

A Novel Technique Based on Deep Learning and a Synthetic Target Database for Classification of Urban Areas in PolSAR Data

Shaunak De^{ID}, *Student Member, IEEE*, Lorenzo Bruzzone^{ID}, *Fellow, IEEE*,
Avik Bhattacharya^{ID}, *Senior Member, IEEE*, Francesca Bovolo^{ID}, *Senior Member, IEEE*,
and Subhasis Chaudhuri, *Fellow, IEEE*

Abstract—The classification of urban areas in polarimetric synthetic aperture radar (PolSAR) data is a challenging task. Moreover, urban structures oriented away from the radar line of sight pose an additional complexity in the classification process. The characterization of such areas is important for disaster relief and urban sprawl monitoring applications. In this paper, a novel technique based on deep learning is proposed, which leverages a synthetic target database for data augmentation. The PolSAR dataset is rotated by uniform steps and collated to form a reference database. A stacked autoencoder network is used to transform the information in the augmented dataset into a compact representation. This significantly improves the generalization capabilities of the network. Finally, the classification is performed by a multilayer perceptron network. The modular architecture allows for easy optimization of the hyperparameters. The synthetic target database is created and the classification performance is evaluated on an L-band airborne UAVSAR dataset and L-band space-borne ALOS-2 dataset acquired over San Francisco, USA. The proposed technique shows an overall accuracy of 91.3%. An improvement over state-of-the-art techniques is achieved, especially in urban areas rotated away from the radar line of sight.

Index Terms—Autoencoder (AE), classification, deep learning, deep neural networks, polarimetric synthetic aperture radar (PolSAR), representation learning, urban remote sensing.

I. INTRODUCTION

REMOTELY sensed images acquired by space- or airborne sensors have revolutionized earth observation allowing an

unprecedented ability to image and observe large areas with high temporal repetitivity.

The active properties of synthetic aperture radar (SAR) data are increasingly exploited in monitoring applications. SAR sensors form an image by transmitting radar pulses either in horizontal or vertical polarization and receiving coherently [1]. These sensors are not dependent on external sources for illumination. They have cloud cover penetrating capabilities, thus, operate without restriction of time and season. Polarimetric SAR (PolSAR) imaging is an advancement made on the single-polarization SAR imaging [2]. The polarization of the incident wave is controlled so that it is restricted alternatively to either h or v transmission. Similarly, reception is performed alternatively in h or v polarizations, so that for every imaged pixel, four combinations of polarization (i.e., hh , hv , vh , and vv) are captured. PolSAR imaging can capture more information about the physical properties of the target than single- or dual-polarization imaging, and consequently, it can lead to better target characterization and classification [3]. However, it introduces unique challenges in the characterization of urban areas, which is increasingly becoming important for systematic monitoring.

The first attempt at the classification of PolSAR images was made in [4]. The authors proposed a maximum-likelihood based classification method for a PolSAR image based on a complex Gaussian distribution and showed that the addition of phase information from multiple channels improves the classification accuracy. Subsequently, a neural network classifier for SAR data was proposed in [5], where a feedforward network with a single hidden layer having sigmoidal activation functions was showed. It was observed that polarimetric and machine learning theory were complementary. However, at that time neural networks were considered computationally intensive and would not be widely adopted until subsequent advancements in computing technology [6] were made.

A popular classification scheme was proposed in [7], which uses a maximum-likelihood classifier based on the Wishart distribution to perform terrain classification. Various studies have explored the effect of speckle and sensing frequency on the accuracy of this technique [8]. An unsupervised scheme based on the Wishart distribution was proposed in [9]. This has been extended by using several techniques for urban classification. In [10], the Chernoff and in [11] the Jeffries–Matsushita distance

Manuscript received April 28, 2017; revised August 10, 2017; accepted September 11, 2017. Date of publication October 15, 2017; date of current version January 12, 2018. This work was supported by Italian Ministries MAECI/MIUR under the India-Trento Program for Advanced Research (ITPAR), Phase III. (Corresponding author: Shaunak De.)

S. De and A. Bhattacharya are with the Centre of Studies in Resources Engineering, Indian Institute of Technology Bombay, Mumbai 400076, India (e-mail: shaunakde@iitb.ac.in; avikbhat@iitb.ac.in).

L. Bruzzone is with the Department of Information Engineering and Computer Science, University of Trento, Trento 38050, Italy (e-mail: lorenzo.bruzzone@unitn.it).

F. Bovolo is with the Center for Information and Communication Technology, Fondazione Bruno Kessler, Trento 38123, Italy (e-mail: bovolo@fbk.eu).

S. Chaudhuri is with the Department of Electrical Engineering, Indian Institute of Technology Bombay, Mumbai 400076, India (e-mail: sc@ee.iitb.ac.in).

This paper has supplementary downloadable multimedia material available at <http://ieeexplore.ieee.org> provided by the authors. The document contains demonstration of the proposed method on additional data-sets. The total size of the file is 1.3 MB.

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/JSTARS.2017.2752282

measures are used in conjunction with the complex Wishart distribution. The complex Wishart distribution model for the covariance or coherency matrices is valid for homogeneous areas. However, in extremely heterogeneous regions, like urban areas, the Wishart assumption is seldom true. Hence, in this case, it is desirable to use a nonparametric classification scheme.

Polarimetric scattering power decomposition parameters can serve as low-level features in classification [12]. Both supervised [13] and unsupervised [9] schemes have been proposed, which perform urban area segmentation based on decomposition parameters. Despite the availability of several sophisticated methods, the selection of an appropriate decomposition approach for a scene is not a trivial task [14]. An alternative strategy for urban classification is to use the textural information in the SAR images [15], [16]. However, in texture-based schemes, the rich scattering information available from polarimetry is often not considered.

Interest in neural networks was renewed with the improvement in technology. In [17], a Karhunen–Loeve transform is used to extract features from a fully polarimetric dataset and a neural network is used to perform classification of the scene into urban, vegetation, and terrain classes. A study about the landcover and soil-type mapping capabilities using a probabilistic neural network was reported in [18] using L-band space-borne data. Multitemporal/multifrequency datasets were exploited using a neural-statistical kernel in [19] for urban classification. An adaptive resonance theory neural network was shown to perform the computationally efficient classification of urban areas using polarimetric data. Hybrid approaches, such as the combination of neural networks and fuzzy logic, were demonstrated to have good performance for SAR classification in [20] and later specifically for the identification of urban areas in [21]. A system for the classification of multitemporal SAR images based on the integration of the SAR system physics and a radial basis function neural network architecture is proposed in [22]. Another multitemporal approach using single-channel SAR datasets to monitor urban landcover in major Italian cities is proposed in [23].

Hinton *et al.* introduced the concept of deep learning in [24], which found successful application in multiple areas [25]–[28] establishing it as a new field of machine learning research. The motivation lies in the capability of neural networks to break down complex learning problems into simpler representations. Deep learning has been recently adopted by the remote sensing community for the improved performance in several applications [29]–[31]. Multiple natural land cover classes were identified from a hyperspectral dataset in [32], whereas objects detection was performed in optical datasets in [29] using deep-learning techniques. Additional information is often added to the classification scheme from auxiliary data sources like optical and multispectral areas [33], [34]. Deep learning techniques are quickly gaining popularity to perform multisensor data fusion as well [35], [36]. A deep-learning-based approach can simplify feature and classifier design, and allow the utilization of augmented datasets. This can help mitigate some of the challenges faced in the classification of urban areas from PolSAR data.

The classification of urban areas in PolSAR datasets is especially challenging, because of the problems related to orientation of targets with respect to the radar line of sight. When targets that exhibit even bounce scattering are oriented away from the radar line of sight, the co-polarized return is reflected away from the sensor, while the cross-polarized return level remains the same [37], [38]. This makes such areas hard to distinguish from natural targets like forest and vegetation. This represents a major challenge for the classification algorithm. In [39], a polarimetric matrix rotation theory framework was proposed, which develops a physically significant set of rotation angle parameters. An approach based on the rotation of the coherency matrix was proposed in [40] to compensate for the overestimation of volume in rotated urban areas. In [41], a link is established between eigen-decomposition method and model-based decomposition methods to solve the overestimation of the volume scattering power problem. Another decomposition technique proposed in [42] uses a mapping to create an improved representation and further helps improve target characterization.

Alternatively, some approaches utilize a separate class for identifying rotated targets, and merge them with the nonrotated class as a postclassification step. However, this makes it mandatory to input training samples from rotated target areas, which may not always be possible. The generalization ability and autonomy of such an approach is therefore low.

In this paper, a novel technique based on the target scattering properties along with a deep learning approach is proposed for urban area classification using PolSAR data. The properties of polarimetric scattering are used to generate synthetic targets, which simulate the effects of rotation about the radar line of sight in uniform angular steps. In general, polarization orientation angle (POA) shifts are observed in polarimetric SAR data, which are induced by terrain slopes and complex scattering in urban areas. In turn, POA shifts can be induced by appropriate Euler rotations to form synthetic targets. They can be suitably leveraged for better generalization in machine learning algorithms. The process is used to generate a database containing simulated urban targets. The synthetic urban target database is used to train and extract features from a sparse stacked autoencoder (AE). Thus, the information about the targets and rotation is accounted for in the classification scheme, which helps to discriminate rotated urban areas without requiring explicit representation in the training stage. The advantage of using a deep-learning-based scheme is that the network can be trained incrementally on a large volume of data, and subsequently used as a generalization for unseen data. This leads to a computationally efficient and fast classification phase. Given the commercial interests in deep learning at large scale, several GPU-based specialized peripherals have been developed. These can be used to further speed up calculations and drastically reduce the training time over CPU-bound computation.

The urban synthetic target database generation technique applied to simulate the effects of rotation about the radar line of sight (see Section II-A). This forms the midlevel features that act as an input to a stacked AE stage. The training of this stage is carefully monitored using metrics derived from information theory (*viz.* cross-entropy error) to prevent overfitting. After

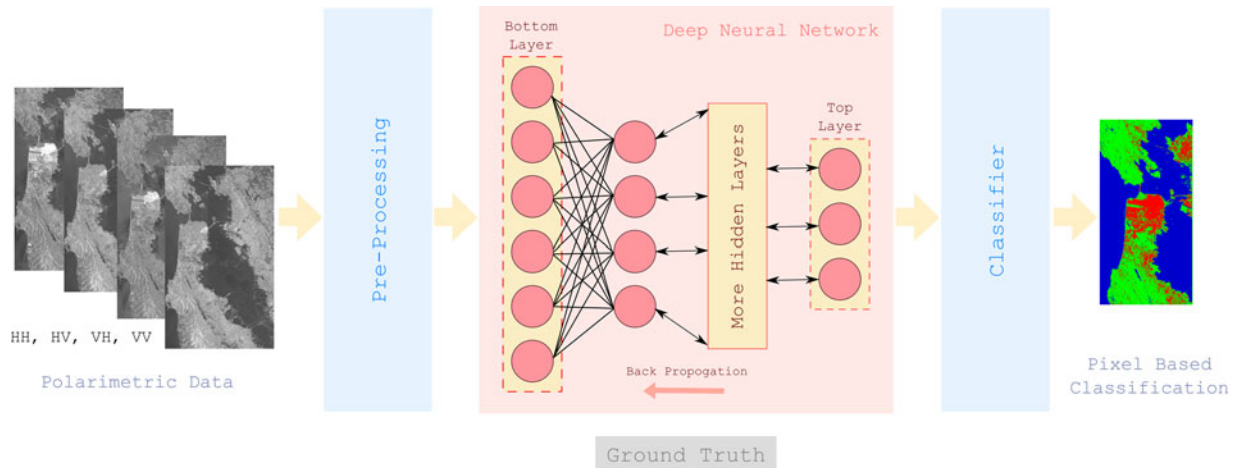


Fig. 1. Framework for the classification of PolSAR images using a deep-learning architecture.

the training phase, the extracted features are described and discussed in Section II-B. Finally, a multilayer perceptron (MLP) is used to perform the final supervised classification using both the extracted features and estimated statistical parameters (see Section II-C). The overall stagewise procedure allows an efficient tuning of hyperparameters for optimizing the classification. The technique is demonstrated on PolSAR data acquired over the city of San Francisco by the UAVSAR and ALOS-2 platforms (see Section III). It is shown that both rotated and nonrotated urban areas can be discriminated with an overall accuracy of about 91.3%. Moreover, the proposed approach shows a better qualitative and quantitative performance when compared with contemporary techniques.

II. PROPOSED CLASSIFICATION APPROACH

The proposed approach for pixel classification of PolSAR data is shown in Fig. 1. The polarimetric channels can be considered as low-level input features that describe the geometry and dielectric properties of the target. However, these features are not robust enough to obtain accurate and stable classification performance. A preprocessing step is applied to transform them into more robust midlevel features, but this expands the data volume manifold. A deep learning network is used to extract high-level features that are both efficient and robust with respect to the midlevel ones. In the literature, both supervised structures, such as CNNs, [43] and unsupervised structures, such as AEs [32] and DBNs [44], are used to this end. The mid- or low-level features are given as an input to the network at the bottom layer and features are extracted from either the top or intermediate layers. The features are then classified using a suitable supervised classifier. In general, there are two kinds of classifiers: Hard classifiers, such as SVMs, that directly produce the classification label as an output at every pixel, and soft classifiers, such as logistic regression, that can fine-tune the whole pretrained network and predict the class labels as a probability. The proposed method aims at classifying PolSAR images to identify urban areas and separate them from their surrounding landcover classes, forest, bare soil, and water bodies. Segmented

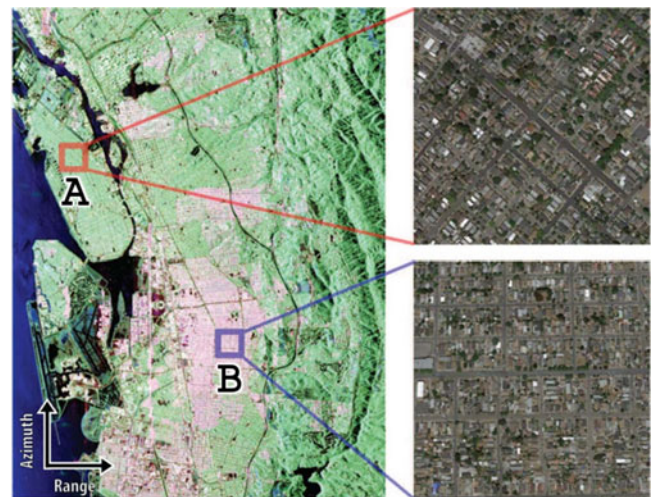


Fig. 2. Pauli composite generated from UAVSAR polarimetric data acquired over Oakland, CA, is shown along with optical data from Google Earth. Area A is oriented with the radar line of sight while area B is not.

urban area maps can subsequently be used to characterize urban density, land area usage, urban sprawl, and landcover change in temporal data amongst other applications.

The difficulty in the classification of urban areas from PolSAR data arises because of the complex scattering of the polarized radar pulse from urban structures [45]. This is especially pronounced when they are oriented away from the radar line of sight. Range and azimuth slopes can cause the polarization to rotate about the radar line of sight, which causes the erroneous identification of the scattering mechanism. For instance, scattering observed from urban structures oriented away from the radar line of sight exhibit high cross polarization, as shown in Fig. 2. This presents a confusion between urban areas and vegetation leading to incorrect classification. To overcome this challenge, a novel synthetic rotation based approach is proposed, which simulates other rotated synthetic targets based on those in the scene. This allows the machine learning algorithm to better generalize, be more robust and successful in discriminating oriented urban areas.

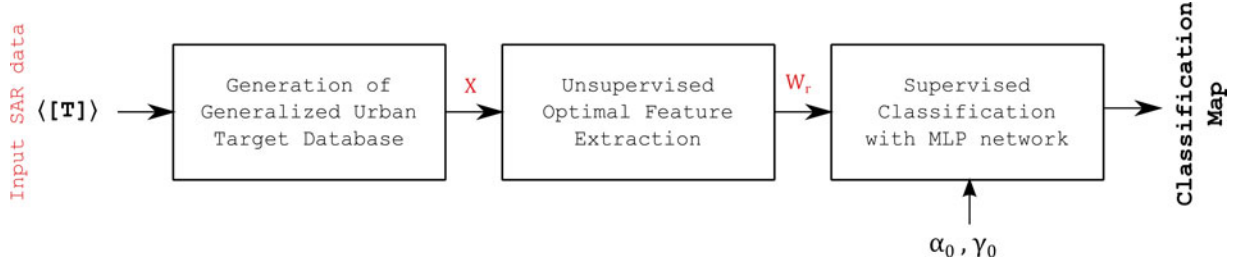


Fig. 3. Block diagram of the proposed approach.

The classification process consists of three stages, as shown in Fig. 3. The first stage deals with the preprocessing of the PolSAR data and the generation of a generalized synthetic urban target database by transforming the input data to obtain a more general representation of the target (see Section II-A). The second stage is an unsupervised feature learning step, which automatically learns an optimal representation of the generalized database using a stacked AE network (see Section II-B). Finally, the third stage extracts optimal features from the previous stages, along with some statistical parameters and performs a supervised classification by using an MLP network (see Section II-C).

A. Generation of Generalized Synthetic Urban Target Database

Measurement by a fully PolSAR system commonly involves the transmission of horizontally (h) and vertically (v) polarized radar pulses followed by their coherent reception. The polarization state of an incident electromagnetic wave is altered on scattering from a complex radar target. This alteration is a function of the physical and geometrical properties of the target itself, which in turn can serve to characterize it. This scattering (or Sinclair) matrix is in the (h, v) polarization basis and has the following form:

$$\mathbf{S} = \begin{bmatrix} S_{hh} & S_{hv} \\ S_{vh} & S_{vv} \end{bmatrix}. \quad (1)$$

The full polarimetric scattering matrix is available for each imaged pixel and consists of four independent measurements (hh , hv , vh , and vv), with the phase relations between them recorded [46]. Generally, in polarimetry, it is assumed that the targets are reciprocal. In the case of symmetric monostatic scattering, we can thus constrain the scattering matrix to be symmetric, i.e., $S_{hv} = S_{vh}$.

The corresponding \mathbf{k}_p -target vector is expressed as

$$\mathbf{k}_p = \frac{1}{\sqrt{2}} [S_{hh} + S_{vv} \quad S_{hh} - S_{vv} \quad 2S_{hv}]^T \quad (2)$$

where the superscript T denotes transpose. In the monostatic backscattering case, we can construct a 3×3 coherency matrix

\mathbf{T} as

$$\mathbf{T} = \mathbf{k} \cdot \mathbf{k}^{*T} \quad (3)$$

$$\mathbf{T} = \begin{bmatrix} t_1 & t_2 & t_3 \\ t_2^* & t_4 & t_5 \\ t_3^* & t_5^* & t_6 \end{bmatrix} \quad (4)$$

where

$$\begin{aligned} t_1 &= |k_1|^2 = \frac{1}{2} |S_{hh} + S_{vv}|^2 \\ t_2 &= k_1 k_2^* = \frac{1}{2} |(S_{hh} + S_{vv})(S_{hh} - S_{vv})^*|^2 \\ t_3 &= k_1 k_3^* = (S_{hh} + S_{vv}) S_{hv}^* \\ t_4 &= |k_2|^2 = \frac{1}{2} |S_{hh} - S_{vv}|^2 \\ t_5 &= k_2 k_3^* = (S_{hh} - S_{vv}) S_{hv}^* \\ t_6 &= |k_3|^2 = 2 |S_{hv}|^2. \end{aligned} \quad (5)$$

L independent and identically distributed samples are averaged to enhance the signal-to-noise ratio while forming the 3×3 L -looked \mathbf{T} as

$$\mathbf{T} = \langle [\mathbf{T}] \rangle = \frac{1}{L} \sum_{i=1}^L \mathbf{k}_{p_i} \cdot \mathbf{k}_{p_i}^{*T} \quad (6)$$

where $\langle \dots \rangle$ denotes temporal or spatial ensemble averaging, under the assumption that the signal is ergodic. Radar images, as a consequence of their coherent nature, are subject to a speckle pattern. Multilooking allows the exploitation of the second-order statistical information while providing preliminary speckle suppression. A refined Lee filter [47] is applied to the dataset as a preprocessing step to further suppress speckle and improve the classification performance. This filter is based on the local statistics of the data and has a good edge preservation ability allowing the retention of details.

The orientation of the target from the radar line of sight or the presence of azimuth slopes can induce POA shifts. These can lead to misinterpretation of the scattering characteristics of the target. This is especially common in urban areas oriented at an angle with respect to the radar line of sight and affects the classification performance. The rotation of \mathbf{T} can help us to optimize and extract more polarimetric information, which, in turn, helps us to improve the characterization of the target scattering mechanism [12]. If we consider a target that is sym-

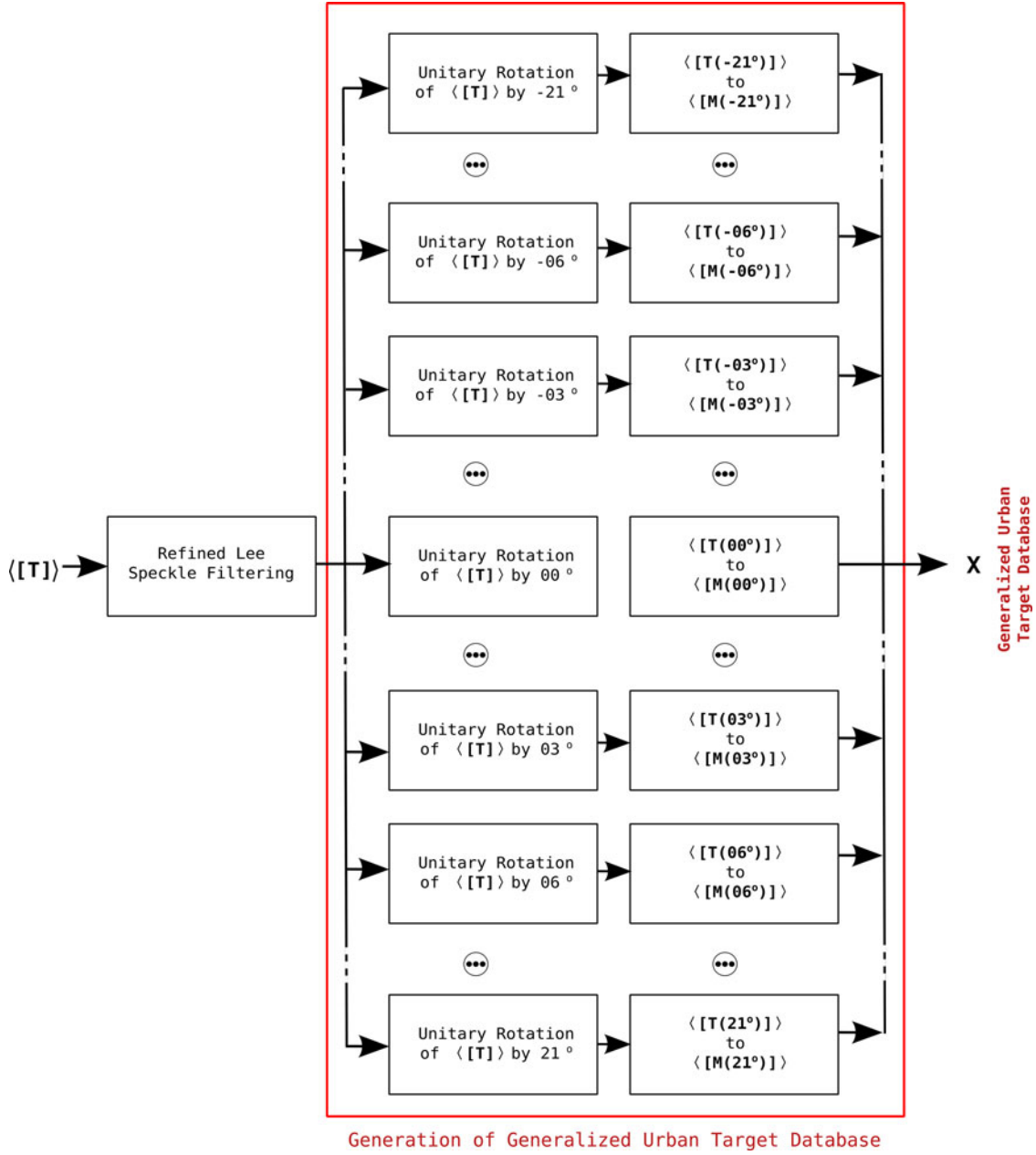


Fig. 4. Schematic diagram of the PolSAR data preprocessing and physics-based generalized urban target database generation step.

metrical about the radar line of sight, the rotation of such a target about the radar line of sight can be represented by the unitary rotation of \mathbf{T} by a matrix rotation model expressed as

$$\mathbf{T}(\theta) = \mathbf{R}(\theta)\mathbf{T}\mathbf{R}(\theta)^{-1} \quad (7)$$

where the special unitary rotation matrix $\mathbf{R}(\theta)$ is given by

$$\mathbf{R}(\theta) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos 2\theta & \sin 2\theta \\ 0 & -\sin 2\theta & \cos 2\theta \end{bmatrix}. \quad (8)$$

The orientation angle θ is estimated either by minimizing the t_6 element given in (5) [48] or by maximizing a stochastic distance measure between t_6 and t_4 [49].

Conversely, if the \mathbf{T} matrix is rotated using unitary rotation as in (7), it would simulate the effect of physically rotating the target while measuring it with a polarimetric antenna. That is, the result of rotating the target \mathbf{T} by $\mathbf{R}(\theta)$ is the \mathbf{T} matrix that would be obtained with target having orientation θ from the radar line of sight during measurement. Thus, it is expected that by synthetically rotating and storing the rotated \mathbf{T} , we can generate a database of synthetic targets based on those detected in the scene. The rotations are done in discrete steps, as shown in Fig. 4, before being collated in a database. The granularity of the rotation steps is directly proportional to the size of training dataset and the capacity of generalization. If the rotations are closely spaced, the database will require more memory, but the generalization ability will be enhanced. However, beyond

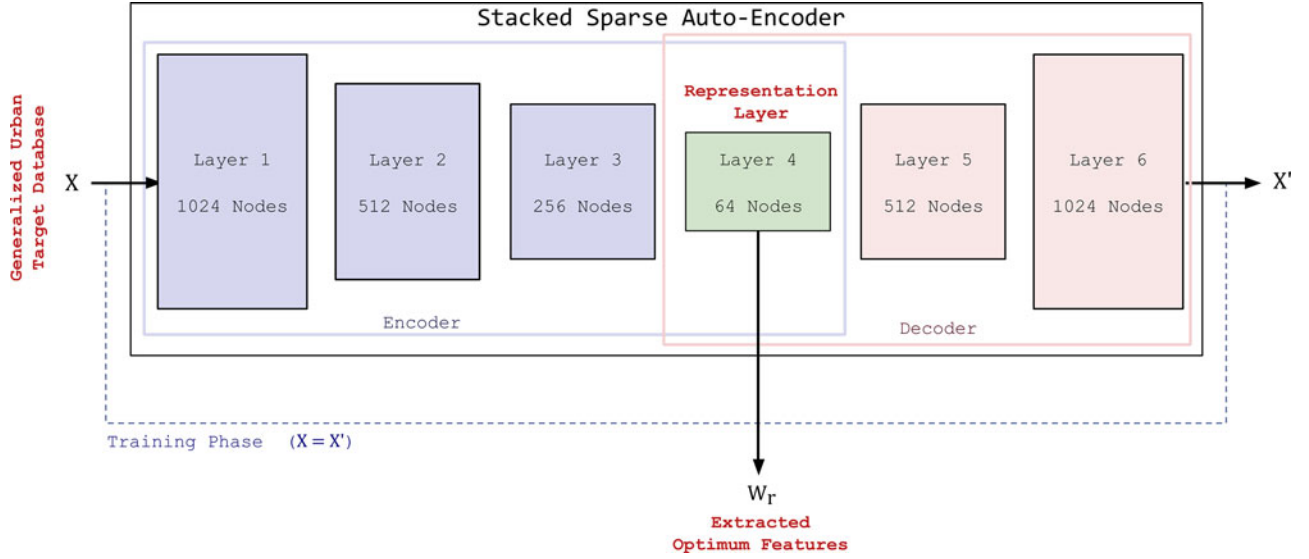


Fig. 5. AE block scheme with Layer 4 extracted as the output feature set.

a point, the gain in performance is not justifiable by the rise in computational cost. Therefore, a step of 3° is selected as a tradeoff between performance gain and memory requirement, but other choices are possible. When these are used in the training of the learning algorithm, they improve the generalization capability of the network, allowing it to recognize targets with orientation not present in the original training data. Thus, the rotation-based synthetic target database creation strategy in conjunction with a deep-learning network architecture improves the classification accuracy of urban areas.

The 3×3 coherency matrix \mathbf{T} , which is Hermitian by the nature of its construction, contains complex-valued quantities in its off-diagonal elements. In the complex-number space, the solution of differentiation is not guaranteed to be analytic. This poses a problem in the backpropagation step while using neural networks. An alternative approach would be to separate the real and imaginary parts of the elements of \mathbf{T} and use them as an input for the network. However, splitting a complex number to process in real-valued neural networks leads to a suboptimal representation of the domain of the problem [50]. To get around it, we convert the complex-valued \mathbf{T} matrix to the real-valued Mueller representation. The Mueller matrix \mathbf{M} is a 4×4 real matrix that can be obtained by a linear transformation of the \mathbf{T} matrix as

$$\mathbf{M} = \frac{1}{2} \begin{bmatrix} t_1 + t_4 + t_6 & t_2 + t_2^* & t_3 + t_3^* & -i(t_5 - t_5^*) \\ t_2 + t_2^* & t_1 + t_4 - t_6 & t_5 + t_5^* & -i(t_3 - t_3^*) \\ t_3 + t_3^* & t_5 + t_5^* & t_1 - t_4 + t_6 & i(t_2 - t_2^*) \\ -i(t_5 - t_5^*) & -i(t_3 - t_3^*) & i(t_2 - t_2^*) & -t_1 + t_4 + t_6 \end{bmatrix}. \quad (9)$$

Here, \mathbf{M} is a real-valued matrix with 16 elements, which reduces to 10 unique real elements assuming reciprocity conditions. This property makes the representation suitable for the purpose of feature learning by neural network structures. Additionally, the Mueller matrix is also closely related to the physical

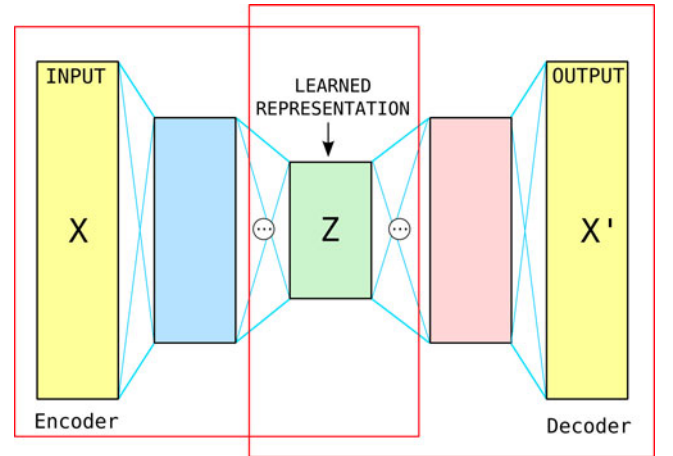


Fig. 6. Structure of a general AE with fully connected hidden layers.

properties of a target [51]. This makes it a reasonable choice for identifying the scattering mechanism from targets. Thus, each rotated \mathbf{T} matrix is converted to its \mathbf{M} representation. \mathbf{M} matrices are collated together as a database of synthetically generated generalized urban targets to serve as an input to subsequent stages of the learning algorithm.

B. Unsupervised Feature Extraction From a Stacked AE Network

The second stage of the proposed classification method consists of an unsupervised sparse stacked AE that is represented in Fig. 5. An AE is a feedforward, fully connected, nonrecurrent neural network with an input layer, an output layer, and one or more hidden layers, as shown in Fig. 6. The output layer has the same number of nodes as the input layer. During the learning phase, the AE is trained to reconstruct for each pixel i the input X_i in the output X'_i . The error is backpropagated and minimized over multiple iterations.

The collated synthetic urban target database from the previous stage is made available to the AE as an input. The training is carried out in an unsupervised manner. At the end of the training phase, the weights \mathbf{W}_r of the central hidden layer Z are extracted to form an efficient feature representation of the dataset. This stage takes the $n \times 10$ dimensional collated target dataset X , where n is the number of synthetic rotations, and reduces it to an m dimensional feature set \mathbf{W}_r , where m is the number of nodes in the representational layer Z and $m < n \times 10$. In practice, the AE stage integrates information about the rotational response of the target and encodes it into a compact representation. Thus, AE acts as an unsupervised feature learning stage and is able to automatically extract an optimized feature set from the input data. This self-learning of the features based on the training data is a hallmark of deep-learning-based architectures. This step gives the network greater generalization ability allowing it to respond to different orientation configurations of the targets. Thus, once trained on a target that is perpendicular to the radar line of sight, it can respond to the same target encountered at an orientation, even if such a configuration was not present in the training samples.

To understand the action of the AE, let us consider a learning problem where a labeled training set (X_i, l_i) is available, where X_i represents the input vector and l_i is the corresponding label for the i th pixel. The label is not required for the unsupervised AE stage, but it is necessary for the subsequent classification. The network consists of individual neurons that are interconnected such that it is possible to define a complex, nonlinear, nonparametric hypothesis $h_{W,b}(x)$. W is the weight matrix and b is the vector of bias terms for the network. Both these parameters are fitted to the data over the training process. A neuron is a computational unit that takes as an input $\mathbf{X}_i^n = X_1, X_2 \dots X_n$ and an intercept term b , and produces the output, called hypothesis, $h_{W,b}(X) = f(\mathbf{W}^T \mathbf{x}) = f(\sum_{i=1}^n W_i x_i + b)$, where $f: \mathbb{R} \Rightarrow \mathbb{R}$ is called the activation function or nonlinearity.

Traditionally, the sigmoid $\sigma(x) = 1/(1 + e^{-x})$ has been used as the activation function. It takes a real-valued number x and returns a value within $[0, 1]$, according to the magnitude of the input. In practice, this has some drawbacks. The output of the sigmoid saturates at either tail of 0 or 1 and the gradient at these regions tends to zero, which causes a very small output to be backpropagated. Additionally, if the initial weights are large, the neuron can quickly become saturated during training. The sigmoid function is not zero centered. During backpropagation, if the input data to the neuron are always positive, i.e., $x > 0$, then the gradient weights will either become all positive or all negative. This could introduce undesirable oscillation dynamics in the gradient updates for the weights. To overcome the nonzero centering problem, one may use the tanh activation function $f(x) = \tanh(x) = (e^x - e^{-x})/(e^x + e^{-x})$, which has the limits $[-1, 1]$. However, this function is subject to saturation when the input has a large dynamic range as it is common in SAR data. A nonsaturating nonlinearity like the rectified linear unit (ReLU) activation function [52] can be used to overcome the saturation problem. The ReLU $f(x) = \max(\epsilon, x)$ simply thresholds the data at ϵ , typically $\epsilon = 0$. The differentiation of the function is defined as $\frac{df}{dx} = \{1 : x > \epsilon, 0 : x < \epsilon\}$. Another advantage is

that the training time for saturating activation functions is larger with the gradient descent algorithm than for its nonsaturating counterparts [53]. Since deep learning algorithms use several layers, faster training of ReLU units (as compared to sigmoid or tanh) translates to multifold reduction in training times [54]. The computation of the output of a ReLU is simple as it does not involve the evolution of exponentials like in the sigmoid and tanh, allowing for in-place computation and requiring less computer memory. A disadvantage of ReLU units is that when subjected to large gradients they can update their weights to such a state that they cannot be activated by subsequent inputs. The neuron is said to “die” in the training. To overcome this, the ReLU units can be modified to include a “leakage” term ϵ , $f(x) = 1(x < \epsilon)(\alpha_r x) + 1(x \geq \epsilon)(x)$. This modification causes the units to have a small negative slope α_r when the input is below the threshold, i.e., $x < \epsilon$. This allows the neuron to recover even if its weight has been updated to a high value by a particular input over subsequent inputs. The differentiation can be defined as

$$\frac{df}{dx} = \begin{cases} 1 & : x > \epsilon \\ \alpha_r & : x \leq \epsilon. \end{cases}$$

The data must be given as an input into the AE in random order to prevent it from memorizing the sequence in which the data are presented. The cross-entropy error in each iteration between the output of the AE and the input data is used to monitor the training. A progression toward zero indicates that the network is properly learning, whereas a stable value indicates that the learning stage is complete. The rate of adaptation of the network is determined by the learning rate (α). α is gradually reduced as the training progress. This is done by multiplying it by a set multiplier (Γ). After a given number of iterations α is updated as $\alpha = \alpha\Gamma$. The sparsity of the network is controlled by setting ρ , which is the expected activation of a hidden unit averaged across the training samples. As $\rho \rightarrow 0$, the representation becomes increasingly sparse, controlled by the adjustment of b . The performance of the stochastic gradient descent algorithm can be improved by introducing of a momentum term (μ). Deep architectures tend to have steep slopes in the objective function near the local optima. This causes the gradient descent to oscillate and leads to a slow convergence. By introducing the μ term, these oscillations are damped.

The AE structure is now a close representation of the input data, and the internal nodes of the AE can be used as features in the classifier. The weights of the representation layer \mathbf{W}_r of the AE are used as features after completion of the training phase. They are made available as an input to the subsequent stage of the method. \mathbf{W}_r is a sparse representation incorporating possible rotations of the targets present in the scene. This gives it a better generalization ability over the original data space. A sparse representation also simplifies subsequent classifier design and improves accuracy due to improved interclass separation [55].

C. Classification Using Multilayer Perceptron (MLP) Network

The third stage of the method consists of an MLP fully connected feedforward neural network (see Fig. 7). The network

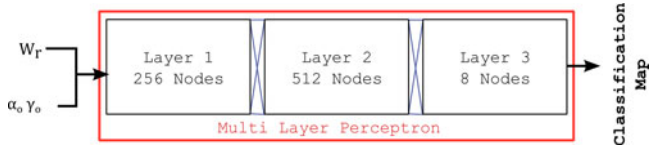


Fig. 7. Stage 3: Final classification is done using an MLP network with parameters extracted from the data.

is trained by a supervised learning algorithm that takes training data and labels as an input and attempts to classify unlabeled additional unseen data points. Here, the input to the MLP is the feature vector W_r extracted from the previous stage and two statistical parameters α_{ij}^0 and γ_{ij}^0 computed from the data. These parameters are generated by fitting the \mathcal{G}_I^0 distribution to the hh , vv , and hv intensity bands [56] over a moving window for each pixel (i, j) in the image. The three α_{ij}^{pq} and γ_{ij}^{pq} parameters are averaged to compute the final values:

$$\begin{aligned}\alpha^0 &= (\alpha_{ij}^{hh} + \alpha_{ij}^{vv} + \alpha_{ij}^{hv})/3 \\ \gamma^0 &= (\gamma_{ij}^{hh} + \gamma_{ij}^{vv} + \gamma_{ij}^{hv})/3.\end{aligned}\quad (10)$$

Here, pq represents the individual polarization bands hh , vv , and hv and parameters α_{ij}^{pq} and $\gamma_{ij}^{pq} \in \mathbb{R}$. These statistical parameters add textural context to the classification step improving the accuracy. The $\alpha_{ij}^0 < 0$ serves as a measure of the homogeneity (smoothness) of the area while γ_{ij}^0 is the scale parameter of the distribution and, thus, is related to the brightness of the area [57]. For values of α_{ij}^0 near zero, the imaged area is very heterogeneous, as in the case of urban areas. The value of α_{ij}^0 diminishes to its lowest value for homogeneous areas. The parameter γ_{ij}^0 can help further discriminate between various types of target.

The training labels are derived from ground truth information about the area. A three-layer network is used to generate the final classification. The network weights and biases are randomly initialized at each iteration using the Xavier strategy [58]. Since the inputs to this stage have smaller dynamic range, and because the goal is to classify the data into labels, we use sigmoid saturating nonlinearities. The network undergoes a training phase using the labeled samples, extracted features, and the statistical parameters. It is iterated to maximize the training accuracy on the test samples. Once the network weights are finalized, the unlabeled pixels in the dataset are classified to generate the thematic map.

III. DATASET AND EXPERIMENT DESIGN

Two datasets are used in this study. A subset (image subfragment) of the UAVSAR L-band SAR dataset has been used to study the experiment parameters and actions. The classification is carried out on a larger ALOS-2 L-band SAR image with extensive performance analysis.

A. Dataset Description: UAVSAR L-Band Dataset

The first dataset is a Single Look Complex L-Band UAVSAR dataset acquired over the San Francisco bay area, USA, on

November 13, 2012, with a resolution of 0.6 m in azimuth and 1.6 m in range. The look angle ranges from 25° to 60° with a range swath of 20 km [59]. This is an airborne sensor capable of acquiring very high-resolution images. The data are multilooked 3 times in range and 12 times in azimuth to bring them to a resolution of 7.2 m in azimuth and 5 m in range to make them comparable in spatial resolution to the space-borne ALOS-2 dataset. A 1000×1000 pixel representative portion is used for these experiments. This dataset is freely available and has similar properties to the space-borne ALOS-2 sensor, so it was used to analyze the action of the network. The subset has a prominent urban area rotated with respect to the radar line of sight, which was used to check the effectiveness of the technique.

B. Dataset Description: ALOS-2 L-Band Dataset

The second dataset is an ALOS-2 image acquired on March 24, 2015 over the city of San Francisco, USA, with a resolution of 3.2 m in range and 2.8 m in azimuth. The dataset is multilooked three times in range and two times in azimuth to generate an image of resolution 9.6 m in range and 5.6 m in azimuth. The radar operates in L-band with $\lambda = 24$ cm. The area is heavily urbanized and has high coverage with airborne and space-borne polarimetric radar creating a sizable well-documented data-pool that can be exploited for various deep-learning techniques. Consequently, this area has been featured in several studies on PolSAR [12], [60], [61].

By operating in L-band, ALOS-2 has increased penetration capability and is less susceptible to scattering due to surface roughness. As a result, return from areas of smooth vegetation, such as parks and golf courses, appears to have undergone specular reflection and, consequently, have very low return power levels. The average backscattered power levels in these areas for the copolarized channels are $\sigma_{hh}^0 \approx -16$ dB and $\sigma_{vv}^0 \approx -15$ dB. This is comparable to that of still water, which is about $\sigma_{hh}^0 \approx \sigma_{vv}^0 \approx -19$ dB, and is much lower than the power returned from forest areas, where $\sigma_{hh}^0 \approx 6$ dB and $\sigma_{vv}^0 \approx 7$ dB. Consequently, the areas having very low backscatter power tend to be classified as belonging to the water class. This phenomenon is a function of the incidence angle of the sensor and the relative roughness of the surface with respect to the sensing wavelength.

C. Experimental Setup

“Urban,” “Forest/vegetation,” and “Water” are the classes chosen in this study. These classes are usually an important input for various applications related to urban classification. Disaster management studies, urban sprawl estimation, target, and infrastructure recognition are examples of applications that benefit from accurate identification of these classes. The urban class comprises of man-made structures whose height is equal to or greater than the sensing wavelength causing the incident EM wave to undergo even number of bounces from the surface surrounding the structure, and the structure itself before returning back to the sensor. This includes structures like houses, walls, masts, etc. These targets are characterized in PolSAR data by a strong double-bounce return because of their constituent materials and geometry [62]. In the context of this

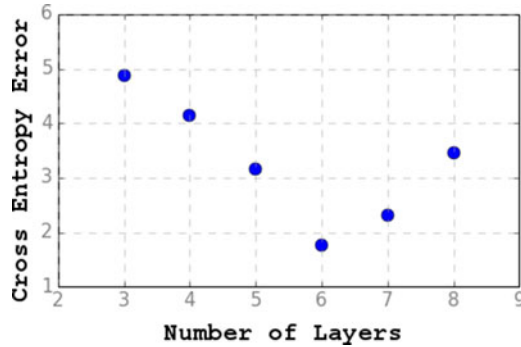


Fig. 8. Effect of layer depth on training of AE in terms of cross entropy error versus the number of layers.

study, the “Forest/Vegetation” class is used to refer to areas of natural vegetation, open spaces between urban areas with light to almost no vegetation, parks, etc. In PolSAR, this class is characterized by diffused scattering that does not change significantly with rotation about the radar line of sight. The “Water” class is made up of water bodies that are characterized by specular or surface reflection depending on the condition of the surface. At longer wavelengths, some smooth natural areas have a very low return and tend to be misclassified as water. The proportion of such pixels is, however, under 0.2% of the total as estimated from the ground truth.

As a part of the SAR preprocessing, a refined Lee filter with a window size of 3×3 is applied to the dataset after the application of appropriate multilooking for the suppression of speckle. The areas under consideration have relatively flat topography, and hence no radiometric terrain correction is applied. Appropriate compensation [63] may be applied in case of highly undulated areas. The synthetic target database is generated by rotating \mathbf{T} with a step size of 3° in the range -22° to 22° . Considering a finer sampling of 1° increases the data volume manyfold without a corresponding appreciable increase in classification accuracy.

The network used in stage 2 of the proposed method is a six-layer stacked sparse AE. To determine the optimal number of layers, the dataset was trained on different AE topologies of increasing depth, but with the same hyperparameters ($\alpha = 0.01$, $\Gamma = 0.1$, and $\mu = 0.95$). The cross-entropy error at the end of training was analyzed to compare the architectures. From the plot shown in Fig. 8, it is seen that a six-layer AE is the most suitable for this dataset. From input X to output X' , the network has 1024, 512, and 256 nodes in layers 1, 2, and 3, respectively, forming the encoder, and 512, 1024 nodes in layers 5 and 6 forming the decoder. The central 64 node in layer 4 contains the learned representation. The nonlinearity of each encoder is a ReLU with a leakage parameter $\alpha_r = 0.01$ allowing the handling of large dynamic range input without saturating or becoming nonresponsive over many iterations. The sparsity hyperparameter of the network is set to $\rho = 0.15$, which ensures that the average activation of each hidden neuron is approximately near 0. The network weights and biases are randomly initialized at the beginning of each iteration using the Xavier strategy [58]. The network is iterated 200 000 times with a batch size of 1000. This ensures that the complete dataset is passed

multiple times through the AE for better generalization. To ensure that the learning machine is not memorizing the order in which the data are being fed, this order is randomized. The base learning rate is set to $\alpha = 0.001$ and the update step multiplier is set to $\Gamma = 0.1$. The step size is chosen to be 40 000 to allow for sufficient number of epochs before the learning rate is updated. The gradient descent momentum is set to $\mu = 0.85$. The value of batch size is adjusted depending on the size of the image.

In stage 3, a three-layer MLP network consisting of 256 hidden nodes in the first layer, 512 in the second layer, and 8 in the last layer is used for the final classification. When common PolSAR visualization techniques (i.e., Pauli RGB) and decompositions are applied, vegetated and rotated urban areas appear to be visually similar. This makes it more difficult to successfully discriminate and delineate rotated urban areas in a scene as compared to areas that are perpendicular to the radar line of sight. Hence, no training areas are considered from areas that do not show dihedral (red) scattering in the Pauli RGB image.

The weights from the representational layer of the previous stage are given as an input to the MLP along with two statistical texture parameters, α_0 and γ_0 (see Fig. 9). A window of 9×9 has been used for the extraction of the parameters. The size of the window is determined by the resolution of the image and dimensions of the targets on the ground. These add more texture information to the classification scheme helping discern urban areas better. For this network, $\alpha = 0.01$ (not to be confused with the statistical parameter $\alpha_{i,j}^0$), $\Gamma = 0.1$, and $\mu = 0.95$. About $\sim 5\%$ of the total pixels in the dataset is used for training the MLP. The MLP network is iterated a maximum of 100 000 times.

Five experiments are carried out on UAVSAR and ALOS-2 datasets. First, the impact of textural parameters is examined. The AE is trained and the weights of the representational layer are extracted. The MLP classification is performed both with and without the introduction of textural parameters. The difference in the classification performance is evaluated in Section IV-A. Second, the impact of the synthetic target generation is studied. The AE is trained on \mathbf{T} as usual. After the completion of training, the weight update is stopped and a \mathbf{T} matrix rotated synthetically by 5° and 9° is given as an input. The activation pattern of the neural network in the case of permuted data is compared with the pattern when no rotation is applied (see Section IV-B). Third, the classification and performance analysis is conducted on the ALOS-2 dataset (see Section IV-C). Accuracy assessment is performed both quantitatively and qualitatively. Fourth, the response of one node of the representational layer is extracted from the AE and plotted to visualize the learned representation (see Section IV-D). Fifth, the results of the proposed method are compared with other classification techniques (see Section IV-E).

IV. RESULTS

The impact of textural parameters and synthetic rotation for target generation on the performance of the network is examined on a subset of the UAVSAR dataset. The analysis of the classification performance using the proposed method was conducted on the full ALOS-2 scene along with the analysis

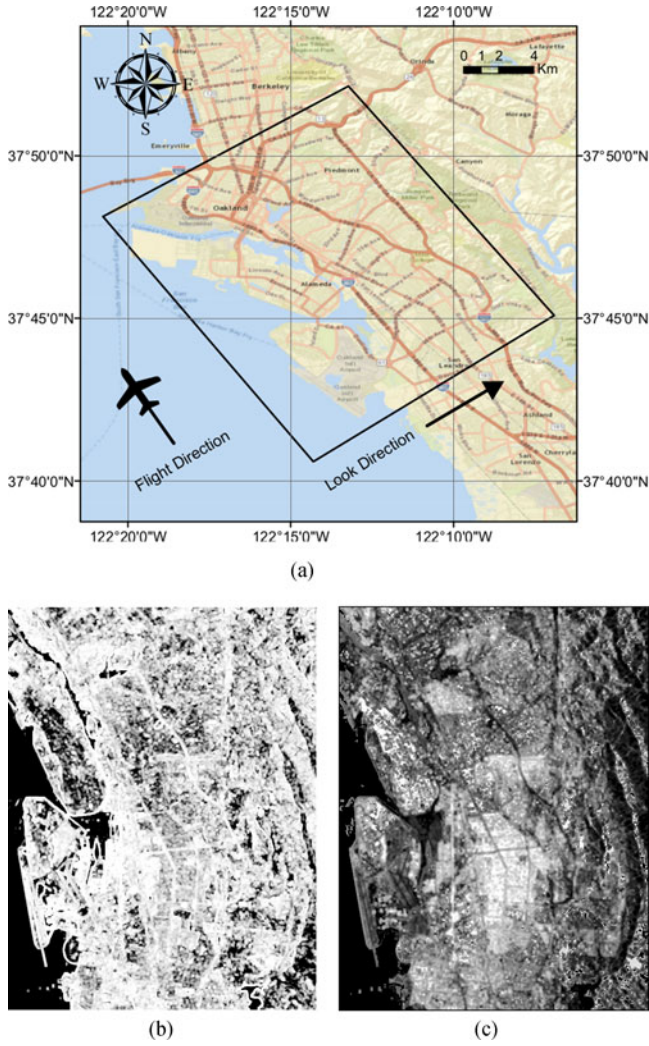


Fig. 9. (a) A map with the dataset footprint shown with the (b) α^0 and (c) γ^0 images extracted from a UAVSAR L-band dataset collected over Oakland.

of the learned representation. Finally, the performance of the proposed technique was compared with state-of-the-art classifiers for PolSAR data.

A. Impact of Textural Parameters

One of the advantages of the deep-learning-based approach is that useful parameters can be included in the later stages of the process. This is in contrast to approaches, where the information must be included at the beginning, causing an increase in dimensionality and classification difficulty, making a feature selection step necessary [64], [65]. Although the inclusion of the textural parameter only slightly improves the overall classification accuracy, its inclusion can help discern low-density-oriented urban targets. In Fig. 10, the classification map of the UAVSAR dataset is presented, both with and without the inclusion of the textural parameters α_0 and γ_0 in the MLP stage. The overall improvement in the classification accuracy due to the inclusion of textural parameters is small ($\sim 3\%$). However, for areas of low density suburban housing interleaved with vegetation and

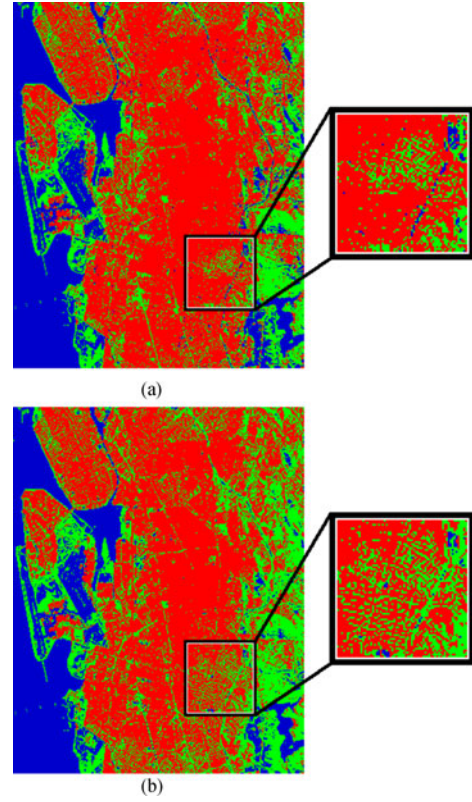


Fig. 10. Classified map derived from an UAVSAR full polarimetric L-band dataset (a) with and (b) without the use of α^0 and γ^0 as textural features.

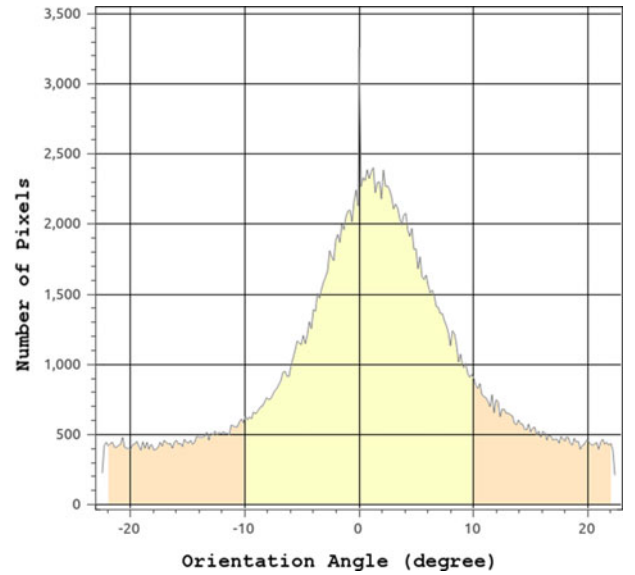


Fig. 11. Distribution of the estimated orientation angles of the region imaged in the UAVSAR L-band dataset acquired over Oakland, USA.

streets, an improvement is observed (black box in Fig. 10) because of benefit from the neighborhood information.

B. Impact of Synthetic Rotation

A histogram of measured orientation angles in the UAVSAR dataset is shown in Fig. 11. It can be seen that most of the targets

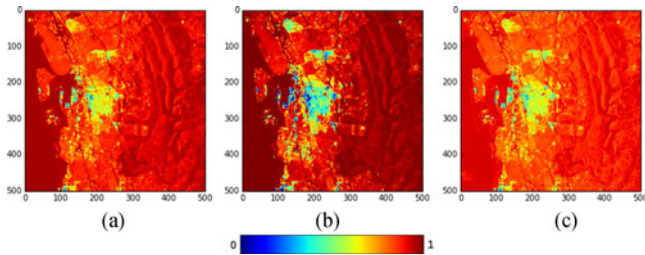


Fig. 12. A section of the UAVSAR image is trained using the AE network and the extracted features are plotted when input corresponds to a synthetic rotation of (a) 0° , (b) 5° , and (c) 9° , respectively. The actual value of the output of the ReLU node has been scaled from 0 to 1 for representation as an image.

are oriented between -10° and 10° . The average orientation angle of this subset is approximately 5° . The orientation angle is computed by the estimation method given in [48]. To visualize the action of synthetic rotation, T is given as an input to the trained AE stage after rotation by 0° , 5° , and 9° separately. As an example, the 11th element of the feature vector is extracted, after completion of the training stage of the AE and is presented in Fig. 12. In the case of the pixels in which the synthetic rotation angle matches the actual target orientation on the ground, the reconstruction error of the AE is lower than the average one over the scene. The areas on the ground that actually correspond to the synthetic rotation (0° , 5° , and 9°) are closer to zero. The AE was only trained on the input dataset. However, it is able to provide an appropriate response when an unseen and arbitrarily rotated version of the input data is introduced. This implies that the application of synthetic rotation to a target is equivalent to the case when such an oriented target is actually present in the training data. Thus, the generalization capability of the AE is improved.

C. Analysis of Classification Performance

The Pauli composite image generated from the dataset is shown in Fig. 13 along with the training areas superimposed on it. They are derived from the interpretation of aerial imagery (ESRI/Bing Maps), topographical maps (USGS), and the Pauli composite and are marked with red polygons for urban areas, green for forest/vegetation areas, and blue for water. The pixels enclosed in these polygons form the ground truth. The proportion of ground-truth pixels in each class is approximately the same, and in total, they constitute less than 10% of the total pixels in the image. These are randomly divided into three groups: training, test, and validation. Pixels in the training set are used in the training of the network. At regular intervals within the training iterations, the process is halted and the test pixels are used to evaluate the accuracy of the network. The difference between the training and test accuracy is used to ensure that the training phase is progressing without overfitting or memorization. To quantify the final performance of the classifier, validation pixels are employed. The confusion matrix and accuracy statistics are presented in Table I.

The dataset is classified with an overall accuracy of 91.6% and a Kappa coefficient of 0.85. Urban areas are classified with an accuracy of 84.2%. The urban class is mostly confused with

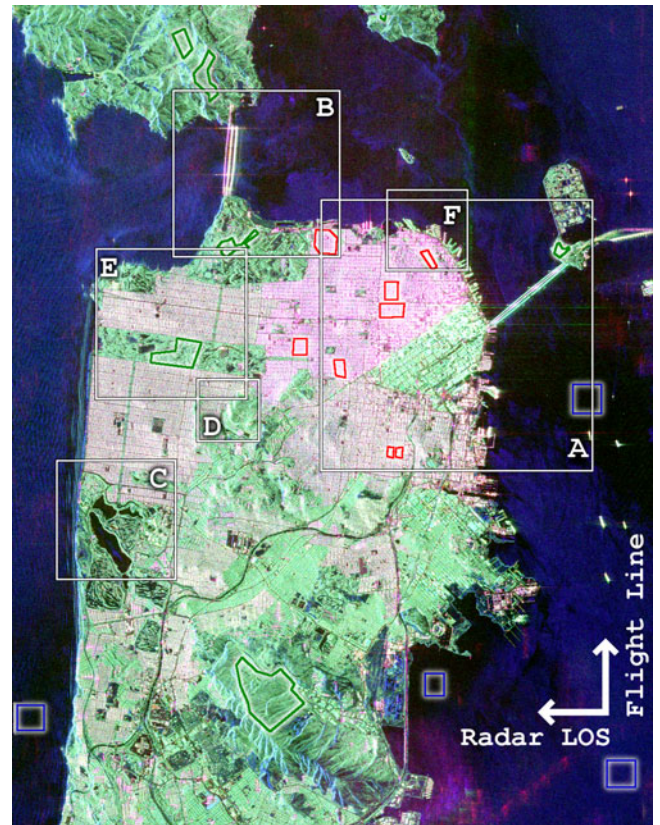


Fig. 13. Pauli RGB image constructed from the ALOS-2 dataset collected over San Francisco, USA, is shown with the ground-truth areas. The urban class is enclosed in red, the forest/vegetation in green, and water in blue polygons.

TABLE I
CONFUSION MATRIX AND ACCURACY STATISTICS FOR ALOS-2 SAN FRANCISCO, CA DATASET USING THE PROPOSED SUPERVISED CLASSIFICATION SCHEME INCORPORATING TEXTURAL INFORMATION

	Urban	Water	Forest	Prod. Acc.
Urban	84.2	0.0	4.4	95.5
Water	1.3	96.9	3.4	96.0
Forest	14.5	3.1	92.2	84.4
User Acc.	84.9	97.0	92.9	
OA (%)	91.6			
Kappa	0.85			

the forest class. This is primarily due to the fact that both urban areas rotated with respect to the radar line of sight and vegetation have higher cross-polarized return. Certain small, smooth vegetated areas like parks have been misclassified as the water class, contributing to the error of 3.1% due to the low backscattered power as previously highlighted.

The resultant map is presented in Fig. 14. Reference aerial photographs are given in Fig. 15. We can see a close correspondence between the areas that appear to be urban in the optical aerial images and the classification result. Detail subsets from selected areas of the classification output along with the corresponding aerial imagery, representing the ground truth, are shown in Fig. 16. This demonstrates the performance of the

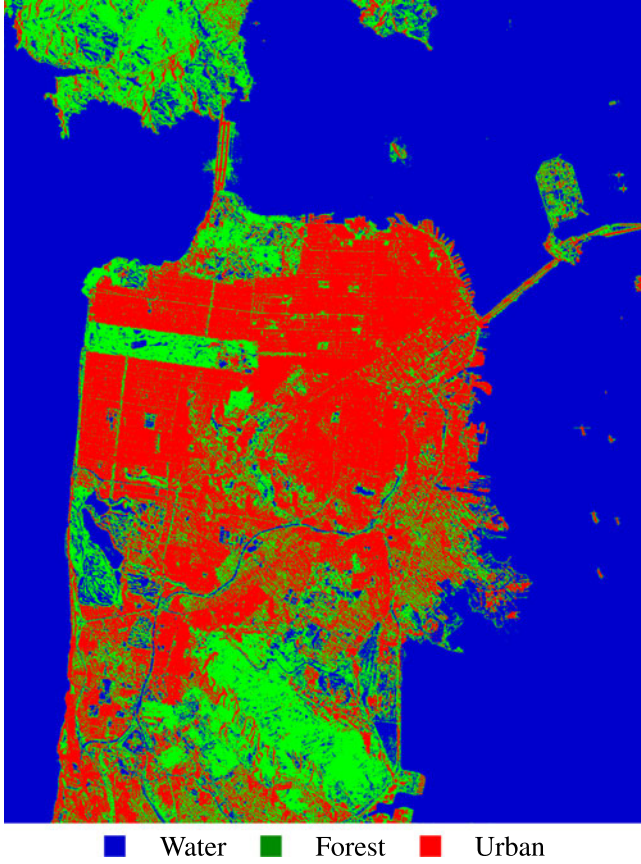
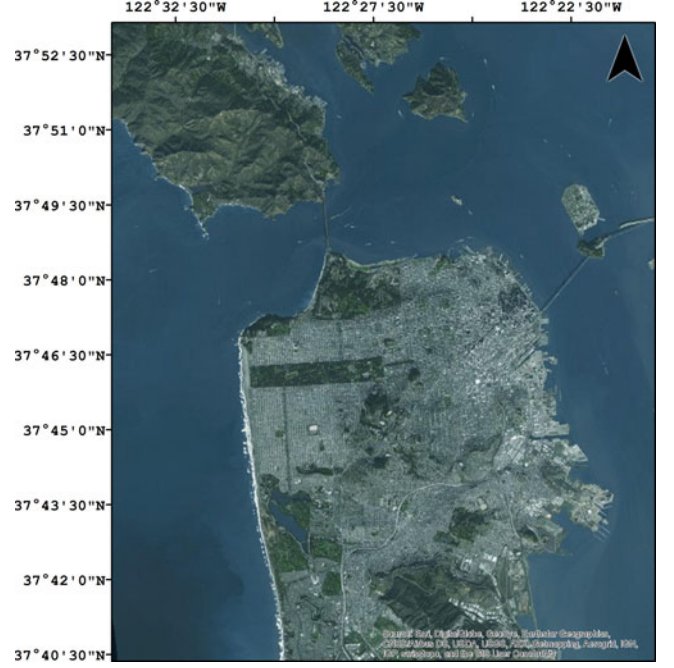


Fig. 14. Classified map of San Francisco derived from an ALOS-2 full polarimetric L-band dataset.

proposed algorithm on the considered dataset. The resolution of this imagery is ~ 0.3 m and is significantly higher than that of the classified map. However, reasonable correspondence can be seen between the two.

The subset shown in Fig. 16(a) represents a dense urban area that is oriented at an angle $\sim -20^\circ$ from the radar line of sight, marked as Area “A1” in the image. Due to this large orientation, this region is often misclassified [66]; however it is classified correctly using the proposed method. The subset shown in Fig. 16(b) shows the Golden Gate bridge, marked as “A2” and some surrounding areas. Apart from the return from the bridge itself, multipath returns from reflections from the bridge and the water surface can also be seen. The radar echoes from the vicinity of the bridge are quite strong and have a high value of γ^0 . This causes the third stage of the classifier, which takes γ^0 as an input, to classify this pixels immediately under the bridge as forest. Fig. 16(c) shows a lake surrounded by urban and vegetated areas. This is marked with a white outline on the aerial images and classification maps. It can be seen that the lake is well discriminated from other classes. As discussed earlier, the smooth vegetated surfaces near the lake are misclassified as water due to near specular reflection at L-band frequencies and low incidence angles.

The area marked as “A3” in the subset shown in Fig. 16(d), consists of a park between two regions of high-density urban



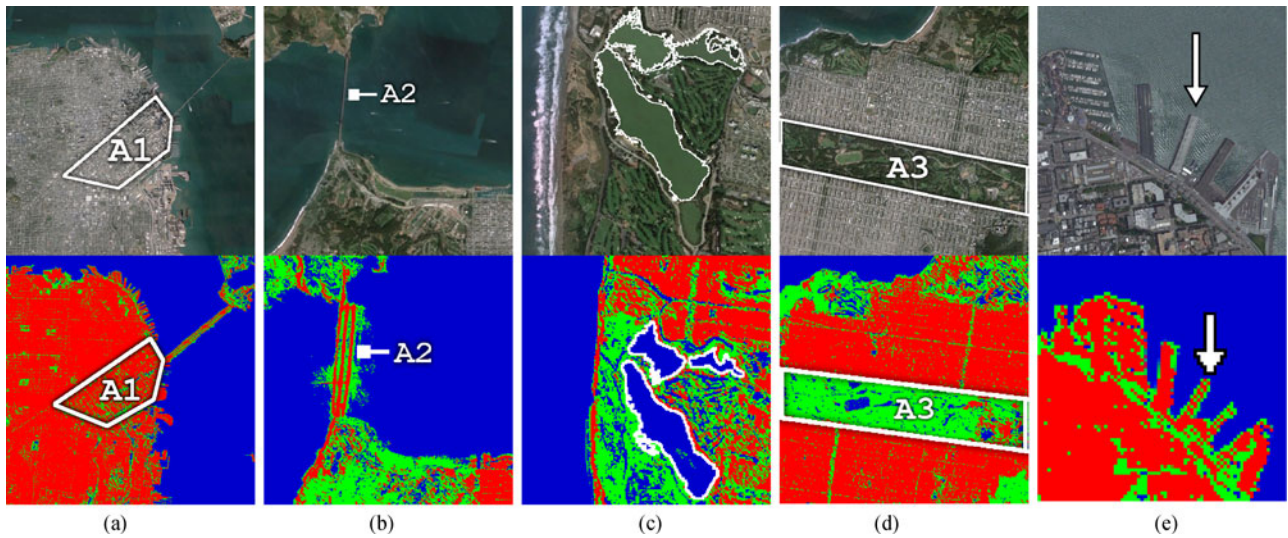


Fig. 16. Crops of the classified map shown alongside corresponding aerial imagery courtesy ESRI/Bing Maps.

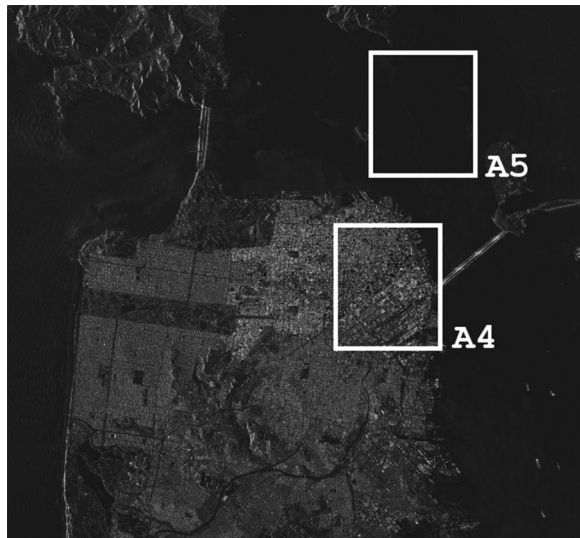


Fig. 17. Feature map for one of the internal nodes in the representation layer from the trained AE.

target orientation. This allows for the accurate classification of the urban areas. In contrast, homogeneous areas like Area “A5” have a lower response allowing them to be distinguished in the final classification.

E. Comparison With State-of-the-Art Classifiers

The proposed method is compared with other commonly used SAR classification techniques. The same training areas were used to train all classifiers.

The class-by-class accuracy is reported in Table II and the results are presented in Fig. 18. The SVM classifier was run with two kernels: a polynomial kernel (SVM-Poly) and an RBF kernel (SVM-RBF). The parameters of the RBF kernel were determined by the method described in [67]. The polynomial

TABLE II
CLASS-BY-CLASS TEST-AREA ACCURACY STATISTICS FOR ALOS-2, SAN FRANCISCO, CA DATASET FOR THE PROPOSED TECHNIQUE, SVM WITH RBF KERNEL, THE SVM WITH POLYNOMIAL KERNEL, AND THE WISHART SUPERVISED CLASSIFIERS

	Proposed	SVM-RBF	SVM-Poly	Wishart + POC	Wishart
Urban	86.2	64.2	58.5	59.1	55.6
Water	98.1	96.4	96.3	92.1	92.1
Forest	86.1	84.2	83.6	83.2	82.8
OA	90.8	81.6	79.4	78.1	76.8

kernel used has a degree of 4. The Wishart classifier was run both with polarization orientation correction (Wishart+POC) and without (Wishart). The proposed classifier has an urban area classification accuracy of 86.2%. It outperforms the SVM-RBF, SVM-Poly, Wishart+POC, and Wishart classifier, which have urban area accuracies of 64.2%, 58.2%, 59.1%, and 55.6%, respectively. The POC is able to improve the discernibility. However, targets that are oriented significantly with respect to the radar line of sight are still misclassified.

The areas shown in the subset, marked “A” and “B,” are highly urbanized as seen in the optical image collected over San Francisco (see Fig. 15). The cross-polarized return from area “A” makes it difficult to discern this from actual areas of vegetation. However, due to the synthetic rotation, the proposed scheme is able to account for the orientation information and more successfully classify the region than the other methods, while maintaining high textural fidelity. The mean orientation angle of area “B” is almost perpendicular to that of area “A,” and the region is characterized by large open spaces between man-made structures. The mixed nature of the pixels makes classification more difficult. As seen in Fig. 18(a), the proposed technique correctly classifies these structures while largely preserving the texture. The other methods erroneously classify most of this region with the RBF kernel SVM performing better than the polynomial

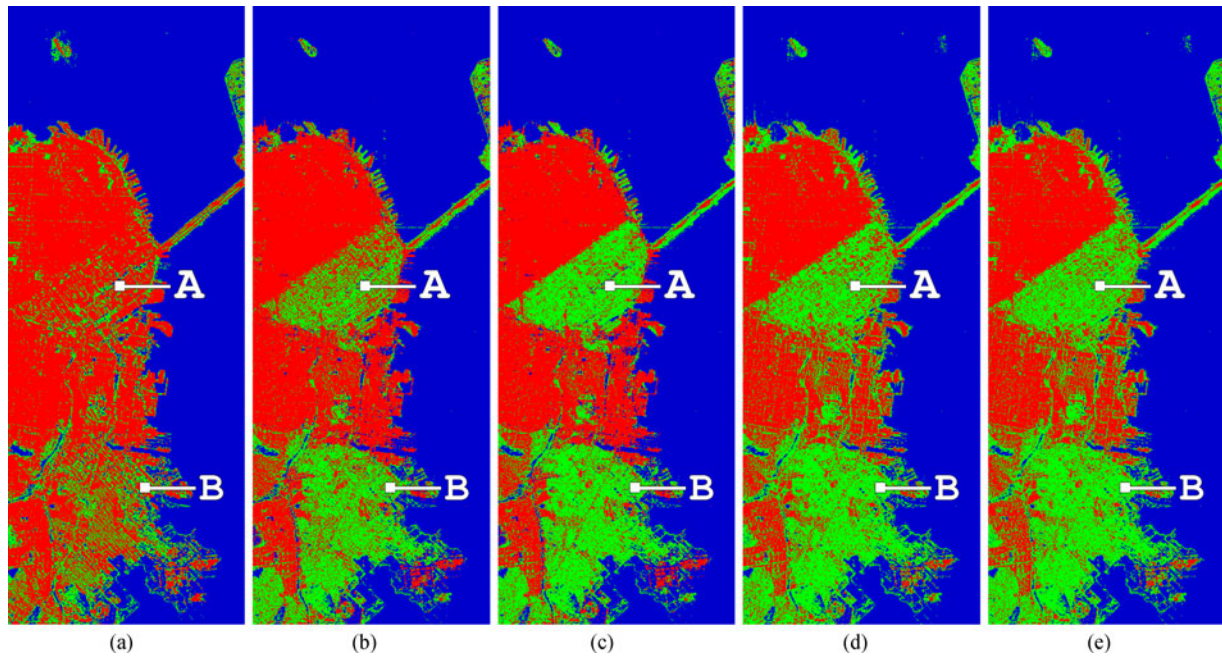


Fig. 18. Classification maps obtained with (a) Proposed method, (b) SVM RBF, (c) SVM polynomial, (d) Wishart with POC, and (e) Wishart.

kernel SVM and Wishart supervised classifier. Wishart+POC performs slightly better than SVM with the polynomial kernel.

V. CONCLUSION

Monitoring of urban areas for growth and disaster relief is an important application of radar remote sensing. However, the identification of urban areas from PolSAR data is challenging due to the misclassification of urban targets that are not perpendicular to the radar illumination. In this paper, we have proposed a novel technique based on a generalized synthetic target database generated according to the principles of the physics of scattering and a deep neural network architecture. The dataset is synthetically rotated by steps of 3° and converted to Mueller matrices. The transformation of the complex-valued coherency matrix to the real-valued Mueller matrix is done to simplify the computation during the backpropagation of the neural network. The creation of the target database serves as a data augmentation step that is integrated into a unified and reduced feature representation by the AE. The increase in information content improves the generalization ability of the network. As a result, it can successfully classify urban targets rotated with respect to the radar line of sight, even though such examples are not presented in the initial training phase.

The generalization ability is demonstrated on a portion of an UAVSAR L-band SAR image. The network is trained on the generalized urban target database, which was created synthetically by simulating the response of targets in the scene through a range of orientation angles. For testing, the data were rotated to a known angle artificially and applied as input. Hence, it is observed that areas on the ground with a certain artificially rotated input angle have a stronger network response. Thus, the synthetic rotation is linked to the physical properties of the target and can be exploited for improving the generalization ability.

The performance of the classification scheme is demonstrated on an ALOS-2 L-band SAR image acquired over San Francisco. Overall accuracy is about 91.63% for the datasets considered. The proposed approach qualitatively and quantitatively outperforms contemporary classification techniques commonly used for PolSAR data. The computation is efficient and fast using GPU techniques and has a potential for further speed up by upcoming hardware improvements. There is no limitation to the size of the dataset as training can proceed in small minibatches that can be scaled to fit the available GPU memory.

In the future, with the advent of sensors capable of multi-frequency and multisquint angle sensing, the synthetic target database generation step can be expanded. Target information at different frequencies and squint angles can be similarly integrated by a modified AE architecture to further improve generalization and classification ability. The angle step size can be adjusted adaptively depending on the scene heterogeneity and data spatial resolutions. The methodology can be adapted to evaluate its ability to learn efficient representations of other polarimetric parameters. The final classifier stage can be modified to incorporate spatial relations allowing for better context sensitive classification.

ACKNOWLEDGMENT

UAVSAR data was provided by NASA/JPL-Caltech. The authors are indebted to the anonymous reviewers for their valuable comments and suggestions, which were instrumental in the improvement of the quality of this paper.

REFERENCES

- [1] J. C. Curlander and R. N. McDonough, *Synthetic Aperture Radar*. New York, NY, USA: Wiley, 1991.

- [2] J. Van Zyl, R. Carande, Y. Lou, T. Miller, and K. Wheeler, "The NASA/JPL three-frequency polarimetric AIRSAR system," in *Proc. Int. Geosci. Remote Sens. Symp.*, 1992, pp. 649–651.
- [3] J.-S. Lee, M. Grunes, and E. Pottier, "Quantitative comparison of classification capability: Fully polarimetric versus dual and single-polarization SAR," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 11, pp. 2343–2351, Nov. 2001.
- [4] J. Kong, A. Swartz, H. Yueh, L. Novak, and R. Shin, "Identification of terrain cover using the optimum polarimetric classifier," *J. Electromagn. Waves Appl.*, vol. 2, no. 2, pp. 171–194, 1988.
- [5] E. Pottier and J. Saillard, "On radar polarization target decomposition theorems with application to target classification, by using neural network method," in *Proc. 7th Int. Conf. Antennas Propag.*, 1991, pp. 265–268.
- [6] J. A. Anderson, *An Introduction to Neural Networks*. Cambridge, MA, USA: MIT Press, 1995.
- [7] J.-S. Lee, M. R. Grunes, and R. Kwok, "Classification of multi-look polarimetric SAR imagery based on complex Wishart distribution," *Int. J. Remote Sens.*, vol. 15, no. 11, pp. 2299–2311, 1994.
- [8] J.-S. Lee, M. R. Grunes, and E. Pottier, "Quantitative comparison of classification capability: Fully polarimetric versus dual and single-polarization SAR," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 11, pp. 2343–2351, Nov. 2001.
- [9] J.-S. Lee, M. R. Grunes, E. Pottier, and L. Ferro-Famil, "Unsupervised terrain classification preserving polarimetric scattering characteristics," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 4, pp. 722–731, Apr. 2004.
- [10] M. Dabboor, M. J. Collins, V. Karathanassi, and A. Braun, "An unsupervised classification approach for polarimetric SAR data based on the Chernoff distance for complex Wishart distribution," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 7, pp. 4200–4213, Jul. 2013.
- [11] M. Dabboor, S. Howell, M. Shokr, and J. Yackel, "The Jeffries–Matusita distance for the case of complex Wishart distribution as a separability criterion for fully polarimetric SAR data," *Int. J. Remote Sens.*, vol. 35, no. 19, pp. 6859–6873, 2014.
- [12] A. Bhattacharya, G. Singh, S. Manickam, and Y. Yamaguchi, "An adaptive general four-component scattering power decomposition with unitary transformation of coherency matrix (AG4U)," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 10, pp. 2110–2114, Oct. 2015.
- [13] C. da Costa Freitas *et al.*, "Land use and land cover mapping in the Brazilian amazon using polarimetric airborne P-band SAR data," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 10, pp. 2956–2970, Oct. 2008.
- [14] S.-W. Chen, Y.-Z. Li, X.-S. Wang, S.-P. Xiao, and M. Sato, "Modeling and interpretation of scattering mechanisms in polarimetric synthetic aperture radar: Advances and perspectives," *IEEE Signal Process. Mag.*, vol. 31, no. 4, pp. 79–89, Jul. 2014.
- [15] C. Tison, J. M. Nicolas, F. Tupin, and H. Maitre, "A new statistical model for Markovian classification of urban areas in high-resolution SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 10, pp. 2046–2057, Oct. 2004.
- [16] F. Dell'Acqua and P. Gamba, "Texture-based characterization of urban environments on satellite SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 1, pp. 153–159, Jan. 2003.
- [17] M. R. Azimi-Sadjadi, S. Ghaloum, and R. Zoughi, "Terrain classification in SAR images using principal components analysis and neural networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 31, no. 2, pp. 511–515, Mar. 1993.
- [18] O. Antropov, Y. Rauste, H. Astola, J. Praks, T. Hame, and M. T. Hallikainen, "Land cover and soil type mapping from spaceborne PolSAR data at L-band with probabilistic neural network," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 9, pp. 5256–5270, Sep. 2014.
- [19] T. M. Pellizzeri, P. Gamba, P. Lombardo, and F. Dell'Acqua, "Multitemporal/multiband SAR classification of urban areas using spatial analysis: Statistical versus neural kernel-based approach," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 10, pp. 2338–2353, Oct. 2003.
- [20] Y. C. Tzeng and K. S. Chen, "A fuzzy neural network to SAR image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 36, no. 1, pp. 301–307, Jan. 1998.
- [21] P. Gamba and F. Dell'Acqua, "Increased accuracy multiband urban classification using a neuro-fuzzy classifier," *Int. J. Remote Sens.*, vol. 24, no. 4, pp. 827–834, 2003.
- [22] F. D. Frate, F. Pacifici, and D. Solimini, "Monitoring urban land cover in Rome, Italy, and its changes by single-polarization multitemporal SAR images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 1, no. 2, pp. 87–97, Jun. 2008.
- [23] P. Gamba and B. Houshmand, "An efficient neural classification chain of SAR and optical urban images," *Int. J. Remote Sens.*, vol. 22, no. 8, pp. 1535–1553, 2001. [Online]. Available: <http://www.tandfonline.com/doi/abs/10.1080/01431160118746>
- [24] G. Hinton, S. Osindero, and Y.-W. Teh, "A fast learning algorithm for deep belief nets," *Neural Comput.*, vol. 18, no. 7, pp. 1527–1554, 2006.
- [25] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2015, pp. 1026–1034.
- [26] L. Deng *et al.*, "Recent advances in deep learning for speech research at Microsoft," in *Proc. 2013 IEEE Int. Conf. Acoust., Speech Signal Process.*, 2013, pp. 8604–8608.
- [27] X. Glorot, A. Bordes, and Y. Bengio, "Domain adaptation for large-scale sentiment classification: A deep learning approach" in *Proc. 28th Int. Conf. Mach. Learn.*, 2011, pp. 513–520.
- [28] V. Mnih *et al.*, "Human-level control through deep reinforcement learning," *Nature*, vol. 518, no. 7540, pp. 529–533, 2015.
- [29] J. Han, D. Zhang, G. Cheng, L. Guo, and J. Ren, "Object detection in optical remote sensing images based on weakly supervised learning and high-level feature learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 6, pp. 3325–3337, Jun. 2015.
- [30] J. Tang, C. Deng, G.-B. Huang, and B. Zhao, "Compressed-domain ship detection on spaceborne optical image using deep neural network and extreme learning machine," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 3, pp. 1174–1185, Mar. 2015.
- [31] X. Chen, S. Xiang, C.-L. Liu, and C.-H. Pan, "Vehicle detection in satellite images by hybrid deep convolutional neural networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 10, pp. 1797–1801, Oct. 2014.
- [32] Y. Chen, Z. Lin, X. Zhao, G. Wang, and Y. Gu, "Deep learning-based classification of hyperspectral data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2094–2107, Jun. 2014.
- [33] Z. Zhu, C. E. Woodcock, J. Rogan, and J. Kellndorfer, "Assessment of spectral, polarimetric, temporal, and spatial dimensions for urban and peri-urban land cover classification using landsat and SAR data," *Remote Sens. Environ.*, vol. 117, pp. 72–82, 2012.
- [34] P. Gamba and M. Aldighi, "SAR data classification of urban areas by means of segmentation techniques and ancillary optical data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 4, pp. 1140–1148, Aug. 2012.
- [35] L. Zhang, G.-S. Xia, T. Wu, L. Lin, and X. C. Tai, "Deep learning for remote sensing image understanding," *J. Sensors*, vol. 501, 2015, Art. no. 173691.
- [36] Y. Li, X. Xu, B.-D. Bai, and Y.-N. Zhang, "Remote sensing image fusion based on fast discrete curvelet transform," in *Proc. 2008 Int. Conf. Mach. Learn. Cybern.*, 2008, vol. 1, pp. 106–109.
- [37] Y. Yamaguchi, A. Sato, W.-M. Boerner, R. Sato, and H. Yamada, "Four-component scattering power decomposition with rotation of coherency matrix," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 6, pp. 2251–2258, Jun. 2011.
- [38] P. Liu and Y. Q. Jin, "A study of ship rotation effects on SAR image," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 6, pp. 3132–3144, Jun. 2017.
- [39] S.-W. Chen, X.-S. Wang, and M. Sato, "Uniform polarimetric matrix rotation theory and its applications," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 8, pp. 4756–4770, Aug. 2014.
- [40] Y. Yamaguchi, A. Sato, W.-M. Boerner, R. Sato, and H. Yamada, "Four-component scattering power decomposition with rotation of coherency matrix," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 6, pp. 2251–2258, Jun. 2011.
- [41] B. Zou, D. Lu, L. Zhang, and W. M. Moon, "Eigen-decomposition-based four-component decomposition for PolSAR data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 3, pp. 1286–1296, Mar. 2016.
- [42] B. Zou, D. Lu, L. Zhang, and W. M. Moon, "Independent and commutable target decomposition of PolSAR data using a mapping from SU(4) to SO(6)," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 6, pp. 3396–3407, Jun. 2017.
- [43] W. Hu, Y. Huang, L. Wei, F. Zhang, and H. Li, "Deep convolutional neural networks for hyperspectral image classification," *J. Sensors*, vol. 2015, Art. no. 258619.
- [44] Y. Chen, X. Zhao, and X. Jia, "Spectral-spatial classification of hyperspectral data based on deep belief network," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 6, pp. 2381–2392, Jun. 2015.
- [45] H. Kimura, "Radar polarization orientation shifts in built-up areas," *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 2, pp. 217–221, Apr. 2008.
- [46] K. Tragl, "Polarimetric radar backscattering from reciprocal random targets," *IEEE Trans. Geosci. Remote Sens.*, vol. 28, no. 5, pp. 856–864, Sep. 1990.
- [47] J.-S. Lee, L. Jurkevich, P. Dewaele, P. Wambacq, and A. Oosterlinck, "Speckle filtering of synthetic aperture radar images: A review," *Remote Sens. Rev.*, vol. 8, no. 4, pp. 313–340, 1994.

- [48] J.-S. Lee, D. L. Schuler, T. L. Ainsworth, E. Krogager, D. Kasilingam, and W. M. Boerner, "On the estimation of radar polarization orientation shifts induced by terrain slopes," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 1, pp. 30–41, Jan. 2002.
- [49] A. Bhattacharya, A. Muhuri, S. De, S. Manickam, and A. Frery, "Modifying the Yamaguchi four-component decomposition scattering powers using a stochastic distance," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 7, pp. 3497–3506, Jul. 2015.
- [50] A. Hirose, *Complex-Valued Neural Networks*. New York, NY, USA: Springer, 2006.
- [51] R. Barakat, "Bilinear constraints between elements of the 4×4 Mueller-Jones transfer matrix of polarization theory," *Opt. Commun.*, vol. 38, no. 3, pp. 159–161, 1981.
- [52] X. Glorot, A. Bordes, and Y. Bengio, "Deep sparse rectifier neural networks" in *Proc. Int. Conf. Artif. Intell. Statist.*, 2011, pp. 315–323.
- [53] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.
- [54] V. Nair and G. E. Hinton, "Rectified linear units improve restricted Boltzmann machines," in *Proc. 27th Int. Conf. Mach. Learn.*, 2010, pp. 807–814.
- [55] Y. Bengio, "Learning deep architectures for AI," *Found. Trends Mach. Learn.*, vol. 2, no. 1, pp. 1–127, 2009.
- [56] A. Frery, H.-J. Muller, C. Yanasse, and S. Sant'Anna, "A model for extremely heterogeneous clutter," *IEEE Trans. Geosci. Remote Sens.*, vol. 35, no. 3, pp. 648–659, May 1997.
- [57] A. C. Frery et al., "Models for synthetic aperture radar image analysis," *Resenhas (IME-USP)*, vol. 4, no. 1, pp. 45–77, 1999.
- [58] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks" in *Proc. Int. Conf. Artif. Intell. Statist.*, 2010, pp. 249–256.
- [59] A. G. Fore et al., "UAVSAR polarimetric calibration," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 6, pp. 3481–3491, Jun. 2015.
- [60] J.-S. Lee, M. Grunes, E. Pottier, and L. Ferro-Famil, "Unsupervised terrain classification preserving polarimetric scattering characteristics," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 4, pp. 722–731, Apr. 2004.
- [61] S. Hensley et al., "The UAVSAR instrument: Description and first results," in *Proc. Radar Conf.*, May 2008, pp. 1–6.
- [62] C. Elachi et al., *Radar Polarimetry for Geoscience Applications*. Norwood, MA, USA: Artech House, 1990.
- [63] D. K. Atwood, D. Small, and R. Gens, "Improving PolSAR land cover classification with radiometric correction of the coherency matrix," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 3, pp. 848–856, Jun. 2012.
- [64] M. Tao, F. Zhou, Y. Liu, and Z. Zhang, "Tensorial independent component analysis-based feature extraction for polarimetric SAR data classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 5, pp. 2481–2495, May 2015.
- [65] B. Banerjee, A. Bhattacharya, and K. M. Buddhiraju, "A generic land-cover classification framework for polarimetric SAR images using the optimum Touzi decomposition parameter subset insight on mutual information-based feature selection techniques," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 4, pp. 1167–1176, Apr. 2014.
- [66] J.-S. Lee and T. Ainsworth, "The effect of orientation angle compensation on coherency matrix and polarimetric target decompositions," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 1, pp. 53–64, Jan. 2011.
- [67] Z. Liu, M. J. Zuo, and H. Xu, "A Gaussian radial basis function based feature selection algorithm," in *Proc. 2011 IEEE Int. Conf. Comput. Intell. Meas. Syst. Appl.*, 2011, pp. 1–4.



Shaunak De (S'15) received the B.E. degree in electronics engineering from the University of Mumbai, Mumbai, India, in 2012, graduating as a gold medalist. He is currently working toward the Ph.D. degree in remote sensing image analysis at the Centre of Studies in Resources Engineering, Indian Institute of Technology Bombay, Mumbai, India.

His current research interests include polarimetric SAR, machine learning, and information theory. He serves as a reviewer for international journals, among

them, the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING and the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS.



Lorenzo Bruzzone (S'95–M'98–SM'03–F'10) received the Laurea (M.S.) degree in electronic engineering (*summa cum laude*) and the Ph.D. degree in telecommunications from the University of Genoa, Genoa, Italy, in 1993 and 1998, respectively.

He is currently a Full Professor of telecommunications at the University of Trento, Trento, Italy, where he teaches remote sensing, radar, and digital communications. He is the founder and the Director of the Remote Sensing Laboratory, Department of Information Engineering and Computer Science, University of Trento. He is the author (or coauthor) of 193 scientific publications in referred international journals (139 in IEEE journals), more than 260 papers in conference proceedings, and 21 book chapters. His current research interests include remote sensing, radar and SAR, signal processing, and pattern recognition. He promotes and supervises research on these topics within the frameworks of many national and international projects.

Dr. Bruzzone is the Principal Investigator of many research projects. Among the others, he is the Principal Investigator of the Radar for icy Moon exploration (RIME) instrument in the framework of the JUPITER ICY moons Explorer (JUICE) mission of the European Space Agency. He is the Editor/Co-Editor of 18 books/conference proceedings and 1 scientific book and a Guest Co-Editor of many Special Issues of international journals. His papers are highly cited, as proven from the total number of citations (more than 19 500) and the value of the h-index (69) (source: Google Scholar). He was invited as a keynote speaker in more than 30 international conferences and workshops. Since 2009, he is a member of the Administrative Committee of the IEEE Geoscience and Remote Sensing Society (GRSS). He ranked first place in the Student Prize Paper Competition of the 1998 IEEE International Geoscience and Remote Sensing Symposium (Seattle, July 1998). Since then he also received many international and national honors and awards, including the recent IEEE GRSS 2015 Outstanding Service Award. He is the co-founder of the IEEE International Workshop on the Analysis of Multi-Temporal Remote-Sensing Images (Multi-Temp) series and is currently a member of the Permanent Steering Committee of this series of workshops. Since 2003, he has been the Chair of the SPIE Conference on Image and Signal Processing for Remote Sensing. Since 2013, he has been the founder Editor-in-Chief of the IEEE GEOSCIENCE AND REMOTE SENSING MAGAZINE. He is currently an Associate Editor of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING. Since 2012, he has been appointed Distinguished Speaker of the IEEE Geoscience and Remote Sensing Society.



Avik Bhattacharya (M'08–SM'16) received the M.Sc. degree in mathematics from the Indian Institute of Technology Kharagpur, Kharagpur, India, in 2000 and the Ph.D. degree in remote sensing image processing and analysis from Télécom ParisTech, Paris, France, and the Ariana Research Group, Institut National de Recherche en Informatique et Automatique (INRIA), Sophia Antipolis, Nice, France, in 2007.

He is currently an Associate Professor at the Centre of Studies in Resources Engineering, Indian Institute of Technology Bombay, CSRE, IITB, Mumbai, India. Before joining IITB, he was a Canadian government Research Fellow at the Canadian Centre for Remote Sensing, Ottawa, ON, Canada. He has received the Natural Sciences and Engineering Research Council of Canada visiting scientist fellowship at the Canadian national laboratories from 2008 to 2011. His current research interests include SAR polarimetry, statistical analysis of polarimetric SAR images, machine learning, and pattern recognition.

He is an Associate Editor of the IEEE GEOSCIENCE AND REMOTE SENSING Letters. He is the founding Chairman of the IEEE Geoscience and Remote Sensing Society of the Bombay Chapter. At present, he is leading the Microwave Remote Sensing Lab, CSRE, IITB (www.msrlab.in).

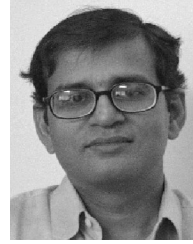


Francesca Bovolo (S'05–M'07–SM'13) received the Laurea (B.S.) and the Laurea Specialistica (M.S.) degrees in telecommunication engineering (*summa cum laude*) and the Ph.D. in communication and information technologies from the University of Trento, Trento, Italy, in 2001, 2003 and 2006, respectively.

She was a Research Fellow at the University of Trento until June 2013. She is the Founder and the Head of the Remote Sensing for Digital Earth Unit at Fondazione Bruno Kessler, Trento, Italy, and a member of the Remote Sensing Laboratory, Trento, Italy.

Her main research interests include the area of remote-sensing image processing. Her interests are related to multitemporal remote sensing image analysis and change detection in multispectral, hyperspectral, and SAR images, and very high-resolution images, in particular. She conducts research on these topics within the context of several national and international projects. She is a referee for several international journals.

Dr. Bovolo ranked first place in the Student Prize Paper Competition of the 2006 IEEE International Geoscience and Remote Sensing Symposium (Denver, August 2006). Since January 2011, she has been an Associate Editor of the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING and of SPIE *Journal of Applied Remote Sensing* since 2015. She has been a Guest Editor for the Special Issue on Analysis of Multitemporal Remote Sensing Data of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING. She is the Technical Chair of the Sixth International Workshop on the Analysis of Multi-temporal Remote-Sensing Images (MultiTemp 2011). From 2006 to 2013, she has served on the Scientific Committee of the SPIE International Conference on Signal and Image Processing for Remote Sensing. Since 2014, she has been the Co-Chair of the same conference. Since 2012, she has been a member of the international program committee of the conference on Pattern Recognition Applications and Methods. She has served on the Scientific Committee of the IEEE Fourth and Fifth International Workshop on the Analysis of Multi-Temporal Remote Sensing Images (MultiTemp 2007 and 2009) and the IEEE GOLD Remote Sensing Conference in 2010, 2012, and 2014.



Subhasis Chaudhuri (M'99–SM'02–F'11) received the B.Tech. degree from the Indian Institute of Technology (IIT) Kharagpur, Kharagpur, India, in 1985, the M.Sc. degree in electrical engineering from the University of Calgary, Calgary, AB, USA, in 1987, and the Ph.D. degree from the University of California San Diego, San Diego, CA, USA, in 1990.

He is currently the KN Bajaj Chair Professor in the Department of Electrical Engineering, Indian Institute of Technology Bombay, Mumbai, India. His primary research interests include pattern recognition, image processing, computer vision, and computational haptics.

Dr. Chaudhuri is a Fellow of the science and engineering academies in India. He has received the Bhatnagar Prize in Engineering Sciences in 2004 and the GD Birla Research Award in 2011. He is a Distinguished Alumnus of IIT Kharagpur.