SuperRF: Enhanced 3D RF Representation Using Stationary Low-Cost mmWave Radar

Shiwei Fang
Department of Computer Science
University of North Carolina at Chapel Hill
shiwei@cs.unc.edu

Shahriar Nirjon
Department of Computer Science
University of North Carolina at Chapel Hill
nirjon@cs.unc.edu

Abstract

This paper introduces SuperRF– which takes radio frequency (RF) signals from an off-the-shelf, low-cost, 77GHz mmWave radar and produces an enhanced 3D RF representation of a scene. SuperRF is useful in scenarios where camera and other types of sensors do not work, or not allowed due to privacy concerns, or their performance is impacted due to bad lighting conditions and occlusions, or an alternate RF sensing system like synthetic aperture radar (SAR) is too large, inconvenient, and costly. Applications of SuperRF includes navigation and planning of autonomous and semi-autonomous systems, human-robot interactions and social robotics, and elderly and/or patient monitoring in-home healthcare scenarios. We use low-cost, off-the-shelf parts to capture RF signals and to train SuperRF. The novelty of SuperRF lies in its use of deep learning algorithm, followed by a compressed sensing-based iterative algorithm that further enhances the output, to generate a fine-grained 3D representation of an RF scene from its sparse RF representation – which a mmWave radar of the same class cannot achieve without instrumenting the system with large sized multiple antennas or physically moving the antenna over a longer period in time. We demonstrate the feasibility and effectiveness through an in-depth evaluation.

1 Introduction

In recent years, many 3D graphics and vision algorithms have been proposed to model and understand 3D scenes using off-the-shelf cameras and depth sensors. These algorithms have found their uses in robotics, virtual reality (VR), augmented reality (AR), and mixed reality (MR) applications. However, the fundamental limitations of these systems are that they are practically useless when there are occlusions, non-ideal lighting, and difficult environmental conditions such as fog and rain. Radio Frequency (RF) signals, on the other hand, can penetrate certain types of obstacles and are more robust to environmental conditions. Therefore, RF sensing has emerged as an alternative or a complementary solution to 3D imaging and vision—leading to interesting applications such as human activity recognition [42], keystroke detection [9], sign language recognition [23], lip motion recognition [43], localization [44], 3D tracking [7], direction finding [21], range estimation [41], heartbeat detection [8], respiration monitoring [4], emotion detection [52], sleep apnea detection [4], and fall detection [45], multi-person gestures [6], and 2D/3D pose estimation [53, 54].

Typically, RF sensing systems work by modeling the changes in an RF environment when a certain event or an object of interest is present in the scene versus when they are not. These systems are trained to identify and learn the least amount of information from RF signals that are sufficient to distinguish between an event A and an event B, where A and B are significantly different, e.g., running vs. standing. When events of interest are similar, these algorithms fail to model their fine-grained differences due to the lack of enough information in the training data. Placing the RF transceiver close to the objects is a common leeway to handle such cases, but in general, there is a lack of research on creating a rich intermediate representation of an RF scene that can be used to infer fine-grained gesture or to describe a scene with minute detail. Such a representation can be used in a wide range of applications such as monitoring a patient, describing an open or a concealed scene, and context-aware navigation of autonomous systems.

Among RF imaging systems, synthetic aperture radar (SAR) [25] is a well-known technique to generate a reasonably high-resolution representation of a scene. However, this technique requires us to move the radar linearly over a sizable distance to simulate a large (synthetic) aperture. A downside of SAR is the time it takes to move the radar, and thus, they are not robust to capturing scenes that contain moving objects. While we can address this issue by building a gigantic radar, such a system will not be scalable, cost-effective, computationally-efficient, and practical for use in indoor mobile environments.

For emerging applications such as augmenting the sensing capability of an indoor social robot, the RF sensor needs to be small in size to fit on the robot, especially when the size of the robot can be constrained by the application scenarios. The ideal size of the sensor should be such that it should be

International Conference on Embedded Wireless Systems and Networks (EWSN) 2020
17–19 February, Lyon, France
ISBN: 978-0-9949886-4-5

easy to fit on most robots. Smaller sensors also allow us to attach multiple of these on the same robotic platform – which increases the sensing coverage and helps explore more surrounding area. In order to enable an adequate 3D sensing and imaging of an indoor environment, the sensor also needs to have a high-resolution since an indoor environment is often cluttered with many irregularly shaped objects which pose additional challenges to high-quality sensing. We also require that these systems should be able to detect moving objects, such as humans and pets, and be able to generate a representation of the scene in near real-time. Fast sensing and computation are crucial to safety-critical applications where a few ms slower response time can be catastrophic.

Inspired by recent works in computer vision domain that aims at increasing the resolution of images to achieve *super-resolution* [27, 22], we propose a framework and an implementation of it, namely the *SuperRF*, that enables fast sensing, and generates high-resolution 3D representation of a scene using mmWave radars – while being low-cost and small in size. Our two-stage framework enhances the RF measurements from a low-cost mmWave radar to achieve a resolution that is close to a SAR's.

- The first stage of our proposed framework includes a training phase that helps the low-cost mmWave radar system learns how to produce SAR-like imagery from the low-resolution SAR imagery. We collect radar snapshots of different objects by sliding the radar linearly like a SAR system. The collected radar data are synthesized to generate SAR imagery. Using these generated images, we train a specially-designed deep neural network that enhances RF images using only *two* low-resolution RF raw snapshots. Our approach mimics the use of an array of radars working simultaneously, which eliminates the time required to move the radar or the physical size requirement of a large antenna.
- The second stage of our framework employs a compressed sensing-based operation to extract the underlying antenna signal for better estimating the desired high-resolution SAR imagery. This iterative method results in the best possible high-resolution imagery that fits the actual measurements we obtain from the limited number of samples. As this problem is ill-posed, given the limited amount of measurements we have, there are numerous possible solutions (RF images) that fits the actual measurements. By exploiting the sparsity of the signals and applying the compressed sensing technique, SuperRF estimates the most likely and a better quality RF imagery than the one generated by the neural network in the earlier stage.

Using empirical data collected with an off-the-shelf 77 GHz mmWave radar, we demonstrate that SuperRF is able to generate RF representations of objects using only two snapshots to achieve similar-quality imagery produced by SAR operation that uses as high as 64 snapshots (or 64 antennas in the SAR direction). This operation is object-invariant and time-saving as snapshots are taken simultaneously. To the best of our knowledge, SuperRF is the first system that enhances RF sensing to be SAR-like—which is impossible to achieve even with existing high resolution angle of arrival estimation algorithms[36, 21, 29].

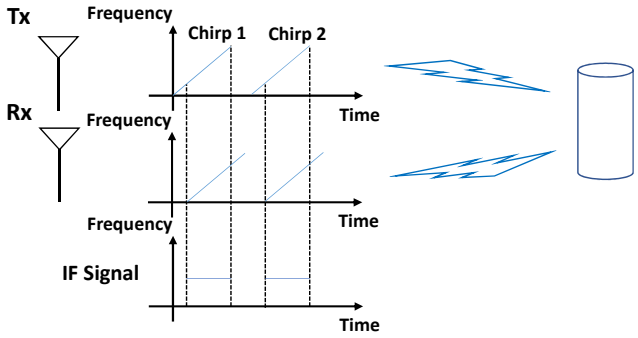


Figure 1: Operational principle of a mmWave radar.

The contribution of this paper are as follows:

- We propose a deep neural network architecture to enhance simultaneously captured two snapshots from mmWave radars into a high-resolution SAR-like imagery which is a low-cost, fast, and software-based alternative to a SAR.
- We propose a second stage of RF signal enhancement that employs a compressed sensing-based method to further improve the neural network generated RF imagery.
- We implement the proposed framework using off-the-shelf low-cost mmWave radar and open-source software. We collect our own data set for training and evaluation.[1]

2 Background

2.1 FMCW mmWave Radar

In recent years, mmWave radars have become popular in the automotive industry and currently they are being used in applications such as advanced cruise control, driver monitoring, and autonomous driving. A mmWave radar is a detection and ranging system that operates in the frequency spectrum between 30GHz and 300GHz. Transmitted signals from these radars have forms including *Continuous Wave* (CW), *Frequency Modulated Continuous Waveform* (FMCW), and *pulsed*. In this paper, we use a commercially available FMCW mmWave radar from which we obtain accurate range information easily and inexpensively [30].

Figure 1 illustrates an FMCW radar in action. The Tx antenna transmits a sequence of *chirps* – a frequency ramp over a short duration. When the signal hits an object, some part of the reflected signal is captured by the Rx antenna which have the same characteristics as the transmitted ones but are delayed in time. As a result, the difference between the transmitted and the received signals of a FMCW radar is a constant frequency – namely the *Intermediate-Frequency* (IF) signal. The IF signal has a linear relation with the distance between the radar and the object. For the radar to distinguish two closely located objects, the objects need to be separated

[1] More info about the dataset can be found at `https://bitbucket.org/embedded_intelligence/superrf_dataset/src/master/`

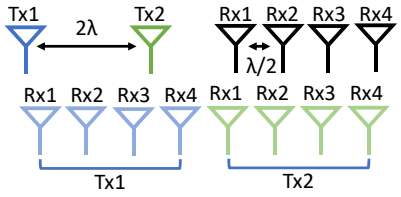


Figure 2: Illustration of a MIMO radar setup.

by an amount larger than the *range resolution*, $d_{res} = \frac{c}{2B}$ of the radar, where c is the speed of light, and B is radar's bandwidth. For a mmWave radar that operates between 77GHz to 81GHz, the bandwidth is 4GHz, which results in a range resolution of around 3.8cm.

To distinguish multiple objects, FFT is applied to the IF signal. For a given number of FFT bins, each bin corresponds to a distance and the index of the bin is called the *range index*. Each bin contains information of all the measurements from that distance which can be considered as an image of a slice in the 3D environment. Multiple such images can be combined to form a 3D representation of the scene, be discussed in Section2.4. The higher the range resolution, the finer the 3D representation will be.

2.2 MIMO Radar

In order to localize or to image an object, a radar needs to calculate the angle of arrival (AoA) of the received signals. This is done by calculating the phase difference across different antennas caused by the difference in traveled distances from the object to the antennas. The *angular resolution* of a radar having N equally spaced antennas is $\theta_{res} = \frac{2}{N}$, where each antenna is separated by a half wavelength, $\lambda/2$.

To increase angular resolution, the most effective way is to increase the number of receiving antennas. However, instead of only increasing the number of receiving antennas, one can also add additional transmitting antennas to create a *multiple input multiple output* (MIMO) radar, one such configuration (the same as the radar we use in this paper) is shown in Figure 2. The top row shows an antenna configuration where two transmitting antennas are separated by 2λ and four receiving antennas are separated by $\lambda/2$. This configuration creates a virtual antenna array, shown in the bottom row of Figure 2, where the number of receiving antennas becomes $2 \times 4 = 8$—which effectively doubles the angular resolution. A more detailed discussion on MIMO radar and multiplexing strategies can be found in [1]. An in-depth discussion on mmWave radar, including how the angle of arrival can be calculated, can be found in [16].

Although increasing the number of antennas increase resolution, such a method requires the radar system to have a complex data processing system to handle a large amount of data. Furthermore, an increased number of transmit antennas means the system requires more time for each transmission antenna to complete the signal generation and transmit sequentially. Hence, large antenna arrays are infeasible for systems that require fast sensing or have limited computational capability.

2.3 SAR Imaging

Instead of using a large physical radar, one can apply *Synthetic Aperture Radar* (SAR) principle to create a virtual, large-aperture radar. SAR works by physically and linearly moving a small-aperture radar. FFT is applied to the received signals across synthesized antennas to determine small differences in the distance where the reflected signal traveled back to the receiving antennas.

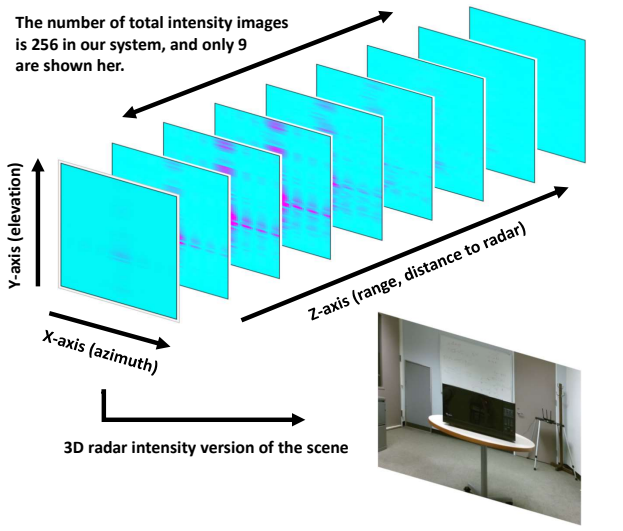


Figure 3: Illustration of 3D representation from intensity images. The intensity images are arranged in order from near to far from the object (microwave). By combining the intensities from all intensity images, a 3D intensity map containing the object (microwave) is constructed.

In the case of a mmWave radar, such as the one we are using in this paper, we can move the radar in a vertical direction to perform SAR operation. By moving the radar in the vertical direction and stopping at 10 positions where consecutive stops are separated by $\lambda/2$, we form a synthetic aperture radar having 10 by 8 antennas. We first apply FFT on each column of the measurements from the synthetic antennas, and then apply FFT on each row of the result from the previous FFT step. By applying the 2D FFT on the measurements, we obtain a 2D matrix where each element corresponds to the signal intensity received by the radar from a specific vertical and horizontal angle. Thus, we obtain the intensity of the corresponding location in the 3D space.

2.4 3D Representation

As each range bin produces an intensity image, a 3D intensity map is produced by combining all the intensity images sequentially. For the intensity image, each axis is the angle defined by the filed of view and FFT bins, for example, the azimuth field of view is 120°, and with 64 FFT bins, each intensity image pixel in the x-axis represents 1.875°. Note that increasing the number of FFT bins does not increase the actual angular resolution of the radar as it is limited by the number of physical antennas. The intensity in the 3D space is calculated by applying trigonometry using the distance (derived from the range bin) and the angle (represented by pixels in the intensity image). An illustration of the 3D representation is shown in Figure 3.

2.5 Compressed Sensing

In digital signal processing, the Nyquist-Shannon sampling theorem dictates that in order to reconstruct a

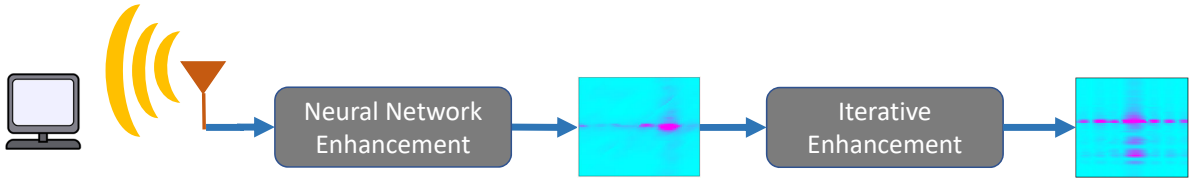


Figure 4: An overview of SuperRF's two-stage signal processing pipeline. The RF snapshot after each stage shows the intensity of RF signals (the X and Y axes represent the horizontal and elevation angles, respectively). Pink (darker color) represents a higher intensity. The (x, y) location of intensity values indicates the angle (horizontal, elevation) where the reflection is coming from.

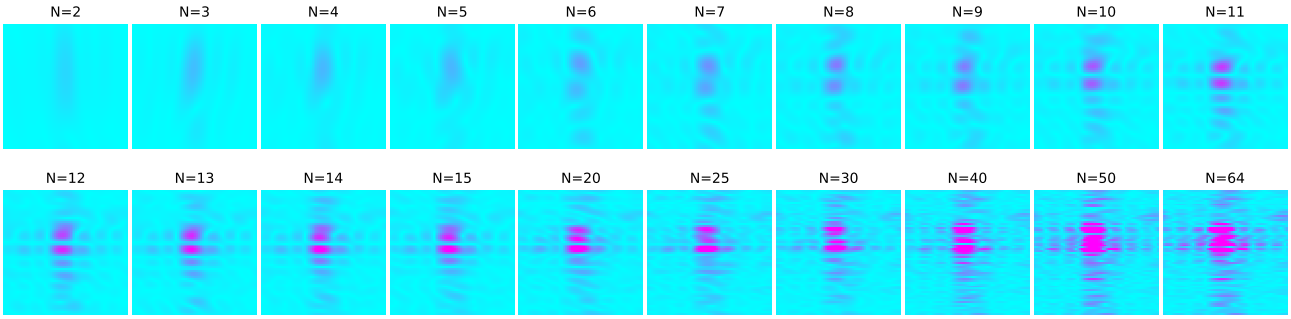


Figure 5: RF signal with different number of snapshots. The x-axis denotes the azimuth angle and the y-axis denotes the elevation angle.

continuous-time signal from its sampled discrete values, the sampling rate has to be at least twice the bandwidth of the signal. However, when we have prior knowledge about the signal, a sampling rate lower than the Nyquist rate can be used to reconstruct the signal. In compressed sensing, the sparsity of a signal is exploited to reconstruct the signal, $\vec{x}$, from a very small number of samples, $\vec{y}$ by solving the underdetermined system of linear equations, $\vec{y} = D\vec{x}$, where the L^1-norm of the signal, $\vec{x}$ is minimized to impose the sparsity constraint.

3 SuperRF Overview

An overview of SuperRF's two-stage RF signal processing pipeline is shown in Figure 4. The goal of SuperRF is to take low-resolution, sparse RF signals from the mmWave radar at the input and to generate a higher-resolution, feature-rich representation of the 3D scene at a near-real-time speed.

Prior to entering the SuperRF's processing pipeline, signals undergo the standard preprocessing step of a mmWave radar. The mmWave radar (hardware) samples the intermediate frequency (IF) signals and performs FFT on the samples. This is often called the *range FFT* since each FFT bin corresponds to a range (i.e., distance) from the radar, and a high value in a bin denotes the presence of one or more objects at that distance. Signals after the range FFT become the input to the SuperRF processing pipeline.

RF signals from each range bin is passed through the first RF enhancement stage of SuperRF– shown as *Neural Network Enhancement* in Figure 4 – where a generative deep neural network (DNN) enhances the resolution of the RF signals. This higher resolution RF signal can be directly used in applications such as object detection and occupancy detection. To enhance the output of the DNN further, the second stage of SuperRF– shown as *Iterative Enhancement* – employs an iterative algorithm that takes a compressed sensing-based approach.

3.1 Stage 1 – Neural Network Enhancement

The goal of the first stage of SuperRF is to increase the resolution of RF sensor data. This step is necessary since the angular resolution of the input signals is generally poor when the number of antennas used in a mmWave radar system is limited. While a large number of antennas can increase the angular resolution, SuperRF's goal is to increase the angular resolution *without* requiring additional antennas beyond what is feasible (space-wise) in the given system.

For a mmWave radar that does not have an elevation antenna, the radar can not differentiate two objects with the same distance to the radar but placed at different heights. Hence, to capture a scene, either the radar has to move along the vertical axis to take and stitch multiple readings – which we call *snapshots*, or we have to use multiple mmWave radars placed on the vertical axis to simultaneously capture multiple snapshots and then stitch them all together to obtain the SAR-like 3D RF representation.

Examples of RF imaging after SAR operation for a different number of snapshots are shown in Figure 5. Consecutive snapshots are taken by moving the radar upwards by $\lambda/2$. We observe that the quality of RF images increases with the number of snapshots, and the improvement in RF imaging quality along the vertical direction is clearly observable.

For a better understanding of the scene, we need high resolution RF image. However, to achieve this, we need either a large number of antennas or SAR operation which means to physically move the radar.

The study above confirms that for an increased resolution of RF images, we either need to run the SAR operation many times by physically moving the radar, or we need to use multiple radars (or, antennas). Neither of these two options is preferred as a large number of antennas increase the radar size and cost, and a SAR operation takes a significant amount of time due to the need to physically move the radar.

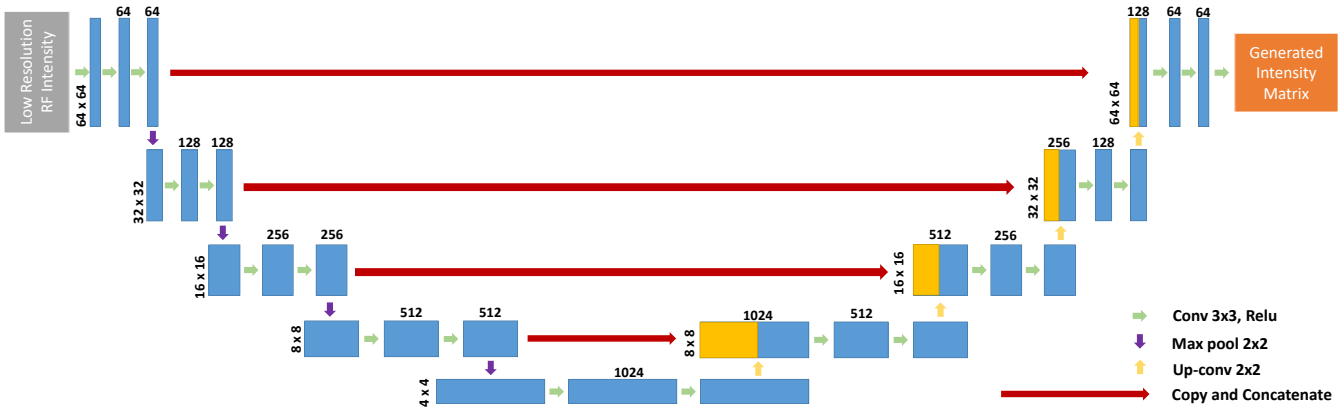


Figure 6: SuperRF's neural network architecture. Each box represent a convolutional feature map. The last 2D convolutional layer has only 1 feature map which generates the enhanced RF intensity matrix. The skip connection model allow global structure being used for later stage reconstruction.

For example, in our experiment, each snapshot takes about two seconds due to the communication delay between the radar and the signal processing board along with the delay due to the mechanical movement of the radar on a slider. For 64 snapshots, the entire process takes nearly two minutes. Even with faster communication and faster mechanical operation of the slider, the expected delay in such a SAR system is typically very long.

In order to generate a high-quality (e.g., similar to 64 snapshots SAR) RF imagery from sparse, low-quality RF input (e.g., only 2 snapshots as input), we employ a generative DNN – which is described in Section 4. The network enables high-resolution RF imaging using a fast, low-cost, and compact radar system, and boosts the RF signal resolution dramatically.

3.2 Stage 2 – Iterative Enhancement

With our deep neural network generated RF images, one can directly compute the 3D RF intensity of the scene. However, no deep learning neural network can generate perfect results even with enough data and training time.

A DNN-based generative approach is several orders of magnitude faster than the traditional SAR approach. However, we observe that the output of the DNN is often a noisy version of the desired high-resolution RF imagery.

To overcome this limitation, we propose to enhance the DNN's output by applying an iterative algorithm that is based on the Griffin-Lim phase recovery algorithm [13] and compressed sensing. We treat the neural network generated results as a noisy model of the true signal and iteratively improves its quality. Given that we have a small number of actual measurements (e.g., 2 snapshots) which is neither enough for compressed sensing nor the RF signal is sparse enough as the radar receives reflected signals from different angels, we optimally enhance the output based on our knowledge of the actual measurements, and then combine them with the information sampled from the inferred values. The details of this method are described in Section 5.

We note that the proposed two-stage RF signal processing framework is application agnostic. The enhanced RF imagery produced by SuperRF can be used in numerous applications such as occupancy detection, obstacle detect, and height estimation that achieves better accuracy and robustness than the state-of-the-art RF sensing systems. In Section 7.4, we provide an evaluation of SuperRF using object recognition as an application.

4 Neural Network Enhancement Algorithm

A typical RF sensor generates coarse signals due to their poor resolution. For high-resolution imaging, one can choose high-frequency RF sensors that operate in the range of over 100 GHz. However, such a high-frequency signal loses its signal power quickly as it propagates through the medium – which makes them unsuitable in many real-world applications that require a reasonable range. Furthermore, due to the propagation loss, high-frequency RF signals normally do not penetrate objects – which makes them less attractive to our intended application scenarios. Hence, in this paper, we use a mmWave radar whose operating frequency is 77 GHz. Algorithms developed in this paper, however, can be applied to other types of radars.

Recall that one can move the radar to collect multiple measurements and then apply SAR operation to obtain a high-resolution RF imagery. However, this operation is both time- and space- expensive. To address this, we propose enhancement of the received signals only from a couple of snapshots to be matched with the ones generated through SAR operation after a large number of snapshots. In other words, we generate 64-snapshot SAR-like RF imagery from only 2 simultaneous snapshots by employing a generative CNN. We focus on improving the resolution in the vertical direction where the resolution is the worst (without any SAR operation). The radar is moved vertically to collect as many as 64 snapshots for training and testing of the CNN.

The CNN learns the information contained in the input signal. We first use the available limited number of snapshots to generate an intensity matrix through FFT operation – which is enhanced by a convolutional neural network (CNN). The CNN is trained with SAR images having a limited number of snapshots as the input and high-resolution RF intensity matrix generated through SAR operation with a large number of snapshots as desired output.

4.1 Analytical Model

We denote the desired intensity matrix generated through SAR operation as I_{SAR}, and the intensity matrix generated by the CNN as I_G. We define the loss function of the CNN as:

$$L = \frac{1}{N}\|I_{SAR} - I_G\|^2 \quad (1)$$

The above loss function represents the mean squared error function where N is the number of intensity pixels. This loss function is used in numerous super-resolution neural network models [51, 22], which minimizes the distance between the generated 2D matrices with the desired ones.

4.2 Network Architecture

The architecture of the proposed deep neural network, which is similar to U-net [35], is shown in Figure 6. We choose an encoder-decoder type network which has been shown to perform well in many super-resolution and denoising problems in computer vision literature [47]. The encoder is a typical convolutional architecture, which consists of 3x3 convolutional filters, followed by ReLU [26] layers. The network is down-sampled with 2x2 max pooling layer with a stride of 2. The feature channels are doubled every time down-sampling happens. The decoder consists of upsampling the feature map with 2x2 up-convolution, which also reduces the feature channel. The 3x3 convolutional filters and ReLU are also employed here. The final layer is a 1x1 convolution which generates the desired high-resolution intensity matrix. We also incorporate skip connection (residual layer) which helps avoid the vanishing gradients problem and reduces the neural network's size. The residual layers are concatenated to the network.

In general, more and complex features can be extracted by deeper convolutional layers. However, a larger neural network causes over-fitting and increases computational complexity. Based on our experiments, we choose five stages with a kernel size of 3x3. The 3x3 sized kernel has been proven to be effective in the literature, and multiple 3x3 convolutional layers together can be used to substitute other kernel sizes (such as 5x5) which has higher computational cost.

5 Iterative Enhancement Algorithm

5.1 Motivation

A neural network is highly dependent on its training and its accuracy beyond what it has seen in the training data is not generally guaranteed. In SuperRF, we observe that while the DNN is fast, its output is often noisy, and depending on the amount of noise, the target application's performance can be significantly degraded. For applications such as robot navigation, in order to ensure that a robot is able to *see* and complement its camera sensors effectively through its RF imaging capability to safely and efficiently navigate the environment, a high degree of accuracy in DNN's output is necessary. However, with limited antenna measurements, the method to further enhance the DNN's output is not straight forward – as one cannot simply make up data out of nowhere. To solve this problem, we take a compressed sensing-based approach in SuperRF. Compressed sensing has recently been proven useful to show that data can be reconstructed with a lower sampling rate as long as the underlying signal is sparse. By treating the number of snapshots as the samples of a compressed sensing problem and by considering the intensity matrix as the intended signal, SuperRF reconstructs a less noisy version of the generated angular intensity representation.

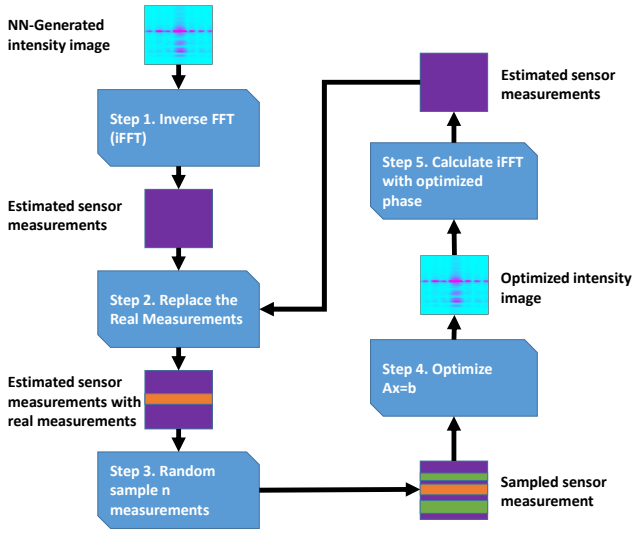


Figure 7: A flowchart of the iterative enhancement step.

5.2 Compressed Sensing

We describe the steps to optimize the intensity matrix generated by the neural network.

Step 1. We estimate the antenna measurements from the neural network generated intensity matrix by performing the inverse Fourier transform (iFFT). The result is an estimation of what the 64-snapshot, full SAR measurement could be.

Step 2. We replace the corresponding antenna measurements with the actual antenna measurements that are available. In our experiment, the middle two snapshots are replaced. This step provides us more accurate antenna measurements.

Step 3. For the newly constructed antenna measurements, we randomly sample n measurements containing the true antenna measurements to form compressed sensing samples. For example, the iFFT step gives us 64 measurements (or, 64 snapshots). We replace the middle two measurements that we have used for the neural network enhancement. The random sampling process takes the two real measurements and samples another $n-2$ measurements from the remaining 62 snapshots to get n synthetic antenna measurements.

Step 4. For compressed sensing, we optimize $Ax = b$, where b denotes the sampled antennas, x denotes the signal we are trying to estimate, corresponding to the intensity matrix. A is the inverse Fourier transform. This step is based on the fact that we have two real measurements. As the real measurements are inserted into the iFFT result, the optimization is performed where the real measurements serve as the constraints.

Step 5. After optimization, we get an estimation of $\hat{x}$ which is a possible explanation of the signals we sampled (note that there is a small number of real measurements while others are estimated from the neural network generated in-

tensity image). Inspired by Griffin-lim algorithm [13], we assume that the amplitude of the neural network generated intensity matrix is correct. Therefore, we replace the amplitude of $\hat{x}$ while maintaining the angle value of it. With the newly constructed $\hat{x}_{new}$, we perform inverse Fourier on it to estimate new antenna measurements.

At this point, we have finished one iteration of our iterative enhancement. In the subsequent iterations, we start from **step 2** to replace the corresponding measurements with the real measurements (the middle two snapshots in our experiments), and then randomly sample another $n-2$ measurements for the remaining 62 snapshots (which could be the same position as the previous iteration but the value will be different due to optimization) for the new optimization step. A flowchart is shown in Figure 7 where we use the intensity image and purple box to represent the intermediate results. The real measurements are shown in orange and the $n-2$ random sampled measurements are shown in green.

With each iteration, we find a possible solution that contains the real antenna measurements. At the end of the user-defined m iterations, we sort the generated $\hat{x}$ by the order of the inverse Fourier transform of $\hat{x}_{new}$ that has the minimal distance to the real measurements. The distance is calculated as the Euclidean distance between the two snapshots from the iFFT of $\hat{x}_{new}$ and the two real measurements. From the top t samples (t is user defined), we find an average of those samples to get an estimate of the true x – which is the desired intensity matrix. The ranking of the top t samples is based on the fact that there are numerous possible explanations during each optimization – some are closer to the real one while some are going to be far away. We use the distance as an indicator to find the most likely ones.

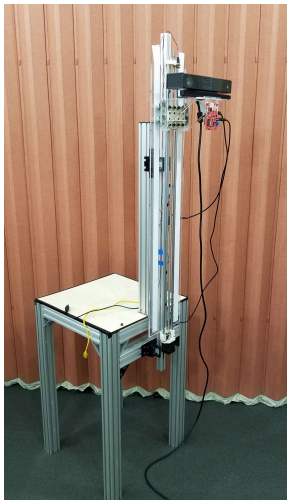
6 Implementation

6.1 Data Collection

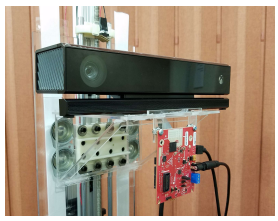
We use an off-the-shelf, low-cost Ti mmWave AWR1443 EVM [2] to collect radar signals and a Kinect V2 sensor to record the ground truth. The setup is shown in Figure 8(a). We build a motorized lead screw linear slide rail that travels at an interval of $\lambda/2$. Both the Kinect and the Ti mmWave radar are attached to a rigid base support as shown in the figure. The distance between the origin of the Kinect's coordinate system and the center of the antennas is 6.4 cm. We develop necessary control software to automatically move the slider, collect the data, and save the data. We include a Kinect sensor to capture high-quality depth images of the scene which is used as the ground truth for scene reference.

The Ti mmWave radar EVM, as shown in Figure 8(c), has 3 TX antennas (2 TX at the same level for azimuth angle estimation and 1 TX for elevation capability) and 4 RX antennas. For SAR operation, we collect data from the mmWave radar with 2 TX and 4 RX, which gives us 8 receiving antennas in MIMO mode[2]. The field of view of the radar along the azimuth and elevation directions are 120° and 30°, respectively. The field of view of the Kinect V2 is 70° and 60°.

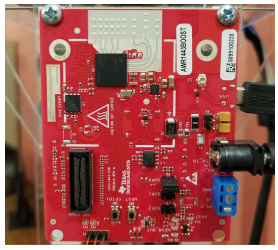
[2]We collect data from all 3 TX and 4 RX antennas, however, in this paper, we only use data from the 2 azimuth TX antennas.



(a) Data collection Setup



(b) Kinect and radar



(c) Ti mmWave EVM

Figure 8: Data collection system: (a) the full setup where the Kinect and Ti mmWave radar EVM are attached to the linear slider; (b) both the Kinect and the mmWave radar are attached to a rigid support; (c) close up of Ti mmWave radar.

During the data collection phase, we save the color image, depth image, point cloud, and camera intrinsic parameters of the Kinect sensor. For the Ti mmWave Radar, we collect raw RF signals from all antennas for each range index and configurations of the antenna. After saving data for a snapshot, the system moves up the linear slider by $\lambda/2$ and collects new data. For the Ti mmWave radar, we modify the out-of-the-box software to expose the full RF data. Due to the bandwidth constraint of the serial communication between the radar and PC, we collect RF data from Doppler bin 0 (static objects). Considering indoor environments, the radar is configured to have approximately 10m maximum working distance.

6.2 Objects and Environments

We use a total of 11 objects in our experiment. These objects are made of different materials such as metal, glass, wood, and fiber. We place these objects at different locations (1-2 meters away from the radar) within the field of view of SuperRF and repeat the experiment in different labs and offices in a typical school building. Some of these objects are shown in Figure 9.

To experiment with the see-through capability of the radar, we collect data in non-line-of-sight situations as well. One such scenario is shown in Figure 10. The mmWave radar sensor and the object are separated by an office wall. Although the Kinect is not able to see the object, the RF signals penetrate the wall and some of the signals are reflected back from the object, and thus we are able to image the non-line-of-sight object. We use different separators such as walls, closed doors, whiteboards, and wood. We note that the system may not see anything in the non-line-of-sight situations when the separators are made of materials such as metal or thick concrete that block RF signals.

6.3 Data Pre-processing

We combine consecutive 64 snapshots to form a 64x8 matrix. The number 8 denotes the number of antennas we have

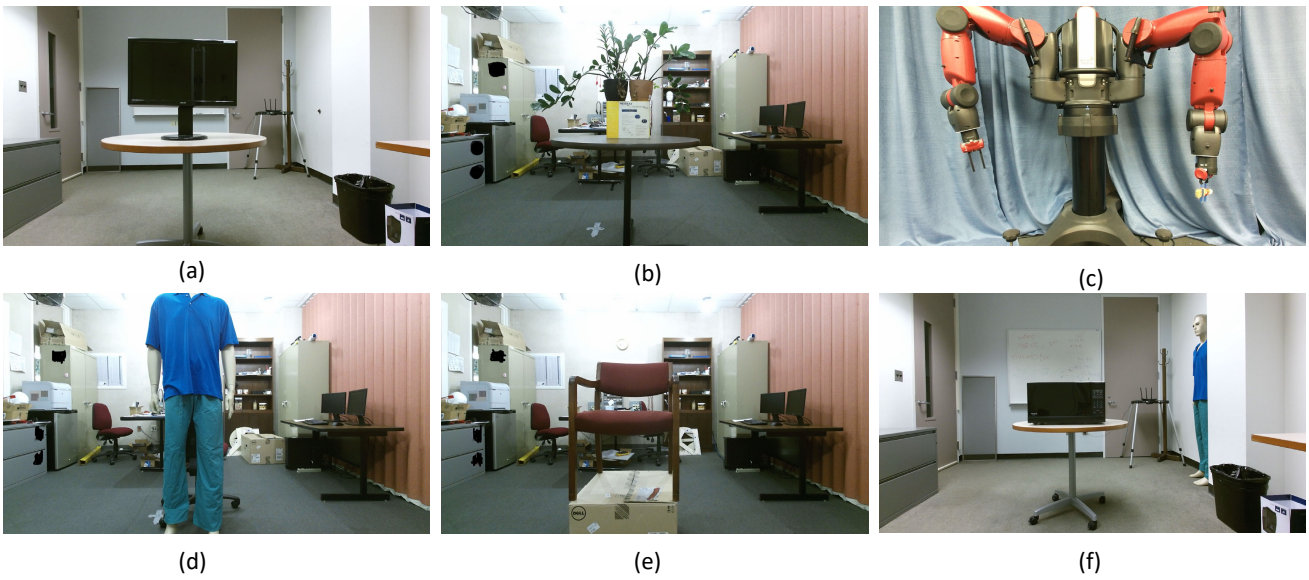


Figure 9: Examples of objects used in the experiment: (a) Computer monitor, (b) plants, (c) robot, (d) dummy, (e) chair, and (f) microwave. Note that the objects are placed in three different indoor environments.

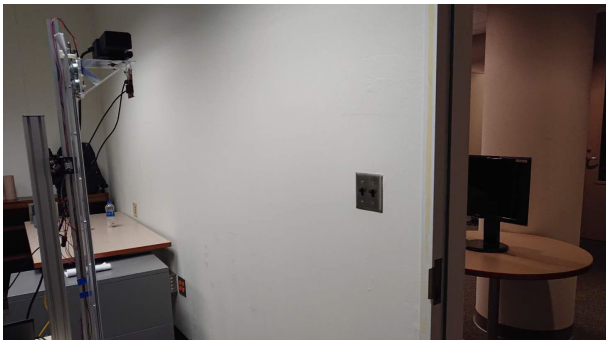


Figure 10: An example of occluded sensing where the radar (on the left) and the monitor (on the right) are separated by the office wall.

in our radar as described in Section 6.1. With 64 bins on each axis, 2D FFT is applied on the antenna data – which results in a 64x64 intensity map where each bin represents an angle of arrival. The 64x64 dimensions denote the azimuth and the elevation for the x-axis and the y-axis, respectively. The generated intensity matrix is used as the desired output of the DNN. The 2D FFT represents the SAR operation.

Out of the 64 snapshots, we choose the middle two snapshots – which is a 2x8 subset of the 64x8 matrix used for the SAR operation. We calculate the 2D FFT of the 2x8 matrix after zero padding to generate 64x64 intensity maps. Zero padding does not increase the angular resolution. The low-resolution intensity matrix becomes the input to the neural network.

An illustration of the data pre-processing is shown in Figure 11. We note that the use of the middle two snapshots does not affect the system's performance. The middle two snapshots are chosen because they cover most of the area that the full 64 snapshots cover and they are easier to extract. One can use any two snapshots as long as they are at the same relative position across all intensity images.

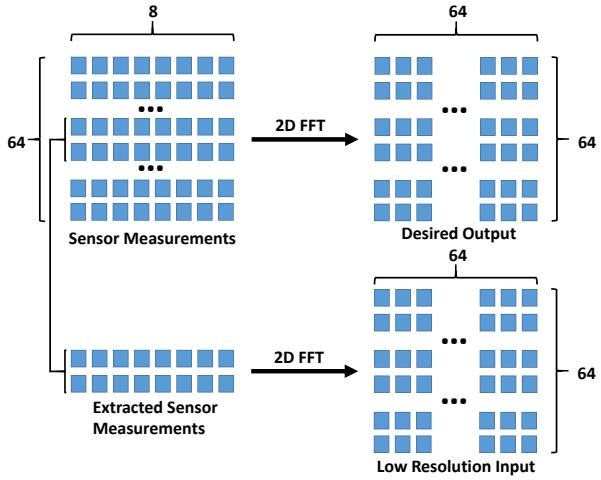


Figure 11: Choosing a subset of the 64 measurements as the input. The *Sensor Measurements* are the 64-snapshot radar measurements (the full SAR operation) and the *Desired Output* is the target output after applying the 2D FFT on the 64 snapshots. We retrieve two snapshots and apply the 2D FFT – which is the input to the neural network.

6.4 Training and Inference

The neural network is trained on a PC having an Nvidia GTX 1080 GPU. We use PyTorch [28] to implement the model. During the training, a batch size of 64 and Adam[20] optimizer are used to optimize the neural network.

During the training, we use only the line-of-sight data and split the data into two subsets: 80% for training and 20% for validation. A total of 2,914,048 intensity images are in the line-of-sight data, of which, 2,331,264 intensity images are used for training. We train for 50 epochs and each epoch takes around 2.3 hours. All training data, which are the same 2,331,264 intensity images, are used in each epoch. The trained model is used for inference in both line-of-sight and non-line-of-sight scenarios.

6.5 Iterative Method

As described in Section 5.2, we use an iteration of $m = 100$ and randomly choose $n = 15$ samples on the synthetic antenna value. This means during each iteration, we randomly sample 13 antenna measurements and combined with the 2 real measurements for optimization. With all the generated samples (here we have 100 samples), we choose the $t = 5$ closest to the real antenna measurements for averaging. With our current non-parallel and non-optimized implementation, the time required to perform SuperRF operation requires around 3 seconds, whereas SAR operation requires over 30 seconds (excluding the time needed for physical movement).

7 Evaluation

7.1 Performance of Neural Network Enhancement

To evaluate the performance of the neural network, we measure the Mean Squared Error (MSE) and the Mean Absolute Error (MAE) of the generated intensity matrix. We report the average error of randomly chosen 15,000 intensity images in Table 1. We observe that the neural network generalizes to both line-of-sight and non-line-of-sight scenarios.

	Mean Squared Error (MSE)	Mean Absolute Error (MAE)
LOS	31589908.8	1079.2
NLOS	55585840.1	1152.7

Table 1: The MSE and MAE for the neural network model in terms of the absolute intensity values.

7.2 Performance of Iterative Enhancement

To evaluate the performance of the iterative method, we measure the MSE and MAE of the neural network only and the end-to-end SuperRF that includes the iterative method respectively. The results are shown in Table 2. We observe that the iterative method greatly improves on MAE while reducing on MSE slightly. This is due to outliers that contribute to large errors as such errors are squared. Lower MAE means that the overall difference between the ground truth and the generated representation is smaller. These results are computed across different types of objects and with different distances between the object and the radar. The ground truth that these results compare to is the intensity images produced by the SAR operation with 64 snapshots.

	Mean Squared Error (MSE)	Mean Absolute Error (MAE)
NN Only	15112451.54	1159.16
SuperRF	15465346.03	1034.23
Percentage	-2.34%	10.78%

Table 2: The MSE and MAE with and without the iterative method. The percentage denotes the performance gain due to the iterative method over neural network-only method.

7.3 Generated Images

We present some examples of the generated images in Figure 12. We observe that SAR operation with only 2 snapshots (column (d)) has very poor resolution, whereas SuperRF generated ones are closer to the ground truth. Note that instead of using 64 snapshots, we only utilized 2 snapshots in SuperRF.

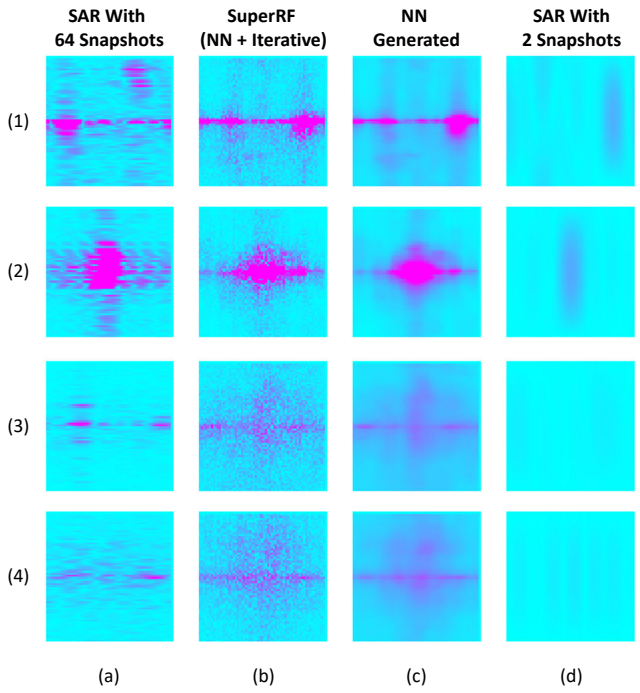


Figure 12: Examples of generated intensity matrices. Each row detnoes one example. Column (a) is the SAR operation with 64 snapshots, column (b) is obtained by the full SuperRF algorithm, column (c) shows the intensity matrices directly from the neural network, and column (d) shows the results of SAR operation with only two snapshots.

7.4 Evaluation of Application Scenario

To better understand the advantage of SuperRF, we train a neural network to perform object recognition on the generated voxelized (3D pixel) representations. The 3D voxel is generated in two ways, one is directly converting intensity matrices to occupancy grid representation through a threshold, another is we trained a separate neural network to perform the conversion.

We divide the objects in our collection into training and testing sets and train the same neural network, which is a standard 3D convolutional neural network (one can use a more advanced neural network such as [14] for better performance), multiple times and measure the object recognition accuracy for each type of data. For the training and evaluation, we use only the area where the objects are placed (thus, no background information is used). The result is shown in Figure 13. *Depth* denotes the case where the depth sensor data is used for training and testing. As expected, it shows the highest accuracy. *AI voxel* denotes the voxelized representation generated a neural network with intensity matrices produced by SuperRF. Its accuracy is very close to what depth sensor achieves. The *AI RF* is the non-voxelized RF representation generated by SuperRF. The *SAR* denotes the neural network voxelized RF representation generated through SAR operation using only 2 snapshots. The *AI RF T* and *SAR T* means to convert the SuperRF generated intensity matrices and SAR generated intensity matrices to the

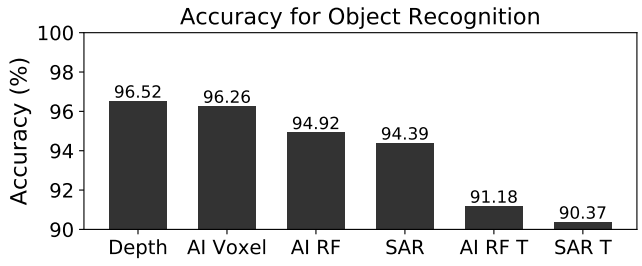


Figure 13: Object recognition accuracy for different types of data. *Depth* is the object recognition with Kinect depth image, others are SuperRF generated or through SAR operation. See Section 7.4 for detail.

occupancy grid respectively.

In this experiment, the SuperRF performance is evaluated in comparison with Kinect and SAR operation. For different comparisons, such as *AI Voxel* vs. *SAR* and *AI RF T* vs. *SAR T*, SuperRF increases the performance of other applications by generating higher quality representations.

8 Related Work

8.1 RF Based Sensing

RF sensing has been employed for different purposes, such as localization [21], vital sign monitoring [8], and gesture recognition [34]. They mainly focus on inferring the desired information for a specific target through RF signal instead of sensing the entire environment. Some works explored capture the human body through wall [5] and estimate human pose [53, 54]. They focus on the human body only and require movements of the subject. Thus cannot be applied to static scenes.

In [15], the author explored the possibility of imaging using WiFi. The result is a low-resolution heatmap and requires the Tx and Rx to be in a different place. 3D voxel generation using WiFi has also explored in [19, 10]. They use two WiFi devices to measure the signal attenuation across the interested area and infer the 3D voxel representation of the scene. Their work requires time-consuming measurements, pre-planned paths and separated Tx and Rx antennas which is undesired. [56] proposed an imaging system with 60GHz transceivers, however, the Tx and Rx antennas are separated and only extract 1D information. They improved the system with [55] to perform localization and object recognition using a single networking device. However, this system requires moving the sensor to generate a 2D map, whereas our system generates 3D representation and does not require movement.

There are numerous work on FMCW imaging, such as [3, 32, 30, 40]. These works focus on sensor design and signal processing where the quality of the measurements is limited by the physic sensor. [31, 40] also investigated SAR operation [25] with FMCW radar to improve imaging resolution. Their SAR operation system requires mechanic components to move the radar which is time-consuming and the subject needs to be still. Our framework generates measurements closer to ones generated through SAR operation but without the need for the time-consuming mechanic operation.

8.2 Other Sensing Techniques

For 3D representation generation, the most straightforward sensor would be depth sensor, and Kinect is one of such sensors that's popular among research communities [17]. Camera is another popular choice for 3D reconstruction [11, 33] and recently LIDAR for 3D sensing [37]. However, these types of sensors suffer limitations from occlusion, non-idea lighting, and difficult environmental conditions such as fog and rain. Our system can perform better than these sensors in such scenarios.

There are some other types of sensors that can do imaging, such as acoustic [24, 18, 38]. While acoustic sensing does not affect by lighting issues, it can be affected by a noisy environment. To achieve high resolution, this type of sensors either utilize large microphone array [18, 38] or SAR operation [24]. In [24], the author demonstrated some see-through capability but the resolution drops when the distance between the target and the sensor becomes larger than 60 cm. Where in our system, the RF sensor has a better penetration capability and can provide better resolution in the farther distance.

8.3 Image Enhancement

In the first stage of our framework, we enhance the RF signal to be closer to the quality of the ones generated through SAR operation. This is inspired by the recent achievements in image super-resolution [27, 48, 46, 22]. [46] uses a deep convolutional neural network to extract features then increases the resolution. [22] utilized GAN [12] to increase the resolution of the picture. The intensity matrix that we wanted to generate in our work is similar to images in which both are 2D matrices. Different from existing super-resolution work, the intensity matrix does not have a bound on the values, and there is no low-resolution image exists to increase resolution. We used a deep convolutional neural network model but there is no upsampling within the network as image super-resolution normally does.

8.4 Point Cloud Upsampling

Another area of work that can be related to ours is the point cloud upsampling [50, 49]. Their work focus on increasing point cloud density which can reduce sparseness and irregularity. Even though the upsampling problem is related to image super-resolution, the lack of special order and regular structure makes the problem hard. They choose a patch-based method to upsampling the point cloud. In our work, we are solving the problem of low resolution and noisy RF sensing enhancement. Where RF signal is too coarse to generate a meaningful point cloud. People also explored the possibility of generating point cloud based on conditional distribution [39]. However, such work cannot be directly applied to ours as they are generating point cloud based on determined category and existing points. For RF signal, there is no such information available.

9 Conclusion

In this paper, we explore the problem of enhancing mmWave radar signals to a higher resolution using a software-based solution. To this end, we propose a two-stage framework. The first stage generates an intensity matrix from only two RF snapshots using a neural network. The

second stage further enhances the RF measurements by using a compressed sensing-based iterative method. From this work, we achieved super-resolution on mmWave radar signal for imaging. The enhanced intensity images can improve performance on different applications such as occupancy detection and object recognition. To the best of our knowledge, this is the first of its kind system that enhances mmWave signals to high-quality intensity matrices to represent a scene.

10 References

[1] http://www.ti.com/lit/an/swra554a/swra554a.pdf. [Online; accessed 30-Nov-2018].

[2] http://www.ti.com/product/awr1443. [Online; accessed 24-Aug-2018].

[3] A 100 ghz fmcw mimo radar system for 3d image reconstruction. In *Radar Conference (EuRAD), 2016 European*, pages 37–40. IEEE, 2016.

[4] H. Abdelnasser, K. A. Harras, and M. Youssef. Ubibreathe: A ubiquitous non-invasive wifi-based breathing estimator. In *Proceedings of the 16th ACM International Symposium on Mobile Ad Hoc Networking and Computing*, MobiHoc '15, pages 277–286, New York, NY, USA, 2015. ACM.

[5] F. Adib, C.-Y. Hsu, H. Mao, D. Katabi, and F. Durand. Capturing the human figure through a wall. *ACM Transactions on Graphics (TOG)*, 34(6):219, 2015.

[6] F. Adib, Z. Kabelac, and D. Katabi. Multi-person localization via rf body reflections. In *NSDI*, pages 279–292, 2015.

[7] F. Adib, Z. Kabelac, D. Katabi, and R. C. Miller. 3d tracking via body radio reflections.

[8] F. Adib, H. Mao, Z. Kabelac, D. Katabi, and R. C. Miller. Smart homes that monitor breathing and heart rate. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, pages 837–846, New York, NY, USA, 2015. ACM.

[9] K. Ali, A. X. Liu, W. Wang, and M. Shahzad. Keystroke recognition using wifi signals. In *Proceedings of the 21st Annual International Conference on Mobile Computing and Networking*, MobiCom '15, pages 90–102, New York, NY, USA, 2015. ACM.

[10] S. Depatla, C. R. Karanam, and Y. Mostofi. Robotic through-wall imaging. *IEEE AntEnnAs & ProPAgAtIon MAgAzInE*, 1045(9243/17):2, 2017.

[11] H. Fan, H. Su, and L. J. Guibas. A point set generation network for 3d object reconstruction from a single image. In *CVPR*, volume 2, page 6, 2017.

[12] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial nets. In *Advances in neural information processing systems*, pages 2672–2680, 2014.

[13] D. Griffin and J. Lim. Signal estimation from modified short-time fourier transform. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 32(2):236–243, 1984.

[14] H. Hu, J. Gu, Z. Zhang, J. Dai, and Y. Wei. Relation networks for object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3588–3597, 2018.

[15] D. Huang, R. Nandakumar, and S. Gollakota. Feasibility and limits of wi-fi imaging. In *Proceedings of the 12th ACM Conference on Embedded Network Sensor Systems*, pages 266–279. ACM, 2014.

[16] C. Iovescu and S. Rao. The fundamentals of millimeter wave sensors.

[17] S. Izadi, D. Kim, O. Hilliges, D. Molyneaux, R. Newcombe, P. Kohli, J. Shotton, S. Hodges, D. Freeman, A. Davison, et al. Kinectfusion: real-time 3d reconstruction and interaction using a moving depth camera. In *Proceedings of the 24th annual ACM symposium on User interface software and technology*, pages 559–568. ACM, 2011.

[18] A. Izquierdo, J. J. Villacorta, L. del Val Puente, and L. Suárez. Design and evaluation of a scalable and reconfigurable multi-platform system for acoustic imaging. *Sensors*, 16(10):1671, 2016.

[19] C. R. Karanam and Y. Mostofi. 3d through-wall imaging with unmanned aerial vehicles using wifi. In *Proceedings of the 16th ACM/IEEE International Conference on Information Processing in Sensor Networks*, pages 131–142. ACM, 2017.

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

[21] M. Kotaru, K. Joshi, D. Bharadia, and S. Katti. Spotfi: Decimeter level localization using wifi. *SIGCOMM Comput. Commun. Rev.*, 45(4):269–282, Aug. 2015.

[22] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. P. Aitken, A. Tejani, J. Totz, Z. Wang, et al. Photo-realistic single image super-resolution using a generative adversarial network. In *CVPR*, volume 2, page 4, 2017.

[23] H. Li, W. Yang, J. Wang, Y. Xu, and L. Huang. Wifinger: Talk to your smart devices with finger-grained gesture. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, UbiComp '16, pages 250–261, New York, NY, USA, 2016. ACM.

[24] W. Mao, M. Wang, and L. Qiu. Aim: Acoustic imaging on a mobile. In *Proceedings of the 16th Annual International Conference on Mobile Systems, Applications, and Services*, MobiSys '18, pages 468–481, New York, NY, USA, 2018. ACM.

[25] A. Moreira, P. Prats-Iraola, M. Younis, G. Krieger, I. Hajnsek, and K. P. Papathanassiou. A tutorial on synthetic aperture radar. *IEEE Geoscience and remote sensing magazine*, 1(1):6–43, 2013.

[26] V. Nair and G. E. Hinton. Rectified linear units improve restricted boltzmann machines. In *Proceedings of the 27th international conference on machine learning (ICML-10)*, pages 807–814, 2010.

[27] K. Nasrollahi and T. B. Moeslund. Super-resolution: a comprehensive survey. *Machine vision and applications*, 25(6):1423–1468, 2014.

[28] A. Paszke, S. Gross, S. Chintala, G. Chanan, E. Yang, Z. DeVito, Z. Lin, A. Desmaison, L. Antiga, and A. Lerer. Automatic differentiation in pytorch. In *NIPS-W*, 2017.

[29] A. Paulraj, R. Roy, and T. Kailath. Estimation of signal parameters via rotational invariance techniques-esprit. In *Nineteeth Asilomar Conference on Circuits, Systems and Computers, 1985.*, pages 83–89. IEEE, 1985.

[30] Z. Peng, J. M. Muñoz-Ferreras, Y. Tang, C. Liu, R. Gómez-García, L. Ran, and C. Li. A portable fmcw interferometry radar with programmable low-if architecture for localization, isar imaging, and vital sign tracking. *IEEE Transactions on Microwave Theory and Techniques*, 65(4):1334–1344, 2017.

[31] J. F. Penner and D. G. Long. Ground-based 3d radar imaging of trees using a 2d synthetic aperture. *Electronics*, 6(1):11, 2017.

[32] N. Pohl, T. Jaeschke, and M. Vogt. An sige-chip-based 80 ghz fmcw-radar system with 25 ghz bandwidth for high resolution imaging. In *Radar Symposium (IRS), 2013 14th International*, volume 1, pages 239–244. IEEE, 2013.

[33] J. K. Pontes, C. Kong, S. Sridharan, S. Lucey, A. Eriksson, and C. Fookes. Image2mesh: A learning framework for single image 3d reconstruction. *arXiv preprint arXiv:1711.10669*, 2017.

[34] Q. Pu, S. Gupta, S. Gollakota, and S. Patel. Whole-home gesture recognition using wireless signals. In *Proceedings of the 19th annual international conference on Mobile computing & networking*, pages 27–38. ACM, 2013.

[35] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.

[36] R. Schmidt. Multiple emitter location and signal parameter estimation. *IEEE transactions on antennas and propagation*, 34(3):276–280, 1986.

[37] B. Schwarz. Lidar: Mapping the world in 3d. *Nature Photonics*, 4(7):429, 2010.

[38] F. Su and C. Joslin. Acoustic imaging using a 64-node microphone array and beamformer system. In *Signal Processing and Information Technology (ISSPIT), 2015 IEEE International Symposium on*, pages 168–173. IEEE, 2015.

[39] Y. Sun, Y. Wang, Z. Liu, J. E. Siegel, and S. E. Sarma. Pointgrow: Autoregressively learned point cloud generation with self-attention. *arXiv preprint arXiv:1810.05591*, 2018.

[40] J.-W. Ting, D. Oloumi, and K. Rambabu. Fmcw sar system for near-distance imaging applicationspractical considerations and calibrations. *IEEE Transactions on Microwave Theory and Techniques*, 66(1):450–461, 2018.

[41] D. Vasisht, S. Kumar, and D. Katabi. Decimeter-level localization with a single wifi access point. In *NSDI*, volume 16, pages 165–178, 2016.

[42] A. Virmani and M. Shahzad. Position and orientation agnostic gesture recognition using wifi. In *Proceedings of the 15th Annual International Conference on Mobile Systems, Applications, and Services*, MobiSys '17, pages 252–264, New York, NY, USA, 2017. ACM.

[43] G. Wang, Y. Zou, Z. Zhou, K. Wu, and L. M. Ni. We can hear you with wi-fi! In *Proceedings of the 20th Annual International Conference on Mobile Computing and Networking*, MobiCom '14, pages 593–604, New York, NY, USA, 2014. ACM.

[44] X. Wang, L. Gao, and S. Mao. Csi phase fingerprinting for indoor localization with a deep learning approach. *IEEE Internet of Things Journal*, 3(6):1113–1123, 2016.

[45] Y. Wang, K. Wu, and L. M. Ni. Wifall: Device-free fall detection by wireless networks. *IEEE Transactions on Mobile Computing*, 16(2):581–594, 2017.

[46] Z. Wang, D. Liu, J. Yang, W. Han, and T. Huang. Deep networks for image super-resolution with sparse prior. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 370–378, 2015.

[47] J. Xie, L. Xu, and E. Chen. Image denoising and inpainting with deep neural networks. In *Advances in neural information processing systems*, pages 341–349, 2012.

[48] J. Yang, J. Wright, T. Huang, and Y. Ma. Image super-resolution as sparse representation of raw image patches. In *2008 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8, June 2008.

[49] W. Yifan, S. Wu, H. Huang, D. Cohen-Or, and O. Sorkine-Hornung. Patch-based progressive 3d point set upsampling. *arXiv preprint arXiv:1811.11286*, 2018.

[50] L. Yu, X. Li, C.-W. Fu, D. Cohen-Or, and P.-A. Heng. Pu-net: Point cloud upsampling network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2790–2799, 2018.

[51] X. Yu and F. Porikli. Ultra-resolving face images by discriminative generative networks. In *European conference on computer vision*, pages 318–333. Springer, 2016.

[52] M. Zhao, F. Adib, and D. Katabi. Emotion recognition using wireless signals. In *Proceedings of the 22nd Annual International Conference on Mobile Computing and Networking*, pages 95–108. ACM, 2016.

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

[54] M. Zhao, Y. Tian, H. Zhao, M. A. Alsheikh, T. Li, R. Hristov, Z. Kabelac, D. Katabi, and A. Torralba. Rf-based 3d skeletons. In *Proceedings of the 2018 Conference of the ACM Special Interest Group on Data Communication*, pages 267–281. ACM, 2018.

[55] Y. Zhu, Y. Yao, B. Y. Zhao, and H. Zheng. Object recognition and navigation using a single networking device. In *Proceedings of the 15th Annual International Conference on Mobile Systems, Applications, and Services*, pages 265–277. ACM, 2017.

[56] Y. Zhu, Y. Zhu, B. Y. Zhao, and H. Zheng. Reusing 60ghz radios for mobile radar imaging. In *Proceedings of the 21st Annual International Conference on Mobile Computing and Networking*, pages 103–116. ACM, 2015.