

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/340010535>

Mitigation of Through-Wall Distortions of Frontal Radar Images Using Denoising Autoencoders

Article in IEEE Transactions on Geoscience and Remote Sensing · March 2020

DOI: 10.1109/TGRS.2020.2978440

CITATIONS

0

READS

85

2 authors:



Shelly Vishwakarma

University College London

18 PUBLICATIONS 42 CITATIONS

SEE PROFILE



Shobha Sundar Ram

Indraprastha Institute of Information Technology

52 PUBLICATIONS 674 CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:



On-Chip Wireless Channel Modelling [View project](#)



OPERA - Opportunistic Passive Radar for Non-Cooperative Contextual Sensing [View project](#)

Mitigation of Through-Wall Distortions of Frontal Radar Images using Denoising Autoencoders

Shelly Vishwakarma, *Student Member, IEEE*, Shobha Sundar Ram, *Member, IEEE*

Abstract—Radar images of humans and other concealed objects are considerably distorted by attenuation, refraction and multipath clutter in indoor through-wall environments. While several methods have been proposed for removing target independent static and dynamic clutter, there still remain considerable challenges in mitigating target dependent clutter especially when the knowledge of the exact propagation characteristics or analytical framework is unavailable. In this work we focus on mitigating wall effects using a machine learning based solution—denoising autoencoders— that does not require prior information of the wall parameters or room geometry. Instead, the method relies on the availability of a large volume of training radar images gathered in through-wall conditions and the corresponding clean images captured in line-of-sight conditions. During the training phase, the autoencoder learns how to denoise the corrupted through-wall images in order to resemble the free space images. We have validated the performance of the proposed solution for both static and dynamic human subjects. The frontal radar images of static targets are obtained by processing wideband planar array measurement data with two-dimensional array and range processing. The frontal radar images of dynamic targets are simulated using narrowband planar array data processed with two-dimensional array and Doppler processing. In both simulation and measurement processes, we incorporate considerable diversity in the target and propagation conditions. Our experimental results, from both simulation and measurement data, show that the denoised images are considerably more similar to the free-space images when compared to the original through-wall images.

Index Terms—Through-wall radar, Denoising autoencoders, sFDTD, Doppler/range enhanced frontal imaging

I. INTRODUCTION

Through-the-wall radar imaging (TWRI) has been extensively researched in recent years, for detecting and monitoring humans and other concealed objects in urban environments. There are varied applications for TWRI such as law enforcement, security, and surveillance, search and rescue, and indoor monitoring of the elderly [1]–[5]. There are broadly two categories of through-the-wall radars: narrowband and broadband. Broadband radars provide excellent downrange resolutions to locate and resolve multiple targets as well as for estimating building layouts [6]. Alternately, narrow band continuous wave (CW) radars have been developed to detect dynamic targets based on their Doppler signatures [2], [3], [7]. Both of these systems can be complemented with two-dimensional array processing to provide either range-enhanced frontal images or Doppler-enhanced frontal images [8], [9]. Frontal images of the humans provide informative signatures of their activities [10]. However, when the radars are deployed in through-wall scenarios, the quality of the radar images significantly deteriorate due to the through-wall propagation artifacts such as - attenuation, defocussing and multipath clutter [3], [8], [11]–[13].

Indoor clutter can be broadly categorized into *target independent static and dynamic clutter*, and *target dependent clutter*. *Target independent static clutter* arise from the reflections off the wall (especially the front face in a through-wall scenario), ceiling, floor, and furniture. Static clutter is easy to eliminate through filtering when the objective is to detect dynamic targets. The

problem becomes more challenging in the context of detection of static and slow-moving targets. Authors in [14] assumed the availability of background data that could be coherently subtracted from the target measurements. Alternately, sparsity-based multipath exploitation methods were explored in [15]–[17]. Here, the algorithm leveraged the orthogonality between the static clutter and the target scattering to mitigate the clutter. *Target independent dynamic clutter* arising from other moving objects in the environment can significantly interfere with Doppler signatures of moving targets. In [18], a method to segregate the Doppler returns from multiple targets was presented. This technique could be used for mitigating dynamic clutter. The third category is *target dependent clutter* that arises from the interactions of the target (static or dynamic) and the complex propagation channel. As a result of refraction and multipath, the radar images are smeared, blurred, and there are shifts in the location of point scatterers in the images [8]. The authors in [19] and [20] used back-projection and sparsity based change detection algorithms, respectively, to track slow moving humans in the range-crossrange space in the presence of target dependent clutter. Both these techniques, however, rely on the availability of accurate knowledge of the through-wall scenarios for detecting static targets. Alternately, the authors in [21], [22] exploited the multipath (instead of suppressing the multipath) to improve the effective signal-to-clutter ratio (SCR) at the original target locations. They removed ghost artifacts by mapping the multipath ghosts to their true targets. Again the technique requires exact information of the room geometry and wall characteristics.

In this paper, we propose an alternate strategy, based on denoising autoencoders, for recovering radar images corrupted by through-wall effects. An autoencoder is a neural network that extracts relevant features from the noisy input data for various tasks such as- dimensionality reduction and data denoising [23], [24]. Autoencoders have been widely used for applications such as anomaly detection, natural language processing, denoising and domain adaptation [23], [25]–[27]. Some preliminary results for clutter mitigation using autoencoders were presented in [28] where the nature of the type of through-wall scenario was assumed to be known during the test phase. The primary advantage of this technique is, however, that the autoencoders require neither prior information regarding the wall characteristics nor any kind of analytic framework to describe the through-wall interference. Instead, the distorted radar signatures due to wall interference are treated as corrupt versions of ideal radar signatures obtained in free space conditions. The algorithm *learns* how to denoise or clean the corrupted signals using training data comprising of both corrupted and clean data. We demonstrate, in this paper, that the autoencoder can be used for removing signal dependent clutter when no information or label of the through-wall scenario is assumed to be known during the test phase. Instead, the autoencoder is trained with a mixture of images gathered in diverse through-wall conditions. Traditional autoencoders have been implemented using back-propagation algorithms such as- gradient descent [29], conjugate gradient descent [30] and steepest descent

[24]. However, they have a very slow learning rate. This translates to long training times and, in some cases, the convergence may not be guaranteed. Instead, we propose to use an alternating direction method of multipliers (ADMM) approach [31], where we break the complex convex optimization problem into smaller sub-problems with closed-form solutions. Thus the convergence is guaranteed and training times are not very long.

We test the performance of the proposed algorithm on two types of radar images - Doppler enhanced and range enhanced frontal images. The *Doppler-enhanced frontal* images of *dynamic human motions* are generated from simulated narrowband radar data of human motions in through-wall environments using the techniques described in [8], [9]. We consider a variety of walls- a dielectric wall, a dielectric wall with metal reinforcements and one with air-gaps. The through-wall propagation phenomenology is modeled using finite difference time-domain (FDTD) techniques [32]. We introduce significant diversity in wall parameters such as dielectric constant and conductivity by incorporating stochasticity in the finite difference equations as suggested by [33]. This is a computationally more efficient technique than running multiple FDTD simulations with varying wall parameters. The second set of images are *range-enhanced frontal images* captured of *static humans* using measurement data gathered with Walabot, a three dimensional programmable, wideband imaging radar [34]. During the training phase, the autoencoder is trained with a diverse mixture of data gathered from different through-wall scenarios. In the test phase, the network denoises the corrupted radar image without requiring any information of the type of wall or its parameters. Both the simulation and measurement results obtained from the autoencoder exhibit very low normalized mean square error and high structural similarity between the denoised reconstructed images and free space images.

To summarize, our contributions in this paper are the following:

- First, we propose a denoising autoencoder to mitigate clutter and distortion in through-wall frontal images of both static and dynamic humans.
- Second, we propose a method to implement the autoencoder using ADMM approach to ensure convergence and fast training times.

In Section. II, we briefly describe the denoising autoencoder structure implemented in our work. Then, in Section III, we use a computationally efficient method based on stochastic FDTD to simulate narrowband Doppler enhanced frontal images of dynamic humans. We present the simulation results where we declutter Doppler-enhanced frontal images of a human subject in Section. IV. Finally, we denoise range-enhanced through-wall radar images of static humans captured by a wideband RF sensor, Walabot, in Section. V. We present the results, analyses and discussion on the strengths and limitations of the proposed approach in the final section. *Notation:* We use the following notation in the paper. Matrices are written in capital bold letters while vectors and variables are written in normal letters.

II. THEORY

Radar images deteriorate significantly due to distortions and multipath clutter signals introduced by through-wall environments. The images may be defocused, blurred or smeared. Ghost targets may appear due to multipath. The objective, here, is to reconstruct clean radar images resembling free space images from corrupted through-wall images. We divide our denoising problem into two stages- training and the test stages.

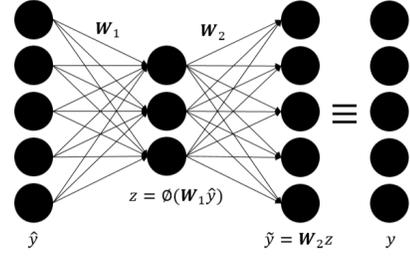


Fig. 1: Denoising Autoencoder

A. Training Stage

A conventional denoising autoencoder shown in Fig.1, first corrupts the clean input data by adding stochastic Gaussian noise, then feeds the corresponding noisy version as input data to the next stage. In this work, we consider the measurements in a through-wall case as our noisy/corrupted data. The main difference is the non-Gaussian nature of interference. During the training stage, M radar images of the target are captured in free space. We vectorise each image of size $\mathbb{R}^{I \times J}$ to obtain $y^{tr} \in \mathbb{R}^{N \times 1}$ where $N = I \times J$, is the total number of pixels in an image. We stack the M images as columns to form a data matrix $\mathbf{Y}^{tr} \in \mathbb{R}^{N \times M}$. We repeat this exercise for the images captured in the corresponding through-wall scenarios to generate $\hat{\mathbf{Y}}^{tr}$ also of size $\mathbb{R}^{N \times M}$. The autoencoder has primarily two stages- encoding and the decoding. In the encoding stage, the algorithm learns a latent/compressed representation $\mathbf{Z} \in \mathbb{R}^{R \times M}$, of the input layer $\hat{\mathbf{Y}}^{tr}$ as shown in (1).

$$\mathbf{Z} = \phi(\mathbf{W}_1 \hat{\mathbf{Y}}^{tr}) \quad (1)$$

Here, ϕ is the mapping function which can either be linear or nonlinear (such as sigmoid, tanh), $\mathbf{W}_1 \in \mathbb{R}^{R \times N}$ the corresponding weight matrix and R is the number of nodes in the hidden layer. Since the hidden layer, \mathbf{Z} , is the compressed representation of the input layer, $\hat{\mathbf{Y}}^{tr}$, it always has fewer nodes than the number of pixels ($R \ll N$). In the decoding stage, the algorithm maps \mathbf{Z} back to obtain a reconstructed signal $\hat{\mathbf{Y}}^{tr} = \mathbf{W}_2 \phi(\mathbf{W}_1 \hat{\mathbf{Y}})$ through weight matrix $\mathbf{W}_2 \in \mathbb{R}^{N \times R}$ such that the error e ,

$$e = \left\| \mathbf{Y}^{tr} - \hat{\mathbf{Y}}^{tr} \right\|_2^2, \quad (2)$$

between the reconstructed images and the free space images is minimized. Therefore the objective in the training stage is to learn weight matrices \mathbf{W}_1 and \mathbf{W}_2 so that the reconstructed images resemble free space images (instead of the corrupted through-wall images).

$$J(\mathbf{W}_1, \mathbf{W}_2) = \min_{\mathbf{W}_1, \mathbf{W}_2} \left\| \mathbf{Y}^{tr} - \mathbf{W}_2 \phi(\mathbf{W}_1 \hat{\mathbf{Y}}^{tr}) \right\|_2^2 \quad (3)$$

The objective function (3) can be solved in multiple ways - gradient descent, conjugate gradient descent, or steepest descent. In some of these ways, the error may become insignificant when back propagated. Additionally, these algorithms have a very slow learning rate. Instead, we propose an alternating direction method of multipliers (ADMM) approach [31]. Here, we introduce a simple variable separation technique to break the complex convex optimization problem into smaller sub-problems which have closed form solutions such that the convergence is guaranteed. The objective function in (3) is reformulated to

$$J(\mathbf{W}_1, \mathbf{W}_2, \mathbf{Z}) = \min_{\mathbf{W}_1, \mathbf{W}_2, \mathbf{Z}} \left\| \mathbf{Y}^{tr} - \mathbf{W}_2 \mathbf{Z} \right\|_2^2 \quad s.t. \quad \mathbf{Z} = \phi(\mathbf{W}_1 \hat{\mathbf{Y}}^{tr}). \quad (4)$$

Since the formulation in (4) is a constrained optimization problem, we relax it using an augmented Lagrangian technique shown below

$$J(\mathbf{W}_1, \mathbf{W}_2, \mathbf{Z}) = \min_{\mathbf{W}_1, \mathbf{W}_2, \mathbf{Z}} \left\| \mathbf{Y}^{tr} - \mathbf{W}_2 \mathbf{Z} \right\|_2^2 + \mu^T \left(\mathbf{Z} - \phi(\mathbf{W}_1 \hat{\mathbf{Y}}^{tr}) \right) + \lambda \left\| \mathbf{Z} - \phi(\mathbf{W}_1 \hat{\mathbf{Y}}^{tr}) \right\|_2^2 \quad (5)$$

Here, the Lagrangian term $\mu^T (\mathbf{Z} - \phi(\mathbf{W}_1 \hat{\mathbf{Y}}^{tr}))$ imposes equality between \mathbf{Z} and the underlying representation $\phi(\mathbf{W}_1 \hat{\mathbf{Y}}^{tr})$ at every iteration, whereas the augmented Lagrangian term $\lambda \|\mathbf{Z} - \phi(\mathbf{W}_1 \hat{\mathbf{Y}}^{tr})\|_2^2$ is minimized over successive iterations and hence ensures the equality only during convergence. We do not require the stringent equality constraint at every iteration and hence ignore the Lagrangian term and reformulate the objective function as shown in

$$J(\mathbf{W}_1, \mathbf{W}_2, \mathbf{Z}) = \min_{\mathbf{W}_1, \mathbf{W}_2, \mathbf{Z}} \|\mathbf{Y}^{tr} - \mathbf{W}_2 \mathbf{Z}\|_2^2 + \lambda \|\mathbf{Z} - \phi(\mathbf{W}_1 \hat{\mathbf{Y}}^{tr})\|_2^2. \quad (6)$$

Here, λ is the regularization parameter between the proxy variable \mathbf{Z} and underlying representation $\phi(\mathbf{W}_1 \hat{\mathbf{Y}}^{tr})$. We divide (6) into a set of smaller sub problems as follows.

Problem1:

$$J(\mathbf{W}_1) = \min_{\mathbf{W}_1} \lambda \|\phi^{-1} \mathbf{Z} - \mathbf{W}_1 \hat{\mathbf{Y}}^{tr}\|_2^2 \quad (7)$$

Problem2:

$$J(\mathbf{W}_2) = \min_{\mathbf{W}_2} \|\mathbf{Y}^{tr} - \mathbf{W}_2 \mathbf{Z}\|_2^2 \quad (8)$$

Problem3:

$$\begin{aligned} J(\mathbf{Z}) &= \min_{\mathbf{Z}} \|\mathbf{Y}^{tr} - \mathbf{W}_2 \mathbf{Z}\|_2^2 + \lambda \|\mathbf{Z} - \phi(\mathbf{W}_1 \hat{\mathbf{Y}}^{tr})\|_2^2 \\ &= \min_{\mathbf{Z}} \left\| \begin{pmatrix} \mathbf{Y}^{tr} \\ \sqrt{\lambda} \phi(\mathbf{W}_1 \hat{\mathbf{Y}}^{tr}) \end{pmatrix} - \begin{pmatrix} \mathbf{W}_2 \\ \sqrt{\lambda} \mathbf{I} \end{pmatrix} \mathbf{Z} \right\|_2^2 \end{aligned} \quad (9)$$

Sub problems in (7)-(9) are all simple least squares problems which already have a closed form solution [35]. At each iteration, we update the network weight \mathbf{W}_1 , \mathbf{W}_2 and proxy variable \mathbf{Z} , till the algorithm converges.

B. Test Stage

During test stage, P corrupted radar images are combined together to form $\hat{\mathbf{Y}}^{test} \in \mathbb{R}^{N \times P}$ and passed through the autoencoder to obtain $\tilde{\mathbf{Y}} \in \mathbb{R}^{N \times P}$. We hypothesize that once the network is trained, we can use weight matrices \mathbf{W}_1 and \mathbf{W}_2 to obtain a denoised form $\tilde{\mathbf{Y}}^{test}$ of the corrupted test data $\hat{\mathbf{Y}}^{test}$ as shown in Fig.2(b).

$$\tilde{\mathbf{Y}}^{test} = \mathbf{W}_2 \phi(\mathbf{W}_1 \hat{\mathbf{Y}}^{test}) \quad (10)$$

Note that the proposed denoising algorithm is significantly faster in generating denoised images at test time as it involves only a simple product operation in (10). This makes the algorithm suitable for real-time applications.

C. Metrics for evaluation

We evaluate the effectiveness of the proposed clutter mitigation algorithm using two metrics- normalized mean square error (NMSE) and structural similarity index (SSIM). We consider the image captured in free space as the clean/ground truth image (\mathbf{Y}^{test}). We calculate the NMSE and SSIM between the through-wall image $\hat{\mathbf{Y}}^{test}$ and ground truth image before denoising (BD). Then we repeat the exercise after denoising (AD). In the second case, the NMSE and SSIM are calculated between the reconstructed / denoised image, $\tilde{\mathbf{Y}}^{test}$, and the ground truth image. The hypothesis, here, is that the NMSE and SSIM will improve after denoising.

The NMSE is computed between \mathbf{Y}^{test} and $\tilde{\mathbf{Y}}^{test}$ using (11).

$$NMSE = \frac{\|\mathbf{Y}^{test} - \tilde{\mathbf{Y}}^{test}\|_2^2}{\|\mathbf{Y}^{test}\|_2^2} \quad (11)$$

NMSE is sensitive to the energy of absolute errors of all the pixels of an image. However, NMSE between two images may be low even if they have drastically different structural features [36]. SSIM [37], on the other hand, is a metric that provides information

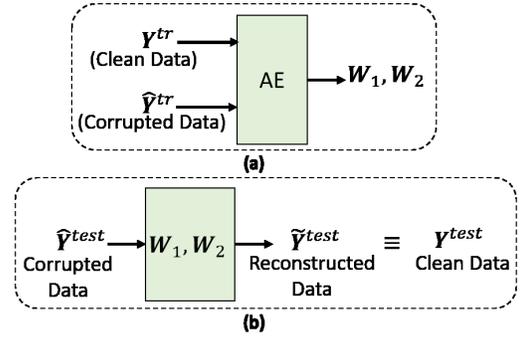


Fig. 2: (a) During training stage, the autoencoder learns coefficients \mathbf{W}_1 and \mathbf{W}_2 from clean free space images (\mathbf{Y}^{tr}) and corrupted through-wall ($\hat{\mathbf{Y}}^{tr}$) images. (b) During test phase, corrupted through-wall images ($\hat{\mathbf{Y}}^{test}$) are denoised to obtain reconstructed images that resemble free space radar images (\mathbf{Y}^{test}).

of the luminance (L), contrast (C) and structure difference (S), between the ground truth image \mathbf{Y}^{test} , and the test image $\hat{\mathbf{Y}}^{test}$. Its value should be 1 if the images are identical. The overall measurement metric becomes the multiplicative combination of three measures shown in (12)

$$SSIM(\hat{\mathbf{Y}}, \mathbf{Y}) = [L(\hat{\mathbf{Y}}, \mathbf{Y})]^\alpha [C(\hat{\mathbf{Y}}, \mathbf{Y})]^\beta [S(\hat{\mathbf{Y}}, \mathbf{Y})]^\gamma \quad (12)$$

We assume $\alpha = \beta = \gamma = 1$. The expressions for L, C, S are

$$L(\hat{\mathbf{Y}}, \mathbf{Y}) = \frac{2\mu_{\hat{\mathbf{Y}}}\mu_{\mathbf{Y}} + C_1}{\mu_{\hat{\mathbf{Y}}}^2 + \mu_{\mathbf{Y}}^2 + C_1} \quad (13)$$

$$C(\hat{\mathbf{Y}}, \mathbf{Y}) = \frac{2\sigma_{\hat{\mathbf{Y}}}\sigma_{\mathbf{Y}} + C_2}{\sigma_{\hat{\mathbf{Y}}}^2 + \sigma_{\mathbf{Y}}^2 + C_2} \quad (14)$$

$$S(\hat{\mathbf{Y}}, \mathbf{Y}) = \frac{\sigma_{\hat{\mathbf{Y}}\mathbf{Y}} + C_3}{\sigma_{\hat{\mathbf{Y}}}\sigma_{\mathbf{Y}} + C_3} \quad (15)$$

Here, $\mu_{\mathbf{Y}}$, $\mu_{\hat{\mathbf{Y}}}$, $\sigma_{\mathbf{Y}}$, $\sigma_{\hat{\mathbf{Y}}}$ and $\sigma_{\hat{\mathbf{Y}}\mathbf{Y}}$ are the local means, standard deviations and the co-variance for the reference \mathbf{Y} and test images $\hat{\mathbf{Y}}$ respectively. Assuming $C_3 = \frac{C_2^2}{2}$, the simplified index becomes

$$SSIM(\hat{\mathbf{Y}}, \mathbf{Y}) = \frac{(2\mu_{\hat{\mathbf{Y}}}\mu_{\mathbf{Y}} + C_1)(2\sigma_{\hat{\mathbf{Y}}\mathbf{Y}} + C_2)}{(\mu_{\hat{\mathbf{Y}}}^2 + \mu_{\mathbf{Y}}^2 + C_1)(\sigma_{\hat{\mathbf{Y}}}^2 + \sigma_{\mathbf{Y}}^2 + C_2)}. \quad (16)$$

We therefore conclude that two images can be regarded as similar only when both NMSE is low and SSIM is close to 1.

III. SIMULATION METHOD

In this section we describe the simulation method to generate a large database of Doppler enhanced frontal images of humans in diverse through-wall conditions. We adopt the technique described in [8] and briefly describe its salient features in the following subsection. We only model the through-wall propagation phenomenology and do not consider multipath scattering from the ceiling, ground and lateral walls. The wall propagation phenomenology, modelled using finite difference time domain techniques (FDTD), and primitive based models of humans are hybridized to generate Doppler-enhanced frontal radar images. There may be considerable variations in the propagation conditions during training and test due to variations in the wall characteristics such as its dielectric constant and loss tangent. Modeling this diversity with independent FDTD simulations is computationally expensive. Therefore, we extend the simulation framework discussed in [4] by incorporating stochasticity in the propagation channel using the stochastic FDTD (sFDTD) technique suggested by [33]. The sFDTD method introduces statistical variations in the electrical properties of the medium. The results of the simulations provide the mean and the variance estimates

of the time-domain electromagnetic fields at every point in the problem space from which numerous samples of the through-wall propagation can be generated. We describe these steps in greater detail in the following sections.

A. Radar Signal Model

We consider a radar with a single continuous wave transmitter of frequency f_c (wavelength λ_c). The receiver consists of an $P \times Q$ uniform planar array along the $X - Y$ plane with element spacing $\lambda_c/2$. We use motion capture data for realistically modeling complex human animation. We interpolate the data from the video frame rate to the radar sampling frequency. The data provides time-varying three-dimensional positions of B markers distributed over the entire human body - head, torso, arms, shoulders, and legs as shown in Fig.3. We assume that these markers correspond to point scatterers each of reflectivity a_b and located at spherical coordinate positions (r_b, θ_b, ϕ_b) where θ_b is the elevation from the XZ plane and ϕ_b is the azimuth with respect to the positive Z axis on the human body. Each body part moves with a Doppler

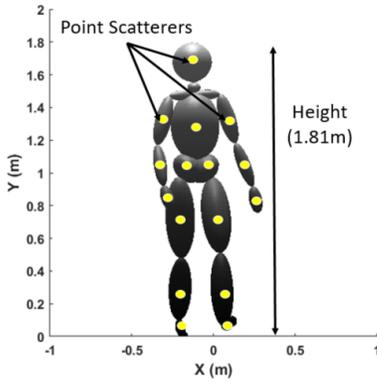


Fig. 3: Radar scattering model of human.

frequency f_{D_b} . The baseband, digitized radar data at each $(p, q)^{th}$ element is given by

$$\mathbf{X}_{p,q,n} = \sum_{b=1}^B a_b[n] e^{-j \frac{2\pi}{\lambda_c} \|\vec{r}_b[n] - \vec{r}_{p,q}\|_2^2}. \quad (17)$$

where n denotes the time sample and $\vec{r}_{p,q}$ is the position of the radar element. The data are processed with discrete short time Fourier transform across the time-domain using a sliding time window centered around τ and two-dimensional Fourier beamforming across the array to obtain a time varying three-dimensional radar cube in terms of azimuth, elevation and Doppler, $\chi_{\phi,\theta,f_D}^\tau$, as shown in (18).

$$\begin{aligned} \chi_{\phi,\theta,f_D}^\tau &= F_{3D} \mathbf{X}_{p,q,n} \\ &= \sum_{b=1}^B a_b H_{3D} [\phi - \phi_b(\tau), \theta - \theta_b(\tau), f_D - f_{D_b}(\tau)]. \end{aligned} \quad (18)$$

Here, $H_{3D}[\cdot]$ is the three-dimensional point spread function across the three domains and F_{3D} indicates the three-dimensional Fourier operator. From the radar cube, we identify peak scatterers of strength $a_m^\tau[f_D]$ located at (ϕ_m, θ_m) at every Doppler bin f_D . We generate Doppler enhanced frontal images $\chi_{\phi,\theta}^\tau$ for every τ by convolving the peak scatterers with a two-dimensional point spread function, $H_{2D}[\cdot]$ as shown in (19).

$$\chi_{\phi,\theta}^\tau = \sum_{f_D} a_m^\tau(f_D) H_{2D} [\phi - \phi_m, \theta - \theta_m]. \quad (19)$$

The incorporation of the additional Doppler dimension enables us to resolve multiple scatterers of the human body along two spatial dimensions thus alleviating the need for larger arrays required for

TABLE I: Simulated Radar Parameters

Radar Parameters	Values
Radar Type	Narrowband
Carrier frequency (f_c)	7.5GHz
Sampling frequency (f_s)	1000Hz
Integration time (T)	0.8s
Dwell time or short time (t_D)	0.1s
Maximum Doppler (f_{Dmax})	± 500 Hz
Doppler resolution (Δf_D)	10Hz
Number of antenna elements ($P \times Q$)	10×10
Azimuth Beamwidth ($\Delta \phi_{azi}$)	10°
Azimuth Beamwidth ($\Delta \theta_{ele}$)	10°
Field of View (ϕ_{azi})	-90° to 90°
Field of View (θ_{azi})	-90° to 90°

successful imaging. The radar signal model parameters that we use in this work are provided in Table.I.

B. Stochastic Model of Through-Wall Propagation

The formulation discussed above describes the radar images generated in free space conditions. In this section, we discuss the simulation of through-wall propagation phenomenology. As shown in Fig.4, we consider a two-dimensional simulation space extending from -1m to 1m and 0m to 4m along the X and Z directions respectively (assuming the wall is invariant along Y height axis). The two-dimensional simulation framework is chosen to reduce the computational complexity of the problem and because most walls show homogeneity along the height. In order to correspond to the radar signal model discussed earlier, we consider a uniform 10 element linear antenna array whose elements are spaced half wavelength apart. We independently simulate the excitation from each element of the array, located at $\vec{\rho}_p$, with a narrowband sinusoidal signal at f_c . The simulation space is bounded by a perfectly matched layer and divided into spatial grids of $\lambda_c/10$ size. We considered three different wall configurations - a homogeneous dielectric wall (Fig.4(a)), a wall reinforced with metal rods (Fig.4(b)), and a wall with air gaps (Fig.4(c)). Each wall type is simulated independently. The dimensions of each wall is 2m x 30cm (X : -1m to 1m, Z : 1m to 1.3m). In all three cases, stochastic variations of 10% standard deviation are introduced in the relative permittivity $\epsilon_r = 4$, and conductivity $\sigma_c = 0.001$ S/m, of each grid point in the wall. Therefore, even the single layer dielectric wall is not truly homogeneous. This is done to mirror real world conditions. For every point in space, $\vec{\rho}_b$, and at every time instant n , the sFDTD simulation gives the mean time-domain electric field $\mu_E[\vec{\rho}_b, \vec{\rho}_p, n]$ and its standard deviation $\sigma_E[\vec{\rho}_b, \vec{\rho}_p, n]$. We use the Gaussian stochastic model to generate 200 samples ($\eta = 1 \dots 200$) of time-domain electric field values $E[\vec{\rho}_b, \vec{\rho}_p, n, \eta] \sim \mathcal{N}(\mu_E, \sigma_E^2)$. The $E[\vec{\rho}_b, \vec{\rho}_p, n, \eta]$ is fast Fourier transformed to obtain the corresponding frequency domain wall transfer function $H_{wall}(\vec{\rho}_b, \vec{\rho}_p, \eta)$ at 7.5GHz.

C. Modeling of electromagnetic radar scatter from dynamic humans in through-wall scenarios

We integrate the free space radar signal model model with the through-wall propagation based on [3], [4]. We considered a scenario where the human is moving behind a wall before a radar. Since our FDTD simulation spatial extent is limited, we removed translational motion of the human and only retained the dynamics of the swinging arms and legs. Therefore, the time-domain scattered returns at each (p, q) antenna element is obtained by hybridizing H_{wall} with the human scattering center model as shown in (20).

$$\mathbf{X}_{p,q,n}[\eta] = \sum_{b=1}^B a_b \gamma_{2D \rightarrow 3D}(p, q, b) (H_{wall}[\vec{\rho}_b[n], \vec{\rho}_p, \eta])^2 \quad (20)$$

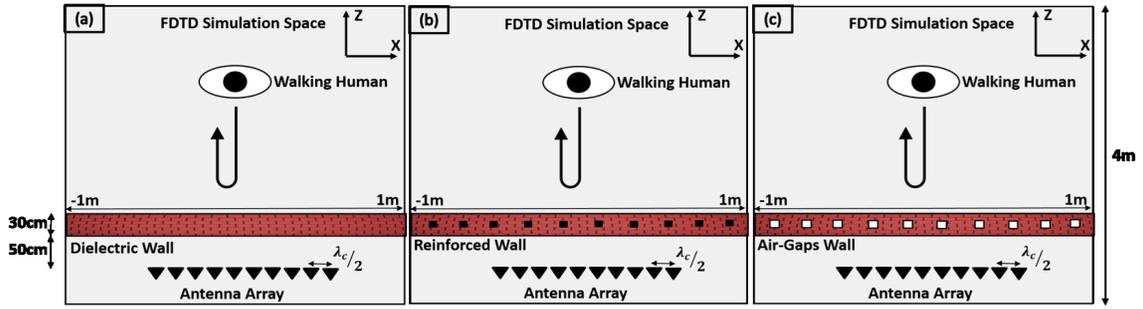


Fig. 4: Room geometry in through-wall scenario (a) Dielectric (b) Reinforced wall and (c) Wall with air-gaps

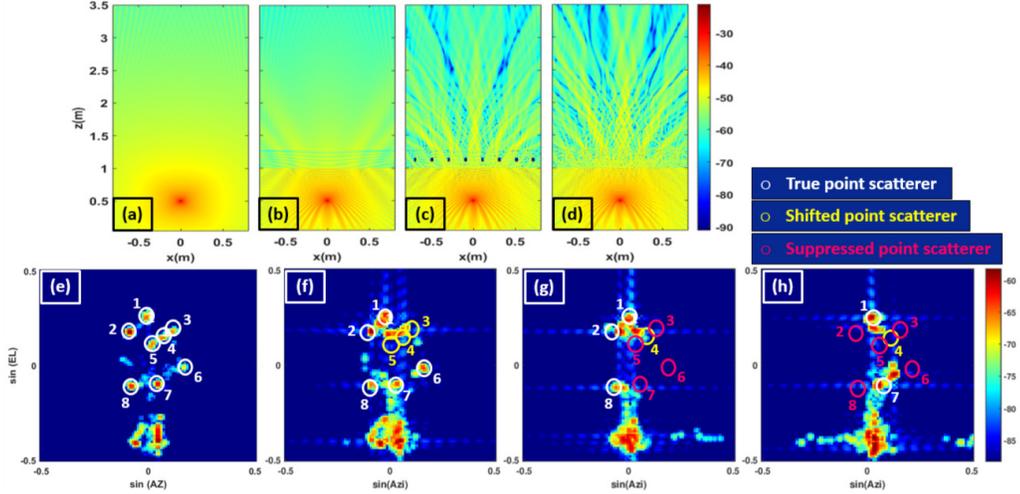


Fig. 5: Magnitude response at 7.5 GHz in (a) freespace, through (b) dielectric wall, (c) reinforced dielectric wall, and (d) dielectric wall with air-gaps respectively. (e)-(h) Corresponding simulated Doppler enhanced frontal image of a walking human

Each point scatterer on the human is projected from its three-dimensional position \vec{r}_b to its two-dimensional counterpart $\vec{\rho}_b$ on the $X-Z$ plane. The term $H_{wall}[\vec{\rho}_b, \vec{\rho}_q, \eta]$ models the propagation from the source $\vec{\rho}_p$ to $\vec{\rho}_b$ and is generated from a stochastic realization (η) of the full wave electromagnetic solver. The scaling factor $\gamma_{2D \rightarrow 3D}$, given by (21), adjusts the phase for modeling three-dimensional physics from the two-dimensional simulation as described in [3].

$$\gamma_{2D \rightarrow 3D}(p, q, b) = e^{-\frac{j2\pi}{\lambda} [\|\vec{r}_b[n] - \vec{r}_{p,q}\|^2 - \|\vec{\rho}_b[n] - \vec{\rho}_p\|^2]} \quad (21)$$

IV. SIMULATION RESULTS

We first consider a single stride of a human walking motion from Sony Computer Entertainment America. The duration of the motion is 0.8s which corresponds to one complete human stride. The human is walking away from the radar at an aspect angle of 180° . We generate the Doppler-enhanced frontal images for free space and the three wall types.

A. Doppler enhanced frontal imaging of dynamic humans in through-wall scenarios

The Doppler enhanced frontal images account for all the electromagnetic phenomenology introduced by the through-wall propagation conditions including attenuation, ringing and multipath. To demonstrate the validity of our claim, we show the magnitude response for the four scenarios at a carrier frequency of 7.5 GHz and the corresponding simulated Doppler enhanced frontal image of a single frame of a walking human, at a radar-target aspect angle of 180° , in Fig.5. Some comments regarding the wall-target interaction phenomenology on the frontal images:

Free Space: Figure 5(a) shows that the magnitude response of the electric field decays as the distance from the source increases. The phase response (not shown here) displays well behaved circular wave-fronts emanating from infinitely long line source excitation. Due to the undistorted phase response, we get the highest quality frontal images in the free space scenario. We can clearly discern both arms, legs and head of the human in the corresponding Doppler enhanced frontal image shown in Fig.5(e). The image falls within the $\pm 30^\circ$ field-of-view of the radar along elevation and azimuth. There is some smearing near the legs due to the limited resolution along azimuth and elevation of the array.

Dielectric Wall: The propagation of a signal through a homogeneous dielectric wall undergoes a two-way attenuation of approximately 12dB when compared to free space Fig.5(b). Hence, the strength of some of the peak scatterers in the frontal image Fig.5(f), become too weak to be visible on the same dynamic scale as that of free space case. This image clearly demonstrates the effect of through-wall attenuation on the images.

Reinforced Wall and Wall with Air Gaps: Figures 5(c), and (d) show that the inhomogeneity inside the wall causes multiple scattering that interfere constructively or destructively in some regions leading to significant distortions. As a result the radar frontal images are significantly distorted - positions of few scatterers get displaced along the azimuth direction due to refraction while some of the point scatterers are not visible at all because these lie at regions of destructive interference shown in Fig.5(g) and (h).

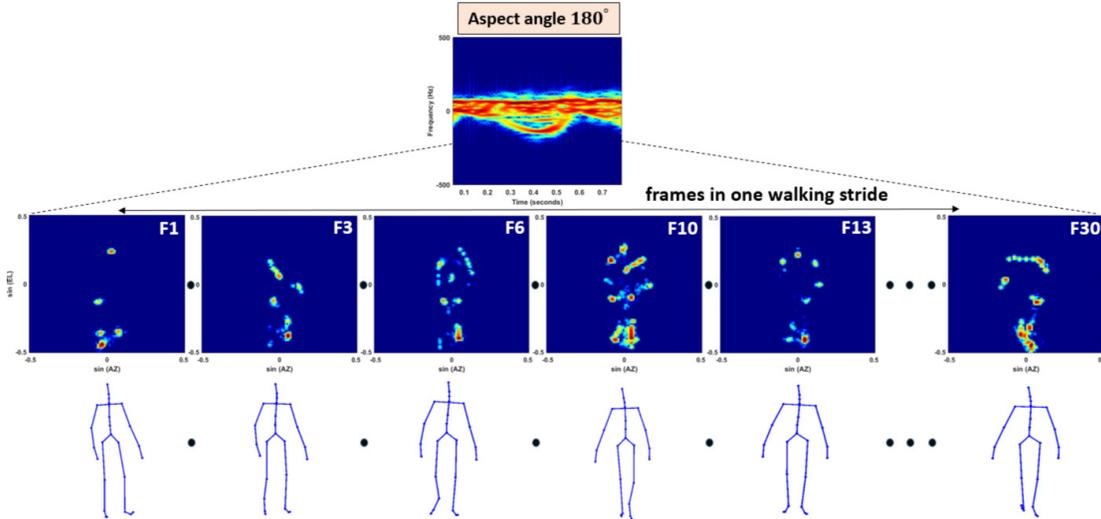


Fig. 6: Simulated Doppler enhanced frontal images over one walking stride of human motion and the corresponding stick figures obtained using ground truth motion capture data.

B. Results from denoising cluttered images

We selected 30 frames of a human subject walking away from the radar spanning a duration of 0.8s as shown in Fig.6. The training data consists of images corresponding to different motion states within a stride as shown in the stick figure models in Fig.6. The size of each image is $[92 \times 92]$. Corresponding to each of these frames, we generated 200 distinct through-wall radar images to capture statistical variations in the wall parameters. Each image is then vectorised to a column vector of size $[8464 \times 1]$, for final processing. Of the total set of 6000 images for each wall type, 80% images are used for training the autoencoder network and the remaining for the test.

The results are obtained by optimizing the number of nodes (R) in the hidden layer and the mapping function connecting the input and the hidden layer. We fixed the hidden layer dimension of the autoencoder network to be 500 and the mapping to be linear between input and the hidden layer. The choice of these parameters are discussed in the appendix A. During training, the weight matrices \mathbf{W}_1 and \mathbf{W}_2 each of size $[500 \times 8464]$ and $[8464 \times 500]$ respectively are first randomly initialized. The weights are updated over successive iterations as discussed in the previous section. The heuristic way of choosing λ is to start with a small value and then keep increasing its value across multiple trials. But this is not an elegant solution. Instead, a Bregman constant, can be introduced between the proxy and the original variable so as to retain a fixed value of λ to ensure equality during convergence [38], [39]. Our experiments showed that the results are not very sensitive to this constant which can subsequently be ignored. Since we give equal importance to both encoding and decoding stage, λ is fixed at 1. We have shown the convergence of the objective function with iterations of the learning algorithm in Fig. 7. Once learned, the weights are used for test.

Tables. II–III show the results for the metrics, SSIM and NMSE, as a function of number of distinct frames of the human walking motion. We compare the metrics obtained from images generated *before denoising (BD)* with those obtained *after denoising (AD)*. To give readers a better understanding of the generality of the solution, we have studied the sensitivity of the performance of the algorithm to different wall types during training and test scenarios. We considered the following three training scenarios.

Case 1: Train and test on data from same wall:

First, we consider the scenario where the autoencoder is trained

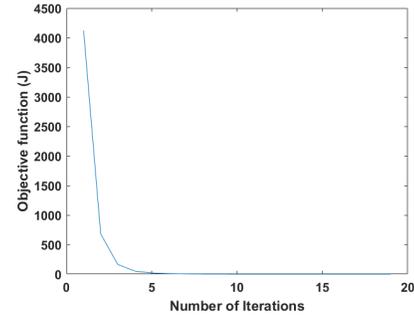


Fig. 7: Convergence curve of the proposed algorithm.

with data from a specific wall configuration and then subsequently tested on images generated from the same wall configuration. The training data set size in this case is $[8464 \times 4800]$. Note that even in the *same wall* case, there is diversity in the training and test data due to the statistical variations in the wall parameters as well as motion characteristics. Before denoising, the dielectric wall case has lowest error when compared to the reinforced and air-gaps walls. This is because the quality of the radar images are a function of the phase and amplitude distortions introduced by the walls to the radar signals. Therefore, based on the magnitude and phase responses shown in Fig.5(c) and (d), we observe the results deteriorate most in the case of the wall with air-gaps. The error between reconstructed and the free space images drop significantly for all wall types after passing through the denoising network. We varied the number of frames from 1 to 30 to increase the diversity in the human motions. Since this is a continuous motion, there may be some slight correlation between images obtained from consecutive frames. However, this is not very evident from visual inspection of the images corresponding to the frames as seen in Fig.6. So, we map the group correlation index across the multiple frames as we increase the diversity of the training data in Fig.8. The figure shows that the group correlation increases till it reaches a plateau of about 10 frames. Hence, the denoising performance seems to improve when we increase from a single frame to 10 number of frames as the training data captures the diversity of motions in the Tables. II–III. Beyond this, the performance of the denoising algorithm slightly deteriorates

TABLE II: Denoising results between clean and corrupted Doppler enhanced frontal images for different through-wall conditions. SSIM: between corrupted and free space image before denoising (BD) and SSIM :between reconstructed and free space image after denoising (AD)

Wall Scenario		Denoising Metric (SSIM)	Number of Frames				
			1	5	10	20	30
Train and test on same wall (Case 1)	Dielectric	BD	0.09	0.55	0.63	0.67	0.64
		AD	0.43	0.87	0.90	0.89	0.86
	Reinforced	BD	0.05	0.24	0.46	0.55	0.47
		AD	0.43	0.81	0.87	0.84	0.80
	Wall With Air-gaps	BD	0.03	0.01	0.34	0.51	0.43
		AD	0.42	0.82	0.83	0.82	0.77
Train and test on different walls (Case 2)		BD	0.05	0.24	0.46	0.55	0.46
		AD	0.41	0.43	0.55	0.08	0.01
Train on multiple walls (Case 3)		BD	0.04	0.26	0.43	0.52	0.45
		AD	0.19	0.81	0.84	0.83	0.78

TABLE III: Denoising results between clean and corrupted Doppler enhanced frontal images for different through-wall conditions. NMSE: between corrupted and free space image before denoising (BD) and NMSE :between reconstructed and free space image after denoising (AD)

Wall Scenario		Denoising Metric (NMSE)	Number of Frames				
			1	5	10	20	30
Train and test on same wall (Case 1)	Dielectric	BD	3.60	3.50	3.18	3.49	3.39
		AD	1.05	0.85	0.99	0.92	1.56
	Reinforced	BD	5.71	5.91	5.57	5.60	5.01
		AD	1.60	1.09	1.08	1.16	1.36
	Wall With Air-gaps	BD	4.42	4.01	3.50	3.25	2.86
		AD	1.14	1.18	1.36	1.32	1.32
Train on multiple walls (Case 3)		BD	5.01	4.49	4.07	4.13	3.78
		AD	0.71	1.16	1.26	1.60	1.58

due to the possible decorrelation between test and training data. However, the deterioration in the performance is very slight. The performance indicates that this algorithm is specifically suited for imaging continuous and periodic motions such as walking. Also note that NMSE and SSIM do not behave in an identical manner for all the cases as they indicate different aspects of similarity of images. In practise, this case seems to be limited since in

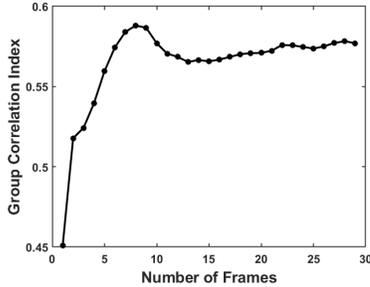


Fig. 8: Group correlation across multiple frames

real world scenarios, we may not know the type of wall available during the test phase. Therefore, we consider a significantly more challenging scenario where there is no information of the type of wall during the test phase.

Case 2: Train using data from a single wall type and test using data from a different wall type:

Next, we analyze the performance of the algorithms when the network is trained on images captured from a dielectric wall and tested with images captured from a reinforced wall type. The resulting denoising performance reported in Table.II deteriorates considerably as the number of frames increases. The reason for the poor performance is because the nature of the clutter in the dielectric case is quite different from that of the reinforced wall case. The algorithm is unable to denoise the clutter due to the lack of diversity in the training data. Henceforth, we do not report results for this case.

Case 3: Train using data from multiple types of walls and test on data from a single wall:

To overcome the limitation of the previous case, we train the

network on images captured from all three through-wall scenarios resulting in total training data set size of $[8464 \times 14400]$. Then, the data from any of these walls is randomly chosen for test. This is a significantly more challenging scenario since no information of the type of wall is available during the test phase. Note that the images used in the test phase have not been used during training. Our algorithm is specifically suited for problems dealing with a great deal of diversity in the target and channel conditions (different wall scenarios). The results, show a very good performance (NMSE and SSIM) comparable to that of the same wall scenarios. Thus, the performance of the algorithms depends on the diversity of test data provided while training.

We performed additional experiments to benchmark the performance of our algorithm with a popular wall clutter mitigation technique based on singular value decomposition (SVD) presented in [40], [41]. In [41], multichannel broadband data were processed to obtain range-DOA images and the objective was to remove static wall clutter. The hypothesis was that the target and wall reflectivities lie in different sub-spaces. Since the wall reflections are typically stronger than target reflections, the first few singular vectors obtained from SVD span the wall subspace while the remaining vectors span the target (human) and noise subspace. We follow the same assumption and reconstruct the frontal images after removing the wall response, arising from the top few eigen values, from the received signal. Then we process the data to obtain the frontal images. The results presented in TableIV clearly indicate that the SVD based approach fails to denoise the radar images in both the homogeneous wall and heterogeneous wall scenarios. This can be attributed to the fact that in our case the target returns are dependent upon the wall characteristics. Therefore, these returns cannot be separated into two sub-spaces using the SVD based approach.

C. Impact of radar-target aspect angle

In order to understand the generality of the proposed denoising solution, we trained our autoencoder network with human radar images captured at different aspect angles with respect to radar line-of-sight conditions. We analyzed the performance of our algorithm for four aspect angles- 0° , 45° , 90° , 180° . The time-

TABLE IV: Denoising results between clean and corrupted Doppler enhanced frontal images for different through-wall conditions. SSIM: between corrupted and free space image before denoising (BD) and SSIM :between reconstructed and free space image after denoising ,AD (proposed) using our proposed approach and AD (base) based on SVD technique.

Wall Scenario		Denoising Metric (SSIM)	Number of Frames				
			1	5	10	20	30
Single wall	Dielectric	BD	0.09	0.55	0.63	0.67	0.64
		AD (SVD)	0.01	0.08	0.38	0.35	0.30
		AD (Proposed)	0.43	0.87	0.90	0.89	0.86
	Reinforced	BD	0.05	0.24	0.46	0.55	0.47
		AD (SVD)	0.01	0.07	0.28	0.13	0.06
		AD (Proposed)	0.43	0.81	0.87	0.84	0.80
	Wall With Air-gaps	BD	0.03	0.01	0.34	0.51	0.43
		AD (SVD)	0.001	0.05	0.12	0.34	0.20
		AD (Proposed)	0.42	0.82	0.83	0.82	0.77

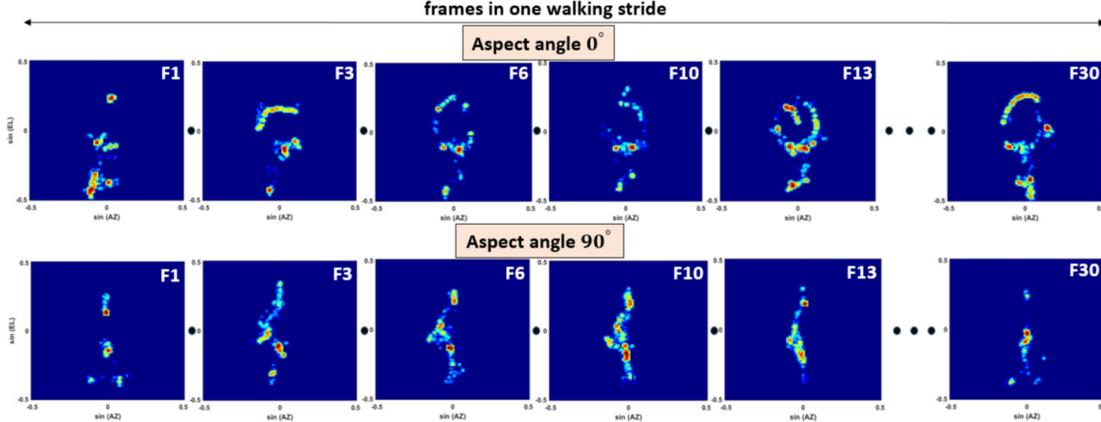


Fig. 9: Simulated Doppler enhanced frontal images over one walking stride of human motion for different aspect angles

TABLE V: Denoising results between clean and corrupted images (captured behind reinforced wall) for different aspect angles. SSIM, NMSE: between corrupted and free space image before denoising (BD) and SSIM, NMSE: between reconstructed and free space image after denoising (AD)

Wall Scenario (Reinforced)		Denoising Metric		
		SSIM	NMSE	
Aspect Angle	0°	BD	0.05	5.64
		AD	0.64	1.16
	45°	BD	0.36	4.01
		AD	0.73	1.02
	90°	BD	0.01	4.33
		AD	0.60	1.67
	180°	BD	0.46	5.01
		AD	0.80	1.23
	0°, 45°, 90°, 180°	BD	0.41	4.60
		AD	0.71	1.44

varying Doppler-enhanced frontal images generated in free space for 0° and 90° are shown in Fig.9. These can be compared to the frontal images for 180° that were shown earlier in Fig.6. Here 0° aspect angle means the person is walking towards the radar and 180° means the person is walking away from the radar. Similarly 90° corresponds to the motion along the tangential direction to the radar. We studied the efficacy of the algorithm for reinforced wall which, as mentioned earlier, is one of the most complex walls. We tested the performance of the denoising autoencoder on two scenarios: In the first scenario, both the training and test data are gathered at the same aspect angle (identical training and test scenario); In the second scenario, we used images captured at multiple different aspect angles for both training and testing the autoencoder. Therefore, during test, the algorithm is not provided information of the aspect angle of the data. We used both NMSE and SSIM to measure the performance and report the results in Table V. We observe highest error when the algorithm is trained with data captured at 90° aspect angle that is when the human

walks in a direction tangential to the radar. This is most likely because of the inherent distortions in these frontal images due to the limited separation of point scatterers on the subject along the azimuth direction which can be clearly seen from Fig.9. Likewise, the Dopplers of the different point scatterers on the human body are not well resolved due to the tangential motion. The results reported for all the aspect angles show significant improvement after denoising. When we consider data from multiple aspect angles, the denoising significantly helps in reconstructing images close to free space images even when the algorithm is not provided any information of the exact aspect angle at which a person is walking. Therefore we can infer that the autoencoder is specifically suited for problems dealing with a great deal of diversity in the target and channel conditions. It can significantly denoise (i) images captured in similar and dissimilar- wall conditions as well as (ii) images captured at different aspect angles of the target provided there is sufficient diversity across training data.

D. Aperiodic human crouch to run motion

We performed additional simulations to evaluate the performance of algorithm when human undergoes a non-periodic motion - when a human transitions from a crouch to a run towards the radar. Fig.10 shows the micro-Doppler signature of this motion. Since our FDTD simulation spatial extent is limited, we removed translational motion of the human and only retained the dynamics of the swinging arms and legs. Here the torso velocity is mostly around 0 since there is no translational motion of the human. We considered 40 consecutive frames of a human subject over a duration of 1.2s. The transition between the motion states can be clearly seen from the frontal images and their stick figure counterparts. This motion is considerably more challenging than the simple human walking motion. As the number of frames increase, there is enough diversity in the motion which is also

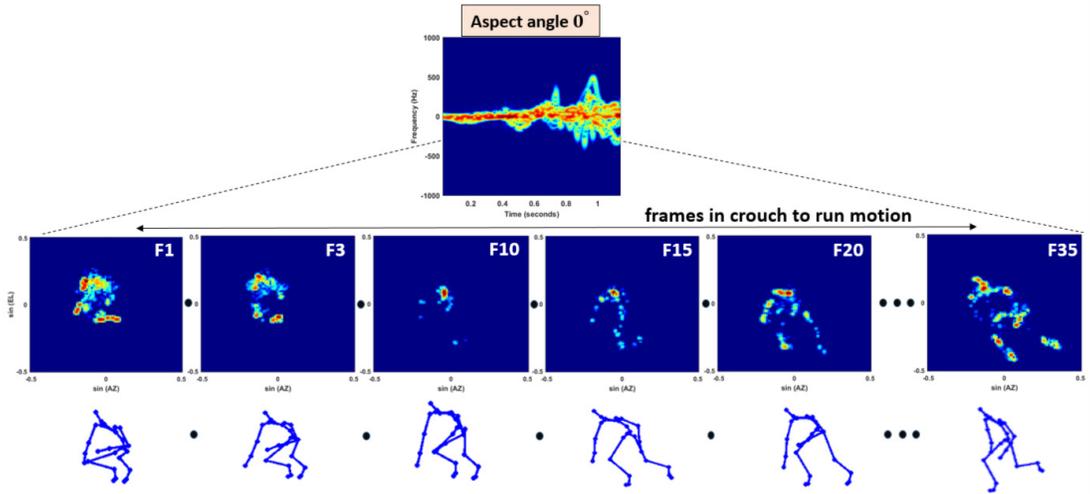


Fig. 10: Simulated Doppler enhanced frontal images over a crouch to run motion of a human and the corresponding stick figures obtained using ground truth motion capture data

reflected in the group correlation index across multiple frames shown in Fig.11(a). We tested the performance of our algorithm on data set captured behind a reinforced wall (a complex wall). Fig.11(b) shows the denoising performance of the proposed algorithm on the complex data set. The figure shows that the denoising algorithm (using the linear mapping function) results in significant improvement in the SSIM after denoising until approximately 15 frames. Beyond this, the performance of the denoising algorithm deteriorates due to the possible decorrelation between test and training data which is also reflected in Fig.11(a). Thus we conclude that the performance depends on how well the algorithm is trained to handle diversity in the test data.

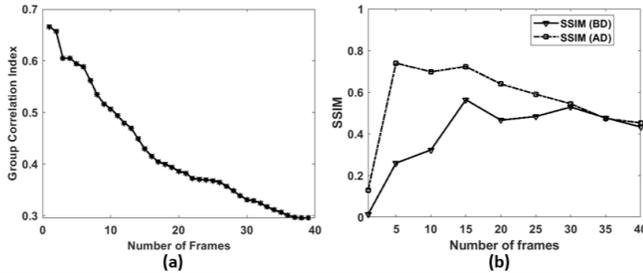


Fig. 11: Simulated human (a) Group correlation across multiple frames and (b) SSIM variation for simulation results with respect to number of frames

V. MEASUREMENT RESULTS

A. Measurement Data Collection

In this section, we evaluate the performance of our algorithm using wideband measurement data captured in both free space and through-wall conditions. The data is collected using Walabot Pro [34], a wideband(3.3-10.3GHz) 3D-programmable RF imaging sensor. Walabot is a low power uncalibrated sensor with limited range in through-wall scenarios. It uses a 4×4 antenna array to illuminate the area in front of it to capture the back-scattered signals. The hardware radar parameters are listed in Table.VI. The range enhanced images are obtained in manner similar to (19) by replacing the Doppler dimension with the range dimension. Here, the peak scatterers across all the range gates are coherently summed to obtain the frontal images of the targets. These steps are performed within the inbuilt processor in the sensor and we

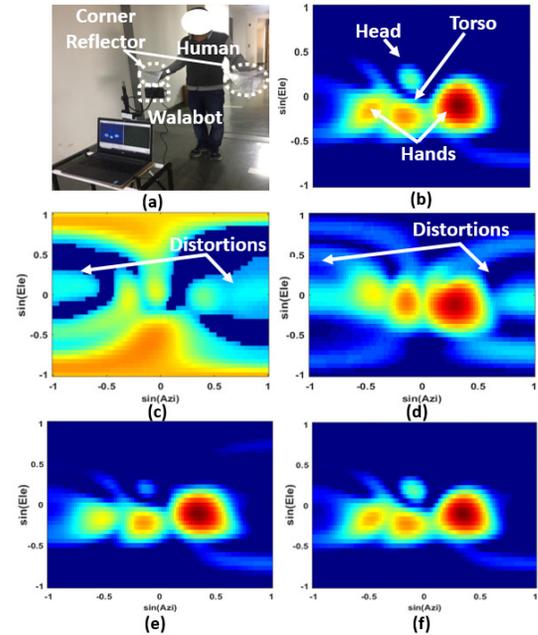


Fig. 12: (a) Measurement setup in free space and measured range enhanced frontal image of a human subject in (b) free space, (c),(d) behind a glass wall, wood wall respectively, (e),(f) Denoised images in through wall scenarios-glass Wall, wood wall respectively using the proposed algorithm.

are provided with output range-enhanced images. The assumption here, is that the targets are still or slow moving.

Our clean measurement data consist of range enhanced frontal images of a human gathered in line-of-sight conditions in an environment mostly free of clutter. The through-wall measurement data comprise of images captured through a 2cm thick glass wall and a 3cm thick wooden wall. The subject stands in front of the radar at a standoff distance of 2m carrying two corner reflectors covered with aluminum tape to enhance the reflectivity from the hands as shown in Fig 12(a). Therefore, the target is an extended target with multiple point scatterers. The experiments are performed on 4 human subjects of different heights and girth at different orientations (-45° to $+45^\circ$) with respect to the radar but always facing the radar. For each of these subjects, we captured

TABLE VI: Parameters of Real Radar Setup. *- values derived from available information

Radar Parameters	Values
Radar Type	Broadband
Bandwidth	3.3GHz – 10.3GHz
Maximum Range (R_{max})	10m*
Range resolution (Δr)	0.2m*
ADC	8 bit
Number of antenna elements ($P \times Q$)	4 × 4
Azimuth Beamwidth ($\Delta\phi_{azi}$)	25°*
Azimuth Beamwidth ($\Delta\theta_{ele}$)	25°*
Field of View (ϕ_{azi})	-90° to 90°*
Field of View (θ_{azi})	-90° to 90°*

75 measurements resulting in a total of 300 images of which 80% are used for training and remaining for test. An example of the resultant radar image in free space is shown in Fig.12(b) where we can clearly discern the torso, legs and two arms of the human. Radar images are corrupted when the measurements are gathered under different through-wall conditions. Some examples of the distortions are presented in Fig.12. Fig.12(c)-(d) correspond to images in through-glass wall and through-wood wall conditions respectively. These images are considerably distorted due to the complex interaction between the wall and the target. Each image of size $[91 \times 37]$ is vectorised to obtain a $[3367 \times 1]$ vector. Then all images are clubbed together to form a training data matrix of size $[3367 \times 240]$ and test data matrix of size $[3367 \times 60]$. Once trained, the weight matrices \mathbf{W}_1 and \mathbf{W}_2 are used to denoise the corrupted test images using the equation (10). Analogous to simulations, we examine the variation of denoising performance for different number of nodes in the hidden layer, for different mapping functions (linear, non-linear- tanh and sigmoid) and for different proportions of training to test data.

B. Measurement Results and Analyses

We tested the performance of the denoising autoencoder for *same wall* and *different wall* scenarios. In the same wall scenario, both the training and test data are gathered from the same type of wall. In the different wall scenario, data from multiple walls are used for training the autoencoder which is subsequently used for denoising images from any of the two walls. The reconstruction results are presented as a function of percentage of training data to test data in Table VII. These results have been obtained using an autoencoder where the hidden layer has 1500 nodes and the mapping function is sigmoid. The choice of these parameters are discussed in the appendix. The table shows SSIM and NMSE between the denoised radar images in through-wall and corresponding radar image gathered in free space conditions. We compare the metrics before denoising (BD) with those after denoising (AD). We observe there is significant improvement in SSIM and reduction of NMSE after denoising. The performance improves as the percentage of training to test data increases for both the same wall and for different wall scenarios. In other words, the performance during test relies on adequate training data. The error for the different wall scenario is only slightly higher than the same wall scenario. This is the scenario when the test algorithm has no knowledge of the wall scenario. Note that in the case of the wideband measurements, we have not presented the result as a function of the number of frames. This is because, the targets are static and each measurement is independent with no correlation between them. Please note that the SVD based algorithm is predicated on the availability of raw data. Since our measurement results are images directly obtained from the Walabot sensor - a cheap sensor that does not provide raw data - we could not apply the SVD on the measurement data.

VI. DISCUSSION ON RESULTS

A. Computational Complexity Evaluation

The real-time performance of the algorithm relies on the test time and the test memory rather than training time. During test, we perform matrix multiplication operations of the trained weights \mathbf{W}_1 and \mathbf{W}_2 with test image test . The sizes of the weight matrices and the image matrix are $R \times N$, $N \times R$ and $N \times 1$ respectively, where N denotes the number of pixels in the image and R denotes the number of hidden nodes in the autoencoder such that the number of nodes is always well below the number of pixels. The computational complexity therefore is $\mathcal{O}(RN)$. We ran our algorithm on Matlab 2015b, where all the variables were stored as 64 bit floats, with an Intel(R) Core(TM) i7-5500 processor running at 2.40 GHz. We report the test and training times of our algorithms as a function of the number of nodes of the hidden layers in Fig.13. Both the training and test times are higher when R is greater. The test time is significantly low even for the highest number of hidden nodes (1500). The computational memory in all of these cases was less than 500MB. Therefore, these test operations can be carried out in easily available processors such as Raspberry PI 3+ (with a 1GB RAM and 1.4GHz clock speed).

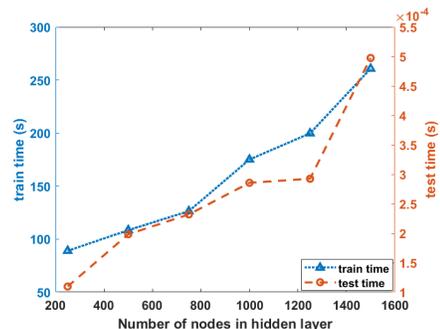


Fig. 13: Computational time as a function of number of nodes in the hidden layer for (left y-axis) training phase (right y-axis) test phase

B. Diversity of training data

The training data must be sufficiently large to handle the diversity of target conditions, channel conditions and any type of labelling errors between free space and through-wall images.

Diversity of target data: In our work, our autoencoder has been trained to handle the diversity in the size, shape and orientation or aspect angle of the target with respect to the radar. In the case of dynamic motions, the correlation between consecutive frames facilitated in improving the denoising performance.

Diversity of channel data: Next, the proposed approach does not require the knowledge of the exact wall conditions or analytical framework during the test phase. Instead, the algorithm was capable of denoising images obtained from diverse through-wall conditions.

Labelling errors between free-space and through-wall images: Finally, in practice, it may be nearly impossible to gather correlated images in free space and through-wall conditions especially for dynamic targets. For example, it may not be possible to replicate human motions in two different scenarios. Therefore, the algorithm must tolerate some degree of diversity in the motion characteristics during test and training phases. A sensitivity analysis of mismatch/labelling error between clean (free space) and the corrupted (through-wall) training images is not considered in this work. Generally in machine learning scenarios, these algorithms are quite robust to reasonable random errors in the training set arising due to incorrectly labelled data. However, the

TABLE VII: Denoising results between clean and corrupted measurement images of real humans for different through-wall conditions under varying percentage of training data. NMSE, SSIM :between corrupted and free space image before denoising (BD) and NMSE,SSIM: between reconstructed and free space image after denoising (AD)

Wall Scenario			Denoising Metric							
			SSIM				NMSE			
			(% of Training Data)				(% of Training Data)			
			20	40	60	80	20	40	60	80
Train and test on same wall	Glass Wall	BD	0.20	0.20	0.22	0.21	36.09	38.47	33.52	38.70
		AD	0.53	0.70	0.88	0.97	7.81	5.10	4.68	3.69
	Wood Wall	BD	0.22	0.22	0.23	0.23	28.21	29.04	27.51	33.76
		AD	0.50	0.55	0.82	0.91	7.59	7.10	4.59	4.24
Train on multiple walls		BD	0.23	0.22	0.23	0.22	29.00	28.63	28.43	29.49
		AD	0.46	0.59	0.70	0.89	8.40	7.39	6.19	4.97

algorithms are less robust to systematic errors when the samples are consistently incorrectly labelled.

VII. CONCLUSION

We demonstrate the efficacy of the denoising autoencoder network at mitigating the distortions and clutter introduced by wall propagation on radar images of humans. The proposed approach requires neither prior information of the wall characteristics nor any kind of analytic framework to describe the wall propagation effects. Instead, the algorithm relies on the availability of a huge training data set comprising of distorted radar images captured in diverse through-wall scenarios and the corresponding clean images in line-of-sight conditions. Once trained, the algorithm is capable of mitigating through-wall effects of similar walls though not necessarily identical walls. This capability makes this approach suitable for tracking humans under diverse propagation environments. We evaluated the performance of the algorithm on both static and dynamic targets. The radar images of dynamic humans were simulated using Doppler-enhanced array processing while the images of the static humans were generated from measurement data using range-enhanced array processing. Before denoising, the images were considerably distorted by through-wall propagation effects. Our algorithm showed that after denoising, the images were structurally similar with low mean square error with respect to the free space images.

ACKNOWLEDGEMENT

This work is sponsored by the Ministry of Electronics and Information Technology, Govt. of India, under the Visvesvaraya PhD Scheme and 5IOA036 FA23861610004 grant by Air Force Office of Scientific Research (AFOSR), AOARD.

APPENDIX

A. Hyper-parameter Selection

We optimized the number of nodes in the hidden layer and mapping functions to obtain the results presented in the previous sections. First we discuss the autoencoder used on the simulation data. We simulated narrowband time domain returns at a carrier frequency of 7.5GHz with the gains of the antennas to be 10dBi, and the transmitted power at +30dBm. As a result, the maximum received signal strength from the human subject (at R = 2.5m) is -57dBm. Therefore, all the pixel values in the simulated images are negative. Fig.14a shows the performance for different mappings- linear, tanh, sigmoid - as a function of number of frames. We observe that the linear and tanh mapping outperform the results obtained using sigmoid mapping functions. This is because both sigmoid and hyperbolic tangent (tanh) functions - are monotonically increasing functions that asymptote at $\pm\infty$. However, the tanh function is symmetric about the origin and produce outputs values between [-1 1] while sigmoid function outputs are always positive [0 1]. The sigmoid mapping function

is thus not suited to handle the negative signal strength in the radar images due to its asymptotic behavior. The linear mapping function is best suited for this case. Our studies also showed that using similar mapping functions in the encoder and decoder result in better performances when compared to the use of dissimilar mapping functions. Fig.14(b), show the variation of SSIM before and after denoising as a function of the number of nodes in the hidden layer. We observe that the performance converges when the number of nodes is approximately 500.

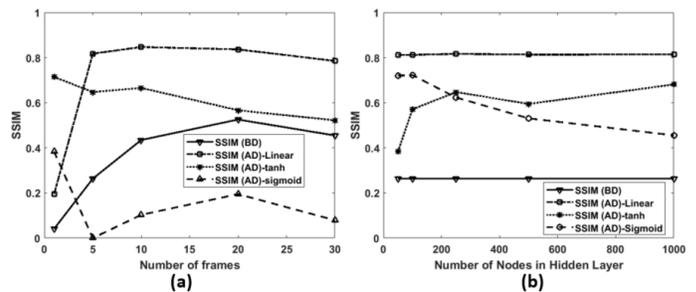


Fig. 14: SSIM variation for simulation results with respect to (a) number of frames, (b) number of nodes in the hidden layer for mapping functions-linear, tanh and sigmoid

Next, we discuss the autoencoder used on the measurement data of real humans in both line-of-sight and through-wall conditions. The measurement data are collected using the Walabot that has an ADC of 8 bits. The digitized data are calibrated to a voltage scale from 0 to 5V. In the radar images, the pixel values are converted to the logarithmic scale and hence consist of both negative and positive values. Due to the dynamic range of the pixel energy values, the sigmoid function is now able to handle the denoising and we get excellent results in Fig.15. The results are in perfect accordance with our previous hypothesis that the performance of these activation functions is sensitive to the signal

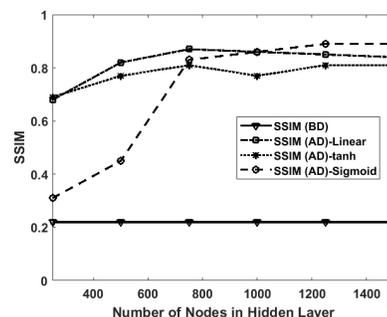


Fig. 15: SSIM variation for measurement results with respect to number of nodes in the hidden layer for mapping functions-linear, tanh and sigmoid for human subjects

strength and dynamic range of the pixels in the radar images. In fact, the non-linear mapping functions are slightly superior to the linear mapping function. The linear mapping function is able to handle the non-linearity in the wall response since the images are inherently sparse. The SSIM varies as a function of the number of nodes in the hidden layer for different mapping functions in Fig. 15. The SSIM improves and tends towards 1 as we increase the number of nodes in the hidden layer to 1500.

REFERENCES

- [1] M. G. Amin, *Through-the-wall radar imaging*. CRC press, 2016.
- [2] R. M. Narayanan, M. C. Shastri, P.-H. Chen, and M. Levi, "Through-the-wall detection of stationary human targets using doppler radar," *Progress In Electromagnetics Research B*, vol. 20, pp. 147–166, 2010.
- [3] S. S. Ram and H. Ling, "Through-wall tracking of human movers using joint doppler and array processing," *IEEE Geoscience and Remote Sensing Letters*, vol. 5, no. 3, pp. 537–541, 2008.
- [4] S. S. Ram, C. Christianson, Y. Kim, and H. Ling, "Simulation and analysis of human micro-dopplers in through-wall environments," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 4, pp. 2015–2023, April 2010.
- [5] C.-P. Lai, R. Narayanan, Q. Ruan, and A. Davydov, "Hilbert-huang transform analysis of human activities using through-wall noise and noise-like radar," *IET Radar, Sonar & Navigation*, vol. 2, no. 4, pp. 244–255, 2008.
- [6] Y. Jia, G. Cui, L. Kong, and X. Yang, "Multichannel and multiview imaging approach to building layout determination of through-wall radar," *IEEE Geoscience and Remote Sensing Letters*, vol. 11, no. 5, pp. 970–974, 2014.
- [7] C. Clemente, A. Balleri, K. Woodbridge, and J. J. Soraghan, "Developments in target micro-doppler signatures analysis: radar imaging, ultrasound and through-the-wall radar," *EURASIP Journal on Advances in Signal Processing*, vol. 2013, no. 1, p. 47, 2013.
- [8] S. S. Ram and A. Majumdar, "High-resolution radar imaging of moving humans using doppler processing and compressed sensing," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 51, no. 2, pp. 1279–1287, 2015.
- [9] —, "Through-wall propagation effects on doppler-enhanced frontal radar images of humans," in *Radar Conference (RadarConf)*, 2016 IEEE. IEEE, 2016, pp. 1–6.
- [10] S. S. Ram, "Doppler enhanced frontal radar images of multiple human activities," in *Radar Conference (RadarCon)*, 2015 IEEE. IEEE, 2015, pp. 1166–1171.
- [11] F. Ahmad, M. G. Amin, and G. Mandapati, "Autofocusing of through-the-wall radar imagery under unknown wall characteristics," *IEEE transactions on image processing*, vol. 16, no. 7, pp. 1785–1795, 2007.
- [12] M. Leigsnering, F. Ahmad, M. Amin, and A. Zoubir, "Multipath exploitation in through-the-wall radar imaging using sparse reconstruction," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 50, no. 2, pp. 920–939, 2014.
- [13] F. H. C. Tivive, A. Bouzerdoum, and M. G. Amin, "A subspace projection approach for wall clutter mitigation in through-the-wall radar imaging," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 4, pp. 2108–2122, 2014.
- [14] F. Ahmad and M. G. Amin, "Multi-location wideband synthetic aperture imaging for urban sensing applications," *Journal of the Franklin Institute*, vol. 345, no. 6, pp. 618–639, 2008.
- [15] M. Leigsnering, *Sparsity-Based Multipath Exploitation for Through-the-Wall Radar Imaging*. Springer, 2018.
- [16] V. H. Tang, A. Bouzerdoum, and S. L. Phung, "Multipolarization through-wall radar imaging using low-rank and jointly-sparse representations," *IEEE Transactions on Image Processing*, vol. 27, no. 4, pp. 1763–1776, 2017.
- [17] V. H. Tang, S. L. Phung, F. H. C. Tivive, and A. Bouzerdoum, "A sparse bayesian learning approach for through-wall radar imaging of stationary targets," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 53, no. 5, pp. 2485–2501, 2017.
- [18] S. Vishwakarma and S. S Ram, "Detection of multiple movers based on single channel source separation of their micro-dopplers," *submitted to IEEE Transactions on Aerospace and Electronic Systems*, 2016.
- [19] A. Martone, K. Ranney, and R. Innocenti, "Through-the-wall detection of slow-moving personnel," in *Radar Sensor Technology XIII*, vol. 7308. International Society for Optics and Photonics, 2009, p. 73080Q.
- [20] F. Ahmad and M. G. Amin, "Through-the-wall human motion indication using sparsity-driven change detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 51, no. 2, pp. 881–890, 2013.
- [21] P. Setlur, M. Amin, and F. Ahmad, "Multipath model and exploitation in through-the-wall and urban radar sensing," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 10, pp. 4021–4034, 2011.
- [22] P. Setlur, G. Alli, and L. Nuzzo, "Multipath exploitation in through-wall radar imaging via point spread functions," *IEEE Transactions on Image Processing*, vol. 22, no. 12, pp. 4571–4586, 2013.
- [23] P. Vincent, H. Larochelle, Y. Bengio, and P.-A. Manzagol, "Extracting and composing robust features with denoising autoencoders," in *Proceedings of the 25th international conference on Machine learning*. ACM, 2008, pp. 1096–1103.
- [24] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *science*, vol. 313, no. 5786, pp. 504–507, 2006.
- [25] M. Sakurada and T. Yairi, "Anomaly detection using autoencoders with nonlinear dimensionality reduction," in *Proceedings of the MLSDA 2014 2nd Workshop on Machine Learning for Sensory Data Analysis*. ACM, 2014, p. 4.
- [26] R. Socher, E. H. Huang, J. Pennin, C. D. Manning, and A. Y. Ng, "Dynamic pooling and unfolding recursive autoencoders for paraphrase detection," in *Advances in neural information processing systems*, 2011, pp. 801–809.
- [27] M. Chen, Z. Xu, K. Weinberger, and F. Sha, "Marginalized denoising autoencoders for domain adaptation," *arXiv preprint arXiv:1206.4683*, 2012.
- [28] S. Vishwakarma, V. Ummalaneni, M. S. Iqbal, A. Majumdar, and S. S. Ram, "Mitigation of through-wall interference in radar images using denoising autoencoders," in *Radar Conference (RadarConf18)*, 2018 IEEE. IEEE, 2018, pp. 1543–1548.
- [29] B. Du, W. Xiong, J. Wu, L. Zhang, L. Zhang, and D. Tao, "Stacked convolutional denoising auto-encoders for feature representation," *IEEE transactions on cybernetics*, vol. 47, no. 4, pp. 1017–1027, 2017.
- [30] I. Sutskever, J. Martens, G. Dahl, and G. Hinton, "On the importance of initialization and momentum in deep learning," in *International conference on machine learning*, 2013, pp. 1139–1147.
- [31] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends® in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011.
- [32] K. Yee, "Numerical solution of initial boundary value problems involving maxwell's equations in isotropic media," *IEEE Transactions on antennas and propagation*, vol. 14, no. 3, pp. 302–307, 1966.
- [33] S. M. Smith and C. Furse, "Stochastic fdfd for analysis of statistical variation in electromagnetic fields," *IEEE Transactions on Antennas and Propagation*, vol. 60, no. 7, pp. 3343–3350, 2012.
- [34] 2017. [Online]. Available: <https://walabot.com/docs/walabot-tech-brief-416.pdf>
- [35] C. M. Bishop, "Pattern recognition," *Machine Learning*, vol. 128, pp. 1–58, 2006.
- [36] Z. Wang and A. C. Bovik, "A universal image quality index," *IEEE signal processing letters*, vol. 9, no. 3, pp. 81–84, 2002.
- [37] A. Hore and D. Ziou, "Image quality metrics: Psnr vs. ssim," in *2010 20th International Conference on Pattern Recognition*. IEEE, 2010, pp. 2366–2369.
- [38] T. Goldstein and S. Osher, "The split bregman method for l1-regularized problems," *SIAM journal on imaging sciences*, vol. 2(2), pp. 323–343, 2009.
- [39] A. Gogna and A. Majumdar, "Discriminative autoencoder for feature extraction: Application to character recognition," *Neural Processing Letters*, vol. 49, no. 3, pp. 1723–1735, 2019.
- [40] F. H. C. Tivive, A. Bouzerdoum, and M. G. Amin, "An svd-based approach for mitigating wall reflections in through-the-wall radar imaging," in *2011 IEEE RadarCon (RADAR)*. IEEE, 2011, pp. 519–524.
- [41] F. H. C. Tivive, M. G. Amin, and A. Bouzerdoum, "Wall clutter mitigation based on eigen-analysis in through-the-wall radar imaging," in *Digital Signal Processing (DSP), 2011 17th International Conference on*. IEEE, 2011, pp. 1–8.

Shelly Vishwakarma received her Ph.D. degree in electrical engineering from Indraprastha Institute of Information Technology Delhi. She is currently a research fellow with Department of Security and Crime Sciences, UCL.

Shobha Sundar Ram completed her Masters and PhD from the University of Texas at Austin, in Electrical and Computer Engineering in 2006 and 2009 respectively. She is currently an Assistant Professor at the Indraprastha Institute of Information Technology (IIIT) Delhi.