studies [34, 81, 82].

As near-field SAR image super-resolution is gaining increasing attention [55, 64, 65, 66, 83, 84, 85, 86, 87], there is a significant need for large quantities of meaningful high-fidelity SAR data. To this end, we design an open-source software platform for high-fidelity near-field dataset generation. Similar existing software implementations in the literature [88, 89, 90, 91] address only the simplified far-field scenario and cannot produce data relevant for near-field imaging. Using the custom framework to generate large, meaningful datasets, we consider several methods for improved imaging using hybrid-learning. First, a SAR super-resolution algorithm is detailed to overcome image distortion caused by positioning errors common to near-field mmWave systems [39]. The proposed framework leverages a mobile-friendly Vision Transformer (ViT) architecture [92, 93] for image-to-image super-resolution. Additionally, an image enhancement network is developed to perform spatial super-resolution on near-field SAR images with irregular sampling geometries common to emerging applications [34, 94]. The proposed algorithm is the first of its kind to pair mobile-suitable neural network

architectures and efficient irregular SAR near-field imaging algorithms, thereby enabling several applications constrained to arbitrary scanning patterns and low computational load.

Based on our expertise in radar gesture recognition and imaging, we propose a hybrid-learning framework for contactless musical interface. Our algorithm leverages high-fidelity spatiotemporal signatures embedded in the radar signal to provide a responsive, precise interface for a host of human computer interaction (HCI) tasks. A hybrid deep learning, signal processing, and computer vision approach yields spatial resolution exceeding the theoretical bound and outperforms state-of-the-art localization performance compared with previous methods.

Finally, we propose an end-to-end interleaved hybrid-learning approach for near-field imaging employing deep learning super-resolution and regression techniques [18, 56] throughout the imaging signal chain. Fully embracing hybrid-learning, this technique allows a network to learn the various characteristics of the signal across the spatial and spectral domains. The proposed hybrid-learning algorithms yield significant performance gains for both image fidelity and computational efficiency.

## 1.3   Contributions and Proposed Work

In response to the challenges and opportunities of hybrid-learning algorithms, we present several studies on data-driven solutions to radar imaging problems and propose novel methods for sensing, tracking, imaging, super-resolution, and multiband radar fusion to achieve the following contributions:

1. We investigate static and dynamic gesture recognition using a small-platform MIMO-FMCW mmWave radar and CNN classifiers. We perform an extensive study of the challenges and opportunities for static gesture recognition by examining several datasets and data preprocessing techniques. We explore the trade-offs of CNN classifiers for

dynamic hand gesture recognition. This contribution is based on the following publication:

- J. W. Smith, S. Thiagarajan, R. Willis, Y. Makris, and M. Torlak, "Improved static hand gesture classification on deep convolutional neural networks using novel sterile training technique," *IEEE Access*, vol. 9, pp. 10893–10902, Jan. 2021.

2. We address another aspect of static gesture recognition to improve the robustness of the classifier given the challenges of static gesture recognition. We propose an efficient data collection approach and a novel technique for deep CNN training by introducing "sterile" data which aid in distinguishing distinct features among the static gestures and subsequently improve the classification accuracy. We provide experimental results demonstrating the ability of this method to improve the classification accuracy of real human hand gestures. This contribution is based on the following publication:

- J. W. Smith, S. Thiagarajan, R. Willis, Y. Makris, and M. Torlak, "Improved static hand gesture classification on deep convolutional neural networks using novel sterile training technique," *IEEE Access*, vol. 9, pp. 10893–10902, Jan. 2021.

3. In addition, we extend the work of [34, 82, 95] by proposing a novel imaging algorithm to enable efficient near-field irregular SAR. This work addresses the need for efficient imaging algorithms for edge applications such as smartphone imaging and automotive SAR. Subsequent deep learning for classification or super-resolution requires high-fidelity SAR images under computational constraints. Whereas conventional mmWave imaging relies on high-precision systems [39], many edge applications, such as freehand imaging, necessitate both irregular array geometries and low computational complexity.

The proposed reconstruction algorithm efficiently projects irregularly sampled multi-static data onto a virtual planar monostatic array achieving image resolution consistent with the computationally prohibitive backprojection algorithm (BPA) with equivalent efficiency to the range migration algorithm (RMA). This contribution is founded on the following publication:

- J. W. Smith and M. Torlak, "Efficient 3-D near-field MIMO-SAR imaging for irregular scanning geometries," *IEEE Access*, vol. 10, pp 10283-10294, Jan. 2022.

4. To enable data-driven algorithms for near-field SAR, we develop a novel software framework for system prototyping, imaging algorithm development, and dataset generation. The proposed software is implemented as an open-source MATLAB toolbox capable of efficiently generating high-fidelity SAR data that can be used for a host of applications. The contribution will be based on the following publications:

- J. W. Smith and M. Torlak, "Survey of emerging systems and algorithms for near-field THz SAR imaging," *Proc. IEEE*, to be submitted.

5. To overcome the positioning errors common in many near-field SAR systems [39], we propose a novel Vision Transformer (ViT) approach for SAR image super-resolution and artifact mitigation. Using data generated from the software toolbox, we train our algorithm on images generated from SAR scenarios with image distortion and defocusing caused by array perturbations. The proposed algorithm employs a mobile-friendly image-to-image enhancement architecture [92, 93] suitable for a host of applications from laboratory environments to edge implementations. We validate the proposed method using both simulations and empirical studies. This contribution is detailed in the following publication:

- 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 *Proc. IEEE Tex. Symp. Wirel. Microw. Circuits Syst. (WMCS)*, Waco, TX, Apr. 2022, pp. 1–6.

6. We then propose extending our work in [94] to develop the first CNN-based SAR super-resolution algorithm for mobile applications. Emerging applications for mobile SAR imaging in the near-field are constrained to irregular sampling geometries and low computational complexity. Previous studies on near-field SAR super-resolution algorithms are limited to conventional SAR geometries and are unsuitable for mobile applications [55, 86]. Using the software toolbox we developed, we generate large synthetic datasets to train a neural processor to perform 3-D SAR image super-resolution. The proposed algorithm employs a generative adversarial network (GAN) architecture for image super-resolution using a patch discriminator technique [96] and efficient depthwise convolution implementation [97]. A thorough discussion of this contribution is provided in the following publication:

   - C. Vasieleiou, J. W. Smith, S. Thiagarajan, M. Nigh, Y. Makris, and M. Torlak, "Efficient CNN-based super resolution algorithms for mmWave mobile radar imaging," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Bourdeaux, France, Oct. 2022, pp. 3803–3807.

7. Additionally, we developed a novel framework for human-computer interaction using a fully convolutional neural network (FCNN) for localization super-resolution in real-time. Our system offers unprecedented high-resolution tracking of hand position and motion characteristics by leveraging spatial and temporal features embedded in the reflected radar waveform. By employing a hybrid-learning approach, we developed a novel spatial super-resolution technique that exceeds the theoretical limitations and a

modified tracking algorithm to optimally leverage the inherent characteristics of radar signatures. This contribution is based on the following publication:

- J. W. Smith, O. Furxhi, M. Torlak, "An FCNN-based super-resolution mmWave radar framework for contactless musical instrument interface," *IEEE Trans. Multimedia*, vol. 24, pp. 2315–2328, May 2021.

8. Finally, we propose a novel hybrid-learning technique for multiband radar image fusion. Using off-the-shelf 4 GHz bandwidth radars at 60-64 GHz and 77-81 GHz, we develop a high-fidelity testbed for collecting multiband radar images. The proposed algorithm achieves an effective bandwidth of 21 GHz and outperforms previous methods, particularly on high-bandwidth targets, in terms of image fidelity and computation time. By leveraging a novel dual-domain architecture, the proposed hybrid-algorithm demonstrates super performance compared to conventional techniques [74, 75, 79] in both simulation and empirical studies. This contribution is based on the following publication:

- J. W. Smith and M. Torlak, "Deep learning-based multiband signal fusion for 3-D SAR super-resolution," in *IEEE Trans. Aerosp. Electron. Syst.*, Apr. 2023.

Through these investigations, we also developed an advanced imaging system and a novel algorithm for near-field cylindrical MIMO-ISAR, which appeared in the following publication

- J. W. Smith, M. E. Yanik and M. Torlak, "Near-field MIMO-ISAR millimeter-wave imaging," *Proc. IEEE Radar Conf. (RadarConf)*, Florence, Italy, Sep. 2020, pp. 1-6.

## 1.4   Outline of the Dissertation

The rest of the dissertation is organized as follows:

- Chapter 2 details the FMCW signal model employed at length throughout this dissertation.

- Chapter 3 presents an investigation of the impact of front-end signal processing techniques on deep learning perception algorithms and details a novel training technique for gesture sensing using mmWave radar.

- Chapter 4 investigates the inclusion of data-driven approaches in the near-field imaging pipeline; details the development and implementation of a software framework for near-field SAR imaging simulation, prototyping, and dataset generation; and presents near-field SAR image super-resolution and restoration algorithms using hybrid-learning techniques.

- Chapter 5 details a real-time deep learning-based framework for contactless musical interface using mmWave radar.

- Chapter 6 presents a multiband radar imaging system built from off-the-shelf 60 GHz and 77 GHz radars and a multiband fusion algorithm that leverages a novel hybrid-learning, dual-domain technique to provide an equivalent bandwidth of 21 GHz from the two 4 GHz bandwidth radars.

- Conclusions, summary, and discussion of proposed work are detailed in Chapter 7.

# CHAPTER 2

# PRELIMINARIES OF FMCW SIGNALING

In this chapter, we detail the fundamentals of frequency-modulated continuous-wave (FMCW) radar to be utilized extensively throughout this dissertation. Over the past several decades, FMCW radars have emerged as an inexpensive option for high-bandwidth systems [39]. FMCW signals contain precise spatial information of the illuminated target and are used for a wide array of applications from gesture recognition [2, 4, 7], concealed threat detection [8, 9], and medical imaging [17]. Throughout this dissertation, we will leverage the characteristics of FMCW signaling for high-fidelity perception and imaging.

## 2.1 FMCW Signal Model

We begin by considering a single bistatic FMCW transceiver, whose transmitter and receiver are positioned at the points $(x_T, y_T, Z_0)$ and $(x_R, y_R, Z_0)$ in a three-dimensional (3-D) space, respectively, and one stationary ideal point reflector in the scene with reflectivity $\sigma$ located at the point $(x_0, y_0, z_0)$. The radar transceiver is positioned on the $x'$-$y'$ plane, located at $z = Z_0$.
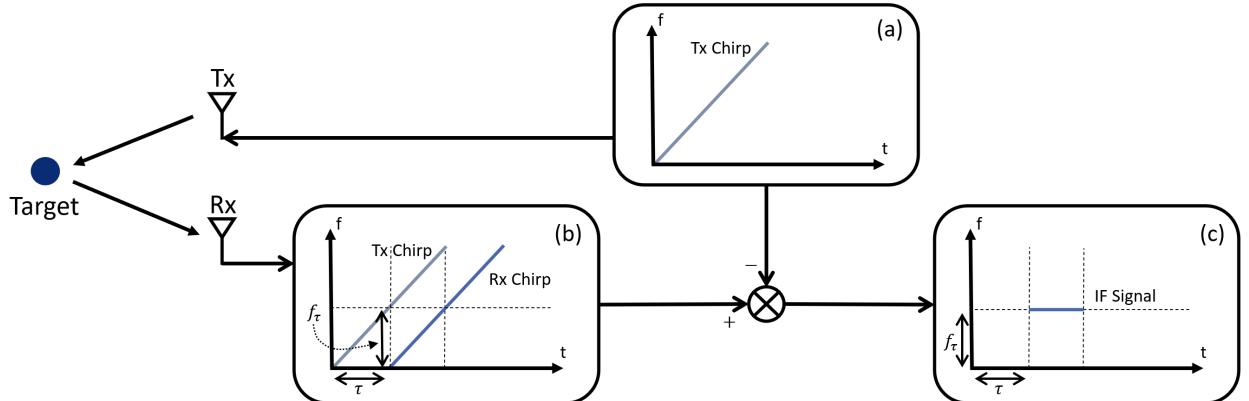


Figure 2.1: FMCW signal chain. (a) FMCW pulse generation. (b) Received signal from single ideal point scatterer. (c) FMCW beat signal after dechirping.

As shown in Fig. 2.1, the FMCW device first generates what is known as a chirp signal, which can be modeled as a complex sinusoidal signal whose frequency increases linearly with time as

$$m(t) = e^{-j2\pi(f_0 t + \frac{1}{2}Kt^2)}, \quad 0 \le t \le T, \tag{2.1}$$

where $f_0$ is the instantaneous frequency at time $t = 0$, $K$ is the chirp slope, and $T$ is the chirp duration. The chirp bandwidth can be computed easily using $B = KT$ [9, 46, 98].

The chirp signal $m(t)$ is transmitted by the transmit antenna, reflects off the ideal point reflector, and returns to the receive antenna as a scaled and time-delayed version of the transmitted signal. Consider the round-trip amplitude decay, the received signal can be modeled as

$$\hat{m}(t) = \frac{\sigma}{R_T R_R} m(t - \tau) = \frac{\sigma}{R_T R_R} e^{-j2\pi(f_0(t-\tau) + \frac{1}{2}K(t-\tau)^2)}, \tag{2.2}$$

where $\tau$ is the round-trip time delay [99] and the values $R_T$ and $R_R$ (see Fig. 2.1) are given by

$$R_T = \left[(x_0 - x_T)^2 + (y_0 - y_T)^2 + (z_0 - Z_0)^2\right]^{\frac{1}{2}}, \tag{2.3}$$

$$R_R = \left[(x_0 - x_R)^2 + (y_0 - y_R)^2 + (z_0 - Z_0)^2\right]^{\frac{1}{2}}. \tag{2.4}$$

Therefore, the round trip time delay $\tau$ can be computed by

$$\tau = \frac{R_T + R_R}{c}, \tag{2.5}$$

where $c$ is the speed of light.

The received signal $\hat{m}(t)$ is demodulated with the transmitted signal $m(t)$ yielding what is known as the IF signal or FMCW beat signal, written as

$$s_0(t) = \frac{\sigma}{R_T R_R} e^{-j2\pi(f_0\tau + K\tau - \frac{1}{2}K\tau^2)}. \tag{2.6}$$

The last phase term in (2.6) is called the residual video phase (RVP) term and is known to be negligible [51]. Finally, the beat signal can be simplified to the expression

$$s_0(x_T, x_R, y_T, y_R, k) = \frac{\sigma}{R_T R_R} e^{-jk(R_T + R_R)}, \tag{2.7}$$

where $k = 2\pi f/c$ is the wavenumber corresponding to the instantaneous frequency $f = f_0 + Kt$ for $t \in [0, T]$.

The continuous-time signal (2.7) is sampled with sampling frequency $f_S$ by the radar analog-to-digital converter (ADC) and can be written in discrete time as

$$s(x_T, x_R, y_T, y_R, n_k) = \frac{\sigma}{R_T R_R} e^{-j(k_0 + \Delta_k n_k)(R_T + R_R)}, \tag{2.8}$$

where $n_k$ is the wavenumber index, $k_0 = 2\pi f_0/c$ is the starting wavenumber corresponding to the starting frequency $f_0$, and $\Delta_k = 2\pi K/(cf_S)$ is the wavenumber step size.

To ease the subsequent signal processing, it is desirable to approximate the multistatic MIMO beat signal, represented in (2.8) as its corresponding monostatic equivalent using the approximation developed in [39, 46, 51] as

$$\hat{s}(x', y', n_k) = s(x_T, x_R, y_T, y_R, n_k) e^{j(k_0 + \Delta_k n_k)\frac{d_x^2 + d_y^2}{4\tilde{Z}}}, \tag{2.9}$$

valid only for small values of $d_x$ and $d_y$, the distances between the transmitter and receiver elements along the $x$- and $y$-directions, respectively, where $\tilde{Z}$ is a reference plane typically given as the center of the target scene. Taking $(x', y', Z_0)$ as the location of the virtual element located at the midpoint between the transceiver pair, as shown in Fig. 2.2a, and $R_0$ as the corresponding distance from the virtual element to the point reflector, the resulting monostatic beat signal is approximately

$$\hat{s}(x', y', n_k) \approx \frac{p}{R_0^2} e^{-j2(k_0 + \Delta_k n_k)R_0}. \tag{2.10}$$

From (2.10), the spatial location, $(x_0, y_0, z_0)$, of the target is embedded in frequency of the radar beat signal, in the form of the radial distance $R_0$, which can be expressed as

$$R_0 = \left[(x_0 - x')^2 + (y_0 - y')^2 + (z_0 - Z_0)^2\right]^{\frac{1}{2}}. \tag{2.11}$$

An example geometry is given in Fig. 2.2a, with a single transceiver pair located at $Z_0 = 0$ m and a point scatterer at $(0, 0, 0.5)$. From (2.10), the FMCW beat signal, shown

13

in Fig. 2.2b, is a single tone sinusoidal signal. Taking the Fourier transform of the FMCW beat signal yields the range profile, which shows a dominant peak at the distance from the radar to the point scatterer.
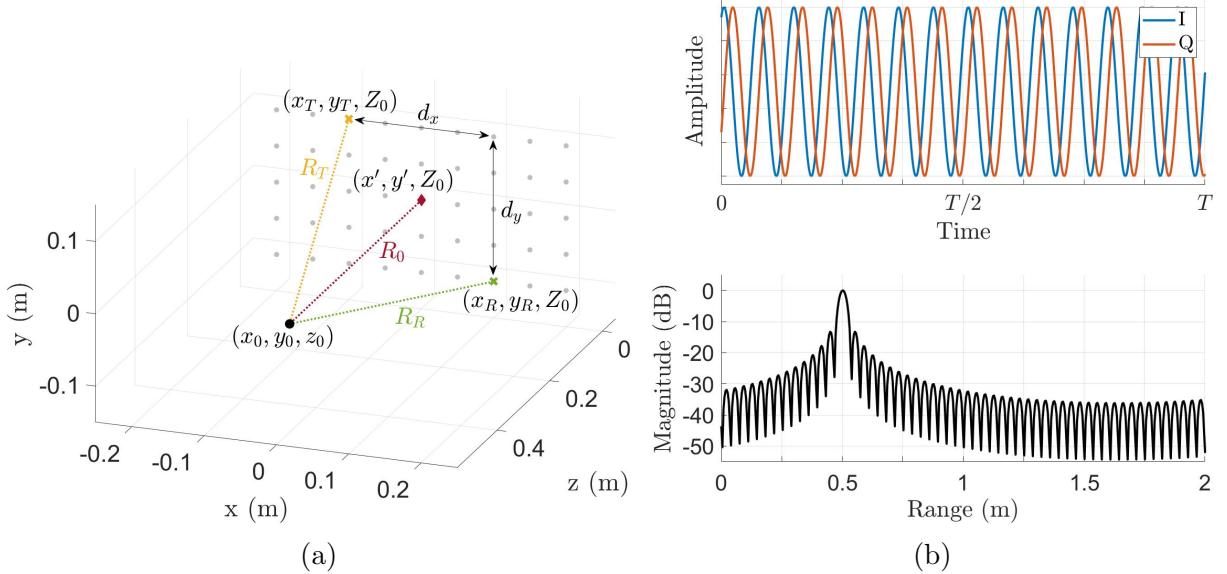


Figure 2.2: (a) Geometry for near-field FMCW radar scenario with Tx and Rx elements located at $(x_T, y_T, Z_0)$ and $(x_R, y_R, Z_0)$, respectively, and a point scatterer at $(x_0, y_0, z_0)$, using $Z_0 = 0$ m and $z_0 = 0.5$ m. (b) FMCW beat signal (top) and corresponding spectral representation using range Fourier transform.

## 2.2   Range-Doppler Processing

The relative velocity of a target can be extracted from the beat signal expressed in (2.10) by exploiting the Doppler effect. As discussed in [26], by transmitting a series of chirp waveforms at a known pulse repetition interval (PRI), $T_{PRI}$, the velocity of a moving target can be identified as the frequency component along the chirp index dimension given by

$$\hat{s}(x', y', n_k, n_c) = \frac{p}{R_0^2} e^{-j(2(k_0 + \Delta_k n_k) R_0 + \frac{4\pi v T_{PRI}}{\lambda_0} n_c)},  \tag{2.12}$$

where $R$ is the initial range of the target, $v$ is the velocity of the target, $\lambda_0$ is the wavelength corresponding to $f_0$, and $n_c$ is the chirp index,

14

Thus, the beat signal sampled across time is a 2-D complex sinusoidal signal with frequencies corresponding to the range and velocity of the target in the first and second dimensions, respectively. Subsequently, to extract the range and velocity, traditional methods perform a 2-D fast Fourier transform (FFT) over a matrix whose rows or columns consist of subsequent chirps. This analysis is known as range-Doppler processing and is commonly applied to many radar signal processing problems [29, 100].

## 2.3 FMCW Response to Distributed Target

Assuming a distributed target occupying volume $V$ in Cartesian $x$-$y$-$z$ space and the same transceiver pair discussed in Section 2.1, the FMCW beat signal can be expressed as

$$s(x_T, x_R, y_T, y_R, k) = \iiint_V \frac{p(x, y, z)}{R_T R_R} e^{-jk(R_T + R_R)} dx dy dz, \tag{2.13}$$

where $p(x, y, z)$ is known as the reflectivity function of the target representing the intensity of reflection from each point of the target throughout volume $V$ and $R_T$ and $R_R$ are the radial distances from the target to the transmitter and receiver, respectively, as

$$R_T = \left[ (x - x_T)^2 + (y - y_T)^2 + (z - Z_0)^2 \right]^{\frac{1}{2}}, \tag{2.14}$$

$$R_R = \left[ (x - x_R)^2 + (y - y_R)^2 + (z - Z_0)^2 \right]^{\frac{1}{2}}. \tag{2.15}$$

By applying the multistatic-to-monostatic conversion in (2.9) [51], the virtual monostatic response can be written as

$$s(x', y', k) = \iiint_V \frac{p(x, y, z)}{R_0^2} e^{-j2kR_0} dx dy dz, \tag{2.16}$$

where $R_0$ is the distance between the virtual monostatic element and the target as

$$R_0 = \left[ (x - x')^2 + (y - y')^2 + (z - Z_0)^2 \right]^{\frac{1}{2}}. \tag{2.17}$$

The virtual monostatic response can be written in discrete-time as

$$s(x', y', n_k) = \iiint_V \frac{p(x, y, z)}{R_0^2} e^{-j2(k_0 + \Delta_k n_k)R_0} dx dy dz. \tag{2.18}$$

15

In many applications, it is desirable to extract the reflectivity function $p(x, y, z)$ from the radar beat signal. This process is known as imaging and requires inversion of the integral in (2.13). However, to achieve this, the radar must be sampled throughout space by utilizing a large array of radar transceivers, known as real array radar (RAR) [101], or the concept of synthetic aperture radar (SAR), in which a small radar platform is scanned throughout space to synthesize a larger array. In this dissertation, orthogonality is leveraged across time by operating the MIMO radar using the time-division multiplexing (TDM) MIMO technique such that each Tx/Rx pair is activated sequentially. Hence, the MIMO-SAR operation involves performing TDM-MIMO at each location in space, but involves its own challenges [39, 51]. High-resolution near-field SAR and MIMO-SAR imaging algorithms and systems for multiple modalities are discussed throughout this dissertation.

FMCW signaling enables low-cost ultra-wideband radar systems for a host of applications. With precise spatial information embedded in the frequency content of the signal, FMCW radars are suitable for many sensing tasks. Extracting and leveraging the spatial information for applications such as classification and perception will be addressed at length in the remainder of this dissertation. Additionally, as radars operate with limited bandwidth, causing a sinc effect in the range domain, improving spatial resolution using machine learning techniques is a promising solution to overcoming system limitations. Throughout this dissertation, we will introduce novel techniques using signal processing, machine learning, and hybrid-learning algorithms leveraging the characteristics of FMCW signals.

# CHAPTER 3

# IMPACT OF FRONT-END SIGNAL PROCESSING TECHNIQUES ON DEEP LEARNING PERCEPTION

In this chapter, we explore various front-end signal processing techniques for improving perception using data-driven algorithms. We investigate signal processing algorithms to extract and process spatiotemporal signatures embedded in the FMCW radar signals, as discussed in Chapter 2. Applications including gesture recognition, SAR image segmentation for concealed weapon detection, and SAR image super-resolution using deep learning display variable performance depending on the signal processing techniques applied to the data prior to the learning algorithm. Here, we explore the impact of front-end signal processing methods by optimizing the fidelity and computational load of hybrid-learning algorithms for perception and imaging. Part of the following work was previously published in [7][1] and [94][2] and will be presented in [102].

## 3.1 Gesture Recognition with mmWave Radar

Accurately classifying human hand gestures has recently received significant attention as non-contact human-computer interaction (HCI) sensors have become increasingly prevalent and desirable. Many efforts have been made to classify moving (dynamic) hand gestures and non-moving (static) hand gestures using optical cameras and many different classifiers [103]. Applications of static gesture classification include augmented/virtual reality (AR/VR) [33], human-computer interaction [104], and even medical applications for range of motion and

---

[1]©2021 IEEE. Reprinted, with permission, from J. W. Smith, S. Thiagarajan, R. Willis, Y. Makris, and M. Torlak, "Improved static hand gesture classification on deep convolutional neural networks using novel sterile training technique," *IEEE Access*, vol. 9, pp. 10893–10902, Jan. 2021.

[2]©2022 IEEE. Reprinted, with permission, from J. W. Smith and M. Torlak, "Efficient 3-D near-field MIMO-SAR imaging for irregular scanning geometries," *IEEE Access*, vol. 10, pp. 10283-10294, Jan. 2022.

therapeutic applications [105]. Such optical systems offer high-resolution two-dimensional (2-D) images but have innate drawbacks, such as requiring specific lighting conditions and lacking depth information. Some solutions have investigated the use of an RGB-D depth camera [32], but these devices suffer from sunlight, restricting their usage to exclusively indoor applications [33]. On the other hand, small form-factor mmWave frequency-modulated-continuous-wave (FMCW) radar offers high-resolution depth information but does not have the cross-range resolution of an optical camera. mmWave radars are advantageous over optical solutions because of the semi-penetrative nature of EM radiation at wavelengths in the mmWave frequency range and independence from ambient temperature effects, allowing for fine measurements in non-ideal lighting and temperature environments including occlusion, fog, indoor/outdoor, etc. Additionally, FMCW mmWave radars enable simultaneous gesture classification and localization. High-resolution spatial information reflected from the human hand is embedded in the FMCW return signal. However, owing to the nature of FMCW radar as a time-of-flight (ToF) sensor and hardware size limitations, an off-the-shelf radar device cannot reconstruct an image reminiscent of the human hand or meaningful to the human eye without employing time-consuming SAR techniques. Thus, a deep convolutional neural network (CNN) approach is adopted to classify dynamic gestures from radar return signals [4].

### 3.1.1 Static Gesture Recognition with mmWave Radar

In this section, we explore the application of hybrid-learning algorithms to the static (stationary) gesture recognition problem. Rather than employing SAR or ISAR to capture images of the hand from many locations using imaging algorithms to recover the reflectivity, we propose classifying a hand gesture using only a single, stationary MIMO-FMCW radar. Similar applications have been employed in commercial products for gesture-based HCI. The most notable example is the use of a 60 GHz radar in the Google Pixel 4 [106]. Fig. 3.1 shows
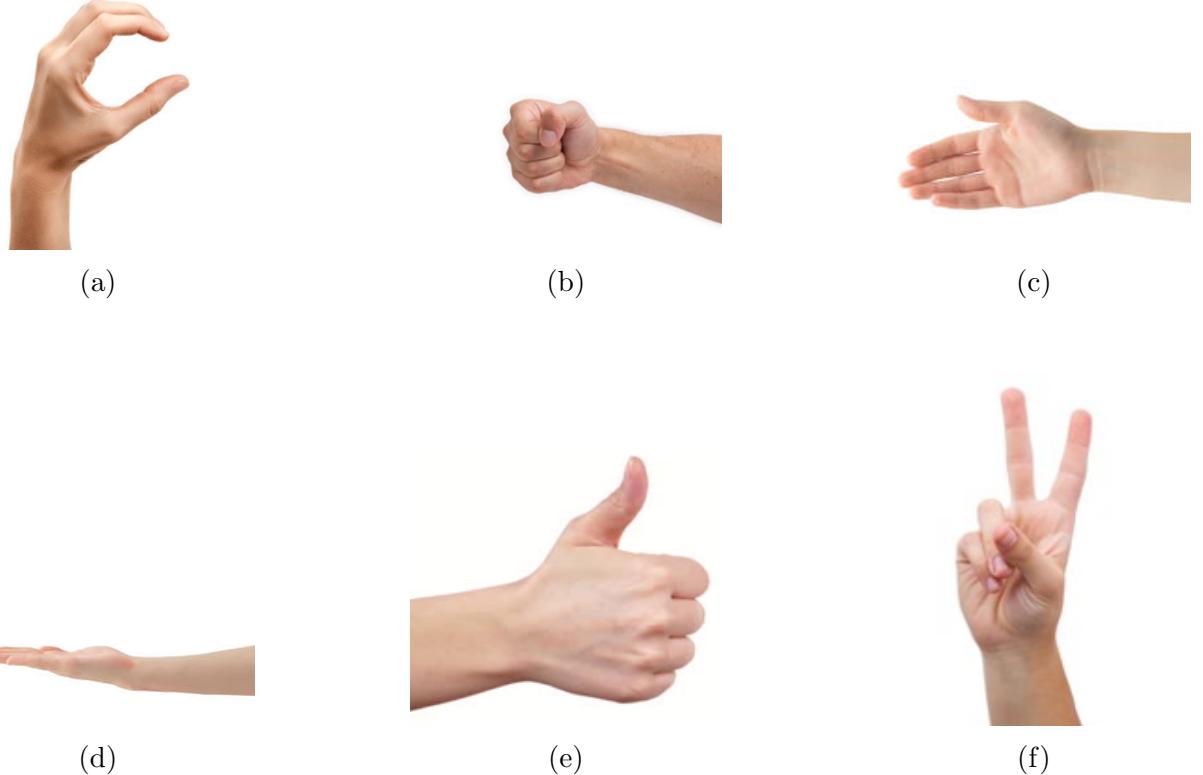
Figure 3.1: Static hand gestures: (a) "c", (b) "fist", (c) "palm", (d) "perpendicular", (e) "thumbs up", (f) "two".

the six gestures employed in the following studies, which shed light on the various degrees of complexity and opportunities for innovation in mmWave radar hand gesture recognition.

Data for subsequent experiments were collected from a diverse set of five participants. The subjects were instructed to position their hand angled forward, backward, left, and right by 30°, resulting in nine Degrees of Freedom (DoF) at each sampling location.

First, a dataset was collected at a single location 50 cm from the radar boresight. Each gesture was captured with the 9 DoFs detailed previously, resulting in a total of 2250 gestures per class, with six gesture classes, as shown in Fig. 3.1. After the preliminary results were promising, a diverse dataset is collected using the setup in Fig. 3.2a.
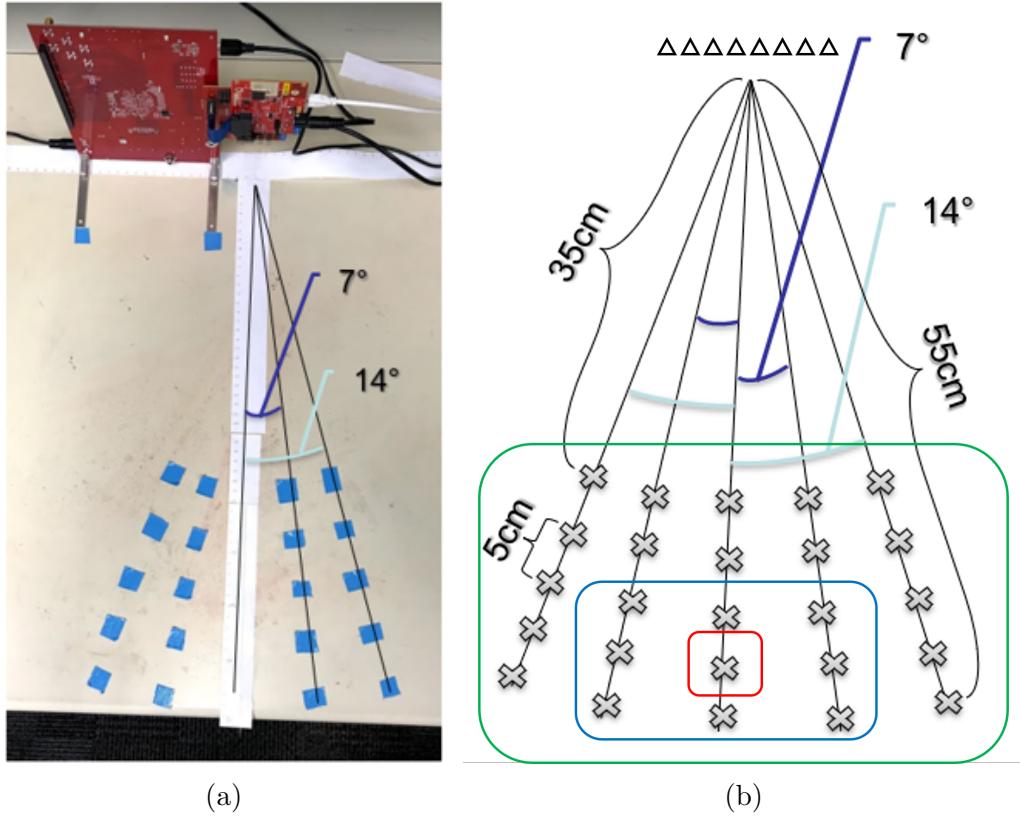
Figure 3.2: (a) Hardware setup consisting of a TI 77 GHz radar and visual indicators for the test subjects. (b) Locations of data captured for each dataset: Dataset 1, Dataset 2, Dataset 3.

As indicated by the green box in Fig. 3.2b, Dataset 2 comprises 25 locations from 35–55 cm, spanning a 28° field of view (FOV), with a total of 4500 captures per class. Finally, a third dataset was collected within a smaller region, indicated by the blue box in Fig. 3.2b, spanning 45–55 cm and a 14° FOV. A summary of these datasets is presented in Table 3.1

Table 3.1: Summary of static hand gesture datasets.

|  | Dataset 1 | Dataset 2 | Dataset 3 |
|---|---|---|---|
| Captures/class | 2250 | 4500 | 405 |
| Ranges | 50 cm | 35 cm − 55 cm | 45 cm − 55 cm |
| FOV | 0° | 28° | 14° |

The data are collected using a Texas Instruments (TI) AWR1443BOOST radar with 4 GHz bandwidth from 77 GHz to 81 GHz is mounted on a TI mmWave-Devpack and TSW1400 data capture card to store the data and transfer it to the PC, where the samples are manipulated in MATLAB. The TI AWR1443BOOST is equipped with a MIMO array consisting of two Tx elements spaced by $2\lambda_c$, one Tx element vertically displaced by $\lambda/2$, and four Rx elements spaced by $\lambda_c/2$ [51], as shown in Fig. 3.3. By orienting the radar in the horizontal direction, a virtual array exists consisting of a row of eight antennas underneath a row of four antennas, as shown in Fig. 3.3. It should be noted that although the radar setup is mounted on a desk, the reflections of the desk are negligible, according an empirical study, owing to the narrow beamwidth of the radar along the vertical direction.
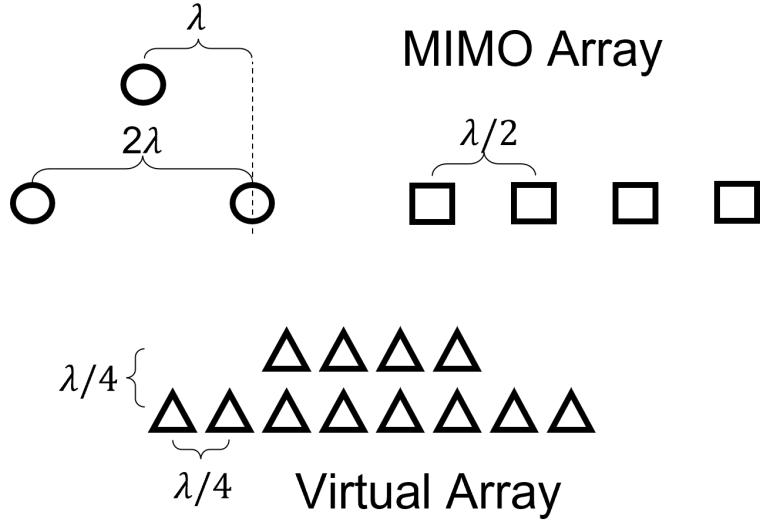


Figure 3.3: TI AWR1443BOOST MIMO antenna array and virtual monostatic array.

After the datasets were collected, preprocessing techniques are applied to investigate the optimal presentation of data to a CNN. CNNs of varying dimensionality are implemented with three hidden layers consisting of a convolution layer with kernel sizes of 5, $5 \times 5$, or $5 \times 5 \times 5$, a batch normalization layer, and a Rectified Linear Unit (ReLU) [19]. After the three convolution layers, a fully connected layer is employed for the six classes and cross-entropy

loss is used to train the network using an ADMM optimizer. The real and imaginary parts of the sample were layered to leverage the signatures embedded in the phase of the data.

64 samples are taken over the 4 GHz bandwidth of the radar; hence, each sample is an array of size $64 \times 12$, owing to the 12 virtual channels. To obtain a baseline, we apply a simple 1-D CNN to the samples of Dataset 1 vectorized as $768 \times 1$ vectors. Even for a simple set of data in Dataset 1, the classification rate is 83%. However, given the underlying mechanics of the problem and format of the data, the samples are not presented to the network in a meaningful way. A range-FFT is performed across the first dimension of the 64 $\times$ 12 array, as described in Section 2.1. After selecting the range bins of interest, a process known as "range-gating," a 2-D CNN is trained using the range-FFT data from Dataset 1, yielding a classification accuracy of 95%.

Here, we note the behavior of the mmWave radar gesture data in the range and angle domains. Two sample range-FFT spectra are computed from random data points selected from the "c" and "fist" classes with the hand at 45 cm and shown in Fig. 3.4. Because the frequency content of the FMCW signal corresponds to the range of the targets, the range-FFT spectrum represents the magnitude (and phase) of the reflection of the target at a given distance. As expected, a significant reflection is observed at approximately 1 m because of the human torso. However, while a reflection from the hand is visible around 45 cm, there is no distinguishing characteristic, to the human eye between the two classes.

Next, an autocorrelation strategy is applied to the radar data using the eight collinear channels. The 2-D autocorrelation matrix is computed from the $25 \times 8$ range-FFT array. The autocorrelation method leverages the spatial relationships between adjacent channels to provide a more learnable representation of the samples; however for Dataset 1, the classification rate remained at 95% using this approach.

Finally, an angle-FFT technique is applied to the samples along with the range-FFT, referred to as "range-angle FFT." After the range-FFT is performed, the sample is rearranged
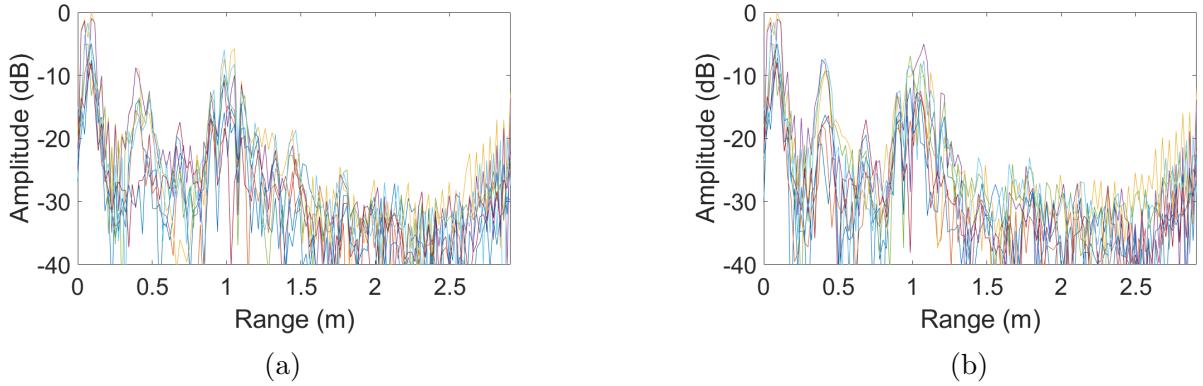
Figure 3.4: (a) Range-FFT of a "c" sample with the hand at 45 cm. (b) Range-FFT of a "fist" sample with the hand at 45 cm.

and zero-padded according to the geometry in Fig. 3.3, with a size of $25 \times 8 \times 2$. An angle-FFT of size 16 is performed across the second dimension yielding a data cube of size $25 \times 16 \times 2$. A 3-D CNN is trained using the range-angle-FFT data, yielding a classification accuracy of 99% for Dataset 1.

Similarly, in Fig. 3.5, we examine the 2-D range-angle-FFT spectra of the same two samples, as shown in Fig. 3.4. We again note the reflection of the torso at approximately 1 m and the hand at approximately 45 cm, close to the center of the FOV; however, the reflections have negligible meaning to the naked eye.
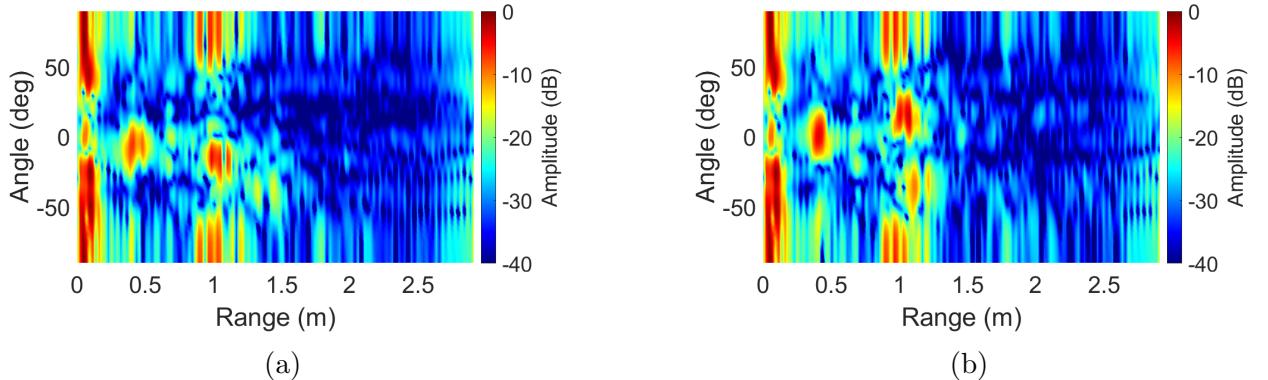


Figure 3.5: (a) Range-Angle-FFT of a "c" sample with the hand at 45 cm. (b) Range-Angle-FFT of a "fist" sample with the hand at 45 cm.

Similarly, the aforementioned preprocessing techniques are applied to Datasets 2 and 3, and the results are summarized in Table 3.2. As expected, Dataset 2, which is the most diverse dataset, is the most difficult to classify. Upon closer inspection, because the test subjects are seated in front of the radar with their hand in front of them, many of the samples in Dataset 2 do not contain a meaningful reflection from the hand as the hand reflection is obscured by the sidelobes from the much stronger torso reflection. Hence, a nulling strategy is employed to project the collected data onto the null space of the peak along the range-FFT corresponding to the torso. However, this method yields a minimal increase in the classification rate for Dataset 2 and reduces the classification accuracy for Datasets 1 and 3. This phenomenon is likely due to the nulling process unintentionally reducing the learnable information about the hand gesture and the proposed nulling procedure is discarded.

However, the autocorrelation and range-angle-FFT methods yield performance increases for all datasets. The autocorrelation technique results in classification rates of 62% and 87% for Datasets 2 and 3, respectively. The range-angle-FFT strategy results in classification rates of 75% and 91% for Datasets 2 and 3, respectively.

Table 3.2: Classification results for various preprocessing techniques of static gesture data.

| | Dataset 1 | Dataset 2 | Dataset 3 |
|---|---|---|---|
| Raw (Vectorized) | 83% | - | - |
| Reformatted Range-FFT | 95% | 61% | 86% |
| Nulled | 91% | 61% | 80% |
| Autocorrelation | 95% | 62% | 87% |
| Range-Angle-FFT | 99% | 75% | 91% |

### 3.1.2 Dynamic Gesture Recognition with mmWave Radar

Similarly, a study is conducted on dynamic (moving) gestures to investigate the impact of data presentation on classification rate. Five dynamic hand gestures are employed, as shown in Fig. 3.6, requiring the user to move their hand in a circle around the boresight of the

radar, push towards the radar, pull away from the radar, wave at the boresight of the radar, or perform the University of Texas at Dallas "whoosh" spirit symbol, pulling their hand from their waist to face level. Five test subjects collect a single dataset while seated at a distances of 1 m from the radar, consisting of 600 captures per class. Each capture consists of 512 FMCW pulses, known as frames, across 2.56 s; hence, the pulse repetition interval (PRI) is 5 ms.



(a)  (b)  (c)

(d)  (e)
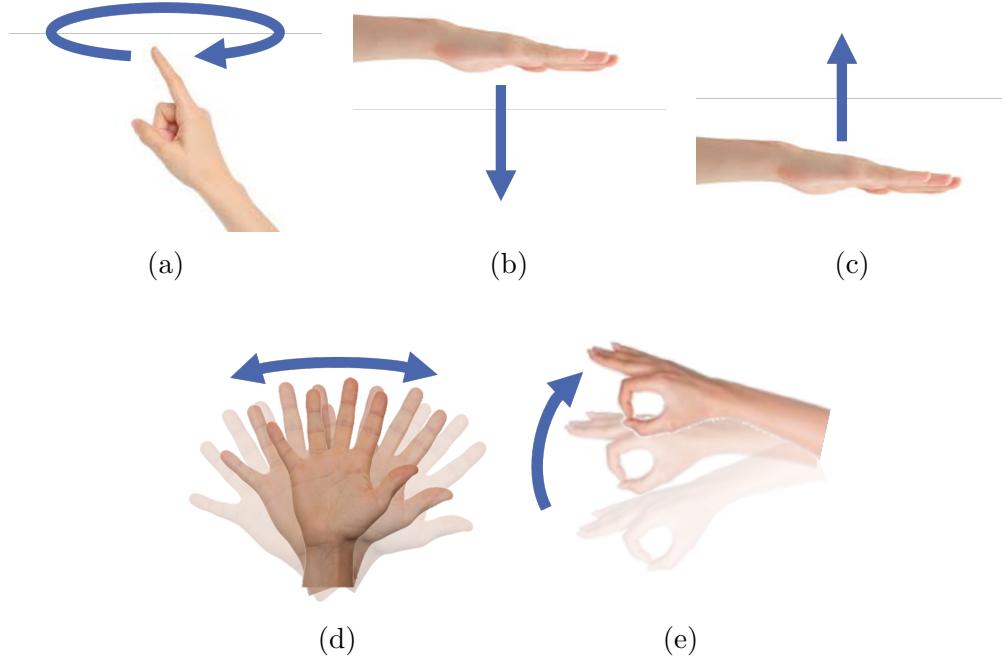
Figure 3.6: Dynamic hand gestures: (a) "circle", (b) "push", (c) "pull", (d) "wave", (e) "whoosh".

The additional dimension of time allows for several new ways of presenting data to the CNN classifier. First, the conventional range-FFT and range gating are applied to the region of interest in which the hand and torso are both located. After the range-FFT, the network can be trained on the range-time data or range-Doppler data, using the Doppler-FFT detailed in Section 2.2. Alternatively, as is commonly employed in speech processing, the short-time Fourier Transform (STFT) can be applied along the time dimension to yield a velocity versus time mapping of the data [107]. In this sense, the CNN will observe the velocity as it changes

across the 2.56 s of the capture, as shown in Fig. 3.7. However, the Doppler-STFT increases the dimensionality of the problem necessitating greater computational power.
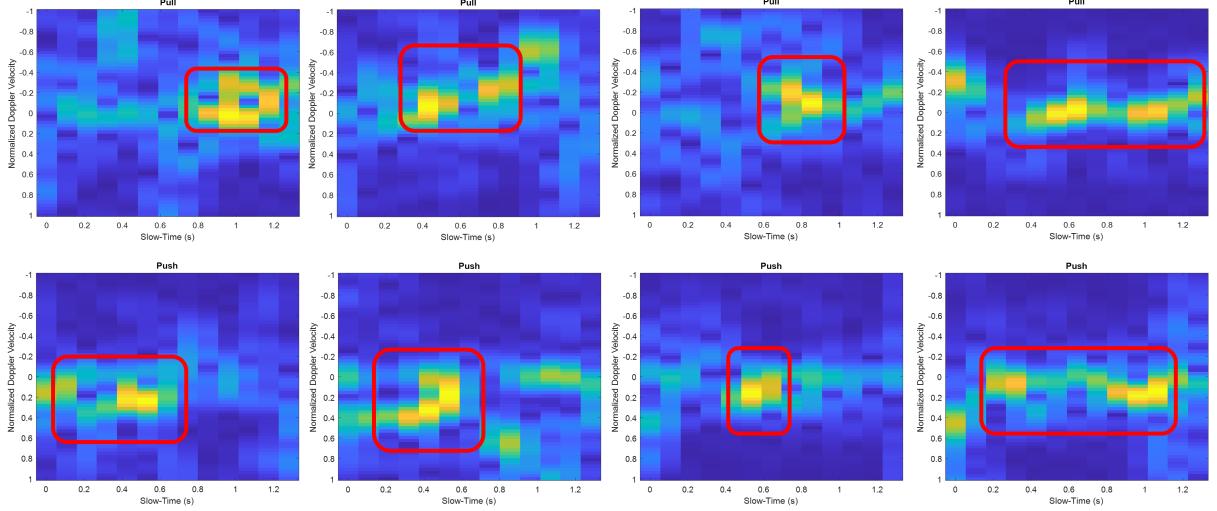


Figure 3.7: Doppler-STFT for pull and push gestures demonstrating where the gesture is performed during the 2.56 s capture.

As expected, given the considerable differences among the gestures over time, the classifier outperforms the static gesture case, in terms of classification accuracy. All combinations of the following preprocessing techniques are compared to evaluate the performance of the CNN: range-FFT, angle-FFT, Doppler-FFT/Doppler-STFT, x8 downsampled in time, 12 channels, and only 1 channel. The x8 downsampling operation is employed to compare the relative classification and computational performance if the gesture is sampled less frequently across time. Utilizing only 1 channel rather than the full 12 channels reduces the dimensionality of the classifier and hence the computational load. Since the most notable variation between classes is in the range-time or range-Doppler domains, employing only a single channel may still capture enough information for robust classification. A comparison of the classification accuracy and required computation time is provided in Fig. 3.8, where the "Efficiency Score," $\eta$, is computed by

$$\eta = 20 \log_{10} \frac{\alpha}{T}, \tag{3.1}$$

where $\alpha$ is the classification accuracy and $T$ is the computation time.

Based on this analysis, the most efficient classifier is the range-Doppler, with a filter in the Doppler domain to only velocities near zero, using only 1 channel. From this result, we can infer that the most meaningful parameters for the neural network to learn are along the range-Doppler domains and the spatial/channel domain offers little insight into classifying dynamic gestures. Additionally, although the same information is present in the range-FFT and range-Doppler-FFT signals, the CNN learns noticeably different features that have a significant impact on algorithm performance.
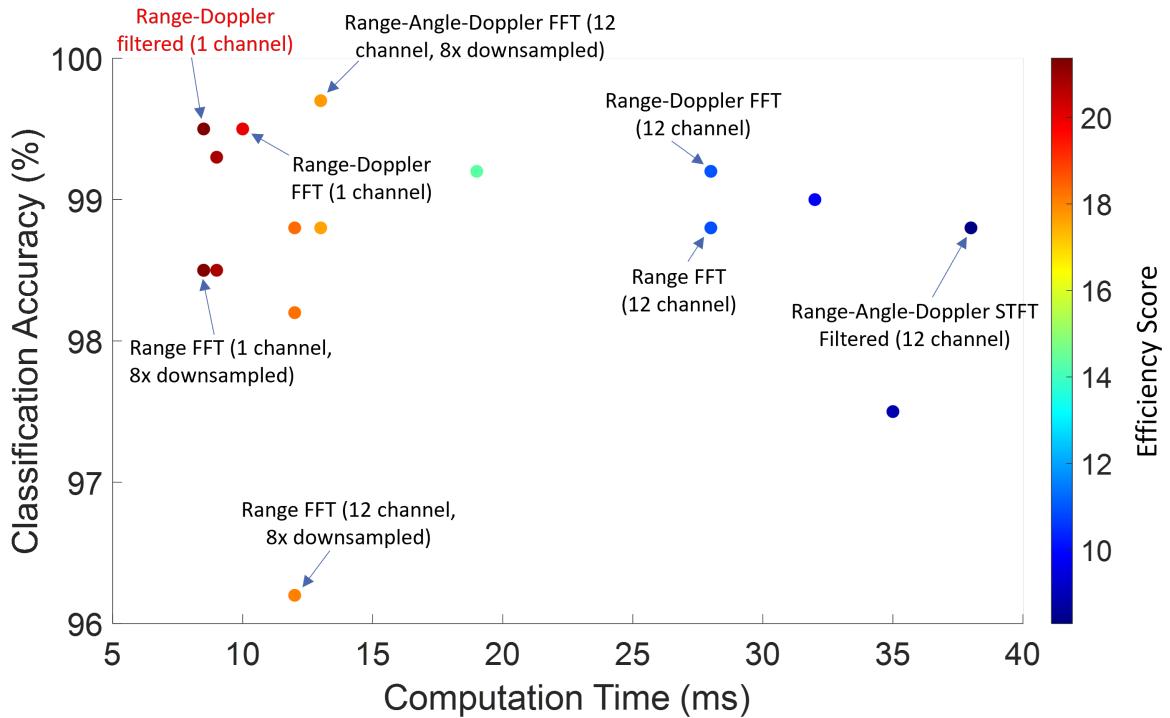


Figure 3.8: Doppler STFT for pull and push gestures demonstrating where the gesture is performed during the 2.56 s capture.

## 3.2 Improved Static Gesture Classification using Novel Sterile Training Technique

Upon closer inspection of the mechanics of the problem, previous results on static and dynamic gesture classification become more apparent. For gesture recognition, a human hand can be mathematically modeled as a distributed target consisting of a continuously varying reflectivity across space. Understanding how radar captures such target scenes provides insight into the difficulty of hand gesture recognition using mmWave signaling.

Assuming a simple linear MIMO array along the $y$-axis, such as the depiction in Fig. 3.9a, and applying the multistatic to monostatic conversion in (2.9), the return signal from a distributed target can be modeled as the superposition of the echo signals from each of the target coordinates scaled by the target's reflectivity function $\sigma(x, y, z)$. The beat signal from each virtual monostatic transceiver at the positions $y'$ can be expressed as

$$s(y', k) = \iiint \frac{p(x, y, z)}{R^2} e^{-j2kR} dx dy dz. \tag{3.2}$$

where $R$ is the radial distance from each virtual monostatic element located at the positions $y'$ to each point in the distributed target domain as

$$R = \left[ x^2 + (y - y')^2 + (z - Z_0)^2 \right]^{\frac{1}{2}}. \tag{3.3}$$

If samples are taken throughout the $x'$-$y'$ plane, the reflectivity function can be reconstructed by inverting (3.2); however, for applications such as hand gesture recognition, the transceiver elements span only a small space along the $y'$-axis. This model provides insight into the simultaneous plausibility and difficulty of the static gesture recognition problem using FMCW radar.

Embedded in the beat signal are high-resolution spatial features describing the shape of the target or static gesture being performed, meaning that different hand poses or static gestures have distinct echo signals unique to that gesture. However, the target scene or hand
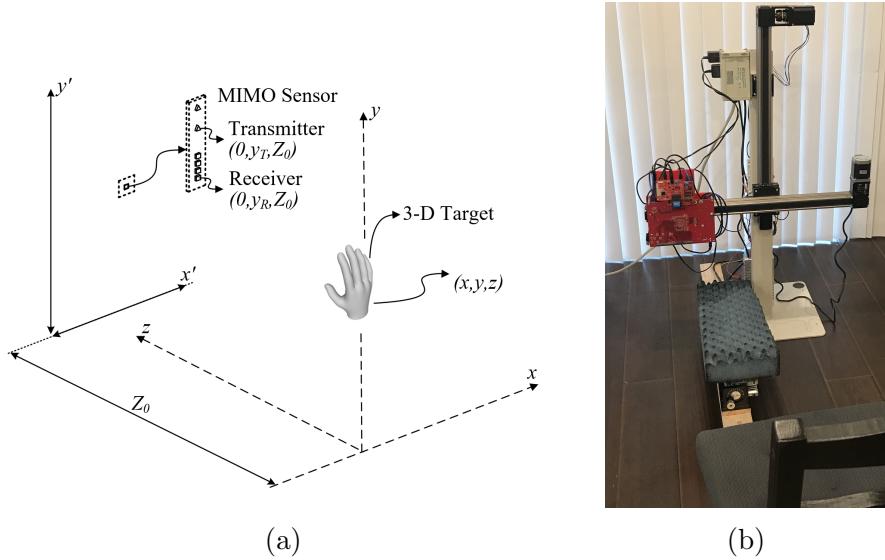
Figure 3.9: (a) A MIMO radar sensor with transmitter and receiver antenna elements located at $(0,y_T,Z_0)$ and $(0,y_R,Z_0)$, respectively captures the return signal from a three-dimensional (3-D) target whose reflectivity function is $p(x, y, z)$. (b) Two-dimensional $x$-$y$ rectangular scanner system with chair for test subject to sit.

cannot be analytically reconstructed as a three-dimensional (3-D) image and can be used to easily classify gestures using traditional optical image approaches. Thus, classifying static hand gestures involves attempting to learn a high-dimensional pattern (hand pose in three dimensions) from low-dimensional radar data.

Another issue inherent to the hand gesture problem is the small radar cross-section (RCS) of the human hand, which results in a low signal-to-noise ratio (SNR). Even with a large amount of data, because the RCS of the hand is low, the features unique to each gesture class are not pronounced. As a result, the CNN has difficulty discerning meaningful features for static gestures.

To overcome these deficiencies, we propose a novel data collection strategy and training technique that employs "sterile" data during network training to improve classification accuracy. First, we employ a 2-D $x$-$y$ SAR scanner, as shown in Fig. 3.9b, to capture data from numerous perspectives, both vertically and horizontally. In this manner, while the user remains stationary, many different views of the hand are captured quickly.

<div align="center">(a)            (b)</div>
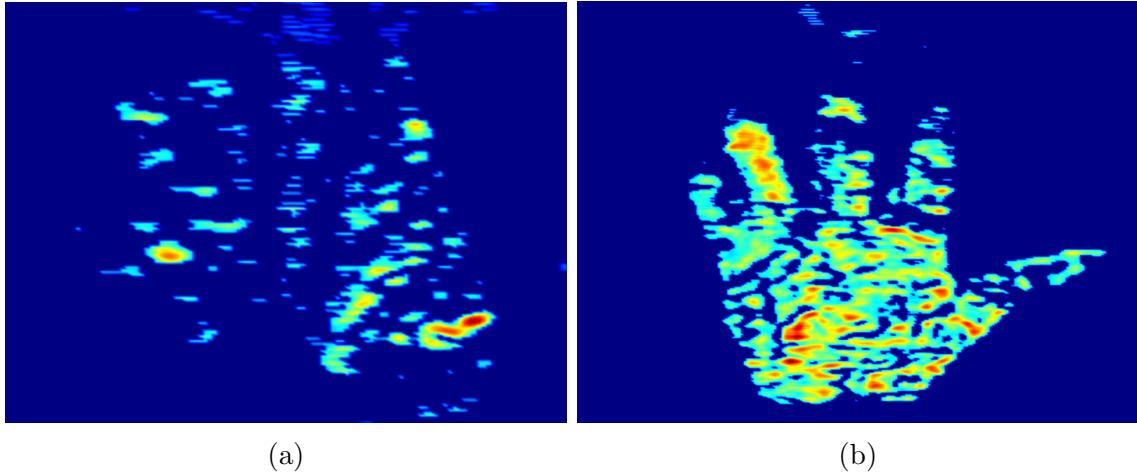
Figure 3.10: Comparison of the reconstructed SAR images from the (a) real human hand and the (b) aluminum cutout of the human hand demonstrating the low RCS of the human hand.

As mentioned previously, the RCS of the human hand is problematically small in comparison to noise and propagation effects. Comparing the range profiles of the different gestures, the differences are mostly indistinguishable to the human eye, as shown in Fig. 3.4. Even though a peak exists in the range FFT at a distance corresponding to the human hand, the features of the gesture reflected back to the radar are not sharply defined and are centered at different places on the human hand.

To demonstrate this phenomenon, a SAR approach is temporarily adopted to reconstruct an image of the human hand using the methods described in [24, 39]. It is important to note that the images shown in Fig. 3.10 are not the data used to train and validate the CNN. These images require all the data (thousands of samples) from the entire horizontal and vertical scan, which takes approximately 5 min to complete.

The reconstructed image of the human hand (Fig. 3.10a) shows a poor image of the hand owing to low RCS and SNR. Comparatively, a SAR image is also reconstructed using an aluminum cutout in the shape of the hand to demonstrate an ideal hand target, as shown in Fig. 3.10b.

<div align="center">30</div>

This empirical analysis reveals the innate difficulty in classifying hand gestures from radar beat signals. Even when employing thousands of radar return signals to construct the SAR image, the hand is barely visible and the gesture is difficult to recognize. From these images, we can infer that the features from a human hand contained in a single beat signal reflected are not pronounced and have a relatively low magnitude compared to the surroundings, noise, etc. In contrast, as shown in Fig. 3.10b, the aluminum cutout demonstrates a high SNR, implying that the features of the gesture are much more prominent and consistent for each static gesture. The novel technique proposed in this section consists of capturing data from many perspectives using a 2-D mechanical scanner from both "real" human hands and "sterile" aluminum cutouts, to improve classification accuracy.

To validate our technique, we collected data from eight participants for three gesture classes: "palm" (Fig. 3.1c), "perm" (Fig. 3.1d), and "thumbs up" (Fig. 3.1e). Similarly, mmWave radar data were collected from the aluminum cutout for each gesture class using a SAR scanner. To compare against a control, we first train two networks using only real human hand data with range and range-angle preprocessing, respectively. For these networks, we use 8000 set aside captures as the validation dataset, making the split between training and validation 80% to 20%. The networks used to classify hand gestures vary based on the preprocessing applied to the dataset. For the range dataset, convolutional layers with kernel sizes of $13 \times 2$, each with 16 filter,s are each followed by a Rectified Linear Unit [19]. These are connected in series, followed by a fully connected layer with three output neurons, softmax layer, and final classification layer using the cross-entropy loss function. The range-angle dataset employs a network with the same architecture, changing only the size of the convolutional layers to $13 \times 4$ to account for larger image sizes. The key to both networks is the complex-valued layering and network architectures. Considering the real and imaginary parts of the radar range data as distinct layers of the image allows the network to identify pixel-to-pixel and layer-to-layer relationships, which correspond to the phase information
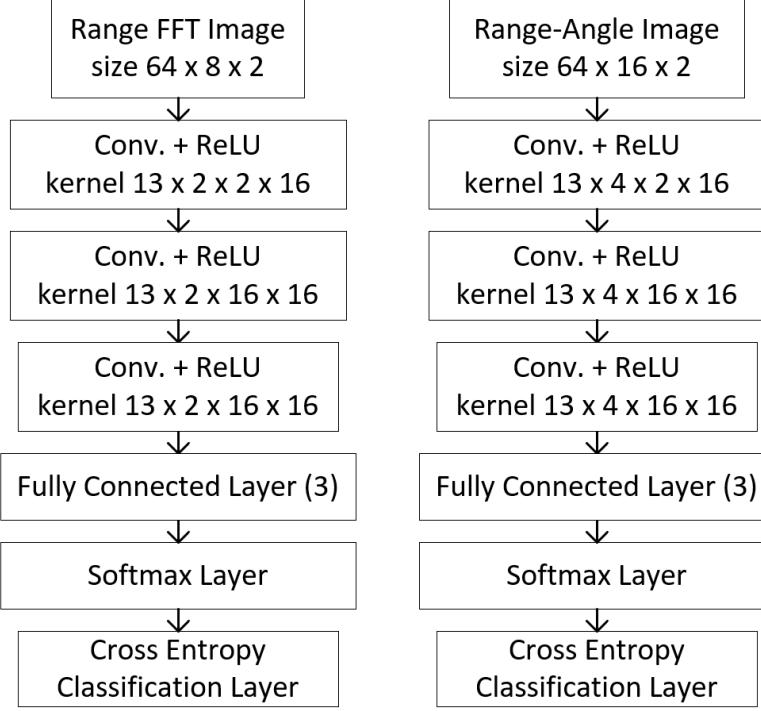
Figure 3.11: The network architecture for the Range FFT DNN and Range-Angle FFT DNN.

of each complex-valued pixel. However, other complex-valued neural network architectures have been explored in the literature [54, 55, 66] and are investigated in Chapter 6. The architectures of both networks are chosen after close inspection of the feature sizes in the observation image domain in both range and channel/angle, in addition to extensive testing to optimize the real-time implementation efficiency and classification rate. Both network architectures are shown in Fig. 3.11.

After training each network with only real human hand data, the range CNN and range-angle CNN yield classification rates of 84.9% and 90.2%, respectively. These networks are named "Human Only" in Table 3.3 since they are trained with only the range and range-angle profiles from human hands. Next, two new networks with identical architectures are trained using the complete datasets, consisting of real human hand data supplemented by "sterile" data from aluminum cutouts. These networks are dubbed "Combined" since they are trained with both real and "sterile" images. It is important to note that the "Combined" networks

are validated with the same validation data as the "Human Only;" the only difference being the training dataset used for each network. These results corroborate our hypotheses on training with "sterile" data, as the classification rates improve to 93.1% and 95.4% for the range and range-angle datasets, respectively.

Table 3.3: Comparison of classification rate between networks trained with only human hand data (Human Only) and networks trained using sterile data to supplement the real human hand data (Combined).

|             | Human Only | Combined |
|-------------|------------|----------|
| Range       | 84.9%      | 93.1%    |
| Range-Angle | 90.2%      | 95.4%    |

Compared to prior work in the literature, our proposed method improves upon gesture recognition by using sterile data while offering a solution to the difficult classification problem of static gestures under three-dimensional spatial translation. Kim *et al.* [6] employ a time-domain gesture recognition approach on an ultra-wideband (UWB) impulse-radio (IR) radar. The approach in [6] considers two scenarios separately: (1) six gestures using human hands 15 cm away from the transceiver and (2) three plaster model gestures rotated at 10° increments. For scenario (1), the hand is kept at a constant position for all captures. Both training and testing are performed using human hand data resulting in a classification rate of 91% using a CNN classifier. In scenario (2), plaster models of each gesture are captured from different perspectives by rotating the plaster model. For this scenario, Kim *et al.* record classification accuracies of more than 90% for three gestures and validated the models using data from the plaster model. Comparatively, our method yields a more robust classifier by including both real human hand reflections and "sterile" reflections in the training processes and validating them with only human hand data. Rather than creating two distinct classifiers for human and sterile data separately, as discussed in [6], the technique proposed in this section unites human and sterile data to construct a robust classifier. Furthermore, our

approach investigates more diverse scenarios by capturing data from multiple test subjects at many locations relative to the hand position.

Extensive work has been carried out towards dynamic gesture recognition using mmWave radar, Doppler radar, and IR-UWB sensors [2, 4, 5, 68, 69, 70, 108]; however, this is an entirely separate problem from the problem addressed in this section as the dynamic gesture case considers only motion. This reduces the dimensionality of the classification to temporal motion features, whereas static gesture recognition on mmWave radar involves classification of a three-dimensional structure using lower-dimensional data, as discussed in Section 3.1.2 previously. Thus, our model is trained for the more difficult problem of static gesture classification under spatial translation and demonstrates superior classification accuracy compared to prior static gesture classification studies [6].

These studies on static and dynamic gesture classification approach hybrid-learning from the perspective of improving deep learning classification techniques by leveraging expertise in signal processing. Similarly, we extend our analysis by examining a similar preprocessing problem in high-resolution imaging.

## 3.3   Efficient 3-D Near-Field MIMO-SAR Imaging for Irregular Scanning Geometries

With the emergence of fifth-generation (5G) and sixth-generation (6G) technologies, UWB mmWave transceivers are enabling unprecedented sensing and communications feats [34, 109, 110]. Small form-factor multiple-input-multiple-output (MIMO) radars are becoming increasingly popular owing to their low cost and power consumption [9, 111]. In addition to emerging 5G communications, mmWave radar has already been realized for high-resolution sensing on the Google Pixel 4 [110]. Of particular interest, recent studies have enabled freehand mmWave imaging by employing positioning sensors commonly employed in smartphones and virtual reality (VR) sensor suites [34, 81, 82, 95, 112]. Sub-wavelength

localization accuracy was previously unachievable by conventional techniques such as 5G mmWave [113] or Bluetooth low-energy (BLE) ranging [114]. Freehand mmWave imaging is a high-resolution imaging technique that relies on conventional synthetic aperture radar (SAR) principles [31, 39, 46, 51, 115] and precise tracking of a handheld radar device as it is moved by a human user throughout space [34, 116, 117, 118]. Whereas traditional mmWave SAR imaging requires precise motion systems to achieve near-ideal synthetic arrays [39], the scanning geometry employed by freehand imaging systems is generally irregular and does not conform to the typical array geometries required for efficient image reconstruction algorithms [102].

While a recent investigation proposes a fast imaging algorithm for irregular SAR geometries using array linearization [119], the proposed technique adopts a simplistic model of the array displacement and does not explore near-field multistatic effects, both of which are addressed in this study. However, efficient algorithms for near-field MIMO-SAR operation under irregular scanning geometries have not been explored in the literature.

Extensive research on freehand mmWave imaging has been conducted by Laviada *et al.* at the University of Oviedo [34, 35, 42, 81, 82, 95, 112]. High-precision localization systems that enable freehand SAR imaging have been investigated using an infrared camera network to accurately track device location over time and recover EM images [81]. Their work was extended to employ an inertial measurement unit (IMU) and depth camera sensors to achieve standalone freehand imaging with promising results [34, 82]. In each of these efforts, the subject attempted to move the hand in a raster pattern to synthesize an approximately rectangular planar aperture using a linear frequency-modulated (LFM) handheld radar [34, 81, 95]. Owing to the subject's inability to move their hand in an ideal planar trajectory and the sensitivity of the mmWave signal to sub-millimeter perturbations, the image was reconstructed using the generalized back-projection algorithm (BPA).

Similar irregular and non-cooperative scanning geometries have been observed in unmanned aerial vehicle (UAV) SAR imaging [35], nonuniform NDT [42], and automotive SAR

imaging [36]. However, for many edge and mobile applications, limitations on power consumption and computational complexity cannot be overcome using existing approaches for irregularly sampled SAR. Although image reconstruction algorithms have been thoroughly investigated in the literature for cooperative synthetic array geometries [9, 23, 31, 39, 46, 51, 102, 115, 120], widely applicable efficient near-field imaging algorithms for applications such as freehand smartphone imaging, UAV imaging, and automotive SAR imaging have not been thoroughly addressed in the existing literature. Furthermore, while MIMO arrays, commonly employed in commercially available radar devices, offer spatially efficient small array sizes, the MIMO-SAR operation introduces a handful of complications in the image reconstruction process and proper handling of the multistatic array is necessary to avoid imaging artifacts [51]. While progress has been made towards projecting MIMO-SAR radar data to virtual single-input-single-output (SISO) monostatic data [51, 102], the analysis is performed on a coplanar assumption that does not generally hold for irregular scanning geometries.

In this section, we propose a novel image reconstruction technique for efficient near-field imaging with irregular scanning geometries, such as those present in freehand imaging, UAV SAR, and automotive scenarios. We examine the system and signal models for UWB MIMO-SAR and develop a multi-planar multistatic approach to mathematically decompose the irregularly sampled synthetic array such that an equivalent virtual planar monostatic array can be constructed. This technique is the first to extend the range migration algorithm (RMA) such that non-cooperative SAR scanning and multistatic effects are simultaneously mitigated. The analysis in subsequent sections provides a novel framework for decomposing irregular SAR scenarios and efficiently projecting irregular MIMO-SAR samples to a virtual planar monostatic equivalent. The proposed algorithm is validated through simulations and empirical experiments, demonstrating robustness to arbitrary scanning patterns and low computational complexity. A thorough study of the relationship between the array irregularity and image resolution of the proposed algorithm is provided. The proposed

technique demonstrates high-fidelity focusing comparable to the traditional planar RMA, even under array perturbation on the order of 10s of wavelengths. Our solution enables the development of emerging technologies that require non-ideal SAR scanning geometries, MIMO multistatic radar, and efficient image reconstruction.
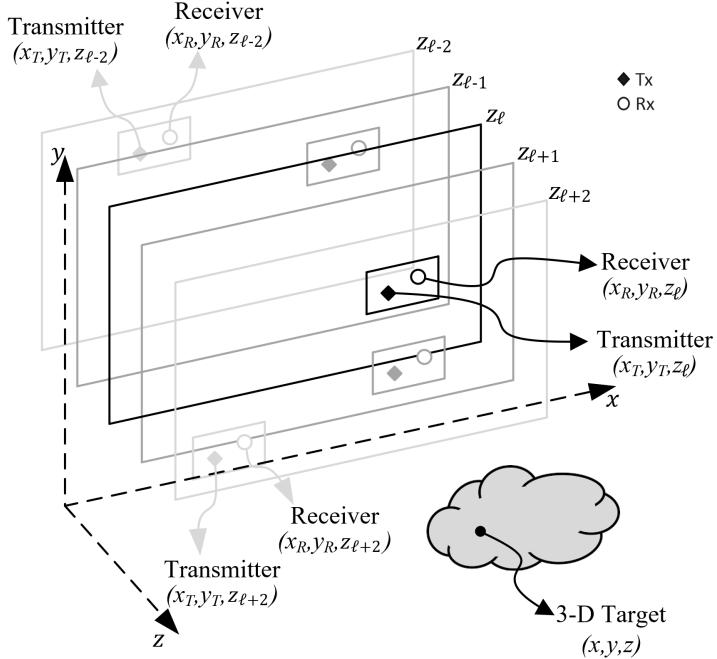


Figure 3.12: Geometry of the multi-planar SAR irregular scanning geometry with a multistatic array.

The remainder of this section is organized as follows. Section 3.3.1 introduces the system model, including the multi-planar multistatic SAR concept, signal model, and a novel compensation technique for planar monostatic SAR. In Section 3.3.3, efficient imaging methods and implementation details are discussed and the Efficient Multi-Planar Multistatic (EMPM) algorithm is proposed. Section 3.3.4 details the hardware and software implementation for collecting multi-planar multistatic SAR data. The results of the simulation and empirical studies are presented and discussed in Section 3.3.5.

### 3.3.1 Near-Field Irregular SAR System Model

In this section, we propose the characterization of irregular or arbitrary three-dimensional (3-D) MIMO-SAR sampling geometry using the multi-planar multistatic scenario shown in Fig. 3.12, where data are collected along different $z$-planes by a MIMO multistatic radar with respect to a stationary 3-D target.

### Multi-Planar MIMO-SAR Configuration

For many emerging SAR applications, as the radar is moved throughout 3-D space, it is generally oriented in the same direction towards some target; however, the samples are taken across several $z$-planes. Because the data are collected during an arbitrary SAR scanning path, the resulting synthetic aperture does not conform to standard scanning regimes, such as rectilinear/planar [8, 39], circular [42, 43], or cylindrical [46, 47, 102]. Hence, the image reconstruction process must consider the irregularity of the spatial sampling, the geometry of which is detailed in Fig. 3.12.

Compared with planar MIMO-SAR, which requires a multistatic MIMO array to be scanned across a planar track [39, 51, 120], multi-planar MIMO-SAR allows the multistatic array to be scanned across a 3-D space. For freehand imaging or automotive SAR, a MIMO array is fixed to a smartphone or vehicle, respectively, and is moved throughout space, generating a multi-planar MIMO-SAR irregular aperture. As shown in Fig. 3.12, because the multistatic array is scanned in an irregular pattern spanning multiple $z$-planes, the locations of the transmit (Tx) and receive (Rx) elements are spatially translated by the movement of the MIMO array. The analyses in the subsequent sections present an efficient solution for irregular MIMO-SAR imaging, such that the position of the radar is known throughout the scan and the planar array assumption does not hold. This scenario is common to many of the aforementioned applications and necessitates both irregular scanning geometries and efficient image recovery.

**The 3-D Multi-Planar Virtual Array Response in Near-Field Imaging**

By the analysis of [51, 102, 121] for the 2-D case, a multistatic MIMO array can be approximated by a monostatic virtual element located at the midpoint of the Tx and Rx elements under the far-field assumption for a small fraction $\epsilon$ as

$$\sqrt{(d_\ell^x)^2 + (d_\ell^y)^2} \leq \sqrt{4\epsilon\lambda R}, \tag{3.4}$$

where $d_\ell^x$, $d_\ell^y$ are the distances between the Tx and Rx elements along the $x$- and $y$-directions, respectively, as shown in Fig. 3.13, $\lambda$ is the wavelength of the carrier frequency, and $R$ is the distance from the midpoint of the antenna elements to a reference point in the scene.

However, under the multi-planar multistatic framework, it is desirable to approximate each Tx/Rx pair using its virtual element located on a $Z_0$ plane in the near-field. Thus, multi-planar data can be projected onto a virtual planar array to ease the subsequent image reconstruction process. As shown in Fig. 3.13, the $\ell$-th Tx/Rx pair located on the $z_\ell$ plane can be approximated by the element located at the midpoint between the Tx and Rx elements migrated to the $Z_0$ plane.

For near-field SAR, the assumption in (3.4) is invalid and the approximation must be handled more delicately. Hence, we derive an efficient compensation algorithm to approximate the multistatic multi-planar array as a monostatic planar array for near-field imaging scenarios.

The transmitter (Tx) and receiver (Rx) of the $\ell$-th multistatic MIMO array are located at $(x_T, y_T, z_\ell)$ and $(x_R, y_R, z_\ell)$, respectively, and the target scene is assumed to be a distributed target whose coordinates are given by $(x, y, z)$. In this study, orthogonality is leveraged across time by operating a MIMO radar using the time-division multiplexing (TDM) MIMO technique such that each Tx/Rx pair is activated sequentially. The round-trip distance
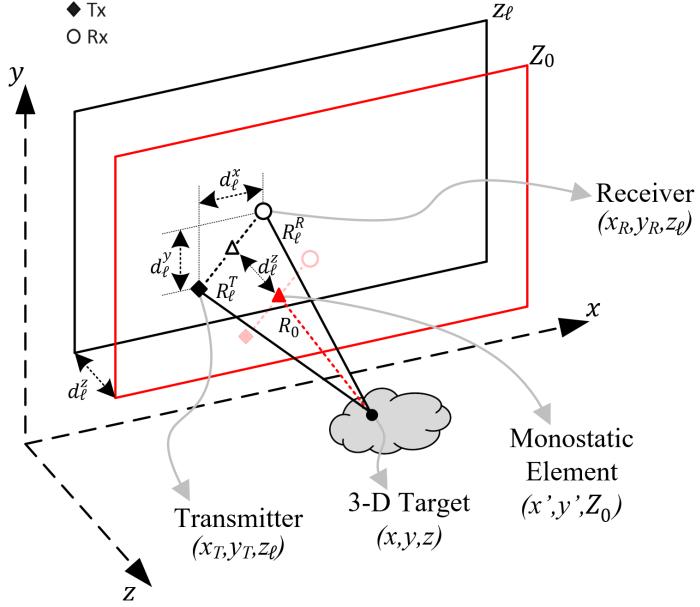
Figure 3.13: Relationship between the multi-planar multistatic elements and virtual planar monostatic elements.

between the $\ell$-th Tx/Rx pair and the point scatter located at $(x, y, z)$ can be written as

$$R_\ell^{RT} = R_\ell^T + R_\ell^R,$$

$$R_\ell^T = \left[(x_T - x)^2 + (y_T - y)^2 + (z_\ell - z)^2\right]^{\frac{1}{2}}, \tag{3.5}$$

$$R_\ell^R = \left[(x_R - x)^2 + (y_R - y)^2 + (z_\ell - z)^2\right]^{\frac{1}{2}}.$$

Denoting the virtual antenna element locations as $(x', y', Z_0)$, the $x$- and $y$-coordinates of the Tx/Rx pair can be expressed as

$$x_T = x' - d_\ell^x/2, \quad y_T = y' - d_\ell^y/2,$$

$$x_R = x' + d_\ell^x/2, \quad y_R = y' + d_\ell^y/2. \tag{3.6}$$

Similarly, denoting $d_\ell^z$ as the distance between the $Z_0$ plane and the $z_\ell$ plane, as shown in Fig. 3.13, the $z$-coordinate of the Tx and Rx elements can be expressed with respect to $Z_0$ as

$$z_\ell = Z_0 + d_\ell^z. \tag{3.7}$$

40

As described in Appendix A, substituting (3.6) and (3.7) into (3.5) and applying the third-order Taylor series expansion of $R_\ell$ for small values of $d_\ell^x$, $d_\ell^y$, and $d_\ell^z$ yields

$$R_\ell^{RT} \approx 2R_0 + \frac{2(Z_0 - z)d_\ell^z}{R_0} + \frac{(d_\ell^x)^2 + (d_\ell^y)^2 + 4(d_\ell^z)^2}{4R_0}$$
$$- \frac{[(x' - x)d_\ell^x + (y' - y)d_\ell^y]^2 + 4(Z_0 - z)^2(d_\ell^z)^2}{4R_0^3}, \tag{3.8}$$

where $R_0$ is the distance between the virtual monostatic element located at $(x', y', Z_0)$ and the point scatterer at $(x, y, z)$, expressed as

$$R_0 = \left[(x' - x)^2 + (y' - y)^2 + (Z_0 - z)^2\right]^{\frac{1}{2}}. \tag{3.9}$$

Centering the target to the origin of the $(x, y, z)$ coordinate system and considering $(x' - x), (y' - y) \ll Z_0$, we can acquire the improved approximation of the round-trip distance between the $\ell$-th Tx/Rx pair and the point scatterer as

$$R_\ell^{RT} = R_\ell^T + R_\ell^R \approx 2R_0 + 2d_\ell^z + \frac{(d_\ell^x)^2 + (d_\ell^y)^2}{4Z_0}. \tag{3.10}$$

### 3.3.2 Multi-Planar Multistatic Signal Model

Consider a multi-planar multistatic array whose Tx and Rx elements are located at $(x_T, y_T, z_\ell)$ and $(x_R, y_R, z_\ell)$, respectively, and a distributed target occupying volume $V$ at locations $(x, y, z)$ in 3-D space with a continuous reflectivity function given by $p(x, y, z)$. The radar beat signal can be written as

$$s(x_T, x_R, y_T, y_R, z_\ell, k) = \iiint_V \frac{p(x, y, z)}{R_\ell^T R_\ell^R} e^{-jk(R_\ell^T + R_\ell^R)} dx dy dz, \tag{3.11}$$

where $k = 2\pi f/c$ denotes the instantaneous wavenumber. Image recovery requires the inversion of (3.11) to produce $p(x, y, z)$. However, given arbitrary sampling locations, the image cannot be computed efficiently using existing techniques [34, 35, 42, 81, 82, 95, 112]. The frequency-domain model of the received signal (3.11) is valid for any UWB radar signaling scheme, including frequency-modulated continuous-wave (FMCW), phase-modulated

continuous wave (PMCW), and orthogonal frequency-division multiplexing (OFDM), which is commonly employed in 5G and IoT applications [122]. Furthermore, prior research on freehand imaging and similar IoT applications has employed a purely stepped-frequency FMCW signal model [81, 82, 95, 112]. Similarly, Google Pixel 4 utilizes a Google Soli 60 GHz mmWave FMCW radar for sensing [110].

However, the derivation of (3.10) enables efficient compensation of multistatic multi-planar data by careful handling of the phase. To achieve the proposed compensation, we express the frequency response of the virtual planar monostatic array, whose elements are located at $(x', y', Z_0)$, as

$$\hat{s}(x', y', k) = \iiint_V \frac{p(x, y, z)}{R_0^2} e^{-j2kR_0} dxdydz, \tag{3.12}$$

where $R_0$ is given by (3.9), $x'$ and $y'$ are the midpoints between each Tx/Rx pair, and $Z_0$ is the plane on which the samples are projected. From the analysis in Section 3.3.1, the relationship between the multi-planar multistatic response and the virtual monostatic array response is given by

$$\hat{s}(x', y', k) \approx s(x_T, x_R, y_T, y_R, z_\ell, k) e^{jk\beta_\ell}, \tag{3.13}$$

where

$$\beta_\ell = 2d_\ell^z + \frac{(d_\ell^x)^2 + (d_\ell^y)^2}{4Z_0}, \tag{3.14}$$

is the near-field residual phase term owing to the arbitrary scanning and MIMO effects in the near-field, as derived in (3.10). Hence, the virtual planar monostatic response can be efficiently acquired from the irregular samples by removing the residual phase to simultaneously account for the multi-planar scanning geometry and near-field multistatic effects. The novel phase compensation technique derived in this section efficiently reduces the dimensionality of the MIMO-SAR imaging problem and projects multi-planar samples onto a single plane to enable computationally tractable algorithms for image reconstruction.

### 3.3.3 Efficient Image Reconstruction Algorithms for Near-Field Planar SAR

In this section, we review traditional planar SAR image reconstruction methods that employ efficient Fourier-based solutions to recover EM images [102] and propose a novel technique for multi-planar multistatic SAR. Existing research on irregularly sampled SAR imaging problems employs the gold-standard back-projection algorithm (BPA) [34, 35, 42, 81, 82, 95, 112]. However, this approach is computationally infeasible for most edge and mobile applications. To overcome this challenge, we employ the approximation in (3.13) to project multi-planar data to a planar-sampled scenario to satisfy the requirements for efficient image reconstruction. The Fourier-based algorithm detailed in the subsequent analysis is known as the range migration algorithm (RMA) or $f$-$k$ algorithm, and has been explored in greater detail elsewhere [23, 52, 102, 115, 120].

The key step to efficiently invert the integral in (3.12) is to represent the spherical wave term as a superposition of plane waves using the method of stationary phase (MSP) [51, 102], such that

$$\frac{e^{-j2kR_0}}{R_0} \approx \iint_A \frac{e^{-j(k'_x(x-x')+k'_y(y-y')+k_z(z-Z_0))}}{k_z} dk'_x dk'_y, \tag{3.15}$$

where

$$k_z^2 = 4k^2 - (k'_x)^2 - (k'_y)^2, \tag{3.16}$$

and $A$ is the region in $k'_x$-$k'_y$ space occupied by the spherical wavefront.

Following the analysis in [39, 102], substituting (3.15) into (3.12) and rearranging the phase terms to leverage the Fourier relationships yields

$$P(k_x, k_y, k_z) = \hat{S}(k'_x, k'_y, k)k_z e^{-jk_z Z_0}, \tag{3.17}$$

where $P(k_x, k_y, k_z)$ and $\hat{S}(k'_x, k'_y, k)$ are the spatial spectral representations of the reflectivity function $p(\cdot)$ and array response $\hat{s}(\cdot)$, respectively. Because the primed and unprimed coordinate systems are coincident, the distinction can be dropped for the remaining analysis.

Hence, the RMA image recovery process can be summarized as

$$p(x, y, z) = \mathrm{IFT}_{\mathrm{3D}}^{(k_x, k_y, k_z)} \left[ \mathcal{S} \left[ \mathrm{FT}_{\mathrm{2D}}^{(x', y')} \left[ \hat{s}(x', y', k) \right] k_z e^{-jk_z Z_0} \right] \right], \tag{3.18}$$

where $\mathrm{FT}[\cdot]$ and $\mathrm{IFT}[\cdot]$ are the forward and inverse Fourier transform operators, respectively, $\mathcal{S}$ is the Stolt interpolation operator required to compensate for the spherical wavefront [102], and $\hat{s}(\cdot)$ is obtained from (3.13). The spatial resolution along each dimension of the recovered image is given by

$$\delta_x = \frac{\lambda_c Z_0}{2 D_x}, \quad \delta_y = \frac{\lambda_c Z_0}{2 D_y}, \quad \delta_z = \frac{c}{2B}, \tag{3.19}$$

where $D_x$ and $D_y$ are the sizes of the aperture along the $x$- and $y$-directions, respectively; $B$ is the system bandwidth; and $\lambda_c$ is the wavelength of the center frequency [51, 52, 102].

Although (3.18) provides an efficient solution for planar array imaging problems, its application to irregular scanning geometries requires a discussion of several key issues. Applying the compensation technique in (3.13) for irregularly sampled data, the multi-planar data can be approximately projected to planar sampling; however, they are likely non-uniform at positions $(x', y', Z_0)$ along the $x$- and $y$-directions. Traditional efficient implementations rely on the common fast Fourier transform (FFT) algorithm; however, recent work on non-uniform planar MIMO-SAR [52, 120] and irregular MIMO real aperture radar (MIMO-RAR) [101] imaging has produced solutions using a non-uniform FFT (NUFFT) approach employing fast Gaussian gridding (FGG), as discussed in [123], for the Fourier transforms and Stolt interpolation step in (3.18). The sampling criteria for the nonuniform planar case are discussed in detail in [52, 101, 120] and apply correspondingly to irregular scanning scenarios after multi-planar compensation. Similarly, the FGG-NUFFT technique is employed in this study to perform the proposed RMA efficiently on irregularly sampled planar data.

For the multi-planar sampling scenario discussed in Section 3.3.1, the RMA cannot be applied directly without multi-planar compensation because the data are sampled on different

$z$-planes, as discussed in Section 3.3.5. If the RMA is applied to the raw multi-planar data, the forward Fourier transform in (3.18) is invalid because the data along the $x'$ and $y'$-directions are not coplanar and the resulting image will suffer from significant distortion, rendering the resulting images unusable in most cases.

```
┌─────────────────────────────────┐
│   Multi-Planar Multistatic Data │
└─────────────────────────────────┘
         ┊╌╌▶ s(x_T, x_R, y_T, y_R, z_ℓ, f)
┌─────────────────────────────────┐
│      Multi-Planar Multistatic   │
│          Compensation           │
│    ⊗◀──── e^{jkβ_ℓ}             │
└─────────────────────────────────┘
┌─────────────────────────────────┐
│   Virtual Planar Monostatic Data│
└─────────────────────────────────┘
         ┊╌╌▶ s(x', y', f)
┌─────────────────────────────────┐
│     Range Migration Algorithm   │
└─────────────────────────────────┘
         ┊╌╌▶ o(x, y, z)
┌─────────────────────────────────┐
│       Reconstructed Image       │
└─────────────────────────────────┘
```
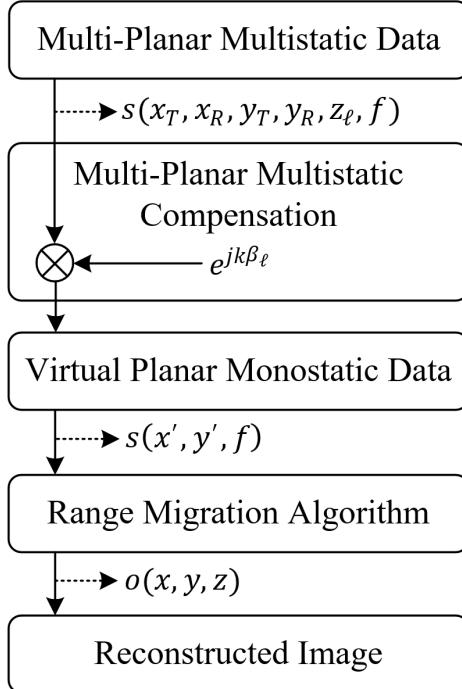
Figure 3.14: The complete EMPM image reconstruction process from irregular sampling compensation to RMA image recovery.

Sampling considerations for image reconstruction remain identical to those in analyses elsewhere [23, 51] after the multi-planar compensation algorithm. Baseband frequency sampling criteria can be determined using the maximum range for a given application. As given in [24], the maximum frequency sampling interval is given by $\Delta_f < c/(2R_{\max})$, where $R_{\max}$ is the maximum target range. Although spatial sampling criteria are not guaranteed for irregular SAR scanning, if the relationship between the capture rate of the radar and the velocity of the radar platform is tuned appropriately during system design, undersampling artifacts are typically minimal [81, 82, 95, 112]. To avoid spatial undersampling, the lower bound of

the pulse repetition frequency (PRF) can be computed using PRF $> 4v_{\max}/\lambda_c$, where $v_{\max}$ is the maximum velocity for a certain application. For example, assuming that the maximum velocity of the human hand for a freehand SAR is 1 m/s and a center frequency of 79 GHz, the lower bound of the PRF is approximately 1.06 kHz. It is important to note that the number of captures increases proportionally with the PRF; hence, at high velocities, a large number of samples are captured. The computational performance of traditional techniques employing the BPA degrades substantially when many samples are captured. On the other hand, the signal-to-noise ratio can be improved by increasing the number of samples, at the cost of increasing the computational burden. Hence, an efficient algorithm for multi-planar MIMO-SAR imaging is required to enable many such technologies.

In terms of computational complexity, the EMPM algorithm offers a significant advantage over existing techniques in the literature [34, 35, 42, 81, 82, 95, 112], which employ the BPA, whose computational complexity is on the order of $O(N^6)$ [39, 120]. The time complexity of the RMA and its FGG-NUFFT variants has been investigated in the literature [101, 120] and the multi-planar compensation step proposed in this section presents negligible computational expense to the RMA, which is on the order of $O(N^3 \log N)$ [23, 31]. Hence, as discussed in Section 3.3.5, the EMPM algorithm offers comparable imaging performance to the BPA with tractable execution time for mobile platforms, similar to the RMA.

The EMPM reconstruction process for efficient near-field SAR imaging with irregular scanning geometries is illustrated in Fig. 3.14. Using the analysis in Section 3.3.1, irregular scanning geometries can be modeled as multi-planar sampling scenarios, as shown in Fig. 3.12, and compensated by removing the residual phase due to the multi-planar multistatic conditions. The key difference between the traditional RMA and the EMPM is the alignment of the multi-planar multistatic (MIMO-SAR) data to virtual planar monostatic data. This crucial step compensates for both the sampling irregularities and multistatic MIMO effects simultaneously, while significantly reducing the dimensionality, from 6-D $(x_T, x_R, y_T, y_R, z_\ell, f)$

to 3-D $(x', y', f)$, and subsequently the computational complexity. Finally, virtual planar monostatic data are used to efficiently recover the image using the RMA. In simulation and empirical studies on irregular SAR scanning geometries, the EMPM algorithm is applied to efficiently produce high-resolution 3-D images previously infeasible due to algorithmic deficiencies.

### 3.3.4   Multi-Planar Multistatic Imaging Hardware Prototype

In this section, we discuss the hardware prototype implementation for empirically validating the proposed imaging algorithm by collecting multi-planar multistatic SAR data. The hardware architecture of the mmWave imaging system is illustrated in Fig. 3.15.
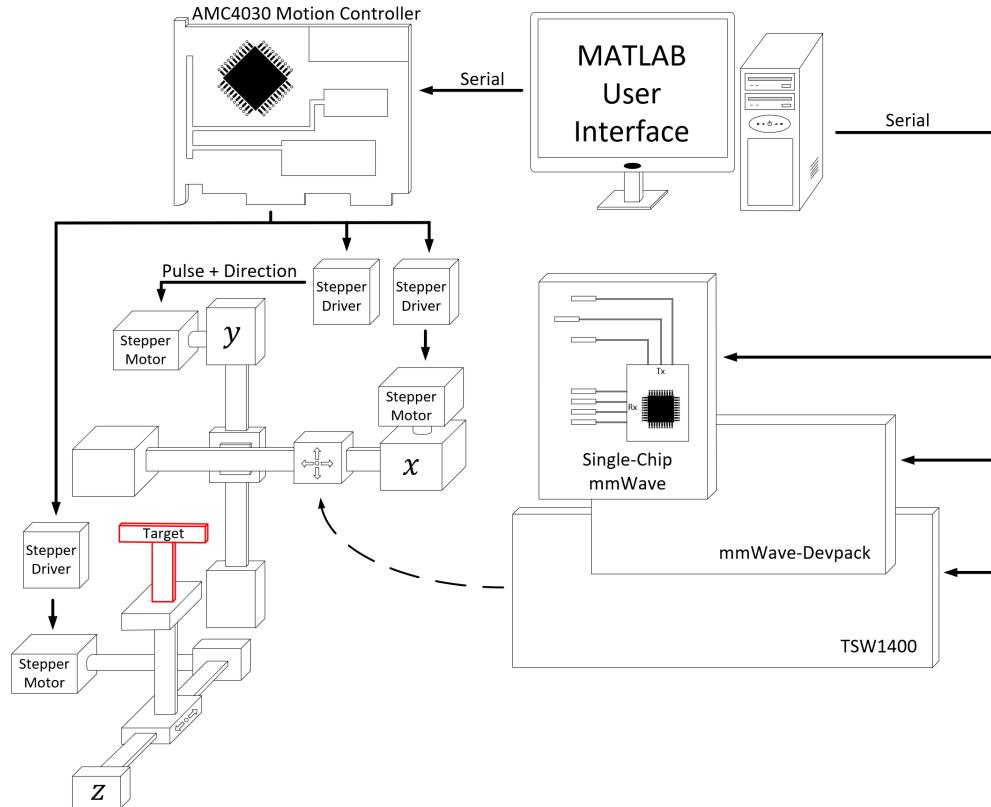


Figure 3.15: System design for 3-D scanner with radar mounted on planar $x$-$y$ rails and target mounted on a linear $z$ rail. The TI radar, data capture card, and mechanical scanner are controlled by MATLAB via USB serial interface.

A Texas Instruments (TI) mmWave MIMO radar is mounted on an $x$-$y$ planar scanner. The TI AWR1443BOOST radar with a bandwidth of 4 GHz from 77 GHz to 81 GHz is mounted on a TI mmWave-Devpack and TSW1400 data capture card to store the data from the SAR scan and transfer it to the PC, where the image recovery algorithm is implemented in MATLAB. The TI AWR1443BOOST is equipped with a MIMO array consisting of two Tx elements spaced by $2\lambda_c$ and four Rx elements spaced by $\lambda_c/2$ [51]. Although 5G and IoT applications commonly employ an OFDM modulation scheme, the TI radar employed for the following experiments utilizes FMCW signaling. However, FMCW radar has been utilized for smartphone applications, notably the Google Pixel 4, which is equipped with a Google Soli FMCW radar [110]. The proposed range migration-based algorithm is applicable to both OFDM and FMCW radars; hence, the results discussed in the following section are relevant for a wide array of 5G, IoT, smartphone, and automotive applications [124].

Additionally, a linear rail is used to move the target along the $z$-direction to collect multiplanar multistatic data under the geometry discussed in Section 3.3.1. All three $x$-$y$-$z$ rails are driven by stepper motors controlled by an AMC4030 motion controller, and the scanning process and radar set up are handled in MATLAB. Additional details on system development and device calibration can be found in [39]. The images are reconstructed using MATLAB implementations running on a desktop PC equipped with a 12-core AMD Ryzen 9 3900X running at 4.6 GHz with 64 GB of memory. Using this hardware prototype, data can be collected for many target scenarios under the multi-planar multistatic scenario by performing multiple planar SAR scans with the target at different $z$-locations. To emulate irregular scanning geometries, data collected throughout the $x$-$y$-$z$ space are subsampled, as discussed in Section 3.3.5. Implementations in the literature employ multi-camera infrared camera systems to track the radar as it is moved through space by the user [81, 95]. Other studies on irregular scanning geometries explored freehand imaging using a stereo camera with an IMU for positioning estimation [34] and UAV near-field imaging with a laser rangefinder and a real-time kinematic (RTK) system for localization [35]. These implementations, among others

[36, 42, 82, 112], demonstrate the viability of high-resolution sensors for precise positioning to enable novel imaging techniques using UWB mmWave radars. Hence, this study focuses on improving the computational efficiency of the imaging technique and assumes that the radar position is known across an irregularly sampled geometry.

### 3.3.5    Measurement Results and Discussion

In this section, we validate the EMPM algorithm derived in Sections 3.3.1 and 3.3.3, as illustrated in Fig. 3.14. Irregular scanning geometries were simulated using the simulation platform developed in [102], and the image reconstruction results are shown by comparing our enhanced method with the gold standard BPA and RMA without multi-planar multistatic compensation. Similarly, using the 3-D mechanical system detailed in Fig. 3.15, irregular scanning geometries were emulated by collecting planar scans with the target at different $z$-locations and subsampling the collected data. Imaging results, comparing the EMPM with the BPA and RMA, demonstrate the computational advantage of our technique while achieving nearly identical spatial resolution. The EMPM algorithm achieves image quality comparable to that of the BPA while offering time and space complexity on par with the RMA.

**Simulated Irregular Geometry SAR Imaging Results**

To validate our proposed algorithm in simulation, we consider three distinct scenarios. First, we investigate the impact of array irregularities on image resolution. We consider the point spread functions (PSFs) of several multi-planar MIMO-SAR scenarios and compare them with an ideal planar scanning scenario to analyze the range and cross-range resolution of the EMPM algorithm. We assume a single ideal point target located at $(0, 0, 0.5 \text{ m})$ in 3-D space for the PSF simulation. For comparison, an ideal linear MIMO-SAR pattern is generated along with several irregular SAR scanning patterns with increasing irregularity.

Each non-cooperative motion track is generated by a semi-smooth, random curve spanning $y' \in [-12.5, 12.5]$ cm with varying $z_\ell$ around $Z_0 = 0$ m with 256 sampling locations, as shown in Fig. 3.17a. To analyze the impact of $d_\ell^z = z_\ell - Z_0$, the distance between the reference plane



(a)
(b)
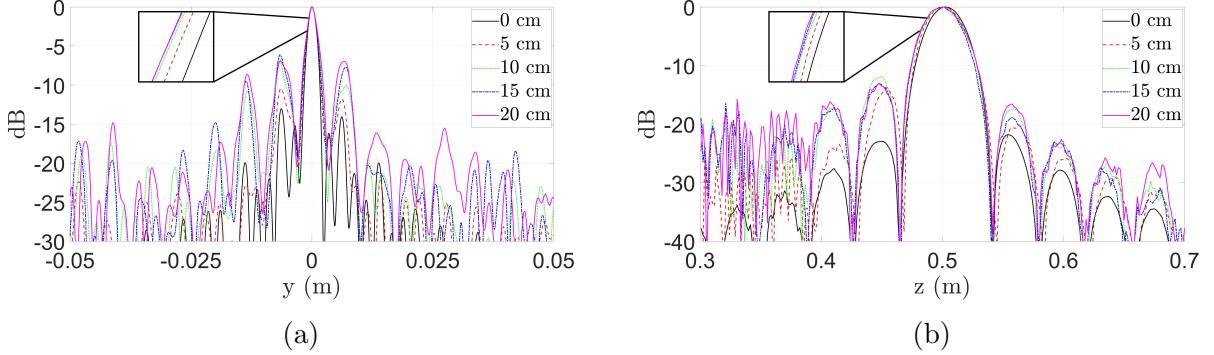
Figure 3.16: Comparison of point spread function (PSF) resolution along the (a) $y$- and (b) $z$-direction with varying maximum distance from the reference plane, $Z_0$, to the samples at $z_\ell$. The distance between the samples and reference plane, $\Delta_z^{\max}$, is varied from 0 cm (linear) to 20 cm with a step size of 5 cm. The linear case (0 cm) is computed with the conventional RMA. Each of the remaining PSFs are computed using the EMPM.

at $Z_0$, and the samples at $z_\ell$, we simulate several multi-planar multistatic with increasing variance of $d_\ell^z$, as shown in Fig. 3.16. The absolute maximum distance, $\Delta_z^{\max} \triangleq \max |d_\ell^z|$, varies from 0 cm, the linear case, to 20 cm with a step size of 5 cm. At a center frequency of 79 GHz, $\Delta_z^{\max} = 20$ cm is more than 50 times the wavelength, $\lambda_c = 3.79$ mm. Prior work on freehand smartphone imaging system design assumes deviations on the order of several centimeters [34]. Along the cross-range dimension, which is symmetric along both the $x$- and $y$-directions, the resolution is minimally affected by the irregular scanning geometry when the algorithm is applied, as shown in Fig. 3.16a. However, a direct relationship between $\Delta_z^{\max}$ and the main beamwidth is observed along with decreased sidelobe suppression compared with the ideal linear case, where the traditional RMA can be employed directly. Along the $z$-direction, the resolution of the EMPM suffers as $\Delta_z^{\max}$ increases but remains quite similar to the resolution of the linear case, as shown in Fig. 3.16b. Hence, the EMPM achieves a focusing performance comparable to that of the ideal linear or planar RMA case while

50

allowing for irregular scanning geometries with large deviations from the reference plane in the $z$-direction.
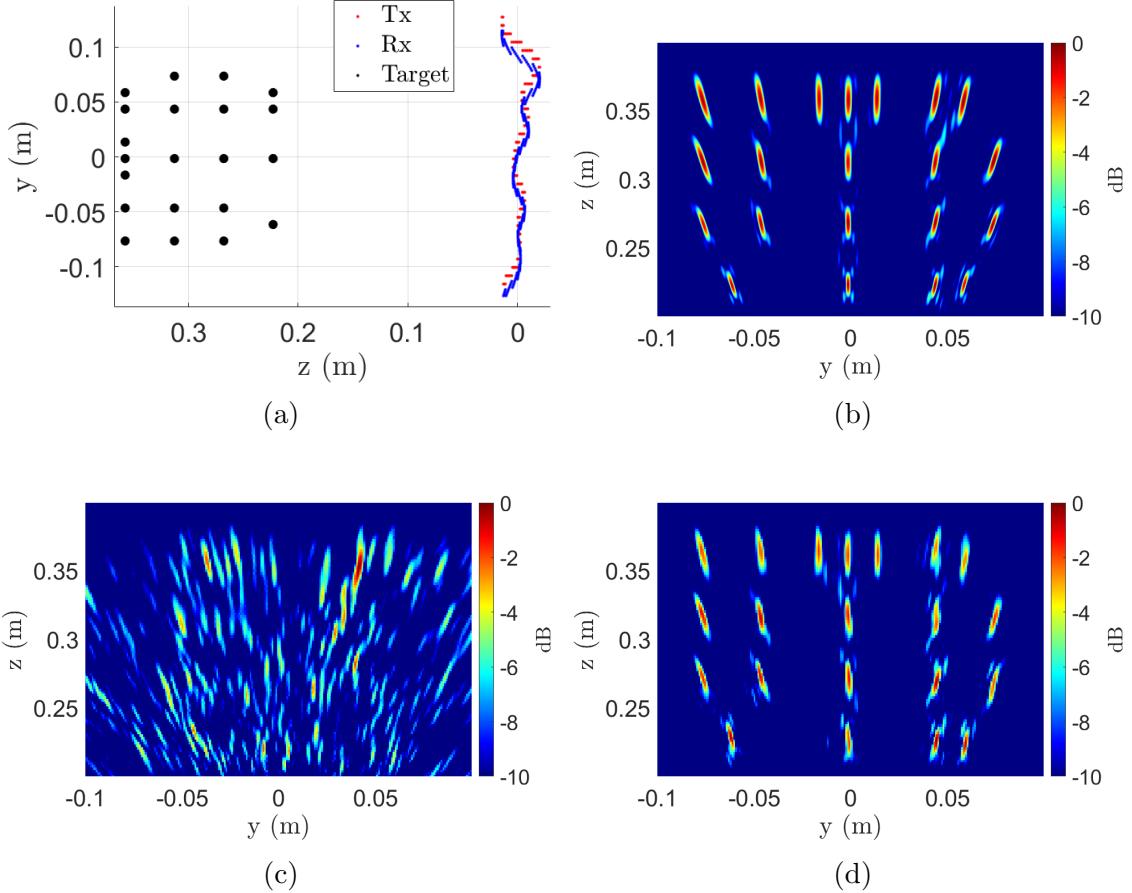


Figure 3.17: (a) Irregular scanning geometry for "UTD" scenario consisting of a multi-linear array in the $y$-direction at $x = 0$ m and corresponding imaging results using the (b) BPA (296.3 s), (c) RMA without multi-planar multistatic compensation (29 ms), and (d) EMPM (30 ms).

To evaluate the performance of the algorithm for more complex targets, a linear array along the $y$-axis is simulated as shown in Fig. 3.17a with 21 point scatterers arranged as the letters "UTD." Again, irregular array locations are generated by a semi-smooth, random curve spanning $y' \in [-12.5, 12.5]$ cm and $z_\ell \in [-2.5, 2.5]$ cm with 256 sampling locations. As shown in Fig. 3.17b, the gold standard BPA recovers each point scatterer without artifacts; however, computing the BPA image requires 296.3 s on our machine. The RMA

without multi-planar multistatic compensation and the EMPM are considerably more efficient, requiring only 30 ms for computation. However, while the RMA image in Fig. 3.17c is significantly distorted and the target shape is lost, the EMPM resolves the point targets comparably to the BPA and requires a fraction of the computation time.

Considering the more broadly applicable 2-D scanning case, a 2-D multi-planar multistatic scenario is simulated with a solid target located at $z = 300$ mm, as shown in Fig. 3.18. The target is a rectangular strip with cutouts of various sizes and the irregular sampling ge-
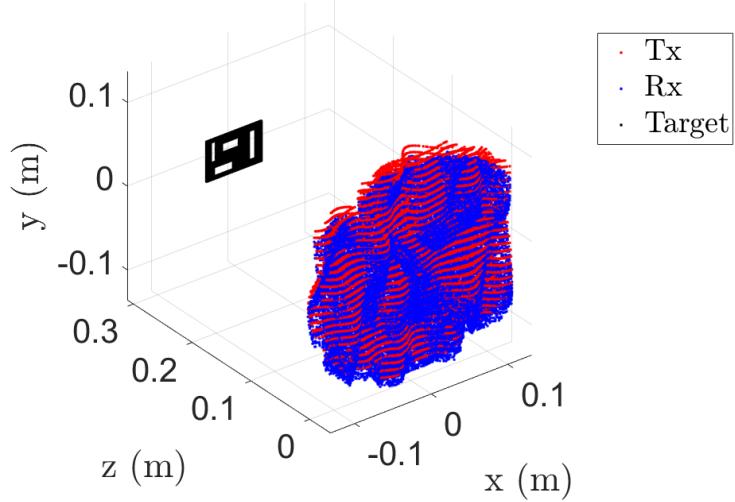


Figure 3.18: Irregular scanning geometry for cutout consisting of a multi-planar array along the $x$- and $y$-directions.

ometry is generated as a 2-D semi-smooth random curve occupying $x' \in [-12.5, 12.5]$ cm, $y' \in [-12.5, 12.5]$ cm, and $z_\ell \in [-2.5, 2.5]$ cm with 102956 sampling locations. The number of sampling locations is selected to approximate a virtual array spanning $[-12.5, 12.5]$ cm along the $x$- and $y$-directions, which is a realistic aperture size for many applications, while satisfying the sampling condition, such that the distance between subsequent sampling points is always less than $\lambda_c/4$. Because the target is located on a single $z$-plane parallel to the planar projection after our compensation technique, a 2-D $x$-$y$ image is recovered at $z = 300$ mm. Again, while the BPA yields a robust reconstruction, the computation time

is excessive for most applications, requiring 1324.8 s on a desktop machine. On the other hand, the EMPM algorithm outperforms the RMA significantly in terms of image quality, nearly matching that of the BPA with only slight artifacting, while demonstrating superior efficiency to the BPA computing a high-resolution 2-D image in only 1.1 s. Similarly, 3-D images can be reconstructed using these methods. The 3-D reconstructed image using the EMPM is shown in Fig. 3.19d, requiring 4.8 s to compute, whereas the RMA and BPA are computed in 4.8 s and 339159.2 s, respectively.
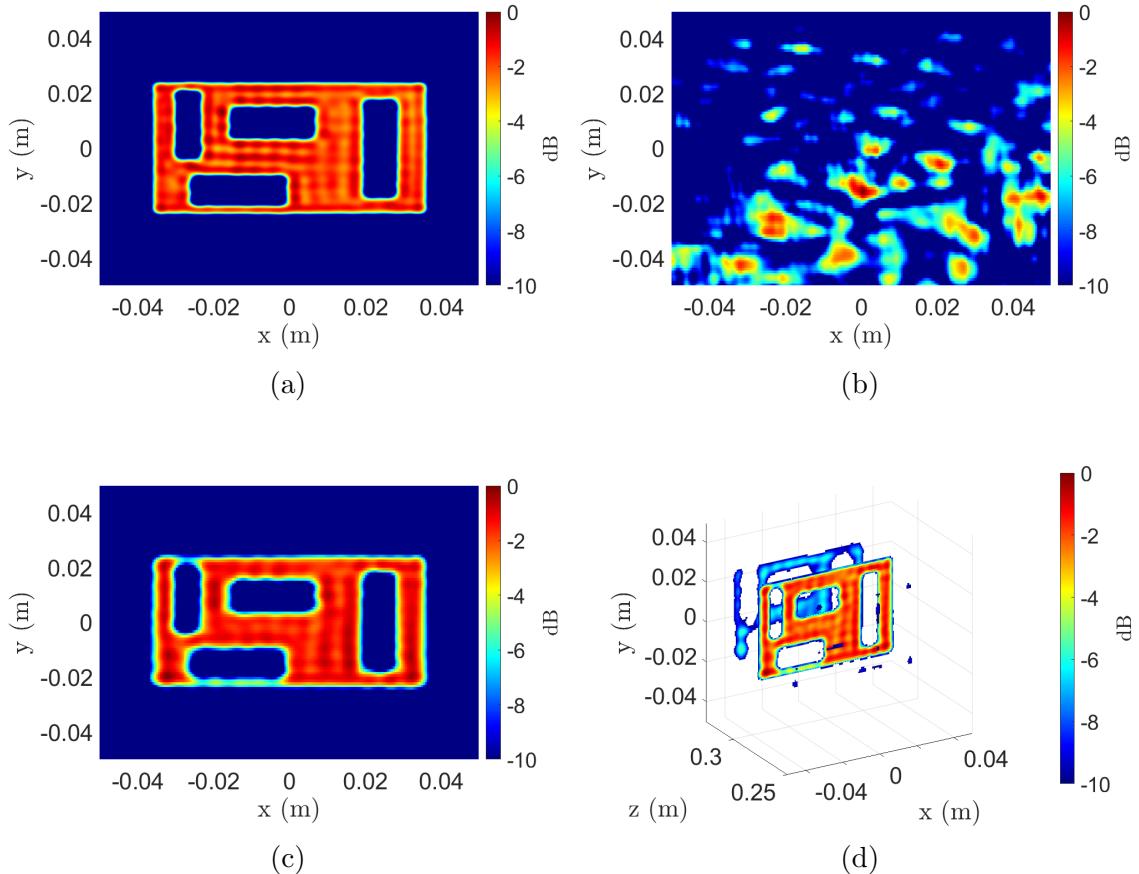


Figure 3.19: Imaging results for the scenario in Fig. 3.18 using the (a) BPA (1324.8 s), (b) RMA without multi-planar multistatic compensation (1.1 s), (c) EMPM at the $z = 300$ mm plane (1.1 s), and (d) the 3-D reconstructed image using the EMPM (4.8 s).

Comparing the results in Figs. 3.17 and 3.19, aberrations appear to be more pronounced along the $z$-dimension or depth. This phenomenon is expected given the analysis in Section

53

3.3.1, where $d_\ell^z$ and the size of the target in the $z$-direction are assumed to be small. Hence, for targets of significant size in the $z$-direction, such as the target in Fig. 3.17c, the proposed compensation suffers from slight artifacting compared with the BPA. However, for many applications, the considerable time savings achieved using our technique is a necessary trade-off compared with the prohibitively slow BPA.

**Empirical Irregular Geometry SAR Imaging Results**

The proposed EMPM imaging technique and system prototype are validated experimentally by capturing SAR data of various target scenes, as shown in Fig. 3.20. The reconstructed images obtained using each method are compared and discussed.



(a)                                                          (b)



(c)
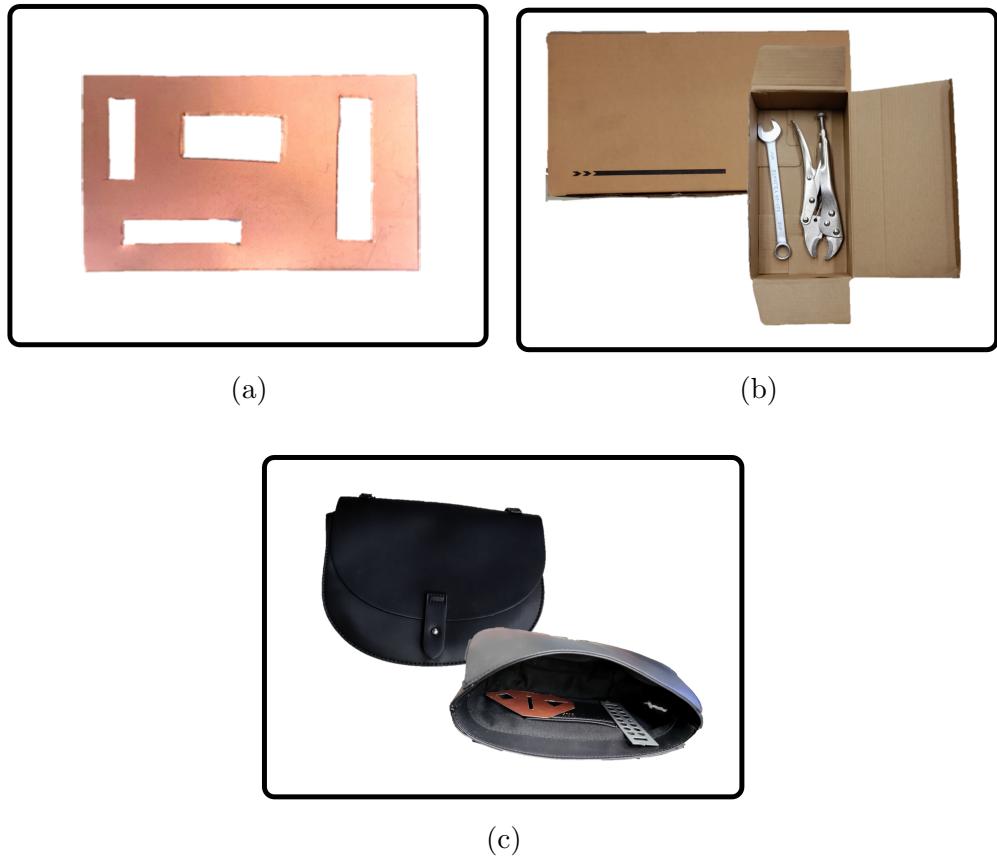
Figure 3.20: Various experimental targets: (a) copper-clad laminate test target, (b) tools hidden inside box, and (c) purse containing metal cutouts.

The test target with several horizontal and vertical rectangular cutouts made from copper-clad laminate (Fig. 3.20a) is illuminated by the $x$-$y$ scanner at the planes $z \in [275, 324]$ mm with a separation of 1 mm. Hence, data are collected throughout the same region discussed previously such that $x' \in [-12.5, 12.5]$ cm, $y' \in [-12.5, 12.5]$ cm, and $z_\ell \in [-2.5, 2.5]$ cm with 102762 sampling locations. After the data were collected, the 50 planar scans were subsampled using a similar random 2-D curve as shown in Fig. 3.18 to emulate the multi-planar irregular sampling scenario. The imaging results and corresponding computation times for each reconstruction algorithm are shown in Fig. 3.21. The EMPM imaging technique demonstrates robustness in projecting the irregular scanning geometry to a planar scenario for more efficient image recovery, as the cutout is recovered cleanly without significant artifacting, as shown in Fig. 3.21c. In contrast, the image recovered using the BPA requires nearly 30 min to compute, and although the RMA is computed efficiently, the RMA without multi-planar compensation cannot resolve the target scene, as shown in Fig. 3.21b. Furthermore, when the target location is unknown in the $z$-direction, 3-D imaging offers an improved solution with a slightly higher computational cost, using the proposed algorithm.

The second target screened by the prototype to demonstrate a hidden target scenario consists of two wrenches (a combination wrench and a vise grip) hidden inside a cardboard box, as shown in Fig. 3.20b. SAR scans of the target are performed with the target at the $z$-planes $z \in [275, 324]$ mm with a separation of 1 mm. To accommodate a larger target size, the aperture is increased to $x' \in [-25, 25]$ cm, $y' \in [-25, 25]$ cm, and $z_\ell \in [-2.5, 2.5]$ cm with 102545 sampling locations. Similarly, the $x$-$y$-$z$ data are sampled to emulate the multi-planar multi-static scenario using a semi-smooth random curve, as shown in Fig. 3.18. The 2-D and 3-D implementations of the BPA and proposed algorithm are applied to the nonuniform data under the irregular scanning geometry, and the recovered images are shown in Figs. 3.22a – 3.22d. Both wrenches are visible in the reconstructed images; however, while the
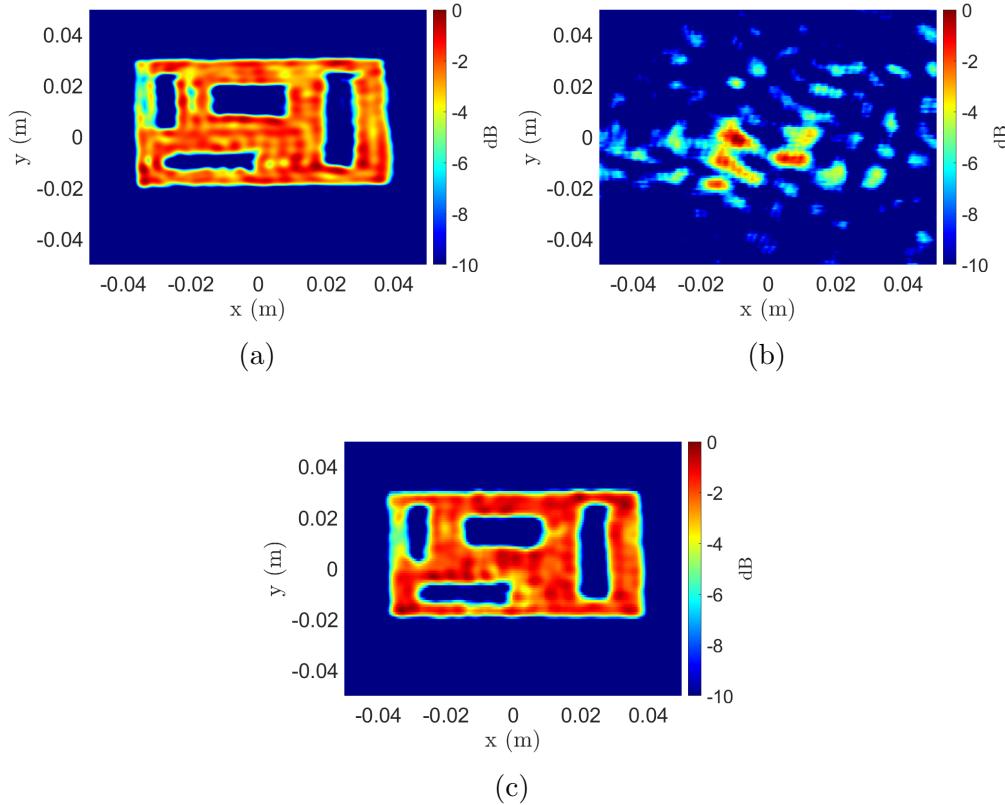
Figure 3.21: Imaging results for the copper test target using the (a) BPA (1324.8 s), (b) RMA without multi-planar multistatic compensation (1.1 s), and (c) EMPM (1.1 s).

2-D image from the BPA and proposed algorithm provide high-fidelity reconstructions of the hidden tools, the 2-D $z$-plane must be carefully selected to obtain such images. The presence and location of targets are generally unknown for concealed item detection problems. Hence, 3-D imaging is preferable for such scenarios and is primarily constrained by computational expense. Our proposed algorithm offers an elegant compromise between the efficiency of the RMA and the image quality of the BPA. In Fig. 3.22d, the 3-D image is computed by the proposed algorithm with an image quality comparable to that of the BPA, with a significantly reduced computational cost.

A third experiment is conducted with several metal cutouts concealed in a purse to emulate a scenario wherein a suspicious personal item is quickly screened with an irregular scanning geometry, such as freehand SAR or drone imaging. Fig. 3.20c shows the purse and two
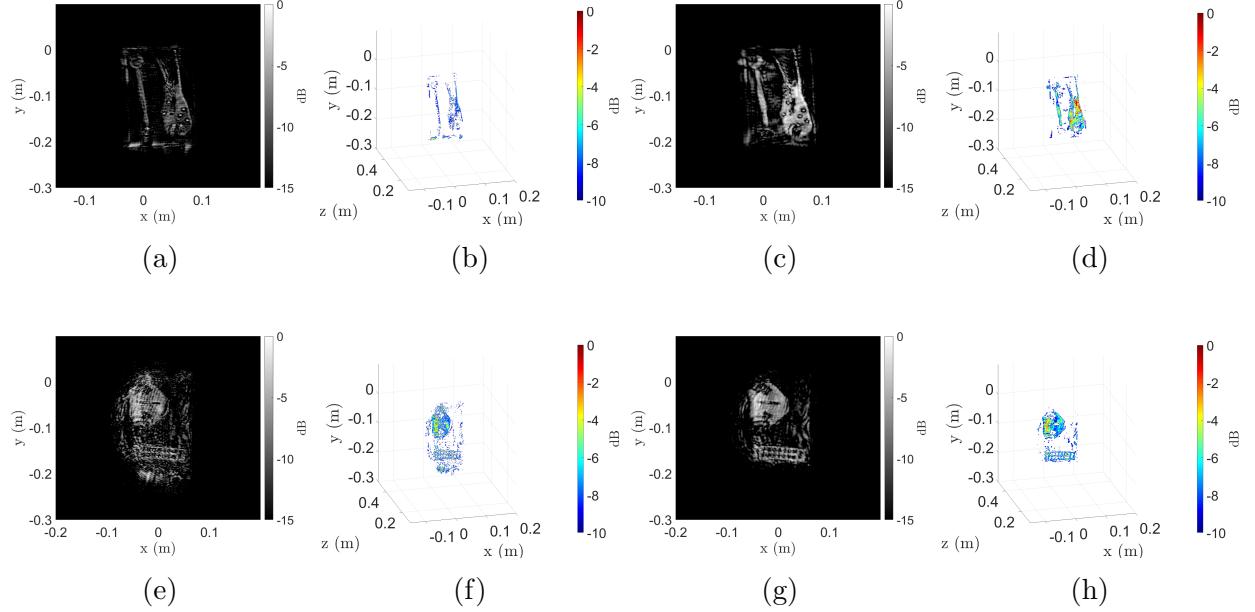
Figure 3.22: Imaging results for the hidden tools target, as shown in Fig. 3.20b, using the (a) 2-D BPA (5299.4 s), (b) 3-D BPA (1356636.9 s), (c) 2-D EMPM (4.3 s), and (d) 3-D EMPM (10.7 s). Imaging results for the concealed items in a purse, as shown in Fig. 3.20c, using the (e) 2-D BPA (5299.4 s), (f) 3-D BPA (1356636.9 s), (g) 2-D EMPM (4.3 s), and (h) 3-D EMPM (10.7 s).

hidden items: a triangularly shaped metal plate with different cutout shapes and a rectangular metal plate with circular holes. The target is scanned by the multi-planar multistatic prototype discussed in Section 3.3.4, and the data are employed to emulate an irregular sampling scenario. Scans are performed with the target at the $z$-planes $z \in [275, 324]$ mm with a separation of 1 mm and an aperture is synthesized within $x' \in [-25, 25]$ cm, $y' \in [-25, 25]$ cm, and $z_\ell \in [-2.5, 2.5]$ cm with 102821 sampling locations. The reconstructed images and corresponding computation times are shown in Figs. 3.22e – 3.22h. Both metal cutouts are resolved using our algorithm, with an image quality comparable to that of the BPA. Again, assuming that the contents of the purse are generally unknown, computing the 3-D image is preferable for concealed item detection. To efficiently recover a 3-D image with irregularly sampled data, existing inversion techniques require excessive computation time and memory

capacity, as shown in Figs. 3.22e and 3.22f. However, the proposed EMPM (Figs. 3.22g and 3.22h) offers an efficient solution that does not compromise image quality.

These experiments demonstrate the advantages of the EMPM and the limitations of the RMA and BPA. A comparison of the computation times required for each algorithm is presented in Table 3.4. Applying the RMA directly to the multi-planar data, as shown in Figs. 3.17c, 3.19b, and 3.21b, yields significant aberrations to the point of failed reconstruction. Therefore, the RMA images are not shown in the other examples. When the target is known to be 2-D and located at a single known $z$-plane, the 2-D BPA implementation can be computed somewhat efficiently in certain instances by employing a graphics processing unit (GPU) and parallelizing the computation [34, 81, 82, 95, 112]. However, particularly for mobile applications, access to high-capacity GPUs is rare or size-prohibitive, and such acceleration is infeasible. Moreover, as the BPA is scaled up to three dimensions, the time and space complexities increase exponentially, requiring excessive computational power and memory. In many emerging applications, efficient 3-D image computation on low-power devices is preferable, if not mandatory, as the precise location of the target is generally unknown. However, efficient algorithms, such as the RMA, require monostatic, planar assumptions that are unachievable by these applications. To enable such technologies, the EMPM imaging algorithm efficiently compensates for the irregular scanning geometry by carefully handling the phase of each sample. This enables image reconstruction under dynamic conditions with computational complexity identical to that of the RMA and image quality comparable to that of the BPA.

In this section, we presented a novel approach for high-resolution, efficient 3-D near-field SAR imaging for irregular scanning geometries. We proposed a multi-planar multistatic framework applicable to a diverse set of applications, including freehand imaging, UAV SAR, and automotive imaging. A novel algorithm, the EMPM, is proposed to efficiently compensate for irregularly sampled multi-planar multistatic data to equivalent planar monostatic

Table 3.4: Computation time, in seconds, required by the various algorithms for each experiment.

| | Metal Cutout | Hidden Tools | Purse |
|---|---|---|---|
| 2-D BPA | 1324.8 | 5299.4 | 5299.4 |
| 3-D BPA | 339159.2 | 1356636.9 | 1356636.9 |
| 2-D RMA | 1.1 | 4.3 | 4.3 |
| 3-D RMA | 4.8 | 10.7 | 10.7 |
| 2-D EMPM | 1.1 | 4.3 | 4.3 |
| 3-D EMPM | 4.8 | 10.7 | 10.7 |

mmWave radar data. Our technique extends the traditional RMA by presenting an algorithm for efficiently aligning multi-planar multistatic data to a virtual planar monostatic scenario. By projecting the data onto a virtual planar monostatic equivalent array, the RMA is extended to account for both irregular scanning and MIMO-SAR effects, resulting in high-fidelity focusing. The EMPM is valid for common radar signaling techniques in 5G, IoT, smartphones, and automotive applications. The simulation results demonstrate the robustness of our approach in the presence of significant spatial deviations among the samples along the $z$-direction. Furthermore, we empirically validated the EMPM by using a custom prototype to capture multi-planar multistatic data for several concealed and obscured scenarios.

In both simulation and experimental studies, our algorithm achieves efficient image reconstruction matching the focusing quality of existing techniques while reducing computational complexity by a considerable margin. The EMPM is valuable for enabling efficient medium-fidelity image recovery; however, additional image enhancement can be pursued via data-driven approaches, such as those discussed in Chapter 4. Without the EMPM imaging algorithm, deep learning-based SAR super-resolution on irregularly sampled SAR images remains infeasible. Herein, we observe the interaction between front-end signal processing techniques and deep learning approaches.

For many problems in perception and imaging, a parallel design of signal processing and data-driven techniques yields impressive results. By examining static and dynamic gestures, we studied the effects of various preprocessing techniques on classification accuracy and proposed a novel "sterile" training technique to improve CNN robustness [7]. We extended our analysis of preprocessing algorithms to the near-field SAR imaging domain for irregular scanning geometries. Many emerging applications require efficient SAR image recovery with non-cooperative arrays in the near-field. The proposed EMPM algorithm achieves image quality comparable to that of the computationally prohibitive BPA with the efficiency of the planar array-constrained RMA [94]. The EMPM enables rapid medium-fidelity image reconstruction for irregularly sampled near-field SAR scenarios. Further investigation into data-driven methods for enhanced imaging is discussed in subsequent chapters.

# CHAPTER 4

# ENHANCED HIGH-RESOLUTION IMAGING ALGORITHMS USING DATA-DRIVEN METHODS

In the previous chapter, we employed signal processing expertise to improve deep learning sensing and perception capabilities. In this study, we employ data-driven techniques to enhance high-resolution near-field SAR image reconstruction algorithms. Part of the following work was previously published in [125][1] and [126][2] and will be presented in [102].

Near-field SAR and ISAR imaging systems are becoming increasingly popular for numerous applications, from security sensing to industrial packaging. Such systems suffer from distortions owing to assumptions and approximations in the image reconstruction process [51] in addition to system limitations [111] and non-idealities [39]. To illustrate these phenomena, we compare the reconstructed SAR image from the points in the shape of the letters "UTD" with its ideal counterpart, as shown in Fig. 4.1. Even in the noiseless case, distortion around the point targets in Fig. 4.1a is present owing to physical limitations and assumptions of the imaging algorithm. Because system cross-range and range resolution are limited by the effective length of the aperture and bandwidth, respectively, any resolution improvement through data processing can enable technologies that are otherwise infeasible due to cost or size constraints. Although some attempts have been made towards SAR image super-resolution using sparse coding and compressed sensing (CS) techniques [127, 128], near-field SAR has only recently received attention for image super-resolution [55, 64, 65, 86, 87]. Additionally, deep learning-based techniques have been shown to outperform CS algorithms for

---

[1]©2022 IEEE. Reprinted, with permission, from 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 *Proc. IEEE Tex. Symp. Wirel. Microw. Circuits Syst. (WMCS)*, Waco, TX, USA, Apr. 2022, pp. 1–6.

[2]©2022 IEEE. Reprinted, with permission, from C. Vasileiou, J. W. Smith, S. Thiagarajan, M. Nigh, Y. Makris, and M. Torlak, "Efficient CNN-based super resolution algorithms for mmWave mobile radar imaging," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Bourdeaux, France, Oct. 2022, pp. 3803–3807.
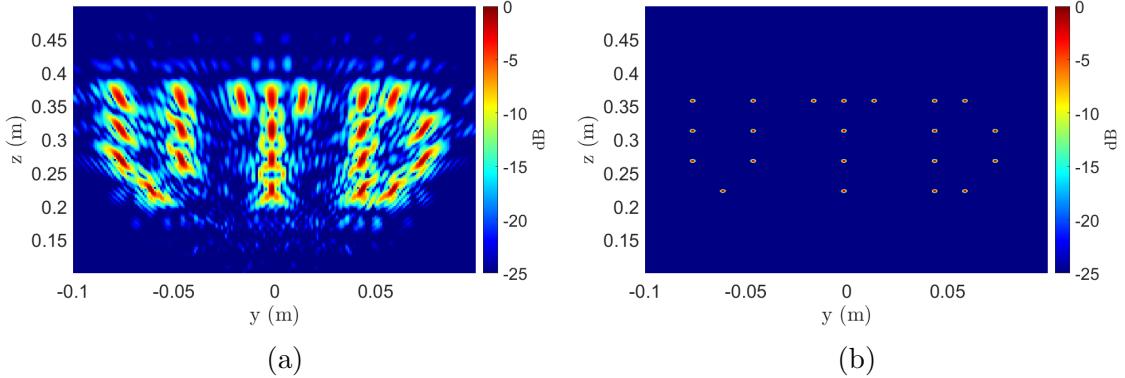
Figure 4.1: 21 ideal point scatters arranged in the letters "UTD." (a) SAR image. (b) Ideal image.

far-field SAR image super-resolution [54]. Prior deep learning-based SAR super-resolution methods are trained exclusively on simulated SAR data consisting of randomly placed points and then tested on real SAR data. Thus, there is an increasing need for meaningful near-field SAR data containing sophisticated targets beyond the simple random points employed in previous studies [66].

## 4.1 High-Fidelity Near-Field SAR Simulation Toolbox

Given that a major obstacle to deep learning on mmWave imaging data is the availability of meaningful data for model training and prototyping, we develop an open-source MATLAB toolbox for system prototyping, imaging algorithm development, and high-fidelity dataset generation.

Many array topologies (real or synthetic) have been investigated for near-field array imaging. The Pacific Northwest National Laboratory (PNNL) is a front-runner in the field of array imaging algorithm development. Sheen *et al.* investigated rectilinear (planar) array imaging [8, 23] and cylindrical SAR (ECSAR) [25] using a switched antenna array time-division multiplexing (TDM) approach by sequentially operating transmitters and receivers and approximating the switched array as a monostatic array. Yanik *et al.* have

driven work towards efficient algorithms for near-field and sparse array image reconstruction [9, 51] as well as system-level design [39, 46, 115]. Similarly, the National University of Defense Technology (NUDT), China, has presented extensive work on MIMO and MIMO-SAR imaging algorithms. Their work included efforts on 3-D imaging using 1-D scanning of a 1-D MIMO array [52, 120], 2-D imaging using circular SAR in the near-field [43], and cylindrical MIMO-SAR [53]. In addition, extensive progress towards image reconstruction has been made using the range migration algorithm (RMA) for rectilinear patterns [31, 37, 38, 39, 115, 127, 129, 130, 131, 132, 133, 134], polar formatting algorithm (PFA) for cylindrical scanning geometries [47, 48, 49, 135], and back-projection algorithm (BPA) for any array geometry [17, 29, 40, 41, 136, 137]. The toolbox presented in this section offers tools for simulating and reconstructing images from 1-D linear, 2-D rectilinear, 1-D circular, and 2-D cylindrical array geometries using the provided BPA, RMA, and PFA image reconstruction algorithms.

Prior work on array imaging simulation tools has largely been limited to two domains: far-field simplified imaging simulation [30, 88, 89, 90, 91, 138] and near-field antenna simulation software [25]. Existing simple far-field packages ignore spherical wave phenomena, thus drastically reducing the simulation and image reconstruction complexity, whereas near-field antenna simulators employ application-specific implementations using computational electromagnetic (CEM) solvers such as HFSS, Xpatch, or FEKO. Efforts falling into the former camp are applicable to many remote sensing applications typical of classical strip-map and spotlight SAR imaging modes but fall short for near-field imaging scenarios where the wavefront curvature must be considered [30, 88, 89, 138]. On the other hand, despite being quite robust, computational EM solvers require a tedious process for setting up each SAR scenario and can be prohibitively slow, requiring a computation time on the order of 10s of hours [25]. The proposed toolbox leverages the advantages of both software types, enabling custom waveform design, complex antenna gain patterns, scanning scenario design, and image reconstruction. This drastically increases the accessibility of imaging simulation and

reconstruction to researchers across numerous fields. In particular, this contribution is key to enabling rapid high-fidelity dataset generation. Although the proposed software platform is designed for high-fidelity THz SAR imaging, it can easily be operated at lower frequencies in the mmWave region. The toolbox user interface and API provide a complete solution for numerous SAR imaging tasks and promote data and benchmark standardization, which is a vital step in the evolution of data-driven signal processing algorithms.

The toolbox comprises five primary modules: (1) setting up the desired waveform parameters, (2) creating the MIMO antenna array, (3) configuring the scanning pattern, (4) constructing the target scene and simulating the beat signal, and (5) reconstructing and displaying the image. The functionality required for each step is encapsulated in five MATLAB classes for the toolbox and displayed within the three tabs of the interactive GUI, as shown in Fig. 4.2. Each step is discussed in detail in the following subsections. Further details on each step of the simulation process and the underlying methodology of the various classes can be found in the documentation of the toolbox, as discussed in Appendix B.

### 4.1.1 Setting the Waveform Parameters

To set up the waveform parameters in the interactive GUI, navigate to the "Device Setup" tab (see (1) in Fig. 4.2) and enter the desired parameters. Although the FMCW modulation scheme is the default waveform, the user can easily employ alternatives such as OFDM or PMCW [122, 124, 139, 140, 141]. As the values are changed by the user in the GUI, the dependent parameters (wavelength, maximum unambiguous range, range resolution, etc.) are automatically updated based on the user-defined fields. The descriptions each parameter are detailed below and in the documentation.

For use with the toolbox API without the GUI, an object of class `THzWaveformParameters` is first initialized. Once the object has been constructed, the user can set the desired chirp parameters by changing the properties of the object, whose names are given below. It is
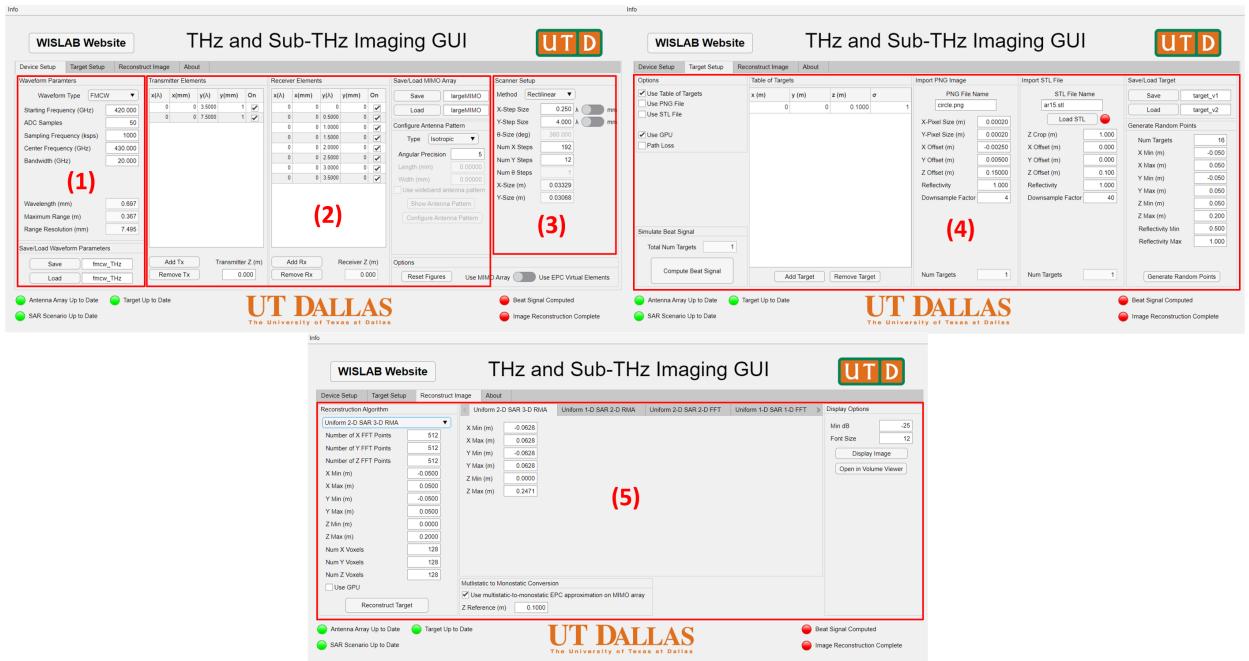
Figure 4.2: Five step process in the interactive GUI.

important to note that the values given to the parameters of the `THzWaveformParameters` object must be in units of s, Hz, and m and must be scaled appropriately rather than the inputs to the interactive GUI, which are given in different units.

- Starting Frequency (GHz): $f_0$/`f0` - Starting frequency of the chirp signal in GHz.

- ADC Samples: $N_k$/`Nk` - Number of samples by the ADC, corresponding to the number of frequency steps in the FMCW chirp.

- Sampling Frequency (ksps): $f_S$/`fS` - Sampling frequency of the beat signal in ksps.

- Center Frequency (GHz): $f_C$/`fC` - Center frequency of the device, used to calculate the wavelength $\lambda = c/f_C$, in GHz.

- Bandwidth (GHz): $B$/`B` - Device bandwidth in GHz.

Hence, the user has complete control over the waveform, including important parameters such as the number of samples and bandwidth, with immediate feedback to relevant

parameters such as the range resolution and maximum resolvable range. In addition, the underlying software is thoroughly documented and straightforward to allow the user to implement custom waveforms of many varieties for simulation and image reconstruction. For system design, having effortless access to such a tool allows for the virtual prototyping of many systems without the need for full hardware prototype construction. Alternatively, the interactive GUI and toolbox both provide an option to save the current waveform parameters for later reuse or reference.

### 4.1.2   Setting Up the Antenna Elements and Array

The second step is to configure an antenna array with transmitters and receivers at the desired locations and optionally use a nonideal antenna pattern. The interactive GUI allows the user to easily set the desired locations of the Tx and Rx elements of the array (see (2) in Fig. 4.2). The MIMO array elements are assumed to be coplanar in an $x$-$y$ plane, whose $z$-coordinate can be specified in the GUI. The $x$ and $y$ positions of each element are set as follows: For example, in the $x$ direction, the value in the column labeled $x(\lambda)$ is scaled by the wavelength, as calculated above and added to the value in the column labeled $x(\text{mm})$ to compute the $x$ position of each antenna element. This functionality is due to the common practice of spacing the antennas by a factor of the wavelength as a result of the spatial sampling criteria [51]. Additionally, the user has complete control over the topology of the radar device and can easily prototype many different array types with various targets and image reconstruction algorithms, as discussed later.

As discussed in Chapter 3, under certain conditions, it is desirable to treat a MIMO array as a virtual monostatic array whose elements are located at the midpoints of the MIMO transceiver pairs. The GUI allows the user to specify whether to use the multistatic MIMO array elements, which can later be approximated as a virtual monostatic array, or to continue the simulation using a monostatic simulation algorithm with ideal SISO elements at the equivalent phase center (EPC) locations of the MIMO array.

66

As the user changes the location or status (On/Off) of the antenna elements, the interactive GUI again provides immediate feedback on the "Graphical Window" showing the array with either the Tx and Rx element or virtual monostatic transceiver elements (Vx), as shown in Fig. 4.3.

Optionally, the user can employ a nonideal antenna pattern. By default, an ideal isotropic antenna pattern of uniform gain is employed to compute the beat signal. However, using the interactive GUI, the user can select several built-in antenna patterns or import a custom antenna pattern from HFSS. The two built-in options, Patch and Dipole antennas employ the MATLAB Antenna Toolbox and are limited to frequencies below 200 GHz. Hence, the option is provided to import a custom antenna pattern simulated in HFSS for use in beat signal computation at THz frequencies.
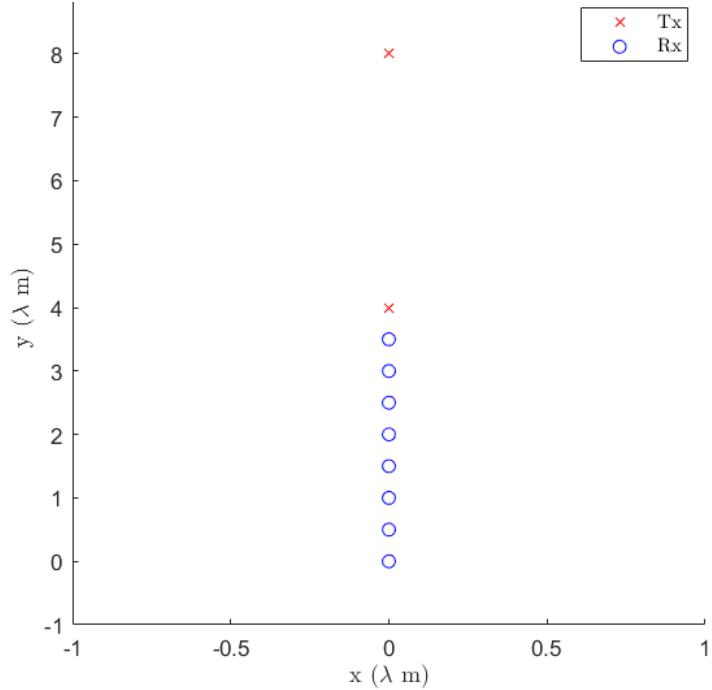


Figure 4.3: Antenna array configuration of a 2 Tx, 8 Rx colinear MIMO array.

The toolbox API employs a custom class called `THzAntennaArray` to compute the relevant antenna parameters from the user inputs. Once the `THzAntennaArray` object is instantiated,

the user can set various parameters, discussed thoroughly in the documentation, determining whether the EPC Vx elements should be used or MIMO Tx/Rx elements (`isEPC`, a Boolean), the $z$ coordinate of the array (`z0_m`), the antenna locations as an array in the same format as the GUI (`tableTx` and `tableRx`), and the antenna pattern properties. Documentation provides thorough instructions and illustrative examples of class usage, methods, and properties. Again, both the interactive GUI and the toolbox allow the user to save the antenna array for later reuse or reference.
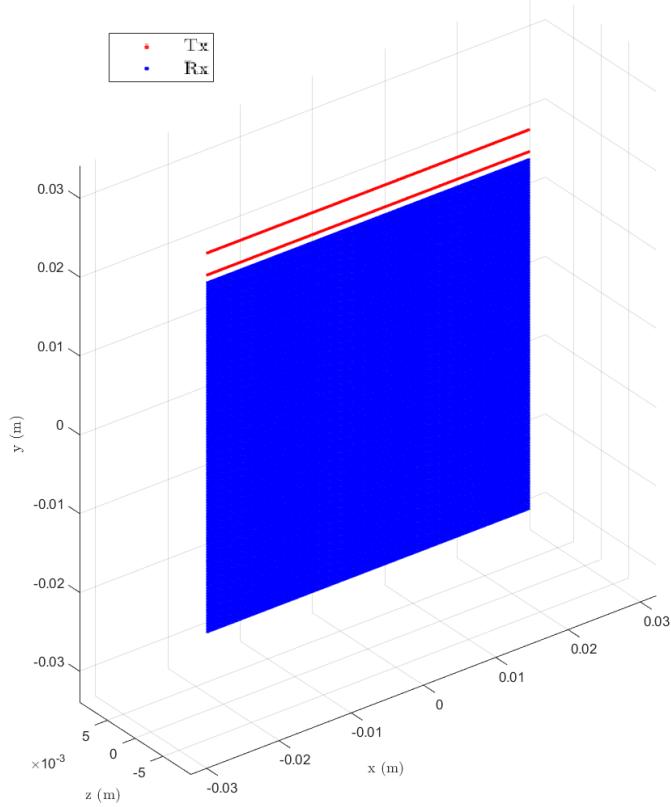


Figure 4.4: Rectilinear scanning geometry with $\Delta_x = \lambda/4$, $\Delta_y = 4\lambda$, $N_x = 256$, $N_y = 16$ using the colinear MIMO array shown in Fig. 4.3.

### 4.1.3   Configuring the SAR Scanning Scenario

After the antenna array is configured, the scanning pattern can be configured using four different scanning modes: linear, rectilinear, circular, and cylindrical. The user can select

the desired scan type under the "Method" drop-down menub and the corresponding fields will be enabled or disabled accordingly depending on the scanning mode. The user has full control over the step size and the number of steps in the $x$ and $y$ directions, denoted previously as $\Delta_x$, $\Delta_y$, $N_x$, and $N_y$, respectively, in addition to the size of the angular scan $\theta_{max}$ and number of angular steps $N_\theta$. As the user changes the parameters in the GUI, the "Graphical Window" updates in real-time showing the current SAR scanning scenario, as shown in Fig. 4.4.

Similarly, using the toolbox API requires initiating an object of the class `THzScanner`. Table 4.1 shows the properties of the `THzScanner` class corresponding to the scanning parameters discussed previously. For more information, please refer to the documentation. Once the user sets the desired SAR scanning parameters, the target scene can be constructed as follows.

Table 4.1: SAR scanning parameters in the interactive GUI and their corresponding properties in the `THzScanner` class.

| Parameter in GUI | Property in Class |
|---|---|
| Scanning Method | `method` |
| X-Step Size - $\Delta_x$ | `xStep_m` |
| Y-Step Size - $\Delta_y$ | `yStep_m` |
| $\theta$-Size (deg) - $\theta_{max}$ | `thetaMax_deg` |
| Num X Steps - $N_x$ | `numX` |
| Num Y Steps - $N_y$ | `numY` |
| Num $\theta$ Steps - $N_\theta$ | `numTheta` |

### 4.1.4 Creating a Target Scene and Simulating the Beat Signal

To create the target scene, the user must switch to the "Target Setup" tab (see (4) in Fig. 4.2). As the user switches tabs, the interactive GUI automatically shows the corresponding figure in the "Graphical Window." On the "Target Setup" tab, the user has three options for creating the target scenario: Table of Targets, Import PNG Image, Import STL File.

The toolbox and GUI can create target scenes using any combination of these methods using the options on the left-hand side of this tab. Additionally, the user can choose to use a graphics processing unit (GPU) to compute the beat signal, which drastically decreases the computation time for large targets, and choose to include two-way path loss in the computation. As the user changes the parameters for any of the target methods, the target scene is updated and shown in real-time in the "Graphical Window."

Using the Table of Targets method, the user can enter the $x$-$y$-$z$ location and reflectivity of any number of desired points, which are treated as ideal point reflectors. Additionally, the user can generate randomly placed point targets that will be filled into the Table of Targets using the panel on the right; more information can be found in the documentation. Figs. 4.5a-b show two scenarios displayed by the interactive GUI using the Table of Targets method, the first with a user-entered point location and reflectivity, and the second with a randomly generated set of points given the user-specified parameters.

The user can also choose to load a PNG image, treated as a 2-D target, in addition to or instead of the point targets entered into the Table of Targets. The user can import various images included in the toolbox or their own custom PNG files if desired. The PNG file is treated as a 2-D $x$-$y$ image consisting of a set of point targets corresponding to the pixels of the PNG file, and the user has full control over the position and scaling of the PNG target. Additionally, the user can change the reflectivity of the PNG target and downsample the target to reduce the computation time of beat signal. For example, the included circle PNG file has been loaded with a downsampling factor of 1, meaning no downsampling, as shown in Fig. 4.5c.

The third method for creating a target scene is to import a Standard Triangle Language (STL) file. An STL file is a standard file type for 3-D computer-aided design (CAD) modeling. With the functionality to import 3-D targets from STL files, our toolbox is capable of simulating highly customizable and complex target scenarios. The STL file is imported as a
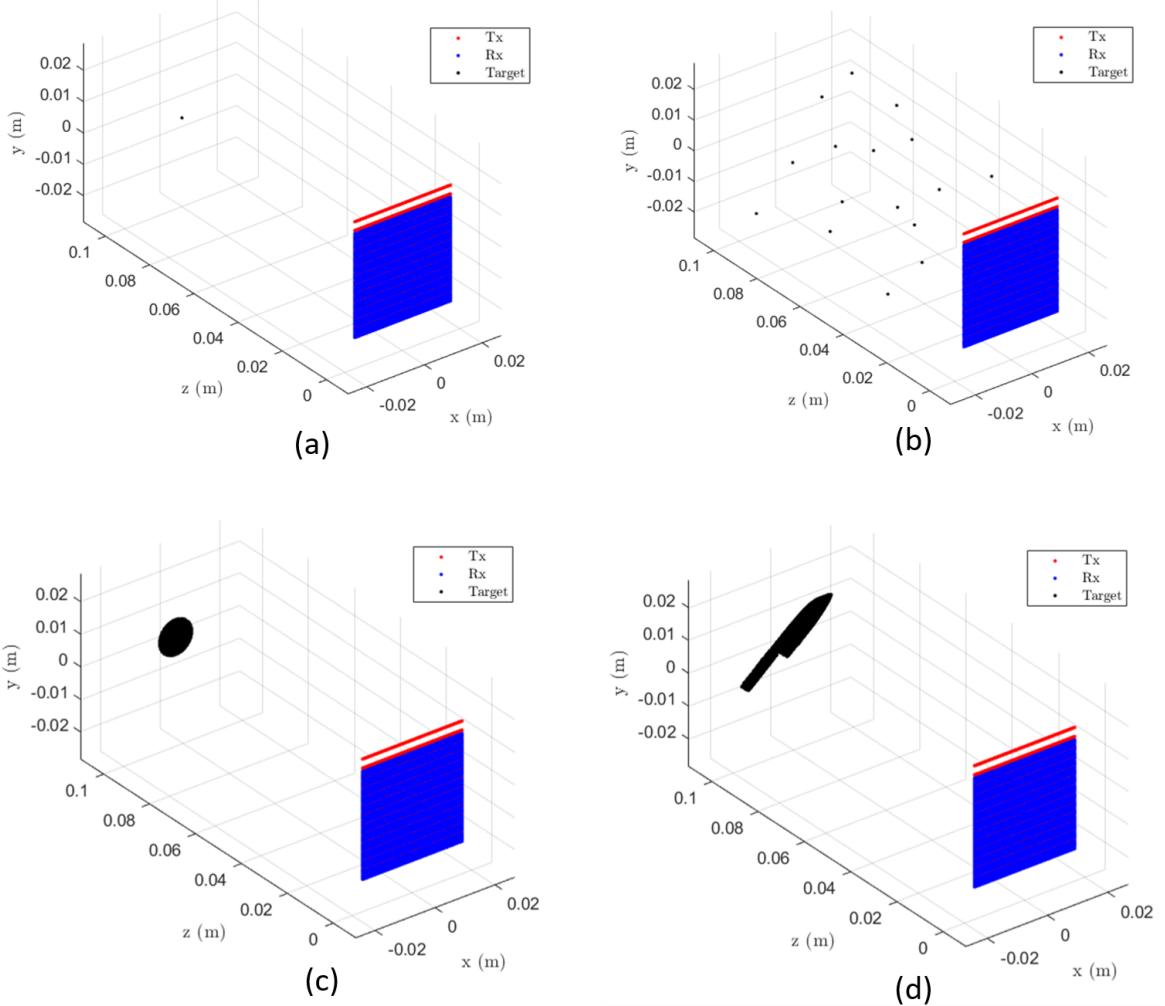
Figure 4.5: Target scenes displayed by the interactive GUI with a rectilinear scanning pattern. a) Single point reflector located at $(0, 0, 0.1 \text{ m})$, b) Random points generated by the GUI, c) `circle.png` with the default positioning parameters, d) `knife.stl` with the default positioning parameters.

3-D target whose relative location, reflectivity, and downsampling factor can be specified by the user. Each vertex of the STL file is treated as an ideal point reflector, meaning that STL targets can comprise thousands to millions of point reflectors. Additionally, the user can specify the maximum value after which to crop the target along the $z$-direction. Given the geometry of the linear and rectilinear scans in this toolbox, cropping a dense 3-D image after some $z$-plane is due to the rationale that the points after that $z$-plane will not be reflected

back to the array as they are blocked by the portion of the target closer to the array. Fig. 4.5d shows the imaging scenario of a knife CAD model loaded into the toolbox as an STL file. The target scene can also be saved by the user in either the interactive GUI or the toolbox for reference or later reuse.

Generating a target scene using the toolbox API requires initializing a `THzTarget` object. The settings for each of the parameters in the GUI are laid out in detail in the documentation with examples but are omitted here due to length and repetition considerations.

Simulation of the THz beat signal for various scanning regimes can be accomplished using a variety of means. To ease the computational load, we adopt a simple approach to simulate the MIMO and EPC-SISO beat signals. First, the target is discretized into a finite number of voxels, $N_{target}$, to be simulated independently as point reflectors using superposition. The locations of the MIMO array elements at each scanning iteration are computed using the `THzAntennaArray` and `THzScanner` objects, where $N_s$ is the number of synthetic elements in the array. Thus, the computation time can be reduced by precomputing the two-way distances between the antenna elements and the target voxels as an $N_s \times N_t$ array $\mathbf{R}$, where $N_s$ is the number of antennas in the SAR array, and $N_t$ is the number of target voxels. Similarly, the discrete reflectivity function $p(x, y, z)$ is treated as a column vector $\mathbf{p}$ of size $1 \times N_t$. The antenna pattern is modeled using a complex weight matrix $\mathbf{W}$ of size $N_s \times N_t \times N_k$ or $N_s \times N_t$. The weight matrix is computed if the user employs a non-isotropic antenna; otherwise, it is ignored. If the user employs a nonideal antenna pattern, the matrix $\mathbf{W}$ is computed by determining the angles between each antenna element and each target voxel. Optionally, the user can simulate the antenna pattern across frequencies and $\mathbf{W}$ is computed for each frequency. The simulation can now be generalized for any scan type as

$$s[n_s, n_k] = \sum_{n_t=1}^{N_t} \frac{\mathbf{p}_{1,n_t} \mathbf{W}_{n_s,n_t,n_k}}{\mathbf{R}_{n_s,n_t}} e^{-j(k_0 + \Delta_k n_k)\mathbf{R}_{n_s,n_t}}, \tag{4.1}$$

72

where $n_s$ is the SAR antenna index, and $n_t$ is the target voxel index. Equation (4.1) can be efficiently computed by vectorizing across $n_k$, $N_t$, or $n_s$. In most cases, $N_k << N_t, N_s$; thus, it is desirable to first attempt to vectorize across $n_t$, $n_s$ for computational efficiency, although it may be the most memory intensive. In our investigations, the costliest operation, computing the complex exponential, can be computed quickly on a GPU because it is highly parallelizable.

Other attempts to improve the computational complexity of this process include writing a custom implementation in a more efficient language, such as C/C++ or CUDA, or calculating the complex exponential more efficiently. The latter method is employed in our implementation in MATLAB when no GPU is available. The procedure leverages the presence of any repeated values in the quantity $(k_0 + \Delta_k n_k)\mathbf{R}_{n_s,n_t}(\mathrm{mod}\ 2\pi)$ by finding unique elements and computing the complex exponential only at these values. However, finding the unique elements of the $\mathbf{R}$ matrix can be computationally prohibitive when $N_t$ and $N_s$ are large. On the other hand, it is often more efficient than computing the entire complex exponential on the CPU for the same large target.

To perform the routine described above, the user can press the "Compute Beat Signal" button on the interactive GUI. During the simulation, the user will be provided with a dialog box showing the simulation progress and estimated time to completion. Once the simulation is complete, the "Beat Signal Computed" lamp will be green, indicating the simulation success. If the user changes the target scene, SAR scanning scenario, antenna array, or waveform parameters, the indicator will return to red, as the current imaging scenario has not been simulated. If the user is using the toolbox API, this process is equivalent to calling the `Compute()` method of the `THzTarget` object, e.g `target.Compute()`. Similarly, a dialog box appears, showing the progress and estimated time until completion.

Creating a target scene using the imaging and simulation toolbox is both flexible and efficient, allowing the user to import target scenes in three different manners and efficiently

compute the beat signal. Thus, our freely-available toolbox provides easy access to many researchers wanting to learn more about near-field THz and sub-THz imaging domains, prototype a system design, develop custom image reconstruction algorithms, or programmatically generate large imaging datasets.

### 4.1.5 High-Resolution Image Reconstruction

The final step in the imaging process is to recover the image from the simulated signal. Numerous image reconstruction algorithms have been investigated in the literature [8, 23, 25, 39, 41, 43, 46, 51, 52, 53, 115, 136]. Our toolbox includes 11 built-in documented image reconstruction algorithms for the four SAR scanning modes. The provided algorithms have been made public to increase the accessibility of the implementation to the research community. In this subsection, we focus on using the interactive GUI and toolbox API to execute the included image reconstruction algorithms.

After navigating to the "Reconstruct Image" tab in the included GUI (see (5) in Fig. 4.2) the user can enter the desired parameters in the panel to the left and select the desired image reconstruction algorithm from the drop-down menu. The application automatically verifies the scan type and parameters to ensure that the selected reconstruction algorithm can be applied to the simulated imaging scenario. The GUI may inform the user of errors in the image reconstruction parameters, allowing the user to change the parameters to recover the desired image. If the user has selected to use the MIMO array rather than the monostatic-equivalent SISO virtual array, there is an option to use the multistatic-to-monostatic approximation detailed in (2.9) at a specific $z$-reference plane, as discussed in Chapter 3. Some parameters on the left are unused for some image reconstruction algorithms. For example, the back-projection algorithm (BPA) does not compute a fast Fourier transform (FFT) and thus does not use the first three fields containing the number of FFT points along the $x$-, $y$-, and $z$-dimensions. The remaining parameters refer to the reconstructed image domain, which is

referred to as the imaging domain, specifically its bounds and the number of voxels along each dimension.

For the included RMA, BPA, and FFT reconstruction algorithms, the theoretical limits on each imaging domain dimension are shown to inform the user before image reconstruction. Once the desired parameters are entered, the user can reconstruct the image using the desired technique. The image reconstruction algorithm is executed, and a dialog box appear to display the computation progress. As expected, any BPA reconstruction requires an exceptional amount of computation time, as discussed in [30, 52, 53]. However, the efficient Fourier-based algorithms are capable of reconstructing images much more quickly. Once the image has been reconstructed, it is automatically displayed in the "Graphical Window." The user also has the option of displaying the image again with a different threshold (in dB) by modifying the parameters on the right or opening the image in the MATLAB built-in Volume Viewer application.

Using the toolbox API, a `THzImageReconstruction` object must be initialized. Table 4.2 summarizes the properties of the `THzImageReconstruction` class and their corresponding parameters in the interactive GUI. Further discussions and examples of the `THzImageReconstruction` class are provided in the documentation. To compute the image reconstruction algorithm, the user must simply call the `Compute()` method of the `THzImageReconstruction` class, e.g. `im.Compute()`. The user can then display the image by calling the `Display()` method with the option of setting the minimum threshold in dB as the property `dBMin`. Alternatively, the user can open the image in MATLAB Volume Viewer by calling `volumeViewer(im.imXYZ)`.

Our open-source simulation and image reconstruction software offers a user-friendly experience for image reconstruction, providing the user with documented, efficient implementations of common image reconstruction algorithms and a seamless platform for algorithm comparison and development. Our toolbox streamlines the process for researchers by providing documented tools with proper interconnections for waveform parameter setup at any

Table 4.2: Image reconstruction parameters in the interactive GUI and their corresponding properties in the `THzImageReconstruction` class.

| Parameter in GUI | Property in Class |
|---|---|
| Reconstruction Algorithm Name | `method` |
| Number of X FFT Points | `nFFTx` |
| Number of Y FFT Points | `nFFTy` |
| Number of Z FFT Points | `nFFTz` |
| X Min (m) | `xMin_m` |
| X Max (m) | `xMax_m` |
| Y Min (m) | `yMin_m` |
| Y Max (m) | `yMax_m` |
| Z Min (m) | `zMin_m` |
| Z Max (m) | `zMax_m` |
| Num X Voxels | `numX` |
| Num Y Voxels | `numY` |
| Num Z Voxels | `numZ` |
| Use GPU | `isGPU` (Boolean) |
| Multistatic-to-Monostatic Conversion | `isMult2Mono` (Boolean) |
| Z Reference (m) | `zRef_m` |
| Min dB | `dBMin` |
| Font Size | `fontSize` |

frequency, bandwidth, etc., creating any MIMO antenna array with an antenna gain pattern, utilizing four different scanning regimes, importing target scenes from a list of point targets, PNG image files, and high-quality CAD models, and image reconstruction. Thus, the entire simulation process is encapsulated into a simple tool designed for a host of applications.

## 4.2 Overcoming Array Perturbation using Vision Transformers

In this section, we develop a novel super-resolution algorithm for near-field SAR under array perturbation. As 5G mmWave devices become increasingly affordable and available, high-resolution SAR imaging is feasible for end-user applications and non-laboratory environments. Emerging applications such as freehand imaging [34], unmanned aerial vehicle (UAV) imaging [35], and automotive SAR [142] face several unique challenges for high-resolution

imaging [94, 112]. First, recovering a SAR image requires knowledge of array positions throughout the scan. Although recent work has introduced camera-based positioning systems capable of adequately estimating the position, recovering the algorithm efficiently is a requirement to enable edge and IoT technologies. Efficient algorithms for near-field SAR have been developed; however, they suffer from image defocusing under position estimation error and can only produce medium-fidelity images when the array is perturbed. In this section, we introduce a mobile-friend vision transformer (ViT) architecture to perform SAR image super-resolution and artifact mitigation under array position errors. The proposed algorithm, Mobile-SRViT, is the first to employ a ViT approach for SAR image enhancement and is validated through simulation and empirical studies.

Traditional SAR imaging requires high-precision laboratory equipment for exact positioning of the antennas throughout the scan. Efficient SAR imaging algorithms have been explored extensively in the literature [23, 39, 46] leveraging the fast Fourier transform (FFT) to recover images from radar data. These efficient algorithms strictly require specific synthetic aperture geometries, such as planar [23, 39], cylindrical [46], etc. to achieve high-resolution imaging. However, with the emergence of 5G and IoT technologies, near-field SAR sensing has received attention at both the system and algorithm levels [34, 94]. Although these applications operate at similar frequencies to traditional laboratory SAR [115], they suffer from two primary constraints: 1) the resulting synthetic aperture has some errors in the estimated position of each array element, and 2) because the image computation typically takes place on a low-power or mobile device, the computational load must be reduced compared to conventional imaging. Consequently, recovering high-fidelity images under such conditions remains an open challenge.

Previous research efforts [94, 112] do not consider errors in position estimation present in practical implementations [34]. In [39], an extensive investigation into practical system design challenges was presented, and a method was developed for synchronizing SAR equipment for

highly precise positioning. However, sophisticated scanners are infeasible for many emerging applications and positioning is degraded owing to noise in positioning or estimation. Such array perturbations cause defocusing and distortion of SAR images as the imaging algorithms improperly compute the matched filter weights based on noisy position estimates. Without knowledge of the exact positions, removing the distortion present in SAR images from the position errors remains an open challenge. For many practical systems, these algorithms can efficiently reconstruct only medium-fidelity images.

Recently, deep learning approaches for optical image super-resolution have been extended to the radar domain for SAR image super-resolution [54, 65, 66, 111]. Using convolutional neural network (CNN) architectures, previous efforts have been successful in improving SAR resolution [54, 66] and removing multistatic artifacts [65]. However, these techniques operate on SAR images collected using traditional techniques in laboratory environments, and do not address the issues that arise because of erroneous array perturbations explored in this study. Nevertheless, deep learning has seen tremendous success in both the optical domain, for image restoration [56] and super-resolution [18], and on conventional SAR images [54, 65, 66]. Hence, deep learning may be a suitable solution for near-field SAR position error artifact mitigation and super-resolution.

Recent advances in computer vision have seen a shift from CNN-based architectures towards the attention mechanism [143] using Vision Transformer (ViT) techniques [144] to achieve performance gains with smaller model sizes [92, 93, 145]. In [92], the MobileViT architecture is presented leveraging a transformer architecture for image classification. Later, the transformer architecture was employed for optical image super-resolution and artifact mitigation [56]. Transformer techniques have appeared in recent work on radar image classification [146] and gesture recognition [147]; however, transformers have yet to be employed for SAR image super-resolution. In this section, we introduce a novel transformer-based architecture for SAR image super-resolution under array perturbation, called Mobile-SRViT.

The proposed algorithm operates on images recovered by the range migration algorithm (RMA) [39] and produces high-fidelity images of intricate targets. We validate our mobile-friendly algorithm using simulation and empirical data from a near-field SAR scenario with positioning errors.
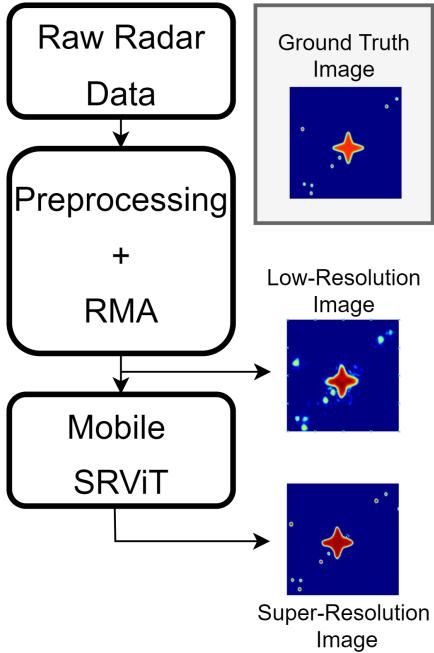


Figure 4.6: Operation of the Mobile-SRViT: the low resolution image produced by the RMA is restored by the Mobile-SRViT algorithm. The ground truth image is shown for reference.

### 4.2.1   ViT-Based Near-Field SAR Super-Resolution

In this section, we detail the proposed transformer-based approach for near-field SAR super-resolution and artifact mitigation. An overview of the Mobile-SRViT algorithm is shown in Fig. 4.6. The raw radar data are preprocessed to a planar monostatic scenario, and the RMA is applied to produce a low-resolution image with distortion, blur, and defocusing caused by imaging non-idealities. The proposed Mobile-SRViT operates on this image to produce a super-resolution image and restore the image quality, while preserving the intricate high-frequency details of the targets.

The Mobile-SRViT architecture is based on the MobileViT network employed for image classification [92]. Because our network is designed for image-to-image processing, the convolution layers are modified to adhere to a fully convolutional neural network (FCNN) framework, similar to the network proposed in [111]. Fig. 4.7 shows the implementation of the Mobile-SRViT algorithm, where "MV2" refers to the MobileNetV2 block proposed in [93]. Our algorithm adopts the approach of [92] such that the image is first processed by several
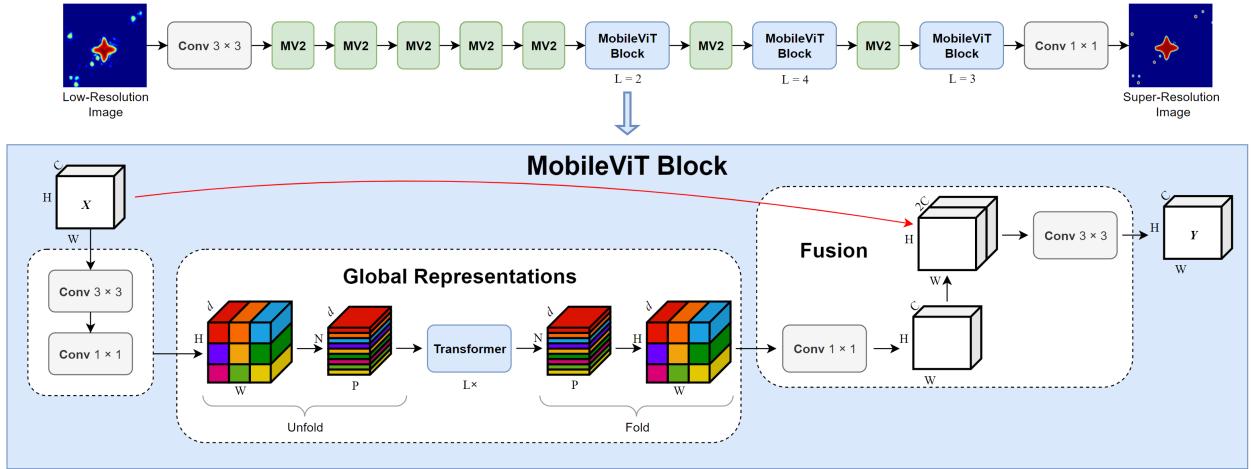


Figure 4.7: Mobile-SRViT architecture.

MobileNetV2 convolution blocks before alternating between MobileViT and MobileNetV2 operations. The MobileViT block is intended to model the global and local information of the input data with fewer parameters than the traditional ViT [144]. Given an input tensor $\mathbf{X} \in \mathbb{R}^{H \times W \times C}$, where $H$ is the height, $W$ is the width, and $C$ is the number of channels, the MobileViT block first applies a $3 \times 3$ convolution layer, followed by a $1 \times 1$ or pointwise convolution layer to produce a tensor $\mathbf{X}_L \in \mathbb{R}^{H \times W \times d}$. The last $1 \times 1$ convolution layer reduces the number of channels to match the input image. To learn global representations, $\mathbf{X}_L$ is unfolded into $N$ non-overlapping patches $\mathbf{X}_U \in \mathbb{R}^{P \times N \times d}$, where $P = 4$ and $N = HW/4$. Each of the $P$ patches is processed using a transformer architecture to encode the inter-patch relationships, yielding

$$\mathbf{X}_G(p) = \text{Transformer}(\mathbf{X}_U(p)), \quad p \in [1, \dots, P]. \tag{4.2}$$

80

Whereas most ViT implementations lose the positional location of each patch [144, 145], the MobileViT retains the patch order and pixel order within each patch. Consequently, $\mathbf{X}_G \in \mathbb{R}^{P \times N \times d}$ can be directly folded to obtain $\mathbf{X}_F \in \mathbb{R}^{H \times W \times d}$. The resulting tensor $\mathbf{X}_F$ is projected onto a low $C$-dimensional space via a $1 \times 1$ convolution layer before being concatenated with $\mathbf{X}$, yielding $\mathbf{X}_O \in \mathbb{R}^{H \times W \times 2C}$. Finally, using a $3 \times 3$ convolution, $\mathbf{X}_O$ is fused to form an output tensor $\mathbf{Y}$ of identical size to $\mathbf{X}$. Interestingly, the receptive field of the MobileViT block is $H \times W$ because $\mathbf{X}_U(p)$ encodes local information from a $3 \times 3$ region via convolutions, and each pixel in $\mathbf{X}_G(p)$ encodes global information over $P$ patches [92]. Our implementation maintains $C = 16$ until the last MobileViT block, where $C = 32$, $d = 2C$ for each MobileViT block, and $L = \{2, 4, 3\}$. The images are of size $256 \times 256$, and the patch size employed is $16 \times 16$. With this architecture, the proposed Mobile-SRViT has 69,122 parameters. Loss is computed using the pixel-to-pixel L1 metric as

$$L_{p2p} = ||\mathbf{X}_{SR} - \mathbf{X}_{HR}||_1. \tag{4.3}$$

Prior attempts at near-field SAR super-resolution have been purely CNN-based [54, 65, 66, 111]; however, the Mobile-SRViT detailed in this section is the first to leverage a transformer architecture for SAR imaging.

**Training the Mobile-SRViT**

The proposed algorithm was trained for 50 epochs using an ADAM optimizer on a single RTX3090 GPU with 24 GB of memory using 4096 samples for the training process and 1024 samples for evaluation. Training data were generated using the procedure detailed in [54, 94, 111]. Each sample was generated with additive white Gaussian noise (AWGN) with a signal-to-noise ratio (SNR) in the range $[-10, 50]$ dB and includes AWGN positioning errors with a standard deviation of 1 mm along the $x$-, $y$-, and $z$-directions to emulate a practical scenario. The training process lasted approximately 6 h with an inference time of 10 ms per sample during validation.

### 4.2.2 Mobile-SRViT Experimental Results

In this section, we conduct simulation and empirical experiments to verify the proposed algorithm. To evaluate the performance of the SRViT algorithm, we first compare the numerical performance of the Mobile-SRViT with the BPA and RMA. The gold-standard BPA has no requirements for SAR array geometry and is well suited for any scanning geometry. However, it is computationally prohibitive, particularly for mobile applications, as it computes the pixel-wise matched filter for every sampling location and frequency and is prone to distortion under array perturbations. The RMA, on the other hand, is more efficient than the BPA by computing the matched filter using Fourier relations. The proposed Mobile-SRViT algorithm attempts to compensate for the distortion present in the RMA images owing to imaging non-idealities.

Table 4.3: Quantitative performance of the Mobile-SRViT compared to the BPA and RMA.

| Metrics | Mobile-SRViT | BPA | RMA |
|---------|--------------|------|------|
| PSNR (dB) | **37.608** | 28.73 | 27.94 |
| RMSE | **0.017** | 0.057 | 0.061 |
| Time (s) | 1.113 | 1324.8 | **1.103** |

Using a test dataset consisting of 1024 samples similar to those in the training dataset but never seen by the network, we apply the Mobile-SRViT, BPA, and RMA to the samples and measure the peak signal-to-noise ratio (PSNR), root-mean-square error (RMSE), and computation time per sample. The results are presented in Table 4.3, where the best evaluations marked in boldface. All experiments were conducted on a desktop PC equipped with a 12-core AMD Ryzen 9 3900X running at 4.6 GHz with 64 GB of memory. The RMA achieves an efficient computation time of 1.103 s but falls short of the BPA in terms of RMSE and PSNR. However, the BPA boasts the highest PSNR and lowest RMSE of the classical algorithms but requires a significantly large amount of computation time, 1324 s. The Mobile-SRViT is superior to the other algorithms, even outperforming the BPA in PSNR

and RMSE, with a total computation time of 1.113 s required to compute the RMA and pass the image through the network. This qualitative analysis demonstrates the superiority of the proposed method in comparison with previous techniques in terms of both computational efficiency and image quality.
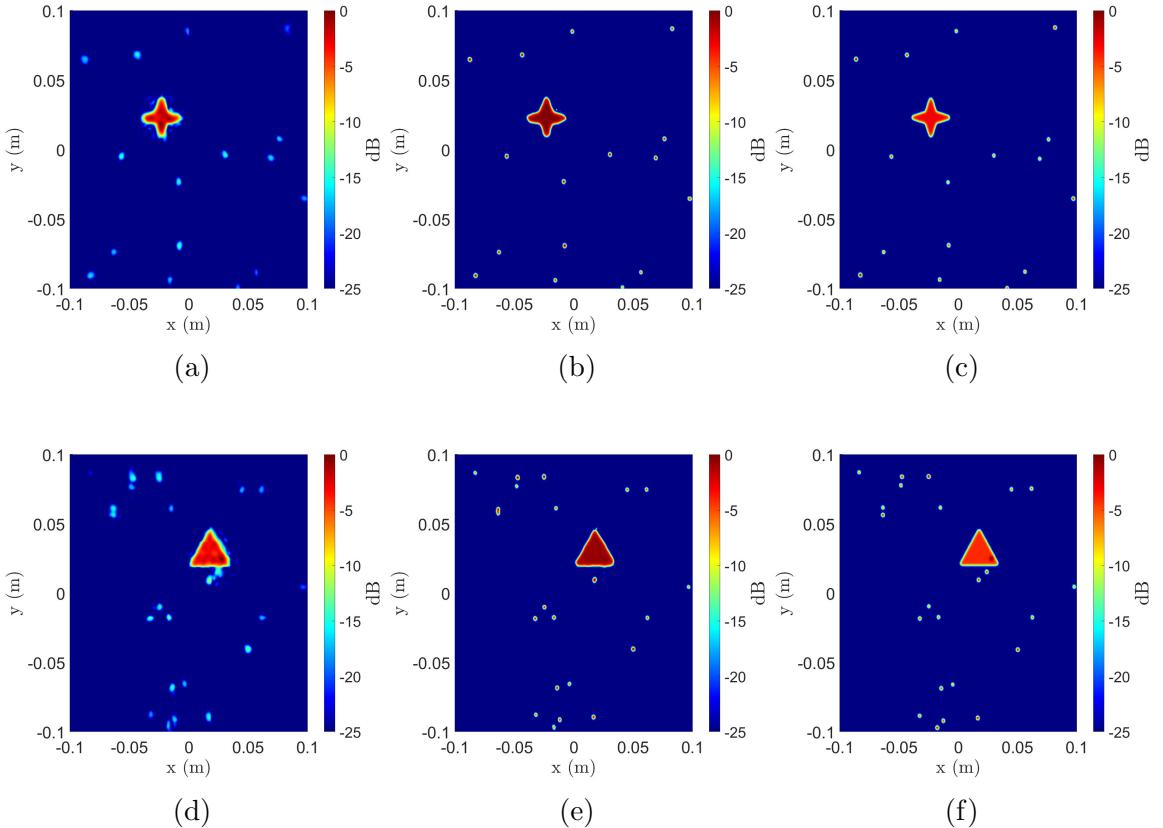


Figure 4.8: Imaging results using the Mobile-SRViT on synthetic data. The images in the first column, (a) and (d), are produced by the RMA and input to the Mobile-SRViT. The images in the second column, (b) and (e), are the super-resolution images output from the Mobile-SRViT. The images in the third column, (c) and (f), are the ground truth images.

We further validate the performance of the proposed algorithm via visual inspection of both the simulated and empirical data. Two samples from the testing dataset are compared in Fig. 4.8. For each sample, the proposed SAR super-resolution network recovers the solid object in addition to the point scatterers and mitigates the distortion caused by position estimation errors and system limitations. The super-resolution images (Figs. 4.8b and 4.8e)

are quite similar to the ideal images (Figs. 4.8c and 4.8f) showing an improvement over the medium-fidelity images recovered by the RMA.

To evaluate the performance of our proposed algorithm on empirical data, we first perform a SAR scan with array perturbations, as shown in Fig. 4.9a. After reconstructing the image with the RMA, as shown in Fig. 4.9b, the Mobile-SRViT is applied to achieve the super-resolution image shown in Fig. 4.9c. The proposed algorithm not only recovers a better-resolved image but also mitigates the multistatic artifacts visible in the RMA image [51, 94].

## 4.3 Lightweight Deep Learning SAR Super-Resolution for Irregular Scanning Geometries

Extending the EMPM preprocessing algorithm proposed in Section 3.3 for efficiently recovering MIMO-SAR images from irregular scanning geometries, we propose a novel CNN-based super-resolution architecture suitable for mobile applications. The EMPM suffers from increased sidelobes and distortion when data are collected across a large volume. Additionally, the results in Section 3.3 assume that the array positions are known without error; however, in the case of position estimation errors, which are common for realistic implementations [34], the image is further defocused for both the BPA and EMPM. In this section, we address the position estimation errors by introducing a novel data-driven super-resolution algorithm suitable for mobile near-field SAR imaging. Whereas previous CNN-based techniques for near-field SAR super-resolution are constrained to regular geometries [54, 55, 65, 66, 148, 149], we propose the first super-resolution algorithm designed for non-cooperative geometries. Furthermore, the proposed super-resolution CNN (SRCNN) is suitable for use in computationally constrained mobile applications. Additionally, previous studies on SRCNNs for near field radar imaging are limited by their datasets of exclusively randomly generated point scatterers [54, 66]. Using the toolbox designed in the previous section, our algorithm is
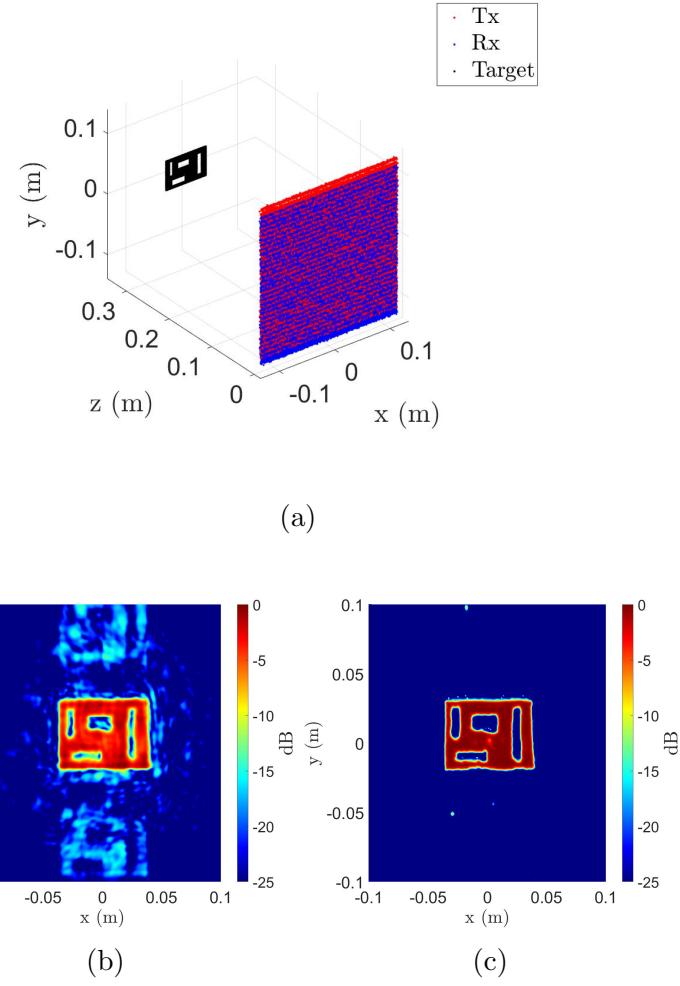
Figure 4.9: Imaging results using the Mobile-SRViT on empirical data from a near-field SAR scenario with array perturbations: (a) SAR sampling geometry, (b) the image reconstructed by the RMA and (c) the super-resolution image produced by the Mobile-SRViT.

trained on sophisticated, realistic targets rather than simple point scatterers. The proposed algorithm, referred to as "Mobile-SRGAN," is trained using a generative adversarial network (GAN) framework [150] and utilizes an efficient depth-wise convolution-based U-Net architecture [151] to obtain high fidelity SAR images with mobile-friendly computational complexity.

The proposed algorithm is summarized in Fig. 4.10. The raw freehand SAR data are processed by the EMPM algorithm, yielding a medium-fidelity image that suffers from im-

85
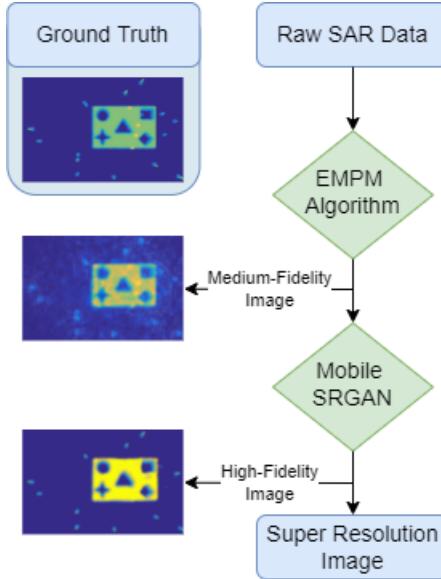
Figure 4.10: Flowchart of proposed design. After the raw SAR data are processed by the EMPM algorithm, the Mobile-SRGAN is applied to achieve a super-resolution image.

age distortion due to positional errors and the small array perturbation assumption in the EMPM. The example image in Fig. 4.10 consists of a solid object with cavities and randomly placed point targets. The proposed Mobile-SRGAN processes the distorted image reconstructed using the EMPM and outputs an enhanced super-resolution image.

The main contributions of this section are:

1) Generation of near-field SAR training datasets with sophisticated targets and irregularly sampled geometries.

2) Development of an efficient SRCNN model on the complex shapes and realistic datasets.

3) Qualitative and quantitative evaluation of the SRCNN model on both real and synthetic images.

### 4.3.1 Mobile Super-Resolution GAN

Using the reconstructed SAR images from the EMPM, we are tasked with mitigating the distortions in these images to generate super-resolution images that are reflective of the

86

original target images. For this purpose, we propose a cGAN-based [150, 152] Mobile-SRGAN framework that trains an encoder-decoder model to aggregate features for denoising, thereby generating a high quality SAR image. This framework consists of two CNNs: the Generator (G) and Discriminator (D), which work in tandem to optimize the model for achieving super-resolution. This particular training framework has previously been employed for medical MRI imaging [153], optical image super-resolution [154] and image-to-image translation [155, 156]. We adapt this architecture with depth-wise convolution operations to develop an efficient methodology that is suitable for mobile applications. The Mobile-SRGAN framework is illustrated in Fig. 4.11.
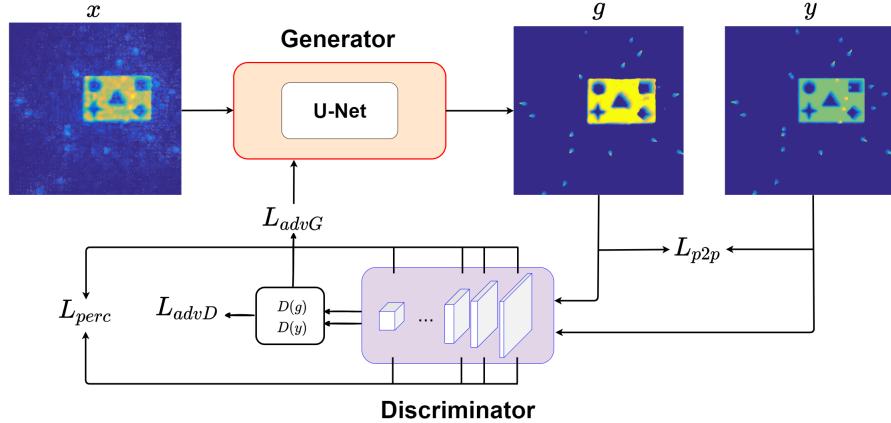


Figure 4.11: Mobile-SRGAN Overview

The training dataset consists of synthetic (reconstructed) SAR images and their corresponding ideal images, generated using the methodology in [54, 111], provided to the generator-discriminator pair. The objective of the generator is to map the distribution of the low-resolution SAR image to the high-resolution ideal image by learning to bridge the gap between these distributions. This learning is guided by the discriminator, which is a classifier established to quantify the probabilistic differences between the "clean" image ($g$) produced by the generator and the ideal target ($y$). The generator attempts to produce enhanced images by transforming the reconstructed SAR image, which is then processed by

the discriminator. This information is then fed back to the generator to guide the training towards the distribution of the ideal image. The network is trained in an adversarial manner, as shown in Algorithm 1, until the discriminator is unable to distinguish between the ideal and clean images, implying that the generator has adequately learned to generate images akin to the high-resolution images.

---

**Algorithm 1:** Training Process for Mobile-SRGAN

> **Data:** $D_{train}, D_{test}$          `// Train-Test Set`
>         $x, g$             `// in and out of G`
>         $y$             `// HR target images`
> **Result:** Trained Model

1   **for** $epoch \leftarrow 0$ **to** $MaxEpochs$ **do**
2      **for** $x \leftarrow 0$ **to** $batches\ in\ D_{train}$ **do**
3          $g \longleftarrow G(x)$
4          `/* Train Discriminator`       `*/`
5          $L_{advD} \longleftarrow log(D(y)) + log(1 - D(g))$
6          $L_{perc} \longleftarrow \sum_{i=0}^{L_D} \|F_{D_i}(y) - F_{D_i}(g)\|_1$
7          backprop D       `// Update D`
8          `/* Train Generator`       `*/`
9          $L_{advG} \longleftarrow log(D(g))$
10        $L_{p2p} \longleftarrow \lambda_{p2p}\|y - g\|_1$
11        backprop G       `// Update G`
12      `/* Evaluate Generator`       `*/`
13      PSNR$(G(D_{test}), y_{D_{test}})$
14      RMSE$(G(D_{test}), y_{D_{test}})$
15   **return** $G$

---

The adversarial training for the generator-discriminator architecture is based on the Binary Cross-Entropy Loss ($BCE$), as described in equations (4.4) and (4.7). The loss terms, $L_{adv}$, $L_{perc}$ and $L_{p2p}$, are combined to train the generator and the discriminator as illustrated in Fig. 4.11. The role of the discriminator is to differentiate between clean (generated) and ideal images based on their features, thereby labeling them as clean (0) or ideal (1), using the perceptual loss as calculated in (4.5). Additionally, the pixel-wise loss term penalizes the discrepancy between the exported image and the clean target (4.6).

$$L_{advD} = BCE(D(y), 1) + BCE(D(g), 0), \tag{4.4}$$

$$L_{perc} = \sum_{i=0}^{L_D} \|F_{D_i}(y) - F_{D_i}(g)\|_1, \tag{4.5}$$

$$L_{p2p} = \lambda_{p2p} \|y - g\|_1, \tag{4.6}$$

$$L_{advG} = BCE(D(g), 1), \tag{4.7}$$

where $F_{D_i}$ is the intermediate activation map extracted from the $i$-th convolution layer of the discriminator, $L_D$ is the number of convolution layers in the discriminator, and $\lambda_{p2p}$ is a constant which weights the contribution of the pixel-wise L1 discrepancy.

We implemented the generator using an encoder-decoder variant [151], as illustrated in Fig. 4.12, by applying a spatial reduction step after every convolutional bottleneck on the decoder and a resolution recovery step in the encoder. On each bottleneck output, a corresponding residual connection forwards the required features, thereby avoiding loss of information. The novelty of this encoder-decoder architecture is that it factorizes the standard $3 \times 3$ convolution into a $3 \times 3$ depthwise convolution (DWC) and a $1 \times 1$ pointwise convolution (PWC) to drastically improve the efficiency [157].
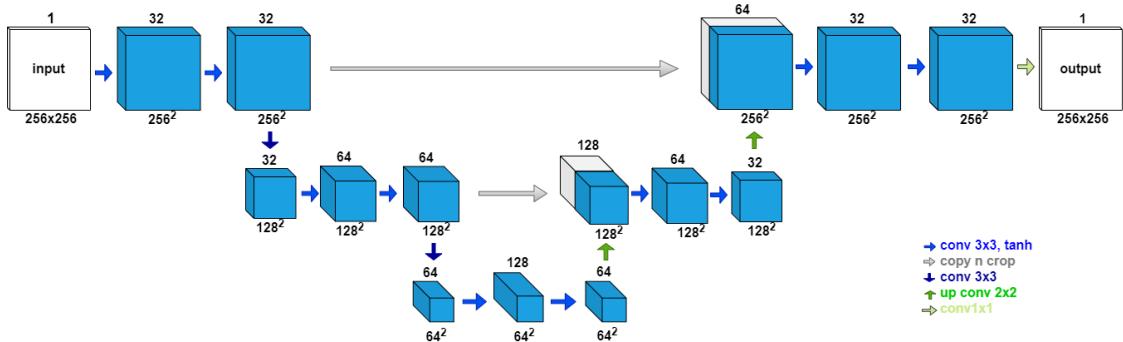


Figure 4.12: Mobile-SRGAN Generator Architecture

DWC and PWC play different roles in generating new features: the former captures spatial correlation information, whereas the latter quantifies channel-wise correlations. Em-

ploying DWC results in a substantial reduction in the CNN model size, with a reduction of up to 75% by applying a DWC on the first layer consisting of 32 feature channels. Hence, the proposed Mobile-SRGAN architecture with 79,233 model parameters is suitable for many mobile and smartphone applications [158]. By incorporating the DWC, resulting parameterization of the spatial reduction step occurs. This influences performance, allowing the training process to utilize spatial reduction as desired. The discriminator is based on the patch discriminator architecture introduced in [156], which divides the input into $16 \times 16$ patches and classifies each patch. The classification score is averaged for all patches [153].

The efficient mobile SAR super-resolution framework was trained for 50 epochs using the ADAM optimizer on a single TESLA P100 GPU with 16 GB of memory using 4096 samples for the training process and 1024 for the evaluation. The training dataset consists of SAR images containing randomly placed point scatterers, solid objects, and hollow objects, which improves over previous studies that only include images from point scatterers [54, 66, 111]. The training process was approximately 8.5 h an inference time during validation process of 14 ms per sample.

### 4.3.2 Performance Analysis of Mobile-SRGAN on Synthetic and Empirical Data

We evaluate the adversarial framework with a test set consisting of 1027 "never-seen-before" images (1024 synthetic images and 3 real images) excluded during the model training. The qualitative results are shown in Fig. 4.13. The network can produce rigid high-resolution images when processing new low-resolution data. By learning the joint distribution mapping between low- and high-resolution images, the generator can create consistent and well-structured objects by mitigating the distortion around the image while enhancing the appropriate local structure. The quantitative performance of the proposed framework can be measured by computing the peak signal-to-noise ratio (PSNR) and root mean square error (RMSE) between the generated images and ideal images of the synthetic testing dataset.

(a) EMPM          (b) Mobile-SRGAN          (c) BPA
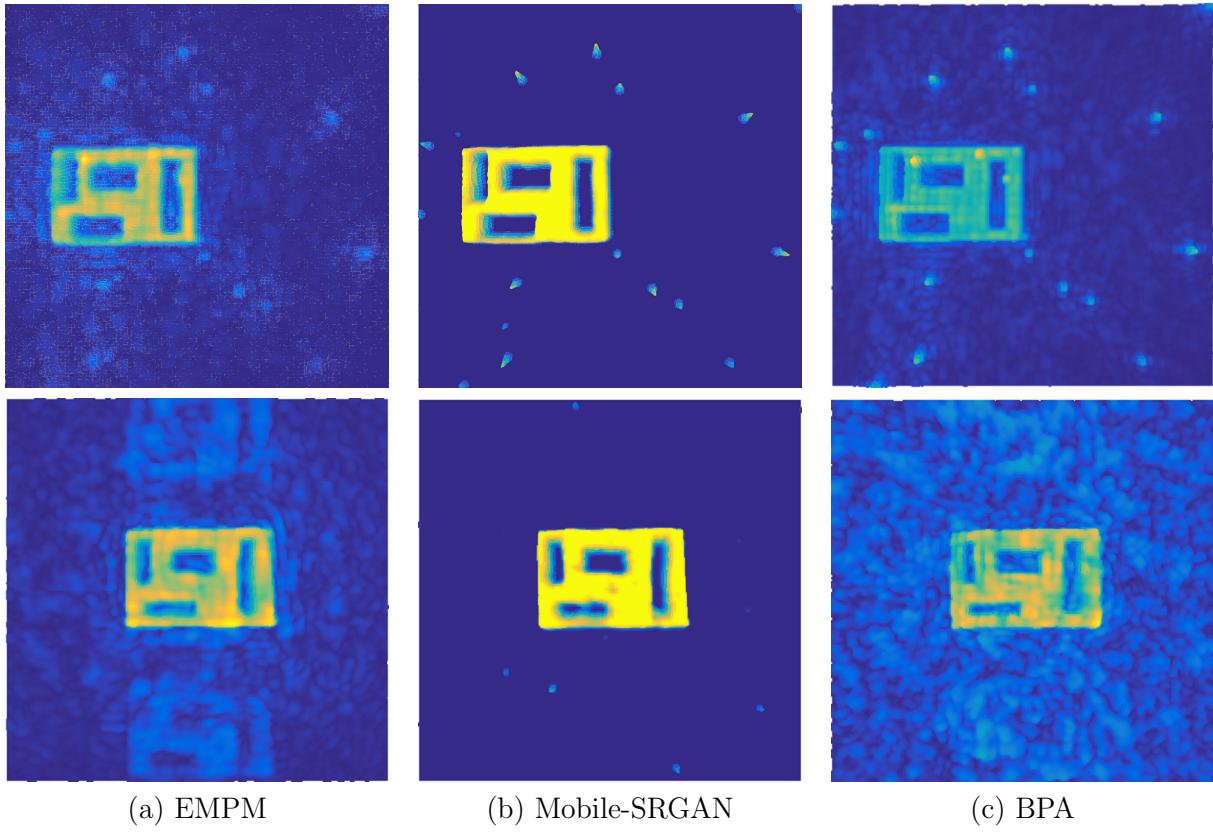
Figure 4.13: Qualitative Results: Comparison of clean images (a) EMPM, (b) Mobile-SRGAN, and (c) BPA. Results from synthetic images are shown on the first row, while results from real images are shown in the second row.

As shown in Table 4.4, the proposed method outperforms existing methods in terms of both computational efficiency and image quality. In particular, our algorithm achieves a higher PSNR and lower RMSE than the gold-standard BPA with only a slight increase in computation time for a single image compared to the EMPM and RMA.

Table 4.4: Quantitative performance of the Mobile-SRGAN compared to the BPA, EMPM, and RMA.

| Metrics | Mobile-SRGAN | BPA | EMPM | RMA |
|---|---|---|---|---|
| PSNR (dB) | **34.926** | 26.33 | 20.20 | 10.158 |
| RMSE | **0.019** | 0.044 | 0.105 | 0.276 |
| Time (s) | 1.117 | 1324.8 | **1.103** | **1.103** |

The Mobile-SRGAN is the first CNN-based super-resolution algorithm for mobile free-hand SAR imaging in the near-field. Rather than training on randomly placed point targets, we improve upon previous work by incorporating solid, intricate objects in the simulation that are more representative of real-world scenarios [54, 66, 111]. The proposed CNN algorithm is applied to images recovered by the EMPM algorithm [94], yielding high-resolution low-noise SAR images and outperforming previous techniques. The Mobile-SRGAN is the first SAR super-resolution algorithm developed for freehand radar imaging, which is a more difficult task than traditional SAR super-resolution, and efficiently recovers high-resolution images with low computational cost, deeming it suitable for computationally constrained applications.

In this chapter, we introduced a novel software platform for accelerating research into data-driven mmWave and THz imaging algorithms. Several deep learning-based algorithms were developed to overcome the non-idealities present in mmWave imaging applications, such as array positioning errors and image distortion from irregular scanning geometries. Furthermore, a focus is placed on computational efficiency, as many applications of interest are computationally limited. Exploring both front-end signal processing techniques and data-driven algorithms for improving mmWave imaging, we observe a lack of integration of signal processing and machine learning methods. The typical signal chain employs signal preprocessing prior to a data-driven algorithm, but separating these operations suffers from several limitations. The signal processing algorithms are unable to provide "context" that data-driven techniques employ to perform information extrapolation tasks such as super-resolution [18]. Additionally, most machine learning algorithms are developed and optimized for optical images and do not take into account the inherent characteristics of the FMCW signals, discussed in Chapter 2. To overcome these limitations, we introduce a novel framework for posing solutions to radar signal processing problems that leverages the strengths of both signal processing and machine learning, called hybrid-learning algorithms. By developing custom data-driven algorithms to exploit the properties of the signals, we notice

a significant improvement in network training convergence and final performance. Using signal processing techniques throughout the machine learning pipeline is shown to improve performance for several tasks including localization [111] and multiband image fusion [159]; however, the concept of hybrid-learning can be applied to a host of applications in signal processing and sensing [160, 161] to improve numerical and computational performance.

# CHAPTER 5

# HYBRID-LEARNING TECHNIQUES FOR CONTACTLESS MUSICAL INSTRUMENT INTERFACE

In this chapter, we extend the efforts in Chapters 3 and 4 to a more balanced approach to mmWave imaging and deep learning. Rather than approaching a problem from a particular perspective, we adopt a fusion of signal processing and machine learning techniques to develop end-to-end hybrid-learning algorithms, allowing the data-driven algorithm to offer insight throughout the image signal processing chain. Part of the following work was previously published in [111][1].

Using the interleaved hybrid-learning methodology, we design and implement a real-time system employing deep learning-based localization, classical signal processing algorithms, and a modified computer vision tracking algorithm. Although the proposed techniques are suitable for a host of tracking applications, this section focuses on their application as a musical interface to demonstrate the robustness of the gesture sensing pipeline and deep learning signal processing chain.

We apply a novel fully convolutional neural network (FCNN) to preserve the geometry of the image and perform super-resolution for improved localization. Radar signal processing using FCNNs is advantageous over other CNN techniques as it allows for data-driven "enhancement" rather than dimensionality reduction, as in classification. Hence, rather than suffering from information loss, the regressive FCNN provides additional "context" learned during the training phase to enhance the radar data. The enhanced data offer several advantages, such as improved SNR, clutter removal, near-field image correction, aliasing suppression, and higher-resolution peaks. In this section, traditional radar signal processing

---

algorithms are shown to achieve considerable performance gains when applied to enhanced data. Our novel approach unifies FCNN-based super-resolution with near-field imaging, which requires more difficult spherical-wave compensation, on a small (8-channel) array and is shown to improve hand-tracking performance significantly. This study is the first documented effort towards near-field radar image super-resolution using an FCNN approach for improved localization. Incorporating our enhancement FCNN in the signal processing chain enables fine motion tracking that is unattainable by existing techniques. Additionally, a particle filter tracking algorithm is presented to further improve tracking robustness by employing the Doppler effect. Compared to prior work on gesture tracking using optical solutions [162, 163], our approach offers precise hand-tracking using a single mmWave sensor, offering higher depth resolution with superior privacy. This section proposes a novel hand-tracking method for musical interface by fusing spatiotemporal algorithms, deep learning-enhanced feature extraction, and robust position tracking algorithms.

## 5.1 System Model for Radar Musical Interface

In this section, we provide an overview of the system model employed by the FCNN-based musical interface and examine the spatiotemporal features of a target in motion. The musician's hand is modeled as a point reflector located at the point $(y, z)$, as shown in Fig. 5.1. To achieve high-fidelity 2-D localization, we employ the range migration algorithm (RMA) over traditional range-angle FFT methods [164, 165], whose localization accuracy is known to be inferior [27].

Furthermore, the aforementioned Doppler principle can be leveraged to extract the velocity of the target by Fourier analysis over successive chirps, as described in Chapter 3. The velocity is extracted from the recovered image, which is denoted by $\hat{p}(y, z)$, given the geometry in Fig. 5.1.
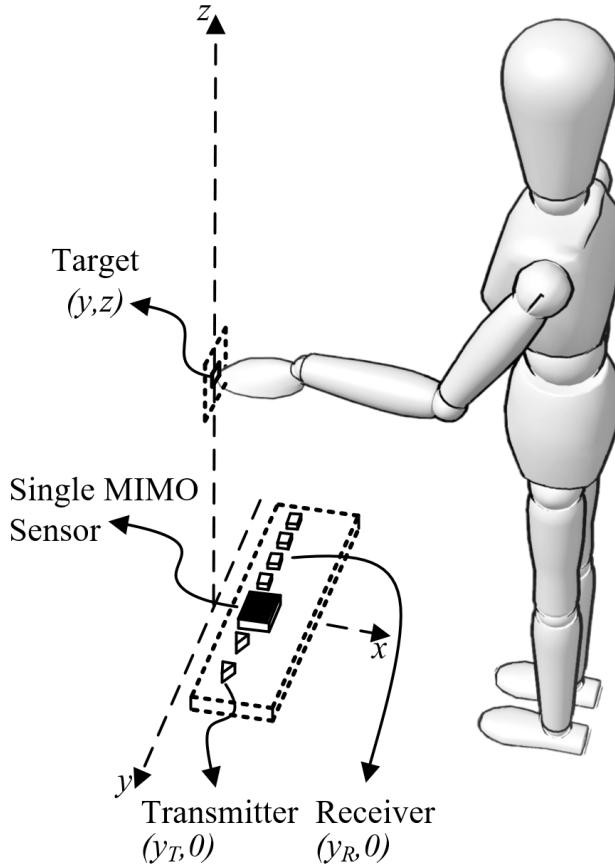
Figure 5.1: The imaging geometry, where the linear MIMO array faces vertically and the musician moves their hand throughout the $y$-$z$ plane.

## 5.2 Classical Spatiotemporal Feature Extraction Techniques

In this section, we introduce a simple approach to spatiotemporal sensing for contactless musical instrument interface. While our system generally tracks the 2-D position and velocity of the user's hand, we have identified three underlying features to achieve fine control of the musical interface: range, cross-range oscillation, and velocity. Based on the geometry shown in Fig. 5.1, we define the range as the position of the hand along the $z$-axis, that is, the vertical displacement between the sensor and the user's hand. Similarly, the cross-range direction is defined as the position of the hand along the $y$-axis. Subsequently, the cross-range oscillation is the rate at which the hand oscillates in the cross-range direction.

The velocity is given by the velocity of the hand with respect to the range $z$-axis. These parameters are selected such that the output musical interface is controlled primarily by the range of the musician's hand and secondarily by the cross-range oscillation and velocity. Throughout the remainder of this section, we refer to these parameters as features extracted from the radar beat signal.

Under the simple gesture tracking regime, the 2-D location and velocity $(\hat{y}, \hat{z}, \hat{v}_d)$ are extracted from the reconstructed image and buffer of recent images. In the next section, the three parameters extracted from the raw data are treated as a vector called the noisy measurement vector $\mathbf{r}$. Even in the ideal case, the spatial resolution of our system along the $y$ and $z$-directions is $\delta_y = 7.5$ cm and $\delta_z = 3.75$ cm, respectively [39]. Several other factors are not considered in the classical direct tracking method, including beam pattern, residual phase errors, and antenna coupling. To address these issues, we present a novel data-driven approach that employs an FCNN for super-resolution and image enhancement.

## 5.3 FCNN-Based Super-Resolution Feature Extraction and Particle Filter Tracking Methods

In this section, we improve upon the simple tracking techniques to overcome noise and foundational non-idealities in the imaging scenario, yielding a more robust algorithm. First, we develop a novel algorithm based on the well-known particle filter [166]. During the particle drift step of the algorithm, rather than shifting by a constant value, our method employs a variable weight factor depending on the corroboration of the new measurement with the Doppler velocity. An illustration of this process is provided in Fig. 5.2, and additional details can be found in [111].

Second, we propose a novel hybrid-learning technique to address issues such as instrumentation delay, ambient/device noise, multistatic effects, and non-spherical beam patterns.
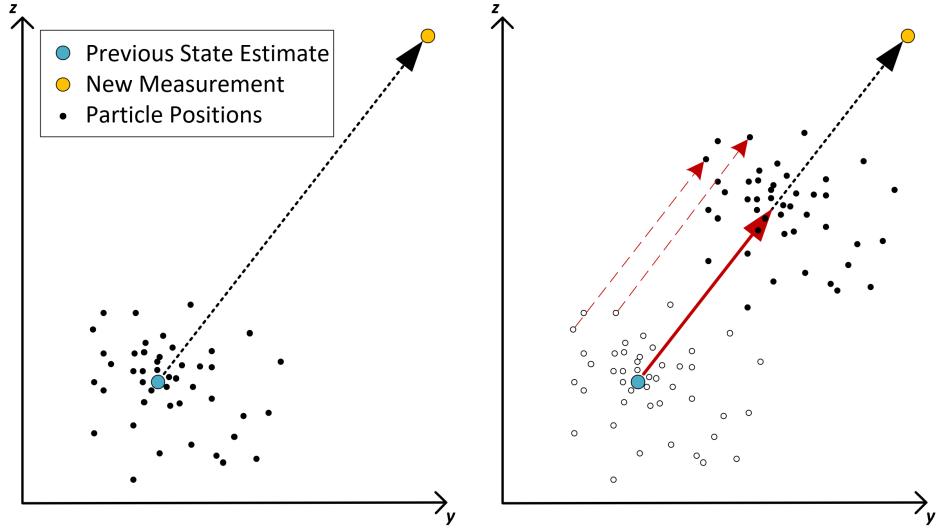
Figure 5.2: A visual example of the modified particle filter algorithm resampling process. The particle locations are resampled by a shift transformation towards the new measurement according to the weight vector $\mathbf{a}$, where $a_y = a_z = 0.5$.

We present a novel FCNN-based technique for image enhancement that improves 2-D position estimation, subsequent tracking accuracy, and Doppler spectrum SNR. To train the enhancement FCNN, we construct a dataset consisting of both real human hand data and synthetically generated data using the proposed MATLAB toolbox. During training, the FCNN learns the highly nonlinear relationship between distorted, blurred RMA images and the ideal images. Our novel training technique results in a robust and generalizable FCNN that improves the image SNR and localization by fitting to the non-ideal imaging constraints.

Additionally, by isolating the peak corresponding to the human hand,the clutter and noise at other positions are mitigated, thereby improving the Doppler spectrum SNR and subsequent velocity estimation. Thus, the FCNN enhances both the spatial and temporal features extracted from the radar beat signal before applying the particle filter. Uniting the proposed particle filter and enhancement FCNN, the range, cross-range oscillation, and velocity are robustly tracked by our novel algorithms and mapped to musical interface controls.
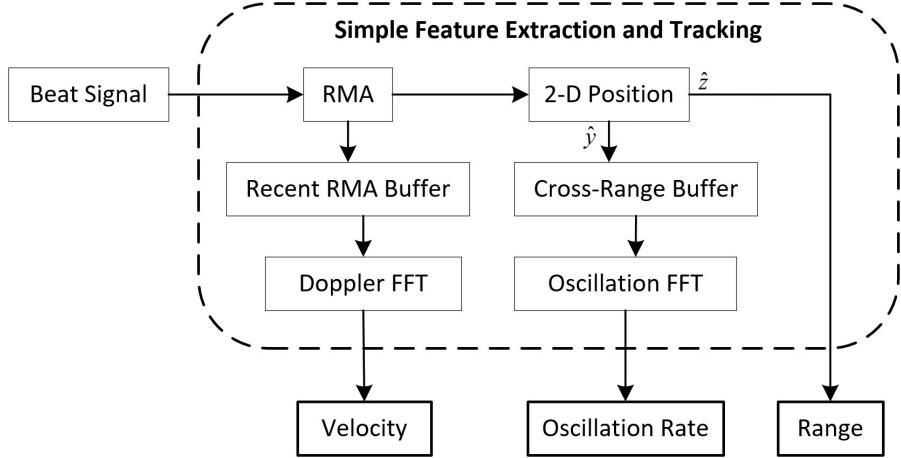
Figure 5.3: Simple tracking signal processing chain. After RMA is performed on the beat signal, features are extracted directly from the raw RMA image.

## 5.4 Simple and Enhanced Methods Compared

The signal processing chain for the simple feature extraction and tracking method is illustrated in Fig. 5.3. The beat signal is loaded into MATLAB, where the preprocessing discussed in the previous section is performed (RMA and peak finding). The reconstructed RMA image and raw features extracted by the classical techniques can be utilized by the particle filter algorithm and super-resolution FCNN to improve the tracking performance. The signal processing chain for the enhanced feature extraction and tracking method is shown in Fig. 5.4. Spatiotemporal features are output from the algorithm and can be used for many tracking applications.

## 5.5 Spatiotemporal Feature Extraction and Tracking Results

In this section, we provide an overview of the results of our novel tracking and feature enhancement algorithms, beginning with the simple, classical techniques, and compare the performance to our proposed methods. Our enhanced tracking regime demonstrates consid-
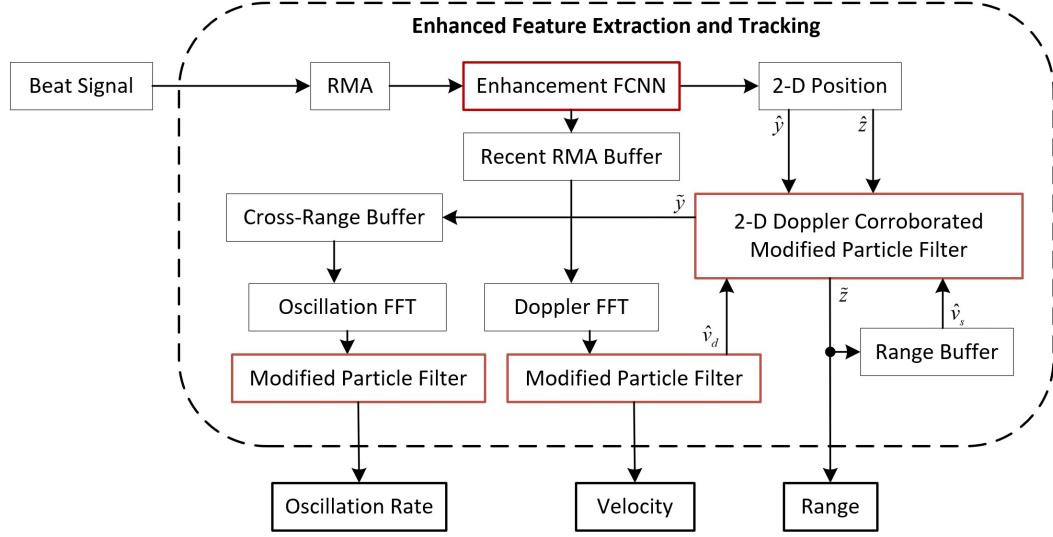
Figure 5.4: Enhanced tracking signal processing chain. Key elements to the enhanced methods are highlighted in red.

erable performance improvement compared with traditional methods and allows for robust super-resolution tracking on a small radar platform unattainable by existing methods.

### 5.5.1 Ground Truth - Ideal Motion Profile

A virtual prototyping approach is adopted to verify the feature estimation techniques. A point target is simulated in motion with $y$-$z$ location and velocity, as shown in Fig. 5.5, using

$$s(y_T, y_R, k) = \frac{p}{R_T R_R} e^{-jk(R_T + R_R)}, \tag{5.1}$$

where $R_T$ and $R_R$ are the distances from the transmitter and receiver to the point target, respectively. This ideal motion profile is employed to compare the tracking performance of our proposed methods with that of the traditional techniques. Empirical noise collected from the radar with an empty scene is added to each synthetic beat signal as

$$\tilde{s}(y_T, y_R, k) = \frac{p}{R_T R_R} e^{-jk(R_T + R_R)} + \alpha \tilde{\omega}(y_T, y_R, k), \tag{5.2}$$

100

where $\tilde{\omega}$ is a complex-valued noise sample corrupting the amplitude and phase of the ideal simulated beat signal and $\alpha$ controls the SNR.
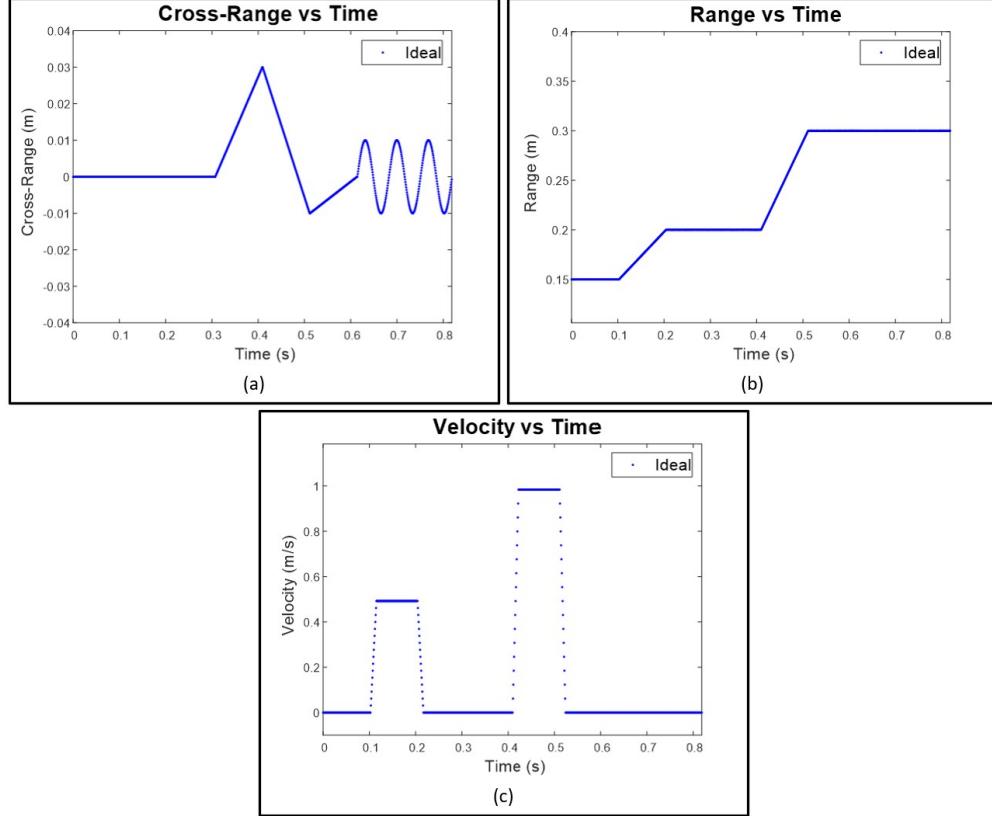


Figure 5.5: Ideal motion profile of the target in the (a) cross-range and (b) range directions as well as the (c) range velocity profile against time.

The motion profile in Fig. 5.5 shows the ideal range $(z)$, cross-range $(y)$, and velocity $(v)$ of the target. The motion profile includes independent and joint movements in the range and cross-range domains, in addition to sinusoidal cross-range oscillation. For our simulations, 4096 time samples are generated using $p \in [0.5, 1]$ to simulate the variance in the hand's empirical radar cross-section (RCS), as observed empirically from prior hand data, and $\alpha \in [1, 3]$ to vary the SNR among samples. Values for $p$ and $\alpha$ are selected randomly within the specified intervals for each time sample and provide a level of stochastic realism to the simulated data.

101

### 5.5.2 Classical Spatiotemporal Imaging Results

First, the simple tracking methods discussed in Section 5.2 are implemented to provide baseline performance metrics. The signal processing chain shown in Fig. 5.3 is performed to extract the spatiotemporal features. At each iteration, the features are extracted directly from the raw RMA images and are therefore prone to erratic behavior.
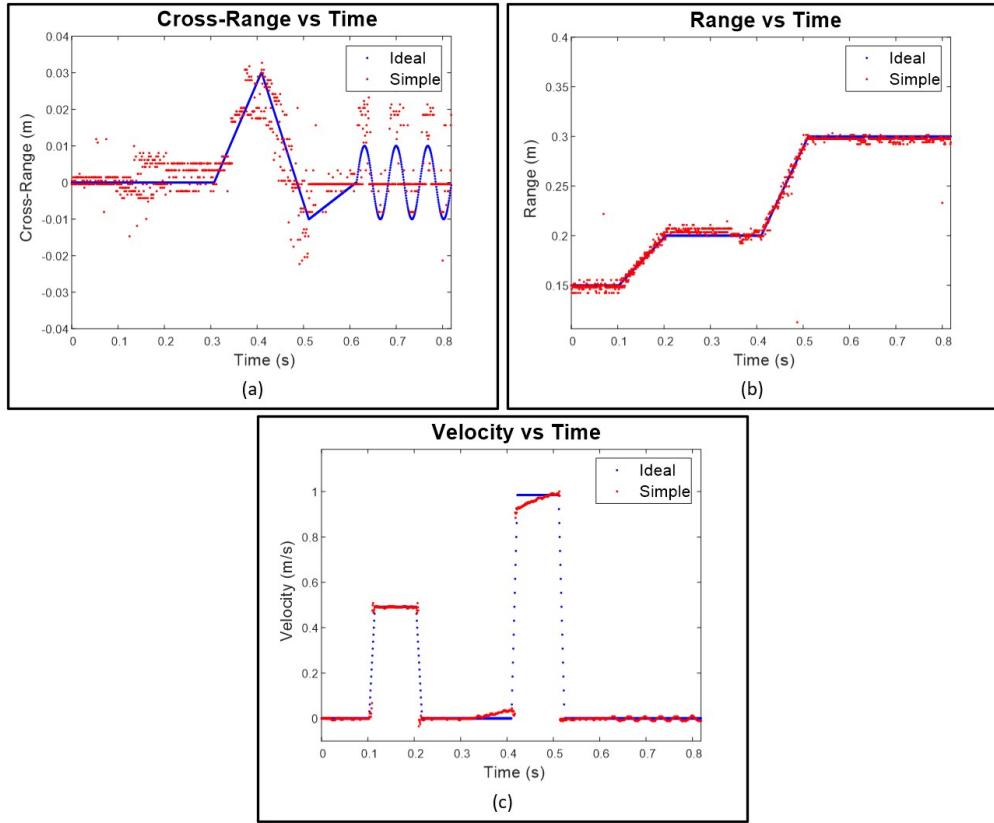


Figure 5.6: Motion profile using simple feature extraction techniques on each frame for every time step (red) compared with the ideal motion and velocity profiles (blue). The (a) cross-range and (b) range are measured directly from the peak of the RMA image of each frame and the (c) velocity is measured using the Doppler FFT of the raw RMA images.

Fig. 5.6 shows the features estimated from the data generated by (5.2) using the simple methods. The real radar noise and varying reflectivity result in outliers and errors in the estimated location and velocity of the target, particularly in the cross-range domain. Without more robust feature extraction and tracking techniques, the performance leaves much to

be desired. In the following sections, the performance of the simple tracking methods is quantitatively compared to the enhanced tracking methods and design considerations are discussed.

### 5.5.3 FCNN-Based Super-Resolution Tracking Results

Assuming the motion profile in Fig. 5.5, our proposed particle filter algorithm is employed in an attempt to robustly track the 2-D position and Doppler velocity of the target across time, significantly improving the user's control over the interface.

First, the particle filter algorithm (PF) without Doppler corroboration is implemented using the data in Fig. 5.6 as elements of the noisy measurement vector $\mathbf{r}$. The PF reduces the effect of noise on position estimation and improves spatiotemporal tracking performance, as shown in Fig. 5.7. Cross-range position tracking is the most improved compared to the traditional methods. Next, the Doppler-corroborated particle filter (DPF) is applied to the same set of data, further improving the estimation of the range. The outliers in Fig. 5.7b are mitigated by the DPF in Fig. 5.7d because the outlying samples result in a sample velocity $\hat{v}_s$ contradicted by the Doppler velocity $\hat{v}_d$ and are weighted as unimportant in the resampling process. The DPF algorithm improves the user experience of our interface by providing a robust and consistent tracking algorithm to smoothly estimate the 2-D position and spatiotemporal signatures of the user's hand. However, the PF and DPF can be further improved by implementing the proposed enhancement FCNN.

After the super-resolution FCNN is trained, a validation dataset identical in size to the training set is collected. Fig. 5.8 shows images enhanced by the enhancement FCNN, demonstrating the robustness of the network. Figs. 5.8a and 5.8b show simulated point targets enhanced by the FCNN, resulting in localization super-resolution. Fig. 5.8c shows an RMA image reconstructed from a real hand capture close to the middle of the cross-range domain. The 2-D position of the hand is located more accurately, as compared with the ideal image.
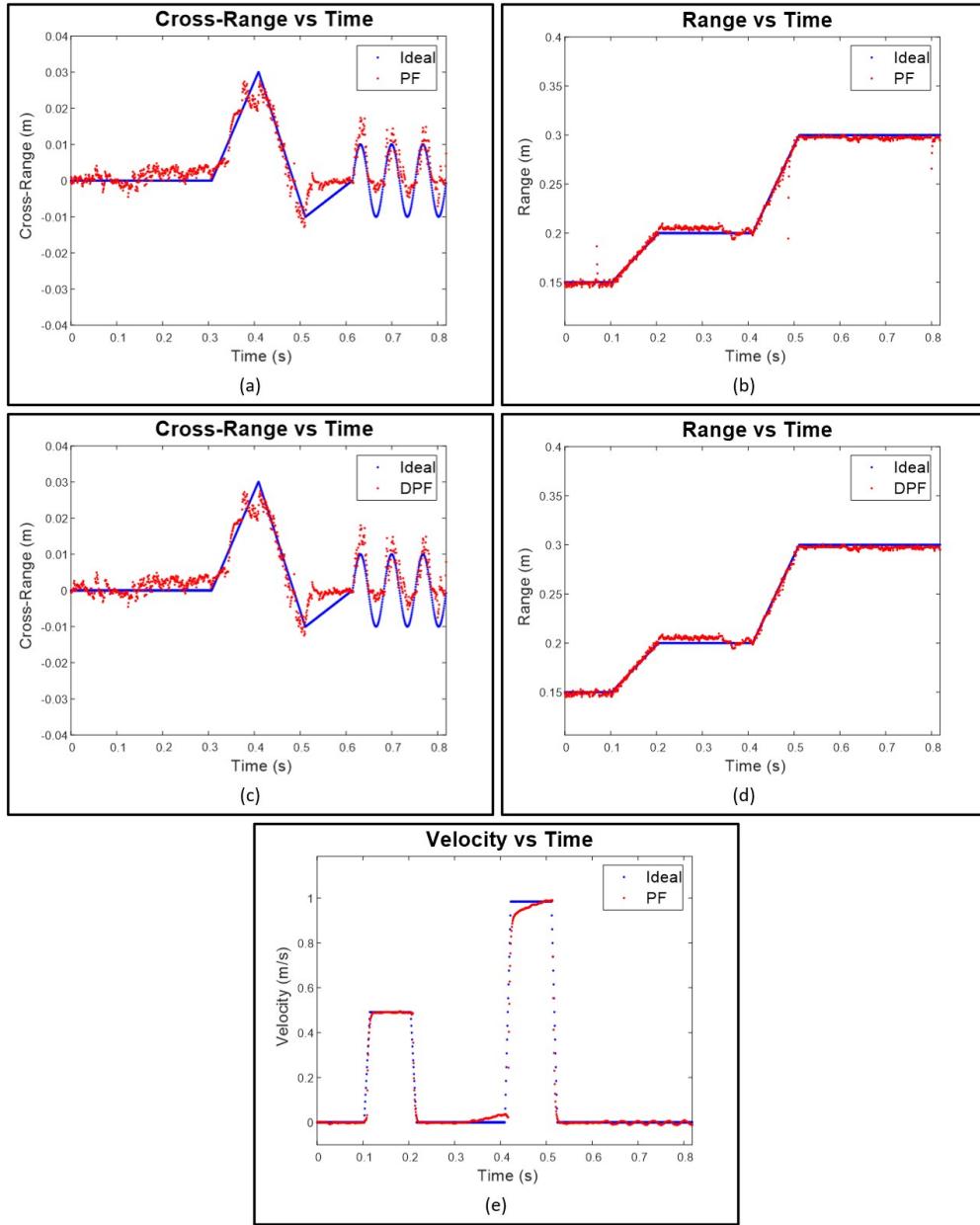
Figure 5.7: Particle filter (PF) and Doppler-corroborated particle filter (DPF) tracking. Improved tracking of the (a)/(c) cross-range and (b)/(d) range positions versus time using the PF/DPF with $N_z = 16$, and (e) Doppler velocity versus time using a PF approach.

Similarly, Figs. 5.8d-5.8f demonstrate the network's ability to enhance images degraded by small hand RCS in comparison to noise, ghosting due to non-ideal beam patterns, ambient and device noise, and other non-idealities. The proposed enhancement FCNN simultane-
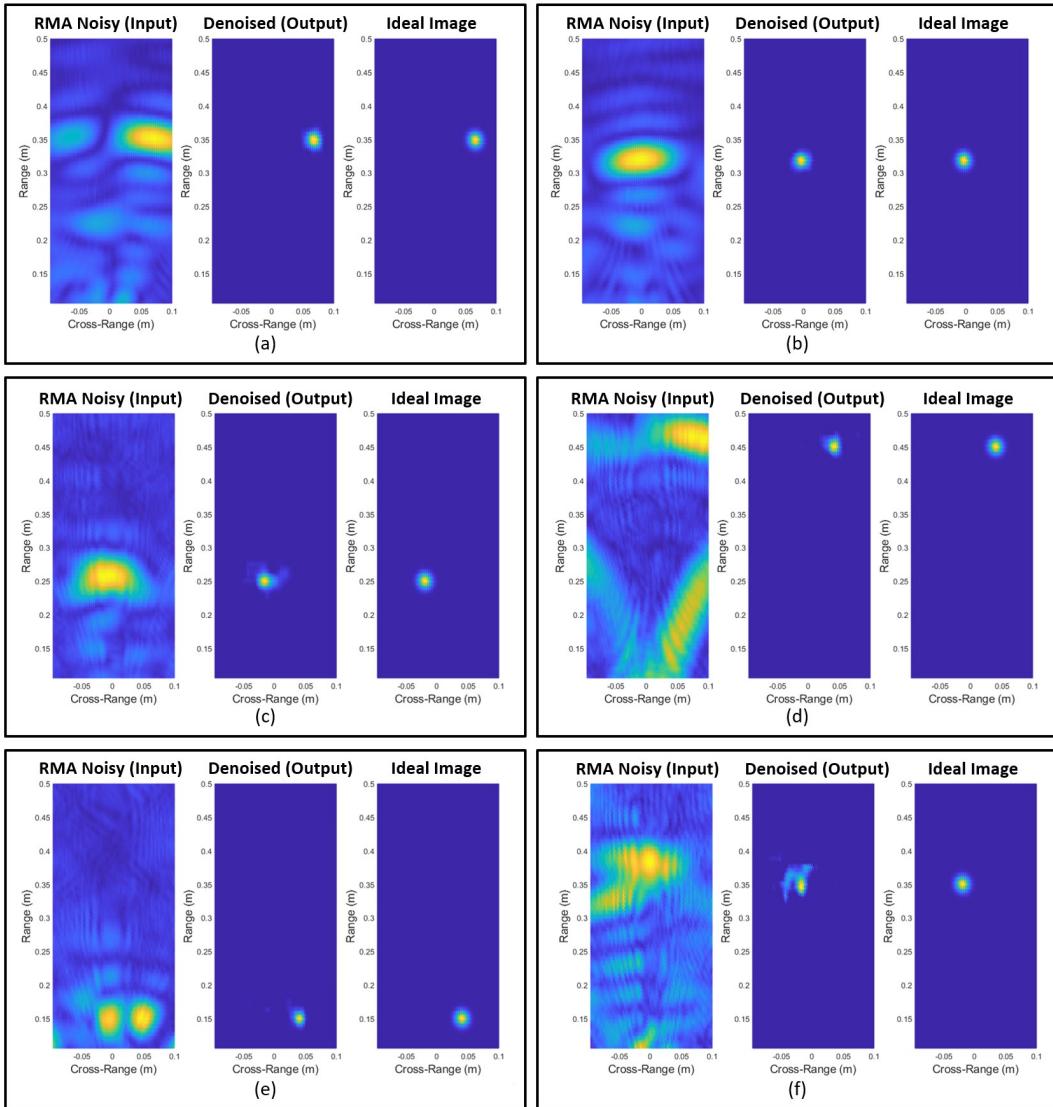
Figure 5.8: Enhancement FCNN applied to simulated (a,b) and real hand (c-f) RMA images for image enhancement and improved localization.

ously enables localization super-resolution and overcomes device and environmental issues. Hence, the features extracted from the enhanced images are much improved compared to the raw RMA images before the FCNN and result in superior tracking performance.

To quantitatively compare the localization improvement of the enhancement FCNN compared to the simple method, the RMSE for the range and cross-range position are computed on the validation dataset using the two techniques and are shown in Table 5.1, where the best

Table 5.1: Simple vs Enhanced Localization RMSE

| | $y$ (m) | $z$ (m) |
|---|---|---|
| Simple | 0.0154 | 0.023 |
| Enhanced | **0.0085** | **0.0083** |

evaluation is denoted in bold face. The enhancement FCNN improves both the resolution of the RMA images and the localization accuracy for both simulated and real data.
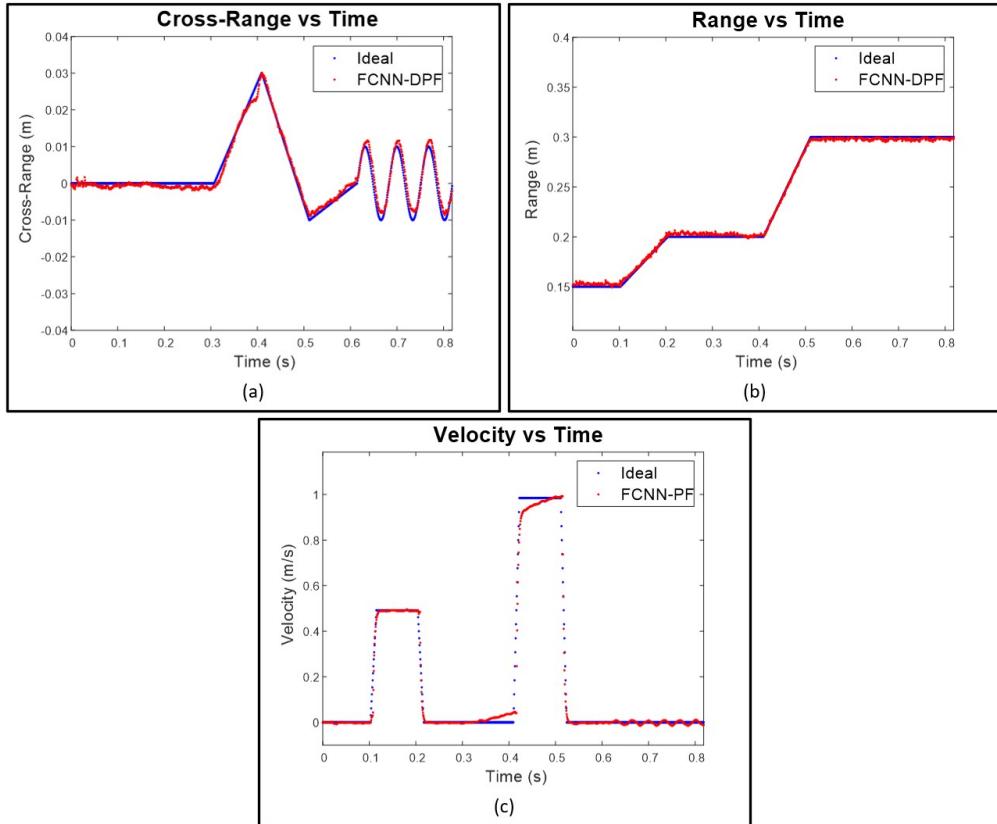


Figure 5.9: Spatiotemporal tracking with the FCNN-enhanced Doppler-corroborated modified particle filter algorithm.

Fig. 5.9 demonstrates the tracking performance of the FCNN and DPF (FCNN-DPF) on the same data as the previous tracking examples, an improvement over the DPF alone. By applying the FCNN-DPF, the range and cross-range tracking of the target is nearly identical to the ideal motion profile and an improvement in the velocity estimation. Using

identical sporadic data resulting in the poorly estimated cross-range positions in Fig. 5.6a, the FCNN-DPF yields an estimation nearly identical to the ideal motion profile. Similarly, the cross-range estimates in Fig. 5.7a and Fig. 5.7c are outperformed by the FCNN-DPF in Fig. 5.9a. Compared with the classical techniques and PF/DPF alone, the localization performance of the FCNN-DPF is considerably superior.

Furthermore, the FCNN is shown to improve the Doppler estimation robustness. As shown in Fig. 5.10, the Doppler spectrum SNR is improved when Doppler processing is performed on the enhanced RMA images as compared to Doppler processing on the raw RMA images. Hence, the enhancement network improves the reliability of the Doppler velocity estimation, aiding spatiotemporal tracking.
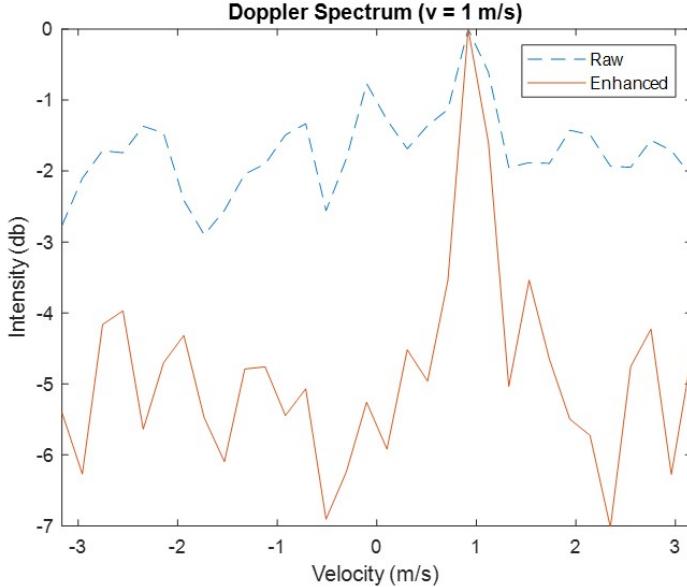


Figure 5.10: Comparison of the Doppler velocity spectrum when the Doppler FFT and video pulse integration steps are performed on the raw RMA images compared to the enhanced RMA images. The simulated data contains 128 frames and uses $\alpha = 3$ for every capture to simulate a low SNR scenario.

To quantitatively compare the tracking performance of the various proposed methods, 4096 unique motion profiles are generated, and the corresponding tracking RMSE is computed for the cross-range, range, and velocity. As presented in Table 5.2, the RMSE for the

cross-range ($y$), range ($z$), and velocity ($v$) improved with the novel algorithms proposed in this section.

As expected, the baseline simple method yields the greatest error for the three features. Comparing PF and DPF, the cross-range and velocity RMSE are identical between the two techniques, but the range RMSE is improved owing to the dynamic weighting technique. The FCNN alone outperforms the simple method but can be improved by including the PF and DPF after image enhancement. Finally, the FCNN-PF and FCNN-DPF yield identical results for the cross-range and velocity RMSE, as expected, but a significant improvement is noted in the range error. The results in Table 5.2 demonstrate the considerably superior tracking performance of the enhanced tracking methods, namely the FCNN-DPF, compared with the simple tracking method. The performance gain realized by implementing the super-resolution FCNN demonstrates the ability of the network to learn system noise and ambiguities during the training phase using both real and synthetic data.

Table 5.2: Performance Comparison of Tracking Methods (RMSE)

|  | $y$ (mm) | $z$ (mm) | $v$ (mm/s) | $\bar{\tau}$ (ms) |
|---|---|---|---|---|
| Simple | 7.86 | 22.0 | 72.4 | 2.29 |
| PF | 5.27 | 13.6 | 52.9 | 2.36 |
| DPF | 5.27 | 6.85 | 52.9 | 2.41 |
| FCNN | 7.74 | 12.3 | 58.4 | 2.67 |
| FCNN-PF | 3.70 | 7.44 | 44.5 | 3.92 |
| FCNN-DPF | 3.70 | 3.07 | 44.5 | 3.96 |

The average latency of each method, $\bar{\tau}$, is measured as the time duration between the new sample being captured and the estimation process being completed on that sample. The resulting estimates are streamed across the MIDI port or sent to the built-in audio signal generation tool. The addition latency contributed by the subsequent synthesis engine is highly dependent on the software used and device under test; thus, it is not considered as part of the latency owing to our methods. A thorough analysis and comparison of the performance of the algorithms is provided in [111].

The contactless musical instrument interface employs an interleaved hybrid-learning approach as insights from the mechanics of the problem are leveraged throughout the algorithm development for optimal localization and tracking. This hybrid approach is employed in the following analysis for multiband signal fusion demonstrating superior performance to previous signal processing and machine learning methods.

# CHAPTER 6

# DEEP LEARNING-BASED MULTIBAND SIGNAL FUSION FOR 3-D SAR SUPER-RESOLUTION

In this chapter, we propose a novel deep learning-based solution for multiband signal fusion to achieve high-resolution synthetic aperture radar (SAR) imaging. This approach, called $kR$-Net, employs a hybrid, dual-domain complex-valued convolutional neural network (CV-CNN) to fuse multiband signals. By exploiting the relationships in both the wavenumber-domain and wavenumber spectral-domain, the framework overcomes the drawbacks of existing multiband imaging techniques for realistic scenarios and achieves high-resolution imaging of intricate targets, enabling technologies such as concealed weapon detection and occluded object classification. The dual-domain architecture demonstrates improved performance over a network operating exclusively in the wavenumber-domain or wavenumber spectral-domain. Furthermore, a fully integrated multiband imaging system is developed using commercially available millimeter-wave (mmWave) radars for efficient multiband imaging. The proposed $kR$-Net is employed to achieve an effective bandwidth of 21 GHz from two radars each with a bandwidth of 4 GHz operating at 60 GHz and 77 GHz. Extensive numerical simulations and empirical experiments are conducted to demonstrate the superiority of our approach over existing methods for a diverse set of realistic 3-D SAR imaging scenarios. Part of the following work was previously published in [159][1].

## 6.1 Multiband Signal Model

In this section, we formulate a signal model for multiband radar signaling. In the multiband sensing scenario, samples are taken across multiple subbands separated by frequency gaps, as

---

shown in Fig. 6.1, where the radar subbands represent the operating frequency ranges of the radars. It is important to note that the subsequent analysis and proposed algorithm assume a weak or constant relationship between the scattering properties and frequency across the entire bandwidth spanned by the subbands. To achieve the desired resolution, multiband signal fusion methods are applied to recover the unoccupied frequency bins and obtain the equivalent wideband signal spanning the entirety of the subbands.
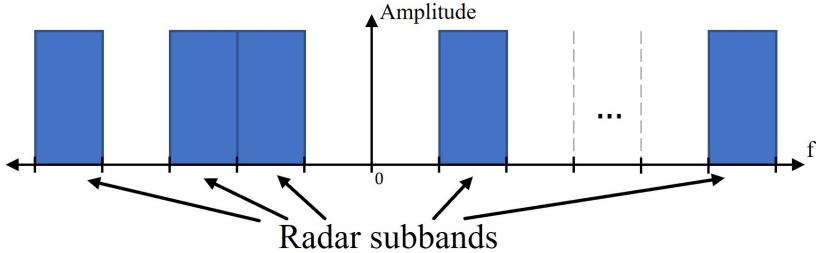


Figure 6.1: A baseband representation of multiband radar signaling where samples are gathered at multiple subbands separated by frequency gaps. The goal of multiband signal fusion is to recover the equivalent wideband signal spanning the entirety of the subbands.

For the remainder of this study, we focus on the dual subband scenario emphasized in the existing literature [72, 73, 74, 75, 76, 77, 78, 79, 80]. However, the following analysis and proposed solution apply to the generalized multiband fusion problem portrayed in Fig. 6.1, as detailed in Section 6.4.2. Additionally, although extrapolation of the signal beyond the highest and lowest subband frequencies has been proposed for previous techniques [76, 77], we focus on developing an algorithm to reliably impute the missing signal in the frequency gap.

In a generic near-field SAR or inverse-SAR (ISAR) scenario, two radars are mounted on a platform that scans a target scene for high-resolution imaging. Hence, provided proper system design, the synthetic aperture elements of both radars can overlap to produce a virtual monostatic element operating in the frequency ranges of both radars. Alternatively, the following fusion signal model can be achieved through other means, such as a monostatic

wideband system sampled at several subbands to reduce sampling bandwidth, a MIMO dual-band system whose virtual elements of each subband overlap, or modeling the problem as fusion in the angular spatial wavenumber domain [75], provided the radars are sufficiently close to each other during scanning.

Suppose the subbands start at frequencies $f_1$ and $f_2$, respectively, and illuminate a target in $x$-$y$-$z$ Cartesian space, where $z$ represents the downrange or range direction, and $x$-$y$ are known as the cross-range directions. Consider the monostatic element operating at both subbands and located at $(x', y', z')$ illuminating $N_t$ targets modeled as point scatterers, where the $i$-th target is located at $(x_i, y_i, z_i)$ with reflectivity $\alpha_i$. The wavenumber domain, or $k$-domain, response to a chirp signal at the first and second radars can be written as

$$s_1(n) = \sum_{i=0}^{N_t-1} \alpha_i e^{-j2(k_1+\Delta_k n)R_i}, \ n = 0, \ldots, N_k - 1, \tag{6.1}$$

$$s_2(n) = \sum_{i=0}^{N_t-1} \alpha_i e^{-j2(k_2+\Delta_k n)R_i}, \ n = 0, \ldots, N_k - 1, \tag{6.2}$$

where $k_1$ and $k_2$ are the wavenumbers corresponding to the starting frequencies $f_1$ and $f_2$, respectively, $k = 2\pi f/c$, $\Delta_k$ is the wavenumber sampling interval, $n$ is the time sample index, $N_k$ is the number of samples in each subband, and $R_i$ is the distance from the radar to the $i$-th scatterer, which is expressed as

$$R_i = \left[ (x' - x_i)^2 + (y' - y_i)^2 + (z' - z_i)^2 \right]^{\frac{1}{2}}. \tag{6.3}$$

Although the sampling conditions, sampling rate $\Delta_k$ and number of samples $N_k$, are considered identical across subbands for simplicity, this is not a strictly necessary condition as the subbands could have different sampling conditions. While different values of $N_k$ raises a trivial issue, different sample rates among subbands will need to be compensated such that the spectral domains are coincident. In addition, the proposed signal model assumes that the scattering parameters, $\alpha_i$, are frequency-independent. However, in a real scenario, scattering

112

properties of the various materials in the target scene vary across subbands, a phenomenon that may not be adequately modeled by (6.1) and (6.4), depending on the material properties and frequency ranges of the subbands. Since the proposed training data scheme assumes the frequency-independence of $\alpha_i$, networks trained on these data are limited because of this assumption, as discussed in Section 6.2.2.

Both $s_1(\cdot)$ and $s_2(\cdot)$ are considered multisinusoidal signals because they are composed of a superposition of scaled complex exponential functions, whose frequencies are determined by the ranges $R_i$. Hence, the wavenumber spectral domain, known as the range domain or $R$-domain, exhibits peaks at positions corresponding to the ranges $R_i$. Defining $\Delta_B \triangleq k_2 - k_1$, the difference between the starting wavenumbers, the signal at the second subband can be rewritten with respect to $k_1$ and different indexing as

$$s_2(n') = \sum_{i=0}^{N_t-1} \alpha_i e^{-j2(k_1 + \Delta_k n')R_i}, \ n' = \tilde{N}, \ldots, N, \tag{6.4}$$

such that $\tilde{N} \triangleq \Delta_B/\Delta_k$ is the offset between subbands 1 and 2, where $N \triangleq \tilde{N} + N_k - 1$, $n' = n + \tilde{N}$, and $\tilde{N} > N_k$. We assume that $\tilde{N}$ is an integer based on the choices of $\Delta_k$ and $\Delta_B$, although the derivation is valid regardless.

From (6.1) and (6.4), the multiband scenario can be understood as a multisinusoidal signal sampled across several disjoint regions offset by $\tilde{N}$. We refer to the two subband scenario as *non-contiguous dual-band*, as shown in Fig. 6.2a. As the two subbands are not coherent in a practical implementation, we implement the algorithm developed in [75] to efficiently estimate the ICP and compensate each subband accordingly. Additional details on the mutual coherency among subbands can be found in [78, 73, 74, 75]. The signal in the $k$-domain can be represented in the $R$-domain, or wavenumber spectral domain, by taking the Fourier transform. Limiting the sampling in each subband to $N_k$ corresponds to a convolution in the $R$-domain with a discrete sinc or Dirichlet kernel of width $1/N_k$, resulting in smearing of the spectral information and causing closely spaced peaks to bend together
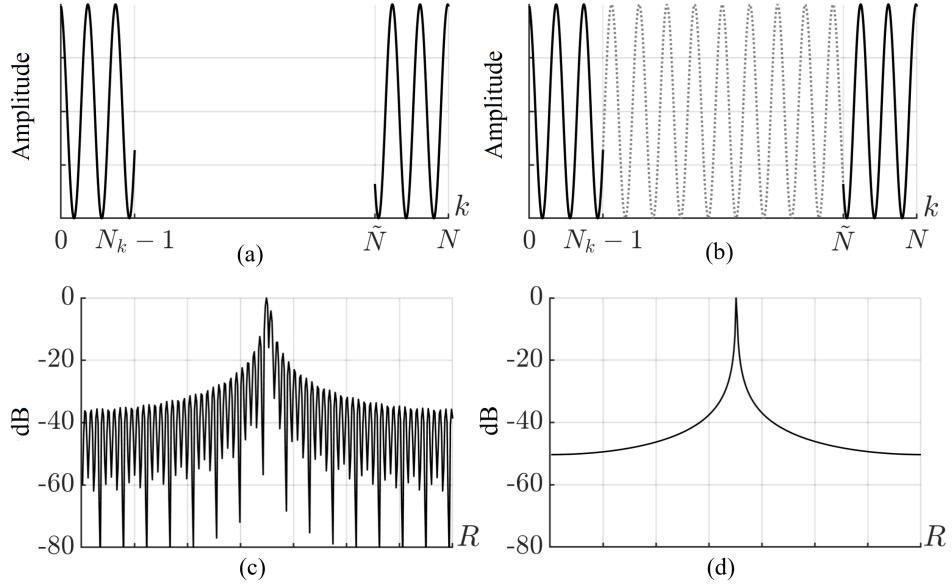
Figure 6.2: Multiband scenario with two subbands in the $k$-domain and $R$-domain. (a) $k$-domain non-contiguous dual-band signal, (b) $k$-domain ideal full-band signal, (c) $R$-domain spectrum of the non-contiguous dual-band signal with zero-padding, (d) $R$-domain spectrum of the ideal full-band signal.

[167]. Given the structure of the multiband signal, the $R$-domain spectrum is the sum of the spectra for each subband if the frequency gap is ignored. Because of the sinc-effect and phase shift in the $R$-domain corresponding to the $\tilde{N}$ sample shift in the $k$-domain for each of the $N_t$ reflectors, the non-contiguous dual-band signal in the $R$-domain suffers from artifacts/sidelobes as the frequency gap between the subbands is neglected [75], as shown in Fig. 6.2c. This analysis is identical to the MFT [79], which results in images degraded by increased sidelobes in the range direction. Comparatively, the $R$-domain spectrum of the ideal *full-band* signal (Fig. 6.2d) does not contain spurious peaks that would distort the images recovered from the fused signal, thereby achieving an improved resolution compared with each subband and the MFT approach.

### 6.1.1  Existing Methods for Multiband Signal Fusion

The objective of multiband signal fusion is to impute the bandwidth between the subbands (interpolating between the subbands or extrapolating the missing samples) in the $k$-domain to acquire the ideal full-band signal shown in Fig. 6.2b, where the dotted portion represents the signal in the frequency gap. Methods for recovering the missing wavenumber domain data from $N_k$ to $\tilde{N} - 1$ apply MUSIC [73] or MPA [74, 75] to estimate the signal poles in an all-pole model. However, these approaches assume that the estimated model order of $s_1(\cdot)$ and $s_2(\cdot)$, $\hat{N}_t \approx N_t$, is small compared to $N_k$. From the analysis in [80], $\hat{N}_t$, the estimated number of targets in the scene, must be chosen such that $\hat{N}_t < \text{round}(N_k/3)$ for the MPA [75]. After the $\hat{N}_t$ signal poles and coefficients are computed, the missing samples can be estimated. However, high-resolution near-field SAR often requires imaging of intricate, continuous objects modeled by thousands or millions of point scatterers, or $N_t \gg N_k$ [46, 134]. As a result, traditional approaches such as the MPA assume simplistic targets, thereby neglecting high-frequency features of the target, and are unable to faithfully recover the multiband signals.

For clarity in the remainder of this chapter, we propose new terminology to describe the spectral composition of radar target scenes based on the portion of the baseband bandwidth occupied by the reflected signal. A target consisting of fewer reflectors than the number of frequency samples ($N_t < N_k$) over the specified frequency range has a *low-bandwidth* relative to the bandwidth of the system. Low-bandwidth targets have a low model order and a low-rank sample covariance matrix, allowing conventional algorithms to adequately approximate signal poles and coefficients. In contrast, a target consisting of a large number of reflectors relative to the frequency sampling ($N_t \gg N_k$) is a *high-bandwidth* target, which is typical in most security and industrial applications of near-field SAR imaging. High-bandwidth targets contain intricate, high-frequency spatial features and have not been addressed for multiband signal fusion in previous studies.
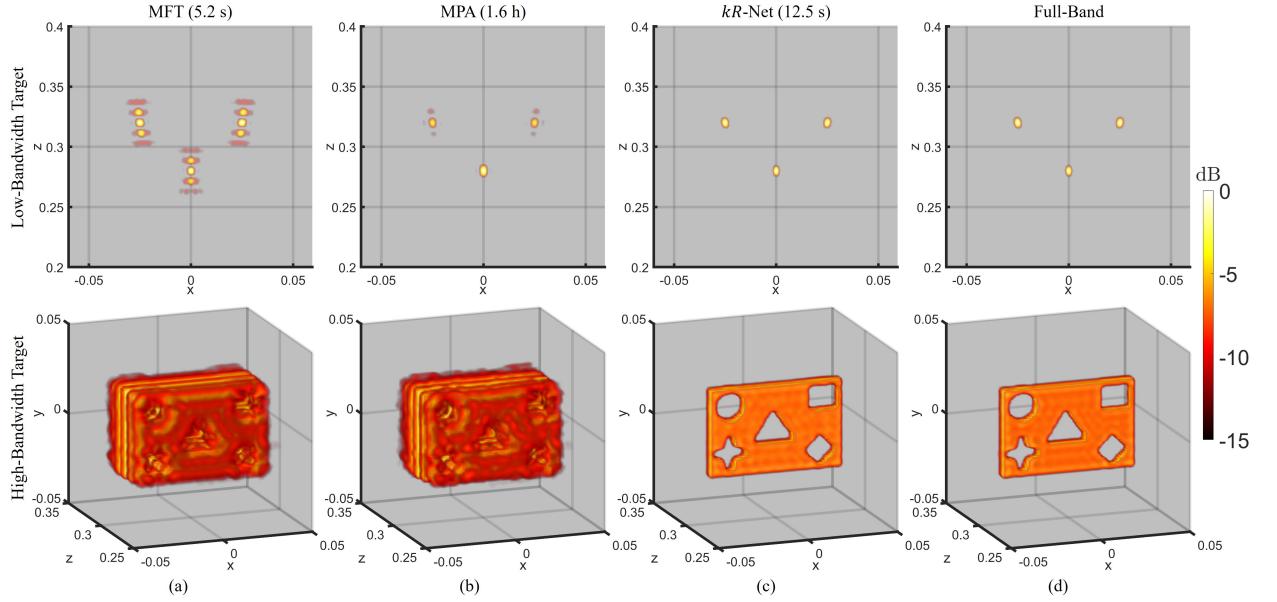
Figure 6.3: Demonstration of the limitations of the matrix Fourier transform (MFT) and matrix-pencil algorithm (MPA) for multiband signal fusion using two targets in simulation: (top) three point targets and (bottom) rectangle with various cutout shapes. (a) The MFT, which requires 5.2 s, suffers from significant sidelobes since it does not compensate for the missing samples in the frequency gap. (b) Although the MPA, which requires 1.6 h for a GPU implementation, achieves adequate reconstruction for the low-bandwidth target (top), its performance degrades for a high-bandwidth target (bottom). (c) The proposed $kR$-Net, which requires 12.5 s, recovers an image that closely resembles the (d) ideal reference image corresponding the full-band scenario as in Figs. 6.2b and 6.2d.

To illustrate this phenomenon, we consider two cases: (1) a low-bandwidth target consisting of 3 point reflectors and (2) a 3-D model of a rectangle with various cutout shapes to constitute a high-bandwidth target, as shown in the first and second rows of Fig. 6.3, respectively. All images are reconstructed using the RMA [39, 23, 94, 168] with the SAR scanning parameters detailed in Section 6.4.1. The MFT [79] is applied to fuse the multiband data, yielding the images in the left column (Fig. 6.3a). As expected, since the MFT does not account for the frequency gap between the subbands, increased sidelobes are observed along the $z$-direction. However, the quality of images recovered using the MPA [75] varies significantly. For the simple low-bandwidth target in the top row ($N_t = 3, N_k = 64$), the MPA recovers each point with minimal undesirable sidelobes compared to the MFT.

116

However, although its performance is better than the MFT, the MPA is plagued by considerable degradation for the high-bandwidth target, where $N_t$ is on the order of thousands and $N_t \gg N_k$, as the high-bandwidth features of the target are not adequately modeled by the MPA. For an equitable comparison, a parallelized GPU implementation of the MPA is employed [74, 75]. Hence, the computation time required for the GPU-implemented MPA is still 1.6 h for this example, deeming it unfit for many applications demanding rapid imaging, such as packing and security screening. By comparison, the MFT and $kR$-Net boast computation times of 5.2 s and 12.5 s, respectively, enabling many common mmWave imaging solutions. In the right column (Fig. 6.3c), the imaging results obtained using the proposed $kR$-Net demonstrate robustness for both targets by achieving focusing performance comparable to the ideal, full-band scenario. Notably, the intricate features of the high-bandwidth target are retained, and the resolution in the $z$-direction is significantly improved compared with the MFT and MPA.

Multiband signal fusion can be posed as a spectral super-resolution/restoration problem in the $R$-domain, the dual to imputation in the $k$-domain. As shown in Figs. 6.2c and 6.2d, $R$-domain super-resolution of the $N_t$ peaks corresponds to imputation of the full bandwidth [169]. Deep learning-based solutions have proven successful in similar spectral-enhancement problems on radar images [65, 66, 111, 126, 170] and multisinusoidal line spectra [167, 171], achieving resolutions exceeding the theoretical limitations. However, data-driven approaches have not been applied to multiband signals to achieve joint $k$-domain imputation and $R$-domain super-resolution. Because the multiband fusion problem has distinct features in the $k$-domain and $R$-domain, we propose a hybrid approach that operates in both domains.
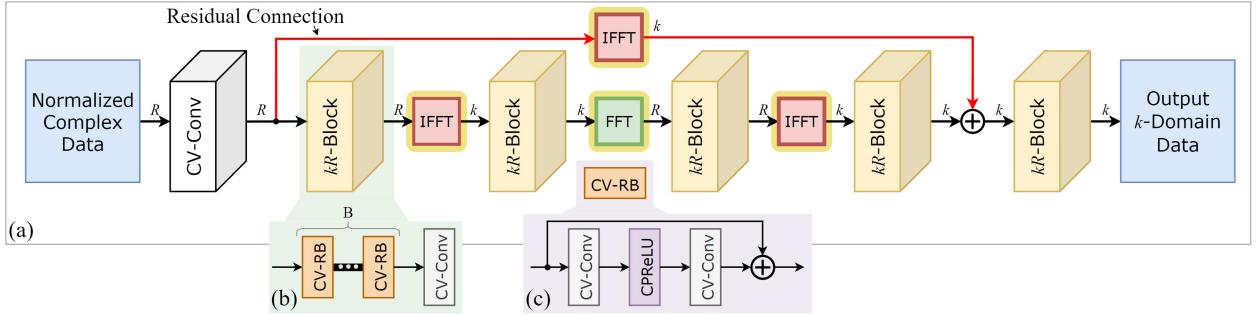
Figure 6.4: Architecture of the (a) hybrid, dual-domain $kR$-Net. Proposed addition of the domain transformation blocks (highlighted: FFT and IFFT) allows the network to learn important features of the signal in both the $k$- and $R$-domains improving on conventional CNN approaches. The novel, hybrid approach using FFT and IFFT blocks throughout the network, achieves superior multiband signal fusion and spectral super-resolution compared with a model without the Fourier operations. Architecture of the (b) $kR$-Block and (c) CV-RB. The $kR$-block consists of $B$ CV-RBs in cascade followed by a CV-Conv layer. Each CV-RB is a residual convolution block with a single CPReLU activation.

## 6.2 Proposed Architecture of $kR$-Net for Improved 3-D Multiband SAR Imaging

In this section, we introduce a novel dual-domain CV-CNN architecture, referred to as $kR$-Net, to perform efficient multiband fusion for improved 3-D near-field SAR imaging. The proposed framework alternates between operating in the $k$-domain and $R$-domain, allowing the network to learn the unique characteristics inherent to each domain. Compared with learning in only one domain, $kR$-Net demonstrates superior convergence and quantitative performance, as discussed in Section 6.4. Additionally, the proposed algorithm is robust for low- and high-bandwidth imaging scenarios, which are common in many realistic applications.

### 6.2.1 Framework of $kR$-Net

The architecture of $kR$-Net is shown in Fig. 6.4a, where the signal domain is denoted at each connection as $k$ or $R$ for the wavenumber domain or wavenumber spectral domain, respectively, and the forward and inverse fast Fourier transform are denoted by FFT and

IFFT, respectively. The input is given as the $R$-domain spectrum of the wavenumber domain samples and is processed in both domains by residual $kR$-blocks. After the residual connection, an additional $kR$-block and complex-valued convolution (CV-Conv) layer processes the signal before it is output in the $k$-domain. The specific designs of each module are detailed as follows.

**Input Layer and Residual Connection**

Rather than layering the real and complex parts of the signal [7, 167] or employing a two-path network [55], the inputs to $kR$-Net are complex-valued signals of length $N$. The complex-valued input signals are normalized in the $R$-domain by the magnitude min-max norm before being passed to the first CV-Conv layer as

$$\mathbf{z} = \frac{\mathbf{x} - ||\mathbf{x}||_{\min}}{||\mathbf{x}||_{\max} - ||\mathbf{x}||_{\min}}, \tag{6.5}$$

where $||\mathbf{x}||_{\min}$ and $||\mathbf{x}||_{\max}$ are the minimum and maximum values of the magnitude of $\mathbf{x}$. Hence, the phase of $\mathbf{x}$ remains unmodified, while the magnitude is scaled to be consistently between 0 and 1.

Since the network expects an input in the $R$-domain, the FFT of the $k$-domain data is computed prior to the input to the network with zero-padding between the subbands. Complex-valued convolution extends the convolution operation employed by CNNs to complex input data. To implement a CV-Conv layer, the convolution kernel matrix must be complex-valued. However, because the convolution between two complex-valued tensors is generally unsupported by deep learning software, we decompose the input signal $\mathbf{x} = \mathbf{x}_R + j\mathbf{x}_I$ into real and imaginary parts. Similarly, by expressing the kernel as $\mathbf{M} = \mathbf{M}_R + j\mathbf{M}_I$, the complex-valued convolution can be written, neglecting the bias terms, as

$$\mathbf{x} \circledast \mathbf{M} = \mathbf{x}_R \circledast \mathbf{M}_R - \mathbf{x}_I \circledast \mathbf{M}_I + j(\mathbf{x}_R \circledast \mathbf{M}_I + \mathbf{x}_I \circledast \mathbf{M}_R), \tag{6.6}$$

where ⊛ denotes the real-valued convolution operation. By decomposing the convolution in this manner, complex-valued convolution can be computed using existing techniques operating on the real and imaginary parts of the input signal and kernel. The real and imaginary parts of the weight matrix $\mathbf{M}$ can be implemented as real-valued matrices according to (6.6), and their values are determined by complex-valued backpropagation following the convention for CV-CNNs [63, 66].

The CV-Conv layer is a general-purpose complex-valued convolution layer defined with a kernel size $K$ and zero-padding such that the signal length of $N$ is preserved at the output, $C_{\text{in}}$ input channels, and $C_{\text{out}}$ output channels. The first layer of $kR$-Net is a CV-Conv layer with 1 input channel and $F$ output channels, where $F$ is the number of feature channels and is constant throughout the network. After the first CV-Conv layer, the intermediate representations are fed through an IFFT block in the residual pass-forward connection, as shown in red in Fig. 6.4a. The residual connection preserves the information at the known subbands, and the network demonstrates superior empirical performance with the proposed configuration than without the pass-forward connection.

## $kR$-Blocks and Domain Transformation Blocks

The $kR$-Block is composed of a cascade of complex-valued residual blocks (CV-RBs) followed by a single CV-Conv layer, as shown in Fig. 6.4b. Each $kR$-Block operates on the signal in either the $k$-domain or $R$-domain, as the signal alternates between the two domains throughout the $kR$-Net. Furthermore, because the domain transformation blocks (FFT and IFFT) are fully differentiable, they can be treated as conventional layers in the network, and gradient backpropagation can be easily implemented [66]. The Fourier operations are performed across each activation map and normalized to make the FFT and IFFT orthonormal. Based on the convolution properties of the Fourier transform, convolution in one domain can be viewed as multiplication in the other domain. In this sense, applying a CV-Conv layer in the

$k$-domain can be considered a fully connected layer in the $R$-domain. However, as illustrated later in Section 6.4.2, our hybrid, dual-domain approach outperforms a network operating exclusively in the $k$-domain or $R$-domain in terms of convergence and numerical performance. Hence, compared with conventional CNN models, the addition of domain transformation blocks throughout the network is key to improving multiband fusion performance.

The CV-RB architecture is shown in Fig. 6.4c. Inspired by [18], the residual block consists of two convolution layers separated by an activation function: the complex parametric rectified linear unit (CPReLU) [66]. Compared with the original ResNet residual block [172], the proposed residual block removes batch normalization and empirically outperforms a bottleneck residual architecture [97] for multiband signal fusion.

The CPReLU activation function is selected over alternatives, such as the complex ReLU (CReLU) [54], which computes the sum of the ReLU operation on the real and imaginary values separately, as the CPReLU has an improved activation over the complex domain. Using the notation employed in (6.6), the CPReLU can be expressed as

$$\begin{aligned} \text{CPReLU}(\mathbf{x}) = \max(0, \mathbf{x}_R) + \eta_R \min(0, \mathbf{x}_R) \\ + j \left( \max(0, \mathbf{x}_I) + \eta_I \min(0, \mathbf{x}_I) \right), \end{aligned} \tag{6.7}$$

where the parameters $\eta_R$ and $\eta_I$ are learned during the training stage of the network for each CPReLU instance [66]. Because $\eta_R$ and $\eta_I$ are learned independently, different layers of the network may learn different representations of the signal in amplitude and phase, aiding network robustness. The CPReLU can be understood as a complex domain parametric rectified linear unit (PReLU), which extends the traditional ReLU into the negative input domain to overcome gradient saturation for negative activation values. In the CPReLU, the real and imaginary parts of $\mathbf{x}$ are independently processed by a PReLU and the output is complex-valued. In the complex domain, this corresponds to retaining information in all four quadrants, corresponding to all combinations of positive and negative real and imaginary activation values, as detailed in [66].

121

After the first four $kR$-Blocks, the residual connection is made in the $k$-domain followed by another $kR$-Block and CV-Conv layer before being output, as shown in Fig. 6.4a. The number of $kR$-Blocks and domain transformations was investigated empirically, and the proposed configuration yielded the optimal numerical performance. However, further investigation of alternate architectures and deeper neural networks is a promising route for future research. In the spirit of [97], multiple values for the convolution kernel size $K$ were investigated, and the optimal value was determined empirically to be $K = 5$. The number of feature maps throughout the network is chosen as $F = 32$, and the number of CV-RBs for each $kR$-block is set as $B = 8$. $kR$-Net comprises 86 CV-Conv layers and 866324 learnable parameters. Multiband signal fusion is performed by $kR$-Net on a signal of length $N$, yielding a fused signal in the $k$-domain of equivalent size. As discussed in Section 6.4.2, a network is trained for a specific multiband signal fusion scenario and must be retrained for application to alternate subband configurations (placement and range of subbands, sampling conditions, length $N$ of full-band signal, etc.).

### 6.2.2 Training Details

The weights of the network are calibrated using an Adam optimizer with a learning rate of $1 \times 10^{-4}$, $\beta_1 = 0.9$, and $\beta_2 = 0.999$. Training is performed on a single RTX3090 GPU with 24 GB of memory with a batch size of 1024 and L1 loss criterion. The complex-valued loss term is defined as

$$\mathcal{L} = \sum_{\ell=0}^{N-1} \left[ |\hat{s}(\ell)_R - s(\ell)_R| + |\hat{s}(\ell)_I - s(\ell)_I| \right], \tag{6.8}$$

where $\hat{s}(\ell)$ are the predicted signals output from the network corresponding to the full-band ground-truth vectors $s(\ell)$ and the subscripts denote the real and imaginary parts of the signals. The complex components are processed separately, similar to the approaches in [66, 63], by traditional L1 distance metrics, and the real-valued result is the sum of the two

L1 losses from the real and imaginary parts of the predicted signal with the ground-truth signal.

More advanced loss functions were investigated, such as the L1/L2 difference between the ground-truth signal and intermediate representations throughout the network or loss between sample covariance matrices. These loss functions were tested in conjunction with alternate configurations, such as varying the number of $kR$-Blocks or removing the residual connection. However, the architecture detailed in Fig. 6.4a demonstrated superior numerical performance in both training and testing with real multiband SAR data. Nevertheless, future investigations into neural network design based on statistical signal processing principles will likely facilitate additional insights and promising results for signal processing problems. Incorporating a hybrid, data-driven signal processing approach is a promising direction for similar future efforts.

**Training and Testing Datasets**

Since there is no publicly available dataset for near-field multiband SAR imaging, we generate training and testing datasets by simulating the response to a multiband LFM radar. The ideal noiseless full-band signals spanning both subbands and the frequency gap in the $k$-domain are generated as

$$s(\ell) = \sum_{i=0}^{N_t-1} \alpha_i e^{-j2(k_1+\Delta_k \ell)R_i}, \ell = 0, \ldots, N-1, \tag{6.9}$$

where $\alpha_i$ values are selected from a complex normal distribution and $R_i$ values are chosen from a uniform distribution spanning the unambiguous range of the radar. After computing the full-band signals $s(\cdot)$ the multiband signals $\hat{s}(\cdot)$ are obtained by nullifying the samples in the frequency gap as

$$\hat{s}(\ell) = \begin{cases} s(\ell) & \ell \in [0, N_k - 1] \cup [\tilde{N}, N-1] \\ 0 & \ell \in [N_k, \tilde{N} - 1] \end{cases}. \tag{6.10}$$

The multiband signals are then corrupted with complex additive Gaussian white noise (AWGN) in each subband. The noisy multiband signals are used as the input to $kR$-Net after taking the FFT and employing the normalization process detailed earlier. Each noisy multiband signal is treated as a feature vector with a corresponding label vector consisting of the ideal full-band signal in (6.9).

To train the network, 1048576 samples are independently generated with $N_t$ target reflectors, where $N_t$ is randomly selected between 1 and 200. The SNR for each sample is selected on a continuous uniform distribution from -10 to 30 dB. A validation set of 2048 samples is generated using the same procedure. Assuming a realistic scenario with two radars with starting frequencies $f_1 = 60$ GHz and $f_2 = 77$ GHz, where each radar has a bandwidth of $B = 4$ GHz, we set $N_k = 64$ and $\Delta_f = 62.5$ MHz. Hence, $\tilde{N} = 272$ and $N = 336$, and the low-rank assumption of the MPA, $\hat{N}_t < \text{round}(N_k/3)$, will often be invalid if the target is high-bandwidth and consists of many reflectors. Although this study employs 60 GHz and 77 GHz radars, the proposed algorithm can easily be extended to other multiband configurations. For practical implementation, federal communications commission (FCC) licensing limits certain combinations of subbands based on application, but the algorithm and concepts derived in this study are applicable across various subband configurations. By training on this dataset, the proposed algorithm learns to perform multiband fusion for high-bandwidth targets.

It is important to note that the proposed signal model and dataset generation scheme impose limitations on the model and its generalizability. In a practical implementation, imperfections that are not modeled in (6.9), such as frequency dependence of the scattering parameters, $\alpha_i$, device non-linearity, clutter, non-Gaussian noises, and different antenna distortions among subbands, may impact the reconstruction quality. Hence, networks trained on these data may display different robustness depending on the imaging scenario, placing additional importance on the system design and calibration. Although a model trained for

a specific multiband scenario will not generalize well to other subband configurations, the proposed model architecture demonstrates the ability to generalize to various multiband imaging schemes when trained with appropriate datasets, as detailed in Section 6.4.2.

An alternative dataset was considered consisting of extended, solid targets, such as those shown in Figs. 6.3 and 6.9. However, because there is no sufficiently diverse dataset of such multiband SAR data, the network did not generalize well across different shapes. Additionally, a network was first trained on a dataset of randomly placed point reflectors, as in (6.9), and then fine-tuned on data from solid targets, but a performance improvement was not observed. Further development of diverse mmWave datasets will be essential to the advancement of joint signal processing and data-driven algorithms and is a promising future direction.
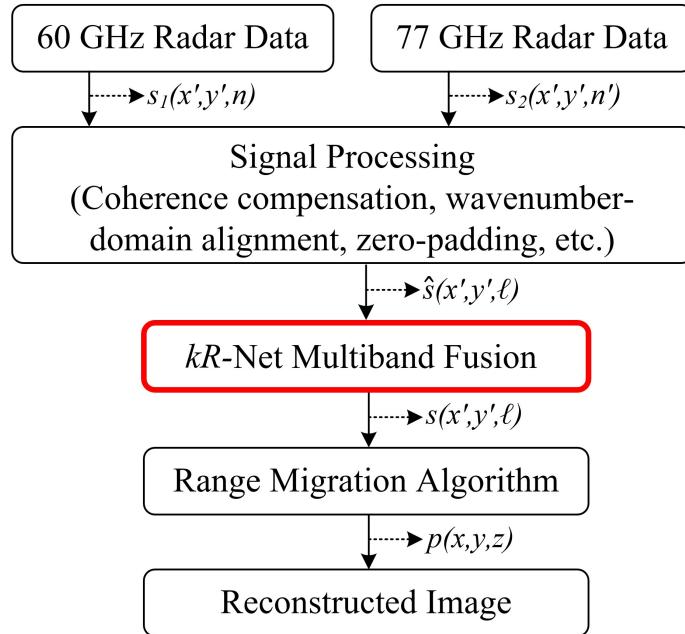


Figure 6.5: Multiband signal fusion pipeline for high-resolution 3-D near-field imaging. The data from both subbands are fused using the proposed $kR$-Net producing a high-fidelity reconstruction for both low- and high-bandwidth targets.

### 6.2.3 Imaging Implementation

After multiband fusion is performed using $kR$-Net, the SAR image is reconstructed from the fused data. A summary of the imaging pipeline is presented in Fig. 6.5. After the data are collected from both radars, preprocessing steps are necessary to ensure signal coherence and align the data in the $k$-domain. We implement the ICP compensation algorithm detailed in [75] and set consistent sampling parameters across radars. The proposed algorithm is valid for both collocated and noncollocated antennas if the spatial wavenumber domains of all radars are sufficiently coincident, implying that both radars have similar illumination of the target. Applying $kR$-Net to the multiband signal is advantageous compared with classical signal processing algorithms as $kR$-Net is highly parallelizable and can efficiently perform signal fusion for many samples. The RMA is applied after the signal fusion step to produce a high-resolution 3-D image [39, 94, 115]. Compared to conventional signal processing-based algorithms for multiband fusion, the proposed $kR$-Net yields superior imaging performance and demonstrates robustness for the realistic case of high-bandwidth, intricate targets.

### 6.3 Multiband Imaging System

This section provides an overview of the implementation of the multiband imaging prototype using commercially available mmWave radars. Whereas prior research on near-field multiband radar imaging has employed sophisticated laboratory equipment [75, 78], which is not suitable for many practical applications, we introduce a highly-integrated system that employs commercially available equipment for multiband near-field SAR. The proposed testbed uses two mmWave radars operating at distinct subbands and introduces a synchronization strategy to achieve efficient data collection.

An overview of the system architecture is presented in Fig. 6.6. The system consists of a 60 GHz radar, 77 GHz radar, two-axis mechanical scanner, motion controller, synchronization module, and a host PC. The Texas Instruments (TI) IWR6843ISK and IWR1642BOOST
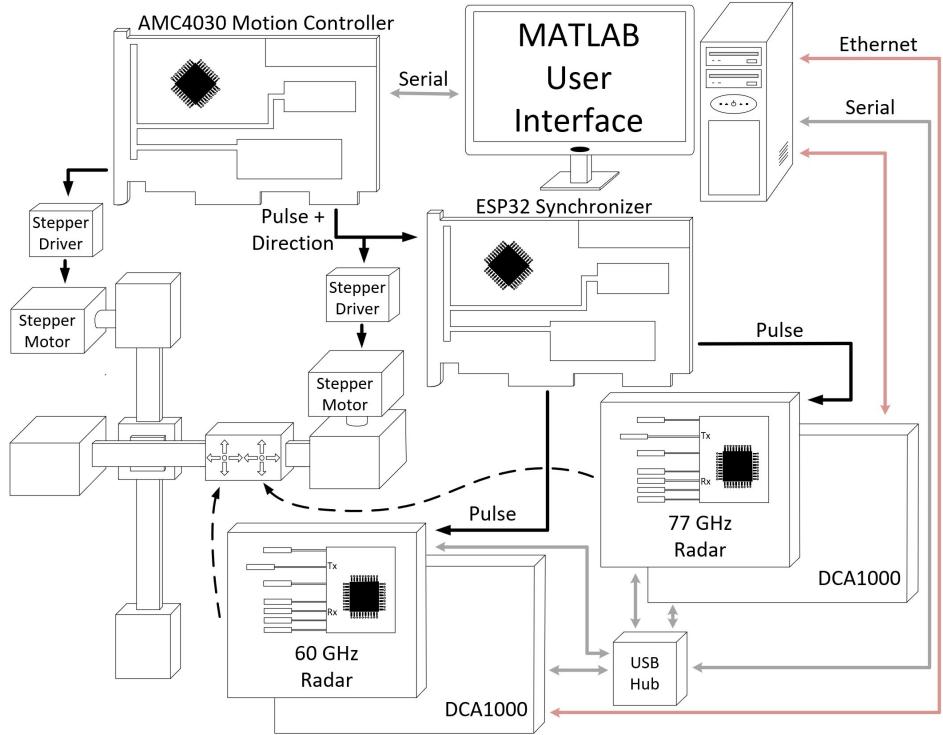
Figure 6.6: System architecture of the proposed multiband imaging testbed prototype.

are chosen as the single chip 60 GHz and 77 GHz radars, respectively. Both radars have an operating bandwidth of $B = 4$ GHz; hence, the subbands span 60–64 GHz and 77–81 GHz. In addition, the LFM radars are configured using the parameters detailed in Section 6.2.2. The data are captured in real-time by the TI DCA1000 evaluation module for each radar and streamed to the host PC over Ethernet. Both radars are mounted onto a belt-driven two-axis mechanical scanner, as shown in Fig. 6.6, such that the lowest Rx antennas on each radar are aligned, and the radars are separated horizontally by a distance of $\Delta_x$.

The two-axis mechanical scanner is driven by stepper motors that receive pulses from a motion controller, and the entire system is controlled using a custom MATLAB user interface running on the host PC. The radars are scanned in the $x$- and $y$-directions at the spatial sampling Nyquist rate of $\lambda/4$ in both the horizontal and vertical directions [51, 23]. Extending the synchronization approach in [39], we design a novel multi-radar synchronizer for the precise positioning of both radars while operating at high scanning speeds. The pro-

posed prototype achieves a speed of 500 mm/s, allowing for short scanning times required by applications such as security screening and packaging. However, scanning multiple physical radar modules at these speeds while achieving precise synthetic element positioning required to mitigate distortion is challenging and has not been addressed previously in the literature. The synchronizer monitors the stepper driver pulses that determine the position of the platform, to which the radars are mounted, as it both accelerates and decelerates. Additionally, the synchronizer tracks the positions of each radar and fires them independently to account for the non-uniform timing required to maintain a uniform synthetic aperture and ensure equivalent illumination of the target scene from both radars. Additional hardware-specific details of mmWave imaging testbeds can be found in [39].

## 6.4 Experimental Results

In this section, the superiority of the proposed $kR$-Net is demonstrated using numerical simulations and empirical experiments. The matrix Fourier transform (MFT) algorithm [79] and matrix-pencil algorithm (MPA) [74, 75] are adopted as comparison baselines for the following experiments. After $kR$-Net is trained using the procedure detailed in Section 6.2.2, we conduct experiments on both synthetic and empirical multiband data to validate the performance of $kR$-Net compared with traditional signal processing approaches. We consider a dual-band system with the radar signaling parameters discussed previously.

### 6.4.1 Visual Comparison of Simulation Results

First, we detail various simulation results obtained using the proposed $kR$-Net algorithm for multiband signal fusion.
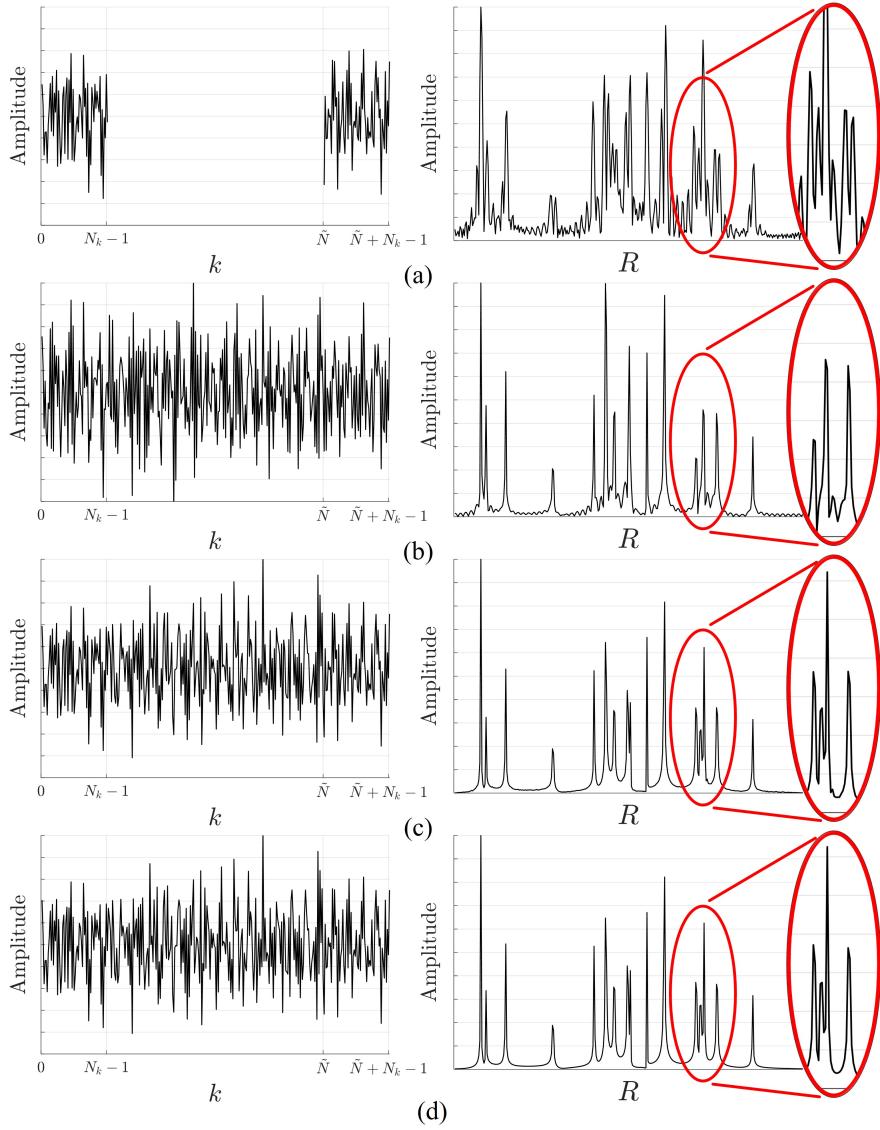
Figure 6.7: Closely spaced peaks resolved using $kR$-Net. Comparison of multiband signal fusion performance using (a) MFT, (b) MPA, and (c) $kR$-Net compared to (d) the ideal full-band scenario for a single simulated signal consisting of randomly placed point scatters. Left: The real part of the $k$-domain signals. Right: The magnitude of the $R$-domain spectra, demonstrating the super-resolution capability of the proposed $kR$-Net.

## Single Multiband Signal

First, we consider the simple case of a single multiband signal captured at the two subbands used throughout the experiments with a target consisting of randomly placed points. The $k$-

domain signals and $R$-domain spectra for the MFT, MPA, and $kR$-Net are shown in Fig. 6.7 and compared with the corresponding ideal full-band signal. Each subband is corrupted with AWGN at an SNR of 20 dB. The MPA imputes the lost signal between the two subbands to recover a signal of length $N$; however, the resulting wideband signal deviates from the ideal signal owing to the assumptions in the MPA. Although it outperforms the MFT, the MPA is unable to recover every peak in the $R$-domain, and $kR$-Net yields the most accurate reconstruction of the full-band signal.
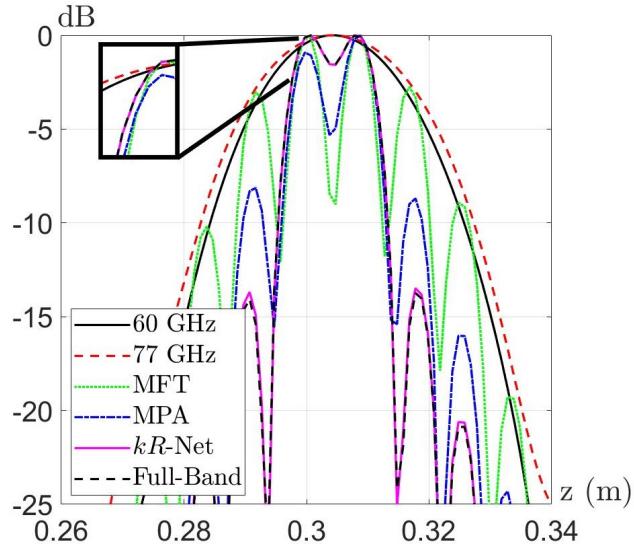


Figure 6.8: Comparison of imaging results from various scenarios sliced along $y = 0$ m with two simulated point scatterers spaced by $\Delta_z = 7.1$ mm, which corresponds to the minimum resolvable distance for an equivalent bandwidth of 21 GHz.

**Effective Bandwidth Study**

Given that the purpose of multiband imaging is to achieve a finer resolution by synthetically increasing the bandwidth, we consider the resolution capability of the proposed algorithm and its corresponding effective bandwidth. The resolution of a radar system in the downrange direction is given by $\delta_z = c/2B$, which determines the ability of the system to resolve two closely spaced reflectors. Hence, we compare the imaging results for various cases with two

closely spaced peaks to evaluate the resolution limit of the algorithms. We simulate a scenario with two point scatterers located at $Z_r$ and $Z_r + \Delta_z$ from the radar boresight with AWGN at an SNR of 20 dB. Using $Z_r = 300$ mm, we evaluate the performance for $\Delta_z = 7.1$ mm, which corresponds to an effective bandwidth of 21 GHz. As shown in Fig. 6.8, the proposed $kR$-Net achieves a nearly identical response to the ideal full-band signal. Since the 60 GHz and 77 GHz radars, subbands 1 and 2, each have a bandwidth of 4 GHz, the two closely spaced reflectors are blurred into a single peak. The MFT resolves the two peaks but has severely increased sidelobes compared to the full-band signal. Because the target consists of two signal components, the MPA resolves both peaks without significant distortion, but demonstrates some minor deviations from the ideal signal. In contrast, $kR$-Net achieves a more accurate signal with lower sidelobes. Thus, the proposed algorithm achieves an effective bandwidth of 21 GHz because the two peaks are clearly resolved.

**High-Bandwidth Target**

To visually evaluate the imaging performance of $kR$-Net and demonstrate the deficiencies of the MFT and MPA on a high-bandwidth target, two shapes are considered under the multiband scenario in the near-field with a planar array with dimensions of 0.125 m $\times$ 0.125 m, satisfying the spatial Nyquist criterion [51]. A star shape is placed on the left at the plane $z = 307$ mm and a diamond to the right at the plane $z = 300$ mm, with the synthetic array at the plane $z = 0$ m. For many high-resolution imaging tasks, objects must be localized, classified, or counted to identify concealed weapons, ensure correct packaging, or detect defects. To localize closely spaced targets, a transceiver with a bandwidth of 4 GHz may not be adequate as the low range resolution yields images stretched in the $z$-direction beyond the physical dimensions of the objects. Fig. 6.9 shows the 3-D reconstructed images and slices along $z = 300$ mm and $z = 307$ mm to demonstrate the distortion along the $z$-direction that can contaminate the images and degrade system performance. As the objects
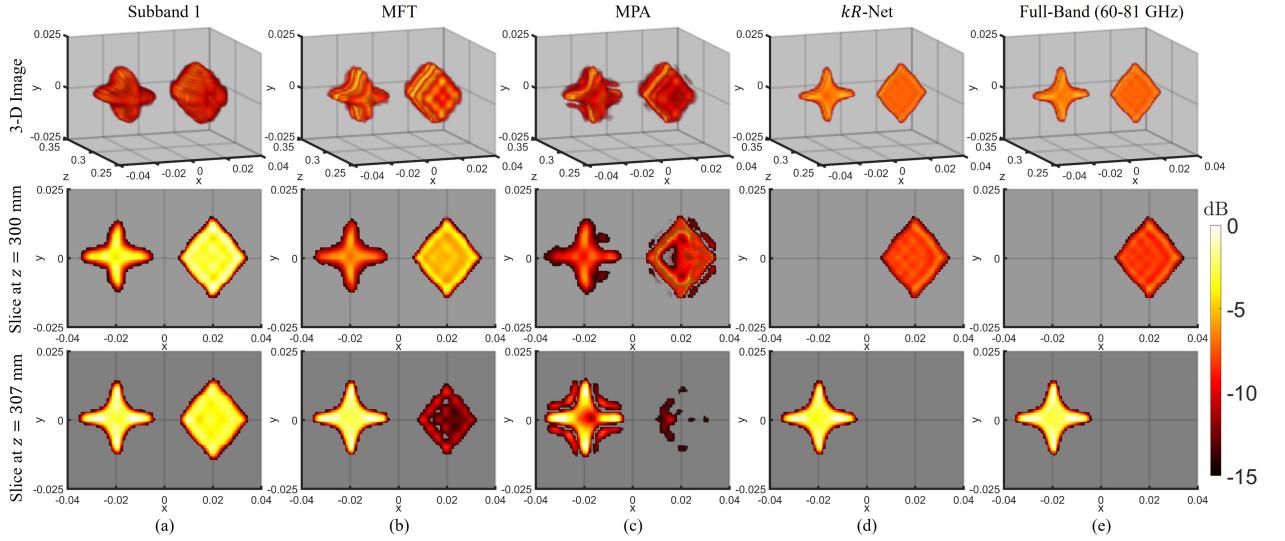
Figure 6.9: High-bandwidth target consisting of a diamond located at $z = 300$ mm and star located at $z = 307$ mm. The proposed $kR$-Net is able to separate the two shapes to their respective $z$-planes, exhibiting comparable imaging performance to an ideal full-band transceiver without feature loss or visible sidelobes present in the MFT and MPA images. Imaging results for simulated high-bandwidth target using (a) subband 1 (60–64 GHz), (b) MFT, (c) MPA, (d) $kR$-Net, (e) ideal full-band (60–81 GHz). First row: 3-D image. Second row: slice at $z = 300$ mm, corresponding to the location of the diamond shape. Third row: slice at $z = 307$ mm, corresponding to the location of the star shape.

are located in distinct $z$-planes, the second and third rows should show only the diamond and star, respectively, as shown in Fig. 6.9e, for the ideal full-band case. The image recovered from subband 1 (60–64 GHz) is shown in Fig. 6.9a and demonstrates the limitations of low bandwidth, as both shapes are clearly visible in both slices. Similarly, the MFT results in powerful sidelobes, as shown in Fig. 6.8, which correspond to ghost shapes along the $z$-direction for solid targets, as shown in Fig. 6.9b. Hence, images recovered using the MFT are not suitable for localization or object counting tasks because they are obscured by spurious sidelobes. For this high-bandwidth target scenario, the image recovered using the MPA, shown in Fig. 6.9c, suffers from a loss of fidelity due to the simplistic multiband fusion model and the assumption of a small number of reflectors. As a result, the MPA images are not only contaminated with sidelobes that degrade the slices corresponding to each object but also fail to retain the high-bandwidth features of the objects. Comparatively, the images

recovered using $kR$-Net, shown in Fig. 6.9d, closely resemble the ideal full-band images. This aligns with the conclusions drawn from the numerical experiments and demonstrates the superior performance of $kR$-Net compared to existing algorithms for realistic high-resolution imaging tasks.

### 6.4.2 Quantitative Investigations

First, we compare the imaging performance of the MFT, MPA, and $kR$-Net numerically, considering a near-field SAR scenario with a planar aperture of $200 \times 200$ synthetic elements. It is important to note that the MFT and MPA are classical signal processing approaches, whereas $kR$-Net employs a hybrid approach that combines data-driven techniques with signal processing algorithms. Fifty Monte Carlo trials are conducted for each experiment. We compare the various algorithms using the structural similarity index measure (SSIM), peak signal-to-noise ratio (PSNR), and normalized root mean square error (NRMSE), where SSIM is defined as

$$\text{SSIM} = \frac{(2\mu_{\mathbf{x}}\mu_{\mathbf{y}} + C_1)(2\sigma_{\mathbf{xy}} + C_2)}{(\mu_{\mathbf{x}}^2 + \mu_{\mathbf{y}}^2 + C_1)(\sigma_{\mathbf{x}}^2 + \sigma_{\mathbf{y}}^2 + C_2)}, \tag{6.11}$$

where $\mathbf{x}$ and $\mathbf{y}$ are the reconstructed image and reference image, respectively; $\mu_{\mathbf{x}}$, $\sigma_{\mathbf{x}}$ and $\mu_{\mathbf{y}}$, $\sigma_{\mathbf{y}}$ are their corresponding mean values and standard deviations, respectively; and $L$ is the dynamic range of the pixel values. For stability $C_1 = (k_1 L)^2$ and $C_2 = (k_2 L)^2$ where $k_1 = 0.01$ and $k_2 = 0.03$ by default [55]. SSIM quantifies the similarity between $\mathbf{x}$ and $\mathbf{y}$, with a larger value indicating better performance and a maximum value of 1 for a perfect reconstruction. Similarly, higher PSNR values (dB) [18] and lower NRMSE values [55] indicate more accurate image reconstruction.

Images are computed using the MFT, MPA, and $kR$-Net for multiband signal fusion of the two subbands, 60–64 GHz and 77-81 GHz, and compared against the image recovered from the ideal full-band scenario spanning the entire bandwidth of 60-81 GHz as in (6.9). First, 3-D images are generated with an SNR of 20 dB for the full-band and multiband

Table 6.1: Comparison of SSIM, PSNR, and NRMSE across different number of targets ($N_t$) using MFT, MPA, and $kR$-Net

| $N_t$ | MFT | | | MPA | | | $kR$-Net | | |
|---|---|---|---|---|---|---|---|---|---|
| | SSIM | PSNR | NRMSE | SSIM | PSNR | NRMSE | SSIM | PSNR | NRMSE |
| 3 | 0.9970 | 48.91 | 1.145 | 0.9997 | 66.92 | 0.1727 | **0.9999** | **90.90** | **0.01779** |
| 10 | 0.9934 | 45.81 | 1.142 | 0.9959 | 48.03 | 0.7904 | **0.9997** | **59.36** | **0.2439** |
| 100 | 0.9703 | 39.56 | 1.108 | 0.9684 | 39.17 | 1.160 | **0.9809** | **41.60** | **0.8779** |
| 400 | 0.9220 | 34.07 | 1.097 | 0.9240 | 33.56 | 1.164 | **0.9425** | **35.44** | **0.9378** |
| 700 | 0.8989 | 32.28 | 1.089 | 0.9061 | 31.94 | 1.130 | **0.9250** | **33.72** | **0.9208** |
| 1000 | 0.8874 | 31.39 | 1.068 | 0.8902 | 30.82 | 1.143 | **0.9094** | **32.45** | **0.9488** |
| 1300 | 0.8678 | 30.03 | 1.055 | 0.8714 | 29.50 | 1.122 | **0.8922** | **31.14** | **0.9302** |
| Avg. | 0.9338 | 37.44 | 1.100 | 0.9364 | 40.14 | 0.9644 | **0.9499** | **46.37** | **0.6967** |

cases, as discussed in Section 6.2.2. The robustness of the various algorithms is compared over a varying number of targets $N_t$, which corresponds to the target bandwidth. The results are presented in Table 6.1. The MFT suffers from increased sidelobes because it does not account for the frequency gap. As expected, the MPA exhibits tremendous performance for low-bandwidth targets but is plagued by imaging degradation as the target bandwidth and $N_t$ increase. In contrast, the proposed $kR$-Net demonstrates robustness compared with classical algorithms, particularly when applied to high-bandwidth targets scenarios. This is expected as $kR$-Net is trained on a dataset containing both low- and high-bandwidth targets. The best evaluations are shown in boldface in Table 6.1, indicating the superiority of $kR$-Net over the MFT and MPA for realistic 3-D imaging scenarios.

Table 6.2: Comparison of SSIM, PSNR, and NRMSE across values of SNR using MFT, MPA, and $kR$-Net

| SNR | MFT | | | MPA | | | $kR$-Net | | |
|---|---|---|---|---|---|---|---|---|---|
| | SSIM | PSNR | NRMSE | SSIM | PSNR | NRMSE | SSIM | PSNR | NRMSE |
| 20 dB | 0.95414 | 33.64 | 1.089 | 0.9743 | 39.72 | 0.5510 | **0.9839** | **42.39** | **0.4064** |
| 15 dB | 0.9487 | 33.08 | 1.153 | 0.9733 | 38.72 | 0.6138 | **0.9801** | **41.06** | **0.4680** |
| 10 dB | 0.9465 | 32.324 | 1.150 | 0.9560 | 37.41 | 0.6580 | **0.9816** | **41.47** | **0.4129** |
| 5 dB | 0.9485 | 34.32 | 1.109 | 0.9573 | 37.25 | 0.8203 | **0.9800** | **41.00** | **0.5375** |
| 0 dB | 0.9457 | 32.48 | 1.131 | 0.9610 | 37.61 | 0.6880 | **0.9780** | **40.76** | **0.5009** |
| Avg. | 0.9487 | 33.17 | 1.1265 | 0.9644 | 38.14 | 0.6662 | **0.9807** | **41.34** | **0.4651** |

Next, images with 200 randomly distributed point scatterers in addition to one solid object, selected from a set of 10 basic shapes including circle, square, triangle, etc., are

generated for performance evaluation. Table 6.2 presents the results for the different algorithms evaluated with SNR values ranging from 0 dB to 20 dB in increments of 5 dB. Because $N_t > 200$ for every image, the targets are considered high-bandwidth, and the MPA is unable to adequately model the target intricacies. In this experiment, the MPA outperforms the MFT because of the addition of the solid target, as the MPA slightly reduces the sidelobes from the solid target, which contribute a larger amount of power than the point scatterers. However, $kR$-Net demonstrates robustness across low and high SNR, yielding high SSIM and PSNR in conjunction with low NRMSE.

These analyses validate the superiority of the proposed algorithm for both low- and high-bandwidth target scenarios. $kR$-Net overcomes deficiencies in the MPA due to a simplistic model that deems the MPA unsuitable for many near-field imaging applications that require high-resolution imaging of intricate objects without feature loss. Furthermore, since $kR$-Net is highly parallelizable, its computation time per image, 12.5 s, is significantly less than that of the iterative MPA [74, 75], which requires 1.6 h for 3-D imaging. As described in Section 6.1.1, the same GPU hardware is used to implement all three algorithms. An efficient, parallelized implementation of the MPA is employed; however, the computation time remains excessive. Alternatively, although the MFT requires only 5.2 s to recover an image [79], it does not account for the frequency gap between subbands and consistently demonstrates inferior performance to that of $kR$-Net in terms of image quality.

**Ablation Study**

To demonstrate the effectiveness of the hybrid, dual-domain approach, we compare two baseline networks with $kR$-Net. First, we remove the FFT and IFFT blocks from the network such that it operates exclusively in the $k$-domain. In contrast to Fig. 6.4a, the signals at the input are in the $k$-domain and remain in the $k$-domain throughout the entire network. For imputation problems, such as image completion or multiband signal fusion, leveraging the

information positioned throughout the input signal, which in our case is the signal at each subband, to complete missing regions is challenging for a CNN. Each pixel of an intermediate representation, that is, the output of a convolution layer, in a CNN depends on a region of the representation in the previous convolution layer. Hence, a pixel in a given representation depends on a certain region of the input signal, which is known as the effective receptive field. However, the effective receptive field typically grows slowly due to small kernel sizes, implying that the inferred signal in the frequency gap will not be aware of the subband signals until later in the network [173]. This is an issue for multiband fusion because samples between subbands require dependence on subband signals for robust estimation. Hence, the performance of wavenumber domain network, called $k$-Net, degrades towards the center of the frequency gap [174] as the information from the subband signals does not adequately impact the prediction of the frequency gap.
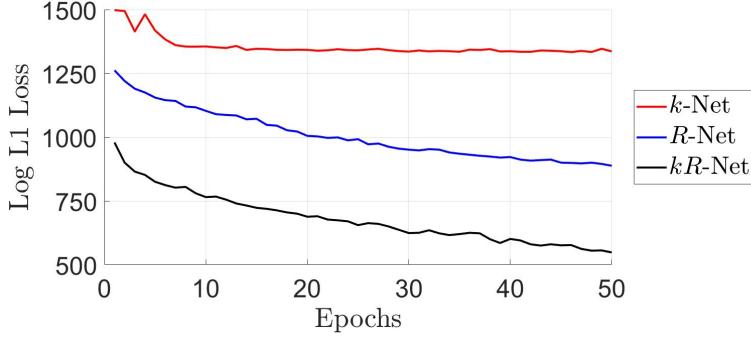


Figure 6.10: The proposed dual-domain $kR$-Net architecture outperforms conventional CNN models operating on $k$- or $R$-domain signals for training convergence and final performance. Comparison of log-scaled validation loss from 2048 samples during training.

Alternatively, the second baseline employs an architecture identical to that of $k$-Net but performs signal fusion in the $R$-domain. This technique, called $R$-Net, represents a spectral super-resolution approach to multiband signal fusion, similar to data-driven line spectra super-resolution algorithms [167, 171]. However, despite being trained and validated on identical signals, there is a noticeable performance difference as each network operates in a different domain.

136

First, we compare the convergence of $kR$-Net against the baseline networks $k$-Net and $R$-Net during training, as shown in Fig. 6.10. The three algorithms are trained for 50 epochs on the same dataset, and the same dataset is used for validation. We observe that the dual-domain architecture of $kR$-Net accelerates network training and improves the final performance. Comparatively, $k$-Net converges at a considerably slower rate and the signal fusion performance of $R$-Net is inferior to that of $kR$-Net.

Table 6.3: Comparison of SSIM, PSNR, and NRMSE across different number of targets ($N_t$) using $k$-Net, $R$-Net, and $kR$-Net

| $N_t$ | $k$-Net | | | $R$-Net | | | $kR$-Net | | |
|---|---|---|---|---|---|---|---|---|---|
| | SSIM | PSNR | NRMSE | SSIM | PSNR | NRMSE | SSIM | PSNR | NRMSE |
| 3 | 0.9972 | 50.05 | 1.01 | 0.9995 | 69.98 | 0.1026 | **0.9999** | **90.90** | **0.01779** |
| 10 | 0.9946 | 47.15 | 0.9809 | 0.9994 | 56.65 | 0.3308 | **0.9997** | **59.36** | **0.2439** |
| 100 | 0.9780 | 40.92 | 0.9475 | 0.9780 | 41.00 | 0.9411 | **0.9809** | **41.60** | **0.8779** |
| 400 | 0.9380 | 35.13 | 0.9715 | 0.9351 | 34.76 | 1.015 | **0.9425** | **35.44** | **0.9378** |
| 700 | 0.9168 | 33.23 | 0.9746 | 0.9167 | 33.05 | 0.9955 | **0.9250** | **33.72** | **0.9208** |
| 1000 | 0.9035 | 32.17 | 0.9767 | 0.8999 | 31.75 | 1.029 | **0.9094** | **32.45** | **0.9488** |
| 1300 | 0.8863 | 30.96 | 0.9500 | 0.8816 | 30.40 | 1.012 | **0.8922** | **31.14** | **0.9302** |
| Avg. | 0.9449 | 38.52 | 0.9732 | 0.9444 | 42.51 | 0.7751 | **0.9499** | **46.37** | **0.6967** |

We repeat the experiment comparing performance across different numbers of targets $N_t$, for each of the three networks, and the results are shown in Table 6.3, where the best evaluation is marked in boldface. Although $R$-Net outperforms the MPA in this experiment, it is unable to achieve numerical performance comparable to that of $kR$-Net. As $N_t$ increases, the performance of $R$-Net begins to degrade. This is likely due to the increased likelihood of closely spaced peaks in the $R$-domain and the resulting spectral blur. On the other hand, $kR$-Net demonstrates improved performance compared with both baselines and substantiates the superior ability of a hybrid approach to overcome spectral blur. The results demonstrate the superiority of $kR$-Net owing to its hybrid, dual-domain approach and verify that learning in both the $k$- and $R$-domains improves the quantitative performance for multiband signal fusion.

**Generalizability Study**

To investigate the generalizability of the proposed technique, we consider three unique multi-band scenarios with different frequency ranges, subband ranges, and number of subbands. Since each configuration requires a new network trained on data specific to those subbands, we simulate datasets with the same parameters as detailed in Section 6.2.2 for each configuration. Configuration $\mathcal{A}$ as two subbands with equivalent bandwidths of 2 GHz starting at 30 GHz and 38 GHz, respectively, resulting in a corresponding full-band of 30 GHz to 40 GHz. In Configuration $\mathcal{B}$, two subbands are used with different bandwidths, 6 GHz and 10 GHz, operating with starting frequencies of 180 GHz and 210 GHz, respectively, resulting in a full-band equivalent of 180 GHz to 220 GHz. Finally, a full-band range of 400 GHz to 460 GHz is achieved in Configuration $\mathcal{C}$ using three subbands with bandwidths of 8 GHz, 10 GHz, and 6 GHz and starting frequencies of 400 GHz, 428 GHz, and 454 GHz, respectively.

Table 6.4: Comparison of average PSNR and imaging time across different multiband configurations results across values of $N_t$ and SNR using MFT, MPA, and the $kR$-Net trained for the respective configuration.

| | Configuration $\mathcal{A}$ | | | Configuration $\mathcal{B}$ | | | Configuration $\mathcal{C}$ | | |
|---|---|---|---|---|---|---|---|---|---|
| | MFT | MPA | $kR$-Net$_{\mathcal{A}}$ | MFT | MPA | $kR$-Net$_{\mathcal{B}}$ | MFT | MPA | $kR$-Net$_{\mathcal{C}}$ |
| $\bar{N_t}$ | 31.60 | 34.12 | **43.57** | 31.10 | 37.41 | **48.50** | 32.67 | 34.07 | **40.52** |
| $\bar{\text{SNR}}$ | 33.50 | 38.82 | **43.82** | 33.64 | 35.18 | **47.86** | 33.48 | 39.19 | **44.79** |
| Time (s) | **4.412** | 4362 | 9.468 | **4.632** | 5276 | 9.673 | **5.086** | 8773 | 11.39 |

Three unique networks are trained on datasets corresponding to each multiband scenario and named $kR$-Net$_{\mathcal{A}}$, $kR$-Net$_{\mathcal{B}}$, $kR$-Net$_{\mathcal{C}}$. To evaluate the numerical performance of these networks, we repeat the experiments by comparing the performance across different numbers of targets $N_t$, and SNR for each configuration. The average results across $N_t$ and SNR are shown in Table 6.4 along with the computation time for each algorithm, where the best evaluation is marked in boldface. For all three configurations, the proposed approach achieves the best numerical performance, demonstrating the ability of the proposed network

to generalize to various realistic conditions. In addition, the MFT has the lowest computation time but the least robust image focusing, and the GPU-implemented MPA requires an excess of 1 h to impute the full-band signal with inferior reconstruction to the proposed algorithm. Since the MPA is only capable of multiband fusion in the case of two subbands, it must be computed twice for Configuration $\mathcal{C}$ to estimate the signal in the two frequency gaps. However, this increases the required computation time for the MPA to 2.4 h for each SAR image for this configuration. In contrast, the proposed algorithm only needs to be run once to perform a more robust reconstruction, requiring 11.4 s for Configuration $\mathcal{C}$. The proposed $kR$-Net demonstrates generalizability to various multisinusoidal imputation problems across a variety of frequency ranges and subband configurations. Provided adequate system design and hardware capable of producing multisinusoidal signals consistent with the pre-designed dataset, our technique can be trained for robust fusion across chip designs, vendors, SAR scanning patterns, etc. However, as described in Section 6.1, the assumption of frequency-independent scattering properties may not be valid for every application and could limit the robustness of the proposed approach depending on the spectral material characteristics of the expected targets at the operating frequencies.



(a)          (b)          (c)
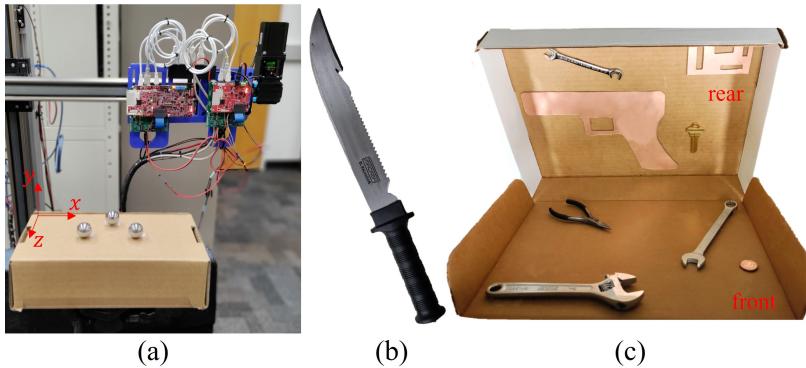
Figure 6.11: Various experimental targets: (a) metallic sphere targets, (b) large knife with serrated edge and notch near handle, and (c) hidden objects inside a cardboard box.
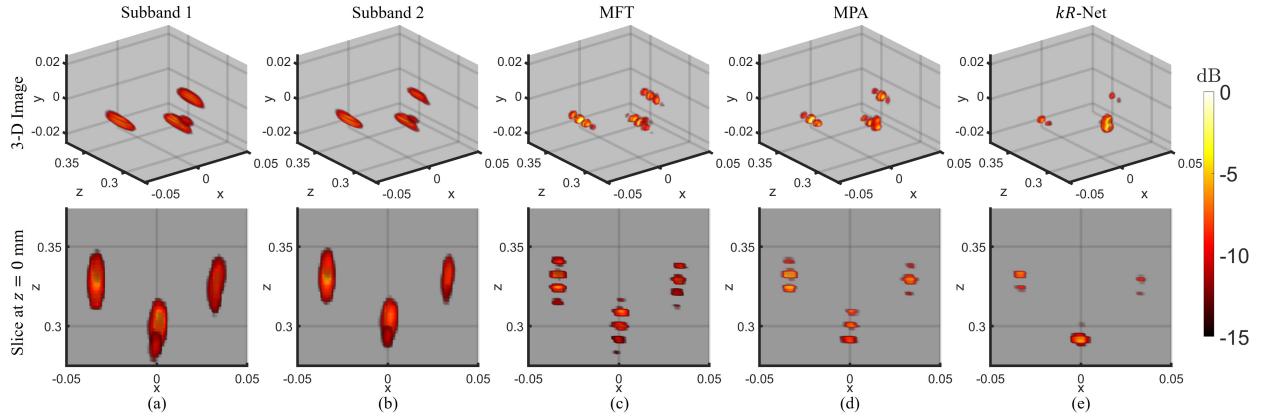
Figure 6.12: Whereas the MFT and MPA demonstrate ghost artifacts in the range direction, $kR$-Net resolves the three peaks with minimal sidelobes. Imaging results from metallic sphere targets, shown in Fig. 6.11a, using (a) subband 1 (60–64 GHz), (b) subband 2 (77–81 GHz), (c) MFT, (d) MPA, (e) $kR$-Net.

### 6.4.3  Empirical Results

Using the multiband imaging system detailed in Section 6.3, we acquire radar data of several objects at the two aforementioned subbands and compare the imaging results of the various multiband fusion algorithms.

First, three metallic spheres, each with a diameter of 1.5 cm, are placed in front of the center of the array, as shown in Fig. 6.11a. The sphere nearest to the planar scanner is separated by 3 cm from the other two spheres in the $z$-direction. The two further spheres are separated by 6 cm such that they are centered around the nearer sphere along the horizontal direction. Along the cross-range directions, the images are well resolved and focused; however, the range resolution varies depending on the approach. After sampling a planar array with dimensions of 0.125 m $\times$ 0.125 m, the recovered images are computed, as shown in Fig. 6.12. The images recovered from the first and second subbands are shown in Figs. 6.12a and 6.12b, respectively, and have low resolution in the $z$-direction because of the bandwidth of 4 GHz.

Applying the MFT to the collected data yields the image shown in Fig. 6.12c. The MFT image is plagued by sidelobes in the $z$-direction, which obscure the location of each
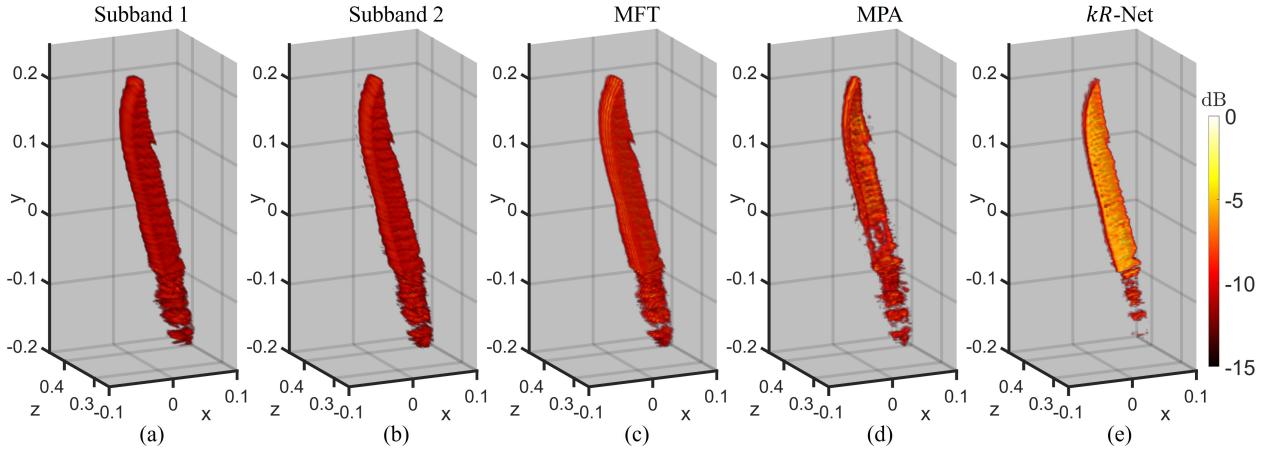
Figure 6.13: $kR$-Net recovers a high-fidelity image of the knife recovering the thin blade with high depth-resolution, while retaining the shape of the knife, which is lost when applying the MPA. Empirical imaging results from large knife, shown in Fig. 6.11b, using (a) subband 1 (60–64 GHz), (b) subband 2 (77–81 GHz), (c) MFT, (d) MPA, (e) $kR$-Net.

metallic sphere. This is due to the fact that the MFT does not account for the missing $k$-domain data in the signal fusion process. In contrast, the MPA attempts to fill the frequency gap and is relatively successful because the number of targets is small compared with the number of samples for each subband. By applying $kR$-Net, the high-resolution image in Fig. 6.12e is recovered. Compared with the image reconstructed using the MPA, the $kR$-Net image has decreased sidelobes for each of the three spheres. Even with a simple target scene, the proposed method demonstrates superior focusing performance compared with the conventional MFT and MPA approaches.

Next, we consider a large knife, as shown in Fig. 6.11b. A large array is synthesized with dimensions of 0.45 m × 0.8 m to scan the knife. The collected data are processed using the MFT, MPA, and $kR$-Net to improve the resolution, and the results are shown in Fig. 6.13. Again, the images from each subband demonstrate comparable focusing performance but are poorly resolved in the range direction because of the limited bandwidth. Without accounting for the missing data in the frequency gap, the MFT image has considerable sidelobes, as shown in Fig. 6.13c, with a similar appearance to the single-radar images in Figs. 6.13a
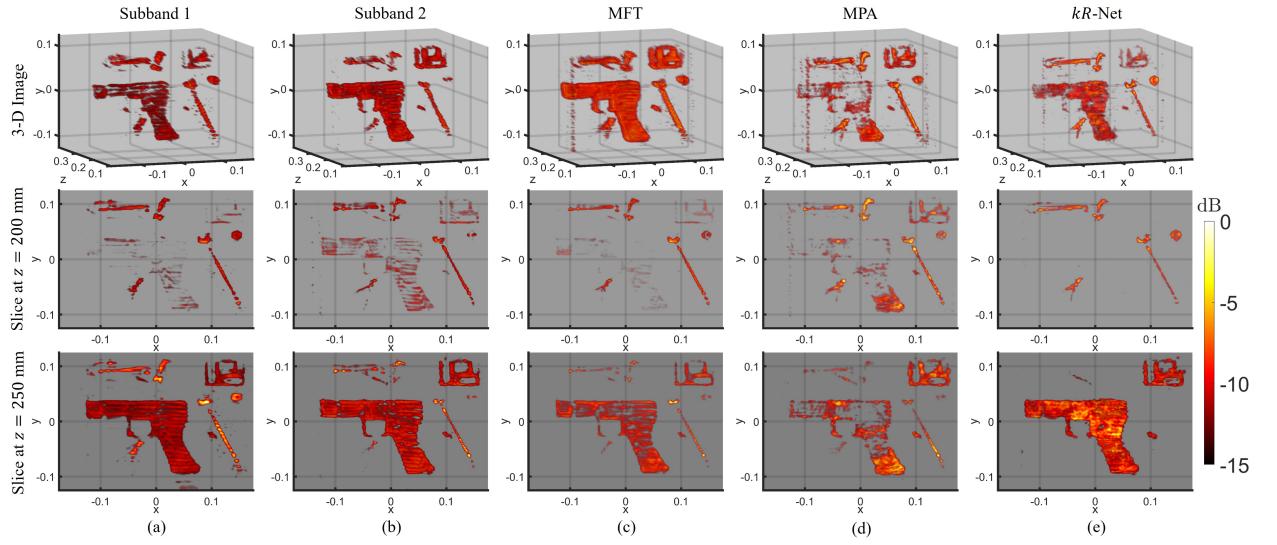
Figure 6.14: The proposed $kR$-Net separates the items at the front and rear of the box demonstrating high-fidelity super-resolution for a realistic hidden item scenario and superior performance compared with the MFT and MPA. Empirical imaging results from hidden objects target, shown in Fig. 6.11c, using (a) subband 1 (60–64 GHz), (b) subband 2 (77–81 GHz), (c) MFT, (d) MPA, (e) $kR$-Net. First row: 3-D image. Second row: slice at $z = 200$ mm, corresponding to the location of the front of the box. Third row: slice at $z = 250$ mm, corresponding to the location of the rear of the box.

and 6.13b. Although the MPA is capable of reducing the sidelobes moderately compared with the MFT, as shown in Fig. 6.13d, the structure of the knife is not retained because of the simplistic model employed by the MPA. As a result, the knife blade is distorted and ghosting is observed along the range direction. For concealed weapon detection or occluded item recognition, the poor reconstruction quality of the MPA for high-bandwidth targets, such as this knife, may prohibitively degrade the system performance. However, $kR$-Net demonstrates the best focusing performance by achieving a fine resolution in the $z$-direction while retaining the intricate features of the target. The serrated edge and notch on the knife are clearly visible in the recovered image shown in Fig. 6.13e, and the handle closely resembles the physical dimensions shown in Fig. 6.11b.

Finally, we consider a hidden object scenario with several items inside a cardboard box, as shown in Fig. 6.11c. The box is positioned such that the front of the box is located 200

mm from the radar boresight and parallel to the synthetic array. The items attached to the front of the box are separated from those in the rear by 5 cm and the box is illuminated by an array with dimensions of 0.25 m × 0.125 m. For more a closer spacing along the $z$-direction, the imaging results are expected to be further degraded. Fig. 6.14 shows the reconstructed 3-D images and slices at $z = 200$ mm and $z = 250$ mm, corresponding to the front and rear of the box, respectively. For accurate high-resolution imaging, the slice at $z = 200$ mm should contain only the objects at the front of the box and the slice at $z = 250$ mm should contain only the objects at the rear of the box.

For the two subbands with bandwidths of 4 GHz, the recovered images are spread across the $z$-direction, and both slices shown are contaminated by objects from the front and rear, as shown in Figs. 6.14a and 6.14b. The small wrench and key, both located at the rear of the box, suffer from weaker reflections and occlusion and are not well resolved by the algorithms tested. A cylindrical SAR approach [46, 43] or image enhancement algorithm [65, 66, 111, 126, 170, 125, 175] may improve image quality in the case of occlusion. The image recovered using the MFT, shown in Fig. 6.14c, exhibits the expected behavior, as sidelobes along the range direction cause ghosting, which results in the objects being visible in both range slices. Similarly, the MPA reduces the sidelobes moderately compared to the MFT, as shown in Fig. 6.14d. However, as expected from prior experiments and the inherent limitations of the MPA, the sidelobes are not mitigated, and some features of the objects are lost. The image recovered using $kR$-Net is shown in Fig. 6.14e and demonstrates improved performance in two key respects. First, the $kR$-Net image retains the high-fidelity features of the target, which are necessary for a host of applications, including image segmentation and object classification. The image quality of the wrenches is particularly notable, as the jaw of each wrench is more clearly visible compared with the images recovered from the existing approaches. Secondly, the ghosting along the $z$-direction is significantly reduced, and the objects at the front side of the box are visible only in the $z = 200$ mm slice. Likewise, the

objects at the rear side of the box are only visible in the $z = 250$ mm slice. Without the contamination observed using the MFT or MPA, the objects can be more easily localized and classified, enabling super-resolution for a host of imaging applications. The proposed hybrid, dual-domain algorithm yields high-resolution, high-fidelity images without feature loss and demonstrates improved performance over existing techniques in realistic scenarios.

Through numerical simulation and empirical analysis, the proposed algorithm demonstrates superior performance to the MFT and MPA in terms of efficiency and image quality. The proposed $kR$-Net offers improved robustness for low- and high-bandwidth target scenarios in addition to low SNR conditions. For practical imaging of complex, sophisticated targets, $kR$-Net achieves spatial super-resolution by improved multiband signal fusion without compromising the intricate features of the target. Hence, $kR$-Net is better suited for high-resolution multiband imaging applications.

## 6.5 Conclusion

In this chapter, we introduce a novel deep learning-based algorithm for multiband signal fusion for 3-D SAR super-resolution. By approaching the signal fusion problem from the wavenumber domain, we observe that imputation in the $k$-domain signal is equivalent to super-resolution in the wavenumber spectral domain. Hence, the proposed network employs a hybrid, dual-domain residual architecture that leverages the relationships in the $k$-domain and $R$-domain for improved performance. We develop a novel residual CV-CNN framework with domain transformation blocks interspersed throughout the network, resulting in superior performance compared with conventional CNN models. Compared with the MPA, which assumes a small number of reflectors in the scene or a low-bandwidth target, the proposed $kR$-Net is robust for imaging scenarios containing intricate targets consisting of many reflectors. Through simulation and empirical validation, $kR$-Net demonstrates superior imaging performance for multiband signal fusion for both low- and high-bandwidth targets. The

hybrid architecture outperforms the equivalent single-domain networks operating in either the $k$-domain or $R$-domain. Extensive numerical investigations validate the superiority of $kR$-Net compared with the conventional MFT and MPA for signal fusion, in addition to single-domain CNN models. Using a custom multi-radar mechanical prototype built from commercially available mmWave radars, we conduct imaging experiments on various targets and observe significantly improved performance of $kR$-Net in terms of image focusing and efficiency.

The hybrid-learning approach demonstrates superior numerical and computational performance compared to previous signal processing and machine learning-based solutions for multiband signal fusion. However, the hybrid concept presented in the chapter can be applied to a host of problems in sensing and signal processing. Specifically, we demonstrated the improvement by using a dual-domain CNN architecture, implying that the learned features in both domains contribute to improve performance. This conclusion alone can have a large impact on a host of signal processing problems. More generally, by exploiting the properties of the signal in conjunction with a data-driven approach, hybrid-learning algorithms can outperform traditional signal processing and machine learning methods.

# CHAPTER 7

# SUMMARY AND FUTURE DIRECTIONS

## 7.1 Summary

In this dissertation, we presented a novel approach to near-field mmWave imaging problems by leveraging machine learning and signal processing techniques in a fusion methodology, allowing for the exploitation of the advantages of both conventional and data-driven algorithms. We examined the impact of front-end signal processing techniques on deep learning perception. We investigated optimal signal processing methodologies for static and dynamic gesture recognition and demonstrated a novel sterile training technique to improve hand gesture classification rates. We developed an efficient algorithm for irregular SAR scanning geometries in the near-field, enabling technologies such as smartphone SAR imaging, near-field automotive SAR, and UAV imaging. We presented an efficient high-fidelity dataset generation platform and discussed several near-field SAR super-resolution efforts. Using the proposed framework and knowledge from previous algorithmic and system-level investigations, we developed two novel deep learning-based algorithms to overcome non-idealities in common and emerging near-field SAR imaging modalities. We designed and implemented a real-time deep learning-based super-resolution contactless interface for musical instrumentation. We presented a novel hybrid, dual-domain architecture for multiband signal fusion. The proposed framework not only outperforms existing traditional techniques, but also demonstrates superiority over single-domain networks for the same tasks, indicating the advantages of hybrid-learning algorithms for RF signal processing problems compared to exclusively signal processing or deep learning-based solutions. In simulations and experiments, we demonstrated the efficacy of hybrid-learning algorithms for improved mmWave perception, sensing, tracking, and imaging.

## 7.2 Future Directions

In this dissertation, we presented hybrid-learning algorithms for improved mmWave imaging; however, there are many future directions for hybrid-learning algorithms applied to imaging and sensing problems across the frequency spectrum. Localization and classification problems using RF sensors were explored in this dissertation by applying CNN architectures directly to radar data. Extending these analyses to include a hybrid approach is a promising direction for further improving computational efficiency and system performance.

In this dissertation, we explored mobile handheld SAR imaging systems applying a conventional technique for efficient imaging and proposed a deep learning algorithm to operate on the resulting images. Uniting these approaches would require designing a data-driven technique for image super-resolution at the data-level, rather than the image level. Such an end-to-end network could benefit from an interleaved hybrid-learning approach by applying neural processing at various stages of the image recovery process.

As demonstrated in this dissertation, an interleaved learning approach can drastically improve the multiband signal fusion performance. Similar problems in imaging and wireless communications may witness success with hybrid algorithms, including bandwidth extrapolation and line-spectra super-resolution.

Finally, as transformer architectures, an alternative to CNNs, have drastically influenced the computer vision community, they are also of interest for hybrid-algorithms in RF signal processing. The transformer can employ multiple-scale architecture, which can be understood from a signal processing perspective as a varying window size, similar to wavelet processing. This technique may be used to learn different aspects of the frequency content of signals to improve sensing, imaging, and communication performance.

The field of deep learning for RF signal processing is ripe with potential and should be explored using a hybrid approach to optimally leverage signal processing theory and data-driven methodologies.

147

# APPENDIX A

## MULTIVARIATE TAYLOR SERIES EXPANSION

Consider an infinitely differentiable real-valued function and an open neighborhood around $(u, v, w) = (u_0, v_0, w_0)$. Let $\mathbf{x} = [u \ v \ w]^T$, and $\mathbf{x}_0 = [u_0 \ v_0 \ w_0]^T$. Hence, the multivariate Taylor series expansion of $f(\mathbf{x})$ in the neighborhood of $\mathbf{x}_0$ can be written as

$$f(\mathbf{x}) = f(\mathbf{x}_0) + (\mathbf{x} - \mathbf{x}_0)^T \nabla \mathbf{f}(\mathbf{x}_0) + \frac{1}{2!}(\mathbf{x} - \mathbf{x}_0)^T \mathbf{H}(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0) + \dots, \tag{A.1}$$

where $\nabla \mathbf{f}$ is the vector of first derivatives

$$\nabla \mathbf{f}(\mathbf{x}) = \begin{bmatrix} f_u(\mathbf{x}) \\ f_v(\mathbf{x}) \\ f_w(\mathbf{x}) \end{bmatrix}, \tag{A.2}$$

and $\mathbf{H}(\mathbf{x})$ is the Hessian matrix of the second derivatives as

$$\mathbf{H}(\mathbf{x}) = \begin{bmatrix} f_{uu}(\mathbf{x}) & f_{uv}(\mathbf{x}) & f_{uw}(\mathbf{x}) \\ f_{vu}(\mathbf{x}) & f_{vv}(\mathbf{x}) & f_{vw}(\mathbf{x}) \\ f_{wu}(\mathbf{x}) & f_{wv}(\mathbf{x}) & f_{vv}(\mathbf{x}) \end{bmatrix}. \tag{A.3}$$

## A.1 Taylor Series Expansion of Round-Trip Distance for Irregular Scanning Geometries

The round-trip distance between the $\ell$-th Tx/Rx pair, whose transmitter and receiver elements are located at $(x_T, y_T, z_\ell)$ and $(x_R, y_R, z_\ell)$, respectively, and the scatterer located at $(x, y, z)$ is expressed in (3.5). Substituting (3.6) and (3.7) into (3.5), $R_\ell^{RT}$ can be expressed as a function of the distances between the Tx and Rx elements along the $x$- and $y$-directions,

$d_\ell^x$ and $d_\ell^y$, respectively, and displacement along the $z$-direction, $d_\ell^z$:

$$R_\ell^{RT}(d_\ell^x, d_\ell^y, d_\ell^z)$$

$$= \left[ (x' - \frac{d_\ell^x}{2} - x)^2 + (y' - \frac{d_\ell^y}{2} - y)^2 + (Z_0 + d_\ell^z - z)^2 \right]^{\frac{1}{2}} \quad \text{(A.4)}$$

$$+ \left[ (x' + \frac{d_\ell^x}{2} - x)^2 + (y' + \frac{d_\ell^y}{2} - y)^2 + (Z_0 + d_\ell^z - z)^2 \right]^{\frac{1}{2}}.$$

The first derivatives of (A.4), evaluated at $d_\ell^x = d_\ell^y = d_\ell^z = 0$, are

$$\frac{\partial R_\ell^{RT}}{\partial d_\ell^x}\bigg|_{(d_\ell^x = d_\ell^y = d_\ell^z = 0)} = \frac{\partial R_\ell^{RT}}{\partial d_\ell^y}\bigg|_{(d_\ell^x = d_\ell^y = d_\ell^z = 0)} = 0,$$

$$\frac{\partial R_\ell^{RT}}{\partial d_\ell^y}\bigg|_{(d_\ell^x = d_\ell^y = d_\ell^z = 0)} = \frac{2(Z_0 - z)}{R_0}, \quad \text{(A.5)}$$

where $R_0$ is the distance between the virtual monostatic element located at $(x', y', Z_0)$ and the point scatterer at $(x, y, z)$, as expressed in (3.9).

The second derivatives of (A.4), evaluated at the point of interest, can be derived as

$$\frac{\partial^2 R_\ell^{RT}}{\partial (d_\ell^x)^2}\bigg|_{(d_\ell^x = d_\ell^y = d_\ell^z = 0)} = \frac{1}{2R_0}\left[ 1 - \frac{(x' - x)^2}{R_0^2} \right],$$

$$\frac{\partial^2 R_\ell^{RT}}{\partial (d_\ell^y)^2}\bigg|_{(d_\ell^x = d_\ell^y = d_\ell^z = 0)} = \frac{1}{2R_0}\left[ 1 - \frac{(y' - y)^2}{R_0^2} \right],$$

$$\frac{\partial^2 R_\ell^{RT}}{\partial (d_\ell^z)^2}\bigg|_{(d_\ell^x = d_\ell^y = d_\ell^z = 0)} = \frac{2}{R_0}\left[ 1 - \frac{(Z_0 - z)^2}{R_0^2} \right], \quad \text{(A.6)}$$

$$\frac{\partial^2 R_\ell^{RT}}{\partial d_\ell^x d_\ell^y}\bigg|_{(d_\ell^x = d_\ell^y = d_\ell^z = 0)} = -\frac{(x' - x)(y' - y)}{2R_0^3},$$

$$\frac{\partial^2 R_\ell^{RT}}{\partial d_\ell^x d_\ell^z}\bigg|_{(d_\ell^x = d_\ell^y = d_\ell^z = 0)} = \frac{\partial^2 R_\ell^{RT}}{\partial d_\ell^y d_\ell^z}\bigg|_{(d_\ell^x = d_\ell^y = d_\ell^z = 0)} = 0.$$

Substituting (A.5) and (A.6) into (A.1), the quadratic approximation of $R_\ell$ can be expressed as

$$R_\ell^{RT} \approx 2R_0 + \frac{2(Z_0 - z)d_\ell^z}{R_0} + \frac{(d_\ell^x)^2 + (d_\ell^y)^2 + 4(d_\ell^z)^2}{4R_0}$$

$$- \frac{[(x' - x)d_\ell^x + (y' - y)d_\ell^y]^2 + 4(Z_0 - z)^2(d_\ell^z)^2}{4R_0^3}. \quad \text{(A.7)}$$

# APPENDIX B

# DOWNLOADING THE TOOLBOX AND

# ACCESSING THE DOCUMENTATION

The proposed toolbox and interactive GUI can be downloaded via GitHub or the authors' website. Using the MathWorks file exchange platform, our toolbox can be downloaded and installed to your MATLAB installation by opening the `THz and Sub-THz Imaging Toolbox.mltbx` MATLAB toolbox installer. To use the functionality of the toolbox from the repository available on our website, simply add the the main folder to the MATLAB search path.

- https://github.com/josiahwsmith10/THz-and-Sub-THz-Imaging-Toolbox

- labs.utdallas.edu/wislab/

To set up and open the directory of the toolbox, call the provided script `THzSimulator()`. (Note: the main folder of the toolbox must be on the MATLAB path and is included when installed via the MathWorks website.) To open the interactive GUI, either call `THzSimulator()` and open the application in MATLAB Application Designer or call the `THzSimulatorGUI()` script to open the stand-alone GUI.

To access the documentation created for this toolbox, after installing the packaged MAT-LAB toolbox or downloading the repository and adding the main folder to the MATLAB path, navigate to the Home tab in the main MATLAB window and select the Help button indicated by a question mark icon. The Documentation Home window will open and the THz Imaging Toolbox will appear under the Supplemental Software section[1]. Additionally,

---

[1]Refer to `www.mathworks.com/help/matlab/matlab_prog/display-custom-documentation.html` for more detailed instructions.

150

the "Info" menu and "About" tab of the interactive GUI contain quick links to the documentation. Inside our documentation, you will find the Getting Started Guide, User Guide, Reference, and Examples sections detailing the usage and functionality of the various tools.

# REFERENCES

[1] Y. Sang, L. Shi, and Y. Liu, "Micro hand gesture recognition system using ultrasonic active sensing," *IEEE Access*, vol. 6, pp. 49 339–49 347, Sep. 2018.

[2] Y. Kim and B. Toomajian, "Hand gesture recognition using micro-Doppler signatures with convolutional neural network," *IEEE Access*, vol. 4, pp. 7125–7130, Oct. 2016.

[3] S. Maragliulo, P. F. A. Lopes, L. B. Osório, A. T. De Almeida, and M. Tavakoli, "Foot gesture recognition through dual channel wearable EMG system," *IEEE Sensors J.*, vol. 19, no. 22, pp. 10 187–10 197, Nov. 2019.

[4] Y. Kim and B. Toomajian, "Application of Doppler radar for the recognition of hand gestures using optimized deep convolutional neural networks," in *Proc. 11th Eur. Conf. Antennas Propag. (EuCAP)*, Paris, France, Mar. 2017, pp. 1258–1260.

[5] J. Park and S. H. Cho, "IR-UWB radar sensor for human gesture recognition by using machine learning," in *Proc. IEEE HPCC/SmartCity/DSS*, Sydney, NSW, Australia, Dec. 2016, pp. 1246–1249.

[6] S. Y. Kim, H. G. Han, J. W. Kim, S. Lee, and T. W. Kim, "A hand gesture recognition sensor using reflected impulses," *IEEE Sensors J.*, vol. 17, no. 10, pp. 2975–2976, Mar. 2017.

[7] J. W. Smith, S. Thiagarajan, R. Willis, Y. Makris, and M. Torlak, "Improved static hand gesture classification on deep convolutional neural networks using novel sterile training technique," *IEEE Access*, vol. 9, pp. 10 893–10 902, Jan. 2021.

[8] D. M. Sheen, T. E. Hall, D. L. McMakin, A. M. Jones, and J. R. Tedeschi, "Three-dimensional radar imaging techniques and systems for near-field applications," in *Proc. SPIE*, vol. 9829, Baltimore, MD, USA, May 2016, p. 98290V.

[9] M. E. Yanik and M. Torlak, "Near-field 2-D SAR imaging by millimeter-wave radar for concealed item detection," in *Proc. IEEE Radio Wirel. Symp. (RWS)*, Orlando, FL, USA, Jan. 2019, pp. 1–4.

[10] X. Zhuge and A. Yarovoy, "Automatic target recognition in ultra-wideband 3-D images for concealed weapon detection," in *Proc. 9th Eur. Conf. Synth. Aperture Radar (EUSAR)*, Nuremberg, Germany, Apr. 2012, pp. 186–188.

[11] L. Carrer and A. G. Yarovoy, "Concealed weapon detection using UWB 3-D radar imaging and automatic target recognition," in *Proc. 8th Eur. Conf. Antennas Propag. (EuCAP)*, The Hague, Netherlands, Apr. 2014, pp. 2786–2790.

[12] L. Chao, M. N. Afsar, and K. A. Korolev, "Millimeter wave dielectric spectroscopy and breast cancer imaging," in *Proc. 7th Eur. Microw. Integr. Circuits Conf. (EuMIC)*, Amsterdam, Netherlands, Oct. 2012, pp. 572–575.

[13] Y. Gao and R. Zoughi, "Millimeter wave reflectometry and imaging for noninvasive diagnosis of skin burn injuries," *IEEE Trans. Instrum. Meas.*, vol. 66, no. 1, pp. 77–84, Nov. 2016.

[14] S. Di Meo, G. Matrone, M. Pasian, M. Bozzi, L. Perregrini, G. Magenes, A. Mazzanti, F. Svelto, P. Summers, G. Renne *et al.*, "High-resolution mm-wave imaging techniques and systems for breast cancer detection," in *Proc. IEEE/MTT-S Int. Microw. Workshop Ser. Adv. Mater. Process. RF THz Appl. (IMWS-AMP)*, Pavia, Italy, Sep. 2017, pp. 1–3.

[15] A. Mirbeik-Sabzevari, S. Li, E. Garay, H.-T. Nguyen, H. Wang, and N. Tavassolian, "Synthetic ultra-high-resolution millimeter-wave imaging for skin cancer detection," *IEEE Trans. Biomed. Eng.*, vol. 66, no. 1, pp. 61–71, Oct. 2018.

[16] A. Mirbeik-Sabzevari, N. Tavassoian, and R. Ashinoff, "Ultra-high-resolution millimeter-wave imaging: A new promising skin cancer imaging modality," in *Proc. IEEE Biomed. Circuits Syst. Conf. (BioCAS)*, Cleveland, OH, USA, Oct. 2018, pp. 1–4.

[17] A. Fedeli, C. Estatico, M. Pastorino, and A. Randazzo, "Microwave detection of brain injuries by means of a hybrid imaging method," *IEEE Open J. Antennas Propag.*, vol. 1, pp. 513–523, Sep. 2020.

[18] B. Lim, S. Son, H. Kim, S. Nah, and K. Mu Lee, "Enhanced deep residual networks for single image super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Honolulu, HI, USA, Jul. 2017, pp. 136–144.

[19] X. Glorot, A. Bordes, and Y. Bengio, "Deep sparse rectifier neural networks," in *Proc. 14th Int. Conf. Artif. Intell. Stat. (ISTATS)*, Fort Lauderdale, FL, USA, Apr. 2011, pp. 315–323.

[20] T. Björklund, A. Fiandrotti, M. Annarumma, G. Francini, and E. Magli, "Automatic license plate recognition with convolutional neural networks trained on synthetic data," in *Proc. 19th Int. Workshop Multimed. Signal Process. (MMSP)*, Dec. 2017, pp. 1–6.

[21] X. Wu, L. Liang, Y. Shi, and S. Fomel, "FaultSeg3D: Using synthetic data sets to train an end-to-end convolutional neural network for 3D seismic fault segmentation," *Geophysics*, vol. 84, no. 3, pp. IM35–IM45, Apr. 2019.

[22] D. M. Sheen, D. L. McMakin, T. E. Hall, and R. H. Severtsen, "Real-time wideband cylindrical holographic surveillance system," Jan. 1999, US Patent 5,859,609.

[23] D. M. Sheen, D. L. McMakin, and T. E. Hall, "Three-dimensional millimeter-wave imaging for concealed weapon detection," *IEEE Trans. Microw. Theory Techn.*, vol. 49, no. 9, pp. 1581–1592, Sep. 2001.

[24] D. Sheen, D. McMakin, and T. Hall, "Near-field three-dimensional radar imaging techniques and applications," *Appl. Opt.*, vol. 49, no. 19, pp. E83–E93, Jun. 2010.

[25] D. M. Sheen, A. M. Jones, and T. E. Hall, "Simulation of active cylindrical and planar millimeter-wave imaging systems," in *Proc. SPIE*, vol. 10634, Orlando, FL, USA, May 2018, p. 1063408.

[26] V. Winkler, "Range doppler detection for automotive FMCW radars," in *Proc. 4th Eur. Radar Conf. (EuRAD)*, Munich, Germany, Oct. 2007, pp. 166–169.

[27] J. Kim, J. Chun, and S. Song, "Joint range and angle estimation for fmcw mimo radar and its application," *arXiv preprint arXiv:1811.06715*, Nov. 2018.

[28] A. B. Baral and M. Torlak, "Joint doppler frequency and direction of arrival estimation for TDM MIMO automotive radars," *IEEE J. Sel. Top. Signal Process.*, vol. 15, no. 4, pp. 980–995, Apr. 2021.

[29] M. Soumekh, "Wide-bandwidth continuous-wave monostatic/bistatic synthetic aperture radar imaging," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Chicago, IL, USA, Oct. 1998, pp. 361–365.

[30] ——, *Synthetic aperture radar signal processing.* New York: Wiley, 1999, vol. 7.

[31] J. M. Lopez-Sanchez and J. Fortuny-Guasch, "3-D radar imaging using range migration techniques," *IEEE Trans. Antennas Propag.*, vol. 48, no. 5, pp. 728–737, May 2000.

[32] J. Lin and C. Huang, "3D hand posture tracking with depth gradient estimation on a RGB-D camera," in *Proc. IEEE Int. Symp. Consum. Electron. (ISCE)*, Hsinchu, Taiwan, Jun. 2013, pp. 109–110.

[33] Y. Son and O. Choi, "Image-based hand pose classification using faster R-CNN," in *Proc. 17th Int. Conf. Control Autom. Syst. (ICCAS)*, Jeju, Korea, Oct. 2017, pp. 1569–1573.

[34] G. Álvarez Narciandi, J. Laviada, and F. Las-Heras, "Towards turning smartphones into mmWave scanners," *IEEE Access*, vol. 9, pp. 45 147–45 154, Mar. 2021.

[35] M. Garcia-Fernandez, Y. Alvarez-Lopez, and F. L. Heras, "3D-SAR processing of UAV-mounted GPR measurements: Dealing with non-uniform sampling," in *Proc. 14th Eur. Conf. Antennas Propag. (EuCAP)*, Copenhagen, Denmark, Aug. 2020, pp. 1–5.

[36] T. Kan, G. Xin, L. Xiaowei, and L. Zhongshan, "Implementation of real-time auto-motive SAR imaging," in *Proc. IEEE 11th Sens. Array Multichannel Signal Process. Workshop (SAM)*, Hangzhou, China, Jun. 2020, pp. 1–4.

[37] S. Paul, F. Schwartau, M. Krueckemeier, R. Caspary, C. Monka-Ewe, J. Schoebel, and W. Kowalsky, "A systematic comparison of near-field beamforming and Fourier-based backward-wave holographic imaging," *IEEE Open J. Antennas Propag.*, vol. 2, pp. 921–931, Aug. 2021.

[38] M. A. Maisto, R. Pierri, and R. Solimene, "Sensor arrangement in monostatic subsur-face radar imaging," *IEEE Open J. Antennas Propag.*, vol. 2, pp. 3–13, Nov. 2020.

[39] M. E. Yanik, D. Wang, and M. Torlak, "Development and demonstration of MIMO-SAR mmWave imaging testbeds," *IEEE Access*, vol. 8, pp. 126 019–126 038, Jul. 2020.

[40] N. Mohammadian, O. Furxhi, R. Short, and R. Driggers, "SAR millimeter wave imag-ing systems," in *Proc. SPIE*, vol. 10994, Baltimore, MD, USA, May 2019, p. 109940A.

[41] 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, Jul. 2010.

[42] B. Wu, G. Álvarez Narciandi, and J. Laviada, "Multilayered circular dielectric struc-ture SAR imaging using time-reversal compressed sensing algorithms based on nonuni-form measurement," *IEEE Antennas Wirel. Propag. Lett.*, vol. 19, no. 9, pp. 1491–1495, Jul. 2020.

[43] J. Gao, B. Deng, Y. Qin, H. Wang, and X. Li, "Efficient terahertz wide-angle NUFFT-based inverse synthetic aperture imaging considering spherical wavefront," *Sensors*, vol. 16, no. 12, p. 2120, Dec. 2016.

[44] S. Demirci, H. Cetinkaya, M. Tekbas, E. Yigit, C. Ozdemir, and A. Vertiy, "Back-projection algorithm for ISAR imaging of near-field concealed objects," in *Proc. 30th URSI Gen. Assem. Sci. Symp. (URSI GASS)*, Istanbul, Turkey, Aug. 2011, pp. 1–4.

[45] G. Jia and W. Chang, "Modified back projection reconstruction for circular FMCW SAR," in *Proc. Inter. Radar Conf.*, Lille, France, Oct. 2014, pp. 1–5.

[46] J. W. Smith, M. E. Yanik, and M. Torlak, "Near-field MIMO-ISAR millimeter-wave imaging," in *Proc. IEEE Radar Conf. (RadarConf)*, Florance, Italy, Sep. 2020, pp. 1–6.

[47] R. K. Amineh, N. K. Nikolova, and M. Ravan, *Real-Time Three-Dimensional Imaging of Dielectric Bodies Using Microwave/Millimeter Wave Holography*. John Wiley & Sons, 2019.

[48] J. Fortuny-Guasch and J. M. Lopez-Sanchez, "Extension of the 3-D range migration algorithm to cylindrical and spherical scanning geometries," *IEEE Trans. Antennas Propag.*, vol. 49, no. 10, pp. 1434–1444, Oct. 2001.

[49] J. Detlefsen, A. Dallinger, S. Huber, S. Schelkshorn, and F. H. F. und Schaltungen, "Effective reconstruction approaches to millimeter-wave imaging of humans," in *Proc. XXVIIIth URSI Gen. Assem. Sci. Symp. (URSI GASS)*, New Delhi, India, Oct. 2005, pp. 23–29.

[50] J. Laviada, A. Arboleya-Arboleya, Y. Álvarez, B. González-Valdés, and F. Las-Heras, "Multiview three-dimensional reconstruction by millimetre-wave portable camera," *Sci. Rep.*, vol. 7, no. 1, pp. 6479–6479, Jul. 2017.

[51] M. E. Yanik and M. Torlak, "Near-field MIMO-SAR millimeter-wave imaging with sparsely sampled aperture data," *IEEE Access*, vol. 7, pp. 31 801–31 819, Mar. 2019.

[52] J. Gao, Y. Qin, B. Deng, H. Wang, and X. Li, "Novel efficient 3D short-range imaging algorithms for a scanning 1D-MIMO array," *IEEE Trans. Image Process.*, vol. 27, no. 7, pp. 3631–3643, Apr. 2018.

[53] J. Gao, B. Deng, Y. Qin, H. Wang, and X. Li, "An efficient algorithm for MIMO cylindrical millimeter-wave holographic 3-D imaging," *IEEE Trans. Microw. Theory Techn.*, vol. 66, no. 11, pp. 5065–5074, Aug. 2018.

[54] 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, Sep. 2018.

[55] M. Wang, S. Wei, J. Liang, Z. Zhou, Q. Qu, J. Shi, and X. Zhang, "TPSSI-Net: Fast and enhanced two-path iterative network for 3D SAR sparse imaging," *IEEE Trans. Image Process.*, vol. 30, pp. 7317–7332, Aug. 2021.

[56] J. Liang, J. Cao, G. Sun, K. Zhang, L. Van Gool, and R. Timofte, "SwinIR: Image restoration using swin transformer," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshop (ICCVW)*, Montreal, Canada, Oct. 2021, pp. 1833–1844.

[57] X. X. Zhu, S. Montazeri, M. Ali, Y. Hua, Y. Wang, L. Mou, Y. Shi, F. Xu, and R. Bamler, "Deep learning meets SAR: Concepts, models, pitfalls, and perspectives," *IEEE Trans. Geosci. Remote Sens.*, vol. 9, no. 4, pp. 143–172, Feb. 2021.

[58] Y. Wei, Y. Li, Z. Ding, Y. Wang, T. Zeng, and T. Long, "SAR parametric super-resolution image reconstruction methods based on ADMM and deep neural network," *IEEE Trans. Geosci. Remote Sens.*, pp. 1–16, Jan. 2021.

[59] M. B. Alver, A. Saleem, and M. Çetin, "A novel plug-and-play SAR reconstruction framework using deep priors," in *Proc. IEEE Radar Conf. (RadarConf)*, Boston, MA, USA, Apr. 2019, pp. 1–6.

[60] 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, Oct. 2020.

[61] Z. Liu, N. Wu, and X. Liao, "SAR image restoration from spectrum aliasing by deep learning," *IEEE Access*, vol. 8, pp. 40 367–40 377, Feb. 2020.

[62] C. Wu, Z. Zhang, L. Chen, and W. Yu, "Super-resolution for MIMO array SAR 3-D imaging based on compressive sensing and deep neural network," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 13, pp. 3109–3124, Jun. 2020.

[63] Z. Zhang, H. Wang, F. Xu, and Y.-Q. Jin, "Complex-valued convolutional neural network and its application in polarimetric SAR image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 12, pp. 7177–7188, Sep. 2017.

[64] M. Wang, S. Wei, J. Liang, X. Zeng, C. Wang, J. Shi, and X. Zhang, "RMIST-Net: Joint range migration and sparse reconstruction network for 3-D mmW imaging," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–17, Apr. 2021.

[65] Y. Dai, T. Jin, H. Li, Y. Song, and J. Hu, "Imaging enhancement via CNN in MIMO virtual array-based radar," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 9, pp. 7449–7458, Nov. 2021.

[66] 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.

[67] S. Z. Gurbuz, A. C. Gurbuz, E. A. Malaia, D. J. Griffin, C. S. Crawford, M. M. Rahman, E. Kurtoglu, R. Aksu, T. Macks, and R. Mdrafi, "American sign language recognition using RF sensing," *IEEE Sensors J.*, vol. 21, no. 3, pp. 3763–3775, Sep. 2020.

[68] S. Zhang, G. Li, M. Ritchie, F. Fioranelli, and H. Griffiths, "Dynamic hand gesture classification based on radar micro-Doppler signatures," in *Proc. CIE Int. Conf. Radar (RADAR)*, Guangzhou, China, Oct. 2016, pp. 1–4.

[69] S. K. Leem, F. Khan, and S. H. Cho, "Detecting mid-air gestures for digit writing with radio sensors and a CNN," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 4, pp. 1066–1081, Apr. 2020.

[70] J. S. Suh, S. Ryu, B. Han, J. Choi, J. Kim, and S. Hong, "24 GHz FMCW radar system for real-time hand gesture recognition using LSTM," in *Proc. IEEE Asia-Pacific Microw. Conf (APMC)*, Kyoto, Japan, Nov. 2018, pp. 860–862.

[71] K. Joshi, D. Bharadia, M. Kotaru, and S. Katti, "WiDeo: Fine-grained device-free motion tracing using RF backscatter," in *Proc. 12th USENIX Symp. Networked Syst. Des. Implement. (NSDI)*, Oakland, CA, USA, May 2015, pp. 189–204.

[72] K. M. Cuomo, J. E. Pion, and J. T. Mayhan, "Ultrawide-band coherent processing," *IEEE Trans. Antennas Propag.*, vol. 47, no. 6, pp. 1094–1107, Jun. 1999.

[73] B. Tian, Z. Chen, and S. Xu, "Sparse subband fusion imaging based on parameter estimation of geometrical theory of diffraction model," *IET Radar, Sonar Navigat.*, vol. 8, no. 4, pp. 318–326, Jul. 2014.

[74] Y. Q. Zou, X. Z. Gao, X. Li, and Y. X. Liu, "A matrix pencil algorithm based multiband iterative fusion imaging method," *Sci. Rep.*, vol. 6, no. 1, pp. 1–10, Jan. 2016.

[75] J. Wang, P. Aubry, and A. Yarovoy, "Wavenumber-domain multiband signal fusion with matrix-pencil approach for high-resolution imaging," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 7, pp. 4037–4049, Apr. 2018.

[76] H. H. Zhang and R. S. Chen, "Coherent processing and superresolution technique of multi-band radar data based on fast sparse Bayesian learning algorithm," *IEEE Trans. Antennas Propag.*, vol. 62, no. 12, pp. 6217–6227, Dec. 2014.

[77] Y. Zhang, T. Wang, H. Zhao, Y. Zhang, and H. Zhao, "Multiple radar subbands fusion algorithm based on support vector regression in complex noise environment," *IEEE Trans. Antennas Propag.*, vol. 66, no. 1, pp. 381–392, Jan. 2018.

[78] J. Tian, J. Sun, G. Wang, Y. Wang, and W. Tan, "Multiband radar signal coherent fusion processing with IAA and apFFT," *IEEE Signal Process. Lett.*, vol. 20, no. 5, pp. 463–466, May 2013.

[79] Z. Li, S. Papson, and R. M. Narayanan, "Data-level fusion of multilook inverse synthetic aperture radar images," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 5, pp. 1394–1406, Apr. 2008.

[80] T. Sarkar and O. Pereira, "Using the matrix pencil method to estimate the parameters of a sum of complex exponentials," *IEEE Antennas Propag. Mag.*, vol. 37, no. 1, pp. 48–55, Feb. 1995.

[81] G. Álvarez Narciandi, M. López-Portugués, F. Las-Heras, and J. Laviada, "Freehand, agile, and high-resolution imaging with compact mm-Wave radar," *IEEE Access*, vol. 7, pp. 95 516–95 526, Jul. 2019.

[82] H. F. Álvarez, G. ÁLvarez-Narciandi, F. Las-Heras, and J. Laviada, "System based on compact mmWave radar and natural body movement for assisting visually impaired people," *IEEE Access*, vol. 9, pp. 125 042–125 051, Sep. 2021.

[83] Z. Long, T. Wang, C. You, Z. Yang, K. Wang, and J. Liu, "Terahertz image super-resolution based on a deep convolutional neural network," *Appl. Opt.*, vol. 58, no. 10, pp. 2731–2735, Apr. 2019.

[84] Y. Li, W. Hu, X. Zhang, Z. Xu, J. Ni, and L. P. Ligthart, "Adaptive terahertz image super-resolution with adjustable convolutional neural network," *Opt. Express*, vol. 28, no. 15, pp. 22 200–22 217, Jul. 2020.

[85] Z. Han, W. Hu, Y. Li, Z. Xu, Y. Zhao, J. Ni, and L. Ligthart, "Terahertz image restoration with "zero-shot" super-resolution," in *Proc. IEEE CSRSWTC*, Fuzhou, China, Dec. 2020, pp. 1–3.

[86] M. Wang, S. Wei, J. Shi, Y. Wu, Q. Qu, Y. Zhou, X. Zeng, and B. Tian, "CSR-Net: A novel complex-valued network for fast and precise 3-D microwave sparse reconstruction," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 13, pp. 4476–4492, Aug. 2020.

[87] M. Wang, S. Wei, J. Liang, S. Liu, J. Shi, and X. Zhang, "Lightweight FISTA-inspired sparse reconstruction network for mmW 3-D holography," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–20, Jul. 2021.

[88] R. J. Jost and A. Uppuluri, "MATLAB-based toolkit for an introductory course in SAR image processing," in *Proc. IEEE Radar Conf. (RadarConf)*, Arlington, VA, USA, May 2005, pp. 685–690.

[89] S. Auer, R. Bamler, and P. Reinartz, "RaySAR - 3D SAR simulator: now open source," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Beijing, China, Jul. 2016, pp. 6730–6733.

[90] R. Kedzierawski, W. Czarnecki, C. Leśnik, and J.-M. Le Caillec, "MATLAB/simulink applications for SAR system design with FPGA," in *Proc. 12th Int. Radar Symp. (IRS)*, Leipzig, Germany, Sep. 2011, pp. 35–40.

[91] R. Deo, A. Jamod, V. D. R. Gopu, and Y. S. Rao, "MATLAB based SAR signal processor for educational use," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Munich, Germany, Jul. 2012, pp. 5318–5321.

[92] S. Mehta and M. Rastegari, "MobileViT: light-weight, general-purpose, and mobile-friendly vision transformer," *arXiv preprint arXiv:2110.02178*, Oct. 2021.

[93] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted residuals and linear bottlenecks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Salt Lake City, Utah, USA, Jun. 2018, pp. 4510–4520.

[94] J. W. Smith and M. Torlak, "Efficient 3-D near-field MIMO-SAR imaging for irregular scanning geometries," *IEEE Access*, vol. 10, pp. 10 283–10 294, Jan. 2022.

[95] G. Álvarez Narciandi, J. Laviada, and F. Las-Heras, "Freehand mm-Wave imaging with a compact MIMO radar," *IEEE Trans. Antennas Propag.*, vol. 69, no. 2, pp. 1224–1229, Feb. 2021.

[96] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Honolulu, HI, USA, Jan. 2017, pp. 1125–1134.

[97] Z. Liu, H. Mao, C.-Y. Wu, C. Feichtenhofer, T. Darrell, and S. Xie, "A convnet for the 2020s," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, New Orleans, LA, USA, Jun. 2022, pp. 11 966–11 976.

[98] C. Wang, Q. Zhang, J. Hu, C. Li, S. Shi, and G. Fang, "An efficient algorithm based on CSA for THz stepped-frequency SAR imaging," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, Dec. 2020.

[99] M. E. Yanik and M. Torlak, "Millimeter-wave near-field imaging with two-dimensional SAR data," in *Proc. SRC Techcon*, no. P093929, Austin, TX, USA, Sep. 2018.

[100] T. Winkler, "Making motion musical: Gesture mapping strategies for interactive computer music," in *Proc. 21st Int. Comput. Music Conf. (ICMC)*, Banff, AB, Canada, Sep. 1995, pp. 261–264.

[101] 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, Jan. 2020.

[102] J. W. Smith and M. Torlak, "Survey of emerging systems and algorithms for near-field THz SAR imaging," in *Proc. IEEE*, to be submitted.

[103] J. Baek, J. Kim, and E. Kim, "Comparison study of different feature classifiers for hand posture classification," in *Proc. 13th Int. Conf. Control Autom. Syst. (ICCAS)*, Gwangu, Korea, Oct. 2013, pp. 683–687.

[104] M. Matilainen, P. Sangi, J. Holappa, and O. Silvén, "OUHANDS database for hand detection and pose recognition," in *Proc. 6th Int. Conf. Image Process. Theory Tools Appl. (IPTA)*, Oulu, Finland, Dec. 2016, pp. 1–5.

[105] A. Anaz, M. Skubic, J. Bridgeman, and D. M. Brogan, "Classification of therapeutic hand poses using convolutional neural networks," in *Proc. 40th Annu. Int. Conf. IEEE Eng. Med. Biol. (EMBC)*, Honolulu, HI, USA, Jul. 2018, pp. 3874–3877.

[106] F. Bernardo, N. Arner, and P. Batchelor, "O soli mio: exploring millimeter wave radar for musical interaction," in *Proc. 17th Int. Conf. New Interfaces for Musical Expression (NIME)*, Aalborg, Denmark, May 2017, pp. 283–286.

[107] D. Griffin and J. Lim, "Signal estimation from modified short-time fourier transform," *IEEE Trans. Acoust. Speech, Signal Process.*, vol. 32, no. 2, pp. 236–243, Apr. 1984.

[108] B. Dekker, S. Jacobs, A. S. Kossen, M. C. Kruithof, A. G. Huizing, and M. Geurts, "Gesture recognition with a low power FMCW radar and a deep convolutional neural network," in *Proc. 14th Eur. Radar Conf. (EuRAD)*, Nuremberg, Germany, Oct. 2017, pp. 163–166.

[109] O. Li, J. He, K. Zeng, Z. Yu, X. Du, Y. Liang, G. Wang, Y. Chen, P. Zhu, W. Tong, D. Lister, and L. Ibbotson, "Integrated sensing and communication in 6G a prototype of high resolution THz sensing on portable device," in *Proc. IEEE Eur. Conf. Netw. Commun. 6G Summit (EuCNC/6G)*, Porto, Portugal, Jun. 2021, pp. 544–549.

[110] K. Basrawi and R. Dill, "Reverse engineering the Soli radar API for military applications," in *Proc. IEEE Radar Conf. (RadarConf)*, Jun. 2021, pp. 1–8.

[111] J. W. Smith, O. Furxhi, and M. Torlak, "An FCNN-based super-resolution mmWave radar framework for contactless musical instrument interface," *IEEE Trans. Multimedia*, vol. 24, pp. 2315–2328, May 2021.

[112] G. Álvarez Narciandi, J. Laviada, Y. Álvarez López, G. Ducournau, C. Luxey, C. Belem-Goncalves, F. Gianesello, N. Nachabe, C. D. Rio, and F. Las-Heras, "Freehand system for antenna diagnosis based on amplitude-only data," *IEEE Trans. Antennas Propag.*, vol. 69, no. 8, pp. 4988–4998, Feb. 2021.

[113] H. Wymeersch, G. Seco-Granados, G. Destino, D. Dardari, and F. Tufvesson, "5G mmWave positioning for vehicular networks," *IEEE Wireless Commun.*, vol. 24, no. 6, pp. 80–86, Dec. 2017.

[114] Z. Hajiakhondi-Meybodi, M. Salimibeni, K. N. Plataniotis, and A. Mohammadi, "Bluetooth low energy-based angle of arrival estimation via switch antenna array for indoor localization," in *Proc. 23rd Int. Conf. Inf. Fusion (FUSION)*, Rustenburg, South Africa, Jul. 2020, pp. 1–6.

[115] M. E. Yanik, D. Wang, and M. Torlak, "3-D MIMO-SAR imaging using multi-chip cascaded millimeter-wave sensors," in *Proc. IEEE Global Conf. Signal Inf. Process. (GlobalSIP)*, Ottawa, ON, Canada, Nov. 2019, pp. 1–5.

[116] C. F. Baumgartner, K. Kamnitsas, J. Matthew, T. P. Fletcher, S. Smith, L. M. Koch, B. Kainz, and D. Rueckert, "SonoNet: Real-time detection and localisation of fetal standard scan planes in freehand ultrasound," *IEEE Trans. Med. Imaging*, vol. 36, no. 11, pp. 2204–2215, Jul. 2017.

[117] J. Blackall, G. Penney, A. King, and D. Hawkes, "Alignment of sparse freehand 3-D ultrasound with preoperative images of the liver using models of respiratory motion and deformation," *IEEE Trans. Med. Imaging*, vol. 24, no. 11, pp. 1405–1416, Oct. 2005.

[118] M. W. Gilbertson and B. W. Anthony, "Force and position control system for freehand ultrasound," *IEEE Trans. Robot.*, vol. 31, no. 4, pp. 835–849, Jun. 2015.

[119] X. Zeng, Y. Ma, Z. Li, J. Wu, and J. Yang, "A near-field fast time-frequency joint 3-D imaging algorithm based on aperture linearization," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Brussels, Belgium, Oct. 2021, pp. 5163–5166.

[120] B. Fan, J. Gao, H. Li, Z. Jiang, and Y. He, "Near-field 3D SAR imaging using a scanning linear MIMO array with arbitrary topologies," *IEEE Access*, vol. 8, pp. 6782–6791, Dec. 2019.

[121] J. H. G. Ender and J. Klare, "System architectures and algorithms for radar imaging by MIMO-SAR," in *Proc. IEEE Radar Conf. (RadarConf)*, Pasadena, CA, USA, May 2009, pp. 1–6.

[122] F. Roos, J. Bechter, C. Knill, B. Schweizer, and C. Waldschmidt, "Radar sensors for autonomous driving: Modulation schemes and interference mitigation," *IEEE Microw. Mag.*, vol. 20, no. 9, pp. 58–72, Aug. 2019.

[123] L. Greengard and J.-Y. Lee, "Accelerating the nonuniform fast Fourier transform," *SIAM Rev.*, vol. 46, no. 3, pp. 443–454, Aug. 2006.

[124] T. Zhang and X.-G. Xia, "OFDM synthetic aperture radar imaging with sufficient cyclic prefix," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 1, pp. 394–404, Jan. 2015.

[125] 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 *Proc. IEEE Tex. Symp. Wirel. Microw. Circuits Syst. (WMCS)*, Waco, TX, USA, Apr. 2022, pp. 1–6.

[126] C. Vasileiou, J. W. Smith, S. Thiagarajan, M. Nigh, Y. Makris, and M. Torlak, "Efficient CNN-based super resolution algorithms for mmWave mobile radar imaging," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Bourdeaux, France, Oct. 2022, pp. 3803–3807.

[127] Q. Guo, J. Liang, T. Chang, and H.-L. Cui, "Millimeter-wave imaging with accelerated super-resolution range migration algorithm," *IEEE Trans. Microw. Theory Techn.*, vol. 67, no. 11, pp. 4610–4621, Jul. 2019.

[128] L. Qiao, Y. Wang, Z. Shen, Z. Zhao, and Z. Chen, "Compressive sensing for direct millimeter-wave holographic imaging," *Appl. Opt.*, vol. 54, no. 11, pp. 3280–3289, Apr. 2015.

[129] H.-S. Shin and J.-T. Lim, "Range migration algorithm for airborne squint mode spotlight SAR imaging," *IET Radar, Sonar, Nav.*, vol. 1, no. 1, pp. 77–82, Feb. 2007.

[130] 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, Feb. 2012.

[131] R. Zhu, J. Zhou, G. Jiang, and Q. Fu, "Range migration algorithm for near-field MIMO-SAR imaging," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 12, pp. 2280–2284, Nov. 2017.

[132] J. Ding, M. Kahl, O. Loffeld, and P. H. Bolivar, "THz 3-D image formation using SAR techniques: simulation, processing and experimental results," *IEEE Trans. Terahertz Sci. Technol.*, vol. 3, no. 5, pp. 606–616, Jul. 2013.

[133] A. Batra, M. Wiemeler, D. Göhringer, and T. Kaiser, "Sub-mm resolution 3D SAR imaging at 1.5 THz," in *Proc. 4th Int. Workshop Mob. Terahertz Syst. (IWMTS)*, Essen, Germany, Jul. 2021, pp. 1–5.

[134] A. Batra, J. Barowski, D. Damyanov, M. Wiemeler, I. Rolfes, T. Schultze, J. C. Balzer, D. Göhringer, and T. Kaiser, "Short-range SAR imaging from GHz to THz waves," *IEEE J. Microwaves*, vol. 1, no. 2, pp. 574–585, Apr. 2021.

[135] X. Yang, Y. Pi, T. Liu, and H. Wang, "Three-dimensional imaging of space debris with space-based terahertz radar," *IEEE Sensors J.*, vol. 18, no. 3, pp. 1063–1072, Dec. 2018.

[136] T. Jaeschke, C. Bredendiek, and N. Pohl, "3D FMCW SAR imaging based on a 240 GHz SiGe transceiver chip with integrated antennas," in *Proc. Ger. Microw. Conf (GeMiC)*, Aachen, Germany, Mar. 2014, pp. 1–4.

[137] J. Moll, P. Schops, and V. Krozer, "Towards three-dimensional millimeter-wave radar with the bistatic fast-factorized back-projection algorithm—potential and limitations," *IEEE Trans. Terahertz Sci. Technol.*, vol. 2, no. 4, pp. 432–440, Jun. 2012.

[138] L. A. Gorham and L. J. Moore, "SAR image formation toolbox for MATLAB," in *Proc. SPIE*, vol. 7699, Orlando, FL, USA, Apr. 2010, p. 769906.

[139] B. Schweizer, C. Knill, D. Schindler, and C. Waldschmidt, "Stepped-carrier OFDM-radar processing scheme to retrieve high-resolution range-velocity profile at low sampling rate," *IEEE Trans. Microw. Theory Techn.*, vol. 66, no. 3, pp. 1610–1618, Mar. 2018.

[140] H. Sarieddeen, N. Saeed, T. Y. Al-Naffouri, and M.-S. Alouini, "Next generation terahertz communications: A rendezvous of sensing, imaging, and localization," *IEEE Commun. Mag.*, vol. 58, no. 5, pp. 69–75, May 2020.

[141] A. Bourdoux, U. Ahmad, D. Guermandi, S. Brebels, A. Dewilde, and W. Van Thillo, "PMCW waveform and MIMO technique for a 79 GHz CMOS automotive radar," in *Proc. IEEE Radar Conf. (RadarConf)*, Philadelphia, PA, USA, Jun. 2016, pp. 1–5.

[142] H. Iqbal, A. Löffler, M. N. Mejdoub, and F. Gruson, "Realistic SAR implementation for automotive applications," in *Proc. 17th Eur. Radar Conf. (EuRAD)*, Utrecht, Netherlands, Jan. 2021, pp. 306–309.

[143] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst. (NeurIPS)*, Long Beach, CA, USA, Dec. 2017, pp. 5998–6008.

[144] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly *et al.*, "An image is worth 16x16 words: Transformers for image recognition at scale," *arXiv preprint arXiv:2010.11929*, 2020.

[145] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo, "Swin transformer: Hierarchical vision transformer using shifted windows," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Montreal, Canada, Oct. 2021, pp. 10 012–10 022.

[146] H. Dong, L. Zhang, and B. Zou, "Exploring vision transformers for polarimetric SAR image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–15, Nov. 2021.

[147] L. Zheng, J. Bai, X. Zhu, L. Huang, C. Shan, Q. Wu, and L. Zhang, "Dynamic hand gesture recognition in in-vehicle environment based on FMCW radar and transformer," *Sensors*, vol. 21, no. 19, p. 6368, Sep. 2021.

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

[149] S. Huang, J. Qian, Y. Wang, X. Yang, and L. Yang, "Through-the-wall radar super-resolution imaging based on cascade U-Net," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Yokohama, Japan, Aug. 2019, pp. 2933–2936.

[150] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Proc. Adv. Neural Inf. Process. Syst. (NeurIPS)*, Montreal, Canada, Dec. 2014.

[151] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. MICCAI*, Munich, Germany, Oct. 2015, pp. 234–241.

[152] M. Mirza and S. Osindero, "Conditional generative adversarial nets," *arXiv preprint arXiv:1411.1784*, Nov. 2014.

[153] K. Armanious, C. Jiang, M. Fischer, T. Küstner, T. Hepp, K. Nikolaou, S. Gatidis, and B. Yang, "MedGAN: Medical image translation using GANs," *Comput. Med. Imaging Graph.*, vol. 79, p. 101684, 2020.

[154] K. Armanious, S. Abdulatif, F. Aziz, U. Schneider, and B. Yang, "An adversarial super-resolution remedy for radar design trade-offs," in *Proc. 27th Eur. Signal Process. Conf. (EUSIPCO)*, A Coruna, Spain, Sep. 2019, pp. 1–5.

[155] C. Wang, C. Xu, C. Wang, and D. Tao, "Perceptual adversarial networks for image-to-image transformation," *IEEE Trans. Image Process.*, vol. 27, no. 8, pp. 4066–4079, Aug. 2018.

[156] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Honolulu, HI, USA, Jul. 2017, pp. 1125–1134.

[157] Y. Guo, Y. Li, L. Wang, and T. Rosing, "Depthwise convolution is all you need for learning multiple visual domains," in *Proc. 35th AAAI Conf. Artif. Intell. (AAAI)*, vol. 33, no. 01, Honolulu, HI, USA, Jan. 2019, pp. 8368–8375.

[158] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "MobileNets: Efficient convolutional neural networks for mobile vision applications," *arXiv preprint arXiv:1704.04861*, Apr. 2017.

[159] J. W. Smith and M. Torlak, "Deep learning-based multiband signal fusion for 3-D SAR super-resolution," *IEEE Trans. Aerosp. Electron. Syst.*, pp. 1–17, Apr. 2023.

[160] S. Xie, R. Girshick, P. Dollár, Z. Tu, and K. He, "Aggregated residual transformations for deep neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Honolulu, HI, USA, Jul. 2017, pp. 1492–1500.

[161] Y. Xie, M. B. Wakin, and G. Tang, "Contaminated multiband signal identification via deep learning," in *Proc. IEEE Stat. Signal Process. Workshop (SSP)*, Rio de Janeiro, Brazil, Jul. 2021, pp. 76–80.

[162] Y. Sun, X. Liang, H. Fan, M. Imran, and H. Heidari, "Visual hand tracking on depth image using 2-D matched filter," in *Proc. UK/China Emerg. Technol. (UCET)*, Glasgow, United Kingdom, Aug. 2019, pp. 1–4.

[163] R. Polfreman, "Multi-modal instrument: towards a platform for comparative controller evaluation," in *Proc. 37th Int. Comput. Music Conf. (ICMC)*, Huddersfield, United Kingdom, Jul. 2011, pp. 147–150.

[164] S. Rao, "Intro to mmWave sensing: FMCW radars," June 26, 2017. Accessed: Apr. 2021. [Online]. Available: https://training.ti.com/node/1139153

[165] H. Kim, S. You, B. J. Jeong, and W. Byun, "Azimuth angle resolution improvement technique with neural network," in *Proc. Int. Conf. Inf. Commun. Technol. Converg. (ICTC)*, Jeju, Korea, Oct. 2020, pp. 1384–1387.

[166] J. García, A. Gardel, I. Bravo, J. L. Lázaro, and M. Martínez, "Tracking people motion based on extended condensation algorithm," *IEEE Trans. Cybern.*, vol. 43, no. 3, pp. 606–618, 2013.

[167] G. Izacard, S. Mohan, and C. Fernandez-Granda, "Data-driven estimation of sinusoid frequencies," in *Proc. Adv. Neural Inf. Process. Syst. (NeurIPS)*, vol. 32, Vancouver, BC, Canada, Dec. 2019, pp. 5127–5137.

[168] V. T. Vu, M. I. Pettersson, A. Batra, and T. Kaiser, "Fourier transform of SAR data cube and 3-D range migration algorithm," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 58, no. 3, pp. 2584–2591, Dec. 2022.

[169] X. Wang and T. Jiang, "Multi-band synthesis of wideband radar based on compressed sensing," in *Proc. Int. Appl. Comput. Electromagn. Soc. Symp. China (ACES-China)*, Nanjing, China, Aug. 2019, pp. 1–2.

[170] H. Zhang, R. Song, S. Chen, G. Wang, Y. Jia, and C. Yan, "Target imaging based on generative adversarial nets in through-wall radar imaging," in *Proc. Int. Conf. Control Autom. Inf. Sci. (ICCAIS)*, Chengdu, China, Oct. 2019, pp. 1–6.

[171] P. Pan, Y. Zhang, Z. Deng, and G. Wu, "Complex-valued frequency estimation network and its applications to superresolution of radar range profiles," *IEEE Trans. Geosci. Remote Sens.*, pp. 1–12, Oct. 2021.

[172] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Las Vegas, NV, USA, Jun. 2016, pp. 770–778.

[173] S. Iizuka, E. Simo-Serra, and H. Ishikawa, "Globally and locally consistent image completion," *ACM Trans. Graph.*, vol. 36, no. 4, pp. 1–14, Jul. 2017.

[174] W. Luo, Y. Li, R. Urtasun, and R. Zemel, "Understanding the effective receptive field in deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst. (NeurIPS)*, Barcelona, Spain, Dec. 2016, pp. 4898–4906.

[175] X. Yin, Z. Hu, J. Zheng, B. Li, and Y. Zuo, "Study on radar echo-filling in an occlusion area by a deep learning algorithm," *Remote Sensing*, vol. 13, no. 9, p. 1779, 2021.

# BIOGRAPHICAL SKETCH

Josiah W. Smith received the BSEE degree (*summa cum laude*) in electrical engineering from The University of Texas at Dallas in 2019, where he is currently pursuing a PhD degree in electrical engineering specializing in communications engineering.

In 2019, he was an undergraduate research intern at the Texas Analog Center of Excellence (TxACE) working on radar reception for hand gesture recognition and high resolution near-field MIMO synthetic aperture radar imaging algorithms. During the summer of 2020, he developed real-time human-computer interaction algorithms for mmWave radar with imec-USA. In the summer of 2021, he developed advanced deep learning and data-driven algorithms for user experience enhancement at Apple in the Display Technologies group.

Mr. Smith was awarded the Texas Instruments Analog Excellence Graduate Fellowship in August 2019 and the Louis Beecherl, Jr. Graduate Research Fellowship in August 2021. In 2020, he was awarded first alternate for the student paper competition in the *IEEE Radar Conf. 2020*, ranking 6/143 and best poster award at the *2020 TxACE Symposium*. He is a student IEEE member. His current research interests include new regime radar imaging algorithm development, ultrawideband radar imaging algorithms, terahertz radar, radar perception, computer vision, machine learning, millimeter-wave sensing, and phased array signal processing.

CURRICULUM VITAE

# Josiah W. Smith

March 31, 2022

## Contact Information:

Department of Electrical and Computer Eng.
The University of Texas at Dallas
800 W. Campbell Rd.
Richardson, TX 75080-3021, U.S.A.

## Educational History:

BSEE, Electrical Engineering, The University of Texas at Dallas, 2019

## Employment History:

Research Assistant, The University of Texas at Dallas, August 2019 – May 2022
Machine Learning Intern, Apple - EE Display Team, March 2021 – August 2021
Computational Imaging Intern, imec-USA, March 2020 – August 2020
Undergraduate Researcher, Texas Analog Center of Excellence (TxACE), January 2019 – August 2019

## Publications :

[J1] **J. W. Smith**, S. Thiagarajan, R. Willis, Y. Makris and M. Torlak, "Improved Static Hand Gesture Classification on Deep Convolutional Neural Networks Using Novel Sterile Training Technique," *IEEE Access*, vol. 9, pp. 10893–10902, Jan. 2021.

[J2] **J. W. Smith**, O. Furxhi, M. Torlak, "An FCNN-Based Super-Resolution mmWave Radar Framework for Contactless Musical Instrument Interface," *IEEE Trans. on Multimedia*, May 2021.

[J3] **J. W. Smith** and M. Torlak, "Efficient 3-D Near-Field MIMO-SAR Imaging for Irregular Scanning Geometries," *IEEE Access*, vol. 10, pp. 10283–10294. Jan. 2022.

[J4] **J. W. Smith** and M. Torlak, "Deep learning-based multiband signal fusion for 3-D SAR super-resolution," in *IEEE Trans. Aerosp. Electron. Syst.*, Apr. 2023.

[J5] **J. W. Smith** and M. Torlak, "Survey of emerging systems and algorithms for near-field THz SAR imaging," *Proc. IEEE*, submitted for publication.

[C1] **J. W. Smith**, M. E. Yanik and M. Torlak, "Near-Field MIMO-ISAR Millimeter-Wave Imaging," in *Proc. IEEE Radar Conf. (RadarConf)*, Florence, Italy, Sep. 2020, pp. 1–6.

[C2] **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 *Proc. IEEE Tex. Symp. Wirel. Microw. Circuits Syst. (WMCS)*, Waco, TX, USA, Apr. 2022, pp. 1–6.

[C3] C. Vasieleiou, **J. W. Smith**, S. Thiagarajan, M. Nigh, Y. Makris, and M. Torlak, "Efficient CNN-based super resolution algorithms for mmWave mobile radar imaging," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Bourdeaux, France, Oct. 2022, pp. 3803–3807.

## Professional Recognitions and Honors:

Louis Beecherl, Jr. Graduate Research Fellowship, Aug. 2021
Best Poster Award, TxACE Symposium, Oct. 2020
First Alternate Student Paper (Ranked 6/143), IEEE Radar Conference 2020, Sep. 2020
Texas Instruments Analog Excellence Graduate Fellowship, Aug. 2019
First Place Senior Capstone Project, UTDesign II Expo, May 2019
First Place Senior Capstone Project, UTDesign I Expo, Dec. 2018
Greater Texas Foundation Removing Educational Barriers Endowed Scholarship, Aug. 2017
Academic Excellence Full-Ride Scholarship, The University of Texas at Dallas, Aug. 2016

## Professional Memberships:

Institute of Electrical and Electronics Engineers (IEEE), 2016 – present