

A Deep Learning Approach for Reconstruction in Millimeter-Wave Imaging Systems

Peyman Rostami, Hojatollah Zamani, Mohammad Fakharzadeh, *Senior Member, IEEE*, Arash Amini, *Senior Member, IEEE*, Farokh Marvasti, *Life Senior Member, IEEE*,

Abstract—In millimeter-wave (MMW) imaging, the objects of interest are oftentimes modeled as 2D binary (black and white) shapes with white pixels representing the reflecting interior of the object. However, due to propagation of the scattered waves, the continuous-domain binary images are convolved with a so-called point-spread function (PSF) before being digitized by means of sampling. As the 2D PSF is both non-separable and non-vanishing in the case of MMW imaging, exact recovery is quite complicated. In this paper, we propose a deep learning approach for image reconstruction. We should highlight that the wave scatterings are suitably represented with complex-valued quantities, while standard deep neural networks (DNN) accept real-valued inputs. To overcome this challenge, we separate the real and imaginary parts as if we had two imaging modalities and concatenate them to form a real-valued input with larger size. Fortunately, the network automatically learns how to combine the mutual information between these modalities to reconstruct the final image. Among the advantages of the proposed method are improved robustness against additive noise and mismatch errors of imaging frequency and object to antenna distance; indeed, the method works well in wide-band imaging scenarios over a wide range of object to antenna distances even in presence of high noise levels without requiring a separate calibration stage. We test the method with synthetic data simulated with software as well as real recordings in the laboratory.

Index Terms—Deep neural networks, inverse problem, millimeter wave imaging, reconstruction, wide-band imaging.

I. INTRODUCTION

MILLIMETER-WAVE (MMW) imaging systems are widely used in different applications such as airport security, non-destructive tests, medical diagnosis and through-wall imaging [1]–[3]. Millimeter waves can penetrate through thin dielectric layers such as plastic, wood and clothing, but are reflected from metallic objects and human body, which makes this band suitable for radar imaging in detecting flaws and concealed objects [4], [5].

In wideband imaging systems, the transmitter antenna sequentially sweeps all frequencies for each object pixel. Based on the geometric position of the TX and RX (i.e. the transmitter and the receiver antennas), the imaging system is categorized as either mono-static (same position for TX and RX) or multi-static (different positions for TX and RX). In this work, we focus on the mono-static case; however, the extension to the multi-static case is straightforward.

The use of deep learning in the literature of millimeter wave (MMW) imaging systems is mainly limited to object detection. For the microwave image reconstruction, a simplified training strategy for deep neural networks is introduced in [6] in which the reconstruction network differs from the trained network.

For 1-D electromagnetic signal inversion problem, an enhanced method of training data generation is studied in [7]. In [8], for solving the inverse scatter problem, traditional iterative techniques are combined with recent convolutional neural network (CNN) architectures. More recently in [9], human pose segmentation alongside a deep CNN is employed to distinguish abnormal patterns in human body parts that may reflect the presence of suspicious objects. The similarity with human body clutter is considered into the training

phase for more robustness. By thresholding the reconstructed image, an object segmentation method is introduced in [10]. For tuning the segmentation threshold, the image histogram is evaluated using a Gaussian mixture model. The expectation maximization (EM) algorithm is then applied to determine the threshold. Evidently, this method is vulnerable to the model mismatch. A deep learning approach for object detection in MMW images is introduced in [11]. While quite stable even in low quality inputs, this method suffers from high computational complexity because of using multiple classifiers. In another work in [12], based on a collected dataset of human millimeter wave images, a faster RCNN is employed alongside the thresholding segmentation technique for object detection. In a different approach, [13] uses the YOLOv3 network for object detection; the outcome is a real-time and highly accurate method.

In [14], a fully convolutional network (FCC) with a general auto-encoder structure is proposed to enhance the recovery quality in CS-ISAR imaging. Unlike MMW imaging, in this modality the measurements are real-valued. Besides, as the method employs compressed sensing (CS) techniques, the quality of the result depends on the sparsity level of the original image.

A neural network structure for an inverse scattering problem with complex-valued input is introduced in [15] that works with complex-valued weights. However, it is more of a hybrid solution in which a traditional linear MMW imaging algorithm is first employed to provide an initial reconstruction, and then, the complex-valued DNN would enhance the initial reconstruction. One issue with such hybrid methods is that they assume a linear approximation of the inverse scattering problem (which is highly non-linear). Further, the DNN does not have direct access to the measurements and cannot comprehensively learn the model. Also, because of the use of the traditional linear image recovery algorithm at the beginning, the method of [15] requires accurate estimates of imaging parameters such as the imaging frequency and the antenna to object distance for high quality reconstruction.

In [16], a convolutional neural network is devised to exploit prior information for improving the performance of a model-based microwave inverse scattering magnetic resonance (MR) imaging system.

A more conventional image reconstruction approach in MMW systems is to employ sparse representations. In [17], two-column sparse arrays with electronically switching antenna elements are designed to reduce sample redundancy. In [18] a hardware-software system is introduced to improve the quality of reconstruction; for this purpose, a dual polarized antenna array is designed to obtain both co- and cross- polarization and a PCA-based method is devised for reconstruction. Also, a random sample imaging mechanism is introduced in [19] that iteratively reconstructs the image by taking into account the sparsity of the hologram in a transform domain.

In this paper, we propose a deep learning reconstruction technique to approximate the inverse of the MMW imaging system. The complex-valued measurements at the receiver antenna are fed into our neural network to obtain a binary real-valued image as the output. As the input is complex-valued, similar to [20], we first decompose it into real and imaginary parts and stack them in depth before feeding

*Authors are with the EE Department of Sharif University of Technology, Tehran, Iran (e-mail:peyman.rostami@ymail.com, hojatollah.zamani@ee.sharif.edu, {fakharzadeh, amini, marvasti}@sharif.edu).

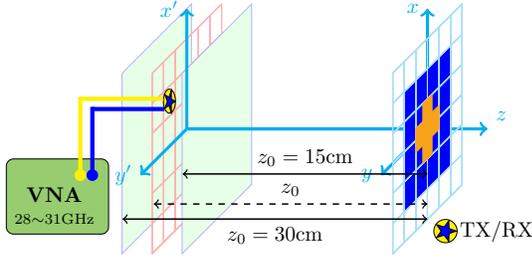


Fig. 1. The mono-static imaging system model. The TX antenna propagates EM waves towards the object and the reflected wave is measured by the RX antenna. An advantage of the method is that no range or frequency calibration is required. As long as the frequency belongs to the range 28 ~ 31GHz and the object is located in 15 ~ 30cm distance to the RX antenna, the method can blindly reconstruct the desired image.

to the network. The real and imaginary parts can be thought of as two imaging modalities with correlated information. It is the task of the network to optimally take advantage of the common information between these two inputs. As deep neural networks are quite capable in learning mixture models, our dataset includes wideband recordings (frequency range of 29^{GHz} to 31^{GHz}) with varying depths (object to antenna distances from 15^{cm} to 30^{cm}) to provide more robustness against model mismatch (other frequency values and depth ranges can also be used). As a result, unlike other MMW reconstruction algorithms which require accurate a priori knowledge of the imaging frequency and depth, our algorithm automatically handles this part which is an attractive option in practical scenarios.

The rest of this paper is organized as follows: In Section II, the imaging system model is described. The details of the designed network are presented in Section III. Simulations and experimental results can be found in Section IV. The advantages and the limitations of the proposed deep-learning-based method compared to traditional approaches are presented in Section IV-B.

II. MMW IMAGING SYSTEM MODEL

The MMW mono-static imaging system can be mathematically modeled as follows: a collocated pair of TX and RX antennas form a transceiver for scanning objects. When scanning the object plane (Fig. 1), the TX transmits millimeter waves towards the object plane and the reflected signals from the illuminated object are captured by the RX. If (x, y, z_0) is the location of an illuminated point on the target, we denote the reflection coefficient from this point in the direction of the transceiver position by $f(x, y)$. Due to the wave propagation phenomenon (assuming far field approximation), $f(x, y)e^{-j2kr}$ is received at the transceiver (normalized by the magnitude decay of $1/(4\pi r^2)$), where $k = 2\pi/\lambda$ is the wave number and r stands for the Euclidean distance between the point on the target and the transceiver. The overall scattered field at the receiver is the linear combination of the reflected waves from all object points, which can be represented as [5]

$$s(x', y') = \iint f(x, y)e^{-j2kr} dx dy, \quad (1)$$

where $s(x', y')$ is the measured signal by RX at position (x', y') (location of the transceiver) and $r = \sqrt{(x - x')^2 + (y - y')^2 + z_0^2}$. A common baseline approach for image recovery is the generalized synthesis aperture focusing technique (GSAFT) described in [5], which is summarized as

$$f(x, y) = \mathcal{F}^{-1}\{\mathcal{F}\{s(x', y')\}e^{jk_z z_0}\}, \quad (2)$$

where $\mathcal{F}\{\cdot\}$ and $\mathcal{F}^{-1}\{\cdot\}$ represent the 2-D Fourier transform and its inverse, respectively, and $k_z = \sqrt{4k^2 - k_x^2 - k_y^2}$. To model the imaging system in discrete domain, let the matrices $\mathbf{F}_{\mathbf{P} \times \mathbf{Q}}$ and

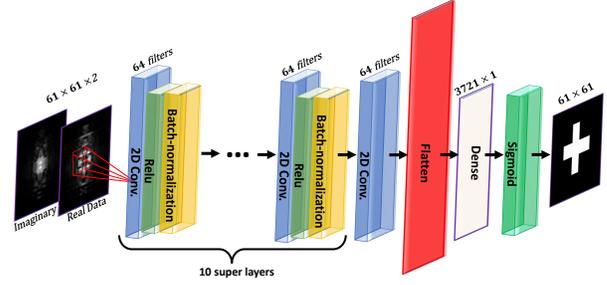


Fig. 2. The proposed deep neural network structure. The real and imaginary parts of the complex measurement matrix are stacked depth-wise and fed into the DNN structure which fuses their information to produce the image of the object at its output.

$\mathbf{S}_{\mathbf{M} \times \mathbf{N}}$ represent the reflection values of the object plane sampled on a grid and the measured data recorded on another grid, respectively. The discrete form of (1) is then given by

$$\mathbf{S}[m, n] = \sum_{p=1}^P \sum_{q=1}^Q \mathbf{F}[p, q] e^{-j2kr[m, n, p, q]}, \quad (3)$$

where $r[m, n, p, q]$ is the Euclidean distance between the transceiver position $[m, n]$ and the object point $[p, q]$. The wideband imaging model can be obtained by extending (3) to N_f frequencies ($f_i, i = 1, \dots, N_f$) as follows

$$\mathbf{S}[m, n, f_i] = \sum_{p=1}^P \sum_{q=1}^Q \mathbf{F}[p, q] e^{-j2\frac{2\pi}{c} f_i r[m, n, p, q]}. \quad (4)$$

III. NETWORK STRUCTURE AND TRAINING PROCEDURE

As explained earlier, our measurements from the object are complex-valued. Hence, for image reconstruction, we design a deep neural network that receives complex-valued input data and returns an image at the output. The network shall be trained by a dataset generated using (4), where \mathbf{S} stands for the measured data (input to the network) and \mathbf{F} represents the desired output of the network. We consider the object under imaging procedure to be represented by a binary (black and white) image, with white area indicating the interior of the object. The structure of the proposed network, the dataset preparation and the training procedure are explained in details in the following subsections.

A. Network Structure

The proposed neural network receives a matrix with 61×61 complex-valued measured entries (pixels) as the input, and returns an image matrix of the same size as the output. The designed network structure is depicted in Fig. 2. For the sake of simplicity, let us define a super-layer with three sequential sub-layers including a 2D convolutional layer with 64 filters of size 3×3 followed by a Relu activation layer, and a batch normalization layer. The complex-valued measurements are first decomposed into real and imaginary parts which are stacked in depth and fed into the DNN structure. The DNN consists of a set of ten super-layers followed by a 2D convolutional layer with 64 filters of size 3×3 and no activation function, before a flattening layer ahead of a dense layer with 3721 neurons and finally, a Sigmoid activation layer, the output of which is de-flattened to form the 61×61 -pixel output image. For the training process, we use Adam optimizer [21]. Binary cross entropy loss function is used to minimize the difference between the binary ground-truth and the estimated images

$$H_p = -\frac{1}{N} \sum_{i=1}^N y_i \log(p(y_i)) + (1 - y_i) \log(1 - p(y_i)), \quad (5)$$

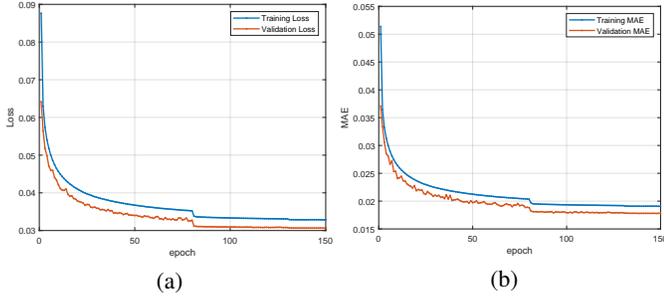


Fig. 3. (a) Training and validation loss, (b) MAE metric.

where y is the label and $p(y)$ is the predicted probability for all N points.

B. Dataset generation and training

The dataset that we used in this work includes 5671 binary images from several open access libraries: [22]–[25]. We split this amount into two subsets: 80% for training, and 20% for test purposes. In each subset, data augmentation is applied with the following strategy: in each subset, all binary images are cropped around their bounding boxes. Each image is then rotated 90, 180, and 270 degrees, and the rotated versions are added to the initial subset. In addition, by keeping the aspect ratio fixed, we scaled each image to $w \times h$ pixels, where $\max(w, h) = 35$ (yielding a $35 \times h$ or $w \times 35$ image). We form the ground-truth binary 61×61 -pixel images by placing each of the scaled images in 15 random locations inside the 61×61 -pixel image. The default value of the pixels is 0; therefore, only those pixels that belong to the interior of the [shifted] scaled shapes are non-zero. This data augmentation strategy increases the volume of the initial dataset with the factor of 60. For each ground-truth binary image in the augmented dataset, we generate its corresponding complex-valued measurements using the forward system model (4). The imaging frequency and the TX to object distance are set uniformly at random in the range 28GHz to 31GHz and 15cm to 30cm, respectively. We further include a complex-valued additive Gaussian noise with varying power such that the resulting signal to noise ratio (SNR) falls between 15dB and 20dB.

With a batch-size of 64, we train the neural network for 80 initial epochs using the learning rate 10^{-3} , another 50 epochs using the learning rate of 10^{-4} , and a final 20 epochs using the 10^{-5} learning rate. The training and validation loss curves (vs. epoch), as well as the mean absolute error (MAE) between the predicted and the desired outputs are shown in Fig. 3.

C. Calibration

Calibration is one of the essential procedures in most real world imaging systems that has a great impact on imaging quality. In most systems, we either need to tune or estimate a number of high-level parameters before operation. MMW imaging systems are no exception and require the knowledge about the exact imaging frequency as well as the distance between the antenna and the object of interest before every single shot. Oftentimes, obtaining such information is tedious and cumbersome. Our approach to overcome this issue is that we train the network on data coming from varying frequencies and distances in the operational range at once (Fig. 1). As a result, the trained network can automatically reconstruct an image from measurements corresponding to frequency range of 28 ~ 31GHz and distance range of 15 ~ 30cm without any further calibration stages. It should be noted that the mentioned ranges are chosen due to their popularity; they could be easily replaced with user-defined ranges.

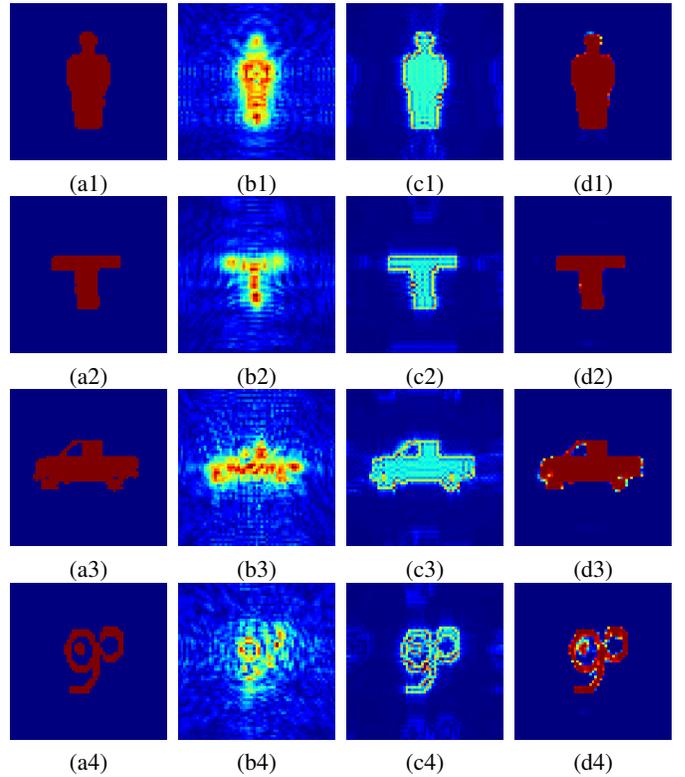


Fig. 4. The noiseless scenario. Columns from left to right correspond to the ground-truth images, measurements, reconstructed images using GSAFT, and the reconstructed images using the proposed method, respectively. The reconstruction qualities in terms of PSNR and SSIM are reported in Table I.

TABLE I
PSNR AND SSIM METRICS FOR IMAGES IN FIG. 4

Object	GSAFT		Proposed	
	PSNR(dB)	SSIM	PSNR(dB)	SSIM
(a1)	9.1733	0.6363	29.9469	0.9826
(a2)	11.2058	0.7137	37.3374	0.9929
(a3)	8.8795	0.5919	25.3446	0.9607
(a4)	11.7696	0.6782	24.8661	0.9820

IV. RESULTS AND DISCUSSIONS

A. Simulation and measurement Results

In this subsection, we study the result of the proposed method in reconstructing images from both noiseless and noisy measurements of various synthetic and real objects. In our experiments, we compare the proposed method with GSAFT [5] as a well-studied baseline in the literature of MMW image reconstruction, as well as a more recent reconstruction method called Delay Multiply and Sum Beamforming Algorithm (DMAS) [26]. We evaluate the reconstruction quality by SSIM metric [27] and PSNR metric which is defined as:

$$\text{PSNR} = 10 \log_{10} \left(\frac{\max(I)^2}{\frac{1}{MN} \sum_{m=1}^M \sum_{n=1}^N (I[m,n] - \hat{I}[m,n])^2} \right) \quad (6)$$

where I is the original image and \hat{I} is its reconstructed version.

In the first experiment, we consider noiseless synthetic measurements obtained in a wide-band mono-static imaging setup with $N_f = 10$ equally spaced frequency points in the range 28GHz to 31GHz, and the object depth (i.e. orthogonal distance between transceiver plane and object plane) set to 15cm. The pair of transmitter and receiver scan an area of 61×61 points and for each point record the data of all frequencies. The sampling interval on both the measurements and the imaging results are 5 millimeters across the

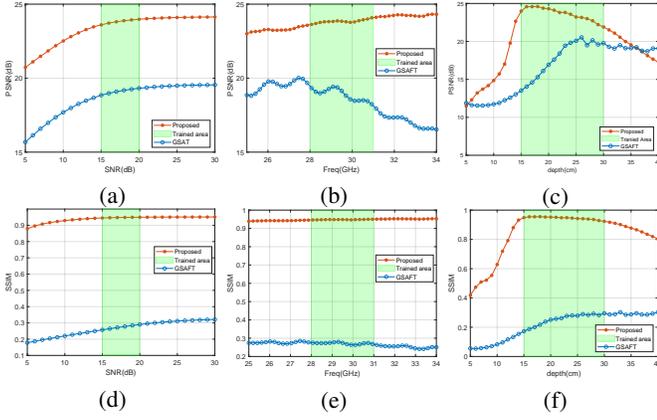


Fig. 5. The sensitivity of the reconstruction to additive noise, imaging frequency and object to TX distance. The PSNR metric in terms of input SNR(dB), frequency and depth are shown in (a), (b) and (c), respectively. Similar curves for the SSIM are shown in (d), (e) and (f).

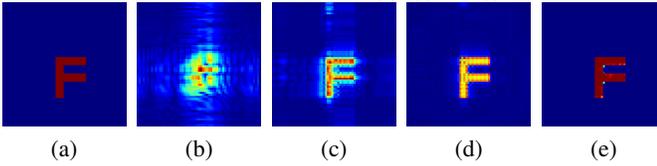


Fig. 6. Reconstruction under loss of calibration. (a) The original object, (b) measured data with depth 18cm, (c) GSAFT recovery with inaccurate depth (loss of calibration) that achieves $\text{PSNR} = 13.49709\text{dB}$, (d) GSAFT recovery with exact depth (calibrated) that achieves $\text{PSNR} = 13.5220\text{dB}$, and (e) the proposed blind recovery that achieves $\text{PSNR} = 37.2754\text{dB}$.

x , and y axes. The results are shown in Fig 4 (and Table I) for four different objects.

In the second experiment, we investigate the sensitivity of the reconstruction to additive noise, imaging frequency and object to sensor distance (depth). For each case, we fix two of the parameters and vary the third beyond the range used for training. In Fig. 5(a) and Fig. 5(d) we respectively plot the PSNR and SSIM metrics of the reconstruction when the imaging frequency and the depth are set to 29.5GHz and 22.5cm, respectively, and the input SNR sweeps the range 5dB to 30dB. We should highlight that the range 15dB to 20dB was used for training. Similar curves are shown in Fig. 5(b) and Fig. 5(e) when the input SNR and the depth are fixed to 17.5dB and 22.5cm, respectively, and the imaging frequency varies from 25GHz to 34GHz (training data was limited to [28GHz, 31GHz]). In the last setup, we sweep the depth from 5cm to 40cm (the used range for training was 15cm to 30cm) while fixing the input SNR and the imaging frequency at 17.5dB and 29.5GHz, respectively (Figs. 5(c)-(f)). We observe that our model is robust to various changes and maintains a high reconstruction accuracy without requiring an exact prior knowledge about these parameters, even in presence of noise. It is worth noting that at MMW frequencies, the wave propagation loss increases with the increase in the depth. This phenomenon can be witnessed at the tail of SSIM curve in Fig. 5 (b).

In the next experiment, we intentionally use the method in a misadjusted setting to mimic loss of calibration. Here, we set the object to sensor distance as 18cm, but use the value 20cm in reconstruction and use various imaging frequencies from 28GHz to 31GHz. Reconstruction results for both the GSAFT and the proposed method are provided in Fig. 6. While the performance of the GSAFT method degrades in the misadjusted setting, the proposed method that works blindly, performs superbly ($\text{PSNR} = 37.2754\text{dB}$).

As another experiment, we use real measurements. The system setup in Fig. 7 shows a pair of antennas connected to a network analyzer for scanning the object at 61×61 predefined points. In this



Fig. 7. The real imaging system setup. The transceiver moves in a 2D plane, scans the object and sends the measured data to the VNA.

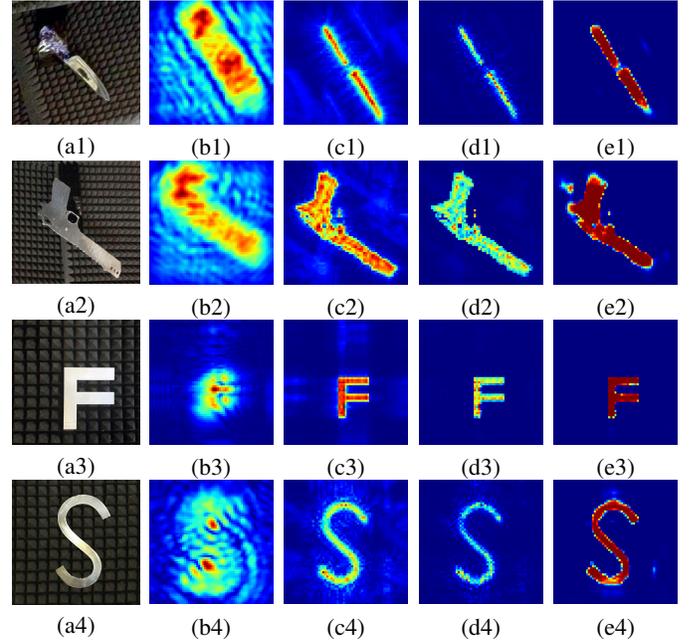


Fig. 8. Each row is related to the measurements of a separate object using the MMW imaging system in Fig. 7. Column (a) shows the optical image of the objects. Column (b) depicts the measured data before reconstruction, and columns (c) and (d) demonstrate the recovery of the fully calibrated GSAFT and DMAS methods, respectively. Finally, column (e) presents the non-calibrated blind recovery results of the proposed method.

wide-band imaging scenario, the frequency is swept from 28GHz to 31GHz and depth is set to 20cm. The results are shown in Fig. 8.

Finally, we compare the run time of our proposed method with GSAFT on a workstation with an Intel(R) Core(TM) i7-8700K @ 3.70GHz CPU and a graphic card of NVIDIA GeForce GTX 1080Ti. A set of 1000 images of size 61×61 pixels were fed to the proposed model as well as the GSAFT method for reconstruction. On average, the proposed method takes 2.55ms for reconstructing each image in contrast to 2.47ms for the GSAFT method. This shows that with parallel processing, the higher computational load of the proposed method can be handled in almost the same time as the simple GSAFT method. It should be noted that we employed the numpy implementation of FFT in this experiment, and our proposed network is implemented in Keras.

B. Discussions

In the previous experiments, we observed that the proposed method had a superior reconstruction accuracy compared to traditional methods like GSAFT and DMAS. Also, while these conventional methods require the knowledge of imaging frequency and object to antenna distance to maintain a good reconstruction performance, the proposed model only assumes a [pre-determined] range for these parameters and does not require the knowledge of the exact values. As measuring these parameters at each imaging shot is not required, we consider

the method as calibration-free. Noise-tolerance is another advantage of the proposed method. As the model is trained also with noisy data (various levels), it is able to partially compensate for the noise without asking for the noise variance (noise level). The latter is rather distinctive when compared to conventional methods.

The two main drawbacks of deep-learning-based approaches such as the proposed method are 1) the need for large datasets for training, and 2) higher computational complexity compared to conventional methods.

Deep-learning-based methods are known to be data hungry, as they need to learn the structures from scratch. Conventional methods, however, rely on certain mathematical models with few parameters that express the structures. In this work, we had to find a workaround to this issue by generating synthetic data for training using (4), instead of gathering a huge volume of experimental data. Since the real-world measurements do not fully match the generated synthetic data, we observe some artifacts in the reconstructed images. In general (not necessarily the studied case in this paper), issues such as insufficient number of layers, limited number of learning parameters and training/test data model mismatch (e.g., the type of noise and distortions) lead to uncertainty in reconstruction results of neural networks. Some recent works such as [28] find it useful to address these issues, and quantify the reconstruction uncertainty by introducing confidence values.

To compare the computational load of the deep-learning-based methods with conventional methods, let us consider the GSAFT method that we used as a baseline in our comparisons. This method is based on FFT, which has considerably less computational load than that of a neural network with multiple layers. Nevertheless, with the availability of current GPUs, parallel processing could substantially compensate for the higher computational cost of a neural network for reducing the run time.

V. CONCLUSION

In this work, we studied the problem of image reconstruction in millimeter wave imaging systems. We proposed a deep-learning-based structure to approximate the inverse system of the MMW imaging device. The designed network receives the measured complex data from the receiver antenna array at its input, and generates the 2D image of the object under test at its output. The proposed neural network architecture is trained on wide-band data, taking depth variations and noise into account. Therefore, it is compatible with real-world wideband imaging systems and capable of maintaining accurate reconstruction without being accurately calibrated. Despite the relative success of the proposed method in reconstructing images both with synthetic and experimental data, there is still room for improvement. The deep-learning methods (like the one employed here) require large amount of training data with close to authentic (real-world) distortions and non-idealities. Besides, they impose higher computational load compared to conventional linear inverse scattering approaches.

REFERENCES

- [1] V. M. Patel, J. N. Mait, D. W. Prather, and A. S. Hedden, "Computational millimeter wave imaging: Problems, progress, and prospects," *IEEE Signal Processing Magazine*, vol. 33, no. 5, pp. 109–118, 2016.
- [2] A. Elboushi and A. Sebak, "MMW Sensor for Hidden Targets Detection and Warning Based on Reflection/Scattering Approach," *IEEE Trans. Antennas Propag.*, vol. 62, no. 9, pp. 4890–4894, 2014.
- [3] H. Zamani and M. Fakhrazadeh, "1.5-d sparse array for millimeter-wave imaging based on compressive sensing techniques," *IEEE Transactions on Antennas and Propagation*, vol. 66, no. 4, pp. 2008–2015, April 2018.
- [4] M. C. Kemp, "Millimetre wave and terahertz technology for detection of concealed threats - a review," in *15th Int. Conf. Terahertz Electronics*, Sept 2007, pp. 647–648.
- [5] D. M. Sheen, D. L. McMakin, and T. E. Hall, "Three-dimensional millimeter-wave imaging for concealed weapon detection," *IEEE Trans. Microw. Theory Tech.*, vol. 49, no. 9, pp. 1581–1592, 2001.
- [6] W. Shao and Y. Du, "Microwave imaging by deep learning network: Feasibility and training method," *IEEE Transactions on Antennas and Propagation*, vol. 68, no. 7, pp. 5626–5635, 2020.
- [7] E. Simsek, "Machine learning exercises on one dimensional electromagnetic inversion," *IEEE Tran. on Antennas and Propagation*, 2021.
- [8] Z. Wei and X. Chen, "Physics-inspired convolutional neural network for solving full-wave inverse scattering problems," *IEEE Transactions on Antennas and Propagation*, vol. 67, no. 9, pp. 6138–6148, 2019.
- [9] Z. Meng, M. Zhang, and H. Wang, "Cnn with pose segmentation for suspicious object detection in mmw security images," *Sensors*, vol. 20, no. 17, p. 4974, 2020.
- [10] W. Yu, X. Chen, and L. Wu, "Segmentation of concealed objects in passive millimeter-wave images based on the gaussian mixture model," *Journal of Infrared, Millimeter, and Terahertz Waves*, vol. 36, no. 4, pp. 400–421, 2015.
- [11] S. L. Tapia, R. Molina, and N. P. de la Blanca, "Detection and localization of objects in passive millimeter wave images," in *24th European Signal Processing Conference (EUSIPCO)*. IEEE, 2016.
- [12] W. Xing, J. Zhang, and L. Guo, "A fast detection method based on deep learning of millimeter wave human image," in *Proceedings of the 2018 International Conference on Artificial Intelligence and Virtual Reality*, 2018, pp. 67–71.
- [13] L. Pang, H. Liu, Y. Chen, and J. Miao, "Real-time concealed object detection from passive millimeter wave images based on the yolov3 algorithm," *Sensors*, vol. 20, no. 6, p. 1678, 2020.
- [14] C. Hu, L. Wang, Z. Li, and D. Zhu, "Inverse synthetic aperture radar imaging using a fully convolutional neural network," *IEEE Geoscience and Remote Sensing Letters*, vol. 17, no. 7, pp. 1203–1207, 2019.
- [15] J. Gao, B. Deng, Y. Qin, H. Wang, and X. Li, "Enhanced radar imaging using a complex-valued convolutional neural network," *IEEE Geoscience and Remote Sensing Letters*, vol. 16, no. 1, pp. 35–39, 2018.
- [16] G. Chen, P. Shah, J. Stang, and M. Moghaddam, "Learning-assisted multimodality dielectric imaging," *IEEE Transactions on Antennas and Propagation*, vol. 68, no. 3, pp. 2356–2369, 2019.
- [17] T. E. H. David M. Sheen, "Reconstruction techniques for sparse multi-static linear array microwave imaging," *Proc.SPIE*, vol. 9078, 2014.
- [18] M. Rezaei, H. Zamani, M. Fakhrazadeh, and M. Memarian, "Quality improvement of millimeter-wave imaging systems using optimized dual polarized arrays," *IEEE Trans. on Antennas and Propagation*, 2021.
- [19] H. Zamani, M. Fakhrazadeh, A. Amini, and F. Marvasti, "Random sampling with iterative recovery for millimeter-wave imaging systems," *IEEE Transactions on Circuits and Systems II: Express Briefs*, 2020.
- [20] Z. Wei and X. Chen, "Deep-learning schemes for full-wave nonlinear inverse scattering problems," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 4, pp. 1849–1860, 2018.
- [21] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.
- [22] P. Kotschieder, M. Donoser, and H. Bischof, "Improving affinity matrices by modified mutual knn-graphs," 01 2009.
- [23] "Available software and databases," <http://vision.ljms.brown.edu/content/available-software-and-databases>, accessed: 2021-03-12.
- [24] X. Bai, W. Liu, and Z. Tu, "Integrating contour and skeleton for shape classification," in *2009 IEEE 12th international conference on computer vision workshops, ICCV workshops*. IEEE, 2009, pp. 360–367.
- [25] "List of kitchen utensil categories," https://homepages.inf.ed.ac.uk/rbf/UTENSILS/kitchen_utensils1.xml, accessed: 2021-03-12.
- [26] H. Been Lim, N. Thi Tuyet Nhung, E.-P. Li, and N. Duc Thang, "Confocal microwave imaging for breast cancer detection: Delay-multiply-and-sum image reconstruction algorithm," *IEEE Transactions on Biomedical Engineering*, vol. 55, no. 6, pp. 1697–1704, 2008.
- [27] Zhou Wang *et al.*, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, April 2004.
- [28] Z. Wei and X. Chen, "Uncertainty quantification in inverse scattering problems with bayesian convolutional neural networks," *IEEE Transactions on Antennas and Propagation*, vol. 69, no. 6, pp. 3409–3418, 2020.