# Sparse Signal Models for Data Augmentation in Deep Learning ATR

Tushar Agarwal, Nithin Sugavanam and Emre Ertin

***Abstract*—Automatic Target Recognition (ATR) algorithms classify a given Synthetic Aperture Radar (SAR) image into one of the known target classes using a set of training images. Recently, learning methods have shown to achieve state-of-the-art classification accuracy if abundant training data is available sampled uniformly over the classes and their poses. In this paper, we consider the problem of improving the generalization performance of learning methods in SAR-ATR when training data is limited. We propose a data augmentation approach using sparse signal models that capitalizes on commonly observed phenomenology of wide-angle synthetic aperture radar (SAR) imagery. Specifically, we exploit the sparsity of the scattering centers in the spatial domain as well as the limited persistence of the scattering coefficients in the azimuthal domain to solve the ill-posed problem of over-parametrized model fitting. Using this fitted model, we synthesize new images at poses not available in training set to augment the training data used by CNN. We validate the performance of the proposed model based data augmentation strategy on subsampled versions of the MSTAR dataset. The experimental results show that for the training data starved region, the proposed method provides a significant gain in the generalization performance of the resulting ATR algorithm.**

***Index Terms*—Deep Learning, Automatic Target Recognition, Data Augmentation, Compressive sensing.**

## I. Introduction

Synthetic aperture radar (SAR) sensing has been widely utilized in obtaining high-resolution imagery of a region of interest, that is robust to weather and other environmental factors. The SAR sensor consists of a moving radar platform with a collocated receiver and transmitter that traverses a wide aperture in azimuth, acquiring coherent measurements. Multiple pulses across the synthesized aperture are combined and coherently processed to produce high-resolution SAR imagery. A significant fraction of the energy in the back-scattered signal from the scene is due to a small set of dominant scattering centers, that are resolved by the SAR sensor. The localization of backscatter energy provides a distinct description of targets of interest [1], typically man-made objects such as civilian and military vehicles. This sparsity structure has been utilized in [2], [3] to design features like peak locations, edges that succinctly represent the scene. These hand-crafted features, in conjunction with template based methods or statistical methods, are subsequently used in solving the target recognition problem. The template based methods exploit the geometric structure and variability of these features in the scattering centers in [4], [5] to distinguish between the different target categories. The scattering center based target signature varies with the pose angle of the target with respect to the sensor platform. Statistical methods can explicitly model and utilize this low-dimensional manifold structure of the scattering center descriptors [6], [7] for improved decisions as well as integrating information across views [8], [9].

However, ATR algorithms based on these hand-crafted features are limited to the information present in these descriptors, and lack the generalization ability to variability in clutter, pose, and noise. With the advent of data-driven algorithms such as artificial neural networks (ANN), an appropriate feature set and a discriminating function can be jointly estimated using a unified objective-function. Recent advances in techniques to incorporate deep hierarchical structures used in ANN [10] has led to the widespread use of these methods to solve inference problems in a diverse set of application areas. Convolutional Neural Networks (CNNs), in particular, have been used as an automatic feature extractor for image data. Such methods have also been adopted in solving the automatic target recognition (ATR) problem using SAR images [11], [12]. These results establish that the CNNs could be effective in radar image classification as well, provided sufficient training data is available. But in general, labeled radar data is not available in abundance, unlike other image datasets. In this paper, we address the problem of scarcity of training data and provide a general method that utilizes a model-based approach to capture the underlying scattering phenomenon to enrich the training dataset.

One of the most effective techniques to tackle the problem of small datasets is the use of transfer learning. Transfer-learning implies using the model parameters, estimated using a similar dataset, as initialization for solving the problem of interest. The most common example of this is, re-using the CNN weights of a model trained on a vast labeled dataset like Image-net [13]. It is conjectured that as natural-images have several similar characteristics, the same feature extractors can be used with little to no fine-tuning. There have been numerous experiments supporting the claim, including the seminal work in [14]. In contrast, radar images are significantly different from the regular optical images. In particular, SAR works in the wavelength of $1cm.$ to $10m.$ while visible light has a wavelength of the order of $1nm.$. As a result, most surfaces in natural scenes are rough for light waves of such small wavelengths, leading to diffused reflections. In contrast, microwaves from radar transmitters undergo specular reflections, especially for man-made objects. This difference

T. Agarwal, N. Sugavanam & E. Ertin are with the Department of Electrical & Computer Engineering, The Ohio State University, Columbus, OH, USA. Corresponding Author: T. Agarwal (agarwal.270@buckeyemail.osu.edu)

978-1-7281-8942-0/20/$31.00 ©2020 IEEE

978-1-7281-8942-0/20/$31.00 ©2020 IEEE

in scattering behavior leads to substantially different images in SAR when compared to optical imaging. Being based on specular reflections also means that the images are much more sensitive to instantaneous factors like the orientation of the imaging device and background clutter. It implies that the readily available optical imagery-based deep neural network models like Alex-net and VGG16 [15] are not suitable for transferring knowledge to this domain. Retraining the network is again inhibited by the limited availability of radar data. So, we explore an alternative strategy to transfer learning, based on data augmentation of radar image datasets using a principled approach of augmenting the dataset that respects the phenomenology of the RF back-scatter data. The paper is structured as follows. We first review relevant research work and outline our contributions in sections I-A and I-B, respectively. In section II-A, we describe the dataset and network architecture in detail. The details of our pose-synthesis methodology is described in section II-B. This is followed by the experiments and corresponding results in sections III and IV, respectively, which provide the empirical evidence for the effectiveness of the proposed data augmentation method. We conclude with some possible directions for future research in section V.

### A. Related Work

The problem of over-fitting in ANNs magnifies when the size of the training set is small. Several methods have been proposed to reduce over-fitting and improve generalization performance. These ill-posed problems are typically solved by using regularizers that impose structure and constraints in the solution space. Apart from standard norm penalties ($\ell_1$ and $\ell_2$), Dropout [16], Batch Normalization [17], over-parameterization [18] and variants of stochastic gradient descent (SGD) [19] induce regularization, either explicitly or implicitly. Transfer Learning is another approach used for improving generalization performance in limited availability of data [20], [21]. For SAR data, there exist several ANN architecture-based approaches to improve generalization [22], [23], [24]. Few other approaches focus on learning special type of features [25]–[29]. Data Augmentation is considered yet another regularization strategy because it serves to reduce the generalization error while not affecting the training error [30]–[35]. This is the main focus of our paper and involves augmenting the training data-set by transforming original data through domain-specific transformations. In [12], Zhong et al. exploited the symmetric nature of most objects in the MSTAR Dataset by adding flip-augmentations with a reversed sign of the azimuthal angle. They also used the pose (azimuthal angle) prediction as a secondary objective for training the network. They empirically showed that this helps by adding meaningful constraints to the network learning. There have been several other models proposed to tackle the ATR problem with the MSTAR data-set, including [36] and [37], but we primarily build upon Zhong et al.'s [12] work.

### B. Contributions

In this paper we introduce a principled approach for pose synthesis based on sparse modeling of available images in the training data. The modeling is performed in phase history domain, after images have been transformed to k-space. We then fit an over-parameterized model inspired by the scattering behavior of canonical reflectors. Specifically, our model captures the viewing-angle dependent anisotropic scattering behavior of scattering centers that makes up the target and realistically model nearby poses. Additionally, our model allows us to generate scintillation effects due to sub-pixel shifts in the augmented images. These sub-pixel shifts are typically not possible in the traditional image domain operations. We hypothesize that these two factors are essential to improving the network's knowledge about the underlying phenomenology of the SAR imaging process. We validate the effectiveness of the proposed method through empirical results, which show a significant boost in generalization performance. It is important to note that our data-augmentation based strategy is generic and decoupled with the network architectures proposed in other works like [23], [22]. We hypothesize that the proposed augmentation strategy, in conjunction with the above-mentioned methods, may further improve results. Our objective here is to demonstrate the benefits of the proposed data-augmentation strategy by fixing the network structure similar to a well studied CNN model by Zhong et al. [12].

## II. A Phase-History Model based Data Augmentation

We want to learn a parametric Neural Network classifier, with parameters $w \in \mathbb{R}^{d_w}$, that predicts an estimate of output labels, $Y \in \mathbb{R}^{d_Y}$ for an input $X \in \mathbb{C}^{d_X}$ , i.e. $\hat{Y} = g(X; w)$ where $d_X$, $d_w$ and $d_Y$ are dimensions of $X$, $w$ and $Y$, respectively. We consider a supervised learning setting, where a labeled training data-set $\mathcal{D}_{train} = \{(X_u, Y_u)\}_{u=1}^{N_{train}}$ is used to estimate classifier parameters $w$, where $N_{train}$ are total number of training samples. The training procedure is the minimization of an appropriate loss function $\mathcal{L} : (w, \mathcal{D}) \to \mathbb{R}$ using an iterative algorithm like Stochastic Gradient Descent. Therefore, the learned $w^*$ are the solution of the following minimization problem.

$$w^* = \mathcal{P}(\mathcal{D}) = \arg\min_{w} \mathcal{L}(w, \mathcal{D}) \tag{1}$$

Data augmentation involves applying an appropriate transformation $T : \mathcal{D}_{in} \to \mathcal{D}_{out}$ to a data-set (only $\mathcal{D}_{train}$ for our purposes) and hence expand it to an augmented data-set $T(\mathcal{D}_{train})$. We also have a validation data-set, $\mathcal{D}_{val} = \{(X_u, Y_u)\}_{u=1}^{N_{val}}$ for cross-validation during training and a test data-set, $\mathcal{D}_{test} = \{(X_u, Y_u)\}_{u=1}^{N_{test}}$ for evaluating $g(X; w)$ post-training. The evaluation can be done using a suitable metric $\mathcal{M} : (w, \mathcal{D}) \to \mathbb{R}$ which maybe different from $\mathcal{L}$ above. Our aim is to find $T$ such that the learnt parameters $w_{aug} = \mathcal{P}(T(\mathcal{D}_{train}))$ perform better than $w_{train} = \mathcal{P}(\mathcal{D}_{train})$ in terms of the chosen metric, i.e. $\mathcal{M}(w_{aug}, \mathcal{D}_{test})$ is more desirable than $\mathcal{M}(w_{train}, \mathcal{D}_{test})$.

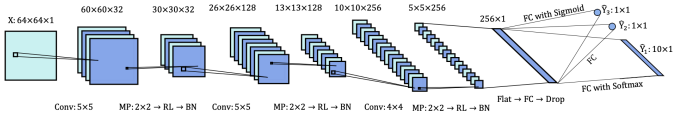


Fig. 1: The Neural Network Architecture. The abbreviations used are as follows. Conv is Convolutional Layer followed by the kernel height × width. MP is max-pooling followed by the pooling-size as height × width. RL, BN, Flat, Drop and FC are ReLU, Batch-Normalization, Flattening, Dropout and Fully-Connected layers respectively. Sizes of feature maps are mentioned at their top as height × width × channels.

### A. Data-set & Neural Network Model

The MSTAR data-set consists of 10 classes of various military vehicles. We use a balanced subset of the data-set, which is referred to as standard operating conditions and is also used by ( [22], [21]). We denote this subset as SOC MSTAR data-set henceforth. Similar to the existing literature, we use the images at depression angle, $\phi = 17^\circ$ for training while images at $\phi = 15^\circ$ form the test set. Similar to [12], we crop the images to $64 \times 64$ with the objects in the center. Note that we crop the images right before feeding it to the ANN. We perform the modeling and augmentation steps on the original images. To investigate the effects of data augmentation, we artificially sub-sample SOC data-set at $\phi = 17^\circ$ by extracting only $R$ ratio of samples from each class where $R \in \{2^{-5}, 2^{-4}, 2^{-3}, 2^{-2}, 2^{-1}, 2^0\}$. We further select $15\%$ of uniformly distributed samples from this uniformly sub-sampled data as the validation set and utilize the remaining $85\%$ as the new training set. The training-data $\mathcal{D}_{train}$ includes real-time pixel-level translations [32] in range and cross-range domain as well as the flip augmentation along cross-range domain [12] and is just referred to as data or data-set. We form our $T(\mathcal{D}_{train})$ by performing the proposed pose augmentation on each radar image in the training data-set as described in the next section II-B. We include flip-augmentation in the final validation set $\mathcal{D}_{val}$ and no augmentations are included in the final test-set $\mathcal{D}_{test}$.

We utilize a network architecture inspired by [12] and shown in Fig. 1. We modify the cosine cost used for pose-awareness in [12] to a pair of simple costs using features $Y_2 = \sin(\theta)$ and $Y_3 = \mathbb{1}_A(\theta)$ where $\theta$ is the azimuthal angle and $\mathbb{1}_A$ is the indicator function over set $A = [\frac{-\pi}{2}, \frac{\pi}{2}]$. The loss function to find the network parameters is now

$$\begin{aligned}\mathcal{L}(w, \mathcal{D}) &= \bar{\mathbb{E}}_{\mathcal{D}}[\mathcal{L}_1(w, X, Y_1) + \mathcal{L}_2(w, X, Y_2) + \\ &\quad \mathcal{L}_3(w, X, Y_3)] \qquad (2)\\ \mathcal{L}_1(w, X, Y_1) &= -\sum_{p=1}^{10} Y_{1,p} \log\Big(\hat{Y}_{1,p}(w, |X|)\Big)\\ \mathcal{L}_2(w, X, Y_2) &= (Y_2 - \hat{Y}_2(w, |X|))^2\\ \mathcal{L}_3(w, X, Y_3) &= -Y_3 \log\Big(\hat{Y}_3(w, |X|)\Big)\\ &\quad - (1 - Y_3) \log\Big(1 - \hat{Y}_3(w, |X|)\Big)\end{aligned}$$

where $|.|$ denotes the absolute value, $|X| \in {\mathbb{R}_+}^{64\times64}$, $Y_1 \in \{0,1\}^{10\times1}$, $Y_2 \in [-1, 1]$, $Y_3 \in \{0, 1\}$ refer to absolute values of complex radar images, the one-hot vector of the 10 classes, $\sin(\theta)$ and $\mathbb{1}_A(\theta)$ respectively. $\bar{\mathbb{E}}_{\mathcal{D}}$ refers to the empirical mean over data-set $\mathcal{D}$ and $Y_{1,p}$ is the $p^{th}$ component of the vector $Y_1$. All the quantities with $\hat{}$ (hat) are the corresponding estimates given by the ANN.

### B. Pose Synthesis Methodology

This section describes the pose synthesis methodology used for data augmentation by estimating a Phase History (PH) model. The strategy is a modified version of our earlier work, which focused on modeling the scattering behavior of targets in monostatic and bistatic setup [38]–[40]. The images in the MSTAR data-set are not correctly registered. Therefore, we construct the model for each image and locally extrapolate the measurements. The SAR operating in spotlight mode has been used to create the images in the MSTAR data-set. The images are translated from the spatial domain to the cartesian frequency domain using the steps described in [41]. Subsequently, we convert the frequency measurements to the polar coordinates to obtain the phase-history measurements as described in [42]. We consider a square patch on the ground of side lengths $L = 30m$ with the target at the center. From the Geometric theory of diffraction, we assume that a complex target can be decomposed into a sparse set of scattering centers. The scattering centers are assumed to be $K$ point targets, described using $\{(x_k, y_k), h_k(\theta, \phi)\}_{k=1}^K$ where $(x_k, y_k) \in [-\frac{L}{2}, \frac{L}{2}] \times [-\frac{L}{2}, \frac{L}{2}]$ are the spatial coordinates of the point targets, $\theta$ is the azimuthal angle, $\phi$ is the angle of elevation of the radar platform and $h_k(\theta, \phi)$ are the corresponding scattering coefficients that depend on the viewing angle. Parametric models for standard reflectors such as dihedral and trihedral were studied in [43]–[45]. These models indicate that the reflectivity is a smooth function over the aspect angle parameterized by the dimensions and orientation of the reflector, which is not known apriori. Therefore, we exploit this smoothness to approximate this infinite-dimensional function using interpolation strategies [46] with the available set of samples, $\Theta$, in the angle domain. We denote the sampled returns from the scene by the matrix $\mathbf{S} = NUFFT(X) \in \mathbb{C}^{N_\theta \times M}$, $X \in C^{64x64}$ is the complex image from the dataset and NUFFT represents the non-uniform fourier transform. The elements of $\mathbf{S}$ are defined as follows

$$\begin{aligned}s(m, i) &= n(m, i) + \sum_{k=1}^{K} h_k(\theta_i, \phi)\\ &\exp\left(-j4\pi \frac{f_m \cos(\phi)}{c} \left(x_k \cos(\theta_i) + y_k \sin(\theta_i)\right)\right). \qquad (3)\end{aligned}$$

where $n(m, i)$ represents the measurement noise, $f_m$ such that $m \in [M]$ are the illuminating frequencies, $M = \frac{2BL}{c}$, $B$ is bandwidth of transmitted pulse, $c$ is the speed of light and the notation $[M]$ denotes enumeration of natural numbers upto $M$. We estimate the function $h_k(\theta, \phi)\ \forall\ k \in [K]$ from the receiver samples.

In order to solve the estimation problem, we assume that the function $h_k$ has a representation in a basis set denoted by the matrix $\mathbf{\Psi} \in \mathbb{C}^{N_\theta \times D}$ of size $D$. For the MSTAR

data-set, the elevation angles we work with are quite close. So we also assume that the variation in $h_k$ with respect to $\phi$ is insignificant. This assumption leads to the following relation $h_k(\theta;\phi) = \sum_{v=1}^{D} c_{v,k}\psi_v(\theta) + \epsilon_P$. Let vector $\mathbf{c_k} = [c_{1,k}\cdots c_{D,k}]^T$. The estimated phase-history matrix is now $\hat{\mathbf{S}}$ whose elements are given by

$$\hat{s}(m,i) = \hat{n}(m,i) + \sum_{k=1}^{K}\sum_{v=1}^{D} c_{v,k}\exp\left(-\left(\frac{\theta-\hat{\theta}_v}{2\sigma_G}\right)^2\right) \exp\left(-j4\pi\frac{f_m\cos(\phi)}{c}\left(x_k\cos(\theta_i)+y_k\sin(\theta_i)\right)\right), \tag{4}$$

where $\hat{n}(m,i)$ consists of the measurement noise and the approximation error. To estimate the coefficients $c_{v,k}$ from the noisy measurements in (4), we discretize the scene with resolution of $\Delta R$ in $x, y$ (range, cross-range) plane to get $K = N_R^2$ grid points, where $N_R = \frac{2BL}{c}$ is the number of range bins. Furthermore, we consider a smooth Gaussian function to perform the noisy interpolation. We partition the sub-aperture $2\Delta\theta$ into smaller intervals of equal length with a corresponding set containing the means of the intervals given by $\{\hat{\theta}_v\}_{v=1}^{D}$, where $D = 12$, which are used as the centroids for the Gaussian interpolating functions. We assume the width of the Gaussian function, $\sigma_G$ as a constant hyper-parameter whose selection is described in section III-A. Hence, $\sigma_G$ is the constant minimum persistence of the scattering center in azimuth that we wish to detect.

Here, the discrete grids for $(x_k, y_k)$ and $(\theta_i, f_m)$ are both known. Let the vectors containing all corresponding grid points for $x_k, y_k, \theta_i, f_m$ be referred as $\mathbf{x}, \mathbf{y}, \boldsymbol{\theta}, \mathbf{f}$ respectively. The problem now is to find the coefficients $c_{v,k}$ that minimizes some metric between $\hat{\mathbf{S}}$ and $\mathbf{S}$. To recover the structured signal $\mathbf{h} = [h_1\cdots h_{N_R^2}]$, which represents the scattering coefficient of a sparse scene that has a sparse representation in an underlying set of functions, we solve the following linear inverse problem using a sparse-group regularization on $\mathbf{c_k} \forall\ k \in [N_R^2]$.

$$\min_{\mathbf{C}}\left(\sum_{k=1}^{N_R^2}\lambda\left\|\mathbf{c}_k\right\|_2 + \left\|\mathbf{S}-\hat{\mathbf{S}}\right\|_F\right) \Longleftrightarrow \min_{\mathbf{C}} J(\mathbf{C},\sigma_G) \tag{5}$$

where $\mathbf{C}$ refers to the matrix $[\mathbf{c_1}\cdots\mathbf{c_{N_R^2}}]$, $\sigma_G$ is a constant hyper-parameter and $\|\cdot\|_F$ refers to the Frobenius norm. The phase-history measurements are converted back to the image using overlapping sub-apertures spanning 3 degrees in the azimuth domain as shown in Fig. 2. We apply the same Taylor window with zero-padding and translate it back to the Cartesian coordinates before applying the Fourier transform to generate the images to augment the data-set.

## III. Experiments

### A. Experimental Setup

The experiments were done using the network and data-sets, as described in section II-A. This model is run on a local machine with a Titan Xp GPU. The Tensorflow (1.10) library is used for the implementation through its python API. We use ReLU activation function everywhere except the final output layers of $\hat{Y}_1$, $\hat{Y}_2$, and $\hat{Y}_3$ where we use Softmax activation, Linear and Sigmoid activations respectively. We first bring the data to K-space and then invert the transformations done when the MSTAR data was prepared, to get the phase history representation. Using available information about the MSTAR dataset, we determine the discrete grids for $(x_k, y_k)$ and $(\theta_i, f_m)$. Now, the model coefficients are determined by solving the optimization problem described in equation (5). As a result, we obtain the model, say $F(\boldsymbol{\theta};\mathbf{f})$. This model is further used to synthesize new columns of phase-history data (or extending the $\boldsymbol{\theta}$ vector) and consequently produce a synthesized image by the procedure described in section II-B followed by transformation of phase history data to complex image data. This pose augmentation along-with sub-pixel shifts constitute our net augmentation transformation $T$. For training the network, we used the magnitude of the complex radar data as input, $X$, as we found it to be the most effective and efficient way for ATR with MSTAR data-set, in agreement with the existing literature. We normalize all input images to unit-norm to reduce some undesired effects due to the use of the Gaussian kernel in extrapolation. We also remove all the synthetic images at poses already in the corresponding training set. Then, the optimization problem in equation (1) is solved using the off-the-shelf, Adam variant of the Mini-batch Stochastic Gradient Descent optimizer with a mini-batch size of 64. In accordance with the early-stopping training regime, the training is run for a large no. of epochs ($> 400$), and the model is saved whenever it gets the best moving average validation performance with a moving average window of 1 to 4 epochs (depending upon the $R$ values). Although we use the accuracy as the performance metric $\mathcal{M}$, the $\mathcal{D}_{val}$ here get really small, especially for small R values. This makes the validation accuracy to easily saturate at $100\%$, yielding this metric useless. So instead, we monitor the minimum classification loss, $\mathcal{L}_1$, as the validation performance metric.

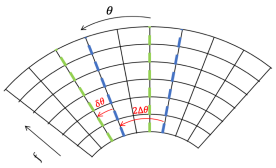


Fig. 2: Extrapolating by $\delta\theta$ using the model estimated from data.

We now explain our choices of a subset of hyper-parameters and mention some others. Rest were kept at the default values of the Tensorflow (1.10) library. The PH model has 2 main hyper-parameters, the $\sigma_G$ and $\delta\theta$. We determine optimum $\sigma_G^* = \arg\min_{\sigma_G}[\min_{\mathbf{C}} J(\mathbf{C},\sigma_G)]$ for every image by using a simple line-search. We choose appropriate $\delta\theta$ by running a grid search over a factor $\eta$ such that $\delta\theta = \eta\sigma_G^*$ where $\eta \in \{1,2,3\}$. For the ANN model, we set dropout rate at 0.2. We add real-time x-y translations to every image where each translation is randomly sampled from the set $\{-6,-4,-2,0,2,4,6\}$, at every epoch.

TABLE I: Test Accuracy from different Methods.

| Ratio (R) | Existing | Proposed | Full |
|---|---|---|---|
| $2^{-5}$ | 67.619 $\pm$ 2.926 | 81.34 $\pm$ 3.218 | 95.454 $\pm$ 0.572 |
| $2^{-4}$ | 81.216 $\pm$ 2.404 | 92.67 $\pm$ 0.543 | 97.928 $\pm$ 0.209 |
| $2^{-3}$ | 89.01 $\pm$ 0.73 | 96.68 $\pm$ 1.311 | 98.784 $\pm$ 0.262 |
| $2^{-2}$ | 95.794 $\pm$ 0.831 | 99 $\pm$ 0.343 | 99.01 $\pm$ 0.168 |
| $2^{-1}$ | 98.557 $\pm$ 0.33 | 99.464 $\pm$ 0 | 99.278 $\pm$ 0.062 |
| $2^{0}$ | 99.505 | 99.629 | 99.505 |

## IV. Results

### A. It's all about the features

We hypothesize that such a data augmentation strategy primarily affects the feature extractors or the convolutional layers of the CNN. The classifier can then be trained and will perform well even with small data-sets. Therefore, we present all our results with respect to the extracted features. To this end, we first learn all network parameters using the specified data-set. Then, we retrain the classifier layers (after and including the first FC layer) from scratch using the corresponding sub-sampled, non-augmented data only. This is repeated for all 6 values of $R$. All the models have the same architecture as described in section II-A and use the same $\mathcal{D}_{test}$. The difference among them is the $\mathcal{D}_{train}$ and $\mathcal{D}_{val}$ used as described in section II-A. For consistency of results, we repeat the process described in section II-A to get four different $\mathcal{D}_{train}$ and $\mathcal{D}_{val}$ for each $R$ (except for $R = 2^{-1}$ and $R = 2^{0}$ where only two and one such unique data-sets were possible, respectively) and report average performance results in table I.

The column referred to as the "Full" learned features from full SOC training data-set. It shows the importance of extracting good quality features, i.e., if we had access to all the poses, we would learn the best features. Having good features makes classification quite easy, which is evident from the high test accuracies even in low data availability. The sub-sampling has little effect on the generalization performance in this case. The "Existing" column learned features from $\mathcal{D}_{train}$ and shows considerably lower test performance, especially in low data availability. This test performance is significantly improved in the "Proposed" column which learned features from $T(\mathcal{D}_{train})$. This shows the effectiveness of our strategy in improving the quality of the features extracted by the CNN. However, there exists a considerable gap between the Full and the Proposed columns in the low data regime. So, there still exists scope for additional improvement in feature extraction. We further did a decoupled investigation on the two augmentation strategies introduced, for observing their respective contributions to the overall results. For this, we carried out two experiments at $R = 2^{-3}$. In the first, we did only sub-pixel level shifts as augmentation in addition to the existing data (as described in section II-A) to see how knowledge about scintillations, affects the performance. We got 92.794 $\pm$ 1.779 test accuracy (%) which is more than 3% improvement over the performance of existing data. Similarly, for the second experiment, we did only pose-based augmentation to the existing data and got 95.814 $\pm$ 0.337 test accuracy (%) which is fairly close to the final performance we get using both augmentations together. These results provide two useful conclusions. First, the gains from these two augmentations are not additive in nature. Second, the majority of gains come from the synthesized pose information.

## V. Conclusion

In this paper, we presented a data-augmentation strategy for training the ANN architectures to solve the ATR problem with limited labeled data. We empirically verified the effectiveness of the augmentation strategy by training an ANN with the augmented data-set synthesized using the phase-history approximation method proposed in [38], [47]. We verified that the proposed model based augmentation strategy gave a significant improvement in the generalization performance of the model compared to the baseline performance over a wide range of sub-sampling ratios. However, the presented augmentation strategy only creates valid images around a small neighborhood of the given azimuth angle. We hypothesize that a global model for each class can be constructed jointly from all images belonging to that class, which in turn can produce a diverse set of SAR images over larger variations of poses. Since the MSTAR target image chips are not correctly registered and aligned across different azimuth angles, such a modeling effort has to incorporate an unknown correction phase ramp term for each image chip to generate a unified model in the phase-history domain successfully. As part of future research, we propose to develop a network architecture to learn a unified model directly in the complex domain that can account for these phase errors and synthesize a larger data-set to improve the performance of the classifier further.

## Acknowledgements

This research was partially supported by NSF Grant CNS-1823070.

## References

[1] L. C. Potter and R. L. Moses, "Attributed scattering centers for SAR ATR," *IEEE Transactions on Image Processing*, vol. 6, no. 1, pp. 79–91, Jan 1997.

[2] M. Çetin, I. Stojanović, N. O. Önhon, K. Varshney, S. Samadi, W. C. Karl, and A. S. Willsky, "Sparsity-driven synthetic aperture radar imaging: Reconstruction, autofocusing, moving targets, and compressed sensing," *IEEE Signal Processing Magazine*, vol. 31, no. 4, pp. 27–40, July 2014.

[3] L. C. Potter, E. Ertin, J. T. Parker, and M. Cetin, "Sparsity and compressed sensing in radar imaging," *Proceedings of the IEEE*, vol. 98, no. 6, pp. 1006–1020, June 2010.

[4] K. Dungan and L. Potter, "Classifying transformation-variant attributed point patterns," *Pattern Recognition*, vol. 43, no. 11, pp. 3805 – 3816, 2010. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S003132031000261X

[5] K. E. Dungan and L. C. Potter, "Classifying vehicles in wide-angle radar using pyramid match hashing," *IEEE Journal of Selected Topics in Signal Processing*, vol. 5, no. 3, pp. 577–591, June 2011.

[6] T. Abdelrahman and E. Ertin, "Mixture of factor analyzers models of appearance manifolds for resolved SAR targets," in *Algorithms for Synthetic Aperture Radar Imagery XXII*, vol. 9475. International Society for Optics and Photonics, 2015, p. 94750G.

[7] C. Hegde, A. C. Sankaranarayanan, W. Yin, and R. G. Baraniuk, "NuMax: A convex approach for learning near-isometric linear embeddings," *IEEE Transactions on Signal Processing*, vol. 63, no. 22, pp. 6109–6121, Nov 2015.

[8] D. Teng and E. Ertin, "WALD-kernel: A method for learning sequential detectors," in *2016 IEEE Statistical Signal Processing Workshop (SSP)*, June 2016, pp. 1–5.

[9] J. Cui, J. Gudnason, and M. Brookes, "Hidden markov models for multi-perspective radar target recognition," in *2008 IEEE Radar Conference*, May 2008, pp. 1–5.

[10] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, May 2017. [Online]. Available: http://doi.acm.org/10.1145/3065386

[11] S. Chen, H. Wang, F. Xu, and Y. Jin, "Target classification using the deep convolutional networks for SAR images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 8, pp. 4806–4817, Aug 2016.

[12] Y. Zhong and G. Ettinger, "Enlightening deep neural networks with knowledge of confounding factors," in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 1077–1086.

[13] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, "ImageNet Large Scale Visual Recognition Challenge," *International Journal of Computer Vision (IJCV)*, vol. 115, no. 3, pp. 211–252, 2015.

[14] M. Oquab, L. Bottou, I. Laptev, and J. Sivic, "Learning and transferring mid-level image representations using convolutional neural networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014, pp. 1717–1724.

[15] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014.

[16] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," *The Journal of Machine Learning Research*, vol. 15, no. 1, pp. 1929–1958, 2014.

[17] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *arXiv preprint arXiv:1502.03167*, 2015.

[18] B. Neyshabur, Z. Li, S. Bhojanapalli, Y. LeCun, and N. Srebro, "The role of over-parametrization in generalization of neural networks," in *7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, May 6-9, 2019*, 2019. [Online]. Available: https://openreview.net/forum?id=BygfghAcYX

[19] N. A. Ruhi and B. Hassibi, "Stochastic gradient/mirror descent: Minimax optimality and implicit regularization," in *7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, May 6-9, 2019*, 2019. [Online]. Available: https://openreview.net/forum?id=HJf9ZhC9FX

[20] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Transactions on knowledge and data engineering*, vol. 22, no. 10, pp. 1345–1359, 2009.

[21] Z. Huang, Z. Pan, and B. Lei, "Transfer learning with deep convolutional neural network for sar target classification with limited labeled data," *Remote Sensing*, vol. 9, no. 9, p. 907, 2017.

[22] S. Chen, H. Wang, F. Xu, and Y.-Q. Jin, "Target classification using the deep convolutional networks for sar images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 8, pp. 4806–4817, 2016.

[23] Z. Lin, K. Ji, M. Kang, X. Leng, and H. Zou, "Deep convolutional highway unit network for SAR target classification with limited labeled training data," *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 7, pp. 1091–1095, 2017.

[24] Z. Fu, F. Zhang, Q. Yin, R. Li, W. Hu, and W. Li, "Small sample learning optimization for resnet based sar target recognition," in *IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium*. IEEE, 2018, pp. 2330–2333.

[25] G. Dong, N. Wang, and G. Kuang, "Sparse representation of monogenic signal: With application to target recognition in sar images," *IEEE Signal Processing Letters*, vol. 21, no. 8, pp. 952–956, 2014.

[26] S. Song, B. Xu, and J. Yang, "SAR target recognition via supervised discriminative dictionary learning and sparse representation of the SAR-HOG feature," *Remote Sensing*, vol. 8, no. 8, p. 683, 2016.

[27] H. Song, K. Ji, Y. Zhang, X. Xing, and H. Zou, "Sparse representation-based SAR image target classification on the 10-class mstar data set," *Applied Sciences*, vol. 6, no. 1, p. 26, 2016.

[28] Y. Huang, G. Liao, Z. Zhang, Y. Xiang, J. Li, and A. Nehorai, "SAR automatic target recognition using joint low-rank and sparse multiview denoising," *IEEE Geoscience and Remote Sensing Letters*, vol. 15, no. 10, pp. 1570–1574, 2018.

[29] Q. Yu, H. Hu, X. Geng, Y. Jiang, and J. An, "High-performance SAR automatic target recognition under limited data condition based on a deep feature fusion network," *IEEE Access*, vol. 7, pp. 165 646–165 658, 2019.

[30] C. Zhang, S. Bengio, M. Hardt, B. Recht, and O. Vinyals, "Understanding deep learning requires rethinking generalization," in *5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings*, 2017. [Online]. Available: https://openreview.net/forum?id=Sy8gdB9xx

[31] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016, http://www.deeplearningbook.org.

[32] J. Ding, B. Chen, H. Liu, and M. Huang, "Convolutional neural network with data augmentation for SAR target recognition," *IEEE Geoscience and remote sensing letters*, vol. 13, no. 3, pp. 364–368, 2016.

[33] D. Marmanis, W. Yao, F. Adam, M. Datcu, P. Reinartz, K. Schindler, J. D. Wegner, and U. Stilla, "Artificial generation of big data for improving image classification: A generative adversarial network approach on SAR data," *arXiv preprint arXiv:1711.02010*, 2017.

[34] M. Cha, A. Majumdar, H. Kung, and J. Barber, "Improving SAR automatic target recognition using simulated images under deep residual refinements," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2018, pp. 2606–2610.

[35] F. Gao, Y. Yang, J. Wang, J. Sun, E. Yang, and H. Zhou, "A deep convolutional generative adversarial networks (dcgans)-based semi-supervised method for object recognition in synthetic aperture radar (sar) images," *Remote Sensing*, vol. 10, no. 6, p. 846, 2018.

[36] M. B. Alver, S. Atito, and M. Çetin, "SAR ATR in the phase history domain using deep convolutional neural networks," in *Image and Signal Processing for Remote Sensing XXIV*, vol. 10789. International Society for Optics and Photonics, 2018, p. 1078913.

[37] J. Wang, P. Virtue, and S. X. Yu, "Joint embedding and classification for SAR target recognition," *arXiv preprint arXiv:1712.01511*, 2017.

[38] N. Sugavanam and E. Ertin, "Limited persistence models for SAR automatic target recognition," in *Algorithms for Synthetic Aperture Radar Imagery XXIV*, vol. 10201. International Society for Optics and Photonics, 2017, p. 102010M.

[39] N. Sugavanam, E. Ertin, and R. Burkholder, "Approximating Bistatic SAR target signatures with sparse limited persistence scattering models," in *International Conference on Radar, Brisbane*, 2018.

[40] N. Sugavanam, E. Ertin, and R. Burkholder, "Compressing bistatic SAR target signatures with sparse-limited persistence scattering models," *IET Radar, Sonar Navigation*, vol. 13, no. 9, pp. 1411–1420, 2019.

[41] A. E. Brito, S. H. Chan, and S. D. Cabrera, "SAR image formation using 2D reweighted minimum norm extrapolation," in *Algorithms for Synthetic Aperture Radar Imagery VI*, vol. 3721. International Society for Optics and Photonics, 1999, pp. 78–91.

[42] M. Cetin, "Feature-enhanced synthetic aperture radar imaging," Ph.D. dissertation, Boston University, 2001.

[43] J. A. Jackson, B. D. Rigling, and R. L. Moses, "Canonical scattering feature models for 3D and bistatic SAR," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 46, no. 2, pp. 525–541, April 2010.

[44] K. Sarabandi and T.-C. Chiu, "Optimum corner reflectors for calibration of imaging radars," *IEEE Transactions on Antennas and Propagation*, vol. 44, no. 10, pp. 1348–1361, Oct 1996.

[45] L. C. Potter, D.-M. Chiang, R. Carriere, and M. J. Gerry, "A GTD-based parametric model for radar scattering," *IEEE Transactions on Antennas and Propagation*, vol. 43, no. 10, pp. 1058–1067, Oct 1995.

[46] H. Rauhut and R. Ward, "Interpolation via weighted L1 minimization," *Applied and Computational Harmonic Analysis*, vol. 40, no. 2, pp. 321 – 351, 2016.

[47] N. Sugavanam and E. Ertin, "Interrupted SAR imaging with limited persistence scattering models," in *2017 IEEE Radar Conference (RadarConf)*, May 2017, pp. 1770–1775.