NOVEL HYBRID-LEARNING ALGORITHMS FOR IMPROVED MILLIMETER-WAVE IMAGING SYSTEMS

by

JOSIAH W. SMITH, BSEE

DISSERTATION PROPOSAL

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

Dec 2021

NOVEL HYBRID-LEARNING ALGORITHMS FOR IMPROVED MILLIMETER-WAVE IMAGING SYSTEMS

Josiah W. Smith, PhD
The University of Texas at Dallas, 2021

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 being 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 for 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 reduce the training load 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 proposal 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 in both the forward synthetic aperture radar (SAR) and inverse synthetic aperture radar

(ISAR) modes; and mmWave image super-resolution using a deep neural network (DNN). 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 work has been conducted on deep learning with images [18, 19, 20, 21] and conventional signal processing on 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 based on the 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 from 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 to 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 currently requires computationally expensive techniques. Efficient algorithms for common SAR patterns (linear [37, 38], rectilinear/planar [8, 39, 40, 41], circular [42, 43, 44, 45], or cylindrical [46, 47, 48, 49, 50]) have been investigated in the existing literature, but computationally tractable algorithms remain unexplored.

On the other hand, while much work on high-resolution near-field mmWave imaging has been conducted on the signal processing front [8, 51, 39, 46, 52, 53], incorporating data-driven techniques into the imaging pipeline remains unaddressed. Optical image super-resolution techniques have gained significant attention in recent years from the machine learning community [18, 54], but SAR image super-resolution has largely been relegated to the far-field imaging regime [55, 56, 57, 58, 59, 60, 61]. However, near-field SAR super-resolution presents several unique challenges, particularly the availability of large, meaningful datasets for training convolutional neural network (CNN)-based algorithms.

Lastly, 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. 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 proposal 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

in both the conventional radar signal processing domain and the 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 resolution.

On 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, taking into account 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 propose 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 research [34, 62, 63].

While existing studies on SAR image super-resolution ignore the spherical-wave phenomena of near-field imaging [55, 56, 57, 58, 59, 60, 61], we propose an open-source software platform for high-fidelity near-field dataset generation. Similar existing software implementations in the literature [64, 65, 66, 67] similarly only address the simplified far-field scenario and are unable to produce data relevant for near-field imaging. Using the proposed framework to generate large, meaningful datasets, we consider several methods for improved imaging using hybrid-learning. An image enhancement network is trained on exclusively simulated data performing spatial super-resolution on near-field SAR images.

Finally, we propose an end-to-end interleaved hybrid-learning approach for near-field imaging employing deep learning super-resolution and regression techniques [18, 54] throughout the imaging signal chain. Fully embracing hybrid-learning, this technique allows a network to learn the various characteristics of the signal across spatial and spectral domains. The proposed hybrid-learning algorithms may pose significant performance gains for both image fidelity and computational load.

1.3 Contributions and Proposed Work

In response to the challenges and opportunities of hybrid-learning algorithms, we present several initial studies on data-driven solutions to radar imaging problems and propose novel methods for sensing, tracking, imaging, and super-resolution 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 on the challenges and opportunities of static gesture recognition by examining several datasets and data preprocessing techniques. We explore the trade-offs of CNN classifiers for dynamic hand gesture recognition. The 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 then return to the static gesture recognition problem in attempt 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 method's ability to improve classification accuracy of real human hand gestures. The 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. Additionally, we extend the work of [34, 35, 42, 62, 63, 68, 69] by proposing a novel imaging algorithm to enable efficient near-field irregular SAR. This work will address the need for efficient imaging algorithms for edge applications, such as smartphone imaging and automotive SAR. Subsequent deep learning for classification or super-resolution demands high-fidelity SAR images under computational constraints. The contribution will be detailed in the following publication:

    - J. W. Smith and M. Torlak, "Efficient 3-D near-field MIMO-SAR imaging for irregular scanning geometries," *IEEE Access*, Nov. 2021, submitted.

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, O. Furxhi, M. Torlak, "Radar Musical Instrument - A Spatiotemporal Real-Time mmWave Sensor for Contactless Human Computer Interaction," in *IEEE Trans. on Multimedia*, May 2021.
    - J. W. Smith and M. Torlak, "Efficient 3-D near-field MIMO-SAR imaging for irregular scanning geometries," *IEEE Access*, Nov. 2021, submitted.
    - J. W. Smith and M. Torlak, "Survey of near-field THz array imaging systems with interactive toolbox," *IEEE Trans. THz Sci. Technol.*, to be submitted.

5. We then propose extending the work of [70] on far-field SAR image enhancement using complex-valued neural networks to the near-field imaging modality. Using the software toolbox, we generate large synthetic datasets on which to train a neural processor to perform 3-D SAR image super-resolution. A thorough discussion of the contribution will be provided in the following publications:

- J. W. Smith, O. Furxhi, M. Torlak, "Radar Musical Instrument - A Spatiotemporal Real-Time mmWave Sensor for Contactless Human Computer Interaction," in *IEEE Trans. on Multimedia*, May 2021.
- J. W. Smith and M. Torlak, "Survey of near-field THz array imaging systems with interactive toolbox," *IEEE Trans. THz Sci. Technol.*, to be submitted.

6. 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 the 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 the radar signatures. This contribution is based on the following publication:

   - J. W. Smith, O. Furxhi, M. Torlak, "Radar Musical Instrument - A Spatiotemporal Real-Time mmWave Sensor for Contactless Human Computer Interaction," in *IEEE Trans. on Multimedia*, May 2021.

7. Finally, we will propose a novel end-to-end hybrid-learning technique with interleaved deep learning throughout the image reconstruction signal chain. Extending our work in [46], we will develop advanced algorithms for MIMO-SAR and ISAR image super-resolution competing with state-of-the-art approaches in terms of image resolution and computational load. The contribution will be covered in the following publications:

   - J. W. Smith, M. E. Yanik and M. Torlak, "Near-Field MIMO-ISAR Millimeter-Wave Imaging," *Proc. IEEE Radar Conf.*, Florence, Italy, 2020, pp. 1-6.

- J. W. Smith and M. Torlak, "End-to-end hybrid-learning for near-field SAR super-resolution using swin transformer," *IEEE Trans. Microw. Theory Techn.*, to be submitted.

1.4 Outline

The rest of the dissertation proposal is organized as follows:

- Chapter 2 details the FMCW signal model employed at length throughout this dissertation proposal.
- 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 preliminary near-field SAR image super-resolution using hybrid-learning algorithms.
- Chapter 5 will address an interleaved approach to perception and imaging using hybrid-learning.
- Conclusions, summary, and discussion of proposed work are detailed in Chapter 6.

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 proposal.

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 σ located at the point (x_0,y_0,z_0). The radar transceiver is positioned on the x'-y' plane located at $z = Z_0$.

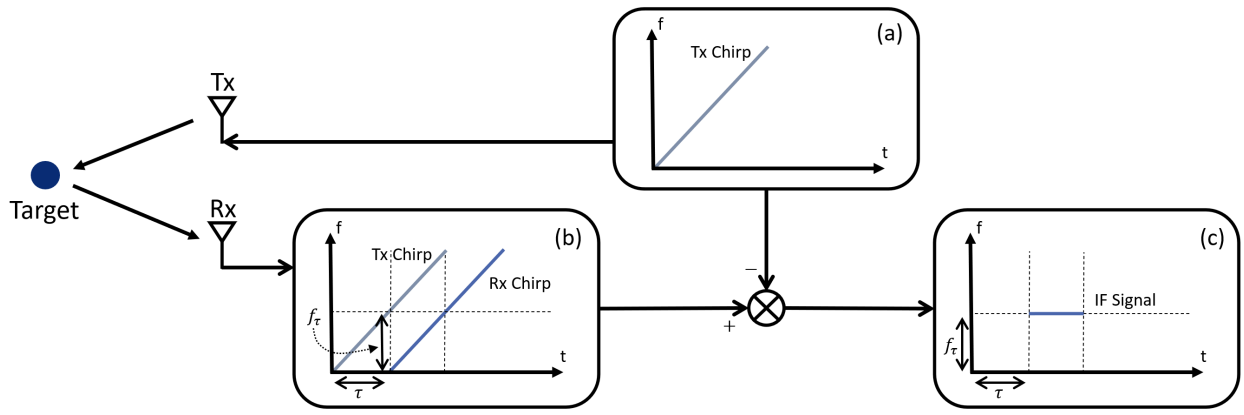


Figure 2.1: FMCW signal chain. (a) FMCW pulse generation. (b) Received signal from single ideal point scatterer. (c) FMCW beat signal after dechirping.

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 whose frequency increases linearly with time as

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

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 easily computed using $B = KT$ [9, 46, 71].

The chirp signal $m(t)$ is transmitted by the transmit antenna, reflected off of the ideal point reflector, and returned to the receive antenna as a scaled and time-delayed version of the transmitted signal. Taking the round-trip amplitude decay into account, 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 τ is the round-trip time delay [72] and the values R_T and R_R (see Fig. 2.1) can be computed by

$$R_T = \sqrt{(x_0 - x_T)^2 + (y_0 - y_T)^2 + (z_0 - Z_0)^2}, \tag{2.3}$$

$$R_R = \sqrt{(x_0 - x_R)^2 + (y_0 - y_R)^2 + (z_0 - Z_0)^2}. \tag{2.4}$$

Therefore, the round trip time delay τ can be computed by

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

where c is the speed of light.

Now, 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 of (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 [46, 39, 51] as

$$\hat{s}(x', y', n_k) = s(x_T, x_R, y_T, y_R, n_k) e^{-j(k_0+\Delta_k n_k)\frac{d_x^2+d_y^2}{4z_0}}, \tag{2.9}$$

valid only for small d_x and d_y, the distances between the transmitter and receiver elements along the x- and y-directions, respectively, where z_0 is a reference plane typically given as the center of the target scene. Taking (x', y', Z_0) as the locatios of the virtual element located at the midpoint between the transceiver pair and R as the corresponding distance from the virtual element to the point reflector, the resulting monostatic beat signal approximates to

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

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

$$R = \sqrt{(x_0 - x')^2 + (y_0 - y')^2 + (z_0 - Z_0)^2} \tag{2.11}$$

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

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

where R is the initial range of the target, v is the velocity of the target, λ_0 is the wavelength corresponding to f_0, and n_c is the chirp index,

Thus, the beat signal sampled across time is a 2-D complex sinusoidal 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, 73].

2.3 FMCW Response to Distributed Target

Assuming a distributed target occupying the volume V in (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 the volume V and R_T and R_R are the radial distances from the target to the transmitter and receiver, respectively, as

$$R_T = \sqrt{(x - x_T)^2 + (y - y_T)^2 + (z - Z_0)^2}, \tag{2.14}$$

$$R_R = \sqrt{(x - x_R)^2 + (y - y_R)^2 + (z - Z_0)^2}. \tag{2.15}$$

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

$$s(x', y', k) = \iiint_V \frac{p(x, y, z)}{R^2} e^{j2kR} dx dy dz, \tag{2.16}$$

where R is the distance between the virtual monostatic element and the target as

$$R = \sqrt{(x - x')^2 + (y - y')^2 + (z - Z_0)^2}. \tag{2.17}$$

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

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

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 the 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) [74], or the concept of synthetic aperture radar (SAR) wherein a small radar platform is scanned throughout space to synthesize a larger array. In this dissertation proposal, 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 scanning geometries are discussed in greater detail in Chapter 4.

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. 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.

3.1 Introduction to 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 [75]. Applications of static gesture classification include augmented/virtual reality (AR/VR) [33], human-computer interaction [76], and even medical applications for range of motion and therapeutic applications [77]. 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 indoors only [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

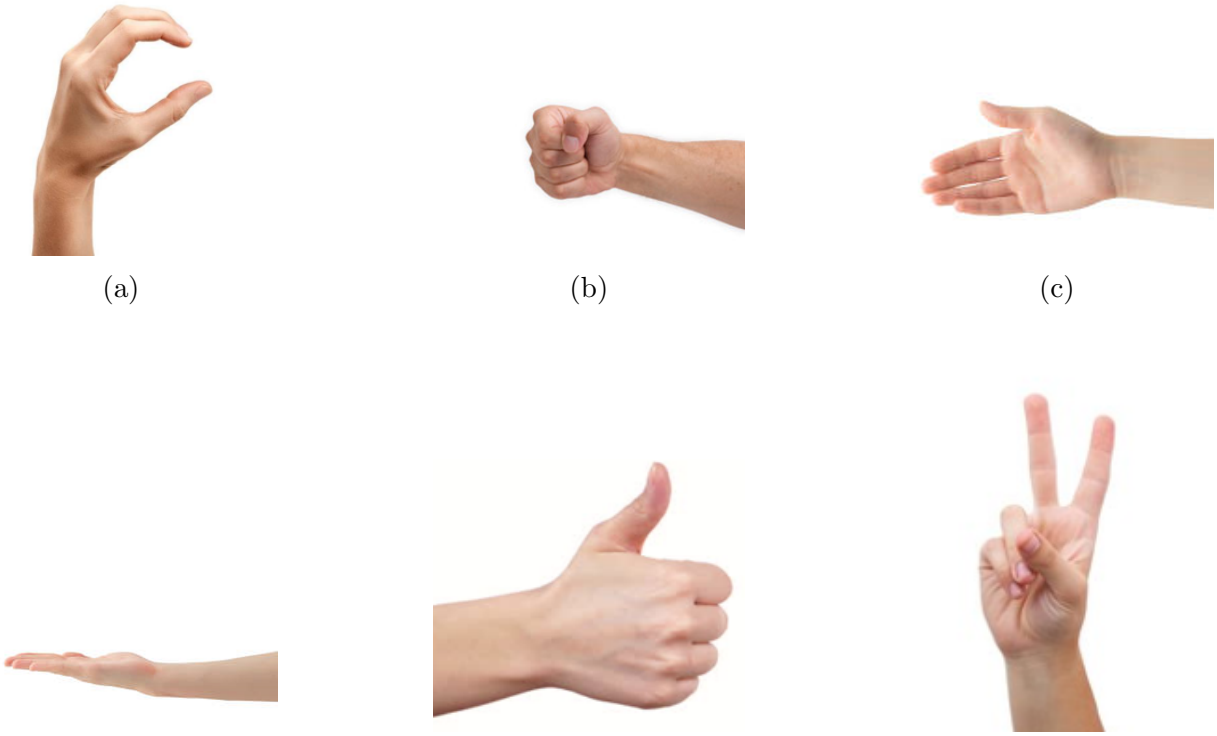
the semi-penetrative nature of the EM radiation at wavelengths in the mmWave frequency range and independence from ambient temperature effects, allowing for fine measurements in non-ideal lighting and temperature environments including occlusion, fog, indoor/outdoor, etc. Additionally, FMCW mmWave radar allows for simultaneous gesture classification and localization. High-resolution spatial information reflected from the human hand is embedded in the FMCW return signal. However, because of the nature of the 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 a 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].

3.2 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 and employ imaging algorithms to recover the reflectivity, as discussed in Chapter 4, we propose classifying a hand gesture using only a single 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 [78]. Fig. 3.1 shows 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 experimentation 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 bore side of the radar. Each gesture was captured with the 9 DoF detailed previously resulting in a total of 2250



(a) (b) (c)

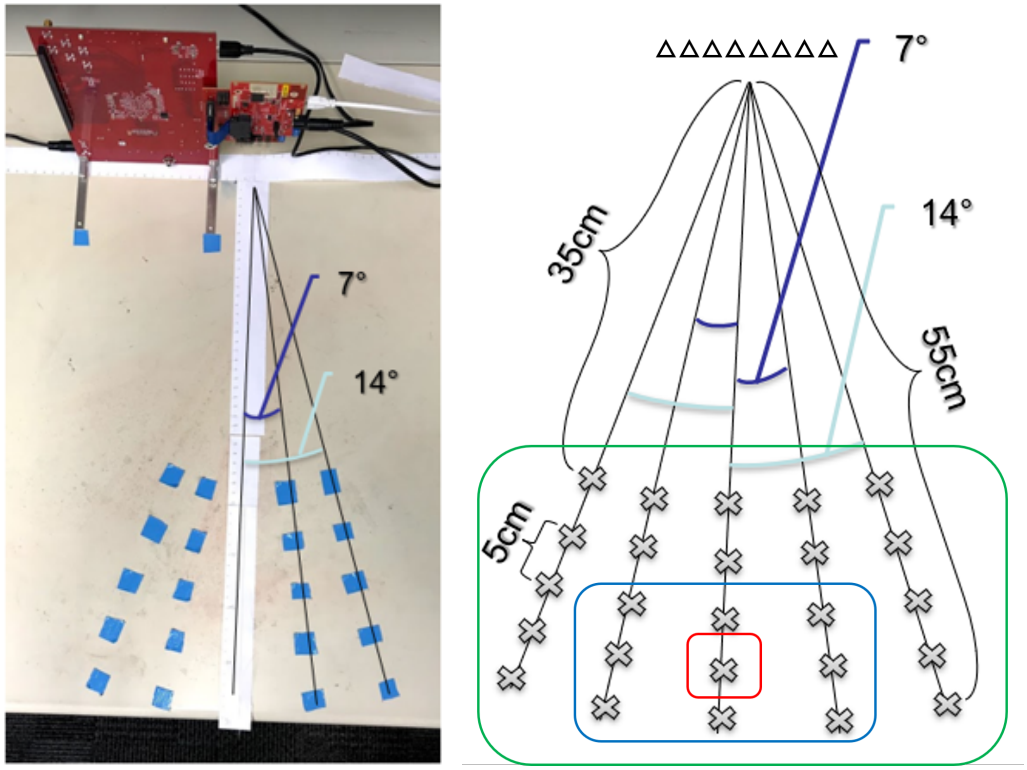
(d) (e) (f)

Figure 3.1: Static hand gestures: (a) “c”, (b) “fist”, (c) “palm”, (d) “perpendicular”, (e) “thumbs up”, (f) “two”.

gestures per class, with six gesture classes, as shown in Fig. 3.1. After preliminary results, which are discussed later, show promising results, a diverse dataset is collected using the setup in Fig. 3.2a.

As indicated by the green box in 3.2b, dataset 2 comprises 25 locations from 35 cm to 55 cm and 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 3.2b, spanning 45 cm to 55 cm and a 14° FOV. A summary of the datasets is provided in Table 3.1

The data are collected using a Texas Instruments (TI) AWR1443BOOST radar with 4 GHz bandwidth from 77 GHz to 81 GHz is mounted to a TI mmWave-Devpack and



(a) (b)

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.

	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°

Table 3.1: Summary of static hand gesture datasets.

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 is created consisting of a row of eight antennas underneath a row of four antennas, as shown in Fig. 3.3. We would also like to note that

although the radar setup is mounted on a desk, the reflections of the desk are negligible, from an empirical study, owing to the narrow beamwidth of the radar along the vertical direction.

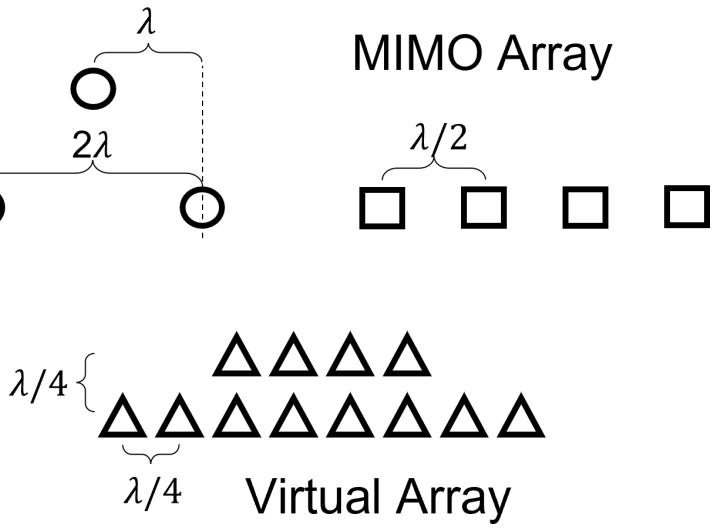


Figure 3.3: TI AWR1443BOOST MIMO antenna array and virtual monostatic array.

After the datasets were collected, preprocessing techniques were applied to investigate the optimal presentation of the data to a CNN. CNNs of varying dimensionality are implemented with three hidden layers consisting of a convolution layer with kernel sizes of 5, 5×5, 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 6 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 64×12, 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 768×1 vectors. Even for a simple set of data in dataset 1, the classification rate is 83%. However, the samples are not presented to the network in a meaningful way, given the underlying mechanics of the problem and format of the data. A range-FFT is performed across the first dimension of

the 64 × 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 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, due to 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.

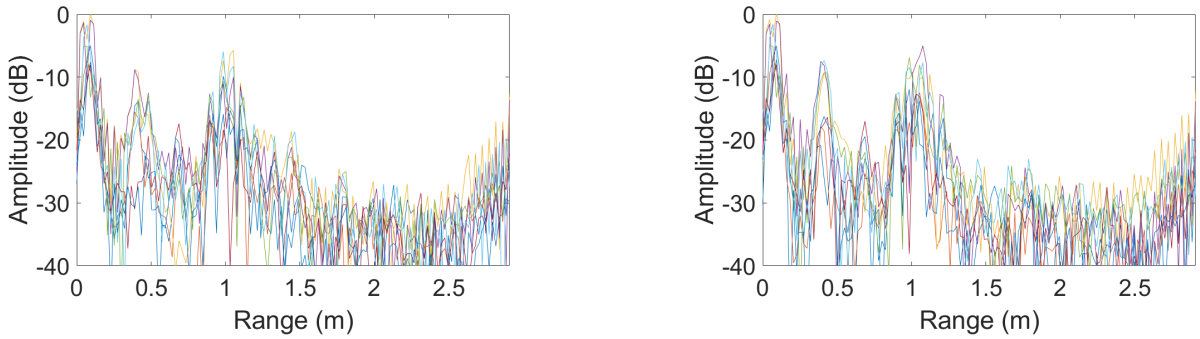


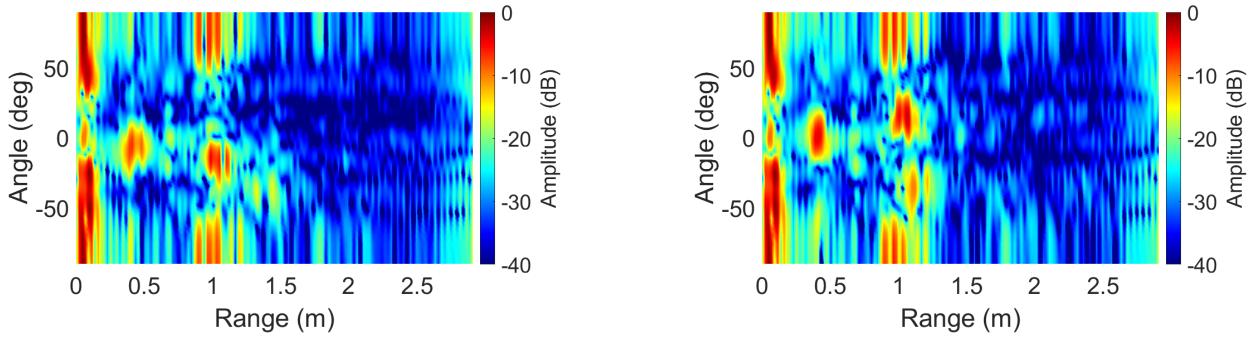
Figure 3.4: (a) Range-FFT of a "c" sample with the hand at 45 cm. (b) Range-FFT of a "fist" sample with the hand at 45 cm.

Next, an autocorrelation strategy is applied to the radar data using the eight colinear channels. The 2-D autocorrelation matrix is computed from the 25 × 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

and zero-padded according to the geometry in Fig. 3.3, with a size of 25 × 8 × 2. An angle-FFT of size 16 is performed across the second dimension yielding a data cube of size 25 × 16 × 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 shown in Fig. 3.4. We again note the reflection of the torso at approximately 1 m and the hand around 45 cm, close to the center of the FOV; however, the reflections have negligible meaning to the naked eye.



(a) (b)

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 projecting 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 in fact 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 the 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.

	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%

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

3.3 Dynamic Gesture Recognition with mmWave Radar

Similarly, a study is conducted on dynamic (moving) gestures to investigate the impact of data presentation on the 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 bore side of the radar, push towards the radar, pull away from the radar, wave at the bore side 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.

The additional dimension of time allows for several new ways of presenting the data. First, the typical range-FFT and range gating are applied to the region of interest in which

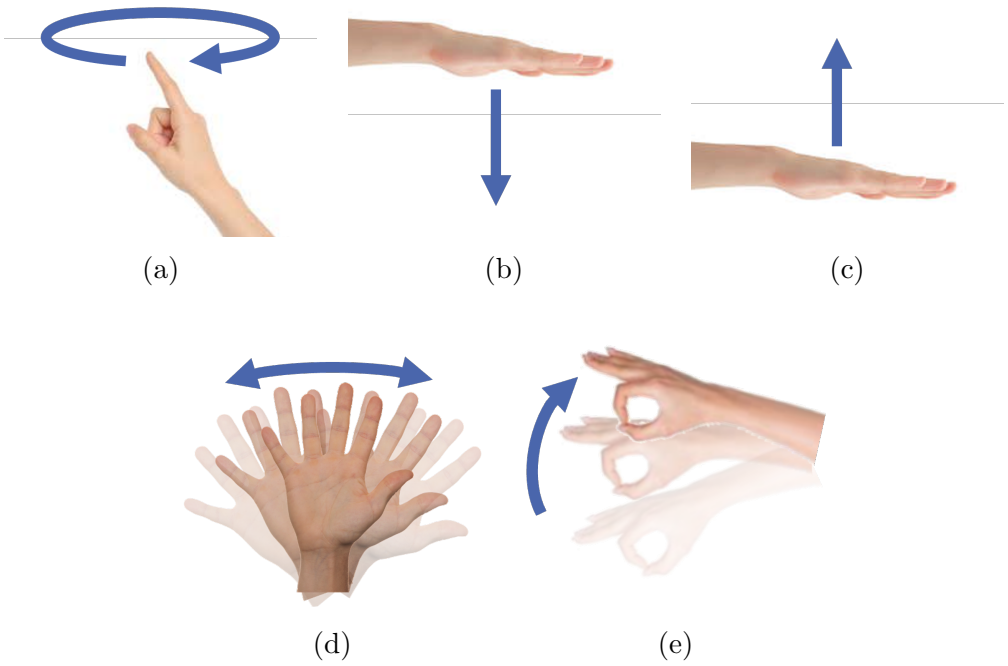


Figure 3.6: Dynamic hand gestures: (a) “circle”, (b) “push”, (c) “pull”, (d) “wave”, (e) “whoosh”.

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 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 [79]. In this sense, the CNN will see 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.

As expected, given the considerable differences among the gestures across 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/only 1 channel. The x8 downsampling is employed to compare the relative classification and computational performance if the gesture is sampled less frequently across time.

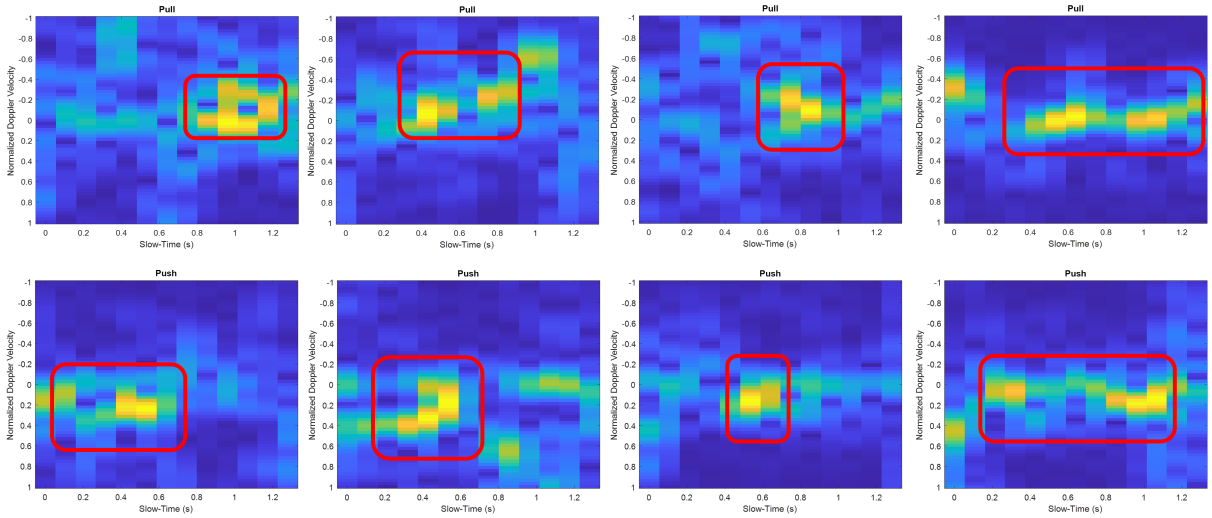


Figure 3.7: Doppler-STFT for pull and push gestures demonstrating where the gesture is performed during the 2.56 s capture.

Because the most notable variation between classes is in the range-time or range-Doppler domains, utilizing only 1 channel, rather than the full 12 channels reduces the dimensionality of the classifier and hence the computational load. A comparison of the classification accuracy and required computation time is provided in Fig. 3.8, where the "Efficiency Score," η, is computed by

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

where α 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 the dynamic gestures.

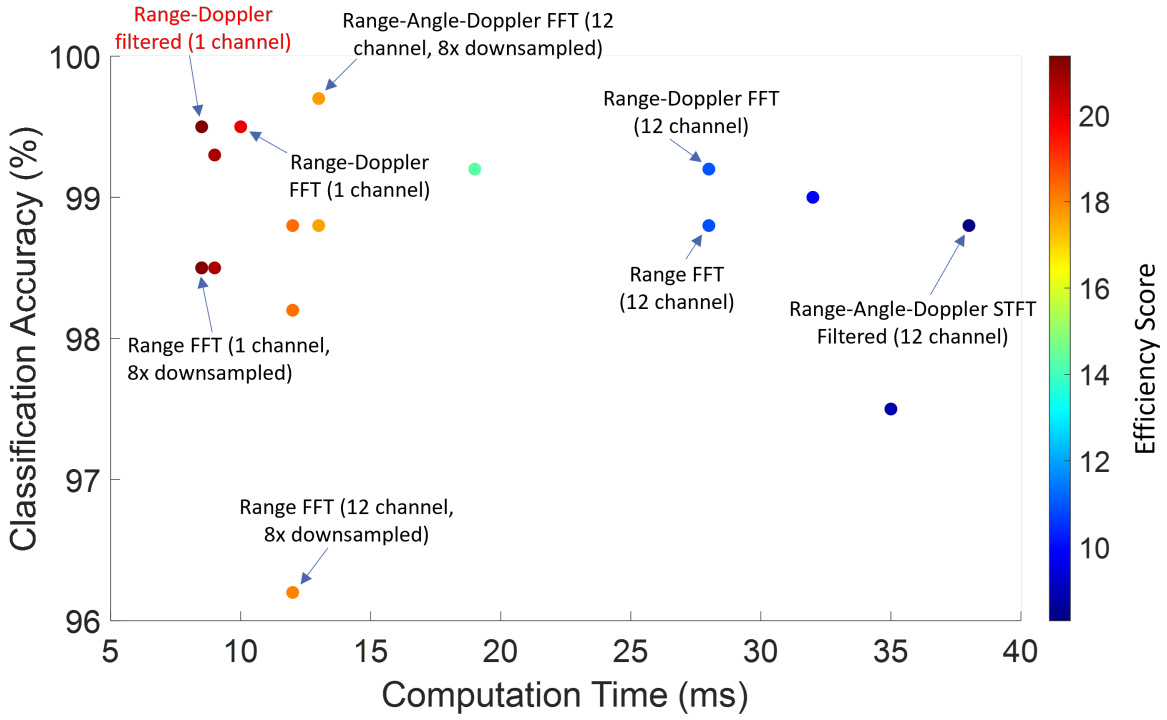


Figure 3.8: Doppler STFT for pull and push gestures demonstrating where the gesture is performed during the 2.56 s capture.

3.4 Improved Static Gesture Classification using Novel Sterile Training Technique

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

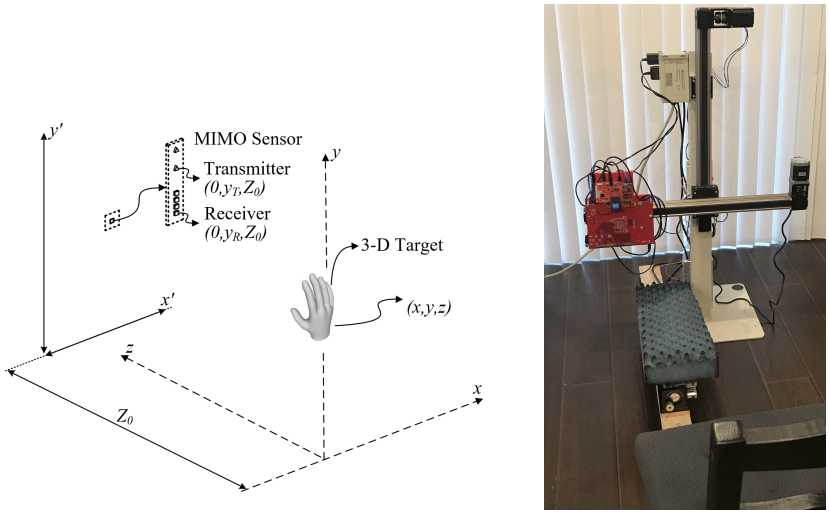
Assuming a simple linear multistatic array along the y-axis, such as the depiction in Fig. 3.9a, after the aforementioned conversion, the return signal from a distributed target can be modeled as the superposition of the echo signals from each of the target coordinates scaled by the target's reflectivity 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

$$R = \sqrt{x^2 + (y - y')^2 + (z - Z_0)^2}. \tag{3.3}$$



(a) (b)

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.

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

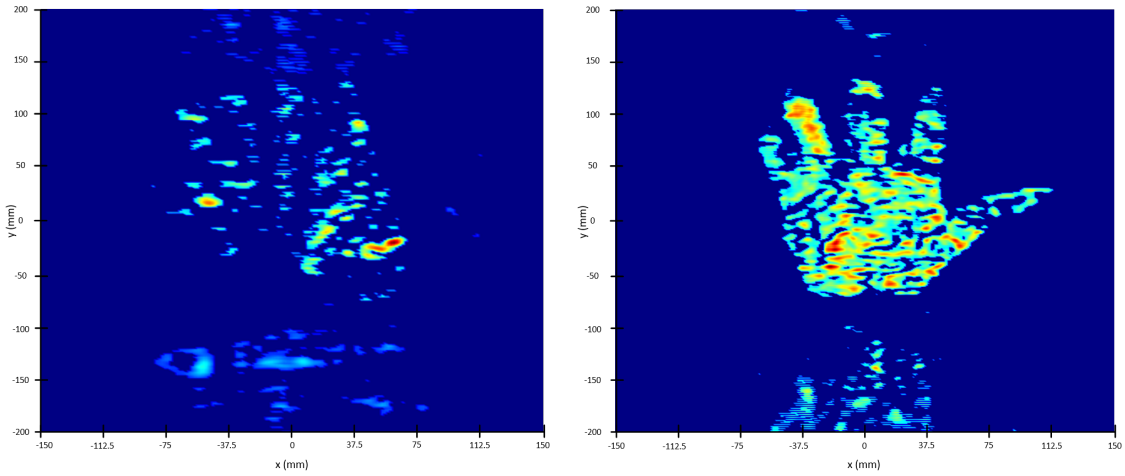
gestures have distinct echo signals unique to that gesture. However, the target scene, or hand, cannot be analytically reconstructed as a three-dimensional (3-D) image and can be used to easily classify gestures using traditional optical image approaches. Thus, classifying static hand gestures involves attempting to learn a high-dimensional pattern (the 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 resulting 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 employing "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 way, while the user remains stationary, many different views of the hand are captured quickly.

As mentioned previously, the RCS of the human hand is problematically small in comparison to the 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 the 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.



(a) (b)

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.

The reconstructed image of the human hand (Fig. 3.10a) shows a poor image of the hand due to the low RCS. 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.

This empirical analysis uncovers the innate difficulty in classifying hand gestures from the radar beat signal. Even by combining 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 quite low magnitude compared to the surroundings, noise, etc. In contrast, from 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 the scanner. To compare against a control, we first train two networks using only the 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%. After training each network with only real human hand data, the range CNN and range angle CNN have 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 range and range-angle profiles from human hands.

Next, two new networks 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. Once trained, the networks 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.

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

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).

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. Likewise, we extend our analysis by examining a similar preprocessing problem in high-resolution imaging

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

With the emergence of fifth-generation (5G) millimeter wave (mmWave) devices, near-field SAR imaging is no longer confined to laboratory environments. Recent advances in positioning technology have attracted significant interest in a diverse set of new applications for mmWave imaging [34, 63]. However, many use cases such as automotive-mounted SAR imaging, unmanned aerial vehicle (UAV) imaging, and freehand imaging with smartphones are constrained to irregular scanning geometries. Whereas traditional near-field SAR imaging systems and quick personnel security (QPS) scanners employ highly precise motion controllers to create ideal synthetic arrays, emerging applications, mentioned previously, inherently cannot achieve such ideal positioning.

Previous work on localization to enable such applications has recently been published employing an inertial measurement unit (IMU) and depth camera sensors, both of which are commonly found in smartphones and automotive applications, for the necessary subwavelength localization required for near-field SAR imaging [34, 63]. In each of these efforts, the subject holds the radar and attempts to move the hand in a raster pattern to synthesize an approximately rectangular planar aperture [34, 62, 68]. However, because of 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 is reconstructed using the generalized back-projection algorithm (BPA).

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. 3.11, where data are collected along different z-planes by a MIMO multistatic radar with respect to a stationary 3-D target. Similar to our previous work [39, 46, 51], we propose to decompose the multiplanar multistatic samples such that the multiplanar multistatic data can

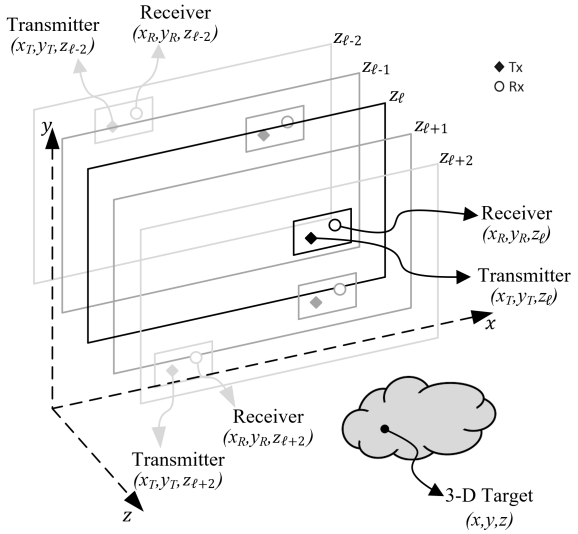


Figure 3.11: Geometry of the multi-planar SAR irregular scanning geometry with a multi-static array.

be projected onto a virtual planar monostatic array for efficient image recovery. Further investigation and results will be presented in the dissertation.

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.

Near-field SAR and ISAR imaging systems are increasing in popularity for numerous applications, from security sensing to industrial packaging. Such systems suffer from distortion due to assumptions and approximations in the image reconstruction process [51]. 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, the distortion around the point targets in Fig. 4.1a is present owing to physical limitations and necessary 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 otherwise infeasible due to cost or size constraints. While some attempts have been made towards image SAR image super-resolution using sparse coding and compressed sensing (CS) techniques [80, 81], data-driven algorithms have yet to be adopted in the image reconstruction process for near-field SAR or MIMO-SAR. Additionally, deep learning-based techniques have been shown to outperform CS algorithms far-field SAR image super-resolution [70]. Naturally, this poses the question: can near-field distortion effects be removed from SAR images?

4.1 High-Fidelity Near-Field SAR Simulation Toolbox

As mentioned previously, a major obstacle to deep learning on mmWave imaging data is availability of meaningful data for model training. In this section, we detail an open-source

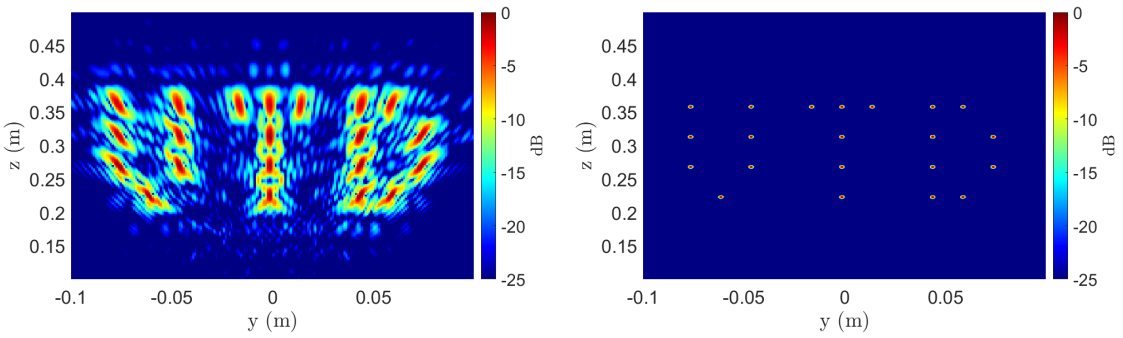


Figure 4.1: 21 ideal point scatters arranged in the letters "UTD." (a) SAR image. (b) Ideal image.

MATLAB toolbox for system prototyping, imaging algorithm development, and high-fidelity dataset generation.

Many array topologies (real or synthetic) have been investigated in the literature for near-field array imaging. The Pacific Northwest National Laboratory (PNNL) is a front-runner in the area of array imaging algorithm development. Sheen *et al.* have 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, 82]. Similarly, the National University of Defense Technology (NUDT), China, has presented extensive work on MIMO and MIMO-SAR imaging algorithms. Their work has included efforts on 3-D imaging using 1-D scanning of a 1-D MIMO array [52, 83], 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, 80, 82, 84, 85, 86, 87, 88, 89], polar formatting algorithm (PFA) for cylindrical scanning geometries [47, 48, 49, 90], and back-projection algorithm (BPA) for any array geometry

[17, 29, 40, 41, 91, 92]. 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, 64, 65, 66, 67, 93] and near-field antenna simulation software [25]. The 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, 64, 65, 93]. On the other hand, despite being quite robust, computational EM solvers require a tedious process of setting up each SAR scenario and can be prohibitively slow, requiring computation time on the order of 10s of hours [25]. Our 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 for enabling rapid high-fidelity dataset generation.

The proposed software interface is functional and near completion. In its current state, the platform provides user control over five key steps in the imaging process. First, the user has full control of the modulation scheme and waveform employed during simulation. Additionally, the antenna gain pattern and multistatic MIMO array are fully customizable to the user's needs. The user can employ a nonideal antenna pattern by selecting between several built-in antenna patterns or importing a custom antenna pattern from HFSS. This allows the user to generate meaningful data that are highly representative of real radar data.

Third, the toolbox is capable of four scanning modes: linear, rectilinear, circular, and cylindrical, although an arbitrary scanning pattern can be specified by the user. After selecting the scanning geometry, any target can be imported from a CAD design or image file for simulation. This enables the simulation of diverse scenarios and can be easily implemented to efficiently generate large datasets of high-fidelity radar data and images of realistic objects. Finally, a host of image reconstruction algorithms is included in the package in addition to a template for custom algorithm implementation and development.

4.2 Towards Near-Field SAR Image Super-Resolution

Using the simulation framework presented previously, we propose a super-resolution technique for near-field SAR imaging, as shown in Fig. 4.2. A neural processor, in the form of a feed-forward FCNN will be trained to perform spatial super-resolution on 3-D near-field SAR images.

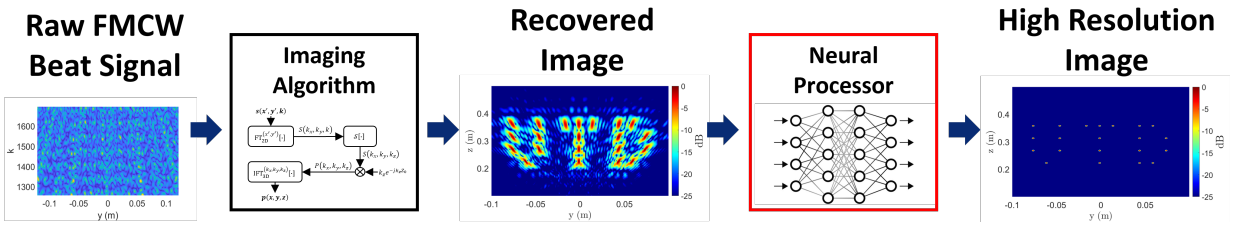
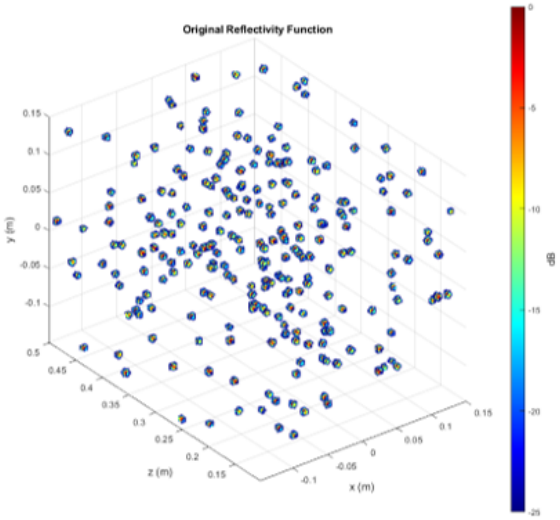
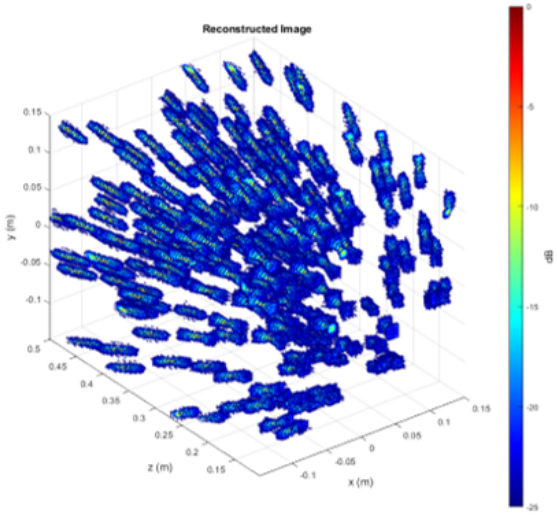


Figure 4.2: Signal chain for proposed near-field SAR image super-resolution technique.

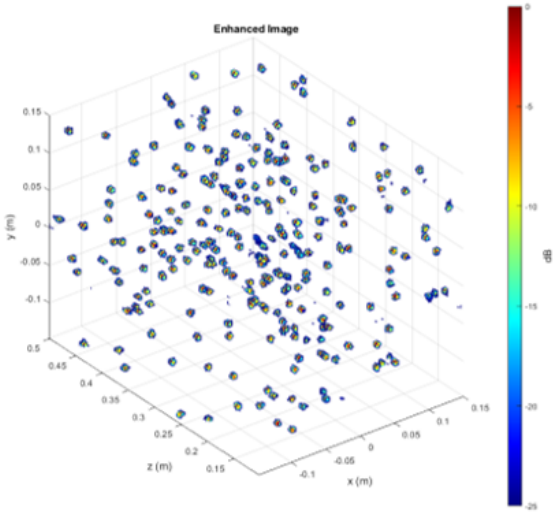
The proposed "enhancement network" will be trained using data from the simulation framework and validated in both simulation and empirical studies. Leveraging a similar approach to [18], we employ residual blocks without batch normalization layers for image enhancement. The training dataset will be comprised of randomly placed point scatterers in space whose number and location vary among samples. Preliminary results are shown here and will be elaborated upon in the dissertation.



(a)



(b)



(c)

Figure 4.3: Randomly generated points to evaluate network performance. (a) Ideal image. (b) Reconstructed SAR image. (c) Enhanced image.

We consider a randomly generated sample similar to those used for training; however, the network has not previously seen this particular SAR image. The raw SAR image in Fig. 4.3b is processed by the neural processor yielding the enhanced image in Fig. 4.3c. As expected, the network considerably reduces the distortion present around each point, even though the distortion appears spatial variant in Fig. 4.3b.

Finally, we consider the point spread function (PSF), which consists of an ideal point scatter located at the bore side of the radar at the position $(0, 0, 0.3 \text{ m})$. The reconstructed image is enhanced by a neural processor and the results are shown in Fig. 4.4. To compare this method with the conventional methods, slices of the PSF along the x-, y-, and z-directions are compared against the range migration algorithm (RMA) and back projection algorithm (BPA), two of the industry standard imaging algorithms, as shown in Fig. 4.4c. The PSF performance of the proposed hybrid-learning technique demonstrates a narrow beamwidth and practically no sidelobes. Notably, along the range (z) direction, the RMA and BPA suffer from a wide beamwidth, hence low resolution; however, the deep learning-based enhanced method shows very fine resolution.

While these preliminary results are promising, several questions remain. A thorough study of the super-resolution capabilities of this approach, its limitations, and performance, in terms of root mean squared error (RMSE) or peak signal-to-noise ratio (PSNR), will be included in the dissertation along with empirical results.

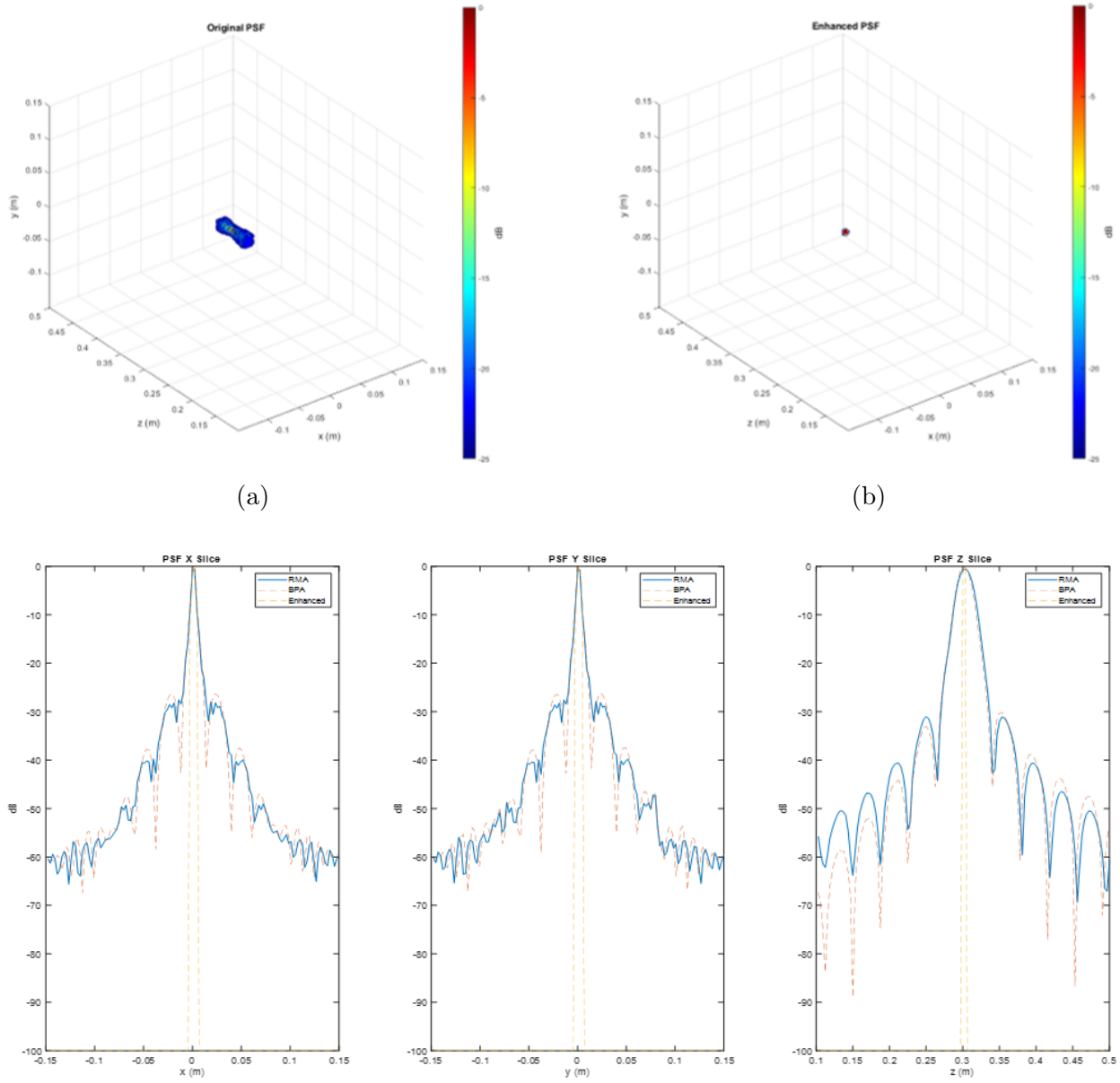


Figure 4.4: Point spread function to evaluate network performance. (a) Reconstructed SAR image. (b) Enhanced image. (c) Comparison of the PSF between range migration algorithm (RMA), back projection algorithm (BPA), and our enhanced algorithm.

CHAPTER 5

INTERLEAVED HYBRID-LEARNING ALGORITHMS FOR MMWAVE PERCEPTION AND IMAGING

In this chapter, we extend the work of 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.

5.1 Hybrid-Learning Techniques for Contactless Musical Instrument Interface

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. While 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 research 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 the tracking robustness by employing the Doppler effect. Compared to prior work on gesture tracking using optical solutions [94, 95], 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.

5.1.1 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 [96, 97], whose localization accuracy is known to be inferior [27].

Further, the aforementioned Doppler principle can be leveraged to extract the velocity of the target by Fourier analysis over successive chirps. The velocity is extracted from the recovered image, which is denoted as $\hat{p}(y, z)$.

5.1.2 Classical Spatiotemporal Feature Extraction Techniques

In this section, we introduce the simple approach to spatiotemporal sensing for contactless musical instrument interface. While our system generally tracks the 2-D position and velocity

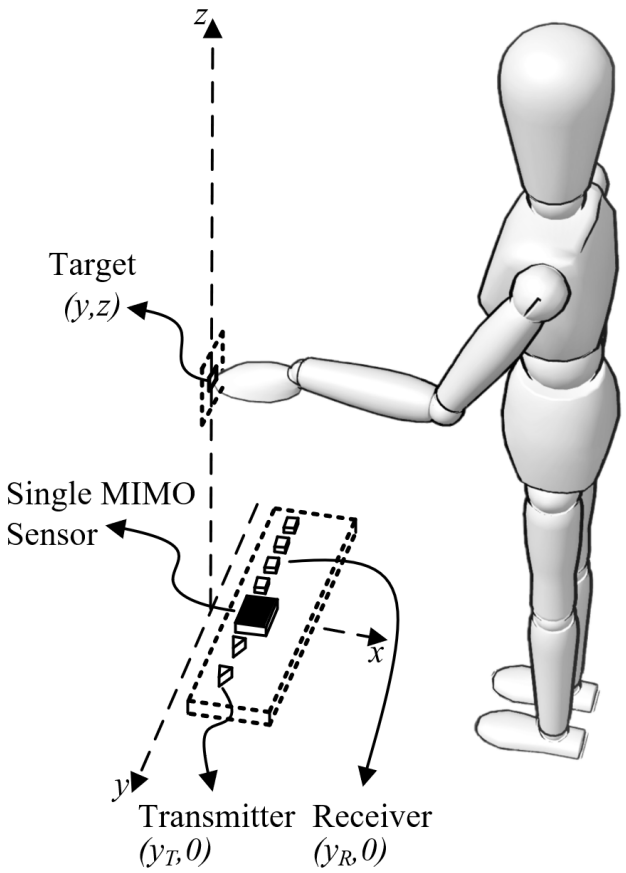


Figure 5.1: The imaging geometry, where the linear MIMO array faces vertically and the musician moves their hand throughout the y-z plane.

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, i.e. the vertical displacement between the sensor and the user's hand. Similarly, cross-range 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 employing an FCNN for super-resolution and image enhancement.

5.1.3 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 [98]. 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 given in 5.2 and additional details can be found in [99].

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. We present a novel FCNN-based technique for image enhancement that improves the 2-D position estimation, subsequent tracking accuracy, and Doppler spectrum SNR. The model is a simplified version of the algorithm proposed in Section 4.2. To train the enhancement FCNN, we construct a dataset consisting of both real human hand data and synthetically generated data using our proposed MATLAB toolbox. During training, the FCNN learns the highly nonlinear relationship between distorted, blurred RMA images and the ideal images.

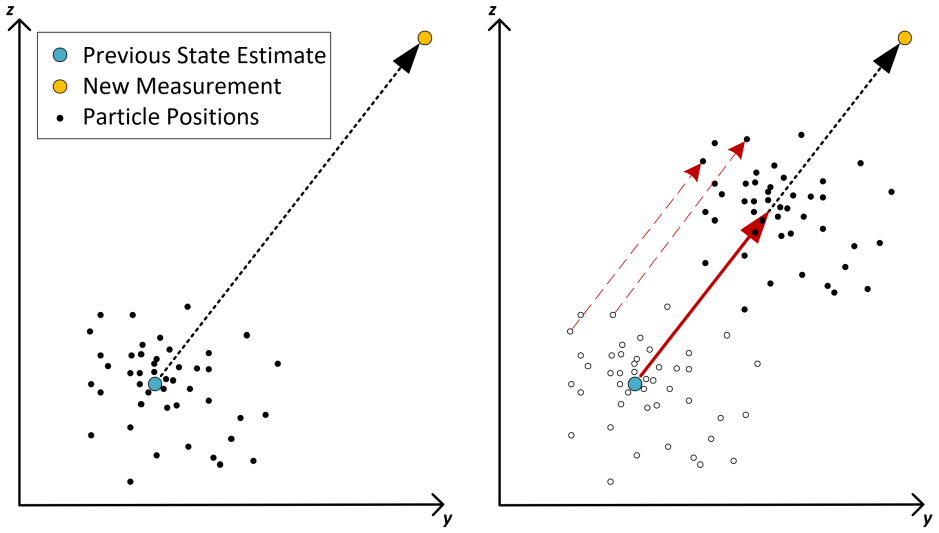


Figure 5.2: A visual example of the modified particle filter algorithm resampling process. The particle locations are resampled by a shift transformation towards the new measurement according to the weight vector $\mathbf{a}$, where $a_y = a_z = 0.5$.

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, clutter and phase 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 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.

5.1.4 Simple and Enhanced Methods Compared

The signal processing chain for the simple feature extraction and tracking method is shown 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

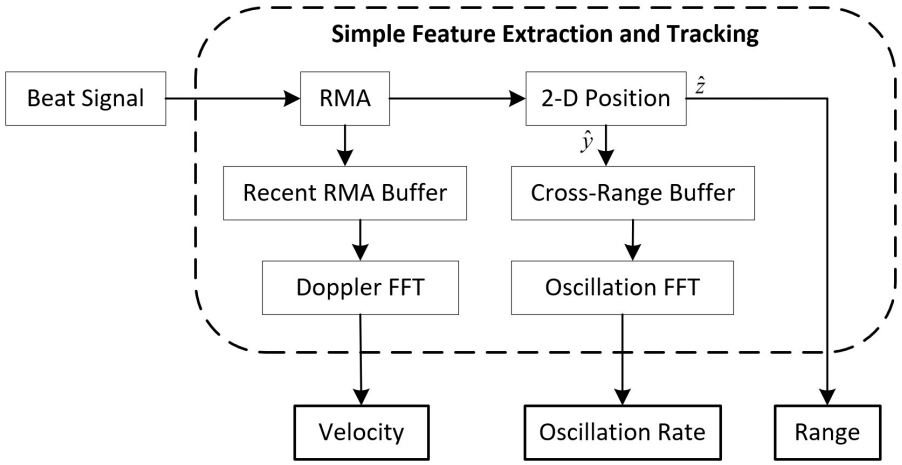


Figure 5.3: Simple tracking signal processing chain. After RMA is performed on the beat signal, features are extracted directly from the raw RMA image.

processing chain for the enhanced feature extraction and tracking method is shown in Fig. 5.4. The spatiotemporal features are output from the algorithm and can be used for many tracking applications.

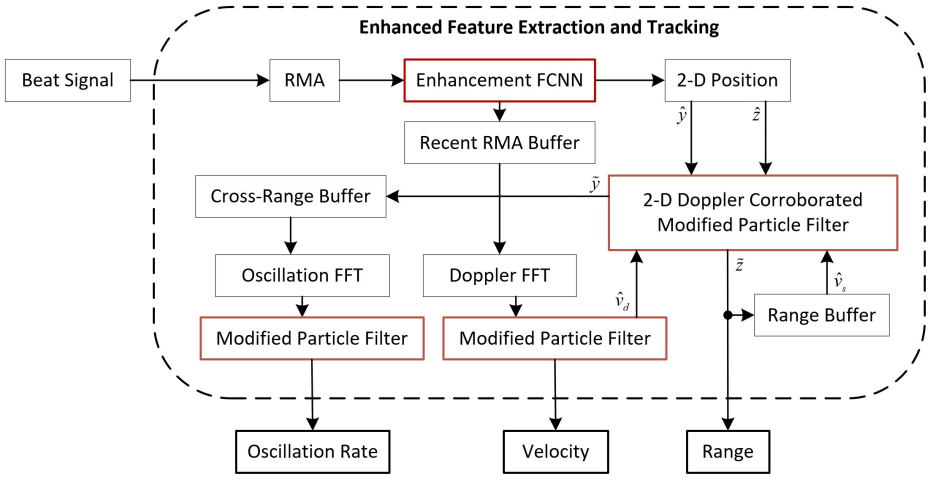


Figure 5.4: Enhanced tracking signal processing chain. Key elements to the enhanced methods are highlighted in red.

To quantitatively compare the tracking performance of the various proposed methods, a total of 4096 unique motion profiles were generated, and the corresponding tracking RMSE was computed for the cross-range, range, and velocity. As shown in Table 5.1, 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 all three features. Comparing PF and DPF, the cross-range and velocity RMSE are identical between the two techniques but the range RMSE is improved due 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, corroborating the findings in Fig. 4.4c. The results in Table 5.1 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 the system noise and ambiguities during the training phase using both real and synthetic data.

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. Additional 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 due to our methods. A thorough analysis and comparison of the performance of the algorithms is given in [99].

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.

Table 5.1: Average RMSE for Tracking Methods

	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

5.2 End-to-End Hybrid-Learning for Near-Field SAR Image Super-Resolution

In this section, we propose a full end-to-end hybrid-learning technique for near-field imaging. Previously, we explored methods for preprocessing data prior to applying a learning algorithm and successfully applied image super-resolution techniques to near-field SAR. Here, however, we propose a novel approach that employs learning algorithms at every step of the image reconstruction pipeline.

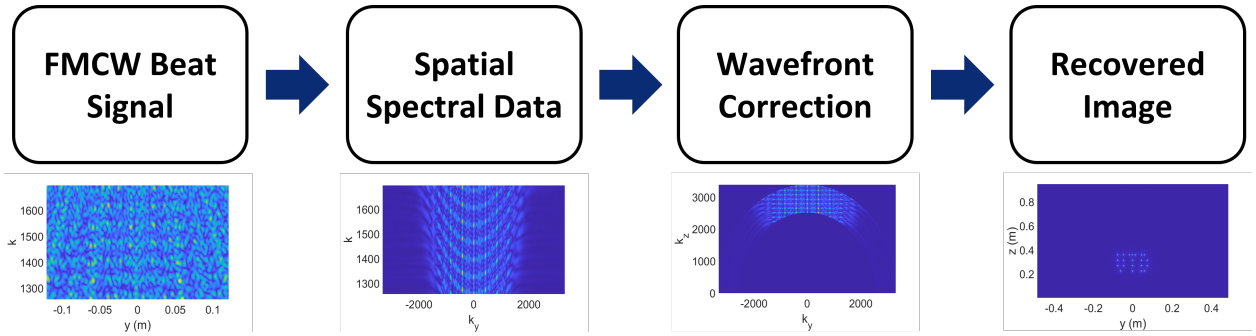


Figure 5.5: Signal chain for range migration algorithm (RMA) applied to linear aperture data.

Fig. 5.5 details the steps of a simple 2-D RMA algorithm applied to a linear aperture and plots the 2-D data at each step. At each stage of the RMA reconstruction algorithm, there is opportunity for leveraging data-driven algorithms to improve image quality. Applying a learning algorithm to the FMCW beat signal could achieve bandwidth extrapolation (the dual of range spectrum super-resolution) or clutter suppression. On the spatial spectral

data, a data-driven approach could perform aperture extrapolation (the dual of cross-range spectrum super-resolution), correct aperture positioning errors common in near-field SAR, or mitigate multistatic effects. Employing a learning method on the wavefront correction step could achieve spectral extrapolation (the dual of image spectrum super-resolution) or improve wavefront correction. And on the recovered image, data-driven algorithms can perform direction image super-resolution or distortion mitigation.

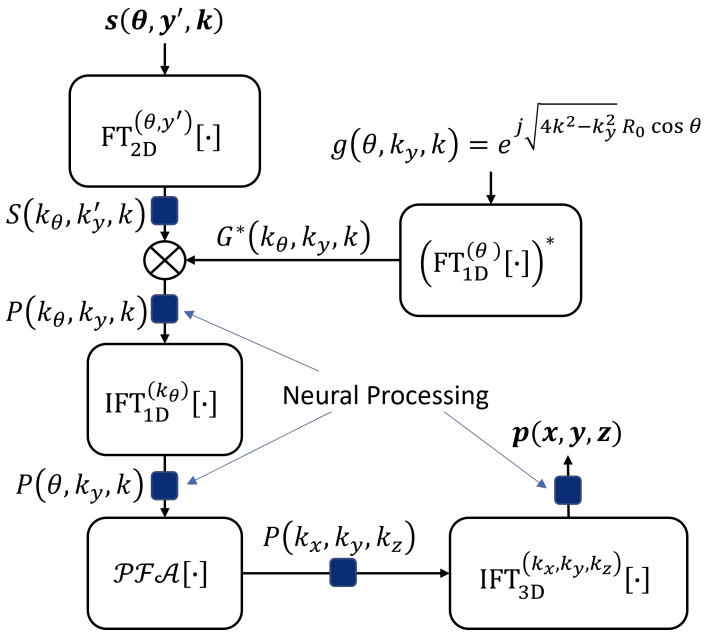


Figure 5.6: Signal chain for range migration algorithm (RMA) applied to linear aperture data.

We propose employing neural processing at each stage of the image reconstruction process, as shown in Fig. 5.6, using a vision transformer-based approach, which has recently emerged as a more robust and less computationally demanding neural processor [54, 100]. In the dissertation, the proposed hybrid-learning imaging algorithm will be applied to multiple imaging problems and will be validated experimentally via empirical studies.

CHAPTER 6

SUMMARY AND PROPOSED WORK

6.1 Summary

In this dissertation proposal, 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 enhanced high-resolution imaging algorithms using data-driven methods. We detailed an efficient high-fidelity dataset generation platform and discussed the preliminary near-field SAR super-resolution results. We presented interleaved hybrid-learning algorithms for perception and imaging. We designed and implemented a real-time deep learning-based super-resolution contactless interface for musical instrumentation. In simulations and experiments, we demonstrated the efficacy of the hybrid-learning approach for improved mmWave perception, sensing, tracking, and imaging.

6.2 Proposed Work

Throughout this dissertation proposal, we presented several opportunities for continued innovation. We are working towards developing our own hardware implementation for smartphone SAR imaging and will complete the imaging algorithm proposed in Section 3.5. We will complete the simulation platform discussed in Section 4.1 and report results for the image super-resolution technique in the dissertation. Finally, we will implement an end-to-end hybrid learning model using the methodology described in Section 5.2 and report our results in the dissertation.

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, 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. IEEE 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 9829 Radar Sensor Techn. XX*, Baltimore, MD, USA, May 2016, pp. 230 – 241.

[9] M. E. Yanik and M. Torlak, "Near-field 2-D SAR imaging by millimeter-wave radar for concealed item detection," in *Proc. IEEE 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. IEEE EUSAR*, 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. IEEE 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. IEEE 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 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. Biomde. 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 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 CVPR*, Honolulu, HI, USA, Jul. 2017, pp. 136–144.

[19] X. Glorot, A. Bordes, and Y. Bengio, "Deep sparse rectifier neural networks," in *Proc. Conf. Artificial Intelligence Statistics*, 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. IEEE MMSP*, 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, 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 10634, Passive and Active Millimeter-Wave Imaging XXI*, Orlando, FL, USA, May 2018.

[26] V. Winkler, "Range doppler detection for automotive FMCW radars," in *Proc. 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*, 2018.

[28] A. B. Baral and M. Torlak, "Joint doppler frequency and direction of arrival estimation for TDM MIMO automotive radars," *IEEE J. Sel. Topics 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 ICIP98*, 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 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. IEEE 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. IEEE EuCAP*, Copenhagen, Denmark, Aug. 2020, pp. 1–5.

[36] T. Kan, G. Xin, L. Xiaowei, and L. Zhongshan, "Implementation of real-time automotive SAR imaging," in *Proc. IEEE 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 subsurface 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 imaging systems," in *Proc. SPIE 10994, Passive and Active Millimeter-Wave Imaging XXII*, Baltimore, MD, USA, May 2019.

[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 structure SAR imaging using time-reversal compressed sensing algorithms based on nonuniform measurement," *IEEE Antennas Wireless Propag. Lett.*, vol. 19, no. 9, pp. 1491–1495, 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. IEEE URSI GASS*, Istanbul, Turkey, Aug. 2011, pp. 1–4.

[45] G. Jia and W. Chang, "Modified back projection reconstruction for circular FMCW SAR," in *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 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," 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. Liang, J. Cao, G. Sun, K. Zhang, L. Van Gool, and R. Timofte, "SwinIR: Image restoration using swin transformer," in *Proc. IEEE ICCV*, Montreal, Canada, Oct. 2021, pp. 1833–1844.

[55] 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 Geosci. Remote Sens. Mag.*, Feb. 2021.

[56] 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.

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

[58] 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.

[59] 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.

[60] 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. Topics Appl. Earth Observ.*, vol. 13, pp. 3109–3124, Jun. 2020.

[61] 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.

[62] 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.

[63] 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.

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

[65] S. Auer, R. Bamler, and P. Reinartz, "RaySAR - 3D SAR simulator: now open source," in *Proc. IEEE IGARSS*, Beijing, China, Jul. 2016, pp. 6730–6733.

[66] R. Kedzierawski, W. Czarnecki, C. Leśnik, and J.-M. Le Caillec, "MATLAB/simulink applications for SAR system design with FPGA," in *Proc. IEEE IRS*, Leipzig, Germany, Sep. 2011, pp. 35–40.

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

[68] 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.

[69] 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.

[70] 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.

[71] 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.*, pp. 1–5, Dec. 2020.

[72] 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.

[73] T. Winkler, "Making motion musical: Gesture mapping strategies for interactive computer music," in *Proc. ICMC*, Banff, AB, Canada, Sep. 1995, pp. 261–264.

[74] 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.

[75] J. Baek, J. Kim, and E. Kim, "Comparison study of different feature classifiers for hand posture classification," in *Proc. IEEE ICCAS*, Gwangu, Korea, Oct. 2013, pp. 683–687.

[76] M. Matilainen, P. Sangi, J. Holappa, and O. Silvén, "Ouhands database for hand detection and pose recognition," in *Proc. IEEE IPTA*, Oulu, Finland, Dec. 2016, pp. 1–5.

[77] A. Anaz, M. Skubic, J. Bridgeman, and D. M. Brogan, "Classification of therapeutic hand poses using convolutional neural networks," in *Proc. IEEE EMBC*, Honolulu, HI, USA, Jul. 2018, pp. 3874–3877.

[78] F. Bernardo, N. Arner, and P. Batchelor, "O soli mio: exploring millimeter wave radar for musical interaction," in *Proc. NIME*, Aalborg, Denmark, May 2017, pp. 283–286.

[79] 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.

[80] 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.

[81] 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.

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

[83] 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.

[84] 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.

[85] 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.

[86] 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.

[87] 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. THz Sci. Technol.*, vol. 3, no. 5, pp. 606–616, Jul. 2013.

[88] A. Batra, M. Wiemeler, D. Göhringer, and T. Kaiser, "Sub-mm resolution 3D SAR imaging at 1.5 THz," in *Proc. IEEE IWMTS*, Essen, Germany, Jul. 2021, pp. 1–5.

[89] 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. Microw.*, vol. 1, no. 2, pp. 574–585, Apr. 2021.

[90] 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.

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

[92] 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. THz Sci. Technol.*, vol. 2, no. 4, pp. 432–440, Jun. 2012.

[93] L. A. Gorham and L. J. Moore, "SAR image formation toolbox for MATLAB," in *Proc. SPIE 7699, Algorithms for Synthetic Aperture Radar Imagery XVII*, Orlando, FL, USA, Apr. 2010.

[94] Y. Sun, X. Liang, H. Fan, M. Imran, and H. Heidari, "Visual hand tracking on depth image using 2-D matched filter," in *Proc. UCET*, Glasgow, UK, Aug. 2019, pp. 1–4.

[95] R. Polfreman, "Multi-modal instrument: towards a platform for comparative controller evaluation," in *Proc. ICMC*, Huddersfield, UK, Jul. 2011, pp. 147–150.

[96] S. Rao, "Intro to mmWave sensing: FMCW radars," June 26, 2017. Accessed on: Apr. 20, 2021. [Online]. Available: https://training.ti.com/node/1139153

[97] H. Kim, S. You, B. J. Jeong, and W. Byun, "Azimuth angle resolution improvement technique with neural network," in *Proc. ICTC*, Jeju, Korea, Oct. 2020, pp. 1384–1387.

[98] 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.

[99] J. W. Smith, O. Furxhi, and M. Torlak, "An FCNN-based super-resolution mmWave radar framework for contactless musical instrument interface," *IEEE Trans. Multimedia*, pp. 1–1, May 2021.

[100] 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," *arXiv preprint arXiv:2103.14030*, Mar. 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 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

January 1, 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.

Voice: (469) 305-9052
Email: josiah.smith@utdallas.edu

Educational History:

B.S.E.E, Electrical Engineering, The University of Texas at Dallas, 2016

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, 2021.
[J2] **J. W. Smith**, O. Furxhi, M. Torlak, "Radar Musical Instrument - A Spatiotemporal Real-Time mmWave Sensor for Contactless Human Computer Interaction," *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*, Nov. 2021, submitted. [C1] **J. W. Smith**, M. E. Yanik and M. Torlak, "Near-Field MIMO-ISAR Millimeter-Wave Imaging," *Proc. IEEE Radar Conf.*, Florence, Italy, 2020, pp. 1-6.

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