

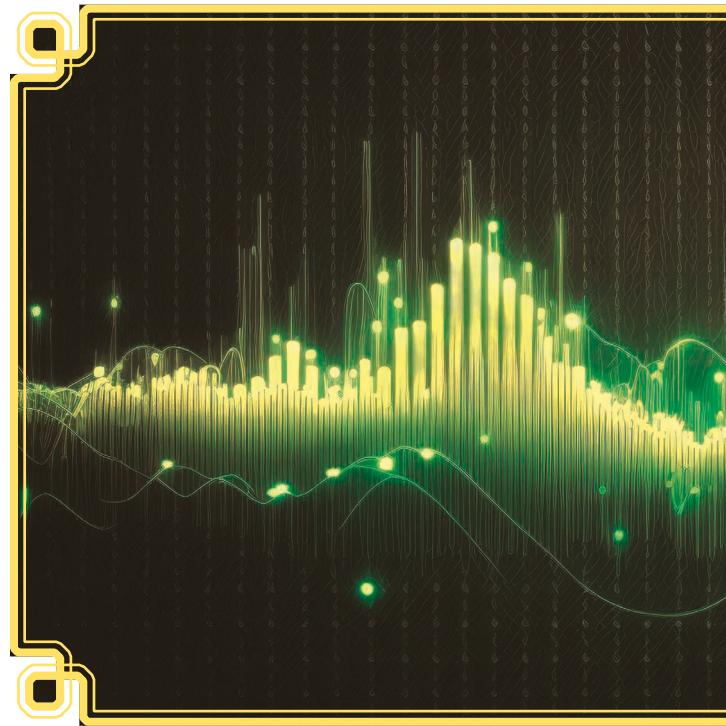
# Scalable Undersized Dataset RF Classification

*Using convolutional multistage training.*

Unlike areas such as computer vision and speech recognition, where convolutional and recurrent neural network (NN)-based approaches have proven effective due to the nature of the respective areas of application, deep learning (DL) still lacks a general approach suitable for the unique nature and challenges of radiofrequency (RF) systems such as radar, signals intelligence, electronic warfare, and communications. Existing approaches face problems in robustness, consistency, efficiency, repeatability, and scalability. Some of the main challenges in radiofrequency (RF) sensing such as radar target identification are the difficulty and cost of obtaining data. Hundreds to thousands of samples per class are typically used when training for classifying signals into 2–12 classes with reported accuracy ranging from 87% to 99%, where accuracy generally decreases with more classes added. In this article, we present a new DL approach based on multistage training (MST) and demonstrate it on RF sensing signal classification. We consistently achieve more than 99% accuracy for up to 17 diverse classes using only 11 samples per class for training, a substantial improvement over standard DL approaches.

## INTRODUCTION

With applications in defense, retail, health care, and tomography [1], [2], [3], [4], [5], RF-based sensor systems that can detect, locate, and identify targets at long distances and under different weather conditions are being developed.



Examples include ground-based and aircraft-mounted radars for the correct identification of targets in the battlefield [6]. Another example is ground-penetrating radar, which has important applications in mineral resources evaluation [7]. Other examples include advanced personnel screening imagers [8]; concealed weapon detection [9]; through-the-wall imaging

[10]; and vehicular access control using RFID [11]. The signal received, which is a result of electromagnetic scattering [12], is difficult to interpret and model using hand-engineered methods, and this is where black box-type approaches such as DL can be useful.

In our previous work [13], we presented a new MST approach to DL for RF transmitter (Tx) identification. MST is a highly distributable structure-parallel DL approach that comprises multiple stages of NN ensembles, each consisting of several small networks, which allows the efficient utilization of Newton-based second-order optimization and provides highly effective data-driven regularization. It was originally designed for the high-fidelity denoising of magnetic resonance images with nonadditive noise [14], [15]. The multistage approach allows very early stopping at each individual stage, where a target error is assigned as a stopping criterion in the first stage and is gradually decreased at successive stages. By systematically assigning specific stopping criteria to each stage, we can control the speed of convergence in the system as a whole to optimize the overall performance and generalization, where a minimal error is reached in the final stage without overfitting. Our second-order MST has proven superior to standard gradient descent-based first-order convo-

We refer to the new implementation as *convolutional MST* (C-MST), and we refer to the network architecture as the *convolutional multistage network* (C-MSN). We demonstrate our method on the classification of radar-like signals, where a linearly polarized electromagnetic wave illuminates an object, which in turn creates a scattered field detected by another antenna. In this experiment, we consistently obtain more than 99% accuracy for 17 classes using only 11 samples per class for training as well as consistency, robustness, wall clock time, and scalability.

The dataset, which contains many acquisitions of scattered electromagnetic waves acquired in the frequency domain using a network analyzer and measuring the S-parameter for all 17 object classes, will be made publicly available. The environment mimics the radar-like detection of objects under ideal conditions, i.e., without any clutter or motion. Our contribution in this article is threefold: 1) providing an overview of existing techniques; 2) extending our method to include a convolutional front end, thereby enabling us to increase the input size while maintaining the excellent generalization properties of MST; and 3) providing a new benchmark dataset to help standardize the comparison between different algorithms for RF classification. Our algorithm is computationally efficient and allows

incremental learning where only part of the network needs to be trained when new targets are added. It can be run on modest computers and may be a good candidate for deployment in the field, for real-time low-shot-number learning.

## NON-DL METHODS FOR TARGET IDENTIFICATION

Target recognition algorithms operate on measured target signature data for comparison with the previously derived computer representations of the targets to provide an estimate of the target's identity. The recognition process is limited by noise in the radar measurements; errors in the generation of signature reference data; and the use of classifiers, which usually involves design compromises. A conventional algorithm is template matching using cross-correlation analysis [16], [17], [18], [19], [20], [21]. The accuracy of template matching can be improved by using statistical pattern recognition techniques designed to determine the class or identity of a measured object by means of the features extracted

from the measured pattern or signature [22], [23], [24], [25], [25], [26], [27], [28], [29], [30], [31], [32], [33], [34], [35], [36].

Features can be hand engineered, including polarization enhancement; resonant-frequency poles; multipath reflection signatures; target structure-induced modulations; microphone effects; jet-engine modulations; or features derived from a

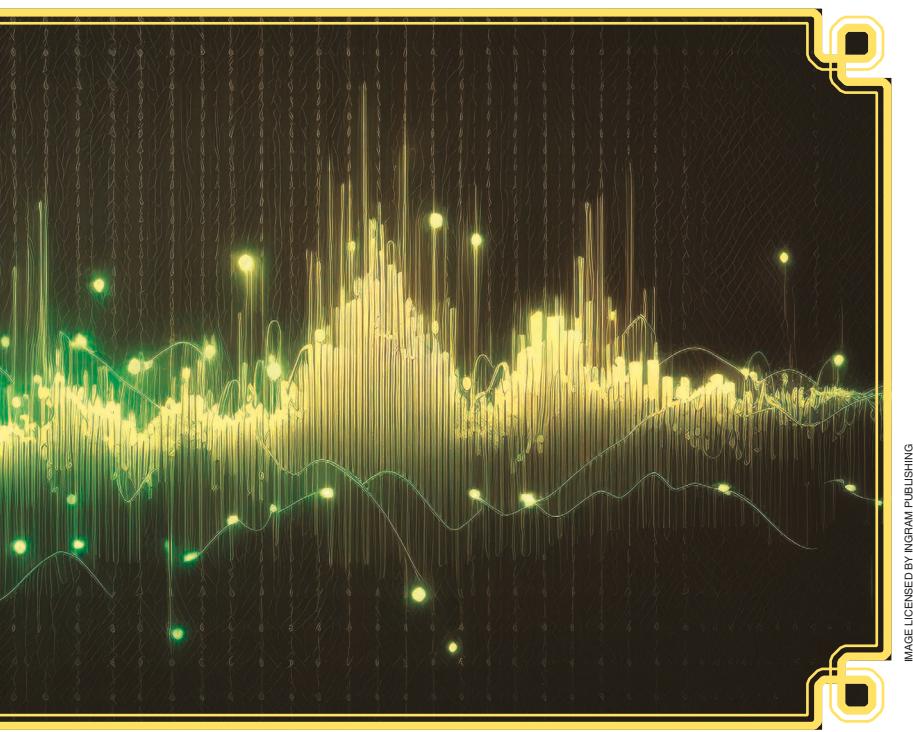


IMAGE LICENSED BY INGRAM PUBLISHING

lutional and deep NNs (DNNs), including MST trained using first-order optimization [13].

Given that MST has proven superior to other DL methods in RF signal classification [13], we investigated an extension of MST employing a convolutional front end as a feature extraction stage for a fully automated end-to-end implementation.

transform domain representation of the signal [37]. Features can also be learned from the data using techniques such as  $k_n$ -nearest-neighbor estimation or Fisher linear discriminant analysis [38]. The use of features for recognition provides advantages by reducing the requirements on the size of reference databases. The extracted features are then compared to a database's content to maximize the target recognition performance using rule-based Euclidean distance or Bayesian techniques [39], [40], [41], [42], [43], [44], [45], [46]. Distance-based methods invoke the sum of all the distance measures for all features, and the minimum distance measure is the best assessment of the target's identity. Unlike end-to-end approaches that would ideally extract optimal features directly from the raw data, manual feature selection can be limited by rough approximations and subjectivity.

### TARGET IDENTIFICATION BY DL

Thanks to the great success of DL methods in recent years in areas such as computer vision and speech recognition, there has been a growing interest in using DL for RF applications. However, unlike computer vision and speech recognition, where standard convolutional and recurrent NN methods have proven to be very effective due to the nature of these applications, DL still lacks a standardized approach suitable to the unique nature and challenges of RF. While an ideal target identification method should be able to operate directly on raw data in an end-to-end fashion, fully connected NN (FCN) implementations trained using standard algorithms are limited by the number of inputs and the number of training samples they need to achieve proper generalization and avoid overfitting. In general, as the input size increases, the number of required training samples increases [47].

Typically, the number of samples becomes impractical after the input size exceeds only a few hundred inputs. However, digitized and demodulated RF signals are complex-valued and may be thousands of samples long. Standard regularization techniques such as  $L_2$ -regularization help remedy the problem but only to a certain extent [48]. For this reason, feature extraction remains a crucial step for an FCN to work, and the shortcomings of feature engineering, which can be limited by subjectivity and crude approximations, inherently remain.

The introduction of convolutional NNs (CNNs) delivered a tremendous advantage to overcoming this DL problem by limiting the number of connections from the inputs to the network [47], [49]. Convolutional neurons scan through the entire input but are connected to only a few inputs at a time. This, in addition to the introduction of DL-specific regularization techniques such as dropout layers [50], revolutionized the field of computer vision. CNNs have the ability to automatically extract features from raw inputs, which is a crucial step toward end-to-end implementations. However, due to the immense computational requirements, CNN training algorithms available today are largely based on first-order gradient descent optimization. While the recent successes of DNNs would be nonexistent if not for first-order methods, it poses limitations on the performance and capabilities of CNNs and DL in general. For example, it

is not uncommon for an effective computer vision application to require tens or hundreds of thousands of training samples, though it is possible for humans to distinguish between different objects after only few encounters [51], [52].

A recent study on using CNNs for RF identification shows that several thousand samples per class are needed to achieve high accuracy [53]. While customized approaches have been proposed for specific applications, performance is very sensitive to hyperparameters, and these approaches require considerable expert effort to tune for proper parameter selection. This makes DL implementations narrowly applicable, mainly limited to the exact application and dataset at hand, which poses significant limitations on repeatability. Furthermore, even with suitable hyperparameters, training with different initial conditions, e.g., initial neuron weights and biases, can cause large variations in performance. A typical approach to address this problem is to train one or more network models several times with different initial conditions and use the average output or vote of a committee of networks as the final result [54], [55], [56], which can help improve performance but is ultimately limited by the success rate of the individual networks.

Second-order Newton-based optimization methods offer several advantages in robustness to hyperparameters, efficiency, accuracy, and convergence speed and require lower network complexity where they can achieve results that are superior to first-order methods with fewer neurons [57], [58], [59], [60], [61], [62], [63]. However, second-order optimization is generally deemed unfeasible for training large networks due to the intractable computational requirements of traditional second-order methods.

Provided a large number of training samples is available, most shortcomings of first-order-based DL approaches are usually manageable. Unfortunately, obtaining data in large quantities is not trivial in RF problems and can be a difficult, time-consuming, and expensive process. Synthetically generated data are often used to compensate for the lack of real data, but they can only approximate the fine details in the signals, which poses a limitation on accuracy and scalability. Generative adversarial networks (GANs) have also been used to generate new training data [64], but GANs can face similar issues depending on their application. Recent work [65] studying the use of GANs for synthetic aperture radar showed promise at improving the quality of synthetically generated data. The study also showed that GANs were difficult to train and were not a perfect substitute for real data [65].

Artificial NNs [66], support vector machines [67], and Markov [68], [69] models have been studied to perform feature extraction and classifier functions [70], [71]. Several attempts have been made to adapt standard DL approaches to radar classification with varying degrees of success [72], [73], [74], [75], [76], [77], [78], [79]. Hundreds to thousands of samples per class were used for training to classify 2–12 classes with reported accuracy ranging from 87 to 99% where accuracy generally decreases with more classes. Existing DL approaches face problems in robustness, consistency, efficiency, repeatability, and scalability.

## EXPERIMENTAL DETAILS, ARCHITECTURE, AND HYPERPARAMETERS

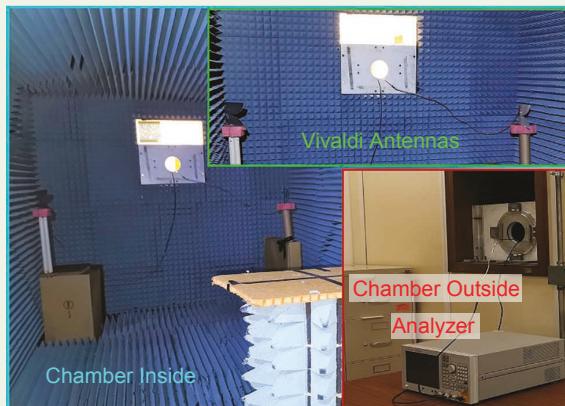
### PHOTOS OF THE OBJECTS AND CHAMBER

Seventeen objects were collected from engineering and chemistry laboratories at the University of California, Los Angeles for use as radar targets. The objects were selected to present a diversity of sizes, shapes, and material composition. The 47.0 by 30.5-cm cardboard platform seen in the pictures provides a scale for each object's size. All 17 targets were placed in the radar's path and rotated through three different angles (0°, 45°, and 90°). Photos of all 17 rotated objects are shown in Figure 5, as seen from the perspective of the radar. Photos of the anechoic setup are shown in Figure S1.

### SIGNAL STRENGTH

Twelve traces were recorded for each of the 17 objects and each of the three orientations per object. Additionally, 112 traces of the empty anechoic chamber were recorded under otherwise equal conditions. Each trace (S21 parameter) was recorded as a string of complex numbers (real, imaginary) representing the complex-valued signal amplitude as a function of the sweep frequency. Signal-to-noise ratios (SNRs) and relative radar cross-sections (rRCSs) were calculated for each of the 17 objects and their three orientations using these data and tabulated later (Table S1). RCSs are provided as dimensionless numbers between zero and one rRCS. The SNR values reflect the maximum amplitude in the frequency domain of the S21 parameter, whereas rRCS values reflect the total signal over the same domain. This is analogous to the pulse radar echo amplitude. Uncertainties in the values represent sample standard deviations, which were calculated using the 12 traces for each object (and orientation).

From each complex-valued trace for each object (and orientation), the magnitude of the complex data was computed and stored in vectors of length 1,600, corresponding to linearly spaced frequency values in the range 675 MHz–8.5 GHz. For the empty anechoic



**FIGURE S1.** The anechoic chamber and experimental setup. The photo outlined in green shows the two Vivaldi antennas in the lower left and right corners, both oriented vertically, as used during the experimental data collection. The photo outlined in red shows the VNA used for all S21 readouts (Agilent model E5071C 9 kHz; 8.5-GHz ENA series). The photo outlined in blue shows the radar's position relative to the cardboard platform.

chamber data, all 112 traces were averaged to provide a clean (low noise) trace. This low-noise trace was used to subtract the background signal for each object (and orientation) to produce traces whose features reflect only the characteristics of an object.

From the background-subtracted traces, the SNR was calculated as follows. The location of the maximum signal was identified, and 100 nearby points were averaged to reduce noise. For each trace, a flat region was identified to estimate the noise as the standard deviation of 100 points taken in the flat region. The SNR was computed as the previously mentioned signal strength divided by the standard deviation of the flat region. The resulting value was inserted into (S1)

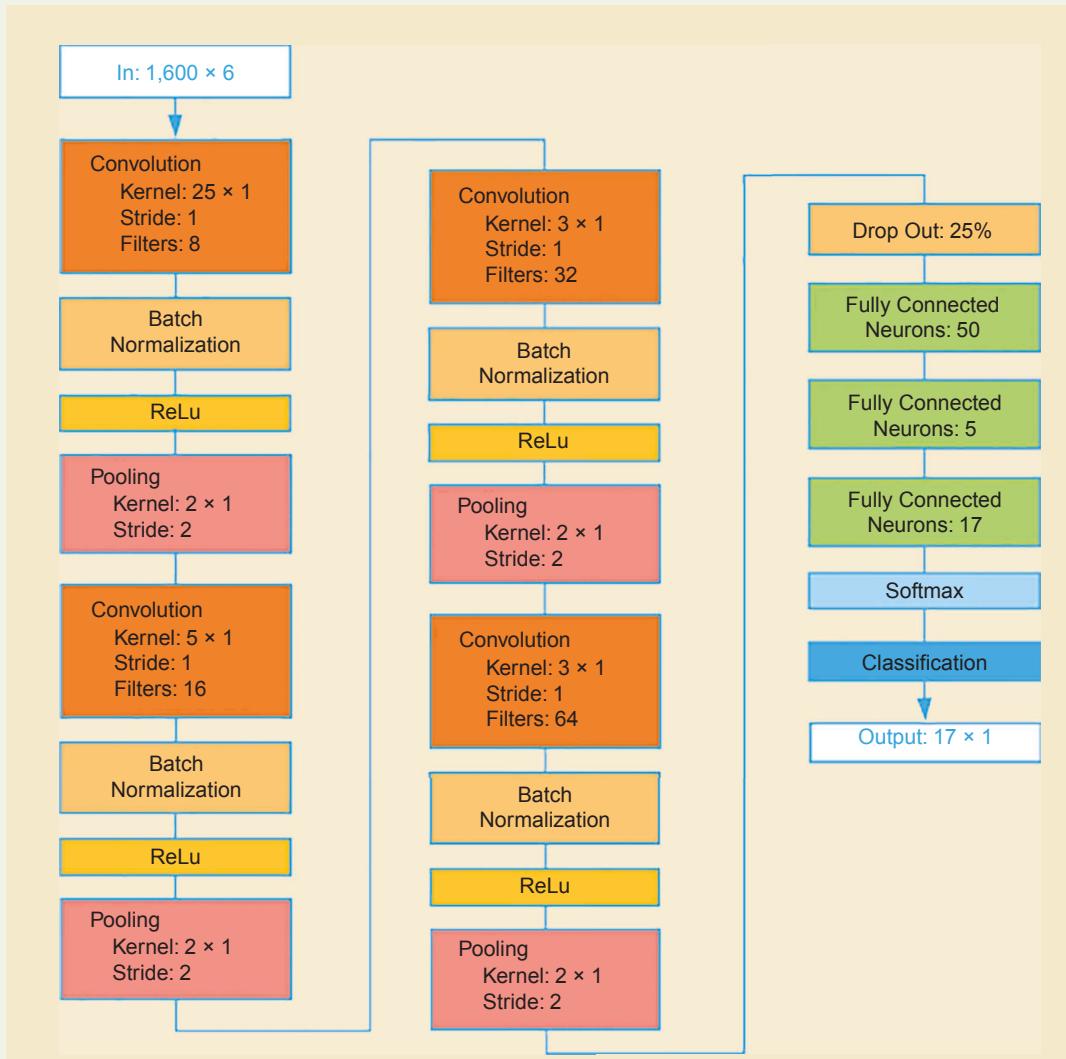
$$\text{SNR}_{\text{dB}} = 20 \log_{10} \left( \frac{S_{21\text{target removed}}}{S_{21\text{with target}}} \right) \quad (\text{S1})$$

to yield a value in dB. rRCS values were estimated from the background-subtracted traces by taking the area under each curve, which represents the total radar signal (integrated across all frequencies). The integral was computed in MATLAB using the trapezoidal rule. This area represents the pulse amplitude in a pulsed radar echo experiment. These pulse amplitudes were then averaged for each object-orientation combination and

**TABLE S1. SNR AND RRCS VALUES FOR EACH OBJECT (1–17) AND THEIR THREE ORIENTATIONS. THE STANDARD DEVIATION OF EACH SIGNAL IS SHOWN IN PARENTHESES. THE SNR VALUES ARE IN dB.**

Obj.	SNR, 0°	SNR, 45°	SNR, 90°	rRCS, 0°	rRCS, 45°	rRCS, 90°
1	20 (10)	13 (8)	19 (9)	0.9 (3)	0.1 (2)	1 (4)
2	14 (7)	14 (7)	8 (7)	0.34 (5)	0.1 (2)	0.4 (1)
3	13 (7)	14 (8)	12 (7)	0.1 (2)	0.1 (2)	0.14 (2)
4	12 (7)	14 (7)	11 (6)	0.22 (6)	0.09 (3)	0.17 (3)
5	14 (8)	13 (7)	14 (8)	0.1 (2)	0.1 (2)	0.17 (4)
6	12 (8)	13 (7)	10 (6)	0.23 (3)	0.1 (3)	0.17 (2)
7	13 (8)	13 (8)	14 (9)	0.1 (3)	0.1 (2)	0.5 (2)
8	12 (7)	14 (8)	13 (8)	0.1 (3)	0.1 (3)	0.2 (3)
9	12 (6)	14 (7)	10 (6)	0.1 (3)	0.1 (3)	0.11 (3)
10	10 (6)	13 (7)	12 (7)	0.12 (3)	0.09 (3)	0.09 (3)
11	10 (6)	14 (8)	11 (6)	0.12 (2)	0.09 (3)	0.1 (3)
12	10 (6)	13 (7)	10 (6)	0.14 (3)	0.09 (3)	0.11 (3)
13	11 (7)	12 (7)	10 (6)	0.23 (2)	0.1 (3)	0.11 (3)
14	11 (6)	13 (7)	12 (7)	0.11 (3)	0.09 (3)	0.1 (3)
15	14 (8)	11 (6)	11 (7)	0.44 (5)	0.1 (3)	0.14 (2)
16	11 (7)	13 (7)	11 (6)	0.1 (3)	0.1 (3)	0.1 (3)
17	12 (7)	13 (7)	13 (7)	0.1 (3)	0.1 (3)	0.1 (3)

(Continued)



**FIGURE S2.** The CNN architecture and hyperparameters. ReLU: rectified linear unit.

normalized to the largest object-orientation average to yield a number between zero and one (relative RCS).

#### CNN ARCHITECTURE AND HYPERPARAMETERS

The detailed CNN architecture and hyperparameters are shown in Figure S2. The hyperparameters used for Adam and Levenberg-Marquardt are provided in Table S2.

#### WALL CLOCK TIME

Figure S3 shows the wall clock time speedup relative to one CPU core. Measurements were obtained by activating a different number of CPU cores and measuring the training time.

#### DEMONSTRATION OF ADDITIONAL REGRESSION APPLICATION

An artificial dataset for RCS versus frequency was created in MATLAB R2021b (MathWorks, Natick, MA) for sphere and disk geometries using the commands `rcssphere` and `rcsdisc`,

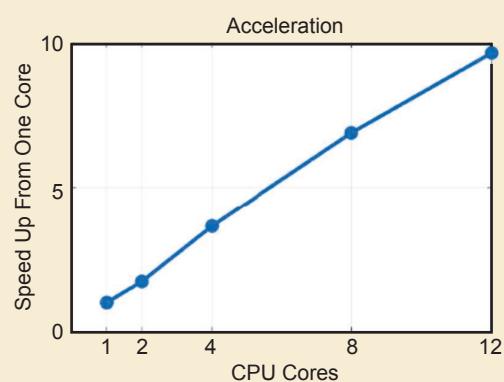
respectively. The command `rcssphere` takes as inputs the sphere radius (m); signal propagation speed [(m/s), set to  $c$ , the speed of light]; frequency range (set to 100–10,000 MHz in steps of 10 MHz); and azimuthal and elevation angles. The azimuthal and elevation angles were uniformly randomly sampled from  $-45^\circ$  to  $45^\circ$ . The sphere radius was uniformly randomly sampled from 0.2 to 1.0 m. The command outputs RCS ( $\text{m}^2$ ) versus frequency, to which was added uncorrelated normally distributed random noise in absolute value with a standard deviation of 0.1  $\text{m}^2$ , i.e.,  $(0.1 \cdot |X|, X \sim \mathcal{N}(0, 1))$ . The signal was then converted to power (dBsm) and used for training the NN. Likewise, the command `rcsdisc` takes as input the disk radius (m); signal propagation speed; frequency; and azimuthal and elevation angles. The radius, speed, frequency, and angles were set as in the case of

(Continued)

**TABLE S2. THE LEARNING PARAMETERS AND OPTIONS OF EACH ALGORITHM. OPTIONS THAT ARE NOT LISTED HERE WERE KEPT AT THE DEFAULT SETTING IN MATLAB.**

Levenberg-Marquardt	
Performance goal	0
Maximum validation failures	6
Minimum performance gradient	1E-07
Initial $\mu$	0.001
$\mu$ decrease factor	0.1
$\mu$ increase factor	10
Maximum $\mu$	1E+10
Adam (CNN)	
Minimum batch size	12
Initial learning rate	1E-03
Learn rate schedule	piecewise
Learn rate drop factor	0.1
Learn rate drop period	5
Shuffle	Every epoch
SGDM (SqueezeNet)	
Minimum batch size	10
Initial learning rate	3E-04
Maximum epochs	100
Shuffle	Every epoch

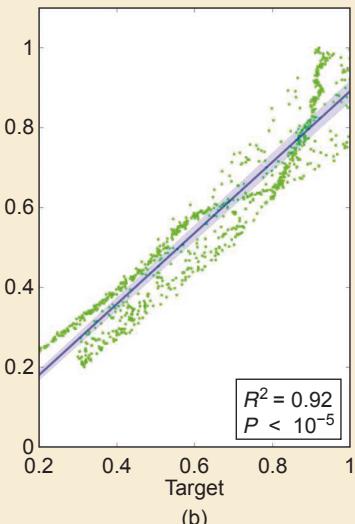
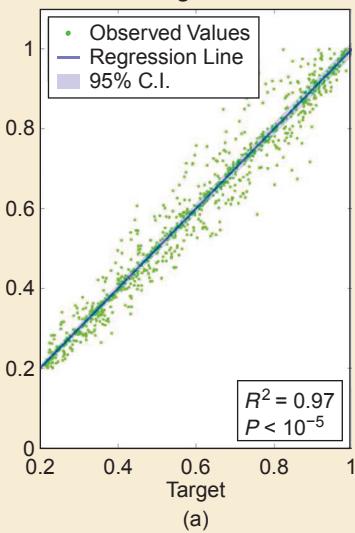
SGDM: stochastic gradient descent with momentum.



**FIGURE S3.** The wall clock acceleration.

the sphere. Noise was added, which had the same distribution as for the sphere except for lower magnitude to compensate for the smaller RCS of the disk compared to the sphere, i.e.,  $(0.001 \cdot |X|, X \sim \mathcal{N}(0, 1))$ . Here, 3,000 samples with random

**Radius Length Correlation**



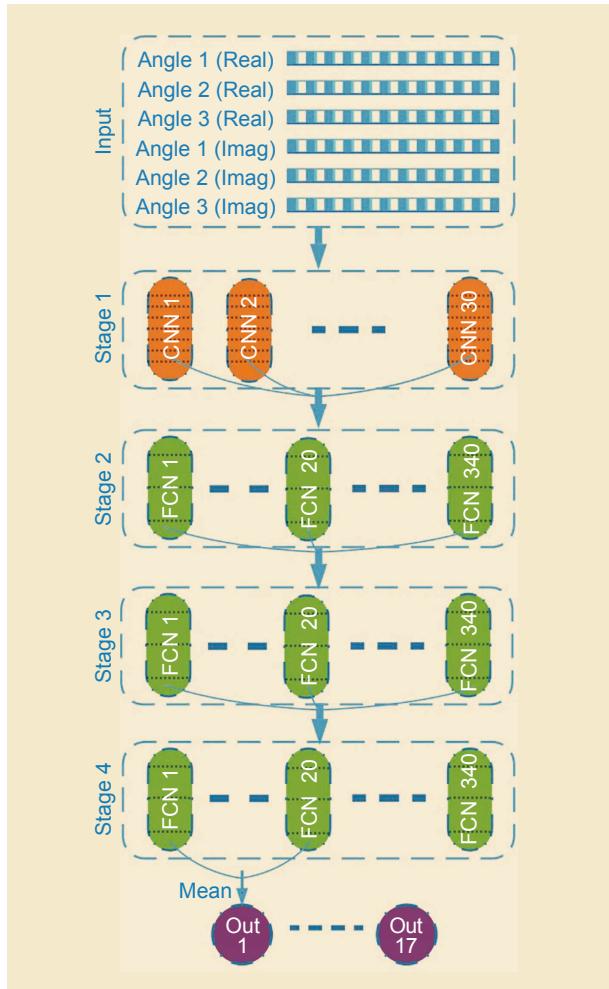
**FIGURE S4.** Correlation plots for the (a) C-MSN and (b) the CNN committee outputs versus target (object size). These graphs correspond to the testing dataset, which contains both disks and spheres. C.I.: confidence interval.

parameters were generated for training, and 1,000 samples were generated for testing. Training and testing datasets each consist of 50% disks and 50% spheres.

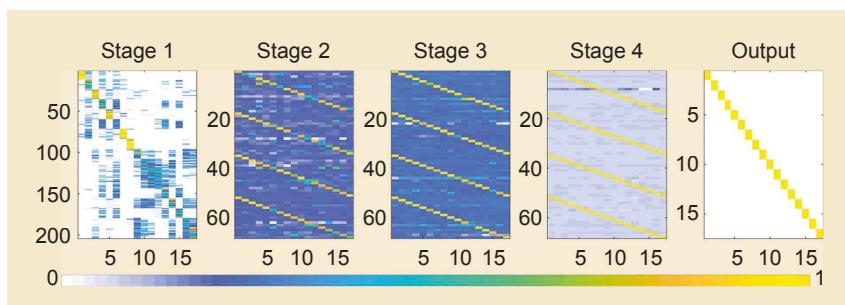
Vectors of object (disk or sphere) RCS versus frequency were used for training the NN to estimate the radius of the object. A C-MSN similar to the one described in the article was used here, with the exception of the softmax and classification layers, which were replaced with regression layers. A committee of regression CNNs was trained for comparison, where the mean value of all CNNs in the committee is used as the final output. The correlation results are shown in Figure S4.

## THE IMPLEMENTATION OF DL BY C-MST

This article presents a new variation of MST with a convolutional front end and its application in the classification of radar-like signals. Given the inherent regularization properties of CNNs



**FIGURE 1.** The C-MSN architecture for 17 object classes. A representative vector of the output of each stage in response to one sample is shown for further illustration. Actual stage outputs in response to all samples are shown in Figure 2.



**FIGURE 2.** An example of outputs from different C-MSN stages. The y-axis is the stacked stage output number, and the x-axis is the sample ID. In this example, stage 1 has 204 outputs. Stages 2–4 each have 68 outputs. The final output represents a confusion matrix. This figure illustrates the successive evolution of the output, where accuracy gradually improves and noise decreases with each stage.

for large inputs [47], [49] and their robustness to translational variance, which has been demonstrated in radar antitransmit receive [80], we have developed C-MST to further improve the performance and applicability of our original MST method. The addition of CNNs not only helps with translational invariance but also helps in situations where training data availability is scarce. Unlike our previous work, where only part of the RF signal is used (the onset) by the MST, the convolutional front end allows using the entire RF signal for a fully automated end-to-end implementation.

We tested the new method's generalization ability when trained with an undersized dataset (11 samples per class) for 17 classes. We term the new method *C-MST*. Here we describe the implementation of our method as well as conventional DL methods used for benchmarking purposes. The C-MST method and its implementation are described in the “C-MST” section later. The C-MST method is validated against two types of CNN committees. The first comparison is with the CNN committee constituting the C-MST front end. The second comparison is against a committee of SqueezeNets [81] that are pretrained on the 1,000-class ImageNet Large Scale Visual Recognition Challenge; these are then trained on our data using state-of-the-art transfer learning. In the “Radar Experiments” section, we describe the experimental methods for acquiring the radar data.

All networks were implemented in MATLAB® R2020a using the Deep Learning Toolbox (MathWorks, Natick, MA) on a CentOS 7-based server featuring two Xeon processors, each with 10 cores, and 128 GB of RAM. Hyperparameters for all methods were selected to maximize generalization. Architecture details and hyperparameter values are provided in “Experimental Details, Architecture, and Hyperparameters.” The dataset will be made available for download at <https://doi.org/10.21227/842w-xk80>.

### C-MST

Thirty CNNs constitute the first stage of the C-MSN illustrated in Figure 1. Each CNN consists of four inner blocks and one outer block. Each inner block contains a convolutional layer followed by a batch normalization layer, a rectified linear unit

(ReLU) layer, and a pooling layer. The outer block consists of a dropout layer, followed by three fully connected layers, a softmax layer, and a classification layer. The architecture details are illustrated in Figure S2 in “Additional Architecture Details and Values of Hyperparameters,” and the hyperparameters for each layer are listed in Table S2 in “Additional Architecture Details and Values of Hyperparameters.” The second fully connected layer in the CNN is a bottleneck layer consisting of five neurons to perform feature dimensionality reduction. Each CNN is trained for 20 epochs, and the

outputs from the bottleneck layer are the features that are passed on to stage 2.

The remaining C-MSN stages are similar to [13], where the original MST approach is described in detail. Multiple identical FCNs are stacked together at each stage, where each FCN is randomly assigned different initial conditions. Herein, each FCN consists of two fully connected layers with 10 neurons in each layer. The FCNs in each stage are separately trained for three epochs using the second-order Levenberg-Marquardt algorithm; hyperparameters are listed in Table S2 in “Additional Architecture Details and Values of Hyperparameters.” Each FCN has one output and is trained to fire in response to one of the object classes only, where 20 FCNs in each stage are assigned to each object class. In the case of classifying 17 objects, the number of FCNs in each stage is equal to  $20 \times 17$ , which yields 340 FCNs per stage. The FCNs are independently trained as regression models, and the concatenated outputs of all FCNs in each stage are passed on as the input to the next stage. In the final stage, the outputs from each group of 20 FCNs corresponding to an object class are averaged together to obtain the final response to that object class. Figure 2 shows an example that demonstrates the gradual improvement in the outputs of subsequent C-MSN stages.

### THE CNN COMMITTEE

The classification layer outputs of the 30 CNNs in the C-MSN front end are collected. The final output is the majority vote of all CNNs.

### THE SqueezeNet COMMITTEE

SqueezeNets pretrained on the 1,000-class ImageNet Large Scale Visual Recognition Challenge were adjusted to accommodate two classes and trained on our data using state-of-the-art transfer learning. To accommodate the input format of SqueezeNets, we transformed our 1D input signals into 2D time-frequency images. Transfer learning was performed through the following steps:

- 1) adjusting the pretrained SqueezeNet input layer size from the original  $227 \times 227 \times 3$  red, green, blue format into  $227 \times 227$ , i.e., the size of our 2D wavelet transformed image
- 2) modifying the last learnable layer in SqueezeNet (layer 64) to have the same number of convolutions as our new number of classes instead of the original 1,000 classes
- 3) modifying the output classification layer in SqueezeNet (layer 68) to our new number of classes instead of the original 1,000 classes
- 4) tuning the adjusted SqueezeNet.

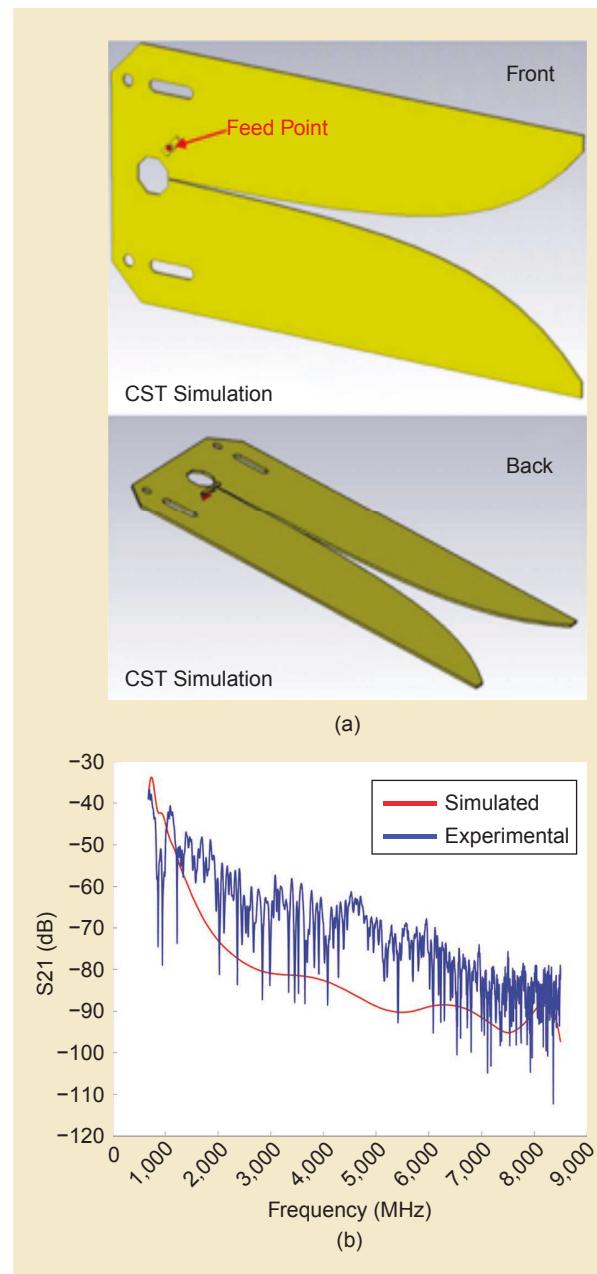
Learning parameters are listed in Table S2 in “Additional Architecture Details and Values of Hyperparameters.”

## RADAR EXPERIMENTS

### DATA COLLECTION

Radar data were collected inside an anechoic chamber (band rejection 1 MHz–10 GHz) using a pair of TSA900

900-MHz–12-GHz printed circuit board (PCB) Vivaldi antennas (RFSPACE Inc., Atlanta, GA) connected to a vector network analyzer (VNA) model E5071C 9 kHz–8.5 GHz ENA Series (Agilent, Santa Clara, CA) (Figures 3 and 4). The VNA was operated in S21 parameter mode with ports 1 [transmitter (Tx)] and 2 [receiver (Rx)] connected to antennas 1 and 2. Frequency sweeping was performed in the range 675 MHz–8.5 GHz (corresponding to wavelengths in free space: 0.44 m–3.5 cm) with IF bandwidth 70 kHz, 1,600 points/trace,



**FIGURE 3.** (a) A Vivaldi antenna model as used in the CST Studio Suite simulation of the empty chamber. (b) The experimental and simulated empty room S21 parameters for the empty chamber. This plot compares simulated versus experimental data. The experimental S21 parameter trace shown was collected on the first day of data collection.

and 512 averages/trace. Both Vivaldi antennas were mounted vertically to make the readout from the Rx antenna sensitive to waves of the same polarization as the Tx antenna. The experimental setup is shown in Figure 4. Each trace (S21 parameter) was saved as a file with 1,600 real/imaginary (complex) data points versus frequency on a linear amplitude scale from the Smith Chart mode of the VNA.

Next, 17 objects were placed approximately 10 ft from the pair of antennas, which were themselves 5 ft apart. The triangular configuration was kept fixed, the objects were rotated through three angles (0, 45, and 90°), and the experiments were repeated 12 times (for each object and angle) on different days over a total period of three months. All targets in subsequent trials and angle rotations were placed at the same locations with intentional positioning errors of up to 10 cm in-plane and 5° for the angles. Other uncertainties in the measurements are due to the limited signal-to-noise ratio (SNR) and possible drifts in the VNA S21 parameter calibration over the three-month span as the VNA was calibrated only once on the first day of experiments. Photos of the 17 objects and their positioning from the antennas' perspective are shown in Figure 5.

## DATA PROCESSING

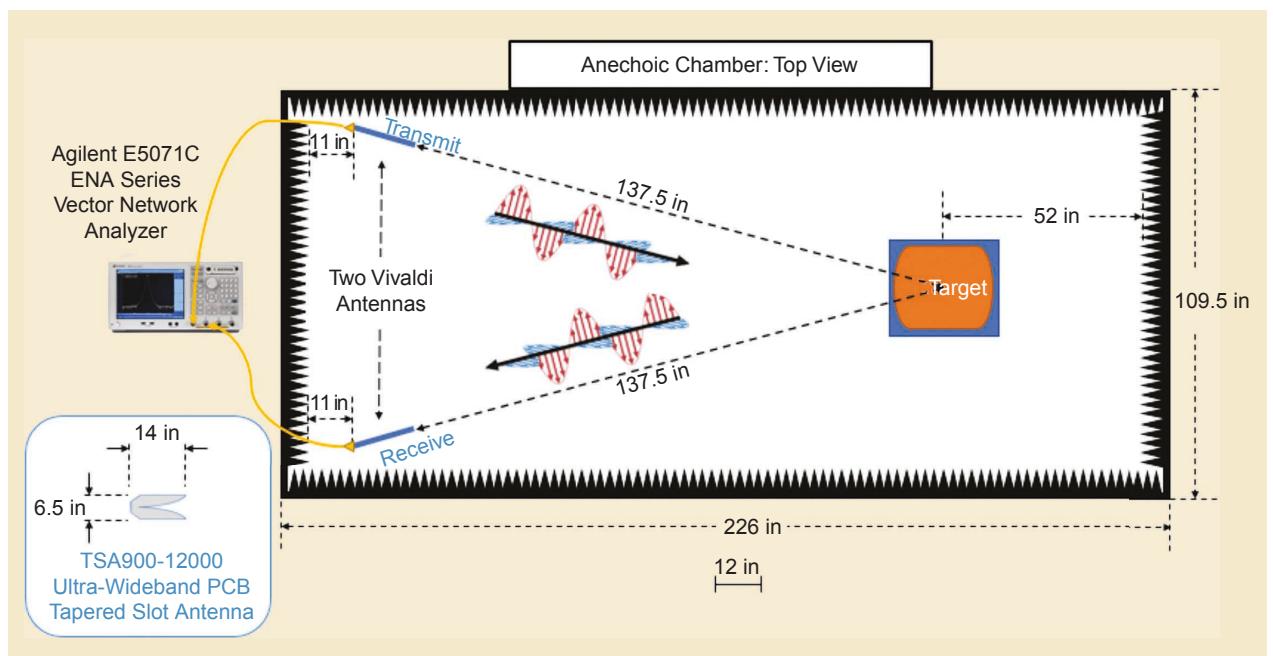
Since the purpose of these data is to assess end-to-end performance, minimal preprocessing was conducted. Fingerprints used for training the NNs were created directly from the raw data. Each fingerprint consisted of traces from three angles (0°, 45°, and 90°), with each trace consisting of 1,600 real and imaginary values, for a total of 9,600 data points/fingerprint. Real and imaginary components

of each trace were normalized separately by subtracting the mean and dividing by the standard deviation. The dataset consisted of 204 fingerprints: 17 object classes with 12 fingerprints per class. Different augmentation and preprocessing strategies can improve the results for all approaches. We omit such strategies herein to more accurately assess the role of the DL approaches in the performance comparison.

## ANALYSIS

Manual identification of targets from raw data proved difficult, as shown in Figures 6(b) and 7(b). Figure 6 shows the log-magnitude plots of the S21 parameter for all 17 objects in the 0° orientation. (The objects and our definition of orientation are shown in Figure 5.) In the low-frequency limit of the S21 parameter, most objects were indistinguishable. Several targets whose S21 parameters are shown in Figure 6(a) displayed traces that can be manually distinguished over the medium-to-high frequency range. Each of these objects was metallic with large cross-sectional areas. This is expected from objects with high conductivity and a large enough area to create conditions for high reflectivity and scattering. On the other hand, several objects, shown in Figure 6(b), have no readily identifiable characteristics in the S21 parameter that allow us to manually distinguish and identify them. We note that the high-frequency limit (6–8 GHz) of our experiment was unreliable for manual target identification due to the high variance between experiments in the cases of low radar cross-sections (RCSs).

Figure 7 examines the angle dependence of the S21 parameter for each object. Certain objects, such as object 1 in



**FIGURE 4.** A schematic of the experimental setup. The top view of the anechoic chamber is shown. The “target” indicated in the figure represents any of the 17 different objects studied in this project; the 17 objects are shown in Figure 5. Relative proportions are drawn to scale. Scale bar: 1 ft. PCB: printed circuit board.

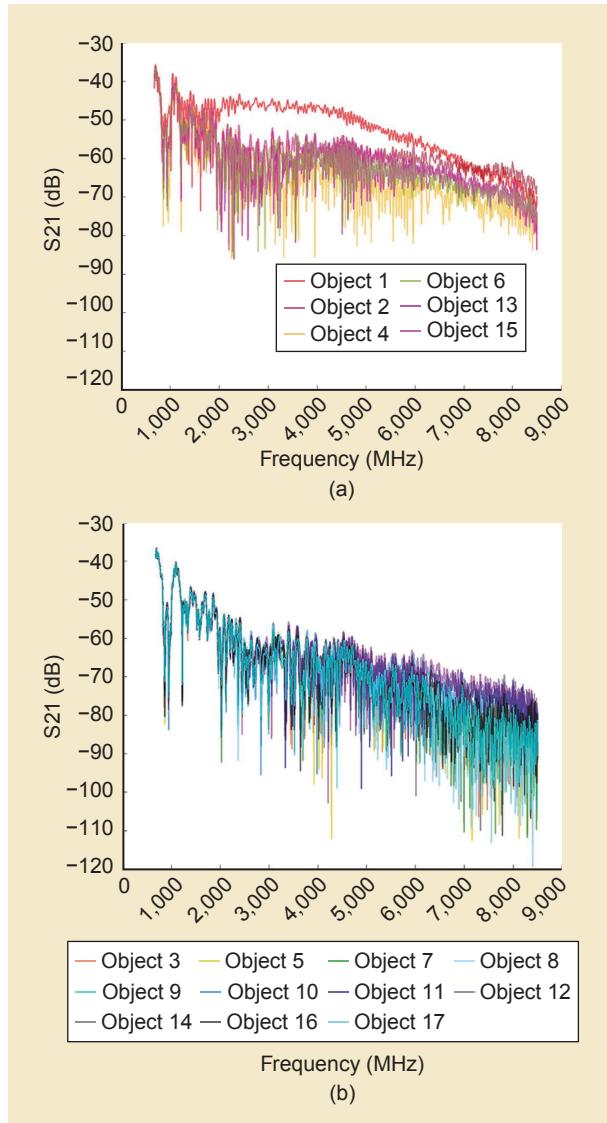
Figure 7(a), reflect the radio waves in different ways depending on orientation. Therefore, once an object is identified, its orientation can be determined from the log-magnitude plot of the S21 parameter. This was the case for relatively large



**FIGURE 5.** The 17 targets are shown on the platform. Object size can be gauged by the 47-cm-wide and 30.5-cm-deep cardboard platform. The 0°, 45°, and 90° orientations are shown in (ii), (iii), and (iv), respectively. The red arrows (i) depict the orientations on the cardboard. Unless otherwise noted, the photos are taken from the radar's perspective at a 45° orientation. The objects vary in size and composition, affecting the rRCS. (0) Platform: Roughly 60° from the radar's perspective. (1) Object 1: An empty metal (copper) box with a hole on the top face. (2) Object 2: A metal box (closed cover) with a home-built circuit. (3a) Object 3: A plastic toolbox (closed). (3b) Object 3: Open. Data were collected with the toolbox closed. (4) Object 4: A metal box (open cover) with a custom circuit. (5a) Object 5: A plastic box (closed). (5b) Object 5: Open. Data were collected with the box closed. (6) Object 6: A dc power supply (metal cover). (7) Object 7: The front cover of a power amplifier (metal) at 90° orientation. (8) Object 8: Rogers duroid laminate (copper) at 90° orientation. (9) Object 9: A data transfer switch box (with a plastic cover). (10) Object 10: A variable capacitor box (with a metal cover). (11) Object 11: A data transfer switch box (with a metal cover). (12) Object 12: A port converter (metallic). (13) Object 13: A data transfer switch box (with a plastic and metal cover). (14) Object 14: A vise (metal) at 0° orientation. (15) Object 15: A metal box. (16) Object 16: A chemistry hotplate stirrer. (17) Object 17: A Black & Decker drill at 0° orientation.

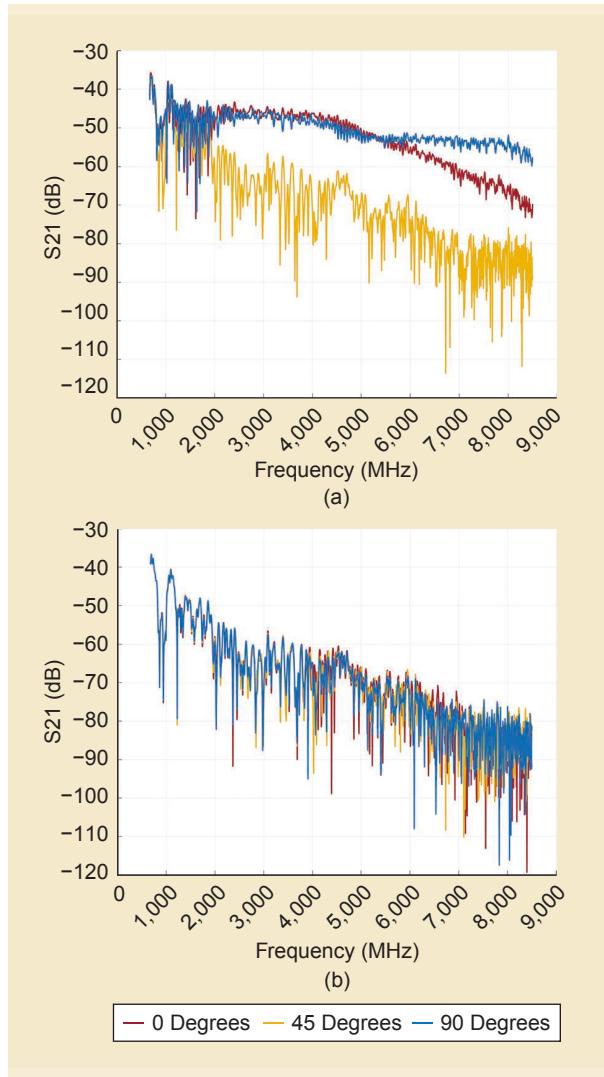
metallic objects that possess low symmetry and present a unique RCS at each orientation. An exception was object 1, whose 0° and 90° orientations have similar cross-sections while exhibiting unique traces (S21 parameters in log-magnitude mode). A possible explanation for this exception is that a crimped copper seam was present on the right and left sides of the object for the 0° and 90° orientations, respectively. This seam likely reflects and scatters radio waves differently depending on its orientation, allowing each orientation's trace to be manually distinguished.

However, other objects displayed few or no clear differences between orientations in the low-frequency low-noise region,



**FIGURE 6.** S21 parameters in log-magnetic mode for all 17 objects at their respective 0° orientations. Traces show the raw data as acquired without any postprocessing or averaging. (a) Several of the targets (objects 1, 2, 4, 6, 13, and 15) have S21 parameters with obvious differences in the log-magnitude plot. (b) Other targets (objects 3, 5, 7, 8, 9, 10, 11, 12, 14, 16, and 17) exhibit S21 parameters with no obvious visual differences.

as shown in Figure 7(b). In our experiment, objects with the least-distinguishable orientations tended to be those presenting small RCS areas, regardless of material composition. Small RCSs imply weak scattering intensity and overshadowing of the object's characteristics in the log-magnitude plot. Such objects with traces that cannot be manually identified present a unique challenge to radar target identification. This is generally the case for targets that present small RCSs due to their sizes, reflective properties, or range of the measurement. This apparent difficulty drives the need to develop DL algorithms for radar target identification.



**FIGURE 7.** S21 parameters for two targets (objects 1 and 17) at three different orientations (the objects and a definition of orientation are indicated in Figure 5). No noise removal or averaging was performed on the traces shown. (a) Object 1 is shown at 0°, 45°, and 90° orientations, each of which is easily identifiable. Objects 2, 4, 6, and 13 (not shown here) also demonstrate easily distinguishable orientations. (b) Object 17 is shown at 0°, 45°, and 90° orientations, each of which is indistinguishable. Orientations of objects 9, 10, 11, 14, and 16 (not shown here) were also visually indistinguishable from their log-magnitude plots.

## RESULTS

Experiments were designed to test the aspects of accuracy, consistency, robustness, and wall clock time of the different DL approaches. C-MSN demonstrated substantial improvement in all aspects over other approaches, with accuracy exceeding the SqueezeNet committee by 9% and the CNN committee by 20%.

### ACCURACY

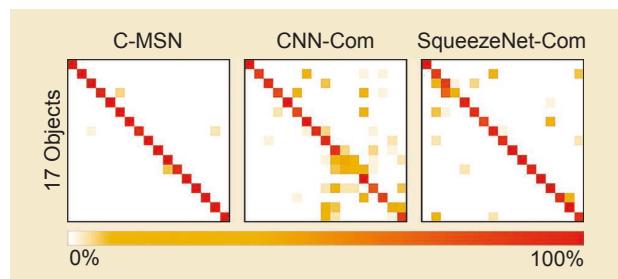
Leave-one-out cross-validation (LOOCV) was used for testing the accuracy of each approach. LOOCV is a special case of  $K$ -fold cross-validation where one sample is left out at a time, and  $K$  is equal to the number of samples [82]. This is not to be confused with the validation data partition used during training. We performed 12 combinations of 11 samples per object for training/one sample per object for testing. The process was repeated five times for each combination for a total of 60 trials per accuracy measurement. The average accuracy of the 60 trials is presented in Table 1. The confusion matrices for C-MSN, the CNN Committee, and the SqueezeNet Committee are shown in Figure 8.

### CONSISTENCY

The standard deviation of the individual accuracies from the 60 trials was calculated to assess the consistency of each approach, and the values are presented in Table 1. The C-MSN standard deviation is more than threefold lower compared to the other methods. Consistency is a highly desirable feature that reduces the number of times a model must be trained to achieve desirable performance, which increases implementation efficiency. C-MST clearly excels in this aspect, outperforming the other approaches by a large margin.

**TABLE 1. ACCURACY AND STANDARD DEVIATION COMPARISONS FOR ALL APPROACHES. EACH VALUE IS CALCULATED FROM 60 TRIALS. THE LOWEST AND HIGHEST PERFORMANCE RESULTS ARE HIGHLIGHTED IN RED AND GREEN COLORS, RESPECTIVELY.**

Accuracy	C-MSN	CNN-Com	SqueezeNet-Com
Average	99.02%	79.9%	90.29%
Standard deviation	2.21%	7.45%	7.25%



**FIGURE 8.** Confusion matrices for C-MSN, the CNN committee, and the SqueezeNet committee. Com: committee.

## ROBUSTNESS TO HYPERPARAMETERS

While second-order optimizers are robust to hyperparameter settings, their computational complexity is prohibitive for DL. Common DL algorithms instead rely on first-order (gradient descent)-based optimization, which is sensitive to the settings of hyperparameters such as the learning rate and the number of layers and neurons in the network [83]. As a distributive training algorithm, C-MST allows the efficient utilization of second-order optimization, making it inherently robust to hyperparameter settings.

## WALL CLOCK TIME

Despite the seeming complexity of C-MST, which involves hundreds of individual FCNs and CNNs, and although second-order training—which is typically associated with high computational complexity—is used, the approach is highly computationally efficient. The computational time in C-MST accelerates virtually linearly with the number of computational cores in a processor, as shown in Figure S3 in “Additional Architecture Details and Values of Hyperparameters.” Given enough cores, C-MST can run in much less wall clock time than the SqueezeNet committee approach despite the fact that transfer learning is used with a pretrained SqueezeNet. This is due to the highly distributable nature of C-MST, where individual networks in each stage can be trained independently and in parallel. This is also due to the gradual multistage convergence in C-MST, as demonstrated in Figure 2, where individual networks require a very small number of training epochs relative to standard approaches. Furthermore, the partial connectivity in C-MST typically results in a relatively small number of parameters for each individual network in inner stages.

## CONCLUSION

Herein we presented a new DL approach for RF classification and collected a new benchmarking dataset for proof-of-concept radar applications. The experiments conducted in this study confirm that while standard CNN approaches can work sufficiently well in some scenarios, their performance drops dramatically as the classification complexity increases. Since classification complexity increases with the number of classes, it is clear that standard DL approaches do not scale effectively in such RF classification applications. In contrast, C-MST is more stable and demonstrates significantly higher performance across all experiments. Most notably, C-MST achieves a 99% accuracy, representing a substantial advantage over the other approaches. Additionally, C-MST is robust, computationally efficient, and highly distributable. Therefore, C-MST effectively scales with computational complexity and training time as the number of classes increases. We propose C-MST as a scalable end-to-end

**When simple noise augmentation is applied, C-MSN is capable of achieving 100% accuracy.**

approach suitable to the nature and challenges of RF data.

We note that while the main purpose of this study is to compare the accuracy and generalization ability of DL approaches when trained with an under-sized training dataset, C-MST is equally applicable to preprocessed data and to other radar modalities, including 2D modal-

ties such as synthetic aperture radar, where the input dimensions of the front-end CNN stage can be adjusted accordingly. An additional demonstration applying our approach to a regression problem where the RCS of an object is estimated from a synthetic (simulated) radar dataset has been included in “Additional Architecture Details and Values of Hyperparameters.” Hyperparameters such as the number of stages and number of networks per stage can also be further adjusted according to the problem complexity.

We also note that additional preprocessing and augmentation techniques can improve the performance of all approaches discussed herein. When simple noise augmentation is applied, C-MSN is capable of achieving 100% accuracy. Such techniques were intentionally omitted in this study to isolate the model performance. (The comparison of models whose performance approaches 100% efficiency would be meaningless.)

Furthermore, our newly collected benchmarking dataset will be made publicly available to enable other groups to validate their work independently when applying our method to other challenging radar classification problems. Data collected from pulsed radar in the field generally include additional factors such as polarization; background clutter; time-domain acquisition; variable target-to-radar distance (range); moving targets; additional sources of noise; and interference and radar jamming. The effects of radar clutter were not included here as the main point of the study was to compare the classification accuracy of the new algorithm (C-MSN) to existing state-of-the-art algorithms under identical conditions. The numerous advantages offered by our approach will help improve RF-based signal classification performance under these challenging scenarios.

## ACKNOWLEDGMENT

This work was partially supported by NSF Grants CHE-2002313 and 1936375. The dataset is available at <https://doi.org/10.1109/MAP.2022.3208813>.



## AUTHOR INFORMATION

**Khalid Youssef** (kyoussef@ucla.edu) is with the Department of Chemistry and Biochemistry, University of California, Los Angeles, Los Angeles, CA 90095-1569 USA. He is a member of IEEE.

**Greg Schuette** (gregschuette32@gmail.com) is with the Department of Chemistry and Biochemistry, University of California, Los Angeles, Los Angeles, CA 90095-1569 USA.

**Yubin Cai** (micaohcai@g.ucla.edu) is with the Department of Electrical and Computer Engineering, University of California, Los Angeles, Los Angeles, CA 90095-1594 USA. He is a Member of IEEE.

**Daisong Zhang** (daisong@ucla.edu) is with the Department of Electrical and Computer Engineering, University of California, Los Angeles, Los Angeles, CA 90095-1594 USA. He is a Member of IEEE.

**Yikun Huang** (yikunh@gmail.com) is with the Department of Electrical and Computer Engineering, University of California, Los Angeles, Los Angeles, CA 90095-1594 USA. She is a Member of IEEE.

**Yahya Rahmat-Samii** (rahmat@ee.ucla.edu) is with the Department of Electrical and Computer Engineering, University of California, Los Angeles, Los Angeles, CA 90095-1594, USA. He is a Life Fellow of IEEE.

**Louis-S. Bouchard** (bouchard@chem.ucla.edu) is with the Department of Chemistry and Biochemistry, University of California, Los Angeles, Los Angeles, CA 90095-1569 USA.

## REFERENCES

- [1] S. M. Patole, M. Torlak, D. Wang, and M. Ali, "Automotive radars: A review of signal processing techniques," *IEEE Signal Process. Mag.*, vol. 34, no. 2, pp. 22–35, Mar. 2017, doi: 10.1109/MSP.2016.2628914.
- [2] K. W. O'Haver, C. K. Barker, G. D. Dockery, and J. D. Huffaker, "Radar development for air and missile defense," *Johns Hopkins APL Tech. Dig.*, vol. 34, no. 2, pp. 140–153, 2018.
- [3] J. Muñoz-Ferreras, Z. Peng, R. Gómez-García, and C. Li, "Review on advanced short-range multimode continuous-wave radar architectures for healthcare applications," *IEEE J. Electromagn., RF, Microw. Med. Biol.*, vol. 1, no. 1, pp. 14–25, Jun. 2017, doi: 10.1109/JERM.2017.2735241.
- [4] N. Joshi et al., "A review of the application of optical and radar remote sensing data fusion to land use mapping and monitoring," *Remote Sens.*, vol. 8, no. 1, p. 70, Jan. 2016, doi: 10.3390/rs8010070.
- [5] W. W.-L. Lai, X. Derobert, and P. Annan, "A review of ground penetrating radar application in civil engineering: A 30-year journey from locating and testing to imaging and diagnosis," *NDT E Int.*, vol. 96, pp. 58–78, Jun. 2018, doi: 10.1016/j.ndteint.2017.04.002.
- [6] P. Tait, *Introduction to Radar Target Recognition*, vol. 18. London, U.K.: IET, 2005.
- [7] J. Francke, "A review of selected ground penetrating radar applications to mineral resource evaluations," *J. Appl. Geophys.*, vol. 81, pp. 29–37, Jun. 2012, doi: 10.1016/j.jappgeo.2011.09.020.
- [8] S. Ahmed, A. Genghammer, A. Schiessl, and L.-P. Schmidt, "Fully electronic e-band personnel imager of 2 m<sup>2</sup> aperture based on a multistatic architecture," *IEEE Trans. Microw. Theory Techn.*, vol. 61, no. 1, pp. 651–657, Jan. 2013, doi: 10.1109/TMTT.2012.2228221.
- [9] A. Agurto, Y. Li, G. Tian, N. Bowring, and S. Lockwood, "A review of concealed weapon detection and research in perspective," in *Proc. 2007 IEEE Int. Conf. Netw., Sens. Control*, pp. 443–448, doi: 10.1109/ICNSC.2007.372819.
- [10] S. Sadeghi, K. Mohammadpour-Aghdam, R. Faraji-Dana, and R. Burkholder, "A DORT-uniform diffraction tomography algorithm for through-the-wall imaging," *IEEE Trans. Antennas Propag.*, vol. 68, no. 4, pp. 3176–3183, Apr. 2020, doi: 10.1109/TAP.2019.2952002.
- [11] D. Almanza-Ojeda, A. Hernandez-Gutierrez, and M. Ibarra-Manzano, "Design and implementation of a vehicular access control using RFID," in *Proc. 2006 Multiconf. Electron. Photon.*, pp. 223–225, doi: 10.1109/MEP.2006.335668.
- [12] M. Mishchenko, "Measurement and modeling of electromagnetic scattering by particles and particle groups," in *Polarimetry of Stars and Planetary Systems*, L. Kolokolova, J. Hough, and A.-C. Levasseur-Regourd, Eds. Cambridge, U.K.: Cambridge Univ. Press, 2015, pp. 13–34.
- [13] K. Youssef, L.-S. Bouchard, K. Haigh, J. Silovsky, B. Thapa, and C. Valk, "Machine learning approach to RF transmitter identification," *IEEE J. Radio Freq. Identif.*, vol. 2, no. 4, pp. 197–205, Dec. 2018, doi: 10.1109/JRFID.2018.2880457.
- [14] K. Youssef, N. N. Jarenwattananon, and L. Bouchard, "Feature-preserving noise removal," *IEEE Trans. Med. Imag.*, vol. 34, no. 9, pp. 1822–1829, Sep. 2015, doi: 10.1109/TMI.2015.2409265.
- [15] Feature-preserving noise removal, by K. Youssef and L. Bouchard. (2014). U.S. Patent US9953246B2. [Online]. Available: <https://patents.google.com/patent/US9953246B2>
- [16] D. Pearson, "Introduction," in *Image Processing*, D. Pearson, Ed. London, U.K.: McGraw-Hill, 1991, pp. 1–14.
- [17] L. Novak, G. Owirka, and W. Brower, "Performance of 10- and 20-target MSE classifiers," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 36, no. 4, pp. 1279–1289, Oct. 2000, doi: 10.1109/7.892675.
- [18] R. Heiden, "Aircraft recognition with radar range profiles," Ph.D. dissertation, Doctoral Thesis, Univ. of Amsterdam, Amsterdam, The Netherlands.
- [19] R. Heiden and F. Groen, "Distance based range profile classification techniques for aircraft recognition by radar - A comparison on real radar data," presented at the Mission Systems Panel 3rd Symp., Advisory Group for Aerospace Research and Development (AGARD; NATO), 17/1-1777, Lisbon, Portugal, May 15–18, 1995.
- [20] S. Hudson and D. Psaltis, "Correlation filters for aircraft identification from radar range profiles," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 29, no. 3, pp. 741–748, Jul. 1993, doi: 10.1109/7.220924.
- [21] J. Tang and Z. Zhu, "Comparison study on high resolution radar target recognition," in *Proc. IEEE Nat. Aerosp. Electron. Conf. (NAECON)*, May 1996, vol. 1, pp. 250–253, doi: 10.1109/NAECON.1996.517652.
- [22] R. Boyle and R. Thomas, *Computer Vision: A First Course*. Oxford, U.K.: Blackwell, 1988.
- [23] E. Davies, "Machine vision: Theory, algorithms, and practicalities," 2012. [Online]. Available: [https://books.google.com/books/about/Computer\\_and\\_Machine\\_Vision.html?hl=wl7ZjOwq8gC&source=kp\\_book\\_description](https://books.google.com/books/about/Computer_and_Machine_Vision.html?hl=wl7ZjOwq8gC&source=kp_book_description)
- [24] L. Roberts, *Machine Perception of Three-Dimensional Solids*. New York, NY, USA: Garland, 1963.
- [25] J. Tippett, *Optical and Electro-optical Information Processing*. Cambridge, MA, USA: MIT Press, 1965.
- [26] D. Vernon, *Machine Vision: Automated Visual Inspection and Robot Vision*. Hoboken, NJ, USA: Prentice Hall, 1991.
- [27] A. Jain, *Fundamentals of Digital Image Processing*. Englewood Cliffs, NJ, USA: Prentice Hall, 1989.
- [28] K. Rosenbach and J. Schiller, "Identification of aircraft on the basis of 2-D radar images," in *Proc. Int. Radar Conf.*, May 1995, pp. 405–409.
- [29] F. Zernike, "Beugungstheorie des schneidenverfahrens und seiner verbesserten form, der phasenkontrastmethode," *Physica*, vol. 1, nos. 7–12, pp. 689–704, May 1934, doi: 10.1016/S0031-8914(34)80259-5.
- [30] T. Kohonen, Ed. *Self-Organizing Maps*. Berlin, Heidelberg: Springer-Verlag, 1997.
- [31] R. Hu and Z. Zhu, "Researches on radar target classification based on high resolution range profiles," in *Proc. IEEE Nat. Aerosp. Electron. Conf. (NAECON)*, Jul. 1997, vol. 2, pp. 951–955, doi: 10.1109/NAECON.1997.622757.
- [32] B. Pei and Z. Bao, "Radar target recognition based on peak location of HRR profile and HMMs classifiers," in *Proc. IET Conf.*, Jan. 2002, pp. 414–418, doi: 10.1109/RADAR.2002.1174738.
- [33] A. Maki, K. Fukui, Y. Kawawada, and M. Kiya, "Automatic ship identification in ISAR imagery: An on-line system using CMSM," in *Proc. IEEE Radar Conf. (IEEE Cat. No. 02CH37322)*, Apr. 2002, pp. 206–211, doi: 10.1109/NRC.2002.999720.
- [34] M. Cassabaum, J. Rodriguez, J. Riddle, and D. Waagen, "Feature analysis using millimeter-wave real beam and doppler beam sharpening techniques," in *Proc. 5th IEEE Southwest Symp. Image Analysis Interpretation*, Apr. 2002, pp. 101–105, doi: 10.1109/IAI.2002.999898.
- [35] S. He, W. Zhang, and G. Guo, "High range resolution MMW radar target recognition approaches with application," in *Proc. IEEE Nat. Aerosp. Electron. Conf. (NAECON)*, May 1996, vol. 1, pp. 192–195, doi: 10.1109/NAECON.1996.517639.
- [36] C. Nieuwoudt and E. Botha, "Relative performance of correlation-based and feature-based classifiers of aircraft using radar range profiles," in *Proc. 14h Int. Conf. Pattern Recognit. (Cat. No. 98EX170)*, 1998, vol. 2, pp. 1828–1832, doi: 10.1109/ICPR.1998.712086.
- [37] B. Borden, *Radar Imaging of Airborne Targets: A Primer for Applied Mathematicians and Physicists*. Boca Raton, FL, USA: CRC Press, 1999.
- [38] R. Duda, P. Hart, and D. Stork, *Pattern Classification*. Hoboken, NJ, USA: Wiley, 2012.

[39] T. Bayes, "An essay towards solving a problem in the doctrine of chances," *Philosophical Trans. Roy. Soc.*, vol. 53, pp. 370–415, Dec. 31, 1794.

[40] M. DeGroot, *Optimal Statistical Decisions*, vol. 82. Hoboken, NJ, USA: Wiley, 2005.

[41] D. Sivia and J. Skilling, *Data Analysis: A Bayesian Tutorial*. Oxford, U.K.: Oxford Univ. Press, 2006.

[42] A. Gelman, J. Carlin, H. Stern, D. Dunson, A. Vehtari, and D. Rubin, *Bayesian Data Analysis*. Boca Raton, FL, USA: CRC Press, 2013.

[43] P. Lee, *Bayesian Statistics: An Introduction*. Hoboken, NJ, USA: Wiley, 2012.

[44] G. D'Agostini, *Bayesian Reasoning in Data Analysis: A Critical Introduction*. Singapore: World Scientific, 2003.

[45] J. Yan, X. Feng, and P. Huang, "Bayes-optimality based feature transform for high resolution range profile identification," in *Proc. 6th Int. Conf. Signal Process.*, Aug. 2002, vol. 2, pp. 1396–1398, doi: 10.1109/ICOSP.2002.1180053.

[46] D. Zhou, L. Wu, and G. Liu, "Bayesian classifier based on discretized continuous feature space," in *Proc. 4th Int. Conf. Signal Process. (Cat. No. 98TH8344) ICSP '98*, Oct. 1998, vol. 2, pp. 1225–1228, doi: 10.1109/ICOSP.1998.770839.

[47] Y. LeCun and Y. Bengio, *Convolutional Networks for Images, Speech, and Time-Series*. Cambridge, MA, USA: MIT Press, 1995.

[48] A. Y. Ng, "Feature selection,  $L_1$  vs.  $L_2$  regularization, and rotational invariance," in *Proc. 21st Int. Conf. Mach. Learn.*, New York, NY, USA: Association for Computing Machinery, 2004, p. 78, doi: 10.1145/1015330.1015435.

[49] Hierarchical constrained automatic learning network for character recognition, by J. Denker, R. Howard, L. Jackel, and Y. LeCun. (1991, Oct.). U.S. Patent US5058179A. [Online]. Available: <https://patents.google.com/patent/US5058179A>

[50] System and method for addressing overfitting in a neural network, by G. Hinton, A. Krizhevsky, I. Sutskever, and N. Srivastva. (2016, Aug. 2). U.S. Patent US9406017B2. [Online]. Available: <https://patents.google.com/patent/US9406017B2>

[51] F. Fleuret, T. Li, C. Dubout, E. K. Wampler, S. Yantis, and D. Geman, "Comparing machines and humans on a visual categorization test," *Proc. Nat. Acad. Sci.*, vol. 108, no. 43, pp. 17,621–17,625, Oct. 2011, doi: 10.1073/pnas.1109168108.

[52] S. Dodge and L. Karam, "A study and comparison of human and deep learning recognition performance under visual distortions," in *Proc. 2017 IEEE 26th Int. Conf. Comput. Commun. Netw. (ICCCN)*, pp. 1–7, doi: 10.1109/ICCCN.2017.8038465.

[53] T. Oyedare and J.-M. Park, "Estimating the required training data-set size for transmitter classification using deep learning," in *Proc. IEEE Int. Symp. Dyn. Spectr. Access Netw.*, 2019, pp. 1–10, doi: 10.1109/DySPAN.2019.8935823.

[54] D. C. Ciresan, U. Meier, L. M. Gambardella, and J. Schmidhuber, "Convolutional neural network committees for handwritten character classification," in *Proc. 2011 Int. Conf. Document Anal. Recognit.*, pp. 1135–1139, doi: 10.1109/ICDAR.2011.229.

[55] D. Cireşan, U. Meier, J. Masci, and J. Schmidhuber, "A committee of neural networks for traffic sign classification," in *Proc. 2011 Int. Joint Conf. Neural Netw.*, pp. 1918–1921, doi: 10.1109/IJCNN.2011.6033458.

[56] Z.-Q. Zhao, D.-S. Huang, and B.-Y. Sun, "Human face recognition based on multi-features using neural networks committee," *Pattern Recognit. Lett.*, vol. 25, no. 12, pp. 1351–1358, Sep. 2004, doi: 10.1016/j.patrec.2004.05.008.

[57] B. M. Ozyildirim and M. Kiran, "Do optimization methods in deep learning applications matter?" Feb. 2020, *arXiv:2002.12642*.

[58] Q. V. Le, J. Ngiam, A. Coates, A. Lahiri, B. Prochnow, and A. Y. Ng, "On optimization methods for deep learning," in *Proc. 28th Int. Conf. Mach. Learn.*, Madison, WI, USA: Omnipress, 2011, p. 265–272.

[59] R. Battiti, "First- and second-order methods for learning: Between steepest descent and newton's method," *Neural Comput.*, vol. 4, no. 2, pp. 141–166, Mar. 1992, doi: 10.1162/neco.1992.4.2.141.

[60] H. H. Tan and K. H. Lim, "Review of second-order optimization techniques in artificial neural networks backpropagation," in *Proc. IOP Conf. Series, Mater. Sci. Eng.*, Jun. 2019, vol. 495, p. 012003, doi: 10.1088/1757-899X/495/1/012003.

[61] Y. A. LeCun, L. Bottou, G. B. Orr, and K.-R. Müller, *Neural Networks: Tricks of the Trade*, 2nd ed. Berlin, Heidelberg: Springer-Verlag, 2012, ch. 1, pp. 9–48.

[62] B. M. Wilamowski and H. Yu, "Improved computation for Levenberg–Marquardt training," *IEEE Trans. Neural Netw.*, vol. 21, no. 6, pp. 930–937, Jun. 2010, doi: 10.1109/TNN.2010.2045657.

[63] H. Yu and B. Wilamowski, "Levenberg–Marquardt training," in *The Industrial Electronics Handbook*, vol. 5, 2nd ed. Boca Raton, FL, USA: CRC Press, Jan. 2011, ch. 12, pp. 12-1–12-16.

[64] I. Goodfellow et al., "Generative adversarial nets," in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, pp. 2672–2680.

[65] J. Guo, B. Lei, C. Ding, and Y. Zhang, "Synthetic aperture radar image synthesis by using generative adversarial nets," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 7, pp. 1111–1115, Jul. 2017, doi: 10.1109/LGRS.2017.2699196.

[66] D. Nandagopal, N. Martin, R. Johnson, P. Lozo, and M. Palaniswami, "Performance of radar target recognition schemes using neural networks – A comparative study" in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP '94)*, Apr. 1994, vol. 2, pp. II/641–II/644, doi: 10.1109/ICASSP.1994.389574.

[67] Z. Li, Z. Weida, and J. Licheng, "Radar target recognition based on support vector machine," in *Proc. 5th Int. Conf. Signal Process. 16th World Comput. Congr. (WCC - ICSP)*, Aug. 2000, vol. 3, pp. 1453–1456.

[68] P. Bharadwaj, P. Runkle, L. Carin, J. A. Berrie, and J. A. Hughes, "Multiaspect classification of airborne targets via physics-based HMMs and matching pursuits," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 37, no. 2, pp. 595–606, Apr. 2001, doi: 10.1109/7.937471.

[69] M. Jahangir, K. Ponting, and J. O'Loghlen, "Correction to 'Robust Doppler classification technique based on hidden Markov models,'" *Inst. Electr. Eng. Proc. Radar, Sonar Navig.*, vol. 150, no. 5, p. 387, Oct. 2003, doi: 10.1049/ip-rsn:20030569.

[70] T. Cooke, N. Redding, J. Schroeder, and J. Zhang, "Target discrimination in complex synthetic aperture radar imagery," in *Proc. 34th Conf. Rec. Asilomar Conf. Signals, Syst. Comput. (Cat. No. 00CH37154)*, Oct. 2000, vol. 2, pp. 1540–1544, doi: 10.1109/ACSSC.2000.911248.

[71] P. Han, R. Wu, Y. Wang, and Z. Wang, "An efficient SAR ATR approach," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP '03)*, Apr. 6–10, 2003, vol. 2, pp. II–429, doi: 10.1109/ICASSP.2003.1202392.

[72] M. Seyfoglu, A. Özbayoglu, and S. Gürbüz, "Deep convolutional autoencoder for radar-based classification of similar aided and unaided human activities," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 54, no. 4, pp. 1709–1723, Aug. 2018, doi: 10.1109/TAES.2018.2799758.

[73] D. Avrahami, M. Patel, Y. Yamaura, and S. Kratz, "Below the surface: Unobtrusive activity recognition for work surfaces using RF-radar sensing," in *Proc. 23rd Int. Conf. Intell. User Interf.*, ACM, 2018, pp. 439–451, doi: 10.1145/3172944.3172962.

[74] B. Jokanović and M. Amin, "Fall detection using deep learning in range-doppler radars," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 54, no. 1, pp. 180–189, Feb. 2018, doi: 10.1109/TAES.2017.2740098.

[75] S. Chen, H. Wang, F. Xu, and Y. Jin, "Target classification using the deep convolutional networks for SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 8, pp. 4806–4817, Aug. 2016, doi: 10.1109/TGRS.2016.2551720.

[76] D. Morgan, "Deep convolutional neural networks for ATR from SAR imagery," in *Proc. Algorithms Synthetic Aperture Radar Imagery XXII*, International Society for Optics and Photonics, 2015, vol. 9475, p. 94750.

[77] J. Lien et al., "Soli: Ubiquitous gesture sensing with millimeter wave radar," *ACM Trans. Graph. (TOG)*, vol. 35, no. 4, p. 142, Jul. 2016, doi: 10.1145/2897824.2925953.

[78] S. Wang, J. Song, J. Lien, I. Poupyrev, and O. Hilliges, "Interacting with soli: Exploring fine-grained dynamic gesture recognition in the radio-frequency spectrum," in *Proc. 29th Annu. Symp. User Interf. Softw. Technol.*, ACM, 2016, pp. 851–860, doi: 10.1145/2984511.2984565.

[79] X. Cao, X. Zhang, R. Togneri, and Y. Yu, "Underwater target classification at greater depths using deep neural network with joint multiple-domain feature," *IET Radar, Sonar Navig.*, vol. 13, no. 3, pp. 484–491, Mar. 2019, doi: 10.1049/iet-rsn.2018.5279.

[80] D. Malmgren-Hansen, R. Engholm, and M. O. Pedersen, "Training convolutional neural networks for translational invariance on SAR ATR," in *Proc. 2016 11th Eur. Conf. Synthetic Aperture Radar (EUSAR)*, pp. 1–4.

[81] F. N. Iandola, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally, and K. Keutzer, "SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5MB model size," 2016. [Online]. Available: <https://arxiv.org/abs/1602.07360>

[82] A. Elisseeff et al., "Leave-one-out error and stability of learning algorithms with applications," in *Proc. NATO Sci. Series. III, Comput. Syst.*, 2003, vol. 190, pp. 111–130.

[83] P. Xu, F. Roosta, and M. W. Mahoney, "Second-order optimization for non-convex machine learning: An empirical study," in *Proc. 2020 SIAM Int. Conf. Data Mining*, pp. 199–207, doi: 10.1137/1.9781611976236.23.

