Development of Auto-Encoder Based Methods for Synthetic Aperture Radar Image Analysis

Shaunak De

Contents

Contents

List of Tables

List of Figures

Chapter 1 Introduction

Remotely sensed images acquired by space-borne or airborne sensors have revolutionized earth observation allowing an unprecedented ability to image and observe large areas of the planet's surface with high temporal repetitivity. This has opened new avenues for the monitoring of dynamic earth processes. Long-term earth observation missions have ensured data continuity over decades allowing the characterization of natural and anthropocentric processes and changes. However, in order for the images to be used for these applications, they must be processed so that the information of interest is brought out. Traditionally thematic maps have been produced by a field survey of the region to be mapped. This technique, while accurate, is time-consuming and expensive; Prohibitively so for large areas. The use of remote sensing techniques allows for the production of small to medium scale thematic maps (encompassing continents or even world maps) quickly and economically with a fair degree of accuracy [2].

Classification of remote sensing data involves grouping of pixels into homogeneous groups, usually representing a type of land cover. This allows the generation of thematic maps. These maps depict various land covers depending on the target application. For example, an agriculture map would typically denote various crop types in an area. Thematic maps play a major role in the large-scale monitoring of several geophysical parameters. Forest cover maps are used to quantify and control deforestation, sea-ice maps are crucial for monitoring the motion of ice-bergs thus safeguarding navigation channels, agriculture maps are often used for yield prediction of food grains, flood area maps help in the minimization of loss to property and mitigation of damage due to flood inundation, urban area maps help study the expansion of human settlement over decades etc. Many of these maps, especially those used for disaster management or the tracking of fast-changing geo-dynamics, must be generated rapidly with a fair degree of accuracy. This information helps shape governance policies, aid legislative decision-making processes and help generate a better understanding of the world around us.

Classification of remotely sensed images entails the automated identification of targets like buildings, cars, armament etc. in acquired scenes relying on training libraries, previously identified targets or other target descriptors. The all-weather, day-night imaging capability of a Synthetic Aperture Radar (SAR) system is highly desirable for defense, security and monitoring applications, where time-critical imaging is required. Unmanned aerial vehicle (UAV) borne SAR systems are being operationally used for information gathering. A peripatetic platform like this is capable of high repetitivity and frequent re-visits, thus it generates a high volume of data.

In such a scenario, the ability to extract information from the datasets in an automated manner is essential. However, object recognition and classification in SAR images is not without its challenges. SAR images suffer from geometric distortions and speckle-like noise patterns which are inherent to the process of SAR image formation. Polarimetric

SAR (PolSAR) images contain more information than single or dual channel SAR, but it is important to develop techniques to utilize this information entirely.

Automation in classification and identification tasks, or at least part-automation, can help completely utilize the potential of the large data volumes. Machine learning algorithms can help collate information from the data, which can eventually lead to the extraction of knowledge from the information. Learning machines are designed to tune their internal parameters according to the input data to improve their efficiency and/or accuracy. Machine learning algorithms are able to utilize the extensive data quantities. However, traditional machine learning algorithms usually need specially designed input features, either from $a-priori$ information about the distribution of the data, or domain knowledge, for a particular application, making complete automation prohibitive. Also, beyond a point, increase in the amount of available training data does not lead to a corresponding improvement in task efficacy [3].

Deep Learning is a new approach based on the workings of the human cognitive system pioneered in 2006 by [4]. The principle behind deep learning is to distribute a complex task into simpler subtasks. Different parts of the network can then be responsible for each sub-task, eventually being connected cohesively to solve the original task. It has shown superior performance as compared to 'shallow' approaches in various classification and recognition challenges and is rapidly being applied to diverse fields like image classification, object recognition, speech and language processing, etc. Deep Learning techniques greatly benefit from increased training information, more so than other contemporary methods, making them a good fit for remote sensing applications in which high volumes of similar data is collected.

1.1 Motivation

Optical remote sensing has been primarily used for earth observation, monitoring and disaster management applications. However, being a passive sensor, it's limited to operating in the presence of solar irradiation. Further, it is impeded by the presence of cloud cover. SAR, on the other hand, is an active imaging technique and has an advantage that it is able to operate independently of ambient illumination. It is also insensitive to the weather due to its capability of penetrating cloud cover. SAR is sensitive to both the wavelength scale geometrical and dielectric properties of the target, allowing it to ascertain information about the target that is not possible with reflectance alone. PolSAR is an advancement over SAR imaging, which uses a combination of transmit and receive polarizations in quadrature. Thus, it is sensitive to the wavelength-scale structure of the target and has better discrimination potential. To be able to exploit this increased target information fully, and to efficiently handle the increasing volume of data being collected by present and planned satellite constellations, automated analysis techniques must be developed. Machine learning algorithms are a reasonable choice for this task, however, hand-crafted features need to be designed from domain knowledge and require

expert intervention, making it time-consuming and prohibitive for large data volumes. The aim of deep learning is to develop a simple feature representation internally and automatically without the need for hand-crafted or a specially designed input space. This makes Deep Learning techniques especially synergistic for PolSAR applications. The motivation for this thesis has been to explore the possibilities of enhancing Earth observation applications of PolSAR data using novel deep learning techniques.

1.2 Research Objectives

In this thesis, novel deep learning algorithms and architectures are proposed for various earth observation applications using fully polarimetric SAR data. The methodologies have been developed and verified using airborne L-band UAVSAR, airborne C-, L-, P-band AIRSAR, spaceborne C-band RADARSAT-2 and spaceborne L-band ALOS-2 datasets. Applications have been demonstrated in the field of agricultural monitoring, urban classification, change detection and disaster response,

The following objectives have been accomplished in the course of this thesis:

1. Auto-Encoders have been demonstrated in finding an efficient sparse feature-space representation of PolSAR data.

2. Scattering physics has been incorporated in an auto-encoders architecture using a novel technique for data-augmentation, which improves generalization capacity.

3. A novel deep-learning framework has been developed that outperformed state of the art techniques in the classification of urban areas, especially those not perpendicular to the line of sight of the radar.

4. Development of a novel semi-supervised technique that allows change detection in urban areas with very few training points using a region growing technique in the auto-encoder's feature space.

5. A novel tensorization technique in conjunction with a spatial auto-encoder has been proposed that is able to combine information from multi-frequency bands to improve classification of natural areas.

Chapter 2Literature Review

Computers have been employed to solve complex computations that would be tedious and time consuming for a human to conduct. These typically deal with executing well-defined sequences of instructions on data, very rapidly. However, computers can only be used as problem-solving tools if both, the problem is well specified, and its solution is tractable. Computers are typically designed to allow the evaluation of a complex sequence of mathematical operations rapidly; this gives them their apparent problem-solving ability. However, to obtain a solution, a computer must first be supplied with a set of instructions or an algorithm, the construction of which is only possible if the steps to the solution can be formalized in codified.

The prediction of the landing spot of a projectile is an example of such a problem. The rules concerning the descent of the body are governed by the laws of physics which may be precisely specified to a computing system, thus leading to the calculation of the correct answer. On the other hand, if the solution to the problem is not deterministic, it may not be possible to arrive at an answer. Pattern/outliers recognition, trend prediction are examples of such kinds of problems. Many real-world problems are traditionally solved numerically based on certain approximation models. If the model is not known beforehand, it is not possible to compute an answer. The model is also inflexible to change in input trends. Machine learning based approaches attempt to discover trends in the data and thus construct a data-driven model, make predictions or adjust the internal parameters of the model depending on changing trends in the data to better it's performance or improve prediction.

2.1 Learning Machines

A learning machine is one that has the ability to adjust its internal workings (code, data or model) to allow it to gain efficiency or accuracy in its task in response to external information or feedback [5]. It is a machine that can learn from its errors and minimize them on future input samples. A learning machine in a lot of ways functions exactly the way human beings animals learn about the world around them. We tend to learn from the statistical dependencies of our actions to the consequences. Over time we adapt our behavior to maximize the rewards from the circumstances. Often the judgment of this reward and consequence itself is based on several factors. The human brain has the ability to generalize previous experiences and apply them to the current circumstance to come up with an optimum decision. Learning machines attempt to mimic this soft computing ability in order to come up with solutions.

It is reasonable to question as to why a machine must be designed to learn; why can it not be constructed such that it performs its task optimally in the first place. It is

not always possible to design a machine that can anticipate all variables and variations in input, environment or its own internal workings. Despite the best efforts of human designers, it may not be able to completely characterize the system. Thus, a machine that adapts on the fly will be more suited to handling situations not envisioned initially. The input data volume might be so large that it would take a long time for humans to codify it completely. A machine that gradually learns patterns from the data, in this situation would be able to capture far more information from the data.
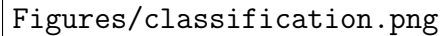
The data itself may be subject to a trend, that is there may be a co-related change in some aspects of the data which are consequences of some external influencing factors. Machines that adapt to changes in inputs need not be resigned to accommodate that. Google designed a system that co-related Internet search terms with the spread of flu, allowing them to make predictions about the spread of winter flu in the United States [6]. However, the system failed to adapt to the change in Internet search behavior during the pH1N1 epidemic in 2009, which caused the system to produce incorrect results [7]. In such cases, machines that can learn from and adapt to change in data trends will exhibit superior performance.

Learning algorithms have been applied to solve a variety of problems including, text or document classification, e.g. spam detection, subject identification, natural language processing, e.g. speech to text, part of speech tagging, computer vision, e.g. image classification, object identification, unassisted vehicle control, e.g. robot navigation, driver aids, optical character recognition (OCR), fraud detection (credit cards, telephones), network security e.g. intrusion detection, denial of service attack prevention, computational biology, e.g. protein function prediction medical diagnosis, recommender systems, e.g. search engines, related content identification, clustering systems, e.g. news collation, related tag identification;

This list is by no means exhaustive, and every day, learning algorithms are being used in new applications every day. It is possible to identify a number of broad templates to address like natured problems. Typical tasks for learning machines can be identified as, [8, 7-10]
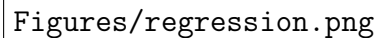
- **Binary Classification** involves the separation of input data into two classes. For instance, the example shown in Figure 2.1. In other words, given a pattern x drawn from a domain χ, the task is to estimate the value of an associated binary random variable $y \in \pm 1$. The final decision may be depended on the input pattern, or on the sequence of input patterns.

- **Multi-Class Classification** is an extension of the binary classification problem, where the set of lables may assume a larger set of values, i.e. $y \in 1, \ldots, n$. For instance we may want to classify emails into categories (Work, Personal, Promotions, Subscriptions, . . .) according to the contents and the sender's address. See Figure 2.1 for an example.

- **Structuted Estimation** is an evolution of the multi-class classification problem. It attempts to improve the multi-classification results based on a known structure.

Figure 2.1: Binary classification (left); separate the stars from triangles. In this example it is possible to do so by drawing a straight line. This serves as an example of a *linear classifier*. Multiclass Classification (right): 3-class classification is shown. Note that in the latter case we have a greater degree for ambiguity. For instance, being able to distinguish stars from diamonds may not suffice to identify either of them correctly, since we also need to distinguish both of them from triangles.

Figure 2.2: We are given a number of instances (indicated by the dots) and would like to find some function f mapping the observations such that $f(x)$ is close to the input values.

For example an ontology may be used while attempting to classify research papers or seasonal temperature information may be used to aid iceberg detection.

- **Regression** aims to estimate a real-valued variable $y \in \mathbb{R}$ given an input pattern x (see Figure 2.2). For instance, we might want to estimate the adrenaline level in a patient based on a blood cortisol measurement.

- **Ranking** is the task of using related parameters to order a set according to some predefined criterion. Internet search rankings are a common example of such a problem where we must rank and sort web pages according to their relevance to the search terms.

- **Clustering** involves the partitioning of input patters into homogeneous groups. For instance, using social media data to group people according to their interests.

- **Dimensionality reduction or Manifold learning** involves the transformation

```
Figures/deeplearning.png
```
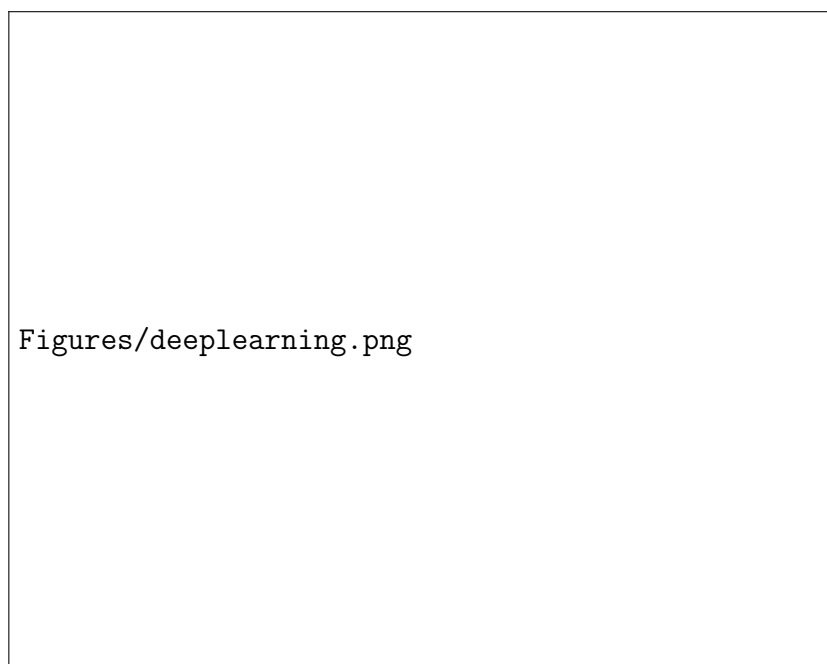
Figure 2.3: Rule based systems are constructed to follow certain hand designed features and generalize poorly when unseen examples are presented. Classical machine learning attempts to learn a mapping from hand designed features, this gives it its generalization ability, however the task of designing the features itself is tedious. Representational learning aims to learn the features from the presented data, however this problem can become as difficult as the original problem itself. The shaded boxes represent the components that are able to learn from data.

of a representation of data into a lower dimensional representation with minimal loss of information. Hyper-spectral images have very high dimensionality (several hundred bands) which can be difficult to work with for computer vision tasks. By discarding irrelevant and redundant information, the dimensionality can be lowered.

2.2 The Problem of Representation

Input data to a machine learning system can be represented in many ways, but some representations help improve the ability of a machine learning algorithm to extract knowledge from the supplied data. For example, numbers represented in the familiar binary notation may not always be successful in machine learning applications, and a one-hot representation may be preferred. The one hot notation involves setting successive bits high for successive numbers. For example, 1 is represented as 0001, 2 as 0010, 3 as 0100 and so on. In a traditional binary representation, numbers that are in proximity to one another have very little similarity, for example 7 (00000111) and 8 (00001000) are

separated by just one position on the number line, yet have no bits in common, while 64 (01000000) and 1 (00000001), which are quite far apart are separated by just one bit in the representation. This causes the learning machine to generalize poorly and will affect the performance of the architecture [9].

A lot of applications of machine learning involve input data that cannot be directly represented numerically. Computer vision for instance deals with images, the classification of spam email deals with raw text written in human languages. It becomes important to convert this input to a system that can be numerically encoded for input to the learning machine. The chosen parameters, which are used to represent a certain property of the data, are called features. The operation thus is a transformation of the data into the feature space. For instance, in an image corner points, lines and edges may be considered to be featured. In voice recognition, the pitch of the audio sample can be considered to be a feature.

Representation of the data is thus a central problem in the design of any machine learning algorithm. For many real-world applications, the selection of an appropriate representation scheme can be dauntingly difficult. One solution is to use a machine learning technique itself to capture not just the mapping from input to output but also the representation. This technique is called *representational learning* [10].

Learned features can offer a higher performance, and save human time and effort that would have been spent in constructing a hand designed model. However, identification of the features is not a trivial task. *Factors of variation* are used to examine the relation between the input data and the data. These factors may not be directly present or sensed in the data, but rather exist in the human mind that allows us to discriminate and make inferences about the data. These may be concepts and abstractions that help us process and understand the data.

In the real world, every data sample is affected by a wide variety of factors, and it becomes important to separate these factors from the data, for instance, a tennis ball may become discolored over the course of a match, and the shape of the ball itself changes on impact. In this situation obtaining a representation of the data requires a human-like understanding of the data. In this situation, representational learning may not be helpful because the complexity of obtaining the representation is comparable to the solution of the original problem in itself.

2.3 Deep Learning

It can be difficult to learn high-level features from data. Deep Learning attempts to solve this by representing the original problem in a series of simpler sub-representations. The depth of the architecture refers to the number of levels of a composition of non-linear operations in the function learned. It allows the learning algorithm to learn a complex abstract concept from a series of more straightforward concepts. The multiple

Figures/deeplearning2.png
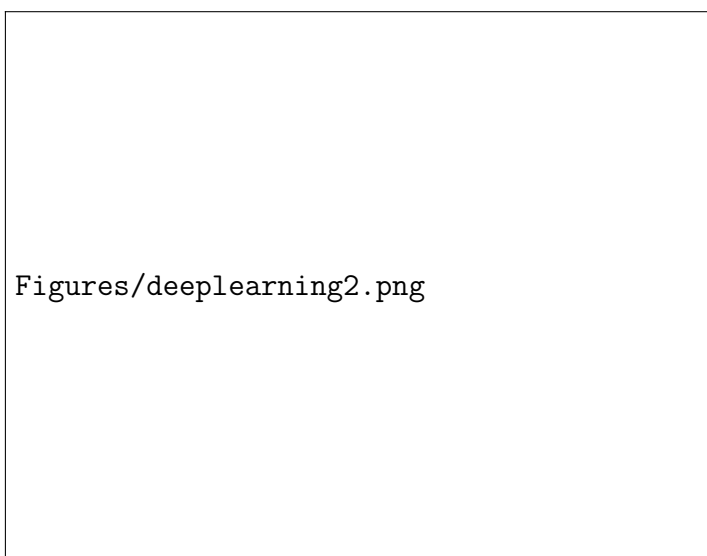
Figure 2.4: Deep learning algorithms attempt to represent the problem as a series of simpler sub-tasks. In the example shown, the task of recognizing the object in the image, is broken down into tasks of recognizing the edges in the image, then the corners in the next layer, then finally object parts in the third layer, which eventually leads to the identification of the object. Image courtesy Zeiler and Fergus, Springer [11].

levels of abstraction will enable a system to learn complex functions mapping the input to the output directly from data, without depending on features crafted explicitly. At each level, features can be added to the representation. Thus it is not necessary to deal with all the features at once, like in shallow learning, allowing a more efficient way of dealing with a more complex representation. Deep learning breaks down the learning of the features into a pipeline where first simpler features are learned, which lead to the processing of increasingly complex features.

In Figure 2.4 the task of object recognition is represented with a deep learning approach. At the first layer, we have the input pixels. This visual information is immediately recognizable to the human vision system, but is of little use to a leaning machine and is too complicated to map the object identity from a set of pixels. Learning at this stage is nearly impossible. We now break this complicated task into simpler sub-tasks, each dealing with a different layer of the representation.

- The first layer of the learning machine looks for the edges in the image by comparing pixel values of individual pixels with that of their neighbors.
- Once the edges have been represented, the next layer can represent the contours and corners as a set of edges.
- This, in turn, allows the third layer to group together these features to begin to identify object parts.
- Finally the output later can relate the parts to identify the object.

Figures/MLtypes.png

Figure 2.5: Broad classification of machine learning algorithms by the nature of their training.

2.4 Classification of Remotely Sensed Images
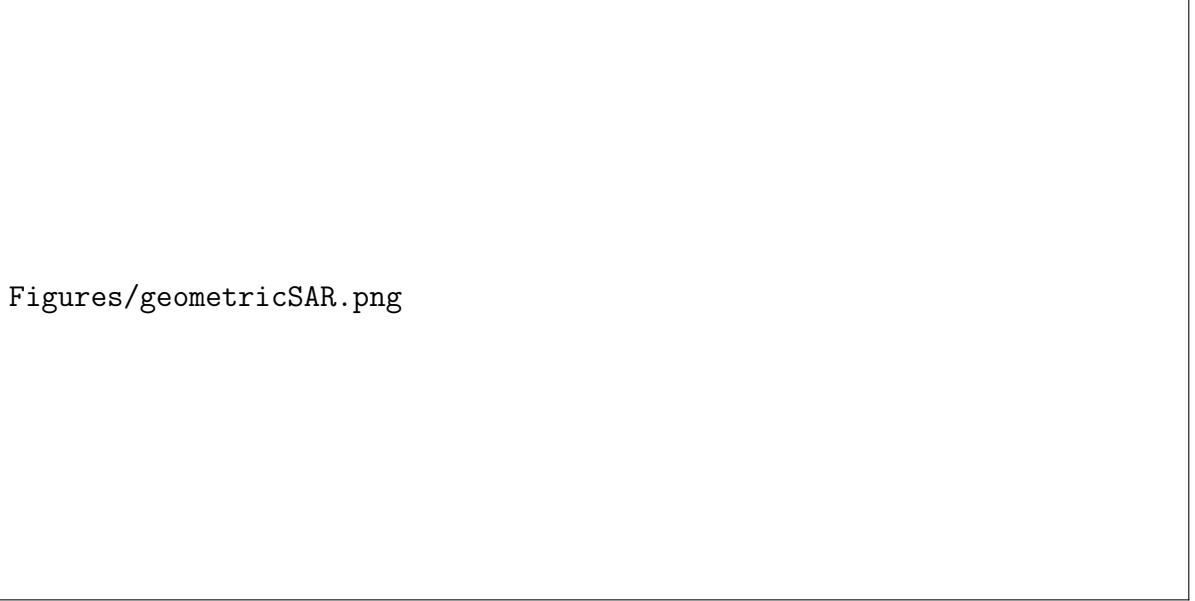
Classification of remote sensing images involves the grouping of pixels a scene or dataset into homogeneous groups, usually representing land cover classes or "themes". This categorized data can then be used to produce thematic maps of the areas represented in the image. Generally this is used to map distinct land-covers types like agriculture, urban and built up, forested, desert areas etc. The selection of the covers classes are application dependent, however certain standardized classification systems like the National Land Cover Data (NLCD) have been used operationally. The NLCD is a 21-class land cover classification scheme applied developed by the USGS Land Cover Institute for use in conjunction with Landsat Thematic Mapper data [12].

Usually multi-spectral data, like that from the Landsat mission, is used to perform classification. The spectral pattern present for each pixel is used to discriminate the classes [13]. These datasets allow for fairly accurate classification of large areas of the Earth's surface. Data from the advanced very-high-resolution radiometer sensor (AVRIS) sensor aboard National Oceanic and Atmospheric Administration's (NOAA) operational series of meteorological satellites was used by [14] to classify land-cover types and monitor changes in vegetation over the continent of Africa. At larger scales, [15] generated a global land-cover classification dataset at a spatial resolution of 8-km using space-borne multi-spectral imagining systems and a decision tree classifier. A higher resolution dataset was derived from the Moderate Resolution Imaging Spectroradiometer (MODIS) sensor on-board Terra by [16]. They generated a global land-cover map with a spatial resolution of 1-Km and using primarily the International Geo-sphere-Biosphere Programme Data and Information System (IGBP-DIS) classification scheme [17].

Data from several medium resolution multi-spectral sensors like AVRIS, MODIS and Landsat TM have been used successfully for such applications due to their continuous global coverage, data availability and high spectral information content. However these sensors do have certain restrictions. They can not image during the night as they are dependent on the incident solar radiation for imaging. Cloud cover can also affect the quality of images as these sensors are unable to image though clouds. This is especially more of a problem while imaging areas covered with snow, as it can become difficult to differentiate the white snow covered land area from the white clouds.

To overcome these problems, increasingly Synthetic Aperture Radar (SAR) data is being used for thematic map generation, especially over the cryosphere [18]. Certain applications also places temporal restrictions on the generation of thematic maps. For example, to generate ice-breg warning maps, the data can be acquired in a limited window and rapid and timely updates of these maps becomes critical [19]. In such cases, it is also desirable to be able to produce these maps, operationally with a high degree of automation.

Polarimetric SAR (PolSAR) imaging is an advancement made on single polarization SAR imaging. The polarization of the incident wave is controlled so that it is restricted

Figure 2.6: Geometric distortions in SAR imaging due to the side looking geometry. Image courtesy VanZyl, Wiley [1].

alternatively to either H or V transmission. Similarly, reception is performed alternativly 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 possible with single or dual polarization imaging, and consequently, this can lead to better classification [20].

2.5 Challenges in SAR Image Classification

SAR imaging introduces its own set of challenges to the classification problem. Being an active sensor, it must supply the illuminating energy, i.e. radar pulses, which necessitates that a power source be tethered to the sensor. Satellite sensors are usually powered by a battery which is rechargeable by solar energy. This limits the sensors continuous operation capability. Unlike optical sensors, which can be operated continuously, a typical modern spaceborne SAR sensor can only be operated for 11 minutes per 90 minute orbit [21]. This makes it difficult to generate the updated global mosaics that are generated using multi-spectral sensors [22]. Additional, in PolSAR mode, because each pixel is imaged twice on two subsequent pulses, the effective swath width of the imaged strip is half as compared to a single polarization SAR system. This somewhat reduces the feasibility of PolSAR in generating thematic maps of large areas.

Unlike optical images, which are acquired with the sensor facing towards the nadir, SAR images are acquired in side look geometry [23]. RADAR images are formed by fundamentally from the echoes returned from targets, thus a RADAR that faces vertically

Figures/shadow.png
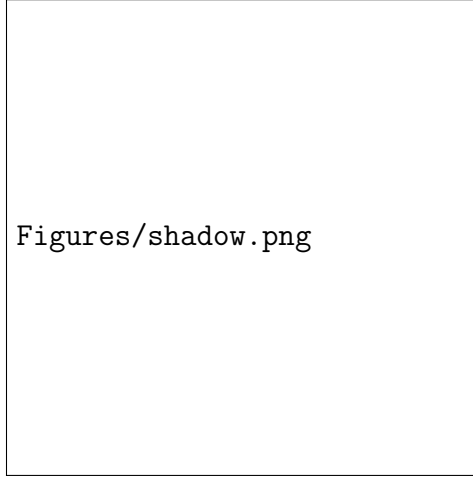
Figure 2.7: Crop of a Pauli RGB image generated from a UAVSAR dataset showing layover (bright) and shadow (dark) areas.

down would not be able to distinguish echos from equidistant targets that lie on either side of the sensor. In non imaging radars with a narrow beam-width, like sounders, this behavior is desirable, but it makes imaging impossible. The side looking geometry introduces geometric artifacts like layover, foreshadow and shadow effects. This causes elevations in the imaged terrain (buildings, hills and mountains) to suffer from certain distortions in the produced image [1, 21] as shown in 2.6. Along with the shape distortion, foreshortening also manifests as an area of bright RADAR return, which can cause the land-cover type in these areas to mis-classified. Similarly, shadow regions appear dark in the image and contain no echo information, making classification difficult as shown in Figure 2.7

Apart from the problems related to the physical constraints of the images, like other coherent imaging techniques, suffers from the problem of speckle [24]. Speckle noise can complicate the task of classification of SAR images. Further, the speckle pattern is non often non-Gaussian in distribution, thus filtering and classification techniques that have been developed for Gaussian optical images can not be directly applied to SAR images. Various non Gaussian models have been proposed to represent SAR data [25]. To represent PolSAR data multivariate K distribution based models have been proposed [26], [27]. This was further extended by [28] who used the G distribution to model the speckle pattern. Non Gaussian model based classification techniques have been applied by [29] to segregate glacier ice types to track changes in the glacier. Various statistical parameters which can be exploited in conjecture with machine learning techniques to alleviate some of the challenges of the classification of PolSAR images.

Figures/lee1.png

(a)

Figures/lee2.png

(b)

Figures/lee3.png

(c)
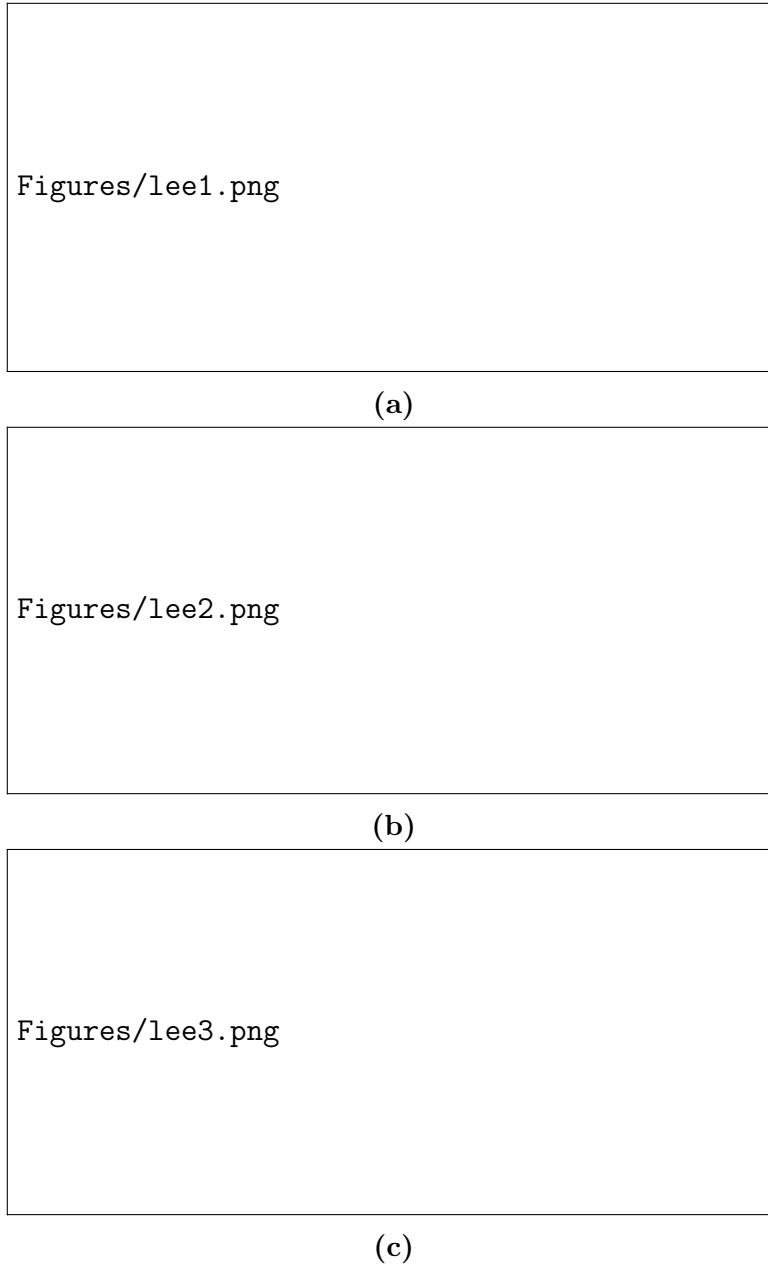
Figure 2.8: SAR speckle distribution are plotted for intensity and amplitude images of different looks. (2.8a) 1 Look intensity SAR image with it's histogram (exponential distribution). (2.8b) 1 Look amplitude SAR image (same area as before) with it's histogram (Rayleigh distribution) (2.8c) 4 Look amplitude SAR image with it's histogram (chi distribution). Figure courtesy Lee and Pottier [23, 104]

15

2.6 Generative and Discriminative Classifiers

Generative classifiers learn a model of the joint probability, $p(x|y)$, where x represents the inputs and y, the labels. They make their predictions according to the $p(x|y)$ computed using the Bayes rule and then selecting the label y with the highest probability. Discriminative classifiers model the posterior $p(y|x)$ directly or learn from a direct map from inputs x to class labels. It is more attractive to use discriminative classifiers are over generative ones. According to [30], the classification problem should be solved directly and not as an intermediate step such as modeling of the probability $p(X|y)$. An experimental comparison of this can be found in [31].

2.7 Supervised Urban Classification

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 repetivity. This has opened new avenues for the monitoring of dynamic Earth processes. Long-term Earth observation missions have ensured data continuity allowing the characterization of anthropocentric processes and changes. Spaceborne optical sensors can deliver continuous global coverage, data availability, and high spectral information content. However, these sensors can not image during the night or inclement weather as they are dependent on the incident solar radiation for imaging.

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 [32]. 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 single polarization SAR imaging [33]. 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 [20]. However, it introduces unique challenges in the characterization of urban areas, which is increasingly becoming important for systematic monitoring.

In the past decades, there has been a rapid inflation of population and the consequent expansion of urban areas. A high concentration of human population lives in urban areas, and it becomes important to monitor the growth and expansion of these regions. In the case of natural calamities like floods, earthquakes, and tsunamis, it is important to derive information at the earliest to guide mitigation actions. In such cases, the availability of up-to-date classification maps of urban areas, which can be referenced

against post-event data, is crucial. The accurate classification of urban areas is the first step towards operational implementation of these applications. For tropical regions, many high population urban centers are shrouded by cloud cover for months at a time during the monsoon season. During this time, these areas are also the most susceptible to flooding. Due to the dense cloud cover, it becomes difficult to perform routine monitoring operations during this period using optical sensors. However, radar data are relatively unaffected by weather phenomena and can image on-demand irrespectively of conditions or illumination providing crucial information in these scenarios.

SAR imaging introduces its own set of challenges to the classification problem. SAR images are acquired in side look geometry, which introduces geometric artifacts like the layover, foreshadow and shadow effects. This causes elevations in the imaged terrain (buildings, hills, and mountains) to suffer from certain distortions in the produced image making classification difficult. Moreover, SAR being a coherent imaging technique, suffers from the problem of speckle which further complicates the task of classification of SAR images. The speckle pattern is often non-Gaussian in distribution [25] making its suppression nontrivial.

Thus filtering and classification techniques that have been developed for Gaussian optical images cannot be directly applied to SAR images. Various non-Gaussian models have been proposed to represent SAR data [25]. To represent PolSAR data multivariate K distribution based models have been proposed [26], [27]. This was further extended by [28] who used the G^0 distribution to model the speckle pattern. Non-Gaussian model-based classification techniques have been applied by [29] to segregate glacier ice types to track changes in the glacier.

The first attempt at the classification of PolSAR images was made in [34]. 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 [35], where a feed-forward 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 [36] were made.

A popular classification scheme was proposed in [37], 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 [38]. An unsupervised scheme also based on the Wishart distribution was proposed in [39]. This has been extended by using several techniques for urban classification. In [40] the Chernoff and in [41] the Jeffries-Matsushita distance 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 non-parametric classification

scheme.

Polarimetric scattering power decomposition parameters can serve as low-level features in classification [42]. In [43] information theory is used to select a subset of the Touzi scattering decomposition parameters, which are then classified using a Support Vector Machine. A multi-decomposition parameter approach was adopted in [44], which used an SVM and a rule-based approach to urban land-cover classification. Both supervised [45] and unsupervised [39] schemes have been proposed that 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 [46]. An alternative strategy for urban classification is to use the textural information in the SAR images [47, 48]. However, in texture based schemes, the rich scattering information available from polarimetry is often not considered. In the classification of urban structures is performed by using the Fisher distribution with Markovian segmentation for identification of ground, buildings, and vegetation. Other texture based schemes have also proved useful for urban segmentation [48].

Interest in neural networks was renewed with improvement in technology. In [49] 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 land-cover and soil type mapping capabilities using a probabilistic neural network was reported in [50] using L-Band space-borne data. Multi-temporal/multi-frequency datasets were exploited using a neural-statistical kernel in [51] for urban classification. An adaptive resonance theory neural network was shown to perform computationally efficient classification of urban areas using polarimetric data. Hybrid approaches such as a combination of neural networks and fuzzy logic were demonstrated to have good performance for SAR classification in [52] and later specifically for identification of urban areas in [53]. A system for 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 [54]. Another multi-temporal approach using single-channel SAR datasets to monitor urban land cover in major Italian cities is proposed in [55].

Hinton et. al. introduced the concept of deep learning in [4], which found successful application in multiple areas [56, 57, 58, 59] 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 improved performance in several applications [60, 61, 62]. Multiple natural land cover classes were identified from a hyperspectral dataset in [63], while objects detection was performed in optical datasets in [60] using deep learning techniques. Additional information is often added to the classification scheme from auxiliary data sources like optical and multi-spectral areas [64, 65]. Deep learning techniques are quickly gaining popularity to perform multi-sensor data fusion as well [66, 67]. A deep learning based approach can simplify feature and classifier design, and allow the utilization of augmented datasets. This can help mitigate

18

some of the challenges faced in the classification of urban areas from PolSAR data.

The classification of urban areas in PolSAR data-sets is especially challenging, because of the problems related to the 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 [68, 69]. This makes such areas hard to distinguish from natural targets like forest and vegetation. This represents a major challenge for the classification algorithm. In [70] 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 [71] to compensate for the over-estimation of volume in rotated urban areas. In [72] a link is established between eigen-decomposition method and model-based decomposition methods to also solve the overestimation of the volume scattering power problem. Another decomposition technique proposed in [73] uses a mapping to create an improved representation and further help improve target characterization.

Various statistical parameters and understanding of the physics of scattering can be exploited in conjecture with appropriate machine learning techniques to alleviate some of these difficulties. In addition to the challenges of the sensing medium itself, the data volumes that will potentially be generated from current and future SAR programs, viz. NISAR and Sentinel, pose additional predicaments. These missions are being designed with systematic global coverage capability that will lead to generation of a spatially and temporally dense collection of data. While this can lead to better understanding of of the Earth, leveraging this information is an additional challenge. Manual analysis of this kind of data-sets are infeasible and automated of semi-automated machine learning techniques which have a relatively low computation time are instrumental in extracting relevant information.

Alternatively, some approaches utilize a separate class for identifying rotated targets and then merge them with the non-rotated class as a post-classification 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.

This might not always be available for a given scene and changes if the flight line of the platform is varied. The generalization ability and autonomy of such an approach is therefore low. Sophisticated models have been developed to account for the rotation problem. While such approaches produce accurate results, they are computationally intensive making them infeasible for high data volumes. In this work, 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 shifts are observed in polarimetric SAR data which are induced by terrain slopes and complex scattering in urban areas. In turn, polarization orientation angle

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 auto-encoder. 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.

2.8 Semi-supervised Urban Change Detection

Urban Change Detection (CD) is an important part of monitoring operations and is done to study changes induced by urbanization or in response to natural disasters for relief operations.

The active imaging capabilities of a Synthetic Aperture Radar (SAR) system allows imaging in inclement weather and illumination conditions. This is highly desirable for urban mapping and change detection applications, especially in the tropical regions where the monsoons obscure urban centers from view of optical sensors for many months. It is also during this period that calamities like floods are most likely to occur making the rapid imaging ability of SAR sensors crucial. Polarimetric SAR (PolSAR) is an advancement over traditional SAR imaging that involves the simultaneous coherent transmission and reception of signals vertically and horizontally polarized in quadrature. This allows a better characterization of the target being imaged leading to improved discrimination capabilities when compared to single polarization acquisitions [74].

CD aims at identifying changes occurred on the ground between two or more acquisitions on the same geographical area at different time instants. The complexity of collecting ground reference samples makes unsupervised approaches the most relevant in the field. The unsupervised change detection procedure can be generalized to consist of three distinct steps: Image pre-processing, Change Index (CI) generation and subsequent thresholding of the CI. Pre-processing includes accurate registration of multi-temporal images which is critical for the accuracy of the technique as a whole. Significant progress has been made in this direction for SAR and sub-pixel registration accuracy is common [75]. For SAR data log-ratio CI is the most commonly exploited [76]. The CI is usually thresholded to form the final binary or multi-class map based on physical properties of the change index or its statistics [77]. PolSAR data have not been widely exploited for change detection applications, but with greater availability of data they are becoming

increasingly important. A framework for joint analysis of the PolSAR features was explored for multiple change detection [78]. Approaches based on PolSAR descriptors [79] and distribution test statistics [80] have also shown to be effective.

DNNs have previously been used for single polarimetric SAR images [81] only. The proposed approach takes an input, a pair of PolSAR images and learns a representation that maximizes the separation of changed areas from unchanged ones in the multi-temporal stacked feature vectors (manifold) domain. This representation is exploited by a weakly supervised MLP technique to generate a change map.

2.9 Multi-frequency Crop Classification

PolSAR is an advanced imaging technique which coherently transmits and receives radar pulses in quadrature polarization. The multi-frequency polarized waves are susceptible to different structure, orientation and dielectric properties of the medium. Various PolSAR systems (viz. AIRSAR, EMISAR, and F-SAR) are capable of simultaneous multi-frequency full PolSAR data acquisition allowing for improved target characterization [82].

The use of single-frequency and single-polarization SAR data has limited success in land cover classification, even when temporal changes in the scattering are exploited, especially for agricultural applications [83]. Techniques like hierarchical segmentation [84] and sketch maps in conjunction with adaptive Markov random fields [85] have demonstrated to obtain an accurate classification for single frequency PolSAR data by leveraging contextual information while preserving edges. Early studies [86, 87] have evaluated the effectiveness of multi-frequency PolSAR data acquisition for crop identification. Furthermore, an exhaustive comparison on the use of multi-frequency single, dual and fully polarimetric acquisition was performed in [38], with the assessment of quantitative classification accuracy. It was found that the integration of polarimetric information from multiple bands leads to an improved classification performance. Different strategies have been adopted to combine information from multi-frequency observations. An early approach for multi-frequency PolSAR classification is described in [82]. It uses a dynamic neural network to combine the information from C-, L- and P-bands for classification. In [88], a wavelet-based textural feature set is derived, which is subsequently classified by a minimum distance classifier. However, it is indicated that a more sophisticated neural network technique might have a better performance. Parameters derived from the decomposition of multi-temporal L-band PolSAR data combined with single polarization C-band data was used in [83] for crop classification. An unsupervised classification technique using dual-frequency PolSAR datasets is introduced in [89]. A 6×6 polarimetric coherency matrix is constructed to combine the multi-frequency information. The methodology uses an iterative algorithm based on a complex Wishart density function to classify the multi-frequency data. Although it is effective for two frequencies, it is challenging to extend this technique to combine three or more bands, since the sta-

tistical parameter estimation equations must be derived. In [90], a matrix reversible transformation approach is developed to combine multi-frequency PolSAR datasets into a representation with a simplified statistical description for classification. However, the approach requires extensive *a-priori* knowledge of the datasets leading to a redundant representation. In PolSAR data analysis, the second order statistics (coherency and co-variance) can be expressed as a Sum of Kronecker Products (SKP) of two matrices [91]. In this, one matrix accounts for the scattering mechanism, while the other character-izes its spatial structures. SKP has been explored for the integration of multi-baseline PolInSAR datasets.

The combination of multi-frequency information however causes a simultaneous increase in dimensionality of the feature set. In classical machine learning, an increase in di-mensionality causes increased complexity for the learning technique. This is undesirable for efficient training which is usually referred to as the "curse of dimensionality" [30]. To this end, dimensionality reduction techniques like the principal component analysis, independent component analysis and non-negative matrix factorization are applied as a pre-processing step before training the learning algorithm which results in a loss of information. However, by exploiting techniques like the Deep Neural Network (DNN), it is possible to incorporate the information without the dimensionality reduction step [92].

The combination of multi-frequency information causes a simultaneous increase in the dimensionality of the feature set. A classification task is often easier to solve in a higher dimensional space. Upcoming space-borne missions like [93] are planned to have multi-frequency systematic PolSAR data collection capabilities, making it important to develop frameworks capable of leveraging the data.

2.10 Snow-Cover Classification

Snow cover variation is a sensitive indicator of environmental change. Therefore, snow cover mapping is a critical exercise which is supposed to be performed at regular intervals. Observations at periodic intervals require a sensor independent of time and weather. Microwave remote sensing is a reliable tool for the timely observation of spatial and temporal variability of snow-pack characteristics.

Approaches like [94] proposed multi-sensor constellations for quantifying snow-cover, especially in the sub-polar maritime regions. They highlighted the problems faced in mapping the ephemeral nature of snow in such regions, in light of lack of access to imagery on occasion due to cloud cover. Although, a multi-sensor approach can be suf-ficient for snow-mapping, it is usually economically infeasible to cover continent scale areas, if the data is to be purchased. Microwave based approaches, which are unper-turbed by the presence of cloud cover, are more practical and efficient for the task. In [95] a seven-channel microwave radiometer was demonstrated for snow-cover map-ping. Such a passive sensor is able to operate continuously, and with its large footprint,

is able to cover large areas of the Earth's surface. At the same time, due to the nature of microwave radiation it is not affected by cloud cover. However, the disadvantage is that the resolution is coarse. Active SAR sensors have a much higher spatial resolution, while also being unaffected by cloud cover [96]. Although active sensors have a smaller coverage than passive ones, they have far increased ability to mapped details. SAR is a side-looking imaging technique, which leads to geometric distortions like layover and shadow. These artifacts can be mitigated by using a hybrid approaches that merges optical and SAR data [97]. Advanced SAR techniques like InSAR have been explored to improve classification performance, especially in challenging conditions like wet snow. In [98] the degree of coherence was used to delineate wet-snow, in areas where backscatter alone was insufficient. PolSAR techniques can also be used to recover information in such conditions. In [99] an advanced technique to estimate snow wetness has been described, which can be extended to improve the accuracy of snow-cover classification maps.

The capabilities assessment of several polarimetric parameters including a supervised classification technique to discriminate snow-covered areas was carried out over the Indian Himalayan region [100]. The proposed polarization fraction (PF) scheme for wet snow mapping, in the absence of any training samples, performed better than the rest of the classification techniques. Although quad-pol ALOS-PALSAR L-band SAR data demonstrated significant improvement, however, an analysis of higher frequencies (S- or C-band) was much desired [101].

Snow-covered areas were segregated from the snow-free/forested areas through optimal values of transmitted and received polarization states. Since the seasonal polarimetric contrast optimization was noted to be limited by significant variability in the polarimetric response of the underlying surface, the PCVE approach was proposed [102]. An adaptive wet snow mapping algorithm using ALOS-2 L-band full-polarimetric image was proposed with a supervised learning approach using quaternion neural networks (QNN) [103]. Muhuri et al. proposed two most recent algorithms for snow cover mapping over the India Himalayas using the Radarsat-2 polarimetric SAR data [104, 105].

In [104], a snow cover mapping algorithm has been proposed by analyzing the variation in the target scattering decomposition parameters [106, 23] in the winter-summer image pairs. In [105], a computationally simple yet an effective snow extent mapping method is proposed by exploiting the ratio of the seasonal variation of co-polarized (*hh-vv*) correlation coefficient and the total scattering power. The difference image is obtained by temporal (winter-summer) ratioing of this index.

Figure 3.1: Pauli composite generated from UAVSAR polarimetric data-take acquired over Oakland, CA is show along with optical data from Google Earth. Area A is oriented with the radar line of sight while area B is not.

Chapter 3Supervised Urban Classification

In this chapter, a novel methodology for classification of urban built-up areas in PolSAR images is highlighted (Section 3.1). Further, experiments are conducted to test and validate the proposed methodology (Section 3.3). Finally, the results of the experiments are discussed (Section 3.4).

3.1 Methodology

The proposed approach for pixel classification of PolSAR data is shown in Figure 3.2. 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 pre-processing step is applied to transform them into more robust mid-level features, but this expands the data volume many-fold. A deep learning network is used to extract high-level features

Figure 3.2: Framework for pixel classification of PolSAR images using a deep learning architecture.

Figure 3.3: Block diagram of the proposed approach.

which are both efficient and robust with respect to the mid-level ones. In literature both supervised structures such as CNNs [107] and unsupervised structures such as autoencoders [63], DBNs [108] are used to this end. The mid- or low-level features are given as 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 which directly produce the classification label as output at every pixel, and soft classifiers such as logistic regression that can fine-tune the whole pre-trained 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 land-cover classes, forest, bare-soil and water bodies. Segmented urban area maps can subsequently be used to characterize urban density, land-area usage, urban sprawl, and land cover 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 [109]. 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 Figure 3.1. This presents a confusion between urban areas and vegetation leading to incorrect classification. To overcome this

Figures/Trento/Method1.png

Figure 3.4: Schematic diagram of the PolSAR data pre-processing and physics based generalized urban target database generation step.

challenge, a novel synthetic rotation based approach is proposed which simulates other rotated synthetic targets based on those present in the scene. This allows the machine learning algorithm to better generalize, be more robust and successful in discriminating oriented urban areas.

The classification process consists of three stages as shown in Figure 3.3. The first stage deals with the pre-processing 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 (Section 3.1.1). The second stage is an unsupervised feature learning step, which automatically learns an optimal representation of the generalized database using a stacked auto-encoder network (Section 3.1.2). Finally, the third stage extracts optimal features from the previous stages, along with some statistical parameters and performs a supervised classification by using a multi-layer perceptron (MLP) network (Section 3.1.3).

Figures/Trento/Method2.png
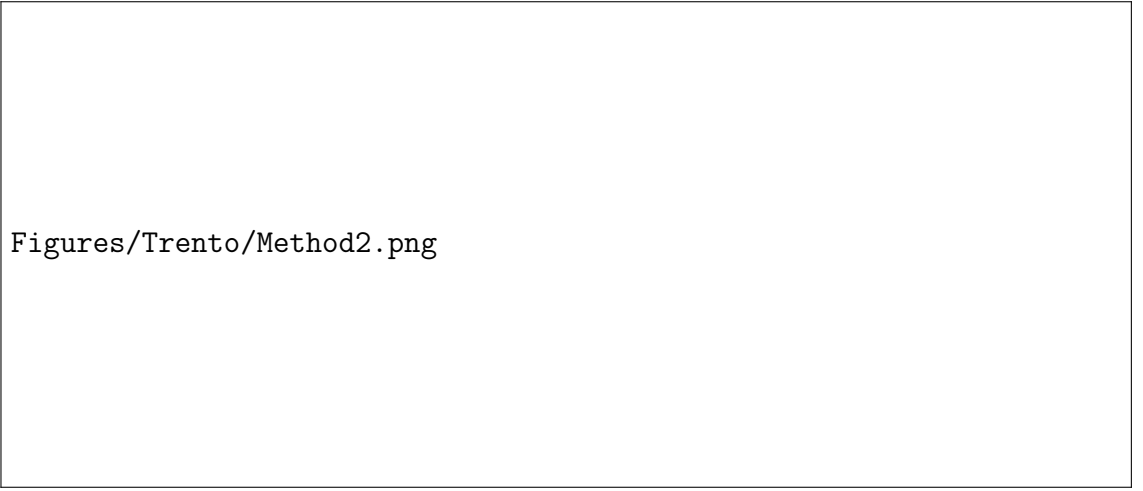
Figure 3.5: Auto-Encoder block scheme with Layer 4 extracted as the output feature-set.

3.1.1 Data Preprocessing

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 form:

$$\mathbf{S} = \begin{bmatrix} S_{hh} & S_{hv} \\ S_{vh} & S_{vv} \end{bmatrix} \tag{3.1}$$

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 [110]. Generally, in polarimetriy it is assumed that the targets are reciprocal. In the case of symmetric mono-static scattering, we can thus constrain the scattering matrix to be symmetric, *i.e.* $S_{hv} = S_{vh}$.

The corresponding $\boldsymbol{k_p}$-target vector is expressed as,

$$\boldsymbol{k_p} = \frac{1}{\sqrt{2}} \begin{bmatrix} S_{hh} + S_{vv} & S_{hh} - S_{vv} & 2S_{hv} \end{bmatrix}^T \tag{3.2}$$

where the superscript T denotes transpose. In the mono-static backscattering case, we can construct a 3×3 coherency matrix, $\mathbf{T}$, as

$$\mathbf{T} = \boldsymbol{k}.\boldsymbol{k}^{*T} \tag{3.3}$$

$$\mathbf{T} = \begin{bmatrix} t_1 & t_2 & t_3 \\ t_2^* & t_4 & t_5 \\ t_3^* & t_5^* & t_6 \end{bmatrix} \tag{3.4}$$

where

$$
\begin{aligned}
t_1 &= |k_1|^2 = \frac{1}{2}|S_{hh} + S_{vv}|^2 \\
t_2 &= k_1 k_2^* = \frac{1}{2}|(S_{hh} + S_{vv})(S_{hh} - S_{vv})^*|^2 \\
t_3 &= k_1 k_3^* = (S_{hh} + S_{vv})S_{hv}^* \\
t_4 &= |k_2|^2 = \frac{1}{2}|S_{hh} - S_{vv}|^2 \\
t_5 &= k_2 k_3^* = (S_{hh} - S_{vv})S_{hv}^* \\
t_6 &= |k_3|^2 = 2|S_{hv}|^2
\end{aligned} \tag{3.5}
$$

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,

$$\mathbf{T} = \langle [\mathbf{T}] \rangle = \frac{1}{L} \sum_{i=1}^{L} \boldsymbol{k_{p_i}}.\boldsymbol{k_{p_i}}^{*T} \tag{3.6}$$

where $\langle ... \rangle$ 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 [111] is applied to the dataset as a pre-processing step to further suppress speckle

and improve 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 Polarization Orientation Angle (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 the $\mathbf{T}$ can help to optimize and extract more polarimetric information which, in-turn helps to improve the characterization of the target scattering mechanism [42]. If we consider a target which is symmetrical 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

$$\mathbf{T}(`) = \mathbf{R}(\theta)\mathbf{T}\mathbf{R}(\theta)^{-1} \tag{3.7}$$

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

$$\mathbf{R}(\theta) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos 2\theta & \sin 2\theta \\ 0 & -\sin 2\theta & \cos 2\theta \end{bmatrix} \tag{3.8}$$

The orientation angle θ is estimated either by minimizing the t_6 element given in (3.5) [112] or by maximizing a stochastic distance measure between the t_6 and the t_4 [113].

Conversely, if the $\mathbf{T}$ matrix is rotated using unitary rotation as in (3.7), it would simulate the effect of physically rotating the target while measuring it with a polarimentric 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 show in Figure 3.4 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 a point, the gain in performance is not justifiable by the rise in computational cost. Therefore a step of 3° is selected as a trade-off 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 back-propagation 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 input for the network. However, splitting a complex number to process in real-valued neural networks leads to a sub-optimal representation of the domain of the problem [114].

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
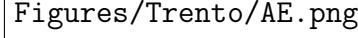
$$\mathbf{M} = \frac{1}{2} \begin{bmatrix} t_1 + t_4 + t_6 & t_2 + t_2^* & t_3 + t_3^* & -i(t_5 - t_5^*) \\ t_2 + t_2^* & t_1 + t_4 - t_6 & t_5 + t_5^* & -i(t_3 - t_3^*) \\ t_3 + t_3^* & t_5 + t_5^* & t_1 - t_4 + t_6 & i(t_2 - t_2^*) \\ -i(t_5 - t_5^*) & -i(t_3 - t_3^*) & i(t_2 - t_2^*) & -t_1 + t_4 + t_6 \end{bmatrix} \tag{3.9}$$

$\mathbf{M}$ 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 properties of a target [115]. This makes it a reasonable choice for identifying the scattering mechanism from targets. In this work, each rotated $\mathbf{T}$ matrix is thus converted to its $\mathbf{M}$ representation. $\mathbf{M}$ matrices are then collated together as a database of synthetically generated generalized urban targets to serve as input to subsequent stages of the learning algorithm.

3.1.2 Unsupervised Representation Learning using Auto-Encoders

The second stage of the proposed classification method consists of an unsupervised sparse stacked auto-encoder (AE) which is represented in Figure 3.5. An AE is a feed forward, fully connected, non-recurrent neural network with an input layer, an output layer and one or more hidden layers as shown in Figure 3.6. 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 back-propagated and minimized over multiple iterations.

The collated synthetic urban target database from the previous stage is made available to the AE as input. The training is carried out in an unsupervised manner. At the end of the training phase, the weights 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 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

Figures/Trento/AE.png

Figure 3.6: Structure of a general auto-encoder with fully connected hidden layers.

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 ith 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, non-linear, non-parametric 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 input $\boldsymbol{X}_i^n = X_1, X_2...X_n$ and an intercept term b, and produces the output, called hypothesis, $h_{W,b}(X) = f(\boldsymbol{W}^\mathsf{T}\boldsymbol{x}) = f(\sum_{i=1}^n W_i x_i + b)$, where $f : \Re \Rightarrow \Re$ is called the activation function or nonlinearity.
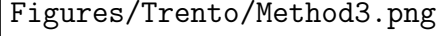
Traditionally the sigmoid $\sigma(x) = 1/(1+e^{-x})$ has been used as the activation function. It takes a real-valued number x and returns a value within $[0, 1]$, according to the magnitude of the input. In practice this has some drawbacks. The output of the sigmoid saturates at either tail of 0 or 1 and the gradient at these regions tends to zero, which causes a very small output to be back-propogated. Additionally, if the initial weights are large, the neuron can quickly become saturated during training. The sigmoid function also is not zero centered. During back-propogation if the input data to the neuron is 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 non-zero 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 non-saturating nonlinearity like the Rectified Linear Unit (ReLU) activation function [116] 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 non-saturating counterparts [117]. Since deep learning algorithms use several layers, faster training of ReLU units (as compared to sigmoid or tanh) translates to multi-fold reduction in training times [118]. Also, the computation of the output of a ReLU is simple as it does not involve 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 can not 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 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 towards zero indicates that the network is properly learning, while a stable value indicates that the learning stage is complete. The rate of adaptation of the network is determined by the set learning rate (α). α is gradually reduced as the training progress. This is done by multiplying it by a set size multiplier (Γ). After a given number of iterations α is updated as $\alpha = \alpha \times \Gamma$. 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 $\rho \to 0$ 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 dampened.

The AE structure is now a close representation of the input data, and the internal nodes

Figure 3.7: Stage 3: Final classification is done using a Multi-Layer Perceptron network with parameters extracted from the data.

of the AE can be used as features in the classifier. The weights of the representation layer $\boldsymbol{W_r}$ of the AE are used as features after completion of the training phase. They are made available as input to the subsequent stage of the method. $\boldsymbol{W_r}$ 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 inter-class separation [3].

3.1.3 Supervised Classification

The third stage of the method consists of a multi-layer percepton (MLP) fully connected feed forward neural network (Figure 3.7). The network is trained by a supervised learning algorithm that takes training data and labels as input and attempts to classify unlabeled additional unseen data-points. Here, the input to the MLP are the features $\boldsymbol{W_r}$ 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 [119] 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:

$$\alpha^0 = (\alpha_{ij}^{hh} + \alpha_{ij}^{vv} + \alpha_{ij}^{hv})/3 \tag{3.10}$$
$$\gamma^0 = (\gamma_{ij}^{hh} + \gamma_{ij}^{vv} + \gamma_{ij}^{hv})/3$$

Here pq represents the individual polarization bands hh, hv, vv 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 the γ_{ij}^0 is the scale parameter of the distribution and thus is related to the brightness of the area [120]. 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 type of targets.

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 [121]. 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.

3.2 Dataset and Study Area

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

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 $0.6\,\text{m}$ in azimuth and $1.6\,\text{m}$ in range. The look angle ranges from $25°$ to $60°$ with a range swath of $20\,\text{km}$ [122]. This is an air-borne sensor capable of acquiring very high resolution images. The data are multi-looked 3 times in range and 12 times in azimuth to bring them to a resolution of $7.2\,\text{m}$ in azimuth and $5\,\text{m}$ 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 data-set 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.

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 $3.2\,\text{m}$ in range and $2.8\,\text{m}$ in azimuth. The dataset is multi-looked 3 times in range and 2 times in azimuth to generate an image of resolution $9.6\,\text{m}$ in range and $5.6\,\text{m}$ in azimuth. The radar operates in L-Band with $\lambda = 24\,\text{cm}$. 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 [74, 42].

Figures/Map.png

(a)

Figures/IGRASS15_WORK/alpha_5_5.png　Figures/IGRASS15_WORK/gamma_5_5.png

(b)   **(c)**
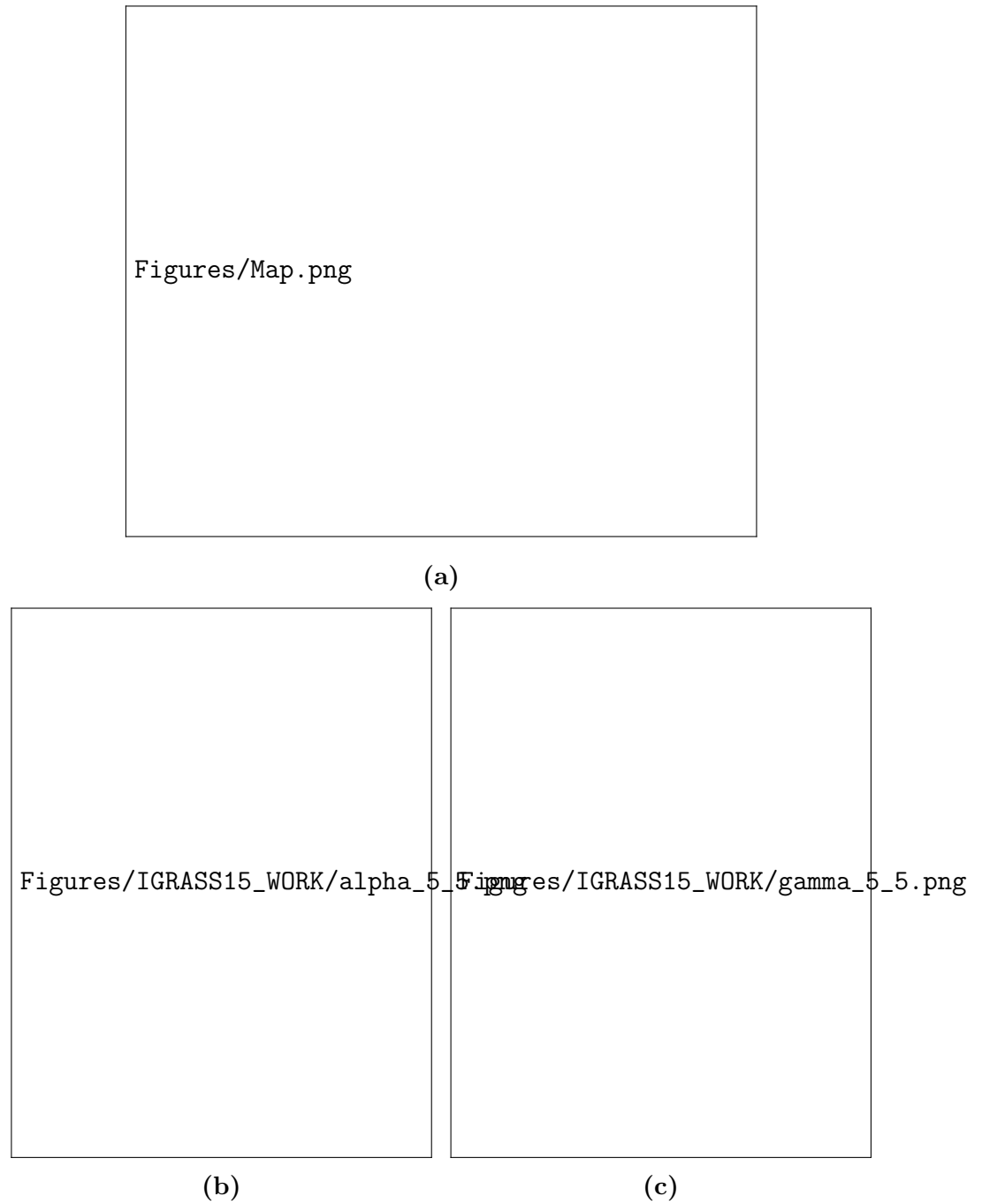
Figure 3.8: (a) A map with the dataset footprint shown with the (b) α^0 and (c) γ^0 images extracted from a UAVSAR L-Band dataset collected over Oakland. These serve as textural information to the classifier.

Figures/layers.png
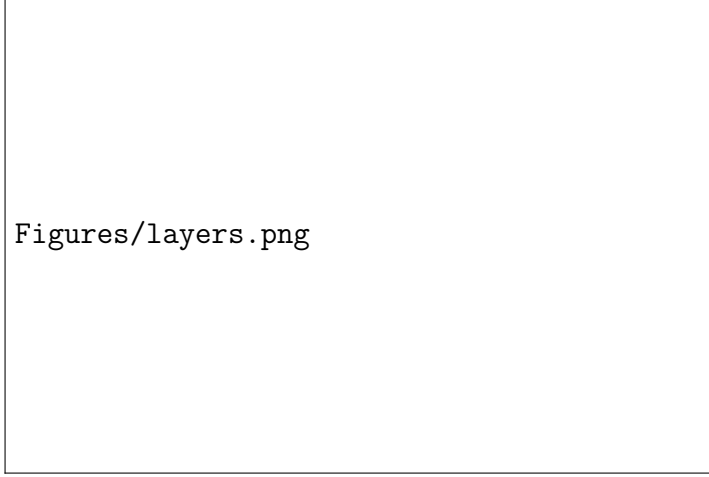
Figure 3.9: Effect of layer depth on training of AE in terms of cross entropy error versus the number of layers.

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 appear to have undergone specular reflection, and consequently have very low return power levels. The average backscattered power in these areas for the co-polarized channels, $\sigma_{hh}^0 \approx -16\text{dB}$, and $\sigma_{vv}^0 \approx -15\text{dB}$. This is comparable to that of still water, which is about $\sigma_{hh}^0 \approx \sigma_{vv}^0 \approx -19\text{dB}$ and is much lower than the power returned from forest areas, where $\sigma_{hh}^0 \approx 6\text{dB}$ and $\sigma_{vv}^0 \approx 7\text{dB}$. 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.

3.3 Experimental Set-Up

'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 which 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 [123]. In the context of this 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 which 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 tends 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 pre-processing, a refined Lee Filter with a window size of 3×3 is applied to the dataset after the application of appropriate multi-looking for the suppression of speckle. The areas under consideration have relatively flat topography, and hence no radiometric terrain correction is applied. Appropriate compensation [124] 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 many-fold without a corresponding appreciable increase in classification accuracy.

The network used in stage 2 of the proposed method is a six layer stacked sparse auto-encoder. To determine the optimal number of layers the dataset was trained on different AE topologies of increasing depth, but with the same hyper-parameters ($\alpha = 0.01, \Gamma = 0.1, \mu = 0.95$). The cross-entropy error at the end of training was analyzed to compare the architectures. From the plot in Figure 3.9 it is seen that a 6-layer auto-encoder is the most suitable for this dataset. From input X to output X', the network has 1024, 512, 256 nodes in layers 1, 2 and 3, respectively, forming the encoder, and 512, 1024 nodes in layer 5 and 6 forming the decoder. The central 64 node layer 4 contains the learned representation. The non-linearity 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 non-responsive over many iterations. The sparsity hyper-parameter 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 [121]. The network is iterated $200,000$ times with a batch size of 1000. This ensures that the complete data set is passed multiple times though the auto-encoder 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 which do not show dihedral (red) scattering in the Pauli

RGB image.

The weights from the representational layer of the previous stage are given as input to the MLP along with and two statistical texture parameters, α_0 and γ_0 (Figure 3.8). A window of 9×9 has been used for 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 data-set 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 classification performance is evaluated in Section 3.4. 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 $9°$ degrees is given as input. The activation pattern of the neural network in the case of permuted data is compared with the pattern when no rotation is applied (Section 3.4). Third, classification and performance analysis is conducted on the ALOS-2 dataset (Section 3.4). 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 (Section 3.4). Fifth, the results of the proposed method are compared with other classification techniques (Section 3.4).

3.4 Results and Discussion

The impact of textural parameters and synthetic rotation for target generation on the performance of the network are examined on a subset of the UAVSAR dataset. The analysis of classification performance using the proposed method was conducted on the full ALOS-2 scene along with the analysis of the learned representation. Finally, the performance of the proposed technique was compared with state of the art classifiers for PolSAR data.

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 [125, 126]. Although

(a) (b)
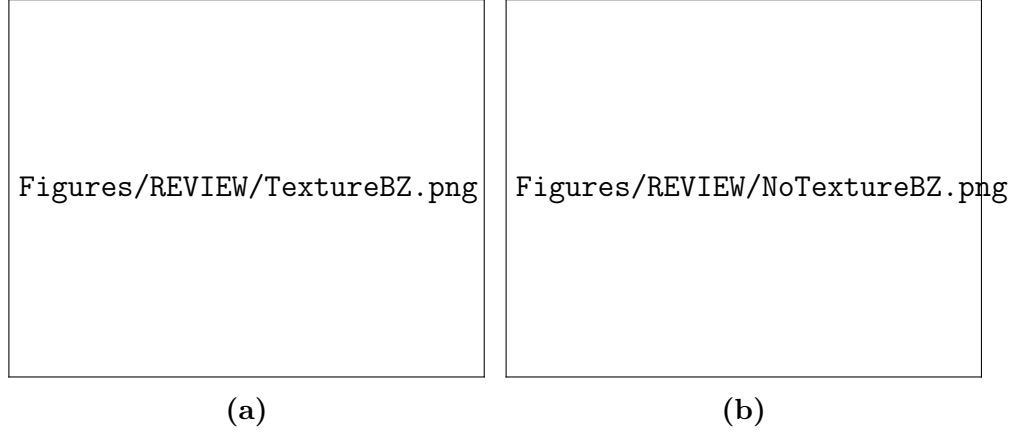
Figure 3.10: Classified map derived from an UAVSAR full polarimetric L-Band dataset (a) with and (b) without the use of α^0 and γ^0 as textural features.

the inclusion of the textural parameter only slightly improves the overall classification accuracy its inclusion can help discern low-density oriented urban targets. In Figure 3.10 the classification map of the UAVSAR data-set 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 classification accuracy due to the inclusion of textural parameters is small ($\sim 3\%$). However, for areas of low density sub-urban housing interleaved with vegetation and streets, an improvement is observed (black box in Figure 3.10) because of benefit from the neighborhood information.

Impact of Synthetic Rotation

A histogram of measured orientation angles in the UAVSAR dataset is shown in Figure 3.11. It can be seen that most of the targets 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 [112]. To visualize the action of synthetic rotation, $\mathbf{T}$ is given as input to the trained AE stage after rotation by $0°$, $5°$ and $9°$ degrees 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 Figure 3.12. 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.

Figures/Trento/orientation_RS2.png
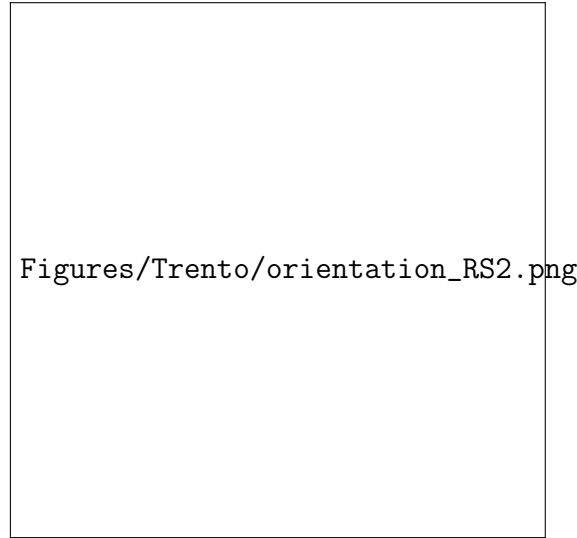
Figure 3.11: Distribution of the estimated Orientation angles of the region imaged in the UAVSAR L-Band dataset acquired over Oakland, USA.

Figures/Rotation_FS/0deg_decode3_11.png   Figures/Rotation_FS/5deg_decode3_11.png   Figures/Rotation_FS/9deg_decode3_11.png

        (a)   **(b)**   **(c)**

Figures/Rotation_FS/JET.png
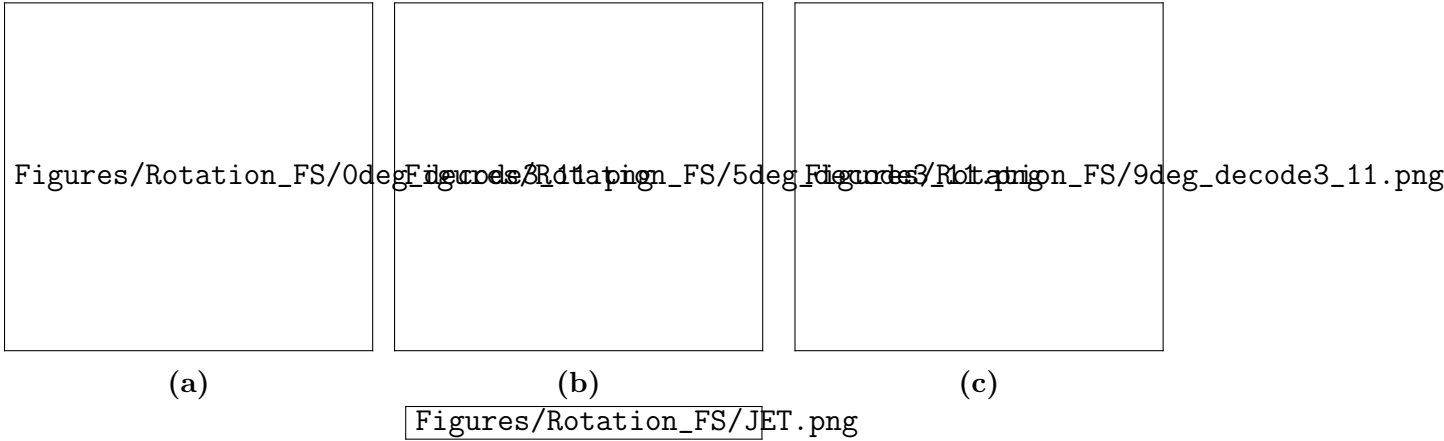
Figure 3.12: A section of the UAVSAR image is trained using the auto-encoder network and the extracted features are plotted when input corresponds to a synthetic rotation of (a) 0° (b) 5° (c) 9° respectively. The actual value of the output of the ReLU node has been scaled from 0 to 1 for representation as an image.

Analysis of Classification Performance

The Pauli composite image generated from the dataset is shown in Figure 3.15 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 $\sim 10\%$ of the total pixels in the image. These are randomly

Figures/ALOS2_SF_3Class/Output_AL2_500k_alt.png

Figures/RS2F3F_3Cl/RS2F_3Cl/RS2FFe_3Cl/RS2GSF_3Class/Red.png
□ = Water □ = Forest □ = Urban

Figure 3.13: Classified map of San Francisco derived from an ALOS-2 full polarimeteric L-Band dataset.

Figures/ALOS2_SF_3Class/ALOS2_SF_3Class/ALOS2_SF_3Class/ALOS2_SF_3Class/ALOS2_SF_3Class

Figures/ALOS2_SF_3Class/ALOS2_SF_3Class/ALOS2_SF_3Class/ALOS2_SF_3Class/ALOS2_SF_3Class

| (a) | (b) | (c) | (d) | (e) |
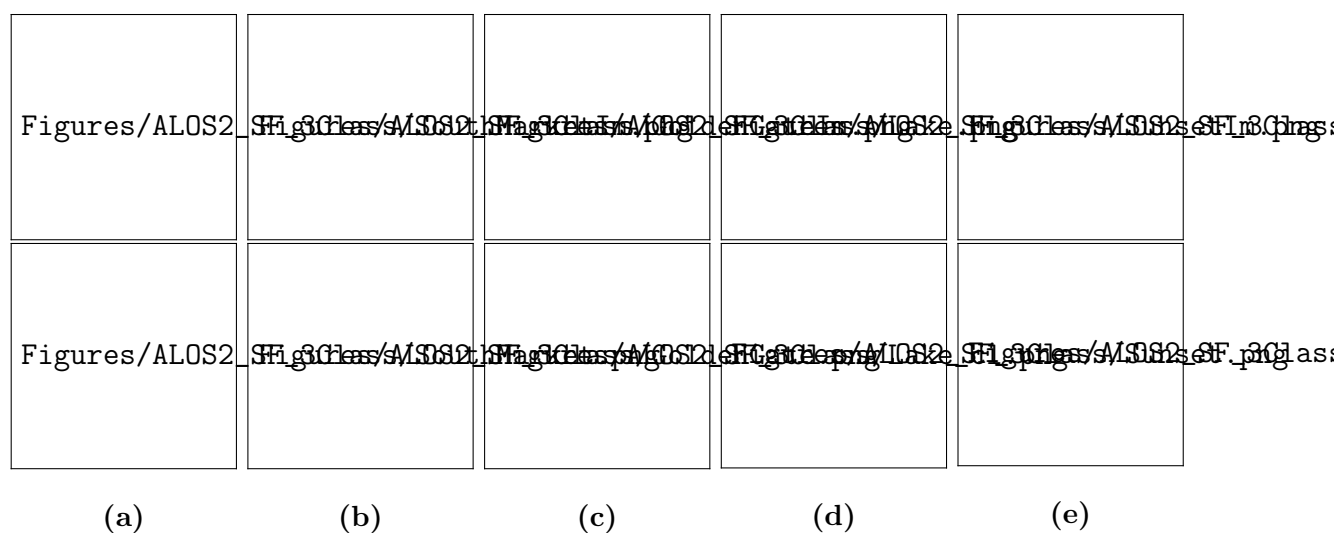
Figure 3.14: Crops of the classified map shown alongside corresponding aerial imagery courtesy ESRI/Bing Maps.

Table 3.1: Confusion matrix and accuracy statistics for ALOS-2 San Francisco, CA dataset using the proposed supervised classification scheme incorporating textural information
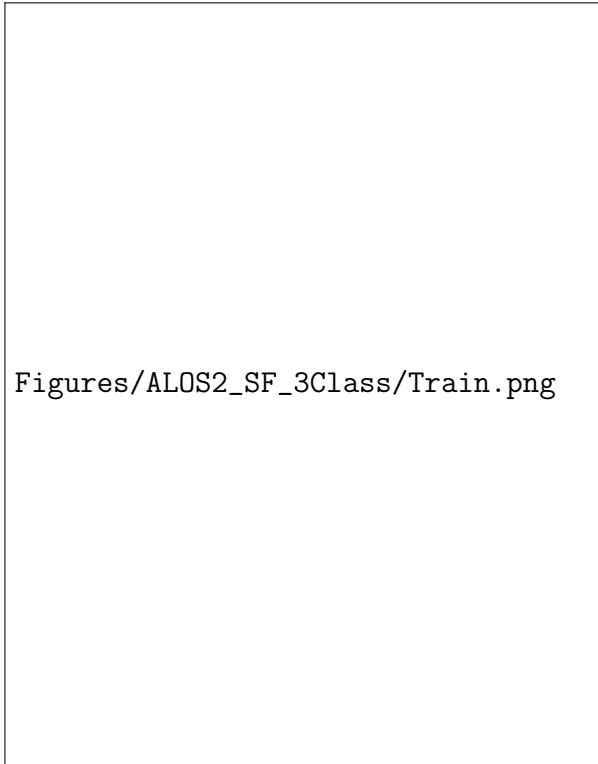
	Urban	Water	Forest	Prod. Acc.
Urban	**84.20**	0	4.39	95.45
Water	1.3	**96.86**	3.41	96.00
Forest	14.5	3.14	**92.20**	84.41
User Acc.	84.85	96.97	92.93	
OA	91.63 %			
Kappa	0.85			

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 over-fitting or memorization. To quantify the final performance of the classifier, validation pixels are employed. The confusion matrix is and is presented in Table 3.1.

The dataset is classified with an overall accuracy of 91.63% and a Kappa coefficient of 0.85. Urban areas are classified with an accuracy of 84.2%. The urban class is mostly confused with 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 3.14% due to the low backscattered power as previously highlighted.
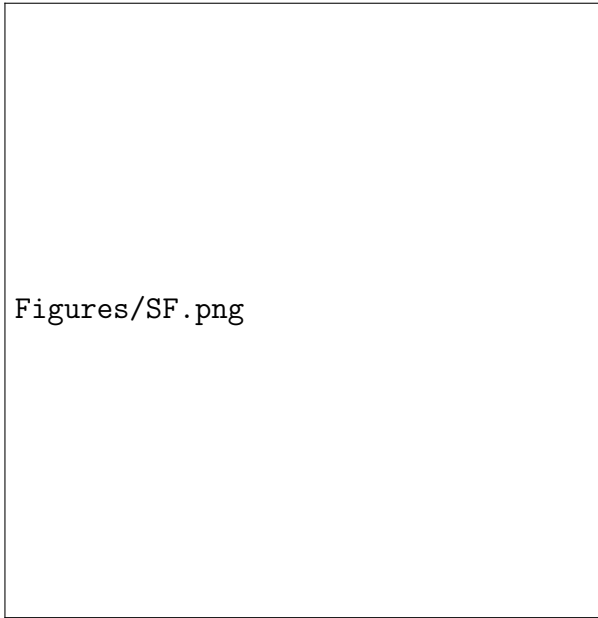
The resultant map is presented in Figure 3.13. Reference aerial photographs are given in Figure 3.16. We can see a close correspondence between the areas that appear to be urban in the optical aerial images and the classification result. Detail subsets from selected areas of the classification output along with the corresponding aerial imagery, representing the ground truth, are shown in Figure 3.14. This demonstrates the performance of the proposed algorithm on the considered dataset. The resolution of this imagery is $\sim 0.3m$ and is significantly higher than that of the classified map. However, reasonable correspondence can be seen between the two.

The subset shown in Figure 3.14a represents a dense urban area that is oriented at an angle $\sim -20°$ from the radar line of sight, marked as Area 'A1' in the image. Due to this large orientation, this region is often misclassified [127], however it is classified correctly using the proposed method. The subset shown in Figure 3.14b shows the Golden Gate bridge, marked as 'A2' and some surrounding areas. Apart from the return from the bridge itself, multi-path 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,

Figure 3.15: 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.

Figure 3.16: Aerial imagery collected over San Francisco, USA depicting the extent covered by the ALOS-2 dataset courtesy of ESRI/Bing Maps.

(a) Proposed   **(b)** SVM RBF   **(c)** SVM Polyno-mial   **(d)** Wishart with POC   **(e)** Wishart
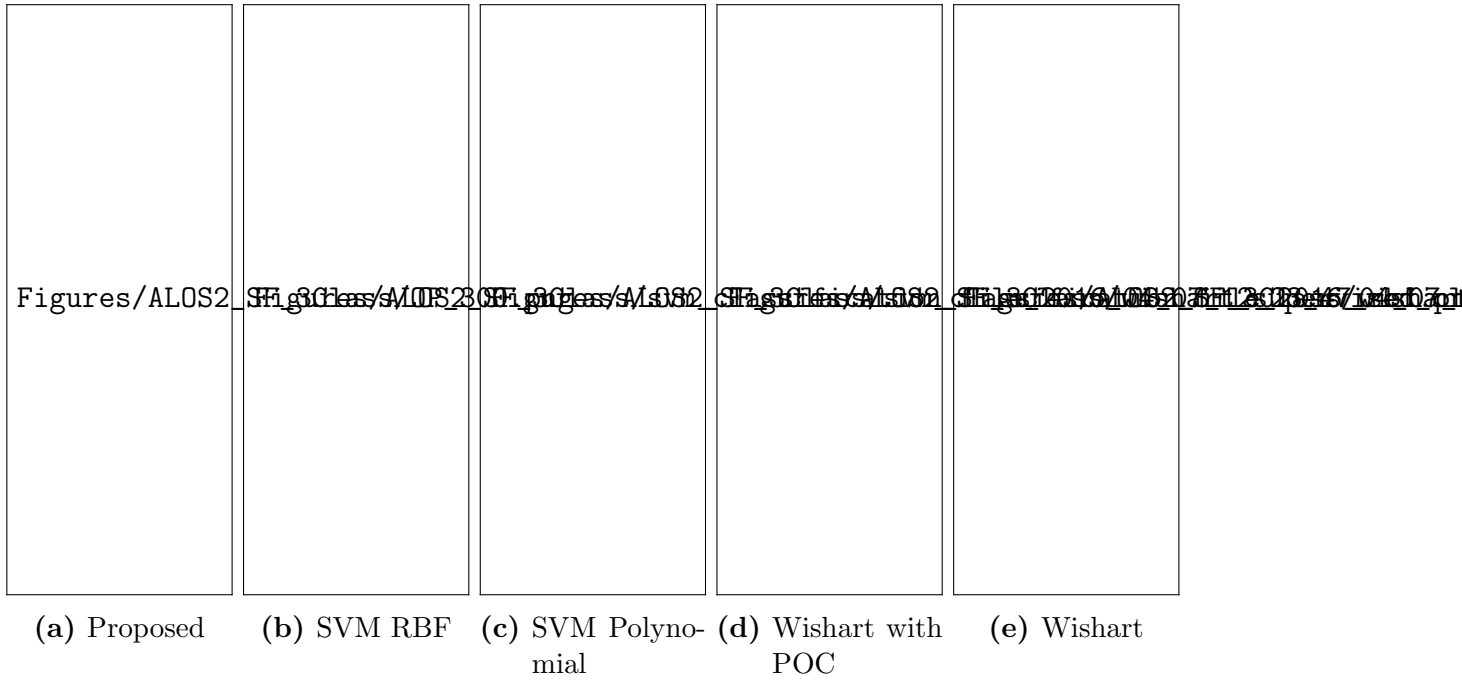
Figure 3.17: Comparison of common classification techniques with the proposed method.

which takes γ^0 as an input, to classify this pixels immediately under the bridge as forest. Figure 3.14c 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 Figure 3.14d consists of a park between two regions of high density urban areas. It can be seen that the area of vegetation is well identified. The high resolution of ALOS-2 allows the faithful classification of small features in this park as well. To further exploit high resolution datasets like ALOS-2 for classification of fine details, it is possible to avoid filtering operations in the SAR pre-processing step. Since, the proposed technique is pixel based, it does not cause a degradation in resolution. Subset shown in Figure 3.14e is of a jetty. The white arrow marks a pier structure which has interleaving perpendicular cylindrical masts. This shows a high double bounce return which does not reduce with rotation of the target about the radar line of sight. This enables the classifier to correctly identify this structure.

Analysis of the Learned Representation

To understand the information encoded in the representational layer of the AE, the response of one of the nodes it is extracted and shown in Figure 3.18 after the completion

Figures/Wr_decode3_11_g.png

Figures/Rotation_FS/GRAY.png

Figure 3.18: Extracted feature map for one of the internal nodes in the representation layer from the trained AE.

Table 3.2: 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	78.14	76.83

of the training phase. This is one of the many internal representations encoded by the network. Since the network is initialized randomly, the exact position of the internal node containing information and its encoding will change each time the network is trained. However, in general, it is observed that each internal node learns a different encoding of the data-set. The representation is sparse in nature with only small percentage of nodes containing useful information for classification. Area 'A4' consists of urban areas both perpendicular to and oriented away from the radar line of sight. Although they appear distinct in the Pauli visualization (Figure 3.15), the trained network evokes a uniformly strong response from the area. Thus the AE is able to combine the information from the synthetically rotated urban target database to form an internal representation that is independent of the 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.

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 3.2 and the results are presented in Figure 3.17. The SVM classifier was run with two kernels: a polynomial kernel (SVM-Poly) and a RBF kernel (SVM-RBF). The parameters of the RBF kernel were determined by the method described in [128]. The polynomial 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 (Figure 3.16). The cross-polarized return

47

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 Figure 3.17a, 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 kernel SVM and Wishart supervised classifier. Wishart+POC performs slightly better than SVM with the polynomial kernel.

Chapter 4 Semi-supervised Change Detection

In this chapter, a novel methodology for semi-supervised classification of changes in urban built-up areas in PolSAR images is proposed (Section 4.1). Further, experiments are conducted to test and validate this methodology (Section 4.3). Finally, the results of the experiments are discussed (Section 4.4).

4.1 Methodology

Figures/CD/BLOCK.png

Figure 4.1: Block scheme of proposed methodology.

In disaster relief and urban monitoring, typically CD analysis is applied in order to highlight the changed urban areas. On the other hand, a further analysis discriminating the unchanged urban areas from natural ones is desirable. Additionally, multi-temporal ground campaigns are often difficult and very few ground truth points may be available. The motivation behind this study is to develop a framework that can simultaneously perform accurate urban change detection and land-cover classification in a weakly supervised manner. The framework takes multi-temporal polarimetric SAR images as input and produces an output with 3 classes: changed urban areas, unchanged urban and unchanged natural areas. This kind of maps are useful for urban monitoring and rescue operations.

The block scheme of the proposed methodology is shown in Figure 4.1. Two PolSAR images acquired over the same geographic area at time instants t_0 and t_1, are pre-processed and normalized (Section 4.1.1) before being given as input to a Stacked Auto-Encoder (AE) network (Section 4.1.2). The AE is trained in an unsupervised manner to create an optimal representation, assuming that only a very small amount of ground truth is available, a label aggregation step is undertaken before classification by a Multi-layer Perceptorn (MLP) (Section 4.1.3). The use of a AE with stacked features has the advantage of preserving both change and multi-temporal polarimetric scattering information which is exploited by the final classifier.
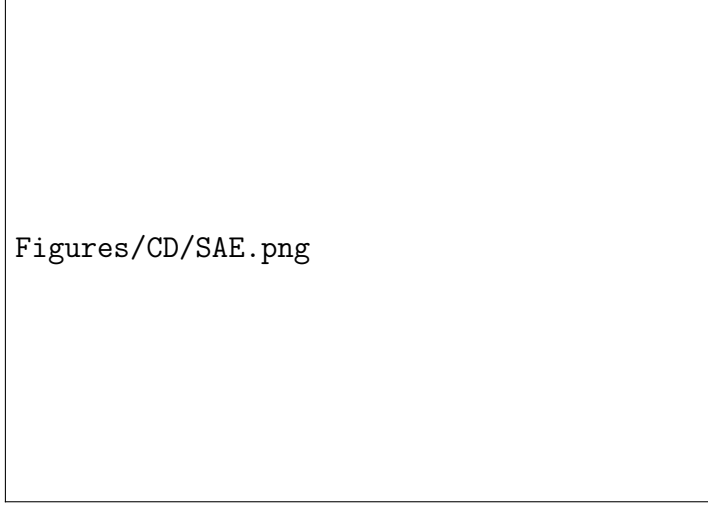
Figure 4.2: Block diagram of the unsupervised Stacked AE stage.

4.1.1 Data Preprocessing

The full PolSAR complex backscattering matrix $\mathbf{S}$ consists of four independent measurements (HH, HV, VH and VV) with the phase relations preserved and is expressed as,

$$[\mathbf{S}] = \begin{bmatrix} S_{hh} & S_{hv} \\ S_{vh} & S_{vv} \end{bmatrix} \tag{4.1}$$

This can be expressed as Mueller matrix $\mathbf{M}$ as follows,

$$\mathbf{M} = \mathbf{A}^* \left(\mathbf{S} \otimes \mathbf{S}^* \right) \mathbf{A}^{-1} \tag{4.2}$$

where,

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & -1 \\ 0 & 1 & 1 & 0 \\ 0 & j & -j & 0 \end{bmatrix} \tag{4.3}$$

Although when expressed in the backscatter convention, the matrix is named Kennaugh $\mathbf{K}$, it is common practice in literature to refer to it uniformly as $\mathbf{M}$.

The $\mathbf{M}$ representation is constructed for images at t_0 and t_1. Since SAR images are affected by speckle [111], it is desirable to perform a smoothing or suppression step to reduce false alarms. A super-pixel (SP) based filtering technique is used. Simple Linear Iterative Clustering (SLIC) algorithm [129] with a perturbation window n_w is applied on the Pauli composite of image acquired at t_0 to form super-pixels (SPs). $\mathbf{M}$ is averaged over SPs for both the t_0 and t_1 images to reduce the effect of speckle on the change detection performance, This approach yields high edge preservation while providing maximum smoothing over homogeneous areas. The dynamic range of the dataset is jointly normalized in the range $[0, 1]$ for faster convergence of the optimization solver

algorithm and for efficient training of the network.

4.1.2 Unsupervised Representation Learning using Auto-Encoders

The pre-processed $\mathbf{M}$ elements from the image pair for each corresponding pixel is stacked together to form the input vector x, which is given to stacked AE (Figure 4.2). The objective of the AE is to reconstruct output x' to match x as closely as possible. The loss function of the AE is modified to be a combination of L2 norm and Cross Entropy error given by,

$$E = \lambda_{w1} \frac{-1}{N} \sum_{n=1}^{N} [x_n \log \hat{x_n} + (1 - x_n) \log(1 - \hat{x_n})] + \lambda_{w2} \frac{1}{2N} \sum_{i=1}^{N} \|x_i^1 - x_i^2\|_2^2 \quad (4.4)$$

where $\lambda_{w1}, \lambda_{w2}$, are the loss weighting parameters, N is the total number of samples in the dataset and E is the total loss evaluated for back-propagation. This modified loss function leads to a representation that maximally separates changed from unchanged areas. Further to prevent over fitting, a weight decay strategy is applied with the regularization term $\lambda \sum_{i=0}^{|N|} (x_i - x_i')^2$, where λ is the decay parameter. Since $|N|$ is quite large for the complete dataset, the optimization is performed in smaller mini-batches of $D << |N|$. Regularization constrains the AE to be less sensitive to the input, but in minimization of the reconstruction error, it remains sensitive to variations along the manifold of high density.

After training, the representational layer z is extracted as a feature from the AE. Principle Component Analysis (PCA) is applied to transform z to a set of linearly uncorrelated variables $PC_1, PC_2...PC_n$ by projection into the eigenvector space. The information content in each PC is evaluated by ordering the set of eigenvalues and considering the corresponding top k PCs for label aggregation.

4.1.3 Weakly Supervised Classification

Ground reference labels are assumed to be available for only a small portion of the scene. This is a realistic assumption for disaster relief or monitoring efforts, where information would be available at a point of the scene where administrative efforts are concentrated. However the total number of labeled pixels is $l << N$ and hence the framework is said to be weakly supervised. The use of very few samples will lead to poor accuracies as the dataset is under-represented. Thus a label aggregation step is performed. In the PC_k space a convex-hull is fitted for each class on the labeled samples, in order to reduce the number of considered points and simplify computation. Then a minimum volume bounding ellipsoid (MVBE) is fitted on these points as illustrated in Figure 4.3.

Figure 4.3: Mechanism of MVBE fitting for labeled sample generation in the PC space.

Thinning along the radial axis of the ellipsoid is performed to exclude points that lie on the periphery of the volume, having a lower probability of belonging to the class than a more internal point. After thinning, points that lie within each ellipsoid are assumed to belong to the corresponding class and are preselected for labeling. From these, an equal number of samples L are chosen per-class using the reservoir sampling technique to preserve the distribution of each class. These samples are then to assigned the corresponding class labels, increasing the number of labeled samples available to the final classifier. A Multi-Layer Perceptron (MLP) network is used to perform the final classification with the aggregated labels. The extracted representation z is used in the classification with the aggregated labels to produce the final output map.

4.2 Dataset and Study Area

A pair of co-registered L-band ($\lambda = 0.23m$) UAVSAR images acquired over Los Angeles, California in 23-04-2009 and 11-05-2015 respectively with a resolution of $0.4m$ in azimuth and $1.66m$ in range are used in this study. The dataset is multi-looked 2×8 times to form an effective pixel-size of $3.2m$. The study-area is a dense urban area with changes occurring primarily due to the effects of urbanization.

The objective of the proposed methodology is to characterize the construction and destruction of urban areas, while simultaneously classifying unchanged urban and natural areas in the rest of the scene. The major changes, i.e. 'Changed Urban' class observed include the construction of suburban neighborhoods, industrial structures and public
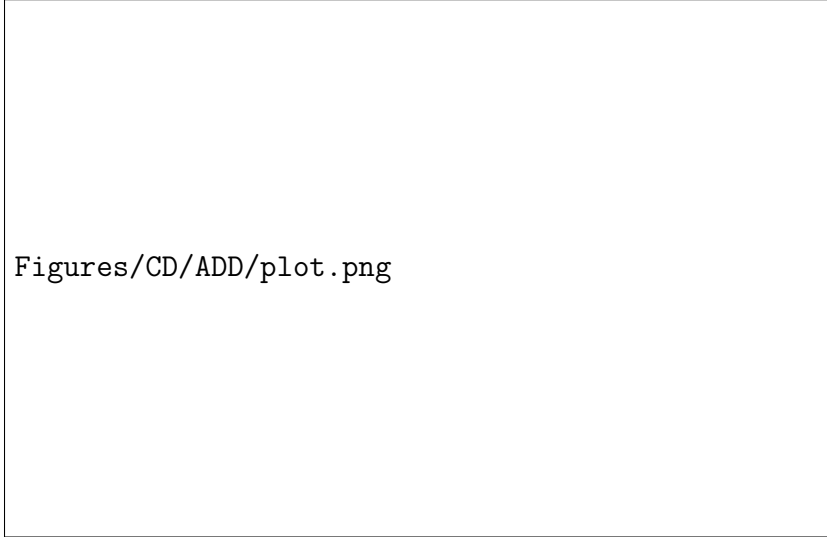
Figure 4.4: Map of all the ground reference information used for CD. As can be seen, the reference information is minimal, and constrained to a small spatial area of the scene.

infrastructure like rail-yards and airports. Natural areas of vegetation comprise of the 'Natural Unchanged' class. The urban infrastructure like houses, buildings, roads etc. that did not change is the 'Urban Unchanged' class.

4.3 Experimental Set-Up

The images are pre-processed and stacked to form the 20 dimensional input vector x. The AE is a sparse 7-layer network comprised of 512, 256, 128 nodes in the encoder, 8 in the representational layer and 128, 256, 512 in the decoder with ReLU nodes. The auto-encoder is trained in an unsupervised manner with learning rate $\alpha = 1e - 5$ at initialization and is dropped by a step factor of $\Gamma = 0.1$ every $100,000$ iterations. After training, z is extracted from the representational layer and a PCA is applied pixel wise. It is seen that $k = 3$ principal components contain most of the useful information.

In a realistic scenario, reference information about the scene would only be available in a small portion of the area. Hence training samples are drawn from the part of the scene bounded by the yellow square in Figure 4.5a. For each class 20 pixels are labeled based on prior knowledge of the region and optical imagery as shown in Figure 4.4. A total of $l = 60$ training samples are used, and $l << N = 9e6$ making the problem weakly supervised. A label aggregation step is performed by the ellipsoidal fitting technique described above. Reservoir sampling is performed to select $L = 300$ labeled pixels per class from the points bounded by each ellipsoid. These aggregated labels are used to train a 3-layer MLP network with sigmoid activation function in a supervised manner. The final classification output is shown in Figure 4.5 along with the Pauli composition of

(a)   (b)   (c)

Figures/CD/RED.png (UC)   Figures/CD/GREEN.png (NU)   Figures/CD/BLUE.png (UU)
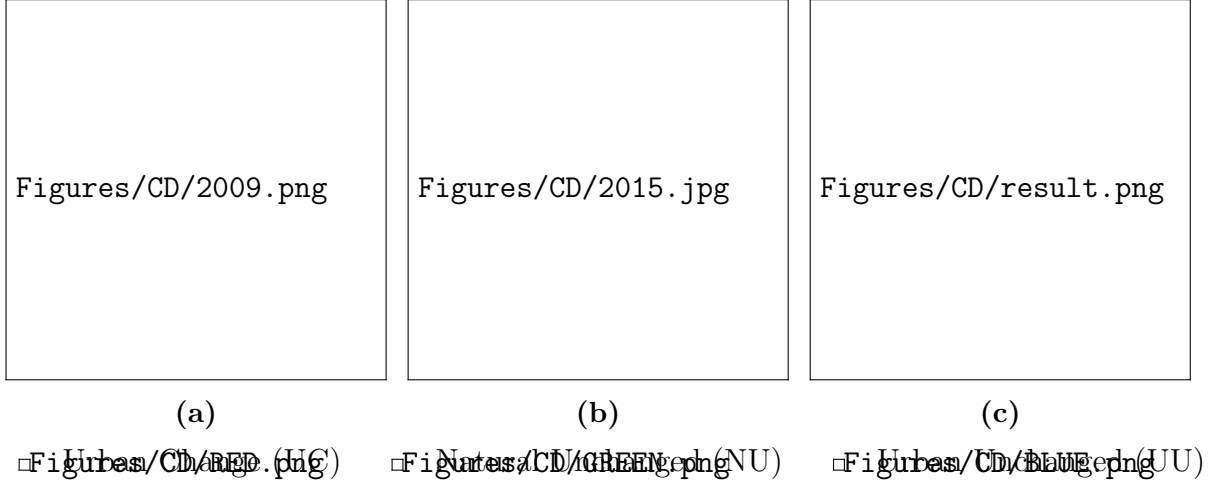
Figure 4.5: The Pauli composite of (a) 2009, (b) 2015 image, and (c) the result of the classi-fication.

the image pair. The output simultaneously maps areas of changed urban ares, unchanged built-up and natural land-cover. This is an improvement over outputs from traditional ratio based methods that are unable to leverage the polarimetric information.

4.4 Results and Discussion

Qualitatively, the output corresponds well with the changed areas observed in the Pauli composition image pair and has low misclassification as shown in Figure 4.5. Comparative crops are shown in Figure 4.6 for three areas of the image. Area 'A' is located close the near range of the radar, and hence is subject to radiometric anomalies due to the steep look angle of the radar, as common with airborne sensors. Despite this, the proposed method performs reasonably well and resilient to the amplitude gradient of the backscatter. The changes in area 'B' are very minor compared to the rest of the area. There is a construction of new housing that has occurred in the vacant lot amidst other built-up areas. The challenge in this crop is that it is close to a train station (right edge), and due to the movement of infrastructure there is inconsistent radar backscatter between the two images. The proposed method is able to capture the change perfectly, ignoring the interference from the railway station. Area 'C' radar artifacts due to return from power-lines can be observed in the two images. Also, the natural areas surrounding the urban structures are significantly changed. The parcel based approach employed in the proposed methodology is able to counter the effects of the artifacts. The method is also able to segregate urban changes from natural ones, something that would not be possible with a traditional change index methodology.

Quantitatively, the confusion matrix is computed and presented in Table 4.1. The overall accuracy $OA = 92.34\%$ with $\kappa = 0.88$. The urban change (UC) areas are most

Figures/CD/ADD/1.png

(a)

Figures/CD/ADD/2.png

(b)

Figures/CD/ADD/3.png

(c)
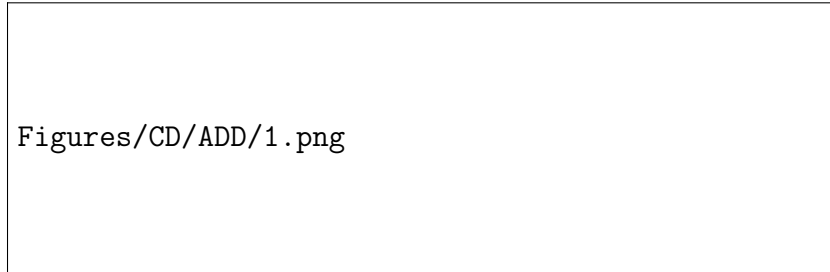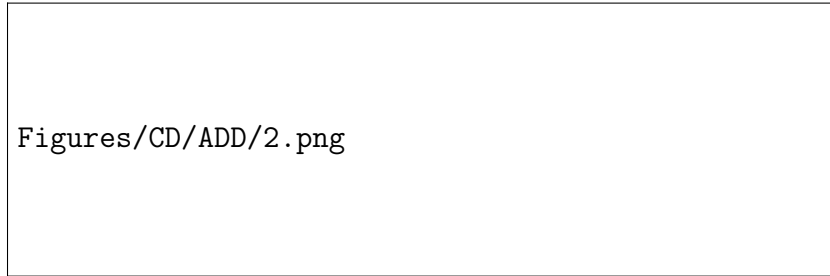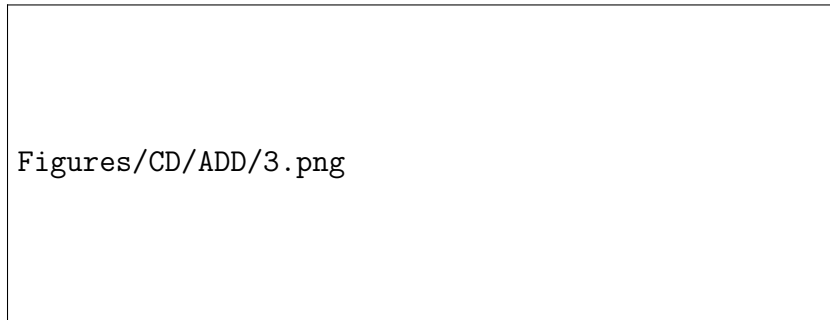
Figures/CD/RED.png (UC)   Figures/CD/GREEN.png (NU)   Figures/CD/BLUE.png (UU)

Figure 4.6: From left to right, the Pauli composite of the 2009 and 2015 images, along with the result of the classification is presented for areas (a) A, (b) B and (c) C.

Figures/CD/A2_2009.png Figures/CD/A2_2015.png Figures/CD/REF_A2.png Figures/CD/OP_1_A2.png

(a)　　　　　　**(b)**　　　　　　**(c)**　　　　　　**(d)**

Figures/CD/BLK.png Figures/CD/WHT.png
No Change　Change

Figure 4.7: The Pauli composite of detail (a) 2009, (b) 2015 image, and the (c) reference
change map, and (d) output change map.

commonly confused with the urban unchanged (UU) ones. This corresponds to areas
where the learned representation was not sufficient to discriminate the two. Another
source of confusion is between natural unchanged (NU) and UU classes. The presence of
vegetation intermixed with buildings in the UU areas makes it more difficult to discern
it from the NU class leading to misclassification.

To evaluate the CD capabilities the two unchanged classes, i.e. UU and NU are grouped
together as a single unchanged class, with the UC representing the change as shown in
Figure 4.7. Performance is evaluated in terms of false positive rate (PFP), false negative
rate (PFN), false alarm rate (FA(%)), detection rate (DR(%)) and kappa (Table 4.2).
The method has a low FA rate of $FA = 2.64\%$. This is due to design of the AE
loss function and the radial thinning performed in the label aggregation step. The
$DR = 82.3\%$ can be improved at the cost of a higher FA, by eliminating the radial
thinning step during label aggregation. The $\kappa = 0.83$ indicates that there is very good
agreement between the reference and output maps. This is confirmed qualitatively by
visual inspection of the output image.

Figure 4.8 depicts a traditional log-ratio change index computed on the dataset. The
index can reasonably capture the change in the scene, but the original scattering mecha-
nism information is discarded. This prevents simultaneous land-cover classification from
being performed. To obtain both the change and land-cover map, two separate processes
must be run on the dataset consuming time and computing resources. In contrast, the
proposed method is able to efficiently compute both the maps in one go.

Further, the change index is unable to differentiate between natural and urban changes
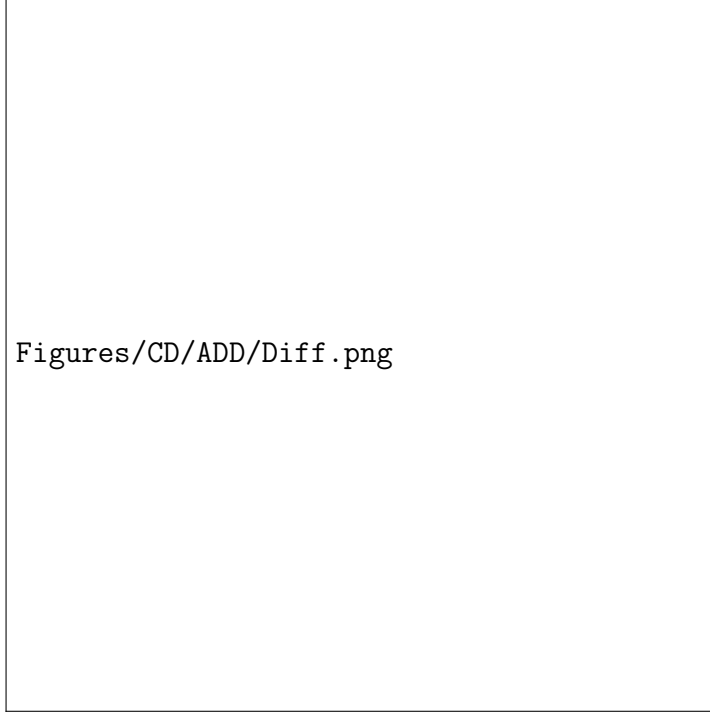
Figure 4.8: A traditional log-ratio change index generated for the dataset.

Table 4.1: Confusion Matrix

	UC (%)	NU (%)	UU (%)	Total
UC (%)	90.28	0.34	0.00	26.11
NU (%)	2.24	88.54	1.10	35.91
UU (%)	7.47	11.12	98.90	37.97
OA	92.34%	κ	0.88	

and only captures the magnitude of the change, disregarding scattering information. Thus, it identifies both the natural and urban changes in the scene as a change, and further analysis is needed to be able to identify the change. In the proposed method, the changes are encoded in the representational layers of the AE, which is then projected in a manifold. Because the scattering information is encoded inherently in this representation, each region of the space corresponds to a particular transition of scattering mechanism between the two images. Due to this it is possible to not only identify the original land cover classes, but also the change information in one step. In a work-flow where a large number of scenes must be automatically processed, this computational advantage can translate to time and resource savings.

Table 4.2: Change Detection Metrics

PFP(%)	PFN(%)	FA(%)	DR(%)	κ
0.89	1.75	2.64	82.3	0.83

| (a) | (b) | (c) | (d) | (e) |
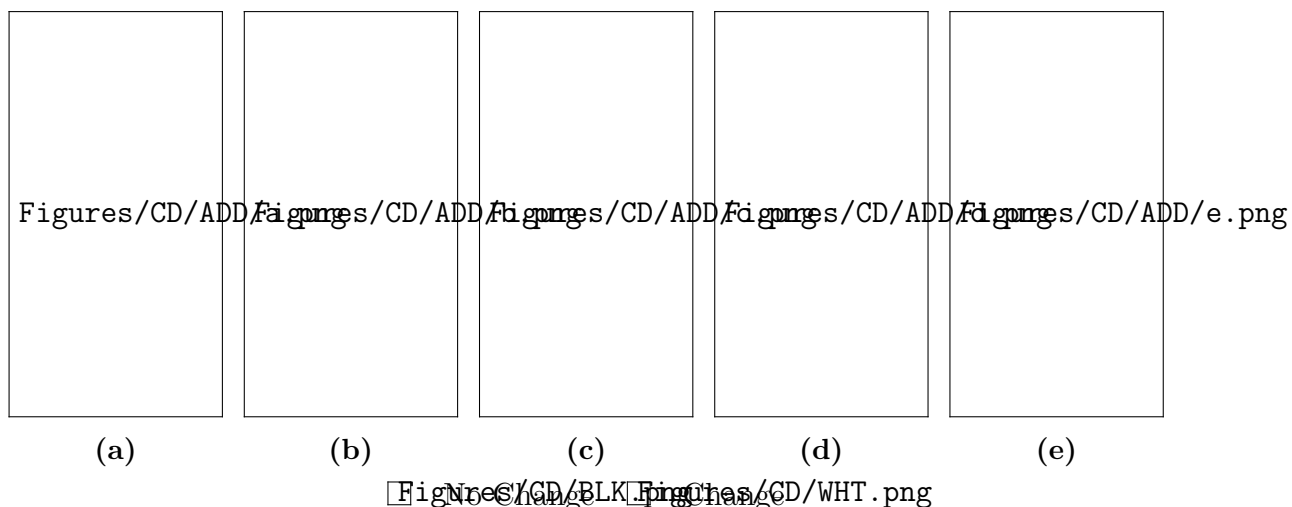
Figure 4.9: The Pauli composite of detail (a) 2009, (b) 2015 image, and the result of change detection by (c) proposed and (d) geodesic distance methods, shown along with the (e) reference map.

Comparative Analysis

The proposed methodology is compared with another recent PolSAR change detection technique [130]. The results are tabulated in Table 4.3 and shown in Figure 4.9. The proposed method is able to outperform the contemporary technique, with an overall accuracy of 92.34% and $\kappa = 0.88$. The section of image used in the comparison is fairly challenging for a change detection algorithm due to the presence of power-lines in the scene. These have a very strong radar return, saturating the receiver and leading the artifacts in the image, as seen. Since both the algorithms use variation in the scattering mechanism to identify changes, they are able to successfully avoid the interference of these power-lines. In contrast, change index and amplitude difference based techniques would be susceptible to this nature of interference.

However, the geodesic distance (GD) measure is still sensitive to changes in natural areas. Over time, vegetation will grow and change, leading to a variation in the observed

Table 4.3: Comparison with other methods

Approach	Overall Accuracy	Kappa
Geodesic Distance (GD) [130]	89.20	0.78
Proposed	92.34	0.88

scattering mechanism. Since the final output is obtained by applying a single threshold to the GD value, it is not possible to fully separate the natural changes from urban ones. In contrast, the proposed method uses ellipsoid fitting to improve the separation of urban changes in the feature space itself. This is followed by a MLP layer, which is further able to flexibly threshold the images to reduce the effect of natural changes. The region growing in the feature space and parcel based approach at input, combined lead to a very low false alarm rate of 2.64% in the proposed technique, and often in practical application of change detection, a lower false alarm rate is more critical than a high accuracy.

It must be noted that although the proposed technique outperforms the GD method, it is much more computationally complex. The training phase of the AE is also time consumptive, however, once trained, the final classification can be performed fast. In contrast, the GD method is computationally simple, and can be computed much quicker, at the cost of a slightly lower accuracy and higher false alarm rate.

Chapter 5Multi-frequency Crop Classification

In this chapter, a novel tensor product and neural network based framework for classification of vegetative areas like agricultural fields and forests is presented (Section 5.1). Further, experiments are conducted to test and validate the proposed methodology (Section 5.3). Finally, the results of the experiments are discussed (Section 5.4).

5.1 Methodology

The methodology is divided into two components: Synthesis of the feature vector using tensorization (Section 5.1.1) and classification using an ANN (Section 5.1.2).

5.1.1 Data Preprocessing

A tensorization framework is proposed which utilizes the Kronecker product of two matrices $\mathbf{A} \in \mathbb{F}^{p \times q}$ and $\mathbf{B} \in \mathbb{F}^{r \times s}$ denoted as $\mathbf{A} \otimes \mathbf{B}$ and is defined as,

$$\mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} a_{11}\mathbf{B} & a_{12}\mathbf{B} & \dots & a_{1q}\mathbf{B} \\ \vdots & \vdots & \ddots & \vdots \\ a_{p1}\mathbf{B} & a_{p2}\mathbf{B} & \dots & a_{pq}\mathbf{B} \end{bmatrix}_{pr \times qs} \tag{5.1}$$

where $\mathbb{F}$ is a field such as $\mathbb{R}$ or $\mathbb{C}$. The matrix block $a_{ij}\mathbf{B}$ is of dimension of $\mathbf{B}$. In general, the Kronecker product of two matrices is non-commutative, meaning that $\mathbf{A} \otimes \mathbf{B} \neq \mathbf{B} \otimes \mathbf{A}$. Each entry of $\mathbf{A} \otimes \mathbf{B}$ (or $\mathbf{B} \otimes \mathbf{A}$) is of the form $a_{ij}b_{kl}$ (or $b_{kl}a_{ij}$). Thus $\mathbf{B} \otimes \mathbf{A}$ can be obtained by permuting the entries of $\mathbf{A} \otimes \mathbf{B}$.

In PolSAR literature, the Kronecker product is used to derive the Kennaugh matrix from the scattering matrix $\mathbf{S}$. A radar target is characterized by a scattering matrix $\mathbf{S}$ which describes the dependence of scattering properties on polarization. It is defined in the hv basis as,

$$\mathbf{S} = \begin{bmatrix} S_{hh} & S_{hv} \\ S_{vh} & S_{vv} \end{bmatrix} \tag{5.2}$$

where each element is a complex quantity composed of the amplitude and the phase of the scattered electromagnetic signal. The Pauli target vector is generated from $\mathbf{S}$ as,

$$\boldsymbol{kp} = \frac{1}{\sqrt{2}} \begin{bmatrix} S_{hh} + S_{vv} & S_{hh} - S_{vv} & 2S_{hv} \end{bmatrix}^T \tag{5.3}$$

which is then utilized to obtain the coherency matrix, $\mathbf{T}$, by spatial ensemble averaging,

$$\mathbf{T} = \frac{1}{N} \sum_{i=1}^{N} \boldsymbol{k_{p_i}} \boldsymbol{k_{p_i}^{\dagger}} \tag{5.4}$$

where $\dagger$ is the conjugate transpose. The coherency matrix $\mathbf{T}$ is a hermitian positive semi-definite matrix, and thus can be diagonalized as,

$$\mathbf{T} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^{\dagger} = \sum_{i=1}^{3} \lambda_i \boldsymbol{u_i} \boldsymbol{u_i}^{\dagger} \tag{5.5}$$

where $\boldsymbol{\Lambda}$ is a 3×3 diagonal matrix with non-negative real eigenvalues, and $\mathbf{U} = [\boldsymbol{u_1}\ \boldsymbol{u_2}\ \boldsymbol{u_3}]$ is a unitary matrix of the SU(3) group, where $\boldsymbol{u_1}$, $\boldsymbol{u_2}$, and $\boldsymbol{u_3}$ are the orthogonal eigenvectors corresponding to the eigen values $\lambda_1, \lambda_2, \lambda_3$.

If $\mathbf{T}_1 = \mathbf{U_1}\boldsymbol{\Lambda_1}\mathbf{U_1}^{\dagger}$ and $\mathbf{T}_2 = \mathbf{U_2}\boldsymbol{\Lambda_2}\mathbf{U_2}^{\dagger}$ are the eigenvalue/eigenvector decompositions of $\mathbf{T}_1$ and $\mathbf{T}_2$ respectively, then,

$$\mathbf{T}_1 \otimes \mathbf{T}_2 = (\mathbf{U_1} \otimes \mathbf{U_2})(\boldsymbol{\Lambda_1} \otimes \boldsymbol{\Lambda_2})(\mathbf{U_1} \otimes \mathbf{U_2})^{\dagger} \tag{5.6}$$

is the eigen decomposition of $\mathbf{T}_1 \otimes \mathbf{T}_2$ utilizing the mixed Kronecker product properties. Here $\mathbf{T_k}$ corresponds to the frequency band f_k. Hence, using the Kronecker product the diagonal matrices corresponding to two or more frequencies may be combined to obtain the tensorized product,

$$\mathbf{X}_{f_1, f_2, \ldots, f_k} = \Lambda_{f_1} \otimes \Lambda_{f_2} \otimes \ldots \otimes \Lambda_{f_k}. \tag{5.7}$$
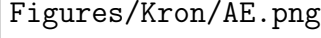
The non-zero elements of the matrix $\mathbf{X}_{f_1, f_2, \ldots, f_k}$ is used as input vector $\boldsymbol{x}$ in the ANN. The eigen decomposition of the Kronecker product of the coherency matrices corresponding to C-, L- and P-band is used to generate feature vectors for classification.

Current technology allows for simultaneous multi-frequency acquisition of PolSAR data of the same scene. Thus, efficient techniques for augmentation and fusion of the acquired data for the analysis of enhanced information content is desirable. The proposed tensorization framework allows for homogeneous composition of information from each frequency-band and prevents an *a-priori* bias towards a dominant band. Hence, such a combination leads to a higher dimensional representation which, in general, is more separable for classification tasks if exploited by suitable learning algorithms.

5.1.2 Unsupervised Representation Learning using Auto-Encoders

The ANN architecture used in this work is divided into two stages as shown in Figure 5.1. First, an unsupervised stochastic sampling sparse stacked Auto-Encoder (AE) stage

Figure 5.1: Schematic diagram of the proposed framework.

learns an efficient intermediate representation of the combined multi-frequency features. Subsequently, this learned representation is classified by a FF network.

The AE is a fully connected, feed-forward, non-recurrent neural network with multiple hidden layers. The number of nodes in the input and output layers is equal. It consists of three parts: the encoder, the decoder and a representational layer (z) as shown in Figure 5.1. During the training phase, the output nodes x'_i are connected to the input nodes x_i. For each pixel i, the AE attempts to reconstruct the x_i as x'_i. The reconstruction error is computed and minimized over multiple iterations in order to learn an optimal reconstruction.

The AE is preceded by a stochastic sampling block which randomly selects a vector as x_i from a $N \times N$ window in each iteration as input vector and target while training the network. This is akin to using a smoothing filter and leads to improved generalization performance of the AE in the presence of speckle [131]. Additionally, the windows are also chosen in random order to prevent the AE from memorizing the order of input.

This representation is built in an unsupervised manner for each pixel in the dataset. The fundamental computational unit of an AE is a connected neuron that takes as input $x_i = [x_1, x_2, \ldots, x_n]$ and a bias intercept term b and produces the hypothesis $h = \mathbf{s}(\sum_{i=0}^{n} \mathbf{W_{r_i}} x_i + b)$, where $\mathbf{s}$ is a non-saturating non-linearity called Parametric Rectified Linear Unit (PReLU). The response of $\mathbf{s}(y_i) = y_i$ when $y_i > 0$ and $\mathbf{s}(y_i) = a_i y_i$ when $y_i < 0$. Here, a_i is a learnable parameter controlling the slope of the negative response of the PReLU [56]. As the width of the input vector increases with the tensorial combination of multiple frequencies, the ReLUs tend to have difficulties with zero gradients during back-propagation. This prevents efficient training of the AE. Hence, the use of PReLUs helps to overcome this problem and improve the convergence rate of the network.

The representation x has relatively high dimensionality with redundancies. This makes classification difficult, which either leads to poor performance or requires a more complex classification scheme. However, an optimal representation learned by the AE simplifies the classification. It is observed that even with a relatively simple FF network, a high level of accuracy is obtained.

On completion of the training phase, the update of weights in the AE are suspended, and the decoder is disconnected. The representation z_i is extracted for each pixel i and is used as input in the FF network stage.

5.1.3 Supervised Classification

The final classification is performed in a supervised manner by the FF network. During this stage, only the labeled training pixels are considered. z_i is applied as an input to the FF network and the labels L_i are applied at the output. The error between the predicted and actual labels is minimized over multiple iterations with only the weights of the FF network being updated while utilizing the representation learned in the training phase from the AE. The terminal layer of the FF uses Softmax saturating non-linearities scaled in the range $[0, 1]$, instead of PReLU. Here a value close to 1 indicates an active response to the corresponding label, while 0 represents an inactive one.

The proposed architecture allows computation of classification confidence for each pixel as a part of the FF network. The number of nodes in the output layer is chosen to be one more than the minimum required to represent every class. This node, called zero (or null) label, remains unmapped to any label and allows the estimation of uncertainty of classification. The null label represents those data points which have not been successfully classified into one of the given set of labels. This is used to estimate the overall confidence level of the classified output by the network and simultaneously exclude pixels with high uncertainty from the final output. In practice, the total labeled pixels are randomly divided into three groups, training, test and validation. The fine tuning is conducted only on the training pixels, with the test pixels used to check for over-fitting at regular intervals. The network performance is measured at the end of classification over the validation pixels.

5.2 Dataset and Study Area

Two multi-frequency polarimetric SAR datasets from the AIRSAR system are used in this study. A C-, L-, P-band dataset acquired over and agricultural area in Flevoland, Netherlands on 3 July 1991, and a C-, L-band dataset acquired over a forested area in Landes, France on 20 June 1991. These 16-look datasets have a slant range resolution of $6.66\,\mathrm{m}$ and an azimuth resolution of $8.20\,\mathrm{m}$. The Flevoland site consists of an agricultural area with large uniform fields with an average area of $\sim$20 ha. Extensive ground truth information from the campaign is available in literature [132]. A 750×700 pixel subset is used to present the results in this paper. The crops present in this subset are Peas, Wheat, Rapeseed (R.Seed), Lucerne, Barley, Potato and Beet. Most of the crops are in the middle of their growth stage. The Landes site is a pine forest, with tree-patches of differing ages. A reference map is constructed from observation of polarimetric signatures.

5.3 Experimental Set-Up

The ANN consists of two stages, an AE and a FF network. The AE consists of 5 hidden layers: 2 encode, 2 decode and 1 representational layer. The FF consists of 3 feed forward fully connected layers. First, the AE is trained in an unsupervised manner under the constraint of minimization of cross-entropy error between input $\boldsymbol{x_i}$ and output $\boldsymbol{x'_i}$ with the FF disconnected (see Figure 5.1). The learning rate is $l_r = 10^{-4}$ with a reduction by a factor of $\Gamma = 0.1$ at every 2 training epochs. Optimization is performed using the ADAM solver with a momentum $\mu = 0.95$ and weight decay $\beta = 0.005$. The total labeled data is divided into 3 parts: training, test and validation which are selected at random. The proportions are 10%, 30% and 60% of the total labeled data respectively. The training is balanced by selecting a fixed number of pixels per class randomly from the 10% training pool. The accuracy statistics are reported as an average of 5 runs on the validation set.

After the training stage of the AE is complete, the representational layer is extracted and given as input to the FF network. The FF network is trained in a supervised manner with $l_r = 10^{-5}$, $\mu = 0.92$ and $\beta = 0.005$. The classification is obtained as output from the terminal layer of the FF network.

For comparison, the individual bands (C, L and P), tensorized band pairs (CL, CP, LP), tensorized triplet (CLP) and a simple band augmentation (CLP_+) are considered. For the individual bands, the input vector $\boldsymbol{x}$ is of dimension 1×3, while for tensorized pairs and triplets, $\boldsymbol{x}$ is of dimension 1×9 and 1×27 respectively. The benefit of using the tensorization framework is demonstrated by comparing it against the simple augmentation of all frequency bands. In the augmented case, the 1×9 resultant input vector is formed by appending individual frequency input vectors. The proposed framework is also compared with standard classification techniques like Support Vector Machine with a Radial Basis Function kernel (SVM-RBF) and Extreme Learning Machine (ELM) network. The parameters used for comparison are automatically determined by standard techniques. For the SVM, $\gamma = 0.125$ and $C = 2^{-4}$. In the ELM, $N = 1200$ with a sigmoid activation function is chosen.

It may be noted, that even though the Kronecker product operation is non-commutative, the operand ordering has no significant effect on the proposed algorithm since the AE is insensitive to the sequence of elements in the input vector. The null label node in the output layer is used to estimate the classification performance while qualitatively improving the results. If the output activation level of the node corresponding to the null label exceeds that of the other nodes, the corresponding pixel is considered unclassified and is excluded from the final classification output. The average normalized activation of the successfully classified pixels is used as an estimation of the classification confidence.

Table 5.1: Class Wise Overall Accuracy

/	C	L	P	CL	LP	CP	CLP	CLP$_+$
Peas	0.40	0.26	0.00	0.84	0.66	0.99	**1.00**	0.86
Wheat	0.70	0.81	0.91	0.84	0.89	**0.95**	0.93	0.90
R.Seed	0.81	0.97	0.88	0.99	0.99	0.99	**1.00**	0.98
Lucerne	0.82	0.94	0.07	0.98	0.91	0.85	**0.99**	0.97
Barley	0.73	0.77	0.76	0.89	0.97	0.93	**0.98**	0.95
Potato	0.71	0.31	0.31	0.89	0.69	0.70	**0.92**	0.90
Beet	0.32	0.27	0.32	0.56	0.68	0.63	**0.90**	0.75

Table 5.2: Table of Accuracy, Kappa and Confidence.

Band(s)	Accuracy (%)	Kappa	Confidence
C	89.78	0.62	0.86
L	90.86	0.64	0.80
P	90.75	0.65	0.87
CL	95.58	0.84	0.92
LP	89.79	0.62	0.92
CP	96.59	0.88	0.91
CLP (Proposed)	98.23	0.94	0.93
CLP_+	95.23	0.89	0.92
$SVM - RBF$	96.48	0.89	—
ELM	95.70	0.82	—

5.4 Results and Discussion

In this section the quantitative results from the classification of the Flevoland and Landes dataset are detailed, and possible explanations for the observed results are examined.

Flevoland Dataset

The result of classification using individual frequency bands is shown in Figure 5.2. The L-band has the best classification accuracy amongst the three individual bands with an overall accuracy of 90.86% and Cohen's kappa (κ) of 0.64. This is followed by the P-band with 90.75% and C-band with 89.79%. Classwise accuracies for each band or band-combinations are shown in Table 5.1. The values tabulated in bold indicate the best performance combination for the particular class. The longer wavelength of the P-band allows it a high penetration capability. This renders broad leaf, sparse crops like peas, potatoes and sugar beet nearly invisible to the radar leading to misclassification. Even at a comparatively shorter wavelength of L-band, the accuracy of classification of the pea crop is only 26.21%. At this frequency band, it is mostly confused with wheat. The

Figures/Kron/Validation_COLOUR.png Figures/Kron/C_COLOUR.png

(a) **(b)**

Figures/Kron/Ls_COLOUR.png Figures/Kron/P_COLOUR.png

(c) **(d)**

Figures/Kron/legend/Luc.png

Figures/Kron/legend/Beet.png

Figure 5.2: (a) Ground Truth. Classification outputs of input vectors derived from individual (b) C (c) L (d) P bands.

Figures/Kron/CL_COLOUR.png Figures/Kron/CP_COLOUR.png Figures/Kron/LP_COLOUR.png

(a)   **(b)**   **(c)**

Figures/Kron/CLP_COLOUR.png Figures/Kron/CLP2_COLOUR.png Figures/Kron/SVM_COLOUR.png

(d)   **(e)**   **(f)**

Figures/Kron/ELM_COLOR.png

(g)

Figure 5.3: Classification outputs of input vectors derived from (a) CL (b) CP (c) LP (d) CPL (e) CLP_+ band combinations with the proposed method and using (f) SVM-RBF (g) ELM methods. The SVM-RBF and ELM methods do not inherently generate a confidence map and are presented unmasked.

Figures/Kron/Conf/C.png Figures/Kron/Conf/L.png Figures/Kron/Conf/P.png

(a) **(b)** **(c)**

Figures/Kron/Conf/CL.png Figures/Kron/Conf/LP.png Figures/Kron/Conf/CP.png

(d) **(e)** **(f)**

Figures/Kron/Conf/CLP.png

(g)

Figures/Kron/Conf/grad.png
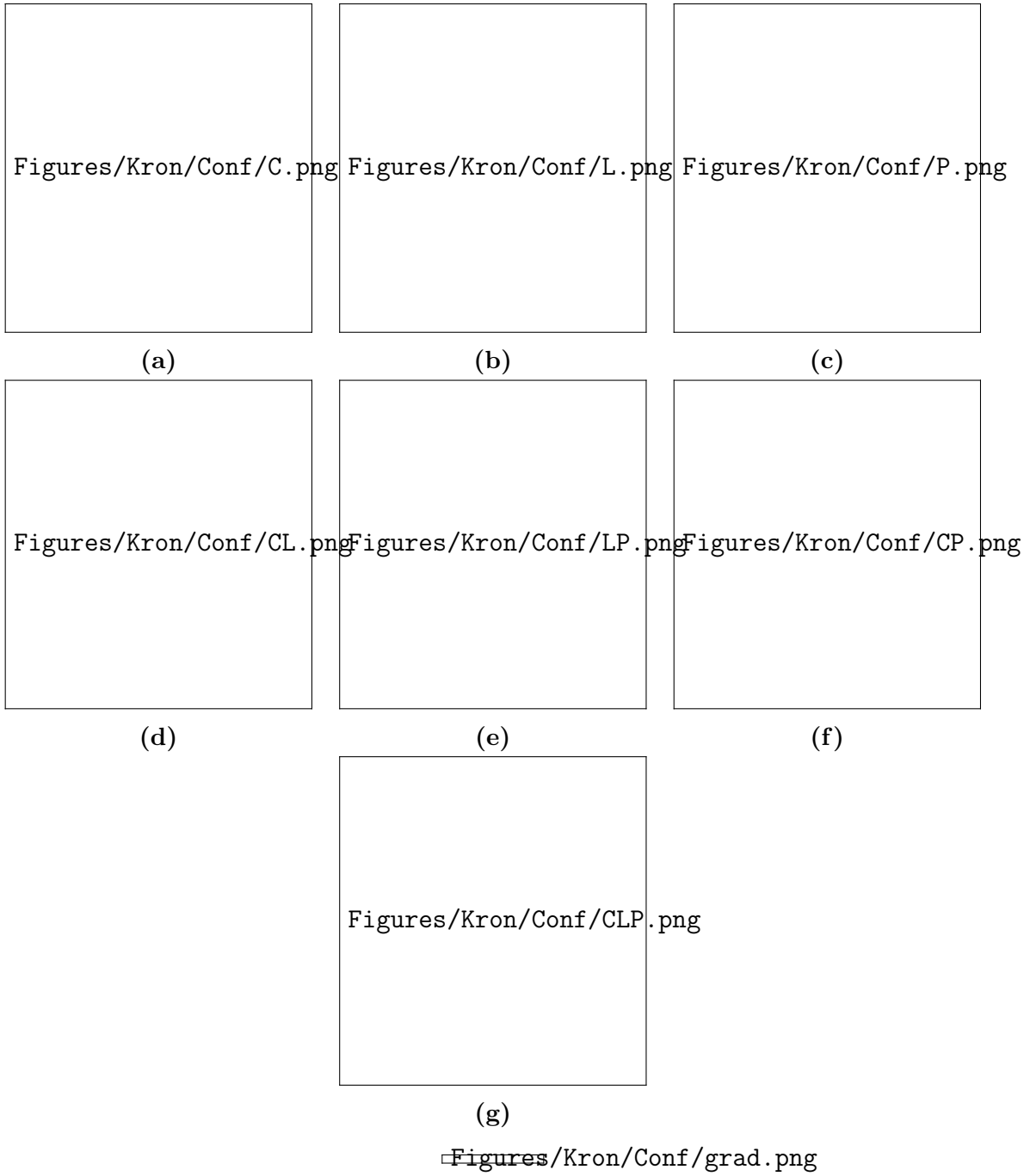
Figure 5.4: Confidence maps generated from the classification using (a) C (b) L (c) P (d) CL (e) LP (f) CP (g) CLP band combinations.

Table 5.3: Confusion Matrix for CLP

(%)	Peas	Wheat	R.Seed	Lucerne	Barley	Potato	Beet	Supp.
Pees	100	0.00	0.00	0.00	0.00	0.00	0.00	2055
Wheat	0.00	93.90	0.04	0.03	5.96	0.00	0.07	56060
R.Seed	0.00	0.00	100	0.00	0.00	0.00	0.00	25994
Lucerne	0.00	0.00	0.00	99.88	0.00	0.00	0.12	4935
Barley	0.00	2.03	0.00	0.00	97.93	0.00	0.04	45497
Potato	0.00	0.00	0.78	0.00	7.15	91.97	0.10	17035
Beet	0.00	3.63	0.07	3.63	2.82	0.04	89.81	14974
Prec.	2055	54107	26160	5489	49537	15673	13529	
Overall	Acc.	98.23%				κ	0.94	
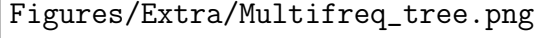
best classification for this crop is obtained using C-band with a classification accuracy of 40.01%.

Sugar-beet which has a height of $24-30\,\mathrm{cm}$ and has a cover of about $40\%-60\%$ is also misclassified by all three bands. Potato crops are taller and have a height of $50-60\,\mathrm{cm}$ with a cover of $90\%-95\%$. These are classified reasonably well with the C-band with an accuracy of 71.16%, but poorly so with L- and P-band with accuracies of 30.80% and 30.93%, respectively. Long stem vegetative crops like wheat, with a cover of $85\%-95\%$ and a height of $85-95\,\mathrm{cm}$, are well classified with all three frequencies. This disparity in performance is indicative of the differences in information content for each crop at different sensing wavelengths. This can be potentially exploited by combining information from different bands to improve the classification performance.

The result of multi-frequency classifications with the proposed approach is shown in Figure 5.3. The combination of information from different bands is able to improve the classification accuracy. The overall accuracies for individual and combined frequency bands along with two standard methods are reported in Table 5.2. It can be seen that the tensor combination of all three bands is the most informative and has the best classification performance with an overall accuracy of 98.23% with $\kappa = 0.94$. In this combination, all classes are correctly identified with the potatoes and sugar beet having the least accuracies at 91.97% and 89.82%, respectively. These classes, however, have not been significantly confused with the others, rather they have remained unclassified due to low confidence. This may be due to broad leaves and the relative size of the crop. The performance of the proposed algorithm is identical to the state-of-the art approach reported in [133]. However, the proposed method being non-parametric makes no *a-priori* assumptions about the distribution of the data.

Among the dual frequency combinations, CP has the highest performance with an overall classification accuracy of 96.59% with $\kappa = 0.88$. This combination contains the most diverse information, with the C band being sensitive to scattering from the vegetation canopy, while the P band is able to penetrate into deeper layers of vegetation. This is followed by the LP combination with an overall classification accuracy of 96.40%

Figure 5.5: Illustration of the differences in received response between sensing frequency bands for varied vegetation cover.

with $\kappa = 0.87$. Finally, the CL combination has the lowest performance, with an overall accuracy of 95.57% with $\kappa = 0.84$. However, it outperforms classification using individual bands. The multilayer information content carried by each (C-, L-, P-) band allows for better crop classification by integrating multi-frequency information. It is also seen that tensorization (CLP) is able to outperform simple band augmentation (CLP_+), and also the SVM-RBF and ELM approaches. Since the confidence map generation was not a part of the SVM-RBF and ELM techniques, they are presented unmasked. However explicit generation of confidence maps might be possible with different strategies in each technique. The confusion matrix for classification using CLP combination by the proposed method is presented in Table 5.3. Wheat is the most significant class in the dataset and is classified with an accuracy of 93.91%, with the rest being confused with barley. Among the crops present, the beet crop is misclassified with most other crops. The confidence maps for the band combinations with the least classification uncertainty for single, double and triple band combinations are shown in Figure 5.4g. In particular, for CLP combination (Figure 5.4g), the field boundaries are automatically masked out due to low confidence, thus improving the quality of the classification.

Landes Dataset

The response of vegetation to different sensing frequencies may be vary, as illustrated in Figure5.5, due to a variety of factors. The species of the tree itself it can cause differences in the density of foliage, size and geometry of leaves in the crown, which would affect the penetration of EM waves into it. Some species of trees shed foliage seasonally, while others retain it around the year, hence, the season of image acquisition affect the returned scattering signal. Similarly the age of the tree influences the biomass content, which in-turn affects the multi-frequency scattering.

This demonstrated in the signature analysis shown in Figure 5.6. The co- and cross-pol polarimetric signatures for L- and C-band acquisitions are shown for the same resolution

Figures/Extra/Signature/C/XpolSignature_50/E30raphignature/C/CopolSignature_50

(a)   (b)

Figures/Extra/Signature/L/CopolSignature_SC/630a/Signpmature/L/CopolSignature_50

(c)   (d)
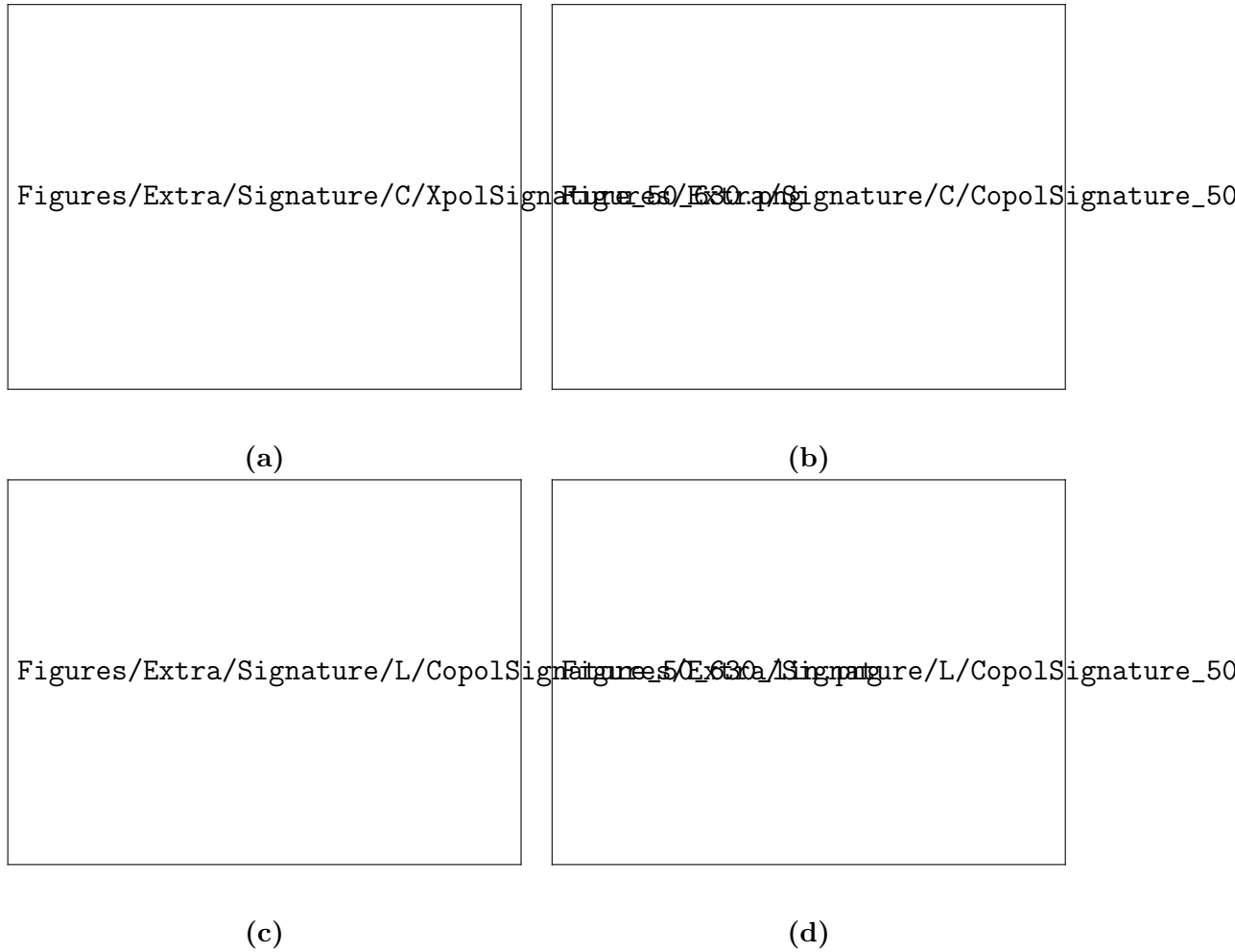
Figure 5.6: Signature analysis [(a) Co-Pol C-band (b) Cross-Pol C-band (c) Co-Pol L-band (d) Cross-Pol L-band] for the same pixel (508, 817) in a forested area at separate frequencies reveal that are different polarimetric information is contained in the same resolution cell at different frequencies.

cell in a forested area. The C-band has lower penetration capabilities and exhibits a signature more in tune with volume scattering, as can be expected from the foliage. The L-band is able to penetrate the foliage cover and return from the ground-trunk interfaces. Consequently the signature observed in the co-pol channel matches more closely with a double bounce scatter. For a coniferous forest, a more detailed study of polarization and radar backscatter mechanisms was conducted in [134]. Hence, for natural areas like crop, or forests, the presence of multiple frequency bands can immensely help target classification.

The Pauli representations of the C-band and L-band data-takes are shown in Figure 5.7. The same area is imaged simultaneously at the two frequencies. Observable difference can be discerned in the two images with respect to perception, scattering mechanism and returned power. Some areas, especially those corresponding to bare soil and specular

(a)   (b)
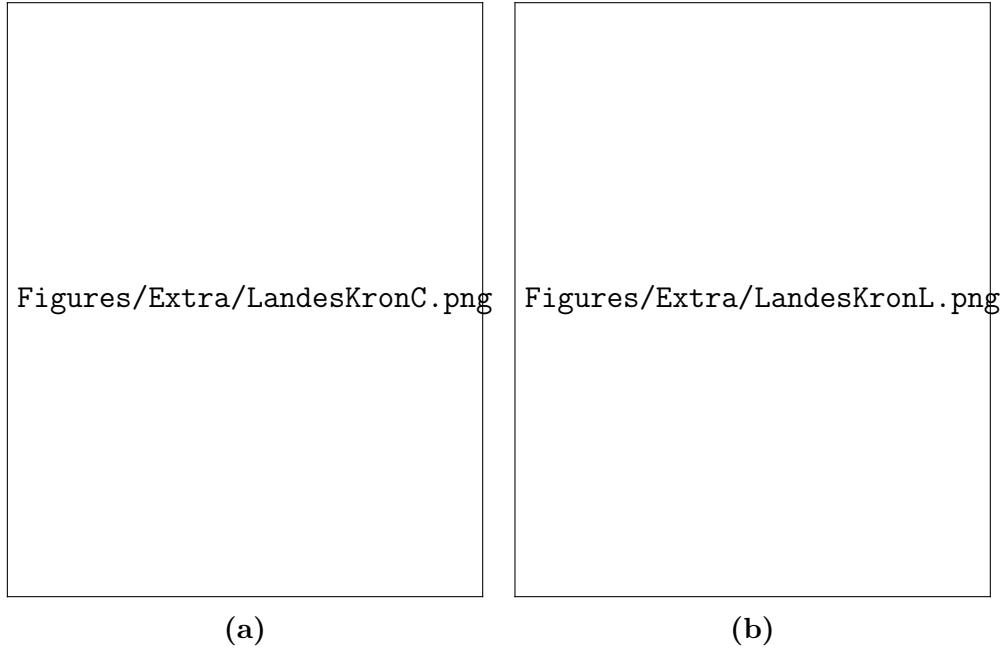
Figure 5.7: Pauli representation of the Landes Dataset at (a) C- and (b) L- band.

reflection have similar return power and scattering type in both images. This is because at the wavelength scale, for both L- and C-band, the scatterer viz. the smooth surface presents a similar reflection phenomenon. However, rough surfaces, with undulations comparable to the wavelength of C-band, undergo Bragg-type scatting for C-band, yet remain specular at L-band. This at C-band, the the returned power is higher, and the surface appears brighter than at L-band. A similar difference is scattering can be noted for voluminous targets like tree and vegetation patches. The phase center of the reflection differs depending on the sensing frequency, which affects the effective sensed returned power and scattering mechanism, leading to differing appearances in the Pauli representation.

Given the fact that the wavelength of the C-band is approximately half of that of the L-band, the inherent resolving capacity of C-band is much higher. The phase center of the returned scatter, i.e. the depth from within the target from which centroid of the signal has returned, also differs with sensing wavelength, with longer wavelengths penetrating deeper into the vegetation and having lower phase centers. As a combined result of these effects, the resolving capability and the average of point of scattering, the perceived spatial texture differs with sensing frequency. This can be exploited by the classification algorithm to tell targets with similar backscatters but different spatial texture properties apart.

Some targets present very similar backscatter independent of frequency. For instance, a wet soil, especially those types with high saturation capabilities like clay, have very little penetration, with all frequencies scattering from almost the same depth, consequently having very similar backscatter, independent of sensing wavelength. This consistency

Figures/Kron/Review/GT.pngFigures/Kron/Review/CL_raw_COLOUR_PR_0_185_1

(a) | (b)

Figures/Kron/Review/C+L_RAW+COLOUR_CROP_MASK.pngCOLOUR_PR_CROP_MASK.pn

(c) | (d)

Figures/Kron/Review/Legend/Barley.png
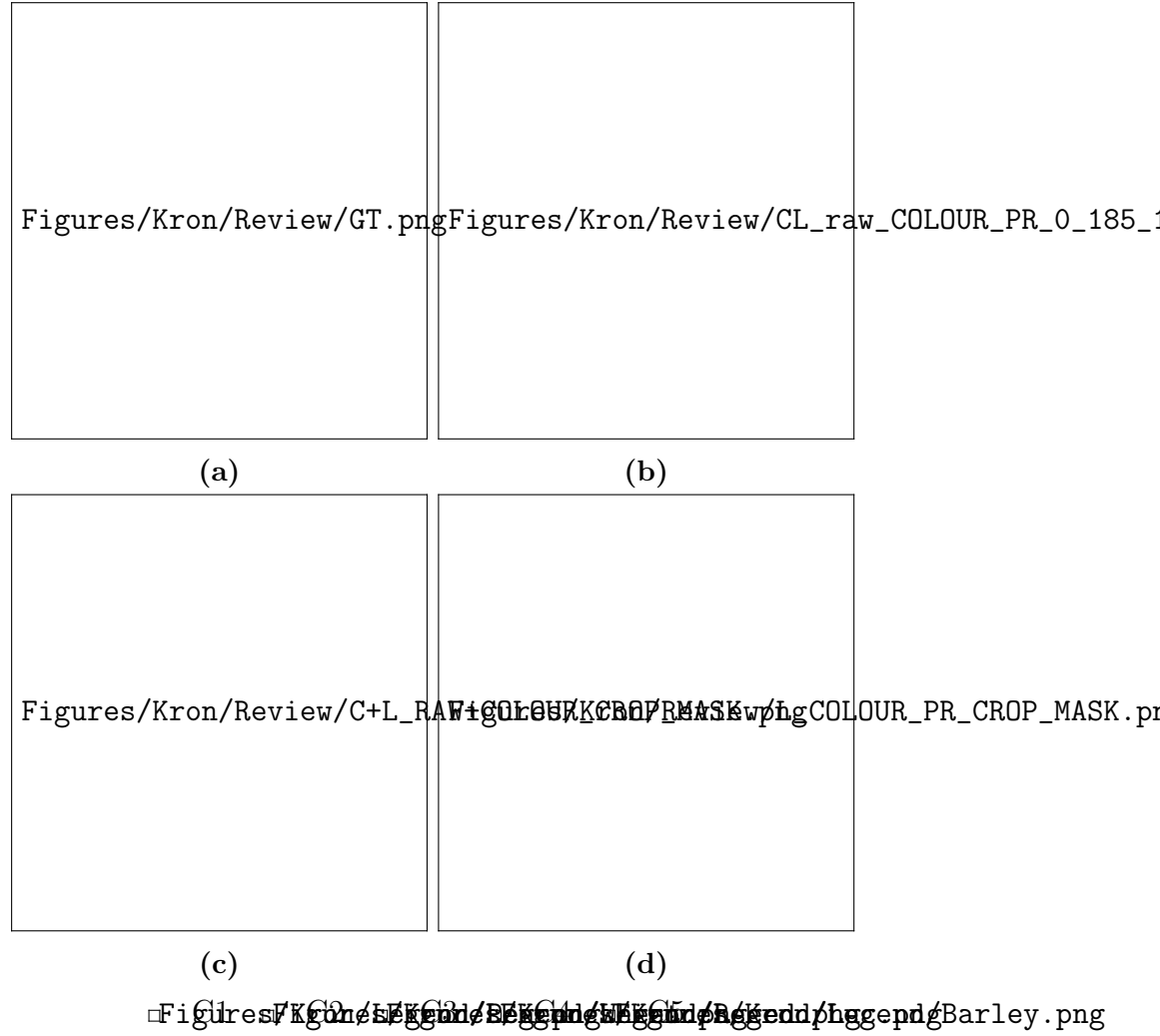
Figure 5.8: (a) Reference Image. Classification outputs of input vectors derived from individual (b) CL (c) CL_+ (d) L bands.

also can be useful in the interpretation of the target. The lack of diversity in the multi-frequency scattering can be used to classify targets of this nature.
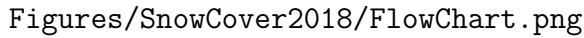
To establish the robustness of the results an experiment is conducted on the Landes dataset with two frequency bands. The classification results are presented in Figure 5.8. The tensorized pair CL has an overall accuracy of 96.23% with $\kappa = 0.94$. This exceeds the classification performance obtained by simple augmentation (CL_+), which has an overall accuracy of 92.15% with $\kappa = 0.91$. For individual bands, the L-band, with an overall accuracy of 89.73% with $\kappa = 0.70$, outperforms the C-band alone, which has an overall accuracy of 77.42% with $\kappa = 0.62$. In this case, the greater penetration ability of L-band is crucial in being able to segregate the age of the trees, since return from the top foliage of the tree is nearly identical irrespective of their ages.

Chapter 6 Snow-Cover Mapping

In this chapter, a novel snow cover mapping framework is proposed which uses the Poincaré sphere parameters obtained from the Radarsat-2 full-polarimetric SAR images in PolSAR images, for which the methodology is highlighted (Section 6.1). Further, experiments are conducted to test and validate the proposed methodology (Section 6.4). Finally, the results of the experiments are discussed (Section 6.5).

6.1 Methodology

An unsupervised Auto-Encoder (AE) is used to learn an optimal representation of these parameters. This is followed by a Feed Forward (FF) network which generates the final snow-cover map as a two-class classification process. The methodology consists of two parts, first the PolSAR data is transformed into basis simulated normalized stokes vectors (Section 6.1.1), followed by encoding and classification by an Auto-Encoder network (Section 6.1.2).

Figures/SnowCover2018/FlowChart.png

Figure 6.1: Overall description of the Neural Network.

6.1.1 Data Preprocessing

In PolSAR, the Mueller matrix is a real representation of backscattering measurement unlike scattering or covariance matrices which are complex. The Mueller matrix for incoherent scattering representation can be obtained from the elements of the coherency matrix $[\mathbf{T}]$ as follows:

Table 6.1: Incident Stokes vector representation (H: Linear horizontal, V: Linear vertical, R: Right circular, L: Left circular)

g_i	H	V	R	L
g_{0i}	1	1	1	1
g_{1i}	1	−1	0	0
g_{2i}	0	0	0	0
g_{3i}	0	0	−1	1

$$[\mathbf{M}] = \begin{bmatrix} \frac{T_{11}+T_{22}+T_{33}}{2} & \Re(T_{12}) & \Re(T_{13}) & \Im(T_{23}) \\ \Re(T_{12}) & \frac{T_{11}+T_{22}-T_{33}}{2} & \Re(T_{23}) & \Im(T_{13}) \\ \Re(T_{13}) & \Re(T_{23}) & \frac{T_{11}-T_{22}+T_{33}}{2} & -\Im(T_{12}) \\ \Im(T_{23}) & \Im(T_{13}) & -\Im(T_{12}) & \frac{-T_{11}+T_{22}+T_{33}}{2} \end{bmatrix} \tag{6.1}$$

The received Stokes vector $\begin{bmatrix} g_{0r} & g_{1r} & g_{2r} & g_{3r} \end{bmatrix}^T$ (6.2) is directly obtained from the observed Mueller matrix $[\mathbf{M}]$ by altering the incident Stokes vector $\begin{bmatrix} g_{0i} & g_{1i} & g_{2i} & g_{3i} \end{bmatrix}^T$ given in table 6.1.

$$\begin{bmatrix} g_{0r} \\ g_{1r} \\ g_{2r} \\ g_{3r} \end{bmatrix} = [\mathbf{M}] \cdot \begin{bmatrix} g_{0i} \\ g_{1i} \\ g_{2i} \\ g_{3i} \end{bmatrix} \tag{6.2}$$

The normalized Stokes vector is given in (6.3) [135] indicates a point on or inside the Poincaré sphere depending on the degree of polarization. This normalized vector is used as an input feature to the neural network, X.

$$X = (\bar{g}_{1r}, \bar{g}_{2r}, \bar{g}_{3r}) = \left(\frac{g_{1r}}{g_{0r}}, \frac{g_{2r}}{g_{0r}}, \frac{g_{3r}}{g_{0r}} \right) \tag{6.3}$$

6.1.2 Unsupervised Representation Learning using Auto-Encoders

The neural network is comprised of two parts, an unsupervised AE for feature representation learning, and a supervised Feed-Forward (FF) network for supervised classification as shown in Fig. 6.1. The AE has five layers: two encoding layers, two decoding layers and one representational layer, whose weights are denoted as Z. During training of the AE encoder, the FF network is disconnected i.e. weight updates for it are suspended. Rectified linear unit (ReLU) is used as the activation function to overcome the issue of

vanishing gradients during training in a deep network. The goal of the AE is to learn a representation X', given input X. It follows that X must be as close to X' as possible. This governed by the cross entropy error between them during the training phase give as,

$$H(X, X') = E_X[-\log X'] = H(X) + D_{KL}(X||X') \tag{6.4}$$

where $H(X)$ us the entropy of X, and $D_{KL}(X||X')$ is the Kullback-Leibler divergence of X from X'.

Training of the AE is considered complete if the change in cross entropy error between successive iterations is negligible, or if the maximum number of iterations are exceeded. The pixels must be presented in a different randomized order each epoch, to prevent the network from memorizing the order pixels, and instead try to learn an optimal representation to simplify classification of the data. After the training phase of the AE, the representational layer is extracted by applying the Poincaré sphere parameters and saving the weights of Z, in a pixel-by-pixel fashion. These extracted weights are used as input for the FF network. The AE is set up such that the dimensionality of Z is greater than that of X. This makes the representation sparse, which can lead to improved separability between classes [136].

6.2 Supervised Classification

The FF network is a three layer fully connected supervised feed forward network. The extracted representation Z is given as input along with training labels to the FF network. It is trained in a supervised manner with Sigmoid nodes used as non-linearities. The training in this case is governed by the $L2$ loss. After training, the unlabeled pixels are presented to the network, and the output response is monitored. The pixel is classified into the class corresponding to the node with the highest response. Thus, a sparse unsupervised representation is learned by an AE, followed by supervised classification of this representation by a FF network. The representation is more separable than the original feature space which improves classification performance. In this paper, we use both full and hybrid polarimetric data to evaluate their effectiveness in snow cover mapping application.

6.3 Dataset and Study Area

The test site is located in the Manali-Dhundi region between the latitudes of $32^o11'N$ to $32^o30'N$, and between the longitudes of $77^o2'E$ to $77^o24'E$. The lower elevations of the area is covered with coniferous forest and hence there is a natural divide between areas
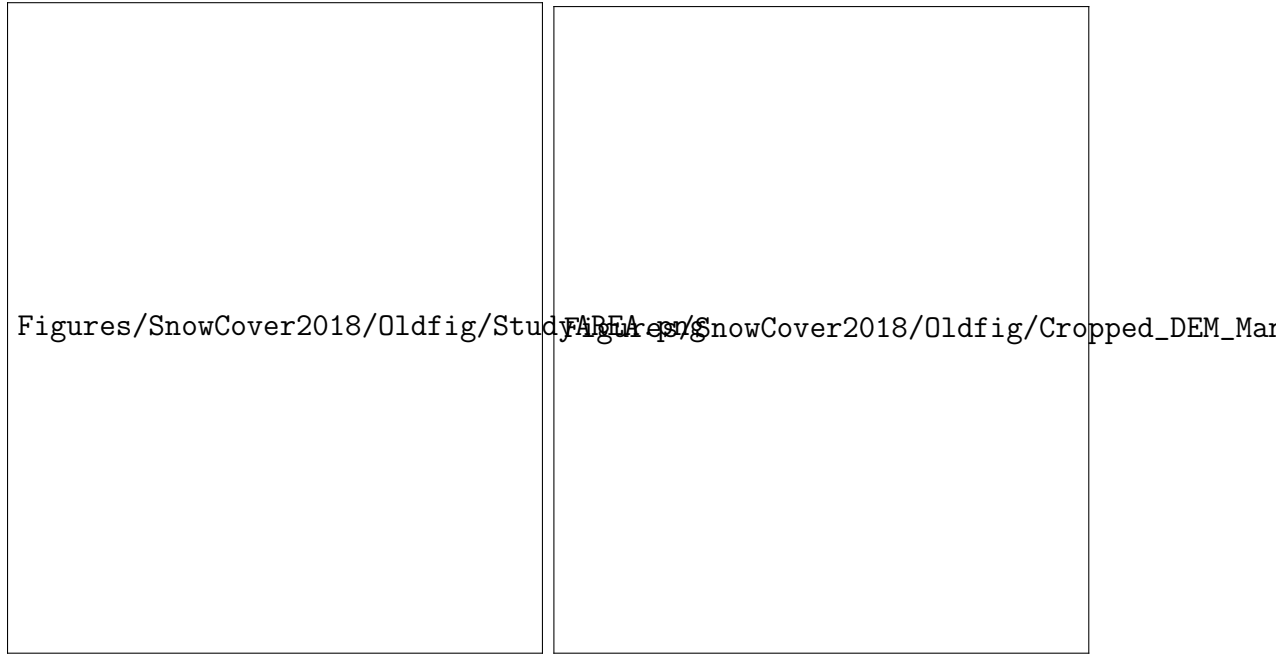
Figure 6.2: (a) SAR image footprint overlaid over the study area, long with markers denoting location of the observatories. (b) Topographic variations in the region.

with/without vegetation. The town of Manali is located in the bank of the river Beas which flows through the valley that contains the forest. A section of river Chenab is also captured in the scene which is housed in another valley (sparsely vegetated) located at a higher elevation. The Snow and Avalanche Study Establishment (SASE) maintains meteorological observatories over this area at differing elevations. Observatory data was used to gain a better understanding of the climate and snow conditions. The location of these observatories are marked in Figure 6.2 (a). Bahang observatory is at a height of 2000 meters above sea mean level (MAMSL), Solang is 2450 MAMSL and Dhundi is located 2900 MAMSL.

In the Indian Himalayan region, the snowfall generally occurs during the month of December to March from an altitude of 2000 m above the mean sea level. The mean minimum temperature in the month of January is around $-15°C$ to $0°C$ and the mean maximum temperature in the month of June is around $20°C$ to $30°C$. The elevation fn the region varies from 1685 MAMSL to 5020 MAMSL, with rapid topographic variations as show in Figure 6.2 (b). This variation in climate and altitude leads to an interesting micro-climate system to be setup in the area.

In this study a multi-spectral scene from Landsat-8 and a SAR data-set acquired by Radarsat-2 is used. The Landsat platform carries two instruments capable of imaging in visible, infrared and near-infrared wavelengths at a spatial resolution of $30m$, panchromatically at $15m$ and in the thermal infrared range range at $100m$ simultaneously. The scene used in this study was acquired on the 23rd of March, 2015 and was the nearest concurrent cloud free acquisition available to the SAR dataset.

The SAR scene was acquired on 18 March 2015 at 6:14 AM in the descending pass configuration with incidence angle in the range $46.0° - 47.2°$ across the swath. The sensor was in the 'Standard Quad-Pol' yielding a 25×25 Km fully polarimetric measurement with HH, VV, HV and VH polarizations at a spatial resolution of 11.8m in range and 5.1m in azimuth. The coverage of the optical image is significantly greater than that of the SAR scene and has been cropped accordingly. In this time of the year the Manali-Dhundi region is mostly snow-covered except for the relatively lower altitude forested zone which remains snow-free at the canopy level.

6.4 Experimental Setup

For the AE, the learning rate is $l_r = 10^{-4}$ with a reduction by a factor of $\Gamma = 0.1$ at every 2 training epochs. Optimization is performed using the ADAM solver with a momentum $\mu = 0.95$ and weight decay $\beta = 0.005$. The total labeled data is divided into 3 parts: training, test and validation which are selected at random. The proportions are 10%, 30% and 60% of the total labeled data respectively. The training is balanced by selecting a fixed number of pixels per class randomly from the 10% training pool. The accuracy statistics are reported as an average of 5 runs on the validation set. The FF network is trained in a supervised manner with $l_r = 10^{-5}$, $\mu = 0.92$ and $\beta = 0.005$. The classification is obtained as output from the terminal layer of the FF network.

The training is proceeded under constraint of the cross-entropy error, which is monitored individually for the training and test datasets. If the cross-entropy error of the two datasets follow very nearly the same trajectory, it can be deduced that the network is under-fitting the dataset. On the contrary, if the trajectories are held at a constant, but significant offset, with the training leading the test, the network is over-fitting the data, leading to a better fit in the training samples, but poor performance in the test-set. In the ideal case, the training leads the curve of the test cross-entropy error with a moderate margin. The aforementioned hyper-parameters are tuned to achieve the same.

6.5 Results and Discussion

In the lower altitude regions near the Bahang observatory, season wooded vegetative forests can be found to grow. The altitude is low enough for the snow-cover to melt in the summer months, hydrating the under-soil and allowing support for vegetative growth. In the winter months, the forests usually lose their foliage, and precipitation leads to accumulation of snow in their branches. Diurnal fluctuations in temperature can lead to freezing as shown in Figure 6.3.

In its frozen state, water is known to present a very different scattering signature than dry or wet snow, hence the forest structure even in the winter months, presents a very

Figures/SnowCover2018/Oldfig/Minimum_Temperature_Vs_Days.png

(a)

Figures/SnowCover2018/Oldfig/Maximum_Temperature_Vs_Days.png

(b)
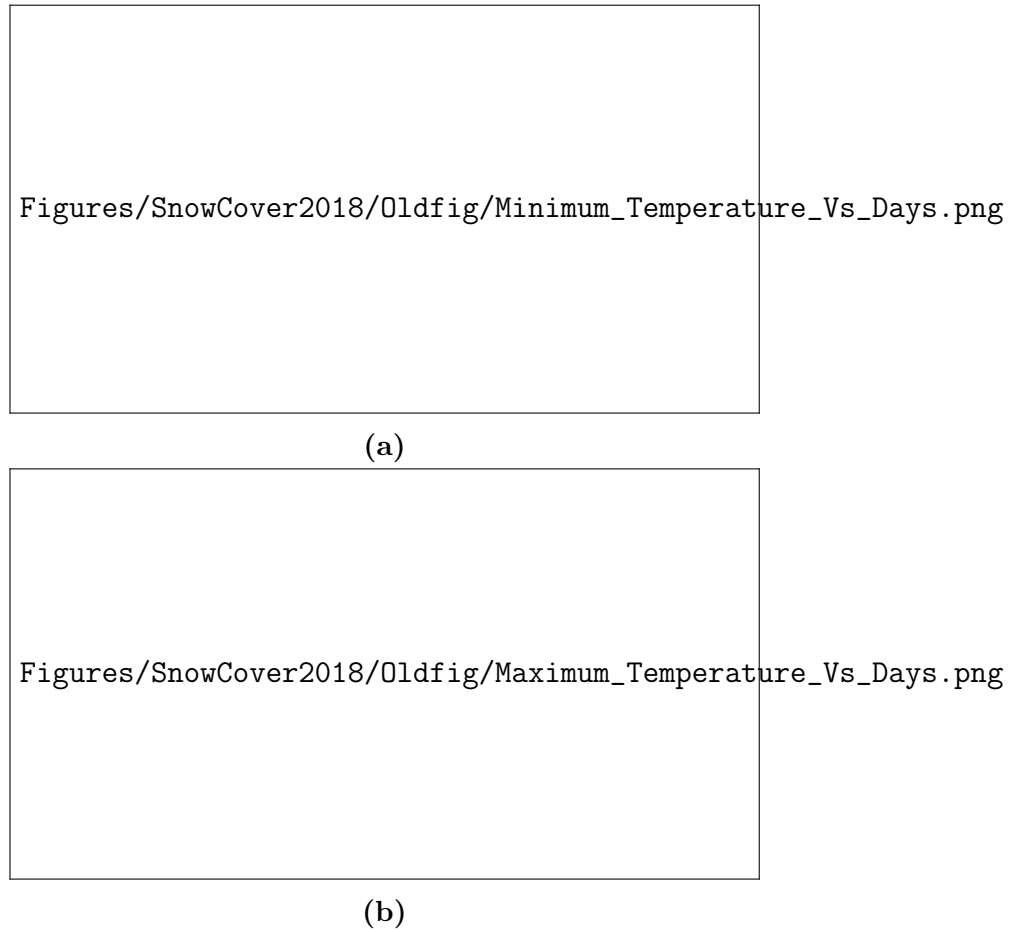
Figure 6.3: Diurnal trends in (a) Minimum and (b) Maximum air temperatures from January to March 2015 with the dashed pink line denoting the day of the acquisition of the SAR scene (18th March, 2015). Data is courtesy the Bahang observatory.

(a)   **(b)**
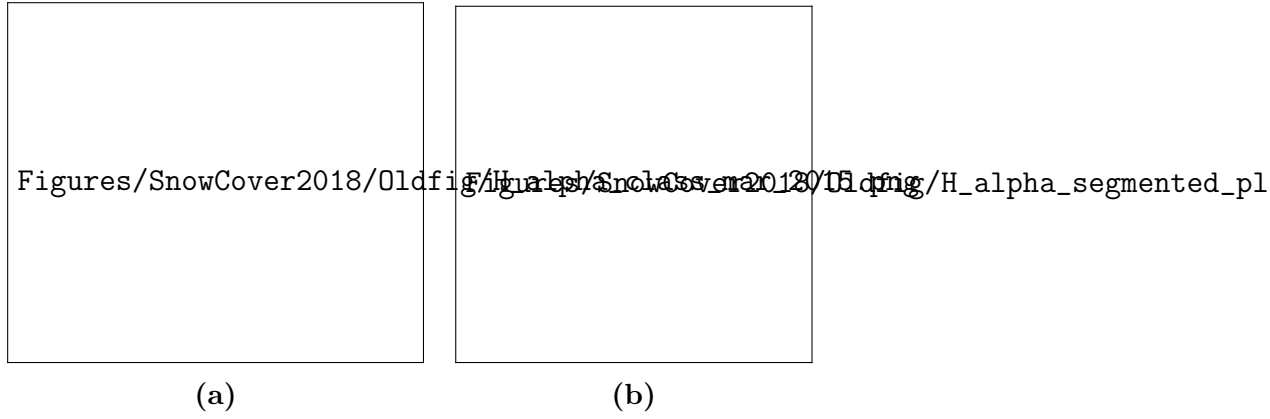
Figure 6.4: (a) H/α class map along with (b) H/α segmented plane for the March 2015 acquisition.

different scattering mechanism than the snow cover, allowing for discrimination. This is can be seen from the H/α analysis.

This transformation can be accredited to drop in the moisture content of the trunk, branches, and twigs due to freezing weather conditions. Due to freezing, the dielectric constant of vegetation distinctively decreases. The conifer species demonstrates higher percentage increase in stiffness since its twigs which are more flexible under thawed conditions exhibits greater percentage increase in Youngs modulus when frozen [137]. Freezing causes an increase in the value of Young's modulus of twigs, branches, and trunks of trees. Sub-zero temperatures are known to increase the stiffness of woody plants. The stiffening of the wood at sub-zero temperatures increases due to the formation of ice crystals in cells of frozen wood which renders the branches more rigid. In the case of coniferous species, under warmer temperatures, the ice crystals within the cells liquefy which allows increased bending of branches to shed the intercepted snow. The thin film of dry snow, if present, on the branches will be transparent at C-band [138].

Two common classes over the region were considered for all the analysis, viz. bare ground over which negligible or extremely sparse vegetation is present, and the forested area. A seasonal variation in the H/α parameters are conducted in this study over the bare snow-covered ground and the forested areas. The seasonal response of the terrain in the H/α plane corresponding to specific scattering characteristics (or zones) with multi-temporal images is shown in Figure 6.4. The figure shows the possible combination of H and α to form specific classes based on the properties of scattering mechanisms.

During February and March, the study area witnessed a substantial amount of precipitation in the form of snowfall. The areas which were bare during rest of the year were covered with snow during winters. An increase in the value of both α and H was observed over the bare-ground. Also, the polarimetric signatures for a forest patch and a bare ground covered by snow are plotted in Figure 6.5 . As can be seen, the scattering mechanisms for the two classes are very different.

(a)

(b)

(c)

(d)
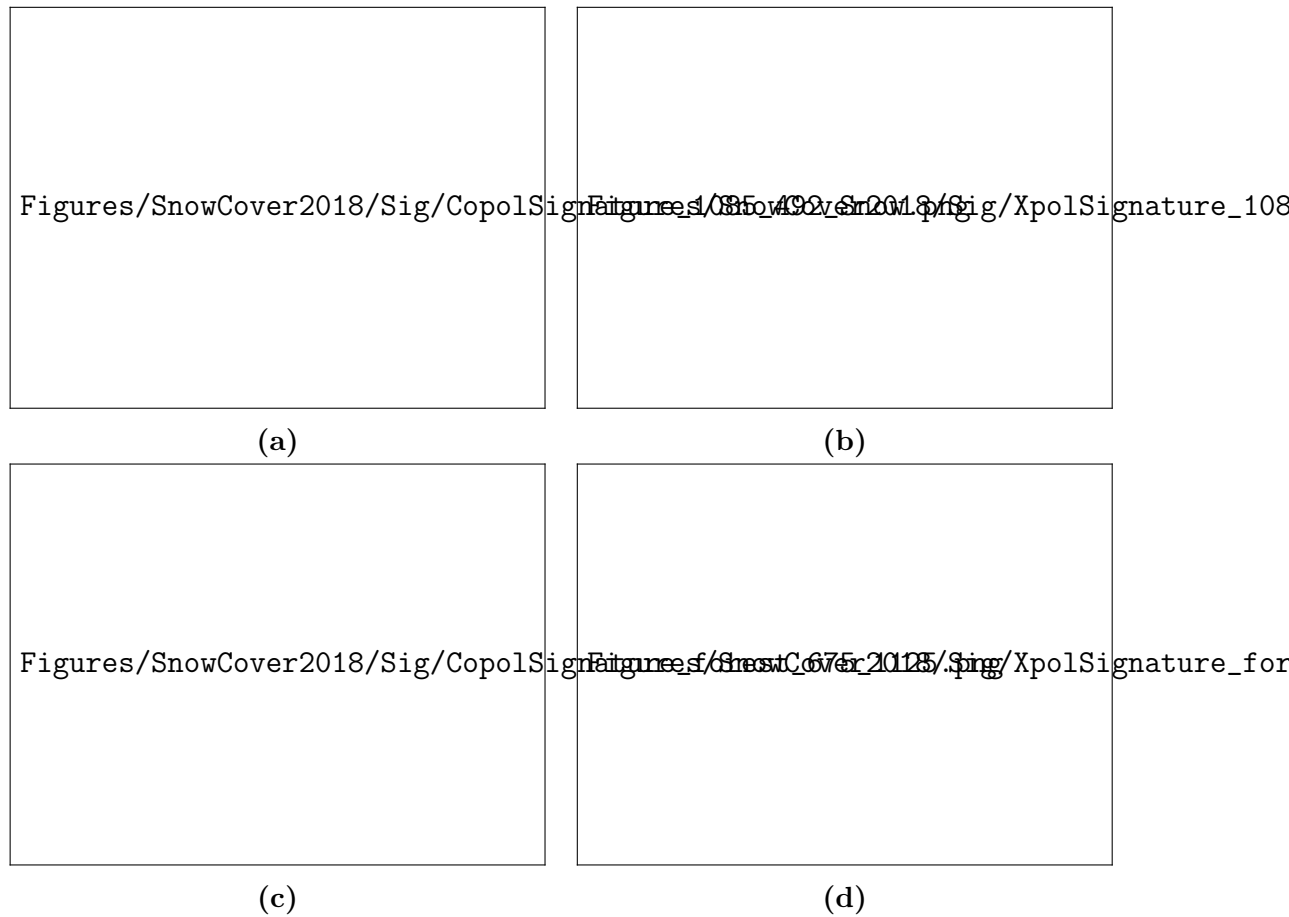
Figure 6.5: Linear Scale plots of (a) Co-pol and (b) Cross-pol polarimetric signatures over snow covered bare ground and (c) Co-pol and (d) Cross-pol polarimetric signatures over forested areas.
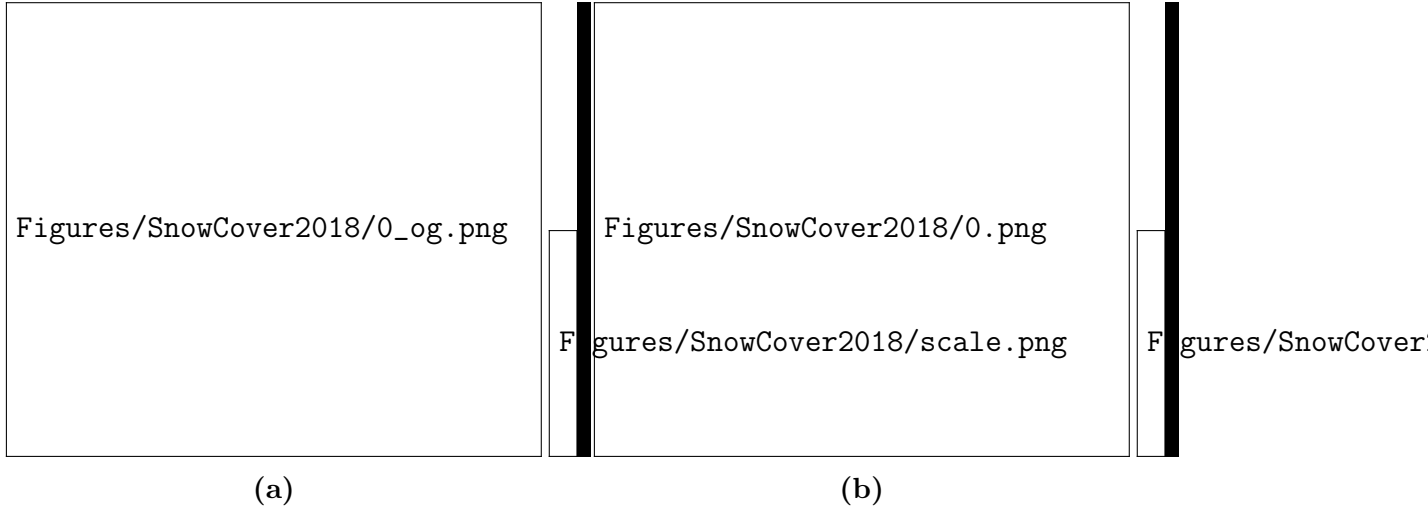
Figure 6.6: A comparison of the (a) original and (b) reconstructed M_{11} element from the AE shown normalized between 0 to 1.

The original and re-constructed M_{11} element is presented in Fig. 6.6. As seen, the reconstruction X' closely follows the original input X, but, with a non-liner transformation. This transformation and happens independently on each dimension of the input. This leads to the generation of a representational feature manifold, which has improved separability for classification over the original input space. For visualization and understanding the action of the AE, two input space features are selected and presented in a scatter plot, along with two features from the representational space in Figure 6.7. The feature combinations are selected at random.

It can be seen in Figure 6.8 that snow is undetected in the forested zone located towards the lower elevations of the area. The results correlate well with the snow-cover maps derived from the optical Landsat-8 data by the NDSI technique. Due to the inherent imaging differences between radar and optical sensors there is disparity in the observed percentage snow-covered/snow-free areas. The algorithm is able to segregate the snow-covered high elevation areas from the snow-free canopies. Due to differences in the scattering mechanisms over the frozen forest and porous snowpack, the representations captured by the AE were different for both these classes. Distinct class representations enabled the network to perform well in the final classification stage.

The primary difference in the snow-cover map obtained from the optical and radar images occurred as a result of layover/shadow effects and selective penetration in radar images. Selective penetration can modify the scattering signatures which can consequently result in training with erroneous representation and thus a misclassification. The images were also acquired on different dates which contributes towards the mismatch.

The difference in the two snow cover maps obtained from the combination of linear polarization and that of the circular polarization are more evident over the forest areas. This can be attributed to the difference in the received scattering mechanism observed

(a)

(b)

(c)

(d)
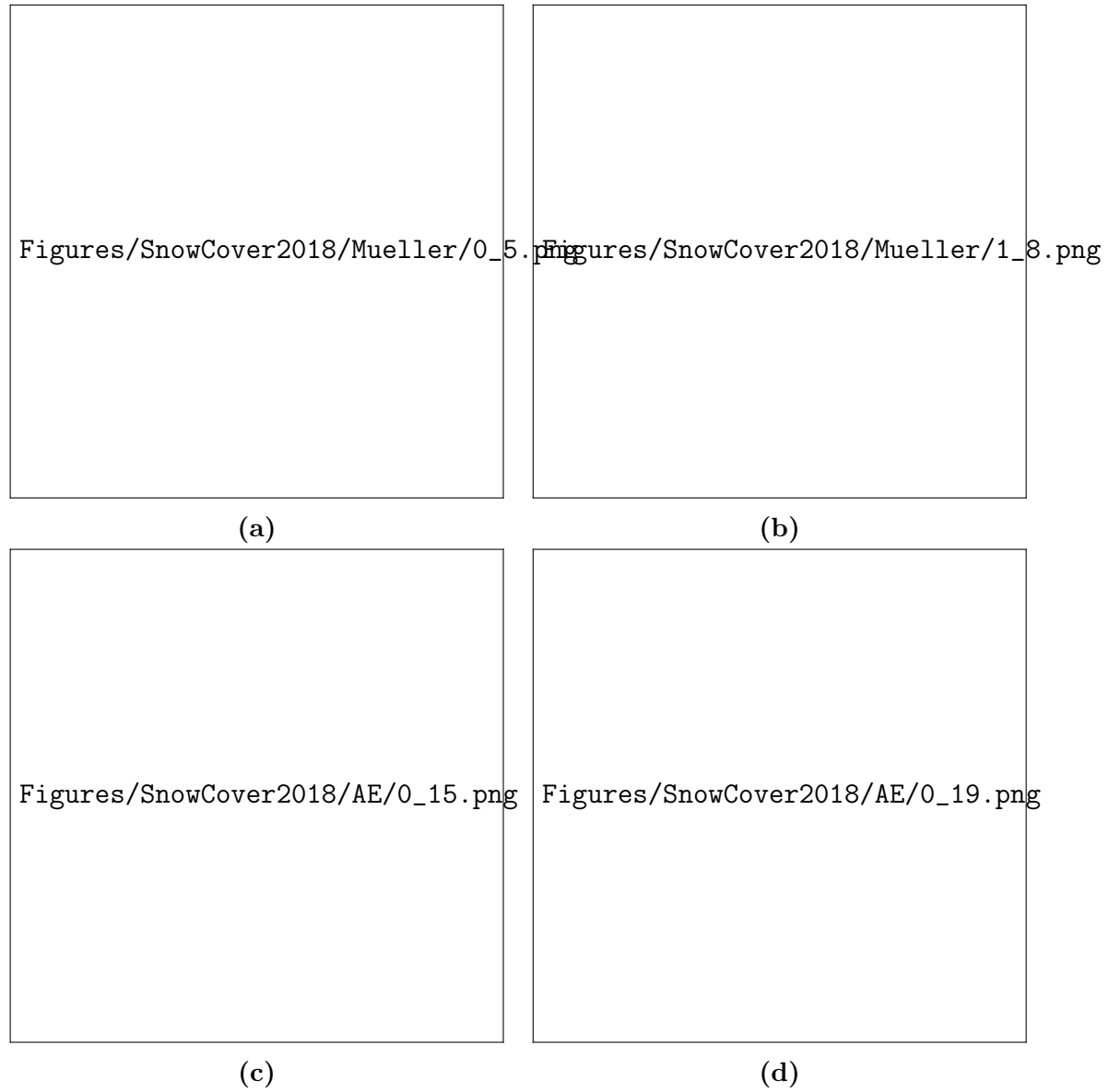
Figure 6.7: A comparison of some (a,b) Mueller (c,d) AE Representation layer features extracted from the full polarimetric image.

Figures/SnowCover2018/FCC.png

Figures/SnowCover2018/IncAng2.png

(a)

(b)

Figures/SnowCover2018/Hyb.png

Figures/SnowCover2018/Lin.png

(c)

(d)

Figures/SnowFigvere20(S8/SnCoverp2018/NoSnow.png
Snow   No-Snow/Shadow
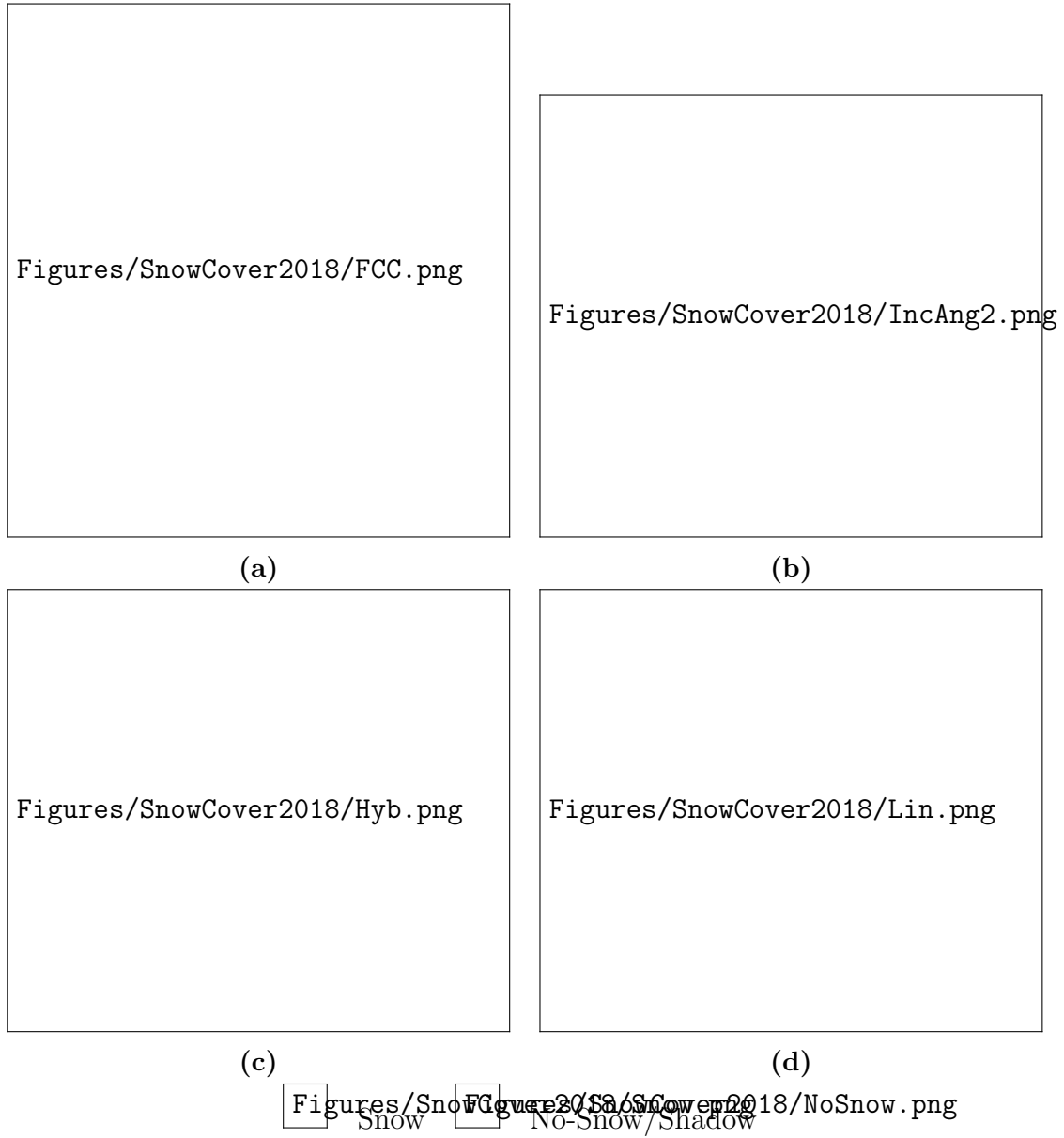
Figure 6.8: (a) False color composite of the area collected by Landsat-8 and the (b) DSM of the area. Resultant snow maps using (c) hybrid polarimetric and (d) full polarimetric data.

at the sensor for the two different incident waves. This could be predominantly because of the scattering types prevailing over the respective areas.

However in the snow laden bare fields, the two different polarizations show similar results. It would be interesting to find an optimized incident polarization that might optimize performance in snow-cover mapping.

Chapter 7 Summary and Future Work

A survey of developments in deep learning to remote-sensing and SAR classification techniques and a review of the state of the art has been presented. We have explored the application of deep learning to address the challenges associated with SAR classification, especially polarimetric SAR. The problem of orientation compensation for rotated targets was identified as a major source of misclassification of urban structures, which could be potentially be alleviated. We applied incremental rotations to the PolSAR data to generate synthetic data-sets. Subsequently, using an auto-encoder, we learn a sparse representation of the entire generated data-set. The feature space thus incorporates the complete target information, including the effect of rotation. These features are then applied to an MLP network which performs the final classification. The resultant classification is significantly better than the commonly used classification algorithms. The output of this, the urban classification can be used in the study of urban sprawl with temporal data.

Further, a weakly supervised change detection technique was developed. Often sufficient ground truth data is not available to use traditional supervised machine learning techniques. A novel Deep Learning based weakly-supervised framework for urban change detection using multi-temporal polarimetric SAR data is developed. A modified unsupervised stacked auto-encoder stage is used to learn an efficient representation of the multi-temporal polarimetric information. Then a label aggregation is performed in the feature space before classification by a multi-layer perceptron. The proposed methodology is validated on an L-band UAVSAR dataset acquired over Los Angeles, CA and performs accurately and effectively with a low false alarm rate.

Subsequently, a novel tensorization framework in conjunction with an ANN algorithm is proposed to leverage the increased information content present in multi-frequency PolSAR data. The Kronecker product is used to combine information from multi-frequency PolSAR data. The resultant combination is efficiently utilized by an ANN. This is demonstrated on a C-, L-, P-band data-set for crop and forest classification. Pairwise tensor combinations of frequency bands (CL, CP, LP) performs better than the individual band, which in turn, is outperformed by the tensorized triplet (CLP). This shows that the tensorized combination of frequency bands have increased information content, which leads to a better classification performance when used in conjunction with an ANN architecture. Moreover, the CLP combination outperforms simple band augmentation (CLP +). In the future, this technique can be exploited for multi-temporal and multi-incidence datasets from advanced sensors for improved classification performance.

Finally, a novel Poincaré sphere and AE based algorithm for generation of snow-cover maps from full and hybrid polarimetric data have been demonstrated. The reconstruc-

tion by the auto-encoder of the original parameters are shown and found to be similar, with an increase in dynamic range. The representational layer of the AE is subsequently extracted and given as input to an MLP layer for classification. This representation has improved separability and leads to a less complex final classification layer, and improved snow-cover classification performance. In the future, spatial inputs can be exploited to the AE to account for inter-relations between the pixels. The normalized DSM can also be included to allow incorporation of the topographic features which affects the returned backscatter. Moreover, an optimal transmission polarization can be determined, which can help further improve separability.

In the future is it intended to expand the functionality of the developed techniques with the following additional concepts:

Transfer Learning

To make the classification technique more generally applicable to large volumes of data, it is desirable that supervision is minimized in the process. This can be done by implementing a unsupervised technique or by transferring the knowledge acquired in the training of one image to the subsequent unseen images. The process of transferring the learning can usually result in a higher overall accuracy on the entire set of images, given that the statistics of the underlying data / scene vary only marginally. This is usually true for images acquired in the same season, using the same or compatible sensor and similar geographic region. Sensors like Sentinel have a set collection strategy that allows them to collect temporally and spatially consistent regular SAR datasets globally. This trend will be continued by future SAR missions like the RCM and NISAR, leading to a large volume of data compliant to the aforementioned criterion. A transfer learning based semi-supervised technique will greatly help extract information from the data. It is seen that the method shows a reasonable transfer of learning, and the performance can be improved by fine tuning.

Variational Auto-Encoders

The Variational Auto-Encoder (VAE) is a generative model that uses bayesian inference principles to create examples based on the samples it has been trained with under probabilistic constrains. It uses $a - priori$ information of the underlying probability distribution of the data to sample apply this constraint and thus embed a latent distribution model in the learned representation. Since the statistical properties of SAR data and speckle is well known, with the coherent returns in a multilooked scene following a Gamma distribution and the speckle pattern following an Inverse Gamma distribution. The VAE can be used to generate radar return samples, under the scattering physics constrains that can drastically improve generalization capabilities of the network. Such a method would find application in urban classification to improve performance in the areas oriented away from the radar line of sight.

List of Publications

Publications in Peer-Reviewed Journals

1. ❀ **Shaunak De**, Debanshu Ratha, Dikshya Ratha, Avik Bhattacharya, and Subhasis Chaudhuri (2018). Tensorization of Multi-Frequency PolSAR Data for Classification Using an Auto-Encoder Network. IEEE Geoscience and Remote Sensing Letters. vol. 15, no. 4, 2018.

2. ❀ **Shaunak De**, Lorenzo Bruzzone, Avik Bhattacharya, Francesca Bovolo and Subhasis Chaudhuri (2017). A Novel Technique Based on Deep Learning and a Synthetic Target Database for Classification of Urban Areas in PolSAR Data. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 11, no. 1, pp. 157-170, 2017.

3. ❀ Debanshu Ratha, **Shaunak De**, Turgay Celik and Avik Bhattacharya (2017). Change Detection in Polarimetric SAR Images Using a Geodesic Distance Between Scattering Mechanisms. IEEE Geoscience and Remote Sensing Letters. vol. 14, no. 7, pp. 1066-1070, 2017.

4. Sanjay Shitole, Mayank Sharma, **Shaunak De**, Avik Bhattacharya, Y. S. Rao and B. K. Mohan (2017). Local contrast based adaptive SAR speckle filter. Journal of the Indian Society of Remote Sensing, 45(3), 451-462.

5. Avik Bhattacharya, Alok Porwal, Swinky Dhingra, **Shaunak De**, G. Venkataraman (2015). Remote estimation of dielectric permittivity of lunar surface regolith using compact polarimetric synthetic aperture radar data. Advances in Space Research, 56(11), 2439-2448.

6. Sanjay Shitole, **Shaunak De**, Y.S. Rao, B.K. Mohan and Anup Das (2015). Selection of suitable window size for speckle reduction and deblurring using SOFM in polarimetric SAR images. Journal of the Indian Society of Remote Sensing, 43(4), 739-750.

7. Avik Bhattacharya, Arnab Muhuri, **Shaunak De**, Surendar Manickam and A.C. Frery (2015). Modifying the Yamaguchi four-component decomposition scattering powers using a stochastic distance. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 8, no. 7, pp. 3497-3506, 2015.

8. Sanjay Shitole, **Shaunak De**, Y.S. Rao, B.K. Mohan and Anup Das (2015). MMSE based seed selection in IDAN speckle filter with point target preservation. International Journal of Imaging and Robotics, 15(3), 14-34.

9. Avik Bhattacharya, **Shaunak De**, Arnab Muhuri, Surendar Manickam, G. Venkataraman and Anup Das (2015). A new compact polarimetric SAR decomposition technique. Remote Sensing Letters, 6(12), 914-923.

Publications in International Conference Proceedings

1. ❀ **Shaunak De**, Davide Pirrone, Francesca Bovolo, Lorenzo Bruzzone and Avik Bhattacharya (2017, July). A novel change detection framework based on deep learning for the analysis of multi-temporal polarimetric SAR images. In Geoscience and Remote Sensing Symposium (IGARSS), 2017 IEEE International, pp. 4554-4557.

2. ❀ Davide Pirrone, **Shaunak De**, Francesca Bovolo, Lorenzo Bruzzone and Avik Bhattacharya (2017, July). Unsupervised change detection in built-up areas by multi-temporal polarimetric SAR images. In Geoscience and Remote Sensing Symposium (IGARSS), 2017 IEEE International, pp. 5193-5196.

3. **Shaunak De**, Abhishek Maity, Vritti Goel, Sanjay Shitole, and Avik Bhattacharya. "Predicting the popularity of instagram posts for a lifestyle magazine using deep learning." In Communication Systems, Computing and IT Applications (CSCITA), 2017 2nd International Conference on, pp. 174-177. IEEE, 2017.

4. ❀ Biplab Banerjee, **Shaunak De**, Surendar Manickam, and Avik Bhattacharya. "An unsupervised hidden Markov random field based segmentation of polarimetric SAR images." In Geoscience and Remote Sensing Symposium (IGARSS), 2016 IEEE International, pp. 1536-1539. IEEE, 2016.

5. Gaurav Kumar Dashondhi, Jyotirmoy Mohanty, Laxmi Narayana Eeti, Avik Bhattacharya, **Shaunak De**, and Krishna Mohan Buddhiraju. "A learning tool for optical and microwave satellite image processing and analysis." In Multispectral, Hyperspectral, and Ultraspectral Remote Sensing Technology, Techniques and Applications VI, vol. 9880, p. 988024. International Society for Optics and Photonics, 2016.

6. ❀ **Shaunak De** and Avik Bhattacharya. "Urban classification using PolSAR data and deep learning." In Geoscience and Remote Sensing Symposium (IGARSS), 2015 IEEE International, pp. 353-356.

7. Avik Bhattacharya, Arnab Muhuri, **Shaunak De**, M. Surendar, and G. Venkataraman. "A New Target Decomposition Technique for Compact Polarimetric SAR Data." In Lunar and Planetary Science Conference, vol. 46, p. 1065. 2015.

8. Avik Bhattacharya, M. Surendar, **Shaunak De**, Gopalan Venkataraman, and Gulab Singh. "Snow wetness estimation from dual polarimetric coherent TerraSAR-X data." In Geoscience and Remote Sensing Symposium (IGARSS), 2014 IEEE International, pp. 2766-2769.

9. Avik Bhattacharya, Arnab Muhuri, **Shaunak De**, and Alejandro C. Frery. "Orientation angle estimation from PolSAR data using a stochastic distance." In Geoscience and Remote Sensing Symposium (IGARSS), 2014 IEEE International, pp. 3474-3477.

10. Siddharth Hariharan, Siddhesh Tirodkar, **Shaunak De**, and Avik Bhattacharya. "Variable importance and random forest classification using RADARSAT-2 PolSAR data." In Geoscience and Remote Sensing Symposium (IGARSS), 2014 IEEE International, pp. 1210-1213.

11. **Shaunak De**, Vineet Kumar, and Y. S. Rao. "Crop Classification using RISAT-1 Hybrid Polarimetric SAR Data." In EUSAR 2014; 10th European Conference on Synthetic Aperture Radar; Proceedings of, pp. 1-4. VDE, 2014.

12. **Shaunak De**, Rao, Y. S., Vineet Kumar, and Anup Das. "Full and Hybrid Polarimetric SAR Data Analysis for Various Land." International Experts Meet on Microwave Remote Sensing,, Ahmedabad, India Features (2013): 16-17.

13. Varsha Turkar, **Shaunak De**, G. G. Ponnurangam, Rinki Deo, Y. S. Rao, and Anup Das. "Classification of RISAT-1 hybrid polarimetric data for various land features." In Synthetic Aperture Radar (APSAR), 2013 Asia-Pacific Conference on, pp. 494-497.

14. Varsha Turkar, **Shaunak De**, Y. S. Rao, Sanjay Shitole, Avik Bhattacharya, and Anup Das. "Comparative analysis of classification accuracy for RISAT-1 compact polarimetric data for various land-covers." In Geoscience and Remote Sensing Symposium (IGARSS), 2013 IEEE International, pp. 3586-3589.

15. **Shaunak De**, Rinky Deo, G. Raghuram, & Y.S. Rao, Software Development for Polarimetric SAR Data Processing: PolSDP. In International Conference on Microwave Antenna, Propagation & Remote Sensing, 2012.

❋   These publications have evolved from the doctoral dissertation.

Bibliography

[1] J. J. van Zyl, *Synthetic aperture radar polarimetry*, vol. 2. John Wiley & Sons, 2011.

[2] P. Gong, J. Wang, L. Yu, Y. Zhao, Y. Zhao, L. Liang, Z. Niu, X. Huang, H. Fu, S. Liu, *et al.*, "Finer resolution observation and monitoring of global land cover: first mapping results with landsat tm and etm+ data," *International Journal of Remote Sensing*, vol. 34, no. 7, pp. 2607–2654, 2013.

[3] Y. Bengio, "Learning deep architectures for ai," *Foundations and trends® in Machine Learning*, vol. 2, no. 1, pp. 1–127, 2009.

[4] G. Hinton, S. Osindero, and Y.-W. Teh, "A fast learning algorithm for deep belief nets," *Neural computation*, vol. 18, no. 7, pp. 1527–1554, 2006.

[5] M. Mohri, A. Rostamizadeh, and A. Talwalkar, *Foundations of Machine Learning*. Adaptive computation and machine learning series, MIT Press, 2012.

[6] J. Ginsberg, M. H. Mohebbi, R. S. Patel, L. Brammer, M. S. Smolinski, and L. Brilliant, "Detecting influenza epidemics using search engine query data," *Nature*, vol. 457, no. 7232, pp. 1012–1014, 2008.

[7] S. Cook, C. Conrad, A. L. Fowlkes, and M. H. Mohebbi, "Assessing google flu trends performance in the united states during the 2009 influenza virus a (h1n1) pandemic," *PLoS ONE*, vol. 6, p. e23610, 08 2011.

[8] A. Smola and S. Vishwanathan, *Introduction to Machine Learning*. Cambridge University Press, 2008.

[9] Y. Bengio, I. J. Goodfellow, and A. Courville, "Deep learning." Book in preparation for MIT Press, 2014.

[10] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: A review and new perspectives," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 35, no. 8, pp. 1798–1828, 2013.

[11] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in *Computer Vision–ECCV 2014*, pp. 818–833, Springer, 2014.

[12] S. Jin, L. Yang, P. Danielson, C. Homer, J. Fry, and G. Xian, "A comprehensive change detection method for updating the national land cover database to circa 2011," *Remote Sensing of Environment*, vol. 132, pp. 159–175, 2013.

[13] T. Lillesand and R. Kiefer, *Remote sensing and image interpretation*. Wiley, 1979.

[14] C. J. Tucker, J. R. Townshend, and T. E. Goff, "African land-cover classification using satellite data," *Science*, vol. 227, no. 4685, pp. 369–375, 1985.

[15] R. De Fries, M. Hansen, J. Townshend, and R. Sohlberg, "Global land cover classifications at 8 km spatial resolution: the use of training data derived from landsat imagery in decision tree classifiers," *International Journal of Remote Sensing*, vol. 19, no. 16, pp. 3141–3168, 1998.

[16] M. A. Friedl, D. K. McIver, J. C. Hodges, X. Zhang, D. Muchoney, A. H. Strahler, C. E. Woodcock, S. Gopal, A. Schneider, A. Cooper, *et al.*, "Global land cover mapping from modis: algorithms and early results," *Remote Sensing of Environment*, vol. 83, no. 1, pp. 287–302, 2002.

[17] T. Loveland and A. Belward, "The igbp-dis global 1-km land cover data set, discover: first results," *International Journal of Remote Sensing*, vol. 65, no. 9, pp. 1021–1031, 1997.

[18] K. C. Jezek, H. Sohn, and K. Noltimier, "The radarsat antarctic mapping project," in *Geoscience and Remote Sensing Symposium Proceedings, 1998. IGARSS'98. 1998 IEEE International*, vol. 5, pp. 2462–2464, IEEE, 1998.

[19] P. Skvarca, "Changes and surface features of the larsen ice shelf, antarctica, derived from landsat and kosmos mosaics," *Annals of Glaciology*, vol. 20, no. 1, pp. 6–12, 1994.

[20] J.-S. Lee, M. Grunes, and E. Pottier, "Quantitative comparison of classification capability: fully polarimetric versus dual and single-polarization sar," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 39, pp. 2343–2351, Nov 2001.

[21] D. M. Simpson, G. T. Burbidge, A. Thornbury, and B. Ghinelli, "Terrasar-l system," in *Remote Sensing*, pp. 17–26, International Society for Optics and Photonics, 2004.

[22] R. Stöckli, E. Vermote, N. Saleous, R. Simmon, and D. Herring, "The blue marble next generation-a true color earth dataset including seasonal dynamics from modis," *Published by the NASA Earth Observatory*, 2005.

[23] J. Lee and E. Pottier, *Polarimetric Radar Imaging: From Basics to Applications*. Optical Science and Engineering, Taylor & Francis, 2009.

[24] L. Gagnon and A. Jouan, "Speckle filtering of sar images: a comparative study between complex-wavelet-based and standard filters," in *Optical Science, Engineering and Instrumentation'97*, pp. 80–91, International Society for Optics and Photonics, 1997.

[25] E. Jakeman and P. Pusey, "A model for non-rayleigh sea echo," *Antennas and Propagation, IEEE Transactions on*, vol. 24, no. 6, pp. 806–814, 1976.

[26] J. Lee, D. Schuler, R. Lang, and K. Ranson, "K-distribution for multi-look processed polarimetric sar imagery," in *Geoscience and Remote Sensing Symposium, 1994. IGARSS'94. Surface and Atmospheric Remote Sensing: Technologies, Data Analysis and Interpretation., International*, vol. 4, pp. 2179–2181, IEEE, 1994.

[27] A. P. Doulgeris and T. Eltoft, "Scale mixture of gaussian modelling of polarimetric sar data," *EURASIP Journal on Advances in Signal Processing*, vol. 2010, p. 2, 2010.

[28] C. C. Freitas, A. C. Frery, and A. H. Correia, "The polarimetric g distribution for sar data analysis," *Environmetrics*, vol. 16, no. 1, pp. 13–31, 2005.

[29] V. Akbari, T. Eltoft, and A. Doulgeris, "Non-gaussian clustering of sar images for glacier change detection," in *Proc. ESA Living Planet Symp*, pp. 1–8, 2010.

[30] V. N. Vapnik and V. Vapnik, *Statistical learning theory*, vol. 1. Wiley New York, 1998.

[31] A. Y. Ng and M. I. Jordan, "On discriminative vs. generative classifiers: A comparison of logistic regression and naive bayes," in *Advances in Neural Information Processing Systems*, pp. 841–848, 2002.

[32] J. C. Curlander and R. N. McDonough, *Synthetic aperture radar*. John Wiley & Sons New York, NY, USA, 1991.

[33] J. Van Zyl, R. Carande, Y. Lou, T. Miller, and K. Wheeler, "The nasa/jpl three-frequency polarimetric airsar system," 1992.

[34] J. Kong, A. Swartz, H. Yueh, L. Novak, and R. Shin, "Identification of terrain cover using the optimum polarimetric classifier," *Journal of Electromagnetic Waves and Applications*, vol. 2, no. 2, pp. 171–194, 1988.

[35] E. Pottier and J. Saillard, "On radar polarization target decomposition theorems with application to target classification, by using neural network method," in *Antennas and Propagation, 1991. ICAP 91., Seventh International Conference on (IEE)*, pp. 265–268, IET, 1991.

[36] J. A. Anderson, *An introduction to neural networks*. MIT press, 1995.

[37] J.-S. Lee, M. R. Grunes, and R. Kwok, "Classification of multi-look polarimetric sar imagery based on complex wishart distribution," *International Journal of Remote Sensing*, vol. 15, no. 11, pp. 2299–2311, 1994.

[38] J.-S. Lee, M. R. Grunes, and E. Pottier, "Quantitative comparison of classification capability: fully polarimetric versus dual and single-polarization sar," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 39, no. 11, pp. 2343–2351, 2001.

[39] J.-S. Lee, M. R. Grunes, E. Pottier, and L. Ferro-Famil, "Unsupervised terrain classification preserving polarimetric scattering characteristics," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 42, no. 4, pp. 722–731, 2004.

[40] 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 Transactions on Geoscience and Remote Sensing*, vol. 51, no. 7, pp. 4200–4213, 2013.

[41] 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," *International Journal of Remote Sensing*, vol. 35, no. 19, pp. 6859–6873, 2014.

[42] 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 Geoscience and Remote Sensing Letters*, vol. 12, no. 10, pp. 2110–2114, 2015.

[43] A. Bhattacharya and R. Touzi, "Polarimetric sar urban classification using the touzi target scattering decomposition," *Canadian Journal of Remote Sensing*, vol. 37, no. 4, pp. 323–332, 2012.

[44] X. Niu and Y. Ban, "Multi-temporal radarsat-2 polarimetric sar data for urban land-cover classification using an object-based support vector machine and a rule-based approach," *International Journal of Remote Sensing*, vol. 34, no. 1, pp. 1–26, 2013.

[45] C. da Costa Freitas, L. de Souza Soler, S. J. S. Sant'Anna, L. V. Dutra, J. R. Dos Santos, J. C. Mura, and A. H. Correia, "Land use and land cover mapping in the brazilian amazon using polarimetric airborne p-band sar data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 46, no. 10, pp. 2956–2970, 2008.

[46] 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 Processing Magazine*, vol. 31, no. 4, pp. 79–89, 2014.

[47] 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 Transactions on Geoscience and Remote Sensing*, vol. 42, pp. 2046–2057, Oct 2004.

[48] F. Dell'Acqua and P. Gamba, "Texture-based characterization of urban environments on satellite sar images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 41, pp. 153–159, Jan 2003.

[49] M. R. Azimi-Sadjadi, S. Ghaloum, and R. Zoughi, "Terrain classification in sar images using principal components analysis and neural networks," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 31, pp. 511–515, Mar 1993.

[50] 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 Transactions on Geoscience and Remote Sensing*, vol. 52, no. 9, pp. 5256–5270, 2014.

[51] 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 Transactions on Geoscience and Remote Sensing*, vol. 41, pp. 2338–2353, Oct 2003.

[52] Y. C. Tzeng and K. S. Chen, "A fuzzy neural network to sar image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 36, pp. 301–307, Jan 1998.

[53] P. Gamba and F. Dell'Acqua, "Increased accuracy multiband urban classification using a neuro-fuzzy classifier," *International Journal of Remote Sensing*, vol. 24, no. 4, pp. 827–834, 2003.

[54] 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 Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 1, pp. 87–97, June 2008.

[55] P. Gamba and B. Houshmand, "An efficient neural classification chain of sar and optical urban images," *International Journal of Remote Sensing*, vol. 22, no. 8, pp. 1535–1553, 2001.

[56] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 1026–1034, 2015.

[57] L. Deng, J. Li, J.-T. Huang, K. Yao, D. Yu, F. Seide, M. Seltzer, G. Zweig, X. He, J. Williams, *et al.*, "Recent advances in deep learning for speech research at microsoft," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 8604–8608, IEEE, 2013.

[58] X. Glorot, A. Bordes, and Y. Bengio, "Domain adaptation for large-scale sentiment classification: A deep learning approach," in *Proceedings of the 28th International Conference on Machine Learning (ICML-11)*, pp. 513–520, 2011.

[59] V. Mnih, K. Kavukcuoglu, D. Silver, A. Graves, I. Antonoglou, D. Wierstra, and M. Riedmiller, "Playing atari with deep reinforcement learning," *arXiv preprint arXiv:1312.5602*, 2013.

[60] 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 Transactions on Geoscience and Remote Sensing*, vol. 53, no. 6, pp. 3325–3337, 2015.

[61] 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 Transactions on Geoscience and Remote Sensing*, vol. 53, no. 3, pp. 1174–1185, 2015.

[62] X. Chen, S. Xiang, C.-L. Liu, and C.-H. Pan, "Vehicle detection in satellite images by hybrid deep convolutional neural networks," *IEEE Geoscience and remote sensing letters*, vol. 11, no. 10, pp. 1797–1801, 2014.

[63] Y. Chen, Z. Lin, X. Zhao, G. Wang, and Y. Gu, "Deep learning-based classification of hyperspectral data," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 7, no. 6, pp. 2094–2107, 2014.

[64] 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 Sensing of Environment*, vol. 117, pp. 72–82, 2012.

[65] P. Gamba and M. Aldrighi, "Sar data classification of urban areas by means of segmentation techniques and ancillary optical data," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 5, pp. 1140–1148, Aug 2012.

[66] L. Zhang, G.-S. Xia, T. Wu, L. Lin, and X. C. Tai, "Deep learning for remote sensing image understanding," *Journal of Sensors*, vol. 501, p. 173691, 2015.

[67] Y. Li, X. Xu, B.-D. Bai, and Y.-N. Zhang, "Remote sensing image fusion based on fast discrete curvelet transform," in *2008 International Conference on Machine Learning and Cybernetics*, vol. 1, pp. 106–109, IEEE, 2008.

[68] Y. Yamaguchi, A. Sato, W.-M. Boerner, R. Sato, and H. Yamada, "Four-component scattering power decomposition with rotation of coherency matrix," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 49, pp. 2251–2258, June 2011.

[69] P. Liu and Y. Q. Jin, "A study of ship rotation effects on sar image," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, pp. 3132–3144, June 2017.

[70] S.-W. Chen, X.-S. Wang, and M. Sato, "Uniform polarimetric matrix rotation theory and its applications," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 8, pp. 4756–4770, 2014.

[71] Y. Yamaguchi, A. Sato, W.-M. Boerner, R. Sato, and H. Yamada, "Four-component scattering power decomposition with rotation of coherency matrix," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 49, no. 6, pp. 2251–2258, 2011.

[72] B. Zou, D. Lu, L. Zhang, and W. M. Moon, "Eigen-decomposition-based four-component decomposition for polsar data," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 9, no. 3, pp. 1286–1296, 2016.

[73] 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 Transactions on Geoscience and Remote Sensing*, vol. 55, pp. 3396–3407, June 2017.

[74] J.-S. Lee, M. Grunes, E. Pottier, and L. Ferro-Famil, "Unsupervised terrain classification preserving polarimetric scattering characteristics," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 42, pp. 722–731, April 2004.

[75] Z. Li and J. Bethel, "Image coregistration in sar interferometry," *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 37, pp. 433–438, 2008.

[76] Y. Bazi, L. Bruzzone, and F. Melgani, "An unsupervised approach based on the generalized gaussian model to automatic change detection in multitemporal sar images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 4, pp. 874–887, 2005.

[77] L. Bruzzone and D. F. Prieto, "Automatic analysis of the difference image for unsupervised change detection," *IEEE Transactions on Geoscience and Remote sensing*, vol. 38, no. 3, pp. 1171–1182, 2000.

[78] D. Pirrone, F. Bovolo, and L. Bruzzone, "A novel framework for change detection in bi-temporal polarimetric sar images," in *SPIE Remote Sensing*, pp. 100040Z–100040Z, International Society for Optics and Photonics, 2016.

[79] A. Marino, S. R. Cloude, and J. M. Lopez-Sanchez, "A new polarimetric change detector in radar imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 51, pp. 2986–3000, May 2013.

[80] K. Conradsen, A. A. Nielsen, J. Schou, and H. Skriver, "A test statistic in the complex wishart distribution and its application to change detection in polarimetric sar data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 41, no. 1, pp. 4–19, 2003.

[81] L. Zhang, L. Zhang, and B. Du, "Deep learning for remote sensing data: A technical tutorial on the state of the art," *IEEE Geoscience and Remote Sensing Magazine*, vol. 4, no. 2, pp. 22–40, 2016.

[82] K. S. Chen, W. P. Huang, D. H. Tsay, and F. Amar, "Classification of multifrequency polarimetric sar imagery using a dynamic learning neural network," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 34, pp. 814–820, May 1996.

[83] H. McNairn, J. Shang, X. Jiao, and C. Champagne, "The contribution of alos palsar multipolarization and polarimetric data to crop classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 47, no. 12, pp. 3981–3992, 2009.

[84] F. Liu, J. Shi, L. Jiao, H. Liu, S. Yang, J. Wu, H. Hao, and J. Yuan, "Hierarchical semantic model and scattering mechanism based polsar image classification," *Pattern Recognition*, vol. 59, pp. 325–342, 2016.

[85] J. Shi, L. Li, F. Liu, L. Jiao, H. Liu, S. Yang, L. Liu, and H.-X. Hao, "Unsupervised polarimetric synthetic aperture radar image classification based on sketch map and adaptive markov random field," *J. of Applied Remote Sens.*, vol. 10, pp. 10 – 10 – 20, 2016.

[86] S. Baronti, F. Del Frate, P. Ferrazzoli, S. Paloscia, P. Pampaloni, and G. Schiavon, "Sar polarimetric features of agricultural areas," *International Journal of Remote Sensing*, vol. 16, no. 14, pp. 2639–2656, 1995.

[87] P. Ferrazzoli, S. Paloscia, P. Pampaloni, G. Schiavon, S. Sigismondi, and D. Solimini, "The potential of multifrequency polarimetric sar in assessing agricultural and arboreous biomass," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 35, no. 1, pp. 5–17, 1997.

[88] S. Fukuda and H. Hirosawa, "A wavelet-based texture feature set applied to classification of multifrequency polarimetric sar images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 37, no. 5, pp. 2282–2286, 1999.

[89] L. Ferro-Famil, E. Pottier, and J.-S. Lee, "Unsupervised classification of multifrequency and fully polarimetric sar images based on the h/a/alpha-wishart classifier," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 39, pp. 2332–2342, Nov 2001.

[90] D. H. Hoekman and M. A. Vissers, "A new polarimetric classification approach evaluated for agricultural crops," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 41, no. 12, pp. 2881–2889, 2003.

[91] S. Tebaldini, "Algebraic synthesis of forest scenarios from multibaseline polinsar data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 47, no. 12, pp. 4132–4142, 2009.

[92] A. Cichocki, D. Mandic, L. De Lathauwer, G. Zhou, Q. Zhao, C. Caiafa, and H. A. Phan, "Tensor decompositions for signal processing applications: From two-way to multiway component analysis," *IEEE Signal Processing Magazine*, vol. 32, no. 2, pp. 145–163, 2015.

[93] P. A. Rosen, S. Hensley, S. Shaffer, L. Veilleux, M. Chakraborty, T. Misra, R. Bhan, V. R. Sagi, and R. Satish, "The nasa-isro sar mission-an international space partnership for science and societal benefit," in *Radar Conference (RadarCon), 2015 IEEE*, pp. 1610–1613, IEEE, 2015.

[94] M. Slater, D. Sloggett, W. Rees, and A. Steel, "Potential operational multi-satellite sensor mapping of snow cover in maritime sub-polar regions," *International Journal of Remote Sensing*, vol. 20, no. 15-16, pp. 3019–3030, 1999.

[95] N. C. Grody, "Classification of snow cover and precipitation using the special sensor microwave imager," *Journal of Geophysical Research: Atmospheres*, vol. 96, no. D4, pp. 7423–7435, 1991.

[96] J. Shi, J. Dozier, and H. Rott, "Snow mapping in alpine regions with synthetic aperture radar," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 32, no. 1, pp. 152–158, 1994.

[97] J. Koskinen, S. Metsämäki, J. Grandell, S. Jänne, L. Matikainen, and M. Hallikainen, "Snow monitoring using radar and optical satellite data," *Remote Sensing of Environment*, vol. 69, no. 1, pp. 16–29, 1999.

[98] T. Strozzi, U. Wegmuller, and C. Matzler, "Mapping wet snowcovers with sar interferometry," *International Journal of Remote Sensing*, vol. 20, no. 12, pp. 2395–2403, 1999.

[99] M. Surendar, G. Singh, A. Bhattacharya, G. Venkataraman, and P. Bharathi, "Improved snow wetness estimation from fully polarimetric sar image," in *Synthetic Aperture Radar (APSAR), 2013 Asia-Pacific Conference on*, pp. 31–34, IEEE, 2013.

[100] G. Singh, G. Venkataraman, Y. Yamaguchi, and S.-E. Park, "Capability assessment of fully polarimetric alos–palsar data for discriminating wet snow from other scattering types in mountainous regions," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 52, no. 2, pp. 1177–1196, 2014.

[101] H. Rott and C. Mätzler, "Possibilities and limits of synthetic aperture radar for snow and glacier surveying," *Annals of Glaciology*, vol. 9, pp. 195–199, 1987.

[102] A. Martini, L. Ferro-Famil, E. Pottier, and J.-P. Dedieu, "Dry snow discrimination in alpine areas from multi-frequency and multi-temporal sar data," *IEE Proceedings-Radar, Sonar and Navigation*, vol. 153, no. 3, pp. 271–278, 2006.

[103] N. Usami, A. Muhuri, A. Bhattacharya, and A. Hirose, "Polsar wet snow mapping with incidence angle information," *IEEE Geoscience and Remote Sensing Letters*, vol. 13, no. 12, pp. 2029–2033, 2016.

[104] A. Muhuri, S. Manickam, and A. Bhattacharya, "Scattering mechanism based snow cover mapping using RADARSAT-2 c-band polarimetric SAR data," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 10, pp. 3213–3224, July 2017.

[105] A. Muhuri, D. Ratha, and A. Bhattacharya, "Seasonal snow cover change detection over the indian himalayas using polarimetric SAR images," *IEEE Geoscience and Remote Sensing Letters*, vol. 14, pp. 2340–2344, Dec 2017.

[106] R. Touzi, "Target scattering decomposition in terms of roll-invariant target parameters," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 45, no. 1, pp. 73–84, 2007.

[107] W. Hu, Y. Huang, L. Wei, F. Zhang, and H. Li, "Deep convolutional neural networks for hyperspectral image classification," *Journal of Sensors*, vol. 2015, 2015.

[108] Y. Chen, X. Zhao, and X. Jia, "Spectral–spatial classification of hyperspectral data based on deep belief network," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 8, no. 6, pp. 2381–2392, 2015.

[109] H. Kimura, "Radar polarization orientation shifts in built-up areas," *IEEE Geoscience and Remote Sensing Letters*, vol. 5, no. 2, pp. 217–221, 2008.

Bibliography

[110] K. Tragl, "Polarimetric radar backscattering from reciprocal random targets," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 28, no. 5, pp. 856–864, 1990.

[111] J.-S. Lee, L. Jurkevich, P. Dewaele, P. Wambacq, and A. Oosterlinck, "Speckle filtering of synthetic aperture radar images: A review," *Remote Sensing Reviews*, vol. 8, no. 4, pp. 313–340, 1994.

[112] 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 Transactions on Geoscience and Remote Sensing*, vol. 40, pp. 30–41, Jan 2002.

[113] A. Bhattacharya, A. Muhuri, S. De, S. Manickam, and A. Frery, "Modifying the yamaguchi four-component decomposition scattering powers using a stochastic distance," *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of*, vol. 8, pp. 3497–3506, July 2015.

[114] A. Hirose, *Complex-valued neural networks*. Springer Science & Business Media, 2006.

[115] R. Barakat, "Bilinear constraints between elements of the 4 x 4 mueller-jones transfer matrix of polarization theory," *Optics Communications*, vol. 38, no. 3, pp. 159–161, 1981.

[116] X. Glorot, A. Bordes, and Y. Bengio, "Deep sparse rectifier neural networks," in *International Conference on Artificial Intelligence and Statistics*, pp. 315–323, 2011.

[117] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, pp. 1097–1105, 2012.

[118] V. Nair and G. E. Hinton, "Rectified linear units improve restricted boltzmann machines," in *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*, pp. 807–814, 2010.

[119] A. Frery, H.-J. Muller, C. Yanasse, and S. Sant'Anna, "A model for extremely heterogeneous clutter," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 35, pp. 648–659, May 1997.

[120] A. C. Frery, A. Correia, C. D. Rennó, C. Freitas, J. Jacobo-Berlles, M. Mejail, K. Vasconcellos, and S. Santanna, "Models for synthetic aperture radar image analysis," *Resenhas (IME-USP)*, vol. 4, no. 1, pp. 45–77, 1999.

[121] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *International Conference on Artificial Intelligence and Statistics*, pp. 249–256, 2010.

[122] A. G. Fore, B. D. Chapman, B. P. Hawkins, S. Hensley, C. E. Jones, T. R. Michel, and R. J. Muellerschoen, "Uavsar polarimetric calibration," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 6, pp. 3481–3491, 2015.

[123] C. Elachi, Y. Kuga, K. McDonald, K. Sarabandi, F. Ulaby, M. Whitt, H. Zebker, and J. van Zyl, "Radar polarimetry for geoscience applications," 1990.

[124] D. K. Atwood, D. Small, and R. Gens, "Improving polsar land cover classification with radiometric correction of the coherency matrix," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 5, no. 3, pp. 848–856, 2012.

[125] M. Tao, F. Zhou, Y. Liu, and Z. Zhang, "Tensorial independent component analysis-based feature extraction for polarimetric sar data classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 5, pp. 2481–2495, 2015.

[126] B. Banerjee, A. Bhattacharya, and K. M. Buddhiraju, "A generic land-cover classification framework for polarimetric sar images using the optimum touzi decomposition parameter subsetan in-

sight on mutual information-based feature selection techniques," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 7, no. 4, pp. 1167–1176, 2014.

[127] J.-S. Lee and T. Ainsworth, "The effect of orientation angle compensation on coherency matrix and polarimetric target decompositions," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 49, pp. 53–64, Jan 2011.

[128] Z. Liu, M. J. Zuo, and H. Xu, "A gaussian radial basis function based feature selection algorithm," in *Computational Intelligence for Measurement Systems and Applications (CIMSA), 2011 IEEE International Conference on*, pp. 1–4, IEEE, 2011.

[129] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Ssstrunk, "Slic superpixels compared to state-of-the-art superpixel methods," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, pp. 2274–2282, Nov 2012.

[130] D. Ratha, S. De, T. Celik, and A. Bhattacharya, "Change detection in polarimetric sar images using a geodesic distance between scattering mechanisms," *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 7, pp. 1066–1070, 2017.

[131] M. D. Zeiler and R. Fergus, "Stochastic pooling for regularization of deep convolutional neural networks," *arXiv preprint arXiv:1301.3557*, 2013.

[132] M. Vissers and J. Sanden, *Groundtruth Collection for the JPL-SAR and ERS-1 Campaign in Flevoland and the Veluwe (NL) 1991*. BCRS rapport, Netherlands Remote Sensing Board (BCRS), 1992.

[133] B. You, B. Xu, J. Yang, C. Yeh, and J. Song, "Novel supervised classification approach for multifrequency polarimetric sar data," *Journal of Systems Engineering and Electronics*, vol. 26, pp. 1216–1221, Dec 2015.

[134] M. C. Dobson, F. T. Ulaby, T. LeToan, A. Beaudoin, E. S. Kasischke, and N. Christensen, "Dependence of radar backscatter on coniferous forest biomass," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 30, pp. 412–415, Mar 1992.

[135] F. Shang and A. Hirose, "Quaternion neural-network-based polsar land classification in poincare-sphere-parameter space," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, pp. 5693–5703, Sept 2014.

[136] A. Coates and A. Y. Ng, "The importance of encoding versus training with sparse coding and vector quantization," in *Proceedings of the 28th international conference on machine learning (ICML-11)*, pp. 921–928, 2011.

[137] M. C. Dobson, K. McDonald, F. Ulaby, and J. Way, "Effects of temperature on radar backscatter from boreal forests," in *Geoscience and Remote Sensing Symposium, 1990. IGARSS'90.'Remote Sensing Science for the Nineties'., 10th Annual International*, pp. 2481–2484, IEEE, 1990.

[138] M. Bernier and J.-P. Fortin, "The potential of times series of c-band sar data to monitor dry and shallow snow cover," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 36, no. 1, pp. 226–243, 1998.