

Springer Optimization and Its Applications 199

Maciej Rysz
Arsenios Tsokas
Kathleen M. Dipple
Kaitlin L. Fair
Panos M. Pardalos *Editors*

Synthetic Aperture Radar (SAR) Data Applications



Springer

Springer Optimization and Its Applications

Volume 199

Series Editors

Panos M. Pardalos , *University of Florida*

My T. Thai , *University of Florida*

Honorary Editor

Ding-Zhu Du, *University of Texas at Dallas*

Advisory Editors

Roman V. Belavkin, *Middlesex University*

John R. Birge, *University of Chicago*

Sergiy Butenko, *Texas A&M University*

Vipin Kumar, *University of Minnesota*

Anna Nagurny, *University of Massachusetts Amherst*

Jun Pei, *Hefei University of Technology*

Oleg Prokopyev, *University of Pittsburgh*

Steffen Rebennack, *Karlsruhe Institute of Technology*

Mauricio Resende, *Amazon*

Tamás Terlaky, *Lehigh University*

Van Vu, *Yale University*

Michael N. Vrahatis, *University of Patras*

Guoliang Xue, *Arizona State University*

Yinyu Ye, *Stanford University*

Aims and Scope

Optimization has continued to expand in all directions at an astonishing rate. New algorithmic and theoretical techniques are continually developing and the diffusion into other disciplines is proceeding at a rapid pace, with a spot light on machine learning, artificial intelligence, and quantum computing. Our knowledge of all aspects of the field has grown even more profound. At the same time, one of the most striking trends in optimization is the constantly increasing emphasis on the interdisciplinary nature of the field. Optimization has been a basic tool in areas not limited to applied mathematics, engineering, medicine, economics, computer science, operations research, and other sciences.

The series **Springer Optimization and Its Applications (SOIA)** aims to publish state-of-the-art expository works (monographs, contributed volumes, textbooks, handbooks) that focus on theory, methods, and applications of optimization. Topics covered include, but are not limited to, nonlinear optimization, combinatorial optimization, continuous optimization, stochastic optimization, Bayesian optimization, optimal control, discrete optimization, multi-objective optimization, and more. New to the series portfolio include Works at the intersection of optimization and machine learning, artificial intelligence, and quantum computing.

Volumes from this series are indexed by Web of Science, zbMATH, Mathematical Reviews, and SCOPUS.

Maciej Rysz • Arsenios Tsokas •
Kathleen M. Dipple • Kaitlin L. Fair •
Panos M. Pardalos
Editors

Synthetic Aperture Radar (SAR) Data Applications

 Springer


Editors

Maciej Rysz
Information Systems & Analytics
Miami University
Oxford, OH, USA

Arsenios Tsokas
Industrial & Systems Engineering
University of Florida
Gainesville, FL, USA

Kathleen M. Dipple
U.S. Airforce Research Laboratory
(AFRL)
Eglin AFB, FL, USA

Kaitlin L. Fair
U.S. Air Force Research Laboratory
(AFRL)
Eglin AFB, FL, USA

Panos M. Pardalos 
Industrial & Systems Engineering
University of Florida
Gainesville, FL, USA

ISSN 1931-6828 ISSN 1931-6836 (electronic)
Springer Optimization and Its Applications
ISBN 978-3-031-21224-6 ISBN 978-3-031-21225-3 (eBook)
<https://doi.org/10.1007/978-3-031-21225-3>

Mathematics Subject Classification: 68Txx, 68T07, 62-XX, 62R07

© The Editor(s) (if applicable) and The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

All rights are solely and exclusively licensed by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors, and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

Synthetic Aperture Radar (SAR) Data Applications presents a diverse collection of state-of-the-art applications of SAR data. Its aim is to create a channel of communication of ideas on ongoing and evolving uses and tools employing machine learning, and especially deep learning, methods in a series of SAR data applications. This book comprises a variety of innovative ideas, original works, research results, and reviews from eminent researchers, spanning from target detection and navigation to land classification and interference mitigation.

Synthetic aperture radar (SAR) is a microwave remote sensing technology which was first conceived in the early 1950s. SAR technology has since seen rapid progress. Today, SAR systems are operated from elevated places on land, from manned and unmanned aircraft and spacecraft. SARs can provide images on a 24-h basis and in all kinds of weather and have the ability to penetrate clouds, fog, and, in some cases, leaves, snow, and sand. They generate maps and data describing features of the surface or reflective object. The advent of machine learning in the SAR community created new opportunities and facilitated tasks in SAR data analysis. Machine learning tools offer an ingenuity to existing and new algorithms.

The editors of this book brought together diverse topics with the aim to spotlight interdisciplinary cutting-edge lines of research that can be useful to a wider audience. *Synthetic Aperture Radar (SAR) Data Applications* is addressed to expert practitioners and general public in industry and academia interested in modern practices and applications using SAR data. Individuals or organizations with intent or ongoing efforts that involve machine learning with SAR data are expected to significantly benefit.

Matthew P. Masarik, Chris Kreucher, Kirk Weeks, and Kyle Simpson in the first chapter, “[End-to-End ATR Leveraging Deep Learning](#),” discuss the need for efficient and reliable automatic target recognition (ATR) algorithms that can ingest a SAR image, find all the objects of interest in the image, classify these objects, and output properties of the objects. Their chapter lays out the required steps in any approach for performing these functions and describes a suite of deep learning algorithms which perform this end-to-end SAR ATR.

Luca Bergamasco and Francesca Bovolo in their chapter “[Change Detection in SAR Images Using Deep Learning Methods](#)” focus on unsupervised deep learning change detection methods that exploit unsupervised deep learning models to extract feature maps. The feature maps retrieved from these models are used to detect changed areas in multi-temporal images and handle the speckle noise. Change detection methods address the regular monitoring of target areas by identifying changes over an analyzed area using bi-temporal or multi-temporal SAR images.

Seonho Park, Maciej Rysz, Kathleen Dipple, and Panos Pardalos in their chapter “[Homography Augmented Momentum Contrastive Learning for SAR Image Retrieval](#)” propose a deep learning image retrieval approach using homography transformation augmented contrastive learning to achieve scalable SAR image search tasks. They introduce a training method for the deep neural networks induced by contrastive learning that does not require data labeling, which, in turn, enables tractability of large-scale datasets with relative ease. The effectiveness of their method is demonstrated on polarimetric SAR image datasets.

Alexander Semenov, Maciej Rysz, and Garrett Demeyer in their chapter “[Synthetic Aperture Radar Image Based Navigation Using Siamese Neural Networks](#)” propose Siamese network models with contrastive and triplet loss that can be used for navigational tasks. They use the SqueezeNet deep neural network as their backbone architecture due to its compact size in comparison to other popular architectures that are often used in SAR image processing tasks. Their experiments demonstrate that their method can be used effectively and holds much promise for future navigational tasks.

Jin Xing, Ru Luo, Lifu Chen, Jielan Wang, Xingmin Cai, Shuo Li, Phil Blythe, Yanghanzi Zhang, and Simon Edwards in their chapter “[A Comparison of Deep Neural Network Architectures in Aircraft Detection from SAR Imagery](#)” compare the performance of six popular deep neural networks for aircraft detection from SAR imagery, to verify their performance in tackling the scale heterogeneity, the background interference, and the speckle noise challenges in SAR-based aircraft detection. Their work confirms the value of deep learning in aircraft and serves as a baseline for future deep learning comparison in remote sensing data analytics.

Yan Huang, Lei Zhang, Jie Li, Mingliang Tao, Zhanye Chen, and Wei Hong in their chapter “[Machine Learning Methods for SAR Interference Mitigation](#)” provide a comprehensive study of the interference mitigation techniques applicable for an SAR system. They provide typical signal models for various interference types, together with many illustrative examples from real SAR data. In addition, they analyze advanced signal processing techniques, specifically machine learning methods, for suppressing interferences in detail. They discuss advantages and drawbacks of each approach in terms of their applicability and future trends from the perspective of cognitive and deep learning frameworks.

Mete Ahishali, Serkan Kiranyaz, and Moncef Gabbouj in their chapter “[Classification of SAR Images Using Compact Convolutional Neural Networks](#)” investigate the performance of compact convolutional neural networks that aim for minimum computational complexity and limited annotated data for the classification of SAR images. Their analysis covers commonly used SAR benchmark datasets consisting

of four fully polarimetric, one dual-, and one single-polarized SAR data including both spaceborne and airborne sensors.

Siddharth Hariharan, Dipankar Mandal, Siddhesh Tirodkar, Vineet Kumar, and Avik Bhattacharya in their chapter “[Multi-frequency Polarimetric SAR Data Analysis for Crop Type Classification Using Random Forest](#)” employ multi-frequency (C-, L-, and P-bands) single-date AIRSAR data using random forest–based polarimetric parameter selection for crop separation and classification. In their study, in addition to polarimetric backscattering coefficients, they also analyzed scattering decomposition powers along with the backscattering ratio parameters and found them vital for multi-frequency crop classification.

Emrullah Acar and Mehmet Sirac Ozerdem in their chapter “[Automatic Determination of Different Soil Types via Several Machine Learning Algorithms Employing Radarsat-2 SAR Image Polarization Coefficients](#)” explore several machine learning algorithms—K-Nearest Neighbor, Extreme Learning Machine, and Naive Bayes—by utilizing Radarsat-2 SAR data in a pilot region in the city of Diyarbakir, Turkey. They collect 156 soil samples for classification of two soil types (Clayey and Clayey+Loamy), compute four different Radarsat-2 SAR image polarization coefficients for each soil sample, and utilize these coefficients as inputs in the classification stage.

Andrea Buono, Emanuele Ferrentino, Yu Li, and Carina Regina de Macedo in their thorough chapter “[Ocean and Coastal Area Information Retrieval Using SAR Polarimetry](#)” describe the role played by synthetic aperture radar polarimetry in supporting the observation of oceans and coastal areas using meaningful showcases. They discuss the capability of generating added-value products in the framework of marine oil pollution by means of experiments of actual polSAR data. Furthermore, they demonstrate the ability of polSAR information to assist in continuous monitoring of coastal profiles for vulnerability analysis purposes.

We would like to acknowledge the support of the U.S. Air Force Research Laboratory at Eglin Air Force Base (task order FA8651-21-F-1013 under contract FA8651-19-D-0037) for the realization of this book. We would also like to express our greatest thanks to all the authors of the chapters in this book as well as the reviewers who provided thorough reports. We likewise express our most sincere appreciation to Springer for their assistance during the preparation of this book and especially to Elizabeth Loew for her encouragement.

Oxford, OH, USA
Gainesville, FL, USA
Eglin Air Force Base, FL, USA
Eglin Air Force Base, FL, USA
Gainesville, FL, USA

Maciej Rysz
Arsenios Tsokas
Kathleen M. Dipple
Kaitlin L. Fair
Panos M. Pardalos

Contents

End-to-End ATR Leveraging Deep Learning	1
Matthew P. Masarik, Chris Kreucher, Kirk Weeks, and Kyle Simpson	
Change Detection in SAR Images Using Deep Learning Methods	25
Luca Bergamasco and Francesca Bovolo	
Homography Augmented Momentum Contrastive Learning for SAR Image Retrieval	63
Seonho Park, Maciej Rysz, Kathleen M. Dipple, and Panos M. Pardalos	
Synthetic Aperture Radar Image Based Navigation Using Siamese Neural Networks	79
Alexander Semenov, Maciej Rysz, and Garrett Demeyer	
A Comparison of Deep Neural Network Architectures in Aircraft Detection from SAR Imagery	91
Jin Xing, Ru Luo, Lifu Chen, Jielan Wang, Xingmin Cai, Shuo Li, Phil Blythe, Yanghanzi Zhang, and Simon Edwards	
Machine Learning Methods for SAR Interference Mitigation	113
Yan Huang, Lei Zhang, Jie Li, Mingliang Tao, Zhanye Chen, and Wei Hong	
Classification of SAR Images Using Compact Convolutional Neural Networks	147
Mete Ahishali, Serkan Kiranyaz, and Moncef Gabbouj	
Multi-Frequency Polarimetric SAR Data Analysis for Crop Type Classification Using Random Forest	195
Siddharth Hariharan, Dipankar Mandal, Siddhesh Tirodkar, Vineet Kumar, and Avik Bhattacharya	

Automatic Determination of Different Soil Types via Several Machine Learning Algorithms Employing Radarsat-2 SAR Image Polarization Coefficients 219
Emrullah Acar and Mehmet Sirac Ozerdem

Ocean and Coastal Area Information Retrieval Using SAR Polarimetry 235
Andrea Buono, Emanuele Ferrentino, Yu Li,
and Carina Regina de Macedo

End-to-End ATR Leveraging Deep Learning



Matthew P. Masarik, Chris Kreucher, Kirk Weeks, and Kyle Simpson

1 Overview

Synthetic aperture radar (SAR) systems are widely used for intelligence, surveillance, and reconnaissance purposes. However, unlike electro-optical (EO) images, SAR images are not easily interpreted and therefore have historically required a trained analyst to extract useful information from images. At the same time, the number of high-resolution SAR systems and the amount of data they generate are rapidly increasing, which has resulted in a shortage of analysts available to interpret this vast amount of SAR data. Therefore, there is a significant need for efficient and reliable automatic target recognition (ATR) algorithms that can ingest a SAR image, find all the objects of interest in the image, classify these objects, and output properties of the objects (location, type, orientation, etc.). This chapter lays out the required steps in any approach for performing these functions and describes a suite of deep learning (DL) algorithms that perform this end-to-end SAR ATR. One novel feature of our method is that we rely on only synthetically generated training data, which avoids some of the main pitfalls of other DL approaches to this problem.

M. P. Masarik · C. Kreucher (✉)
KBR, Ann Arbor, MI, USA
e-mail: matthew.masarik@us.kbr.com; chris.kreucher@us.kbr.com; ckreuche@umich.edu

K. Weeks · K. Simpson
Signature Research, Inc., Navarre, FL, USA
e-mail: kweeks@signatureresearchinc.com; ksimpson@signatureresearchinc.com

© The Editor(s) (if applicable) and The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

M. Rysz et al. (eds.), *Synthetic Aperture Radar (SAR) Data Applications*, Springer Optimization and Its Applications 199, https://doi.org/10.1007/978-3-031-21225-3_1

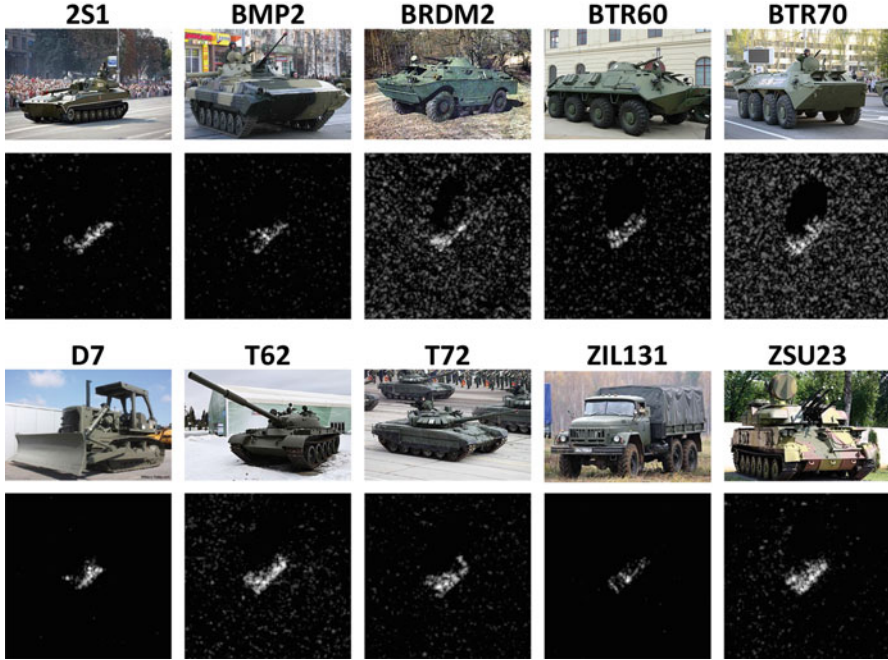


Fig. 1 Comparison of EO images and SAR images of the 10-class MSTAR target set. The significant differences between EO and SAR images make interpretation of SAR imagery non-trivial and traditionally reliant on trained analysts

2 Introduction

Synthetic aperture radar (SAR) is a powerful remote sensing technique that coherently processes a sequence of radar returns to form radar images [1, 2]. SAR is an active imaging mode, transmitting the microwave radio frequency (RF) energy it uses to make images; hence, a SAR sensor can operate day or night in all-weather conditions.

However, SAR images contain a significant amount of speckle noise, and the images have typically much lower resolution than electro-optical (EO) imagery. Moreover, SAR images are sensitive to the direction of illumination because of both shadowing and self-shadowing effects. These features mean that conventional image classification methods applied to EO imagery are not directly applicable to SAR imagery. Figure 1 shows some example EO images and SAR images of the publicly available 10-class Moving and Stationary Target Acquisition and Recognition (MSTAR) [3, 4] target set, providing one illustration of these differences.

Analogously, as shown in Fig. 2, full-scene SAR imagery includes a number of features not present in EO imagery that makes the tasks of target detection, orientation estimation, and classification more difficult. These include residual defocus, non-uniform illumination, and distinct background statistics that stem from



Fig. 2 A SAR image from Mission 78, Pass 1 collected by the Advanced Detection Technology Sensor (ADTS), an airborne SAR/RAR millimeter-wave sensor operated by MIT Lincoln Laboratory. The ADTS operates at Ka band (32.6 to 37 GHz) and was used to collect clutter and armor scenes using its stripmap mode



Fig. 3 A flow diagram describing an example ATR processing chain

the physical nature of the ground cover (e.g., tree cover, water, dirt, road). Because of these factors, SAR imagery has been traditionally interpreted by human experts rather than machine automation.

The number of high-resolution SAR systems and the amount of data they generate are rapidly increasing, which has resulted in a shortage of analysts available to interpret this vast amount of SAR data. There is a significant need, therefore, for efficient and reliable algorithms that can ingest a SAR image, find all the objects of interest in the image, classify these objects, and output properties of the objects (type, size, orientation, etc.). We will refer to this suite of algorithms to as end-to-end SAR ATR (Automated Target Recognition). We note that some authors use the term ATR in reference to just the final step, that of classifying image chips that are known to contain objects of interest. In this chapter, we treat the entire end-to-end problem that contains all aspects starting from a full SAR scene and ending with a collection of classified image chips. A block diagram showing the steps involved in a representative end-to-end SAR ATR algorithm is shown in Fig. 3.

Note that while some ATR algorithms combine a constant false alarm rate (CFAR) detector and target–clutter discriminator into a single algorithm, the algorithm discussed in this chapter splits the detection task into these two stages. Similarly, some classification algorithms do not rely on knowledge of the target orientation, but the algorithm presented here exploits this estimate. Hence, different algorithms will result in slight modifications to the flow diagram in Fig. 3, but the basic structure of detect, characterize, classify, and output underpins all ATR algorithms.

Algorithms for SAR ATR have been studied for many years [5], receiving increased attention in the literature over the past 20 years in large part due to the public release of the high-resolution MSTAR dataset described earlier [3, 4]. The MSTAR dataset consists of SAR image chips of ten military vehicles collected with

an airborne sensor using approximately 600 MHz of bandwidth at X-band with images formed to approximately 1 foot pixel spacing. Because the classification component of SAR ATR is widely regarded as challenging and the MSTAR dataset includes image chips known to contain vehicles, much of the community's focus has been on the classification component of SAR ATR rather than the front-end steps of detection and target characterization.

Many of the early SAR classification algorithms were template-matching approaches, which estimate target class by choosing the class corresponding to the template that best matches the data [6]. Some other early algorithmic approaches include the attributed scattering center model approach [7], support vector machines [8], and neural networks [9]. An excellent summary of the state of the art through approximately 2016 is given in [10], and a updated survey was recently given in [11]. Some of the most important recent efforts employ modern convolutional neural network (CNN) approaches [12–15].

While the MSTAR dataset has provided an excellent testbed for SAR ATR classification that is easily accessible, in recent years it has led to considerable misunderstanding and inflated performance predictions. The issue is that the typical MSTAR experiment presented in the literature uses a training set and a testing set (nominally collected at 17° and 15° elevation, respectively), that are so similar that nearly any reasonable technique should be capable of achieving very high (e.g., $>99\%$) accuracy. The datasets are so similar for several reasons:

1. The targets in the training and testing sets were the exact same vehicle (i.e., the exact same T-72).
2. The targets were on the exact same patch of ground across the training and testing sets.
3. The training and testing data were collected on successive flight passes.
4. The sensors collecting the data and the image formation procedure were identical across the training and testing sets.

This level of similarity between the training and testing sets is unreasonable to expect in practice [11]; hence, the literature tends to exaggerate algorithm performance. Even more damning, however, is that due to the targets being on the same patch of ground across the training and testing sets, the authors have demonstrated that classification accuracies of $>70\%$ are achievable *using only the background clutter of the chips*.

Hence, even if an algorithm demonstrates excellent performance on the standard MSTAR experiment, it is unclear if the algorithm is performing target discrimination that will generalize to more realistic ATR scenarios or is just memorizing particulars of the vehicles and terrain. It is also unclear if the learned algorithms will perform well when applied to data collected by a different sensor.

To address these issues, as well as the relative dearth of data compared to optical images, this chapter describes an end-to-end SAR ATR approach that utilizes synthetically generated training data (e.g., asymptotic ray-tracing predictions based on target CAD models). In addition, our approach is novel as it is a hybrid method that combines the robustness of conventional algorithms (e.g., template matching

and CFAR detectors) with the performance improvements possible using emerging deep learning (DL) techniques.

This chapter proceeds as follows: Sect. 3 describes an approach to the elements of the front-end ATR tasks (target detection and orientation estimation) that combines a conventional CFAR algorithm as a prescreener and a DL algorithm as a final target-clutter discriminator; next, Sect. 4 describes a hybrid conventional/DL approach to target classification that casts a template-matching algorithm into the DL framework and then allows learning to refine the templates (features) and also learns the interclass relationships; and finally, Sect. 5 concludes.

3 Front-End Algorithms

This section discusses the front-end SAR processing chain that begins with a collected SAR image and generates a series of target-centered image chips to be classified, as well as an orientation estimate of the object in the chip. These steps are highlighted in Fig. 4. In our approach, we perform this front-end function by first executing a CFAR detector to separate target-like areas from the background, then carrying out a DL-based target discrimination to separate true targets from target-like clutter, and finally carrying out an orientation estimation to predict the target angle relative to the image.

As is standard in the SAR community, we work with approximately 600 MHz bandwidth data with images formed with approximately 1 ft pixel spacing. For the MSTAR targets and graze angles, chips of size 128×128 are sufficient to capture both the target and its shadow. This chip size is also standard in the literature. All of our algorithms can be scaled appropriately if the chip sizes change either due to resolution or target size.

3.1 Target Detection

A comprehensive review of detection algorithms for SAR ATR can be found in [10]. The target detection algorithm described here employs a two-stage approach. The first stage is a prescreener that is implemented as a cell-averaged constant false alarm rate (CFAR) detector that models the clutter as Rayleigh-distributed. The second stage is a data-driven convolutional neural network (CNN) that predicts a probability that the image contains a target. This algorithm is illustrated in Fig. 5.



Fig. 4 The end-to-end SAR ATR processing chain. This section describes algorithms for the operations outlined in red

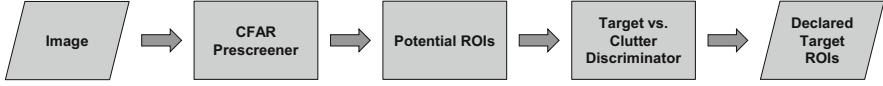


Fig. 5 Flow diagram describing the target detection algorithm

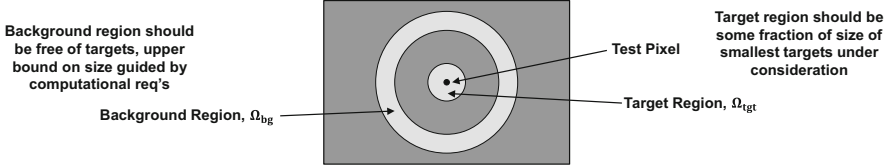


Fig. 6 Definition of target and background regions used in the CFAR

3.1.1 CFAR Prescreener

The prescreener is implemented as a CFAR detector. To derive the CFAR detector, consider Fig. 6, which defines the background region Ω_{bg} and the target region Ω_{tgt} for a test pixel. The regions are defined so that the inner diameter of Ω_{bg} is approximately as large as the largest target to be encountered, the thickness of Ω_{bg} is approximately the expected minimum spacing between targets (though the upper bound on the size of this region is dictated by computational requirements), and the diameter of Ω_{tgt} is approximately the size of the smallest object to be encountered.

The statistics of the non-target (clutter) regions are well studied [16, 17]. Here we have elected to model the clutter as Rayleigh-distributed. The maximum-likelihood estimates (MLEs) of the Rayleigh shape parameter in the background and target regions are given, respectively, in Eqs. (1) and (2):

$$\hat{\sigma}_{bg}^2 = \frac{1}{2|\Omega_{bg}|} \sum_{\Omega_{bg}} |I|^2 \quad (1)$$

$$\hat{\sigma}_{tgt}^2 = \frac{1}{2|\Omega_{tgt}|} \sum_{\Omega_{tgt}} |I|^2. \quad (2)$$

Detection of anomalies is thus reduced to the following hypothesis-testing problem:

$$\begin{aligned} H_0 &: \hat{\sigma}_{tgt}^2 \leq \hat{\sigma}_{bg}^2 \\ H_1 &: \hat{\sigma}_{tgt}^2 > \hat{\sigma}_{bg}^2. \end{aligned} \quad (3)$$

A reasonable test statistic for this problem is

$$T = \frac{\hat{\sigma}_{tgt}^2}{\hat{\sigma}_{bg}^2} = \frac{\frac{1}{2|\Omega_{tgt}|} \sum_{\Omega_{tgt}} |I|^2}{\frac{1}{2|\Omega_{bg}|} \sum_{\Omega_{bg}} |I|^2} = \frac{|\Omega_{bg}| \sum_{\Omega_{tgt}} |I|^2}{|\Omega_{tgt}| \sum_{\Omega_{bg}} |I|^2}. \quad (4)$$

The hypothesis test to declare a detection is then

$$T \underset{H_0}{\overset{H_1}{\gtrless}} \gamma, \quad (5)$$

where γ is some threshold.

It remains then to determine how to set the threshold γ in terms of a desired probability of false alarm (P_{FA}), which requires a model on the distribution of T . First, as commonly done [18], we model the SAR image pixels as 0-mean complex Gaussian with variance σ^2 in I and Q. This gives Rayleigh statistics for the detected pixels, and if the pixel values were independent, the sum over a region would be Gamma-distributed, i.e.,

$$\sum_{\Omega_{bg}} |I|^2 \sim \Gamma\left(|\Omega_{bg}|, \frac{1}{2\sigma^2}\right). \quad (6)$$

However, the SAR image formation process introduces a correlation between the pixels. The exact distribution of the sum of correlated Rayleigh variates is complicated [19, 20], but a useful approximation [21] is

$$\sum_{\Omega_{bg}} \frac{|I|^2}{\sigma^2/2} \sim \Gamma\left(\frac{|\Omega_{bg}|}{u}, 2u\right), \quad (7)$$

where $u = 1 + 2\rho(|\Omega_{bg}| - 1)$, with ρ capturing the average correlation in the region under sum. This is an empirical quantity which we estimate offline from a background dataset. With this approximation, the test statistic T is seen to be a ratio of independent Gamma-distributed variables, which is F -distributed as

$$T \sim F\left(\frac{2|\Omega_{tgt}|}{1 + 2\rho(|\Omega_{tgt}| - 1)}, \frac{2|\Omega_{bg}|}{1 + 2\rho(|\Omega_{bg}| - 1)}\right). \quad (8)$$

Thus, in terms of a desired P_{FA} , the threshold γ can be determined in terms of the inverse F-distribution, and the final hypothesis test is

$$T \underset{H_0}{\overset{H_1}{\gtrless}} F^{-1}\left(1 - P_{FA}, \frac{2|\Omega_{tgt}|}{1 + 2\rho(|\Omega_{tgt}| - 1)}, \frac{2|\Omega_{bg}|}{1 + 2\rho(|\Omega_{bg}| - 1)}\right). \quad (9)$$

Note that this test is performed for every pixel in the image but is still computationally efficient because the summation terms may be computed quickly

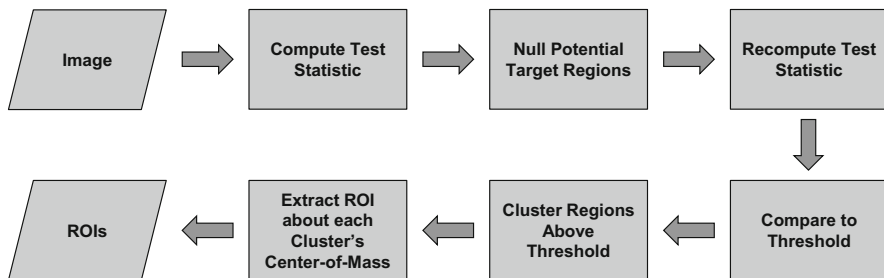


Fig. 7 Flow diagram for the anomaly detector

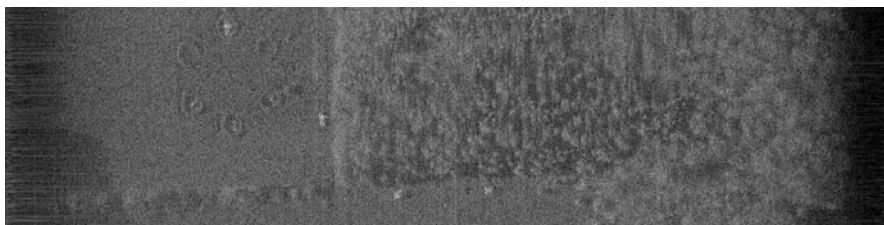


Fig. 8 An example input SAR image from the publicly available Advanced Detection Technology Sensor (ADTS) collection

via convolution. Additional gains in efficiency are possible by making the target and background regions rectangular and using integral images to compute the summations.

Since the background statistics are computed via averaging, they are susceptible to contamination by targets or other objects. This suggests that a censoring procedure should be employed wherein the prescener is run once, and then it is run again while ignoring detections. Figure 7 shows the flow diagram for the anomaly detector.

Figure 8 shows an example SAR image containing an open field populated with four target vehicles (the image is the HH polarization image of frame 22, pass 7, mission 78 from the publicly available ADTS sample set images available on the AFRL Sensor Data Management System). In addition to the targets, the image contains a group of trees on the right side of the image.

Figure 9 shows the (transformed) CFAR test statistic as computed via Eq. (9). As expected, the bright areas correspond to the targets, but the regions around these targets are unnaturally low due to the corruption of the background statistics by the target.

Figure 10 shows the CFAR test statistic that has been recomputed by masking out detections. This results in a significantly cleaner CFAR image.

Finally, Fig. 11 shows detected anomaly regions outlined with a yellow box. Note that, apart from the five detected regions in the trees at the right part of the image, all the other detections are desirable anomaly detections as they correspond to targets.

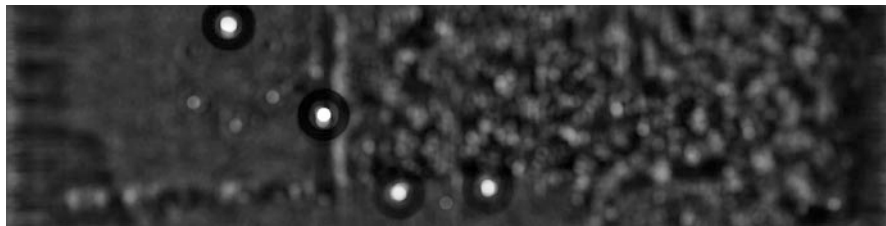


Fig. 9 The test statistic computed on the input SAR image without masking detections



Fig. 10 The test statistic recomputed with initial detections masked

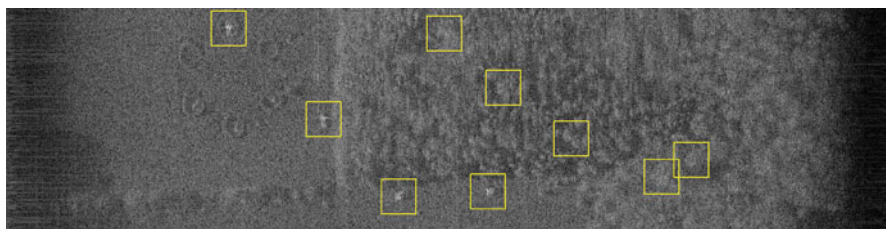


Fig. 11 Image with anomalous regions outlined with a yellow box

A method for dismissing the regions that do not contain targets will be discussed in Sect. 3.1.2.

3.1.2 Target vs. Clutter Discriminator

As seen in Fig. 11, the CFAR will detect some objects that, while they are statistically anomalous, are not targets and therefore should not be passed on for discrimination (e.g., the detections in the trees). This motivates the use of a second stage in the detection algorithm with the purpose of screening out detections that are clearly uninteresting from an ATR perspective. There are many approaches to designing such a screening algorithm (see [22] for a summary of previously applied techniques), but with the recent advances in computer vision via deep learning, it is natural to use deep learning techniques to build a target vs. clutter discrimination algorithm.

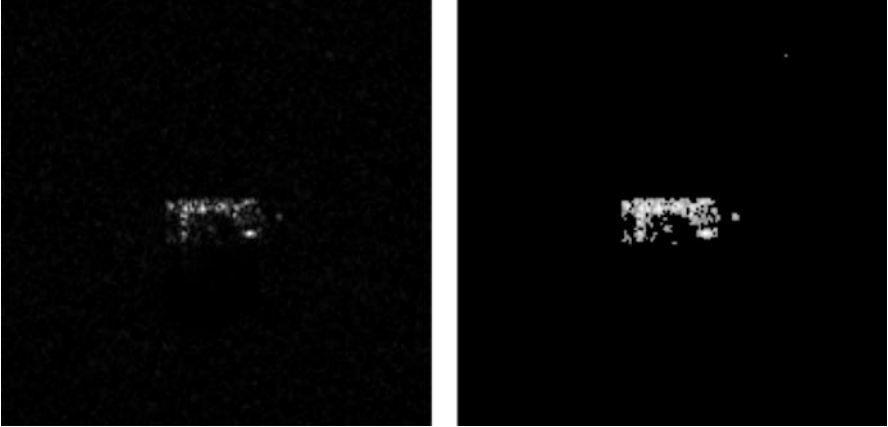


Fig. 12 Comparison of a SAR image chip (left) and preprocessed chip (right)

The target vs. clutter discriminator described here is constructed using a novel CNN architecture trained with synthetic SAR target data and historical measured SAR clutter data to classify a given region of interest (ROI) as containing a target or not. This discriminator is designed to be robust across sensors, terrain, and targets. In addition to training the discriminator on large amounts of diverse target and clutter data and applying data augmentation techniques, robustness is improved by feeding the network thresholded and remapped input images, where only the top- N magnitude pixels are kept and the values are remapped by linearly rescaling the dB -domain image to the interval $[0, 1]$. This preprocessing technique is effective because it maintains the shape of the anomaly detected by the CFAR (which allows for target vs. clutter discrimination), it removes absolute amplitude data (which improves discrimination robustness to sensor, target, and terrain variations), and it maps to a more visually relevant space for the pixel values (which facilitates the use of a CNN). An example target chip and its preprocessed counterpart are shown in Fig. 12. In this example, $N = 400$ and 40 dB of dynamic range was mapped to $[0, 1]$.

The architecture of the target vs. clutter discriminator CNN is shown in Table 1. The CNN is comprised of standard layers ($2D$ convolutions, $2D$ max pooling, batch normalization, dropout, and fully connected layers) and contains 3,548,745 trainable parameters.

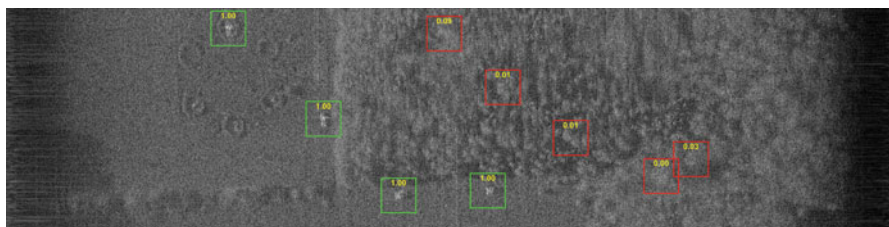
The network is trained using synthetically generated target data and historical clutter data. The network is trained using the binary cross-entropy loss, i.e., for a predicted output y_p and a truth label y_t :

$$\ell(y_p, y_t) = -y_t \log(y_p) - (1 - y_t) \log(1 - y_p). \quad (10)$$

To account for targets that are not centered in the chip, the discriminator is trained using random translational augmentations for each training chip. The ADADELTA optimizer is used to train the discriminator [23].

Table 1 SAR target vs. clutter discriminator CNN architecture

Layer name	Layer parameters	Output size	No. of parameters
Input layer	128×128 input size	128×128	N/A
2D convolution	40 kernels of size 20×20	$109 \times 109 \times 40$	16,040
Batch normalization	N/A	$109 \times 109 \times 40$	160
Dropout	Dropout fraction = 0.2	$109 \times 109 \times 40$	0
2D Max pooling	Pool size = 2×2	$54 \times 54 \times 40$	0
2D convolution	80 kernels of size 15×15	$40 \times 40 \times 80$	720,080
Batch normalization	N/A	$40 \times 40 \times 80$	320
Dropout	Dropout fraction = 0.2	$40 \times 40 \times 80$	0
2D Max pooling	Pool size = 2×2	$20 \times 20 \times 80$	0
2D convolution	160 kernels of size 10×10	$11 \times 11 \times 160$	1,280,160
Batch normalization	N/A	$11 \times 11 \times 160$	640
Dropout	Dropout fraction = 0.2	$11 \times 11 \times 160$	0
2D convolution	320 kernels of size 5×5	$7 \times 7 \times 320$	1,280,320
Batch normalization	N/A	$7 \times 7 \times 320$	1280
Dropout	Dropout fraction = 0.2	$7 \times 7 \times 320$	0
Flatten	N/A	$15,680 \times 1$	0
Fully connected	16 output nodes	16×1	250,896
Batch normalization	N/A	16×1	64
Dropout	Dropout fraction = 0.2	16×1	0
Fully connected	1 output node	1	17

**Fig. 13** Example SAR image with CFAR anomaly detections

The final result of our two-stage detector (CFAR plus target-clutter discriminator) is shown in Fig. 13. Boxes in green have passed both stages, while boxes in red have been flagged by the CFAR stage but rejected at the second stage. In this example, we find that after the second stage, all of the true targets have been detected, while no false targets are detected.

3.2 Target Orientation Estimation

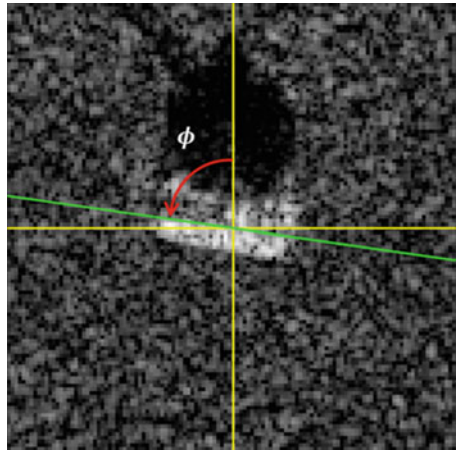
After detection, we next characterize the vehicle by performing target orientation estimation. The goal of the target orientation estimation algorithm is to estimate the aspect angle of the target within the SAR image (c.f., Fig. 14). Robust, accurate estimation of this parameter is important because it can be used to improve target classification performance. In template-matching methods, knowledge of the aspect angle reduces the number of comparisons required to declare a target, which improves efficiency but also reduces the number of potential false matches. In the classification algorithm discussed here, the aspect angle is used to give preference to those nodes of the CNN that correspond to angles close to the input chip’s aspect angle.

There are several traditional image-processing-based algorithms for determining this angle. One such example is based on the Radon transform that preprocesses the image, computes the discrete Radon transform of the image [24], and then selects the angle corresponding to the maximum of the Radon transform of the image. While these algorithms perform well in some cases, their performance leaves much to be desired, especially for target images without a prominent edge. To address this deficiency, we have developed a CNN approach to aspect angle estimation. A high-level description of the algorithm is shown in Fig. 15.

The appropriate loss function for the CNN is not obvious due to angle wrapping. Moreover, the authors have found that achieving accurate angle estimation better than modulo 180° is unrealistic due to vehicle symmetry and a lack of detail in SAR images. Hence, the loss function that was used for the angle estimation algorithm is

$$\ell(\theta_{est}, \theta_{true}) = |\sin(\theta_{est} - \theta_{true})|. \quad (11)$$

Fig. 14 The orientation angle ϕ is useful for the classification algorithm that is the final stage of the end-to-end algorithm



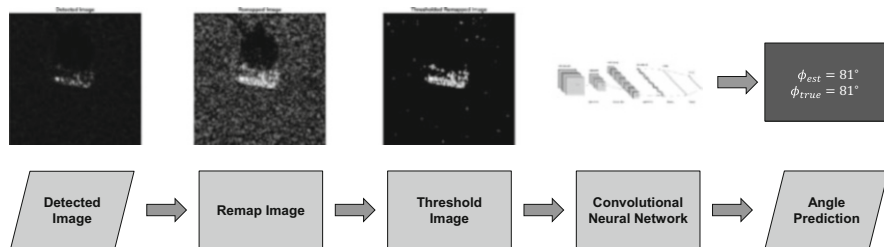


Fig. 15 High-level description of the orientation angle estimation algorithm

Table 2 Architecture of the aspect angle estimation CNN

Layer name	Layer parameters	Output size	No. of parameters
Input layer	128×128 input size	128×128	N/A
2D convolution	10 kernels of size 40×40	$89 \times 89 \times 10$	16,010
Batch normalization	N/A	$89 \times 89 \times 10$	40
Dropout	Dropout fraction = 0.2	$89 \times 89 \times 10$	0
2D convolution	20 kernels of size 40×40	$70 \times 70 \times 20$	80,020
2D Max pooling	Pool size = 2×2	$35 \times 35 \times 20$	0
Batch normalization	N/A	$35 \times 35 \times 20$	80
Dropout	Dropout fraction = 0.2	$35 \times 35 \times 20$	0
2D convolution	30 kernels of size 10×10	$26 \times 26 \times 30$	60,030
Batch normalization	N/A	$26 \times 26 \times 30$	120
Dropout	Dropout fraction = 0.2	$26 \times 26 \times 30$	0
2D convolution	40 kernels of size 5×5	$22 \times 22 \times 40$	30,040
Batch normalization	N/A	$22 \times 22 \times 40$	60
Dropout	Dropout fraction = 0.2	$22 \times 22 \times 40$	0
Flatten	N/A	$19,360 \times 1$	0
Fully connected	32 output nodes	32×1	619,552
Batch normalization	N/A	32×1	128
Dropout	Dropout fraction = 0.2	32×1	0
Fully connected	1 output node	1	33

The loss is ambivalent to $\pm 180^\circ$ estimation errors and is maximized for $\pm 90^\circ$ errors. Note the actual implementation is a softened version of this function to overcome the discontinuity in the derivative at 0° and multiples of $\pm 180^\circ$.

The architecture of the orientation estimation CNN is given in Table 2. The CNN consists of 805,949 trainable parameters and is trained using only synthetically generated training data.

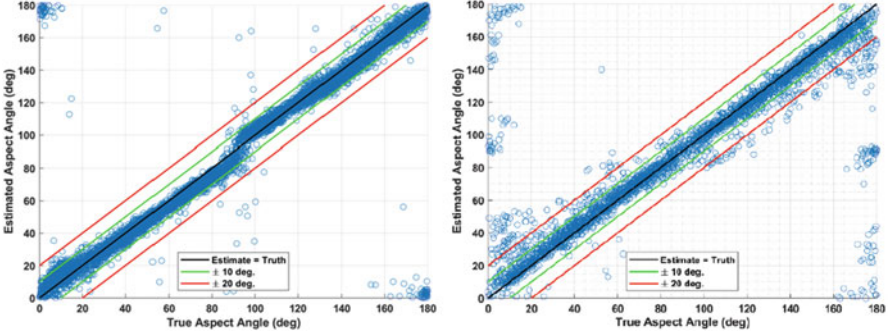


Fig. 16 Left: Deep learning aspect angle estimation performance. Right: Radon transform aspect angle estimation performance

Table 3 Aspect angle estimation algorithm performance summary on the MSTAR dataset

Threshold	Deep learning % within threshold	Radon % within threshold
$\pm 10^\circ$	96.3%	84.4%
$\pm 20^\circ$	99.4%	92.1%

The deep learning algorithm and a Radon transform algorithm performance on the MSTAR dataset are shown in Fig. 16. The performance of the algorithms is summarized in Table 3. The figures and the table show that the deep learning algorithm performs excellently and significantly outperforms the Radon transform algorithm.

4 Classification Algorithm

This section describes the final stage of the end-to-end ATR algorithm, which is the classification (i.e., declaration of object type and class) of a chip nominated by the front-end processing. The classification algorithm presented here is a novel hybrid of a classical template-matching algorithm and a deep learning CNN image classification algorithm. The motivation for this approach is the desire for an algorithm that would:

1. Maintain the robustness of template-matching algorithms to variations in sensor, sensor geometry, clutter, and minor target variations.
2. Improve performance of the template-matching algorithm by using more complex information from each input image as well as information about how an input relates to each of the different target classes.

The new algorithm builds on standard template-matching algorithms [10], which nominally consist of the following steps:

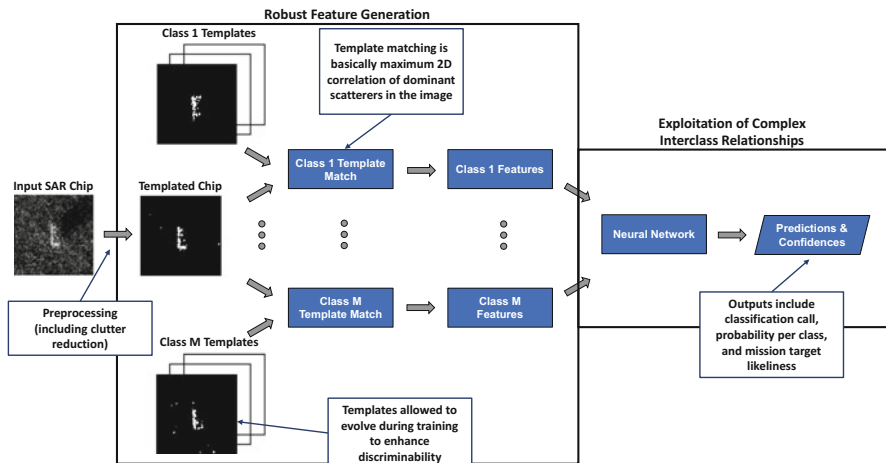


Fig. 17 A high-level overview of the proposed algorithm. The first stage learns features, and the second stage learns interclass relationships

1. Estimate the target orientation.
2. Preprocess the chip by keeping only the top N-valued pixels and then binning the pixels into discrete bins between zero and one.
3. Compute the maximum 2D correlation between the preprocessed test chip and each processed training set template chip near the estimated orientation.
4. Declare the predicted target class as the class of the training chip that achieves the maximum correlation.

Our hybrid deep learning/template-matching algorithm first recasts template matching as a CNN with large kernels (with size on the order of the target size in pixels) and then allows network weights to evolve (i.e., the templates to deform) to improve discrimination. The final step of the algorithm is a fully connected layer that maps feature-match scores to a score indicating the likelihood of belonging to each target class. Note that this is different from a classic template-matching step where the declared target class is the class of the best template match. Instead, each training class contributes to the classification call of the input test chip, allowing both “positive” and “negative” information to play a part in the decision.

With this as background, the classification CNN can be viewed as splitting classification into two interconnected stages: (1) robust feature generation and (2) exploitation of complex interclass relationships. This is illustrated in Fig. 17.

Our CNN implementation incorporates the four steps in the conventional template-matching algorithm described above. Step 1 (orientation estimation) uses the front-end orientation estimate described earlier; Step 2 (preprocessing) is implemented by preprocessing all chips; Step 3 (correlation estimation) is implemented with a convolution layer followed by a max-pooling operation; and Step 4 (maximization) is softened from a maximum operation to a weighted-sum operation via a fully connected layer.

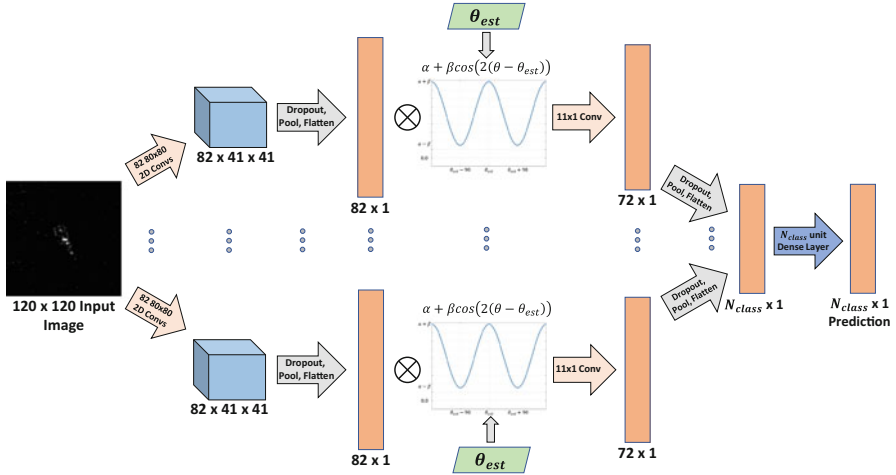


Fig. 18 Detailed architecture of the proposed algorithm

In the absence of training (and with proper initialization on the dense layer), this configuration literally performs conventional template matching. This gives the “epoch 0” (untrained) network a unique advantage over other DL techniques in that it performs fairly well without any training. This can also be viewed as a good initial estimate in the optimization process carried out by training. When the network is trained, both the templates and weights in the final dense layer are updated, resulting in an improvement over template matching that stems from both (i) improved features and (ii) exploitation of interclass relationships. Figure 18 shows a more detailed illustration of the algorithm.

The convolutional kernels in the first stage of the CNN are initialized before training as preprocessed chips at 5° spacing. The estimated azimuth orientation angle is included by applying a weighting function to the template-match scores, which lowers the scores from azimuths far from the estimated azimuth and elevates scores near the correct azimuth (modulo 180°). In addition, to further improve performance: (1) dropout layers have been included after the initial template-match score generation stage as well as just before the final fully connected layer, and (2) a convolution across template-match scores in azimuth with a nominally Gaussian kernel is applied to induce robustness to spurious template matches and errors in the orientation estimation algorithm.

4.1 Algorithm Training

To overcome issues arising from the similarity of the training and testing sets, as well as to demonstrate ATR in the important scenario where collected training data

is either scarce or unavailable, the algorithm is trained using only synthetic data. Synthetic data was generated using asymptotic ray-tracing techniques from 3D CAD models of the MSTAR targets. The data was simulated at X-band, with bandwidth and aperture chosen to achieve 1 foot resolution. The image chips were then formed by backprojection of the synthetic data at 1° increments for each of the 10 targets using HH, HV, VH, and VV polarizations, yielding 1440 synthetic chips per target class. Due to reciprocity, we elected to only use HV polarizations.

The algorithm is trained using the ADADELTA optimizer [23] for 50 epochs. Each training chip undergoes a series of data augmentation transformations, including:

1. A random translation of no more than 10% of image size in each dimension.
2. Addition of zero-mean Gaussian noise ($\sigma = 5^\circ$) to the chip orientation angle.
3. Addition of Rayleigh noise to achieve a target-to-clutter ratio of ~ 10 dB. More complicated clutter models can be used at the cost of computational efficiency.

4.2 Validation Experiment

The classification algorithm was validated using the synthetically generated target chips described above for training, and the trained model was tested on the 10-class MSTAR flight-collected dataset at 15° elevation. The algorithm achieved an overall classification accuracy of $\approx 92\%$ using fully synthetic training and collected testing data. The detailed classification results of the experiment are summarized by the confusion matrix shown in Fig. 19. It can be seen that the classification performance is 80% or better for each of the ten targets. Moreover, six of the ten targets performed above 90%. Performance on the lower-performing classes could likely be improved by improvements to the CAD model and/or signature generation.

4.2.1 The Learned Network

Examples of the learned features (“templates”) of the network are shown in Fig. 20. It is interesting to compare these learned templates with the example chips in Fig. 21. Clearly, there is a strong correspondence between the field-collected targets and the learned templates even though the network was trained on only synthetic data. This shows that the feature extraction stage of the proposed algorithm remains similar to a template-matching algorithm, which is a promising indicator for robustness.

The off-diagonal weights of the fully connected layer mapping from the maximum template-match score for each class to the pre-softmax outputs are shown in Fig. 22. This shows that the dense network has learned a weighted average that is

True Class	2S1	230	4	0	0	25	1	9	5	0	0	% Correct	% Incorrect		
	BMP2	0	573	3	0	0	1	1	4	0	5			97.6%	2.4%
	BRDM2	6	4	225	8	9	6	2	4	0	10			82.1%	17.9%
	BTR60	2	0	15	157	1	8	2	2	0	8			80.5%	19.5%
	BTR70	1	0	1	2	191	1	0	0	0	0			97.4%	2.6%
	D7	0	2	3	0	0	265	0	2	0	2			96.7%	3.3%
	T62	4	0	0	0	3	0	246	20	0	0			90.1%	9.9%
	T72	0	1	0	0	0	1	8	560	9	3			96.2%	3.8%
	ZIL	6	2	8	1	3	4	1	5	233	11			85.0%	15.0%
	ZSU	0	0	1	0	0	1	1	6	0	265			96.7%	3.3%
			2S1	BMP2	BRDM2	BTR60	BTR70	D7	T62	T72	ZIL			ZSU	
		Predicted Class													

Fig. 19 Confusion matrix for MSTAR targets with the algorithm trained on purely synthetic data

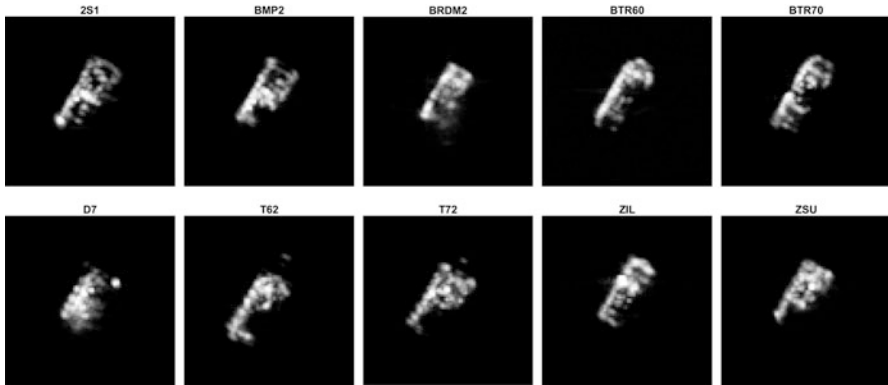


Fig. 20 Examples of the convolutional kernels of the trained network

dependent on the target class, which implies that the network has learned detailed relationships between the target classes. The following observations can be made:

1. The T62 and T72 are strongly attracted to each other.
2. The T72 is attracted to tracked vehicles but is agnostic to or strongly repelled from the wheeled vehicles (BRDM2, BTR60, BTR70, and ZIL131).
3. The 2S1 is strongly attracted to the other tank-like vehicles (BMP2, T62).
4. The BTR70 is strongly attracted to other wheeled vehicles (BRDM2, ZIL131).

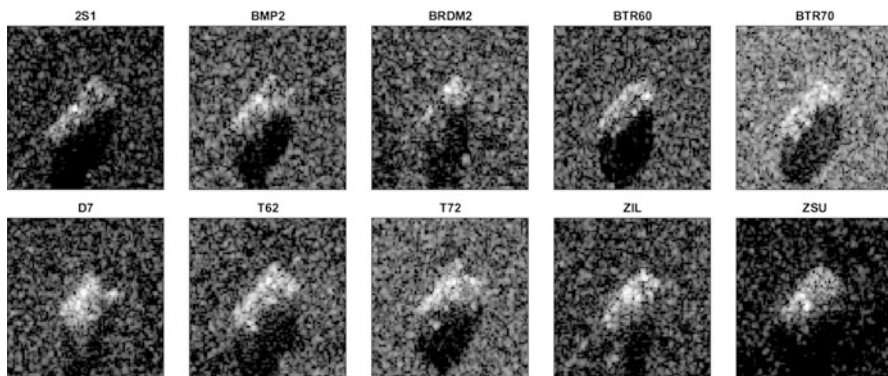


Fig. 21 Examples of the collected MSTAR data



Fig. 22 Off-diagonal weights of the final dense layer

4.2.2 Comparison to the Literature

As discussed previously, the typical MSTAR experiment in the literature [3, 8, 12, 25, 26] is to train on the data at 17° elevation and test on the data at 15° elevation. To baseline our algorithm against the algorithms in the literature, we applied our approach to the standard MSTAR experiment. The algorithm achieved an overall accuracy of 99.3% in 50 training epochs. The confusion matrix is shown in Fig. 23. The results show that the algorithm achieves nearly perfect performance

	2S1	BMP2	BRDM2	BTR60	BTR70	D7	T62	T72	ZIL	ZSU		
True Class	2S1	0	0	0	0	0	0	0	0	0	100.0%	0.0%
	BMP2	585	0	0	0	0	0	2	0	0	99.7%	0.3%
	BRDM2	0	272	0	0	0	0	1	0	1	99.3%	0.7%
	BTR60	0	4	185	0	1	2	0	2	1	94.9%	5.1%
	BTR70	0	0	0	196	0	0	0	0	0	100.0%	0.0%
	D7	0	0	0	0	271	0	1	0	2	98.9%	1.1%
	T62	0	0	0	0	0	273	0	0	0	100.0%	0.0%
	T72	0	0	0	0	0	2	580	0	0	99.7%	0.3%
	ZIL	0	1	0	0	0	1	0	270	2	98.5%	1.5%
	ZSU	0	0	0	0	1	0	0	0	273	99.6%	0.4%
											% Correct	% Incorrect

Fig. 23 Confusion matrix for our algorithm in the standard MSTAR classification setup

in this experiment; however, as discussed previously, this result is not indicative of algorithm performance so much as the similarity between the training and testing data.

4.2.3 Comparison to Template Matching

As the proposed algorithm is supposed to be an improvement over template matching, it is important to compare the proposed algorithm performance to that of a template-matching algorithm. To this end, the primary experiment (train on synthetic data, test on the MSTAR publicly available data) was repeated using a template-matching algorithm to perform the predictions. The template-matching algorithm used is to:

- 1. Process a given test chip and all template chips (keeping top N pixels and then binning the values into discrete bins).
- 2. Find the maximum correlation between the processed test chip and all processed template chips.
- 3. Report the predicted class as the class of the template that produced the best match.

Using the same training data as described above, the template-matching algorithm achieved only $\approx 79\%$ overall classification accuracy. The confusion matrix is shown in Fig. 24. Hence, the proposed algorithm has enabled an overall classification performance improvement of about 13%.

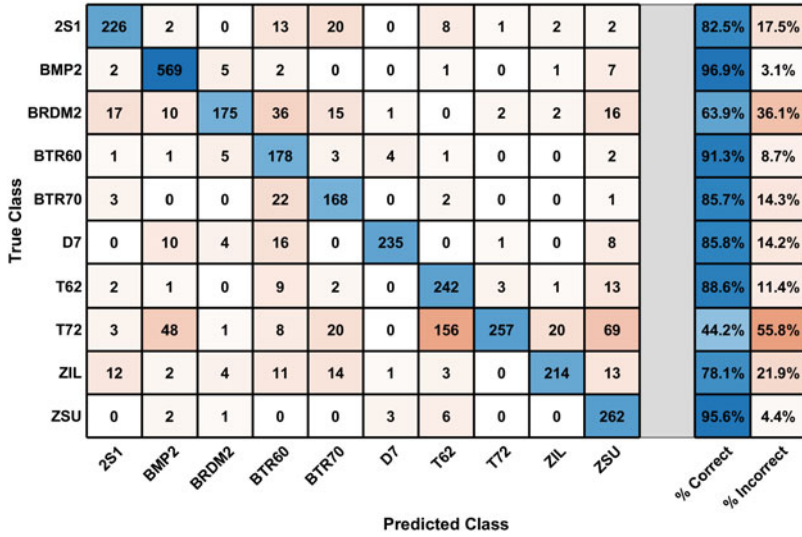


Fig. 24 Confusion matrix for the synthetic MSTAR experiment using template matching

5 Conclusion

This chapter provided a framework for end-to-end SAR ATR and discussed a suite of deep learning algorithms in that framework. We focus on approaches that use synthetic data for training to sidestep some of the issues present in deep learning approaches that use training and testing data that are very similar. In addition, our deep learning algorithms are novel as they build on the success and approaches of legacy algorithms, producing hybrid conventional/deep learning approaches. The algorithms were applied to the publicly available MSTAR dataset and demonstrated excellent performance, even when training only on the synthetically generated data. This promising suite of algorithms is a significant step forward in the state of the art for SAR ATR.

Acknowledgments This material is based upon work supported by the Air Force Research Laboratory (AFRL) under Contract No. FA8651-20-C-0046. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the AFRL.

References

1. Carrara, W.G., Goodman, R.S., Majewski, R.M.: Spotlight Synthetic Aperture Radar: Signal Processing Algorithms. Artech House, Boston (1995)
2. Jakowatz, C.V., Wahl, D.E., Eichel, P.H., Ghiglia, D.C., Thompson, P.A.: Spotlight-Mode Synthetic Aperture Radar: A Signal Processing Approach. Springer (2011). <https://doi.org/10.1007/978-1-4613-1333-5>

3. Ross, T., Worrell, S., Velten, V., Mossing, J., Bryant, M.: Standard SAR ATR Evaluation Experiments using the MSTAR Public Release Data Set. SPIE Conference on Algorithms for Synthetic Aperture Radar Imagery V **3370**(April 1998), 566–573 (1998). <https://doi.org/10.1117/12.321859>
4. Keydel, E.R., Lee, S.W., Moore, J.T.: MSTAR Extended Operating Conditions: A Tutorial. In: E.G. Zelnio, R.J. Douglass (eds.) Proceedings of SPIE Algorithms for Synthetic Aperture Radar Imagery, vol. 2757, pp. 228–242. International Society for Optics and Photonics (1996). <https://doi.org/10.1117/12.242059>. <http://proceedings.spiedigitallibrary.org/proceeding.aspx?articleid=1018994>
5. Brown, W.M., Swonger, C.W.: A prospectus for automatic target recognition. IEEE Transactions on Aerospace and Electronic Systems **25**(3), 401–410 (1989). <https://doi.org/10.1109/7.30795>
6. Novak, L.M., Owirka, G.J., Brower, W.S.: Performance of 10- and 20-target MSE classifiers. IEEE Transactions on Aerospace and Electronic Systems **36**(4), 1279–1289 (2000). <https://doi.org/10.1109/7.892675>
7. Potter, L.C., Moses, R.L.: Attributed scattering centers for SAR ATR. IEEE Transactions on Image Processing **6**(1), 79–91 (1997). <https://doi.org/10.1109/83.552098>
8. Zhao, Q., Principe, J.C.: Support vector machines for SAR automatic target recognition. IEEE Transactions on Aerospace and Electronic Systems **37**(2), 643–654 (2001). <https://doi.org/10.1109/7.937475>
9. Roth, M.W.: Survey of neural network technology for automatic target recognition. IEEE Transactions on Neural Networks **1**(1), 28–43 (1990). <https://doi.org/10.1109/72.80203>
10. El-Darymli, K., Gill, E.W., McGuire, P., Power, D., Moloney, C.: Automatic Target Recognition in Synthetic Aperture Radar Imagery: A State-of-the-Art Review. IEEE Access **4**, 6014–6058 (2016). <https://doi.org/10.1109/ACCESS.2016.2611492>
11. Kechagias-Stamatis, O., Aouf, N.: Automatic target recognition on synthetic aperture radar imagery: A survey. IEEE Aerospace and Electronic Systems Magazine **36**(3), 56–81 (2021). <https://doi.org/10.1109/MAES.2021.3049857>
12. Morgan, D.A.E.: Deep convolutional neural networks for ATR from SAR imagery. In: Proc. SPIE 9475, Algorithms for Synthetic Aperture Radar Imagery XXII, 94750F, vol. 9475 (2015). <https://doi.org/10.1117/12.2176558>
13. Chen, S., Wang, H., Xu, F., Jin, Y.: Target Classification Using the Deep Convolutional Networks for SAR Images. IEEE Transactions on Geoscience and Remote Sensing **54**(8), 4806–4817 (2016). <https://doi.org/10.1109/TGRS.2016.2551720>
14. Zhou F. and Wang, L., Bai, X., Hui, Y.: SAR ATR of Ground Vehicles Based on LM-BN-CNN. IEEE Transactions on Geoscience and Remote Sensing **56**(12), 7282–7293 (2018). <https://doi.org/10.1109/TGRS.2018.2849967>
15. Soldin, R.J., MacDonald, D.N., Reisman, M., Konz, L.R., Rouse, R., Overman, T.L.: HySARNet: a hybrid machine learning approach to synthetic aperture radar automatic target recognition. In: Proc. SPIE 10988, Automatic Target Recognition XXIX, vol. 10988 (2019). <https://doi.org/10.1117/12.2518155>
16. Billingsley, J.B., Farina, A., Gini, F., Greco, M.V., Verrazzani, L.: Statistical analyses of measured radar ground clutter data. IEEE Transactions on Aerospace and Electronic Systems **35**(2), 579–593 (1999). <https://doi.org/10.1109/7.766939>
17. Greco, M.S., Gini, F.: Statistical analysis of high-resolution SAR ground clutter data. IEEE Transactions on Geoscience and Remote Sensing **45**(3), 566–575 (2007). <https://doi.org/10.1109/tgrs.2006.888141>
18. Skonik, M.I.: Radar Handbook, Third Edition. McGraw Hill (2008)
19. Alexandropoulos, G.C., Berberidis, K., Sagiias, N.C., Lazarakis, F.I., Datsikas, C.K., Alexandridis, A.A., Dangakis, K.P.: On the sum of squared correlated Rayleigh variates and applications to maximal-ratio diversity. In: 2007 IEEE 18th International Symposium on Personal, Indoor and Mobile Radio Communications, pp. 1–5 (2007). <https://doi.org/10.1109/PIMRC.2007.4394161>

20. Joarder, A.H., Omar, M.H., Gupta, A.K.: The distribution of a linear combination of two correlated chi-squared variables. *Rev. Colombiana Estadística* **36**(2), 209–219 (2013)
21. Ferrari, A.: A note on sum and difference of correlated chi-squared variables (2019). <https://doi.org/10.48550/arXiv.1906.09982>
22. El-Darymli, K., McGuire, P., Power, D., Moloney, C.R.: Target detection in synthetic aperture radar imagery: a state-of-the-art survey. *J. of Applied Remote Sensing* **7**, 71535–71598 (2013). <https://doi.org/10.1117/1.JRS.7.071598>
23. Zeiler, M.D.: ADADELTA: An Adaptive Learning Rate Method. arXiv p. 6 (2012). <http://arxiv.org/abs/1212.5701>
24. Beylkin, G.: Discrete radon transform. *Acoustics, Speech and Signal Processing, IEEE Transactions on* **35**(2), 162–172 (1987). <https://doi.org/10.1109/TASSP.1987.1165108>
25. Chen, S., Lu, F., Wang, J., Liu, M.: Target aspect angle estimation for synthetic aperture radar automatic target recognition using sparse representation. In: 2016 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC), pp. 1–4 (2016). <https://doi.org/10.1109/ICSPCC.2016.7753656>
26. Wilmanski, M., Kreucher, C., Lauer, J.: Modern approaches in deep learning for SAR ATR. In: Proceedings of the 2016 SPIE Defense, Security, and Sensing Symposium, vol. 9843, pp. 98430N–98430N–10 (2016). <http://dx.doi.org/10.1117/12.2220290>

Change Detection in SAR Images Using Deep Learning Methods



Luca Bergamasco and Francesca Bovolo

1 Introduction

Synthetic aperture radar (SAR) images can be acquired in any weather and light condition. This feature is useful when associated with change detection (CD) methods for civil protection tasks, such as disaster monitoring [1] and flood detection [2–4], or agricultural monitoring [5, 6]. In civil protection tasks, CD using SAR images allows to promptly identify the damaged areas even in cloudy conditions since SAR signal can penetrate clouds. However, SAR images are peculiar data with unique problems, such as the speckle noise and side looking acquisition distortions, that need to be addressed to detect changes accurately.

Some CD methods perform a multi-temporal classification to detect changes between images (i.e., post-classification comparison [73]). However, these methods require many labeled data challenging to gather for multi-temporal acquisitions [7]. Thus most SAR CD methods exploit the fusion or comparison of bi-temporal or multi-temporal images to retrieve change indexes and enhance the differences between images. This approach type is the most common since it is unsupervised and does not require labeled data [7]. The log ratio is a change index widely used for the CD of SAR images [8–10]. Another family of change indexes is the similarity measures that evaluate the similarity of bi-temporal images from various points of view. Kullback–Leibler (KL) divergence is exploited as change index to measure the similarity of bi-temporal SAR image probability densities [11, 12]. The normalized compression distance (NCD) is a non-negative number representing the difference between two images where a small value means unchanged areas and high values changed ones [13]. Mutual information is used to evaluate the common information

L. Bergamasco · F. Bovolo (✉)
Fondazione Bruno Kessler, Trento, Italy
e-mail: lbergamasco@fbk.eu; bovolo@fbk.eu

© The Editor(s) (if applicable) and The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

M. Rysz et al. (eds.), *Synthetic Aperture Radar (SAR) Data Applications*, Springer Optimization and Its Applications 199, https://doi.org/10.1007/978-3-031-21225-3_2

between couples of bi-temporal SAR images (e.g., polarimetric SAR images) [14, 15], whereas the variational information quantifies their difference [16]. These two change indexes are merged in the mixed information through a linear combination using the trade-off parameter α that balances the two contributions [17]. These methods extract features that can well characterize the change between images. However, they retrieve the change indexes of a single scale of the images. This limits the CD performance, especially when SAR images requiring an extended analysis of the spatial context information are examined.

Multi-scale features provide multiple representations of the analyzed images that allow improving the modeling of geometrical details and homogeneous areas. Multi-scale approaches improve the modeling of the spatial context information with respect to single-scale methods in high-resolution images. Multi-scale versions of the same scene can be retrieved using various strategies, such as pyramid [18], wavelets [10, 19], morphological filters [20], and object-based methods [21, 22]. The multi-scale representations of the discrete wavelet transform (DWT) are used in CD methods to preserve the geometrical details of the scene by considering the reliable areas of the various scale levels [10] or to select the most informative representations and include the spatial context information using the Markov random fields (MRFs) [23]. Multi-scale features are also exploited in building CD methods where they are used to detect the changed buildings at the proper scale level to avoid superfluous details that can badly affect the performance [24]. This method identifies new and destroyed buildings by evaluating the increased–decreased backscattering pattern of candidates retrieved from the change detection of the optimal scale-level feature. Changed buildings are examined using four fuzzy rules that evaluate spatial properties and alignments of the candidates. These rules are created by considering the real backscattering patterns occurring in new and destroyed building cases. The previously presented methods exploit handcrafted features designed ad hoc for specific scenarios and need to be re-designed when the latter ones change.

Deep learning (DL) methods alleviate this problem since they automatically learn features from the input data during a training phase. Many DL CD methods are based on convolutional neural networks (CNNs) [25–29] since they automatically learn features that model the spatial context information on images. The spatial context features extracted from convolutional-based models allow improving the capability of CD methods to detect the changes accurately [26, 29, 30]. However, most of the DL CD methods are supervised and require large multi-temporal labeled datasets for the training [26, 28]. The gathering of many multi-temporal labeled data is challenging. Domain adaptation (DA) methods address this problem by fitting the characteristics of a DL model pre-trained with given labeled samples (i.e., source domain) with the ones of the data to analyze that might be acquired by another sensor or in another geographical area (i.e., target domain) [31]. Many DA CD methods exploit generative adversarial networks (GANs) or adversarial models to reduce the difference between source and target domains. This allows developing CD methods using the information derived from multi-sensor data [32–34]. CycleGANs are used to extract feature maps with a common domain between

multi-spectral and SAR images that improve the CD in bi-temporal SAR images [35]. Some DA CD methods change the domain of a DL model trained with a given type of RS data to process data with heterogeneous characteristics [36, 37] or transfer the knowledge of a pre-model to process unlabeled RS data [38]. However, the DA of pre-trained DL models to perform applications or process RS data different from the source ones is still a challenging open issue in the literature.

CD method can exploit transfer learning to process target RS data with a pre-trained model. Unlike DA, transfer learning methods do not require fine-tuning or re-training steps before the CD. These methods achieve accurate results in bi-temporal CD tasks when they process images with similar characteristics with respect to the ones used for the training [27, 29, 39, 40]. Performance of transfer learning CD methods decreases when the difference between the characteristics of target and source images increases [39].

Unsupervised DL methods can be used to avoid the dependence on labeled data in DL CD methods. These methods achieve accurate results in the CD between bi-temporal images [29, 30, 41]. Autoencoders (AEs) can be used to reduce the differences between pre-change and post-change images due to factors uncorrelated with the ground changes [42]. Unsupervised DL model can automatically learn and extract features providing information about changes that can be used in CD tasks [43, 44]. However, most of these methods do not capture the spatial context information of images that is relevant to find changes and can improve the CD outcome [45]. Convolutional autoencoders (CAEs) automatically learn spatial features during an unsupervised training that allow managing the spatial context information. CAEs merge the capabilities of AEs to automatically learn features from the input data with the capacity of convolutional neural networks (CNNs) to analyze and encode the spatial context information [46]. CAEs are often used to pre-trained DL models in an unsupervised way that are then fine-tuned with labeled data to perform a specific task or classify input data [47–49]. Thus the amount of labeled data needed to train the model is reduced with respect to supervised DL methods. CAEs can be used to transform pre-change images into post-change ones, and vice versa, and compare the transformed image with the other one to detect the changes [53]. CAEs extract more informative multi-scale feature maps than other SoAs, such as DWT. DWT only retrieves the multi-scale representations of input images, whereas CAEs also capture their semantic information [50–52]. CAEs can produce bi-temporal multi-scale feature maps by processing bi-temporal images that can be used by a multi-scale CD method to improve the detection of the changed areas with respect to single-scale approaches [51, 52].

We focus this chapter on an unsupervised DL CD method on bi-temporal SAR images using CAE to retrieve multi-scale feature maps. This method trains a CAE in an unsupervised way using unlabeled data sampled from the pre-change SAR image. The trained CAE processes the bi-temporal SAR images to extract from a given number of CAE layers bi-temporal multi-scale feature maps that are compared to enhance changes between images. Many of the comparisons provide poor information about changes. Thus a standard-deviation-based feature selection step is applied to keep only the most informative comparisons. The multi-scale

selected comparisons are processed by a multi-scale detail-preserving CD method that allows obtaining an accurate change map while handling the speckle noise [10].

The chapter has the following outline. Section 2 presents the state of the art of DL CD methods used to examine SAR data. Section 3 describes the methodology, and Sect. 4 presents the experimental settings and the results. Finally, we draw our conclusion in Sect. 5.

2 State of the Art

In this section, we present the convolutional autoencoders (CAEs) that are the DL models on which the method presented in Sect. 3 is based, the formulation of the change detection (CD) problem, and the state-of-the-art (SoA) methods regarding the detection of changes in bi-temporal and multi-temporal synthetic aperture radar (SAR) images using statistical, machine learning (ML), and deep learning (DL) approaches.

2.1 Convolutional Autoencoders

Convolutional autoencoders (CAEs) merge the capability of autoencoders (AEs) and convolutional neural networks (CNNs) to automatically learn spatial context features from the input data during unsupervised training [46]. CAEs:

- Have the capability of AEs to learn features from the input data during an unsupervised training phase while reconstructing the input.
- Exploit the capability of convolutional layers to analyze and encode the spatial context information of images.

CAEs can be divided into an encoder and a decoder (Fig. 1). The encoder compresses the spatial information of the input and increases the number of features. The decoder upsamples the spatial information and increases the feature complexity by aggregating the previous feature maps. The objective of the unsupervised training is the minimization of a loss function that usually is a mean squared error (MSE):

$$J(W, b) = MSE = \frac{1}{N} \sum_{n=1}^I (X_n - X'_n)^2. \quad (1)$$

N represents the number of samples composing the training set, X_n is the n th input image sample, whereas X'_n is its reconstruction. The MSE minimization reduces the reconstruction error and retrieves an output image as similar as possible to the input one. The CAE automatically learns spatial context features to reconstruct the input image during the training.

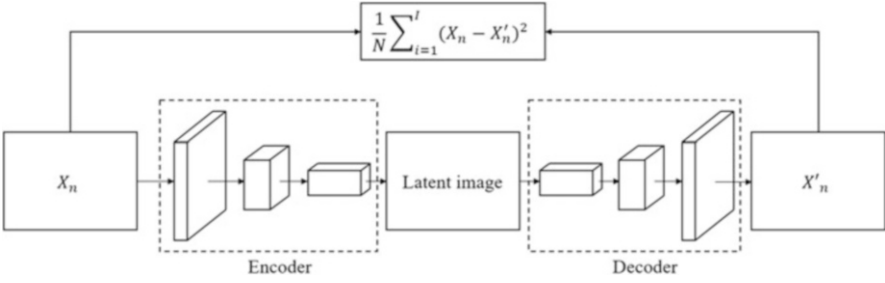


Fig. 1 The block scheme of a convolutional autoencoder

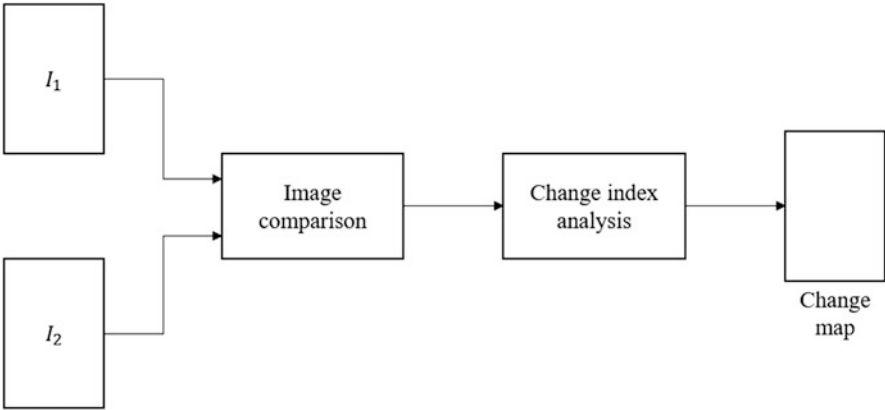


Fig. 2 Block scheme of the standard change detection approach

2.2 Change Detection

Change detection (CD) is a process that aims to identify changed objects by comparing two (bi-temporal case) or more images (multi-temporal case) I_1, I_2, \dots, I_m (where m is the total number of analyzed images) acquired over the same geographical area at different times (see Fig. 2). In this chapter, we focus the analysis on the bi-temporal CD analysis. In the bi-temporal CD case, a pre-change I_1 and a post-change I_2 image acquired over the same geographical area at different times t_1 and t_2 are compared and analyzed. If the CD method is supervised, the image comparison is exploited to discriminate between the class ω_{nc} of unchanged pixels and the different classes $\Omega_c = \{\omega_{c_1}, \omega_{c_2}, \dots, \omega_{c_B}\}$ of changed pixels, where B is the total number of classes in the scene. In the unsupervised case, the image comparison is used to distinguish between unchanged ω_{nc} and changed Ω_c pixels. The landcover transitions can also be distinguished with unsupervised approaches, so Ω_c can be further detailed in B change types. However, no explicit from-to label can be assigned to the transitions without labeled training samples.

2.3 CD Methods Using SAR Images

One of the most simple yet effective operators to detect the changes in bi-temporal SAR images I_1 and I_2 is the logarithm ratio:

$$LR = \log \left(\frac{I_2}{I_1} \right). \quad (2)$$

The ratio between bi-temporal SAR images reduces the multiplicative error components, while the logarithm transforms the multiplicative speckle noise into additive noise and makes the distribution of the classes of interest more symmetrical [8–10]. Another operator family often used as change indexes for SAR images is composed of theoretical similarity measures [54], such as the Kullback–Leibler (KL) [11, 12] divergence and normalized information distance (NID) [13]. The KL divergence measures the similarity between two probability distributions characterizing the neighborhood of a pixel. Assuming that $f_1(x)$ and $f_2(x)$ are the probability density functions of I_1 and I_2 , the KL divergence from I_2 to I_1 is given by

$$KL(I_2|I_1) = \int \log \frac{f_1(x)}{f_2(x)} f_1(x) dx. \quad (3)$$

It has a small value if the probability distributions in I_1 and I_2 are similar (no changes); otherwise, the value is high. The KL divergence is an asymmetric function, so the summation of two asymmetric divergences is needed to obtain a symmetric version ($KL(I_2|I_1) + KL(I_1|I_2)$). The estimation of the KL divergence requires knowing the statistical behavior of the multi-temporal SAR images. The KL divergence assumes a closed form if the multi-temporal images have a Gaussian distribution; otherwise, a non-parametric estimation is needed, which is computationally demanding. NID evaluates the similarity between two objects, and it is defined as the minimum amount of energy required to transform an image into the other one. Since NID is based on a non-computable notion of Kolmogorov complexity, a practical version of NID is exploited: the normalized compression distance (NCD) [13]. The NCD values represent the difference between the two images: small values mean no changes, and high values represent changes. NID and NCD capture the non-linear dependencies between images. The fusion of features derived by SAR images for CD purposes is performed using other statistical similarity measures, such as the mutual information [55], variational information [16], and mixed information [17]. The mutual information evaluates the common information or the independence between bi-temporal SAR images, whereas the variational one quantifies the different information between images. The mixed information merges the mutual and variational information using a trade-off parameter $\alpha \in [0, 1]$ resulting in better performance than these two similarity measures considered independently. The mixed information effectively processes multi-temporal or multi-sensor SAR data having heterogeneous radiometric characteristics.

Multi-scale CD methods extract from the input images features with various spatial resolutions used to reduce the noise of homogeneous areas and preserve the geometrical details in the heterogeneous ones. [10] apply two-dimensional stationary DWT on the log ratio LR of two bi-temporal images to retrieve multi-scale feature maps. This method exploits the low-pass component of DWT to obtain a set of feature maps LR_{MS} with L' resolution levels $LR_{MS} = \{LR_{LP}^0, \dots, LR_{LP}^{l'}, \dots, LR_{LP}^{L'-1}\}$ ($l' = 0, \dots, L' - 1$), where $LR_{LP}^{l'}$ is the DWT low-pass component of the l' th scale level obtained by processing the log ratio LR , and applies a multi-scale CD method. Each $LR_{LP}^{l'}$ is used to retrieve a reliability map that indicates the reliable and unreliable areas and allows handling the speckle noise of SAR images. The reliable areas are the homogeneous ones, whereas the unreliable areas are the heterogeneous ones. The heterogeneity of a resolution level l' is calculated by comparing a local coefficient of variation (LCV) computed on pixels included in a moving window centered in (i, j) (4) with the global coefficient of variation (CV) of that resolution level (5). The LCV and CV formulas do not only calculate the homogeneity of the area but detect the speckle noise.

$$LCV^{l'}(i, j) = \frac{\sigma(LR_{LP}^{l'}(i, j))}{\mu(LR_{LP}^{l'}(i, j))} \quad (4)$$

$$CV^{l'} = \frac{\sigma(LR_{LP}^{l'})}{\mu(LR_{LP}^{l'})} \quad (5)$$

The pixel (i, j) of the feature map at resolution level l' is homogeneous and therefore reliable with small speckle noise if $LCV^{l'}(i, j) < CV^{l'}$. The method averages all possible combinations of LR_{MS} to merge the multi-resolution information of the multi-scale feature maps:

$$\overline{LR_{MS}^{l'}} = \frac{1}{l' + 1} \sum_{h=0}^{l'} LR_{LP}^h \quad (6)$$

Change maps are retrieved for each resolution level l' by applying to each $\overline{LR_{MS}^{l'}}$ a threshold $T^{l'}$ that can be found using manual or automatic approaches. The pixel (i, j) of the final change map M is assigned to the changed (ω_c) or unchanged (ω_{nc}) class, according to the class detected in the pixel of position (i, j) of $LR_{MS}^{S_{i,j}}$, where $S_{i,j} \leq L' - 1$ is the most reliable resolution level of pixel (i, j) , i.e.,

$$M(i, j) \in \begin{cases} \omega_{nc}, & \text{if } \overline{LR_{MS}^{S_{i,j}}}(i, j) \leq T_{S_{i,j}} \\ \omega_c, & \text{if } \overline{LR_{MS}^{S_{i,j}}}(i, j) > T_{S_{i,j}}. \end{cases} \quad (7)$$

These multi-scale CD methods retrieve multi-scale features providing heterogeneous representations of scenarios allowing better examining the image context and improving the CD in challenging areas (e.g., changed area borders). However, these methods obtain multi-scale representations using handcrafted features that might not capture further multi-scale and spectral information to optimally detect changes.

Multi-scale methods can perform the building CD in SAR images using fuzzy rules [24]. The two-dimensional stationary wavelet transform is applied to the log ratio LR of two SAR images to find the optimal scale to detect destroyed buildings. A set $L' - 1$ of multi-scale representations $LR_{MS} = \{LR_{LP}^0, \dots, LR_{LP}^{l'}, \dots, LR_{LP}^{L'-1}\}$, where $l' = 0, \dots, L' - 1$ is the resolution level and $LR_{LP}^{l'}$ is the DWT low-pass component of the l' th scale level, is computed from LR . The method selects the optimal resolution level that reduces the geometrical details and removes the small changes but preserves the changes of a given dimension (i.e., destroyed buildings). Once the optimal scale level is selected, a change map derived from LR_{LP}^{opt} , where LR_{LP}^{opt} is the representation of LR at the optimal resolution level, is retrieved. The change map shows the unchanged pixels ω_{nc} and the changes with a decreasing (ϵ^-) and an increasing (ϵ^+) of backscattering. From the change maps, a set of H changed building candidates $\Gamma = \{\gamma_1, \gamma_2, \dots, \gamma_H\}$ by taking the regions with near ϵ^- and ϵ^+ areas. The destroyed buildings are detected by applying four fuzzy rules to the candidates Γ :

- **Completeness:** Evaluates the presence of both ϵ^+ and ϵ^- areas in the same region belonging to a candidate.
- **Proportionality of areas:** Controls that the ϵ^+ area does not prevail on the ϵ^- and vice versa. It finds the minimum between the comparison of increased (s_I) and decreased (s_D) backscattering area extensions $r_s = \min\{s_I/s_D, s_D/s_I\}$. The nearer is r_s from 1, the more proportional the increased and decreased areas are.
- **Equivalence of length:** Checks that the length in the azimuth direction between the ϵ^+ and ϵ^- areas is similar. The length of ϵ^+ (l_I) and ϵ^- (l_D) are computed by considering their extrema. The minimum between the ratio of l_I and l_D is calculated ($r_l = \min\{l_I/l_D, l_D/l_I\}$) to estimate the similarity of the two area lengths. The farther is r_l from 1, the smaller is the probability to have a candidate with an increased–decreased pattern.
- **Alignment:** Controls the alignment of the increased–decreased pattern in the range direction. In particular, it checks that the centroids of ϵ^+ and ϵ^- areas are on the same line, so the angle α included in the line connecting the centroids and the range direction should be small.

Each of the fuzzy functions results in a value that represents the probability of the candidate being a new/destroyed building. The four results are multiplied, so if one of the four fuzzy operators provides small results, the overall one will be small. The candidates with values higher than a defined threshold T are considered new/destroyed buildings accordingly to the order of appearance of ϵ^+ and ϵ^- .

2.4 Unsupervised DL CD Method Using SAR Images

Convolutional-based deep learning (DL) models automatically learn spatial context features from the input data during a training phase. These features can be used to detect changes in bi- and multi-temporal remote sensing (RS) data. However, most of the DL models are trained in a supervised way and require many labeled data. The gathering of labeled RS data is challenging in bi- and multi-temporal cases, especially for SAR images. A solution to this problem is the use of pre-trained models that can be used as feature extractors. Yet few DL models are pre-trained with SAR images. In [1], the authors propose a CD method that exploits the transfer learning to extract informative feature maps to detect the changes that occurred in bi-temporal SAR images. The method adapts a convolutional neural network (CNN) pre-trained using VHR airborne orthophotos [56] using adaptive batch normalization (AdaBN) layers [57]. AdaBN states that the domain information of models is in the parameters of batch normalization (BN) layers, whereas the class information is in the weights and biases. Thus it aligns the statistical parameters of BN layers between the source and target domains to adapt the pre-trained model to the target data, i.e., SAR images. The authors exploit the adapted model to process bi-temporal VHR SAR images and retrieved feature maps from L layers. The bi-temporal multi-scale feature maps are upsampled using a bi-linear interpolation to the input SAR image spatial resolution and combined into two hypervectors that are compared to retrieve a difference hypervector δ enhancing the changes in the scene. However, many of the compared feature maps of $\delta_l \in \delta$, where $l = 1, \dots, L$, do not provide information about changes. Thus a variance-based feature selection step is applied. Each feature comparison δ_l is divided into S splits δ_{l_s} , where $s = 1, \dots, S$, and the variance is computed for each δ_{l_s} . The feature selection keeps only a percentile of the feature maps of δ_{l_s} having high variance values. For each layer l , the subsets of feature maps retrieved by each split δ'_{l_s} are combined to obtain the difference hypervector of the l th layer δ'_l :

$$\delta'_l = \bigcup_{s=1}^S \delta'_{l_s}. \quad (8)$$

The difference vector δ'_l of each layer is concatenated to the deep change hypervector $G = \delta'_1, \dots, \delta'_l, \dots, \delta'_L$. The magnitude (9) of G composed by D elements g^d , where $d = 1, \dots, D$, is computed and used to detect the changes in the scene.

$$\rho = \sqrt{\sum_{d=1}^D (g^d)^2}. \quad (9)$$

The magnitude ρ is thresholded using Otsu's threshold [58] (other manual or automatic thresholding methods can be used) to obtain a binary change map that

shows changed (Ω_c) and unchanged areas (ω_{nc}). The K-mode algorithm clusters the binary changes Ω_c into two increased ϵ^+ and decreased ϵ^- backscattering. The fuzzy method presented in Sect. 2.3 exploits the increased and decreased backscattering areas to detect the new and destroyed building. This method achieves good performance when the analyzed images have a similar spatial resolution with respect to the ones used to pre-train the model. However, its performance decreases when the spatial resolution of the input image differs or the area to analyze has different characteristics with respect to the pre-trained images (i.e., the pre-trained image represents forest areas and the target image urban areas).

Unsupervised DL methods are preferred in CD tasks since they do not require labeled data but can use the huge amount of unlabeled multi-temporal images that can be gathered easily. In [35], the authors propose a CD method that exploits the features extracted by a cycle generative adversarial network (CycleGAN) that are learned from the input data during an unsupervised training. This DL model aims to transcode SAR images into optical images to learn semantic features that can improve the CD between the bi-temporal SAR data. The CycleGAN is composed of two generators and two discriminators. The generators (G_{YZ} and G_{ZY}) learn, during the unsupervised training, to transcode optical images (Y) into the SAR domain (Z) and vice versa, while the discriminators (D_Z and D_Y) learn to distinguish the generated images (\tilde{Y} or \tilde{Z}) from the original ones (Y or Z). The adversarial loss function between generators and discriminators allows the generators to transcode images from one domain to another one.

$$\min_{G_{YZ}} \max_{D_Z} \mathbb{E}[\log D_Z(Z)] + \mathbb{E}[1 - \log(1 - D_Z(\tilde{Z}))] \quad (10)$$

$$\min_{G_{ZY}} \max_{D_Y} \mathbb{E}[\log D_Y(Y)] + \mathbb{E}[1 - \log(1 - D_Y(\tilde{Y}))], \quad (11)$$

where G_{YZ} and G_{ZY} are the generators that transform the image from Y into Z and vice versa, whereas D_Z and D_Y are the discriminators of the Z and Y domains, respectively. After the transcoding of an image into another domain through one of the two generators (e.g., G_{YZ}), it is processed from the other one (e.g., G_{ZY}), transformed back into the original one, and compared with the input image to check the preservation of the semantic information.

$$G_{ZY}(G_{YZ})(Y) \approx Y \quad (12)$$

$$G_{XZ}(G_{ZY})(Z) \approx Z. \quad (13)$$

The cycle consistently loss function makes the training of the model more robust. After the training of the CycleGAN, the bi-temporal SAR images are processed through one of the two generators G_{ZY} to extract bi-temporal features f^1 and f^2 from L intermediate layer of G_{ZY} . The features are extracted from the middle layers since they are the most suitable for the transfer learning tasks [39, 59, 60]. For each layer $l = 1, \dots, L$, the bi-temporal feature maps f_l^1 and f_l^2 are compared to enhance the changes in the scene, and a variance-based feature selection method

is applied to keep, for each layer l , only the most informative features. The selected features of each layer are concatenated in a D -dimensional hypervector G that captures multi-scale change information. For each component, g^d of G , where $d = 1, \dots, D$, the authors retrieve a change map using the Otsu's threshold [58]. Thus D change maps are obtained, one for each g^d . The final change map is obtained using a majority vote strategy [20] that assigns a suitability score τ to each pixel that agrees on the change of a given pixel. The pixels having a τ greater than a given threshold are considered changed, while the others are unchanged. The change map is exploited to retrieve increased ϵ^+ and decreased ϵ^- backscattering that are used to detect new and destroyed buildings using the same fuzzy approach used in [1]. CycleGANs provide good results but are complex to train because of the many model parameters and the complex loss function.

In [51], the authors proposed a DL CD method that exploits the features extracted from a convolutional autoencoder (see Sect. 2.1) to detect flood areas in bi-temporal SAR images. CAEs automatically learn spatial context features from the input data during unsupervised training. The CAE is composed of L layers and trained using a training set X composed of unlabeled patches x_n , where $n = 1, \dots, N$ sampled from the pre-change image I_1 . The aim of the CAE training is the minimization of the reconstruction error achieved using a sum-squared error (SSE) loss function defined as

$$SSE = \sum_{n=1}^N (X_n - X'_n)^2, \quad (14)$$

where X'_n is the reconstruction of X_n . In this way, the CAE learns spatial context features representing the input data. After the training, the authors separately process the bi-temporal SAR images I_1 and I_2 and retrieve bi-temporal multi-scale features from $L' < L$ layers. For each layer $l' \in L'$, the bi-temporal features $f_{l'}^1$ and $f_{l'}^2$ are compared to obtain L' comparisons $c_{k,l'} = (f_{k,l'}^2 - f_{k,l'}^1)^2$, where k is the k th feature of the layer l' , that enhance the change between bi-temporal features. Assuming the comparisons with high variance values provide more information about changes, a variance-based feature selection method is applied to select a fixed number m of features with the highest variance values. The selected comparisons are aggregated for each layer l' to retrieve a difference image (DI).

$$DI_{l'} = \sqrt{\sum_{k=1}^m c_{k,l'}}. \quad (15)$$

The L' multi-scale DIs are used in a detail-preserving multi-scale change detection method [10] (see Sect. 2.3) that exploits the multi-scale information to preserve the geometrical details of the change map and improve the detection of the change area borders. CAEs allow learning informative features using unlabeled data easily. However, they do not preserve the geometrical details of hidden-layer features since

the contribution of hidden layers is not considered in the loss function during the training. The feature selection of a fixed number is suboptimal since informative features might be excluded or non-informative ones kept.

3 Unsupervised Change Detection in SAR Images Using Deep Learning Multi-scale Features

Change detection (CD) task is studied in many state-of-the-art (SoA) methods, either deep learning (DL) approaches or not (see Sect. 2). Most of the DL models are supervised and require many labeled data (especially DL methods) that are challenging to gather in the multi-temporal case. Domain adaptation (DA) methods alleviate this problem since they exploit the DL model trained with given labeled samples (source domain) to process data acquired in another geographical area (target domain). Some DA CD methods transfer the knowledge between heterogeneous RS data [36, 37] or between existing labeled and unlabeled RS images [38]. However, the DA of pre-trained models is still a challenging open issue in the literature and still requires labeled data. For this reason, unsupervised CD methods are preferred.

Unsupervised CD DL models can exploit transfer learning to process target RS data with pre-trained models. However, their performance decreases using data with heterogeneous acquisition characteristics since they have different behaviors for similar objects [39]. The design of DL methods that provide information about changes is challenging. Autoencoders learn features from the input data during the unsupervised training, but they do not capture the spatial context information of images [45], which is critical for CD tasks. Convolutional autoencoders (CAEs) address this problem since they learn spatial context features from the input images in an unsupervised way [50, 51] (see Sect. 2.1). DL models provide many features, but not all of them are informative. Thus the selection of only the most informative features is a critical aspect of the DL CD method. However, many methods exploit all or a fixed number of features to detect the changes. The selection of a fixed number of features may remove informative features or include non-informative ones [51]. CD methods using all the features decrease their performance since many features contain irrelevant information about changes. Finally, many DL CD methods exploit only model output or single-scale features, so objects with various dimensions are not well detected.

We present an unsupervised CD method to detect the changes of bi-temporal SAR images that exploits multi-scale feature maps extracted from CAE layers that dynamically select and use all the informative spatial features retrieved by the hidden layers of a CAE. The CAE is trained in an unsupervised manner using unlabeled data.

The method extracts bi-temporal multi-resolution feature maps from multiple layers of the models that are compared to define multi-resolution difference feature maps. Since most of the difference feature maps have no information about changes, a standard-deviation-based feature selection [29, 68] is applied to dynamically

choose only the ones providing relevant information about the changes. The selected feature maps are aggregated to retrieve multi-resolution difference images that emphasize the changes. The multi-resolution difference images are analyzed through a detail-preserving multi-scale CD method, inspired by [10]. The method provides two reliable area detection alternatives. The first one exploits [10] to compute the coefficient of variation and retrieve the reliability maps, whereas the second alternative exploits a gradient-based method to retrieve the reliability maps.

3.1 Unsupervised Change Detection Based on Convolutional Autoencoder Feature Extraction

The presented method performs change detection (CD) of bi-temporal SAR images I_1 and I_2 acquired at times t_1 and t_2 , respectively. A set of N unlabeled samples $X = \{X_n, n = 1, \dots, N\}$ extracted from the pre-change I_1 is available. The method uses X to train from scratch a convolutional autoencoder (CAE) with L layers in an unsupervised way. I_1 and I_2 are processed through the trained model to extract bi-temporal deep feature maps of the images from $L' < L$ model layers. The feature maps are compared and fused to detect changed (ω_c) and unchanged (ω_{nc}) pixels, where ω_c includes all the relevant changes occurred in the image, while ω_{nc} represents the no change (Fig. 3).

3.1.1 Unsupervised CAE Training

CAEs have the property to produce an output image that is as similar as possible to the input one by unsupervised learning of spatial context features from a set of

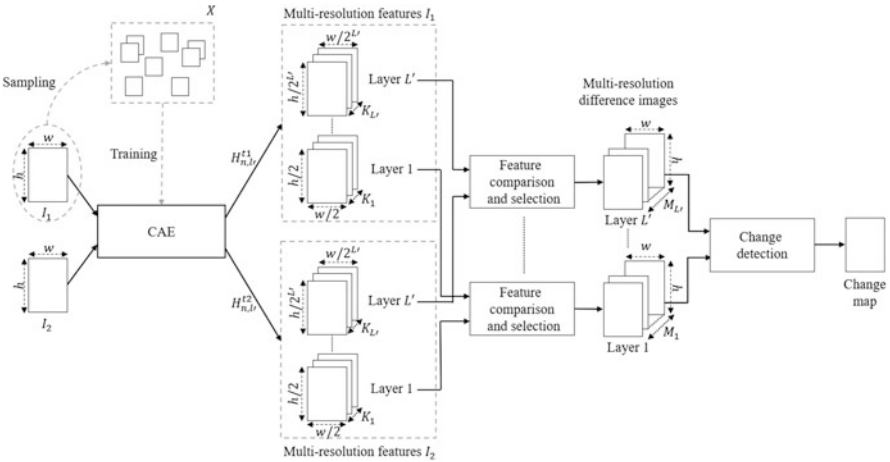


Fig. 3 Block scheme of the unsupervised CD method based on CAE

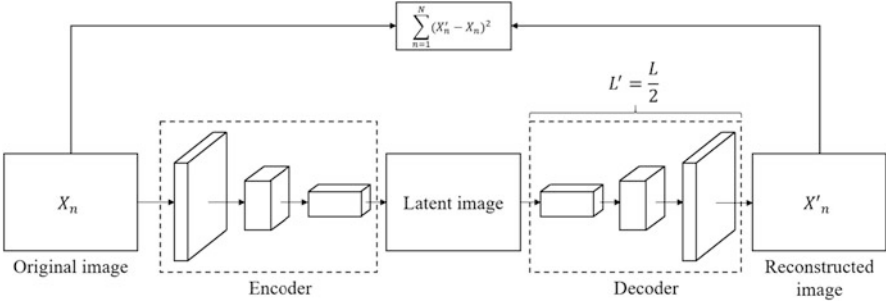


Fig. 4 Block scheme of the CAE model used in the presented method

unlabeled training samples (X). The CAE contains L convolutional layers divided into an encoder and a decoder (Fig. 4). The CAE includes strided convolutional (in the encoder) and deconvolutional layers (in the decoder), batch normalization (BN) layers, and leaky rectified linear unit (ReLU) activation functions. BN [69] layers normalize the values within a batch during its processing in the model and increase the learning speed of the model. ReLU may lead to the saturation issue [70] that badly affects the training performance since it imposes all negative values to 0. Leaky ReLU [70] improves the ReLU by keeping the non-linearity of ReLU and improving the handling of negative values. Leaky ReLU transforms the negative values into values close to 0, according to the function $y = \alpha x$. The feature maps of a layer l , where $l = 0, \dots, L$, for an input sample $X_n \in X$, where $n = 1, \dots, N$, are defined by $H_{n,l} = \phi(W_{l-1,l} * H_{n,l-1} + b_l)$, where $H_{n,0} = X_n$, $W_{l-1,l}$ is the weight matrix of the layer l processing with a convolution operation $*$ the feature maps of the layer $l - 1$, $H_{n,l-1}$. b_l represents the biases of layer l , and $\phi(\cdot)$ is the leaky ReLU activation function. The training minimizes a sum-squared error (SSE) (16) that aims to reduce the reconstruction error and retrieve an output as similar as possible to the input. Thus the model learns spatial context and semantic features from the training set X during the training. The training is performed through the back-propagation strategy that trains the model using the error between the original samples (X) and the predicted ones from CAE (X') computed with the SSE:

$$SSE = \sum_{n=1}^N (X_n - X'_n)^2. \quad (16)$$

3.1.2 Feature Extraction

Since the patches included in X are sampled from the pre-change image I_1 , it is reasonable to assume that the CAE generates spatial context feature maps representing both bi-temporal images I_1 and I_2 since they were acquired over

the same geographical area. The CAE separately processes I_1 and I_2 and extracts from a priori chosen $L' = L/2$ layers of the trained model the bi-temporal multi-resolution feature maps. In the literature [29], it was observed that the initial encoder layers extract simple features (i.e., edges), whereas the last decoder layers retrieve features providing more semantic information about the change. Thus features are not selected from the encoder layers but are retrieved from L' layers of the decoder. Thus L' is also the number of scale levels used during the multi-scale CD. The number of CAE layers L defines the number of scale levels L' . Hence, L should be carefully chosen to determine the trade-off that allows obtaining the best performance in terms of noise reduction, informative content of the learned features, and loss of geometrical details. Through the processing of the bi-temporal images, the method retrieves for each layer $l' = 1, \dots, L'$ bi-temporal feature maps representing the pre-change image I_1 , $H_{n,l'}^1$, and the post-change image I_2 , $H_{n,l'}^2$. Since the feature maps are extracted by the same model, they are in the same feature space and can be compared. The bi-temporal feature maps extracted from L' layers produce multi-scale feature maps with various spatial sizes. Thus a bi-linear interpolation method is applied to uniform the spatial dimensions of the multi-scale feature maps retrieved by the L' layers. For each layer l' , the bi-temporal feature maps $H_{n,l'}^1$ and $H_{n,l'}^2$ are compared to enhance information about changes. The k th feature maps ($k = 1, \dots, K_{l'}$) retrieved by the layer $l' = 1, \dots, L'$ of the model are compared as follows:

$$DF_{n,k,l'} = (H_{n,l',k}^2 - H_{n,l',k}^1)^2 \quad (17)$$

where

$$l' = 1, \dots, L'$$

$$k = 1, \dots, K_{l'}.$$

Comparing the bi-temporal feature maps, if no change occurs, the feature comparison results in small values close to 0 since the feature maps are similar. On the contrary, where the change occurs, the feature maps representing I_1 and I_2 are heterogeneous since their reconstructions differ. Hence, the feature map comparison highlights the changes assuming values far from 0. It is worth noting that the CAE can accurately reconstruct only objects or areas of images learned during the training phase and existing in X . If I_2 has a change with structures not in X , it will be reconstructed unpredictably. However, unpredictable reconstructions do not negatively affect the capability of CD since those structures of I_2 will be differently reconstructed with respect to I_1 in any case.

The $K_{l'}$ resulting comparisons emphasize the change between feature maps representing I_1 and I_2 , respectively, where $K_{l'}$ can be in the order of hundreds. However, not all of the comparisons provide information about changes. In each layer l' , only a limited amount of the difference feature maps $DF_{n,k,l'}$ are informative. Thus the method keeps only the most informative difference feature maps. For each layer l' , the method applies a feature selection (FS) inspired by [29, 68] to remove the

difference feature maps with a low probability of providing relevant information about changes. Assuming that comparisons with high standard deviation values carry more relevant information than the ones with low standard deviation, the $M_{l'}$ difference feature maps having a standard deviation higher than the other ones are chosen. The difference feature maps are sorted into descending order according to their standard deviation value and selected until the gap between the standard deviation of two difference feature maps is larger than 10% of the maximum standard deviation value range of layer l' :

$$|\sigma(DF_{n,k,l'}) - \sigma(DF_{n,k+1,l'})| > 0.1|\sigma_{max}(DF_{n,l'}) - \sigma_{min}(DF_{n,l'})|. \quad (18)$$

Figure 5 shows an example of the standard deviation value behavior retrieved by the difference feature maps of a single layer l' sorted in descending order. We expect a drop in the standard deviation when difference feature maps provide less change information. In this example, the gap in the standard deviation values between the first two difference feature maps and the third one is greater than 10% of the maximum standard deviation value range. Thus we assume that the first two feature maps provide the most relevant information about the change and are preserved for the CD. This feature selection step can be conducted either manually or automatically. The number of selected feature maps $M_{l'}$ can be different in each layer l' . The feature selection can choose no difference feature maps when the layer l' does not have $DF_{n,l'}$ providing a sufficiently high $|\sigma(DF_{n,k,l'}) - \sigma(DF_{n,k+1,l'})|$.

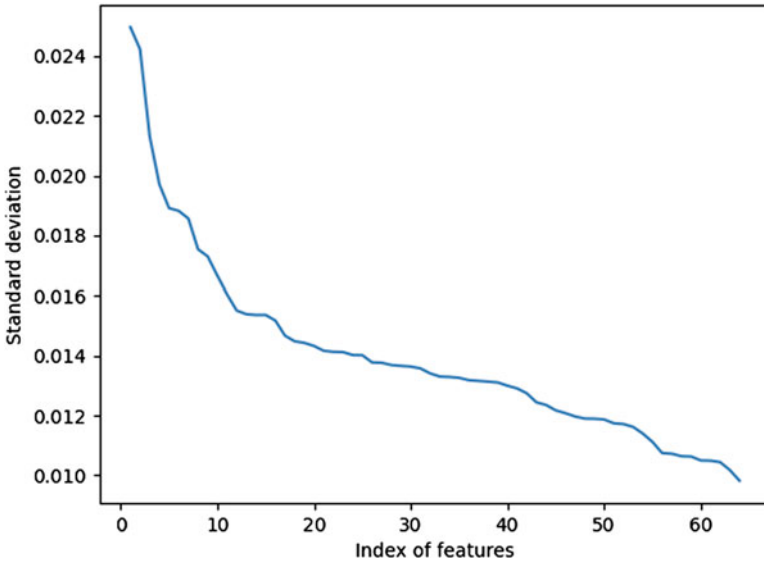


Fig. 5 Behavior of the standard deviation values of the quadratic difference between the feature maps of a layer l' sorted in descending order. In this case, only the first two difference feature maps are chosen

The selected feature maps have the highest probability of containing changes because of their high standard deviation values. Hence, they maximize the CD probability. This feature selection process analyzes all the spectral bands of the input multi-spectral RS images without using any a priori band selection step. The latter is often required to keep only the spectral bands that optimize the detection of a given change type and maximize the performance of SoA unsupervised methods with respect to this kind of change. The feature selection extracts the crucial information content from all the input spectral channels and makes the method independent of the change type. The method aggregates the selected $M_{l'}$ difference feature maps of a layer l' to compute the difference image (DI) (19) of that layer.

$$DI_{l'} = \sqrt{\sum_{k=1}^{M_{l'}} DF_{k,l'}}, \quad (19)$$

where

$$l' = 1, \dots, L'.$$

3.1.3 Change Detection

The presented method computes $DI_{l'}$ for each layer l' of the CAE with $M_{l'} > 0$ to retrieve L'_{sel} multi-resolution difference images, where L'_{sel} is the number of considered layers and difference images with $M_{l'} > 0$. Thus, it processes the L'_{sel} multi-resolution DIs with a detail-preserving multi-scale approach [10]. This method applies a multi-scale analysis to the multi-resolution $DI_{l'_{sel}}$, where $l'_{sel} = 1, \dots, L'_{sel}$, to handle the noise of the bi-temporal images and produce a change map that preserves the geometrical details and homogeneous areas. The multi-scale CD method associates to each pixel the label ω_c, ω_{nc} of the most reliable level (i.e., the lowest resolution level in which the pixel has a homogeneous behavior). The reliable levels can be identified by two strategies. The first strategy compares, for each layer l'_{sel} , the local coefficient of variation (LCV) $LCV = \sigma(DI_{l'_{sel}}(i, j)) / \mu(DI_{l'_{sel}}(i, j))$, where σ is the standard deviation and μ is the mean, computed on pixels included in a moving window centered in (i, j) of $DI_{l'_{sel}}$ with the global coefficient of variation (CV) $CV = \sigma(DI_{l'_{sel}}) / \mu(DI_{l'_{sel}})$ of the DI of that layer to find the most reliable resolution level for the pixel (i, j) :

$$\frac{\sigma(DI_{l'_{sel}}(i, j))}{\mu(DI_{l'_{sel}}(i, j))} < \frac{\sigma(DI_{l'_{sel}})}{\mu(DI_{l'_{sel}})}. \quad (20)$$

This approach assigns the pixels composing high-resolution change maps to the most heterogeneous and unreliable areas along change borders (e.g., the object contours) and the pixels of low-resolution change maps to homogeneous change areas. This method tends to overestimate the unreliable areas, thus reducing the

performance during the change detection. The second strategy for identifying the most reliable scale level is based on a gradient-based approach, i.e., the Canny filter [71]. The Canny filter detects the object edges better than other SoA edge detection methods and is less affected by various noise conditions [72]. For each of the L'_{sel} layers, the method applies a Canny filter to DI'_{sel} to retrieve a reliable map (RM'_{sel}) (21).

$$RM'_{sel} = \text{Canny}(DI'_{sel}). \quad (21)$$

Each RM'_{sel} shows the areas with the highest gradient in DI'_{sel} that represent the transitions between ω_c and ω_{nc} . These transition areas are the least reliable ones. The transition areas are thinner than the other strategy to mask fewer changes in borders. L'_{sel} possible combinations of the L'_{sel} DIs [10] are averaged:

$$\overline{DI'_{sel}} = \frac{1}{l'_{sel} + 1} \sum_{h=0}^{l'_{sel}} DI_h, \quad l'_{sel} = 0, 1, \dots, L'_{sel}. \quad (22)$$

An automatically retrieved threshold T'_{sel} (for the experiments we used an Otsu's threshold [58]) is applied to each $\overline{DI'_{sel}}$ to obtain a change map for each resolution level l'_{sel} . Each pixel of the final change map CM is assigned to ω_c or ω_{nc} , according to the class detected in position (i, j) in the change map at the resolution level l'_{rel} , where $l'_{rel} = 0, \dots, L'_{sel}$ corresponds to the most reliable resolution level for the position (i, j) [10]:

$$CM(i, j) \in \begin{cases} \omega_{nc}, & \text{if } \overline{DI'_{rel}}(i, j) \leq T'_{rel} \\ \omega_c, & \text{if } \overline{DI'_{rel}}(i, j) > T'_{rel}. \end{cases} \quad (23)$$

4 Experimental Design and Results

In this section, we present the datasets used for the tests and the experimental design. We also show and discuss the results.

4.1 Description of Datasets

As observed in Table 1, we exploited two datasets composed of bi-temporal SAR images to evaluate the flood area detection and building change detection applications. For the first application, we used a dataset composed of two bi-temporal SAR images acquired by Sentinel-1 on January 17th, 2019 (Fig. 6a), and January 29th, 2019 (Fig. 6b), representing a flooded area due to a dam failure that occurred near

Table 1 Summary of the change detection methods for SAR images and the applications suitable to use them

Change detection methods	References	Applications
Image (Log-)rationing	Bazi et al.[8]	Urban area change detection
	Grover et al.[9]	Tropical forest cover change
Kullback–Leibler distance	Inglada et al.[11]	Volcanic eruption
	Mercier et al.[12]	Volcanic eruption
Mutual information	Erten et al.[55]	Agricultural monitoring
Mixed information	Guengen et al.[17]	Agricultural monitoring
Non-deep-learning multi-scale methods	Bovolo et al.[10]	Burned area detection
	Marin et al.[24]	Building change detection
Deep learning multi-scale methods	Saha et al.[1]	Building change detection
	Saha et al.[35]	Building change detection
	Bergamasco et al.[51]	Flood area detection

the city of Brumadinho in Brazil on January 25th, 2019. The bi-temporal Sentinel-1 images have a size of 540×566 with a spatial resolution of 9 m/pixel. The reference map (Fig. 6d) shows the flooded area that points out changed (28,417) and unchanged pixels (277,223). The difference in the backscattering due to the flooded area can be observed in the false-color composition images, where green and magenta pixels represent the backscattering increasing and decreasing, respectively. We randomly sampled the pre-change image I_1 acquired on January 17th 2019 to create an unlabeled dataset composed of 37,319 patches with a size of 64×64 pixels. The dataset was divided into a validation set composed of 3731 patches and an unlabeled training set X composed of 33,588 patches.

For the second application, we used a dataset composed of bi-temporal SAR images acquired by Cosmo-SkyMed on April 5th, 2009 (Fig. 7a) and September 12th, 2009 (Fig. 7b). The bi-temporal images represent the L'Aquila urban area in Italy destroyed by an earthquake occurred on April 5th, 2009. They have a size 1024×1024 with a spatial resolution of 2.5 m/pixel. The false-color composition image (Fig. 7c) shows the backscattering changes confirming the destroyed building in the reference map (Fig. 7e) that indicates changed and unchanged pixels. The dataset is composed of 19,930 unlabeled patches with a size of 64×64 that were randomly sampled by the pre-change image I_1 acquired on April 5th. The dataset was divided into a validation set composed of 1993 unlabeled patches used for CAE validation during the training and a training set with 17,937 unlabeled patches.

In both datasets, we applied a logarithm operator to transform the behavior of the speckle noise from multiplicative to additive. We used a patch size of 64 since it was the best trade-off between the number of possible samples to include in the dataset and the spatial context information. Smaller patch sizes may lead to inadequate spatial context information modeling, while larger patch sizes would lead to fewer patches for the training. We overlapped the patches during the sampling to increase N . Thus each patch shared a part of the spatial information with the neighboring ones.

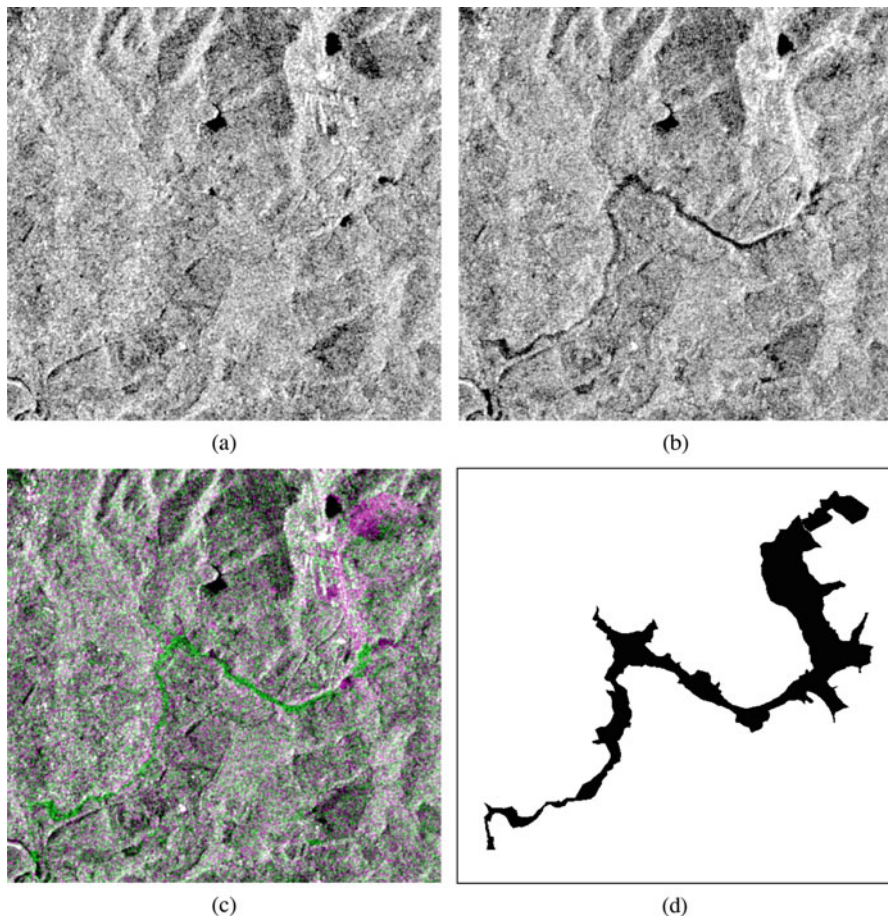


Fig. 6 Sentinel-1 images acquired in an area near Brumadinho, Brazil on (a) January 17th, 2019, (b) January 29th, 2019, (c) the false-color composition highlighting the changes between the bi-temporal images (green represents the backscattering decrease, and magenta is the increase), and (d) the reference map of the flooded area (The white pixels represent no changes and the black pixels the changes)

4.2 Design of the Experiments

We trained the CAE in an unsupervised way with an epoch number E for each dataset. The CAE training had the objective of optimizing the image reconstruction (which differs from the method goal, i.e., CD), and the validation set assessed the quality of the reconstructed image with the validation loss. When the validation loss increases, while training loss decreases, the model overfits, and the CD performance may be affected badly. However, only the multi-scale feature maps retrieved by the CAE were used for CD and not the CAE output. The CAEs for the experiments

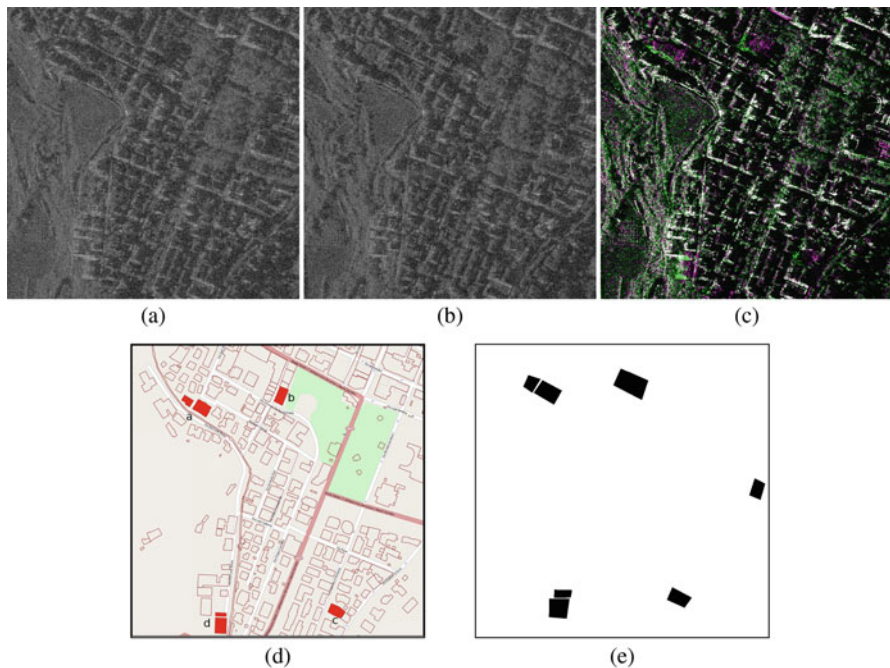


Fig. 7 Cosmo-SkyMed images acquired in L'Aquila, Italy on: (a) April 5th, 2009, (b) September 12th, 2009 (Agenzia Spaziale Italiana, 2009. All Rights Reserved.), (c) the false-color composition highlighting the changes between the bi-temporal images (green represents the backscattering decrease, and magenta is the increase), (d) the cadastral map of the area, and (e) the reference map showing the changes (the white pixels represent no changes and the black pixels the changes)

included batch normalization layers [69] and leaky ReLU activation function [70] with $\alpha = 0.2$. We trained the CAE model with learning rate $lr = 10^{-4}$ and batch size $bs = 100$. We performed several experiments on both the datasets with different parameter setups:

- *Experiment 1:* The objective of this experiment was the analysis of the method performance by varying the number of CAE layers. This hyperparameter is relevant for the retrieving of informative feature maps since the increase of the model depth leads to a better image de-noising and an increase of the model receptive field. A model with a large receptive field can better analyze the spatial context of images, which is critical in the CD of SAR images because of the complex spatial context information. On the contrary, a DL model with too many layers deteriorates the geometrical details of images. This experiment allowed finding a good trade-off between image de-noising and the generation of informative feature maps. We compared the CD performance to choose the optimal number of layers and thus the number of multi-scale levels L' using a CAE with $L = 2, 4, 6, 8$ trained with $E = 100$ for the Brumadinho dataset

and $E = 300$ for the L'Aquila dataset. The standard-deviation-based reliability approach with a window size of 3×3 was used for both datasets. The number of epochs and the reliability strategies were fixed according to the results achieved by Experiment 3, whereas the window size of the standard deviation reliability approach is used according to Experiment 2.

- *Experiment 2:* This experiment compared the CD performance by varying the window size of the standard deviation reliability approach. We tested window dimensions of 3×3 , 5×5 , and 7×7 . We used the CAE model with $L = 8$ trained with $E = 100$ for the Brumadinho dataset and the model with $L = 8$ trained with $E = 300$ for the L'Aquila dataset. The number of model layers and epochs was fixed according to the results of Experiments 1 and 3.
- *Experiment 3:* The experiment analyzed the CD performance by changing the number of training epochs E . We varied the number of training epochs $E = 50, 100, \dots, 300$ using a CAE with $L = 8$ layers for both the datasets (see Experiment 1). We tested the method using the standard-deviation- and Canny-based reliability approaches to observe the strategy providing the most accurate reliability maps. We used a moving window size of 3×3 in the standard deviation reliability approach for both datasets (see Experiment 2).
- *Experiment 4:* The experiment verified that the multi-scale strategy and the feature selection were effective. We compared the presented method against two limit cases. The first one exploited single-scale feature maps retrieved by a single CAE layer: the bottleneck. The second one used all feature maps obtained by the L' layers of the model without any feature selection step. We exploited in both datasets the CAE with $L = 8$ layers (see Experiment 1) and the standard-deviation-based reliability approach with a window size of 3×3 and used CAE trained for $E = 100$ for the Brumadinho dataset and $E = 300$ for the L'Aquila dataset. We fixed the epoch number and the window size of the standard deviation reliability strategy according to the results provided by Experiments 3 and 2, respectively.

SoA comparisons for the L'Aquila dataset were performed against: (i) a fuzzy-based building CD method [24], (ii) a transfer learning method used to extract features for building CD [1], and (iii) a building CD using a CycleGAN to transcode SAR and optical images [35]. For Brumadinho dataset, we compared the presented method using the optimal parameters retrieved from previous experiments (Table 2) with an unsupervised change detection method using feature maps extracted from a CAE [51]. To evaluate the performance of the tested methodologies for the Brumadinho dataset, we considered the number of the truly detected changes (TPs), the false-alarm rate (FAs), the missed-alarm rate (MAs), the overall errors (OE),

Table 2 The optimal parameters to use in the two datasets

Datasets	L	E	Window size	Learning rate	Batch size
Brumadinho	8	100	3×3	10^{-4}	100
L'Aquila	8	300	3×3	10^{-4}	100

the sensitivity ($TP/(TP+FN)$), and the specificity ($TN/(TN+FP)$). For the L'Aquila dataset, we considered the number of correctly, missed, and falsely detected changed buildings, as in [1, 24, 35]. To do so, we exploited the change map to cluster the changed pixels, as in [35], into two classes: increment (ϵ^+) and decrement (ϵ^-) of backscattering. The two classes are used to determine new and destroyed buildings using a fuzzy-based strategy [24, 35].

4.3 Experiment 1: Analysis of the Performance Varying the Number of Layers

Experiment 1 observed the CD performance of the presented method by varying the depth of the CAE to find the optimal trade-off between the generation of informative feature maps and the preservation of spatial information. In the Brumadinho dataset, the method achieved the best performance using a CAE with $L = 8$ layers. The use of CAE with $L = 8$ layers outperformed the other options with $L = 2, 4, 6$ from the point of view of the true detected changed areas and FAs (see Table 3). These observations were confirmed by the qualitative results that showed a sharp improvement in the CD performance from CAE with $L = 2, 4, 6$ to one with $L = 8$ (see Fig. 8). The latter detected most of the changed areas homogeneously and found fewer FAs (see Fig. 8d). This improvement was due to the de-noising capability and the large receptive field achieved with this model depth. The de-noising capacity reduced the noise contribution on the feature maps, whereas the large receptive field allowed the model to analyze a larger spatial context and produce more informative feature maps.

As in the previous dataset, the method applied to the L'Aquila dataset achieved its best performance using a CAE with $L = 8$. Also in this case, the CAE with $L = 8$ allowed to improve the detection of true changed areas by reducing the number of FAs (see Table 4). The number of model layers is strongly related to the size of the changes with respect to the spatial resolution of the data. Considering the same object acquired by sensors with heterogeneous spatial resolution, the one acquired by the sensor with the higher spatial resolution has more details, and it needs a larger receptive field to analyze the spatial context than the one acquired by a

Table 3 FAs, MAs, TPs, OE, sensitivity, and specificity of the presented method vs. the number of layers L in the Brumadinho dataset trained for $E = 100$

L	FA	MA	TP	OE	Sensitivity	Specificity
2	2.32%	83.64%	4648	9.88%	16.36%	97.68%
4	2.25%	81.66%	5212	9.63%	18.34%	97.65%
6	2.06%	80.47%	5549	9.35%	19.53%	97.94%
8	1.96%	66.47%	9527	7.96%	33.53%	98.04%

The best experimental results are in bold

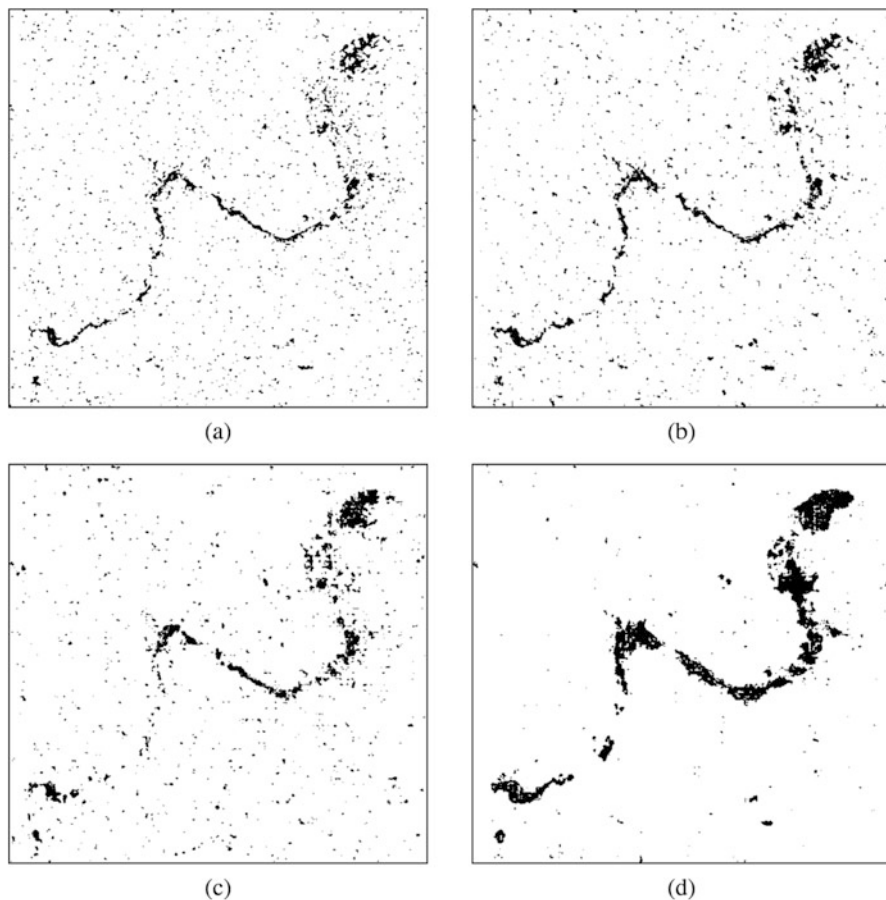


Fig. 8 Change maps computed using (a) $L' = 1$ of the CAE with $L = 2$, (b) $L' = 2$ of the CAE with $L = 4$, (c) $L' = 3$ of the CAE with $L = 6$, and (d) $L' = 4$ of the CAE with $L = 8$. All the CAEs are trained for $E = 100$ using the Brumadinho dataset. Standard-deviation-based reliability approach was used to find the most reliable areas (The white pixels represent no changes, and the black pixels the changes)

sensor with lower spatial resolution. In both datasets, the changes are characterized by many correlated neighboring pixels and cover large areas with respect to the sensor resolution. Thus a wide receptive field achieved by models with many layers was needed to properly catch the spatial context information of the changes. This improvement can also be observed in the qualitative results where the CD method using a CAE with $L = 8$ more changed areas with fewer FAs than using $L = 2, 4, 6$. The model depth is more critical for the L'Aquila dataset than the Brumadinho one since the L'Aquila dataset is composed of very high-resolution (VHR) images. With VHR images, the processing of the spatial context information is more relevant because of the many geometrical details and the strong spatial correlation between

Table 4 FAs, MAs, TPs, OE, sensitivity, and specificity of the presented method vs. the number of layers L in the L'Aquila dataset trained for $E = 300$

L	FA	MA	TP	OE	Sensitivity	Specificity
2	0.94%	88.84%	2894	3.11%	11.16%	99.06%
4	1.52%	88.5%	2984	3.67%	11.5%	98.48%
6	0.7%	89.0%	2853	2.88%	11.0%	99.3%
8	0.69%	71.72%	7336	2.45%	28.28%	99.31%

The best experimental results are in bold

neighboring pixels. In this case, a deep model (e.g., $L = 8$) is necessary to achieve good results since the more a model is deep, the more its receptive field is large. Thus it can better analyze the spatial context information and produce informative feature maps that model the spatial context information. We exploited a CAE with $L = 8$ to process the two datasets since the experimental results showed the model superiority with this layer number in terms of detected changed areas and FAs (Fig. 9). Thus the CAE used in the experiments is composed of four layers for both encoder and decoder (see Table 5).

4.4 Experiment 2: Analysis of the Performance Varying the Window Size of the Standard-Deviation-Based Reliability Approach

Experiment 2 compared the CD performance of the presented method by varying the moving window size of the standard deviation reliability approach. The window size affects the method reliability in the detection of heterogeneous areas. The smaller the window size, the less reliable the local coefficient of variation (LCV) is. The bigger the window size, the lower the preservation of the geometrical details. Thus we can find the optimal window size by preserving the geometrical details without removing changed pixels. In the Brumadinho dataset, the 3×3 moving window size allowed detecting more change areas than the other options by keeping the number relatively low (see Table 6). The method using a 7×7 window size found the lowest number of FAs, but it missed many changed areas. For our experiments, we used a window size of 3×3 since it can detect more changed areas and occurred in the least percentage of OE. In the L'Aquila dataset, we observed the same behavior as the previous dataset. The method using a 3×3 moving window detected more changed areas than the other options with slightly higher FAs and gave the least percentage of OE (Table 7). The method using a 7×7 window size retrieved fewer FAs than the 3×3 option, but it did not preserve geometrical details with the MAs increasing. For these reasons, we exploited a 3×3 moving window size during the experiments.

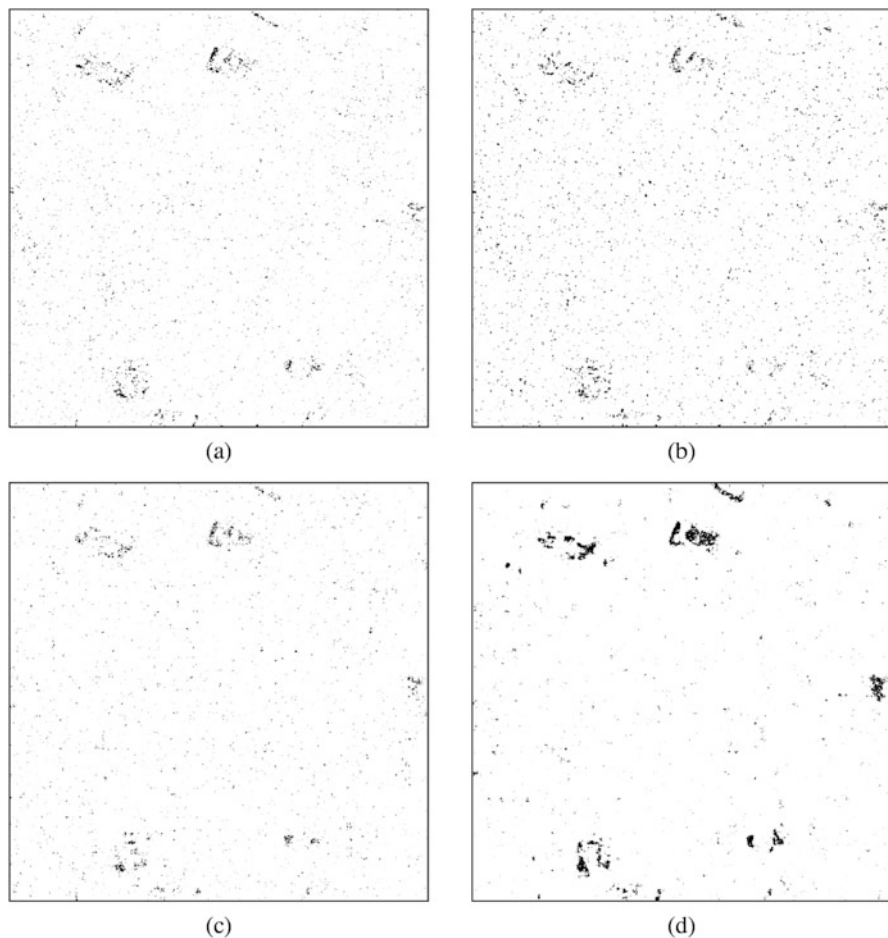


Fig. 9 Change maps computed using (a) $L' = 1$ of the CAE with $L = 2$, (b) $L' = 2$ of the CAE with $L = 4$, (c) $L' = 3$ of the CAE with $L = 6$, and (d) $L' = 4$ of the CAE with $L = 8$. All the CAEs are trained for $E = 300$ using the L'Aquila dataset. Standard-deviation-based reliability approach was used to find the most reliable areas (The white pixels represent no changes, and the black pixels the changes)

4.5 Experiments 3 and 4: Brumadinho Dataset

Experiment 3 proved that the presented method is not significantly affected by the variation of E from the point of view of the FAs. The variation of E influenced the capability of the model to detect truly changed areas (see Table 8). The method achieved the best performance between $E = 100$ and $E = 150$ by maximizing the detection of the changed areas while keeping a small number of OEs. The Canny-based strategy allowed retrieving fewer FAs, but it increased the MAs. For the experiments, we used a CAE trained for $E = 100$ exploiting the standard deviation

Table 5 Structure of the convolutional autoencoder used during tests for the Granada and Elba datasets

Layer type	Kernel size	Output size
Input	–	$64 \times 64 \times 6$
Strided Conv2D	$5 \times 5 \times 32$	$32 \times 32 \times 32$
Batch normalization	–	$32 \times 32 \times 32$
Leaky ReLU	–	$32 \times 32 \times 32$
Strided Conv2D	$5 \times 5 \times 64$	$16 \times 16 \times 64$
Batch normalization	–	$16 \times 16 \times 64$
Leaky ReLU	–	$16 \times 16 \times 64$
Strided Conv2D	$5 \times 5 \times 128$	$8 \times 8 \times 128$
Batch normalization	–	$8 \times 8 \times 128$
Leaky ReLU	–	$8 \times 8 \times 128$
Strided Conv2D	$5 \times 5 \times 256$	$4 \times 4 \times 256$
Batch normalization	–	$4 \times 4 \times 256$
Leaky ReLU	–	$4 \times 4 \times 256$
Deconv2D	$5 \times 5 \times 128$	$8 \times 8 \times 128$
Batch normalization	–	$8 \times 8 \times 128$
Leaky ReLU	–	$8 \times 8 \times 128$
Deconv2D	$5 \times 5 \times 64$	$16 \times 16 \times 64$
Batch normalization	–	$16 \times 16 \times 64$
Leaky ReLU	–	$16 \times 16 \times 64$
Deconv2D	$5 \times 5 \times 32$	$32 \times 32 \times 32$
Batch normalization	–	$32 \times 32 \times 32$
Leaky ReLU	–	$32 \times 32 \times 32$
Deconv2D	$5 \times 5 \times 6$	$64 \times 64 \times 6$

Table 6 FAs, MAs, TPs, OE, sensitivity, and specificity of the presented method by varying the window size with the standard-deviation-based reliability approach in the Brumadinho dataset. The results were retrieved by using $L' = 4$ multi-scale feature maps extracted by a CAE trained for $E = 100$

Window size	FA	MA	TP	OE	Sensitivity	Specificity
3×3	1.96%	66.47%	9527	7.96%	33.53%	98.04%
5×5	1.63%	71.56%	8082	8.13%	28.44%	98.37%
7×7	1.27%	77.65%	6350	8.37%	22.35%	98.73%

The best experimental results are in bold

reliability approach since it is the combination that detects more changed areas while keeping a low OE percentage.

The presented method achieved comparable performance to the SoA CD method using CAE feature maps [51]. The latter detected slightly more changed areas than the presented one. However, the method had a lower OE percentage than the SoA one and found fewer FAs (see Table 9). The strategy using the Canny filter retrieved fewer FAs than the standard deviation one. However, it missed many changed areas and obtained more OEs than the standard deviation strategy. Qualitative results (see

Table 7 FAs, MAs, TPs, OE, sensitivity, and specificity of the presented method by varying the window size with the standard-deviation-based reliability approach in the L'Aquila dataset. The results were retrieved by using $L' = 4$ multi-scale feature maps extracted by a CAE trained for $E = 300$

Window size	FA	MA	TP	OE	Sensitivity	Specificity
3×3	0.69%	71.72%	7336	2.45%	28.28%	99.31%
5×5	0.45%	79.66%	5277	2.41%	20.34%	99.55%
7×7	0.34%	85.46%	3772	2.45%	14.54%	99.66%

The best experimental results are in bold

Table 8 FAs, MAs, TPs, OE, sensitivity, and specificity of the presented method vs. the number of epochs on the CAE training and by using a standard-deviation-based and Canny-filter-based approach to find the most reliable areas (Brumadinho dataset)

E	Reliab. approach	FA	MA	TP	OE	Sens.	Spec.
50	Canny	1.21%	78.16%	6206	8.37%	21.84%	98.79%
	Std. dev.	1.26%	77.19%	6482	8.32%	22.81	98.74%
100	Canny	1.68%	70.1%	8497	8.04%	29.9%	98.32%
	Std. dev.	1.96%	66.74%	9527	7.96%	33.52%	98.04%
150	Canny	1.59%	71.07%	8221	8.05%	28.93%	98.41%
	Std. dev.	1.81%	67.33%	9284	7.91%	32.67%	98.19%
200	Canny	1.17%	76.63%	6642	8.19%	23.37%	98.83%
	Std. dev.	1.28%	74.42%	7269	8.08%	25.58%	98.72%
250	Canny	1.58%	72.9%	7700	8.21%	27.1%	98.42%
	Std. dev.	1.57%	71.17%	8192	8.04%	28.83%	98.43%
300	Canny	1.21%	75.55%	6947	8.13%	22.45%	98.79%
	Std. dev.	1.28%	74.04%	7365	8.05%	25.92%	98.72%

The best experimental results are in bold

Fig. 10) confirmed the quantitative ones. The SoA and presented methods retrieved comparable results. The SoA one detected more changed areas (e.g., the bottom part of Fig. 10b) than the presented method. However, the latter found fewer FAs than the SoA method (e.g., the bottom part of Fig. 10e). We can notice that standard deviation and Canny-based approaches retrieved similar results. However, the latter obtained some artifacts due to the value of the standard deviation of the Gaussian kernel used in the Canny filter (e.g., upper part of Fig. 10f).

Experiment 4 proved the effectiveness of the multi-scale strategy by comparing the results achieved using a detail-preserving multi-scale CD [10] with the one using single-scale feature maps. The use of the multi-scale strategy allowed the detection of many changed areas and preserved many geometrical details (see Fig. 10e). The single-scale strategy detected only the wider and most homogeneous areas and lost many changed areas with small geometrical details (see Fig. 10c). The method performance using the feature selection (see Fig. 10e) and the one without feature selection (see Fig. 10d) provided similar results. However, the method using the feature selection exploited only the most informative feature maps and reduced the computational burden of the CD.

Table 9 FAs, MAs, sensitivity, specificity, and OE (in the number of pixels and percentage) obtained by the SoA methods and the presented one. We tested the presented method using the two reliability approaches, a single-scale feature, and no feature selection (Brumadinho dataset)

Method	FA	MA	Sens.	Spec.	OE	
					Pixels	%
CAE CD [51]	2.89%	60.65%	39.35%	97.11%	25238	8.26
Presented CAE CD w/ std. dev.	1.96%	66.47%	33.53%	98.04%	24323	7.96%
Presented CAE CD w/ Canny	1.68%	70.1%	29.9%	98.32%	24583	8.04%
Presented CAE CD w/ single-scale feat.	1.97%	71.41%	28.59%	98.03%	25764	8.43%
Presented CAE CD no feat. sel.	2.01%	65.36%	34.64%	97.99%	24133	7.9%

The best experimental results are in bold

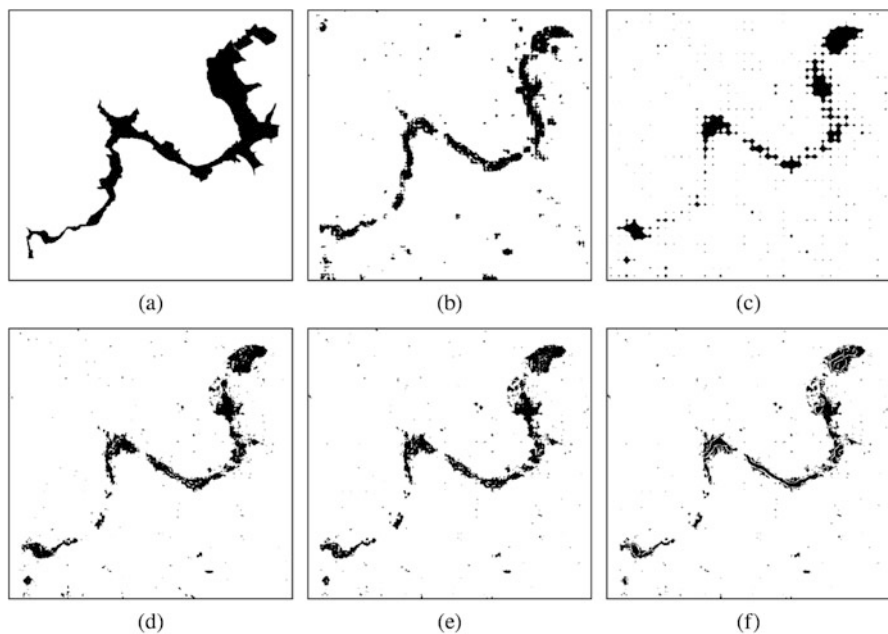


Fig. 10 Comparisons between (a) the reference map of the dataset of Brumadinho and the maps achieved by applying (b) unsupervised SoA CD method using the CAE, the presented method exploiting (c) a single-scale feature, (d) no feature selection, (e) standard-deviation-based reliability maps, and (f) canny filter (The white pixels represent no changes, and the black pixels represent the changes)

4.6 Experiments 3 and 4: L'Aquila Dataset

Experiment 3 showed that the variation of E did not significantly affect the performance of the presented method from the FA point of view. The E variation affected the MA percentage. Thus the increase of E led to a changed area detection increase. The method achieved the best CD performance using a CAE trained for $E = 300$. The Canny-based strategy detected more changed areas than the standard deviation one until $E = 250$ and achieved similar results from the MA point of view with respect to the standard deviation reliability approach (see Table 10). However, the latter produced fewer FAs than the Canny approach. During the experiments, we exploited a CAE trained for $E = 300$ using the standard deviation reliability approach since it improved the detection of changed areas and minimized the OEs.

The method using CAE feature maps retrieves similar results than the other SoA methods (see Table 11). It detected all destroyed buildings with no false alarms, as the other DL-based methods. The DL neural networks better modeled the spatial context information of the scenario than no-DL approaches, such as DWT used in [24]. The better modeling allowed to find less false alarms. We can observe from the qualitative results (see Fig. 11) that DL-based method methods detected less FAs (see Fig. 11d, f, h) than the DWT-based one (see Fig. 11b). In this way, the DL-based method did not detect any false destroyed buildings (see Fig. 11e, g, i). The feature maps extracted from a DL model trained with the data (i.e., CycleGAN- and CAE-based methods; see Fig. 11f, h) to analyze provided a more accurate spatial information than the ones extracted by an adapted DL model (i.e., AdaBN-based model; see Fig. 11d). The CycleGAN- and CAE-based model

Table 10 FAs, MAs, TPs, OE, sensitivity, and specificity of the presented method vs. the number of epochs on the CAE training and by using a standard-deviation-based and Canny-filter-based approach to find the most reliable areas (L'Aquila dataset)

E	Reliab. Approach	FA	MA	TP	OE	Sens.	Spec.
50	Canny	1.07%	76.29%	6151	2.93%	23.71%	98.23%
	Std. dev.	0.87%	79.24%	5203	2.82%	20.06	99.13%
100	Canny	1.26%	78.62%	5545	3.17%	21.38%	98.74%
	Std. dev.	0.84%	81.87%	4704	2.84%	18.13%	99.16%
150	Canny	0.87%	79.37%	5352	2.81%	20.63%	99.13%
	Std. dev.	0.53%	80.08%	5167	2.5%	19.92%	99.46%
200	Canny	1.28%	75.86%	6611	2.84%	25.48%	98.98%
	Std. dev.	0.75%	80.02%	5182	2.71%	19.98%	99.25%
250	Canny	1.02%	74.52%	6611	2.84%	25.48%	98.98%
	Std. dev.	0.73%	73.83%	6788	2.54%	26.17%	99.27%
300	Canny	0.96%	71.75%	7329	2.72%	28.25%	99.04%
	Std. dev.	0.69%	71.72%	7336	2.45%	28.28%	99.31%

The best experimental results are in bold

Table 11 The number of correctly, missed, and falsely detected destroyed buildings in the L'Aquila dataset (200 buildings in total)

Method	Correctly detected destroyed buildings	Missed destroyed buildings	Falsely detected destroyed buildings
Fuzzy-based method [24]	6	0	1
AdaBN-based method [1]	6	0	0
CycleGAN-based method [35]	6	0	0
CAE-based method	6	0	0

Table 12 FAs, MAs, sensitivity, specificity, and OE (in the number of pixels and percentage) obtained by testing the presented method, a single-scale feature, and no feature selection (L'Aquila dataset)

Method	FA	MA	Sens.	Spec.	OE	
					Pixels	%
Presented CAE CD w/std. dev.	0.69%	71.72%	28.28%	99.31%	25967	2.45%
Presented CAE CD w/single-scale feat.	2.2%	89.34%	10.66%	97.8%	45716	4.36%
Presented CAE CD no feat. sel.	1.01%	52.65%	47.35%	98.99%	23940	2.28%

The best experimental results are in bold

achieved similar performance. However, the latter had a lower computational burden than the CycleGAN-based one since it exploited a simpler DL model.

Experiment 4 proved the effectiveness of the multi-scale strategy that can detect more changed areas than the single-scale option (see Table 12). The latter lost many geometrical details. This aspect and the complexity of the urban scenario and VHR SAR data characterized by multiple reflections due to the buildings led to a poor change detection (see Fig. 12b). The multi-scale approach allowed a deeper analysis of the spatial information that improved the CD performance (see Fig. 12d). The method using feature selection found fewer FAs than without feature selection. It detected more changed areas without applying the feature selection (see Table 12), but the increase of the FAs led to not distinguishing adjacent buildings (see the bottom-left side of Fig. 12c). Thus the feature selection provided better performance to identify single constructions and achieve better performance in the building detection.

5 Conclusions

This chapter presented a series of change detection (CD) methods to analyze synthetic aperture radar (SAR) data. We presented the main change indexes and the single-scale and multi-scale approaches to handle the speckle noise and the spatial

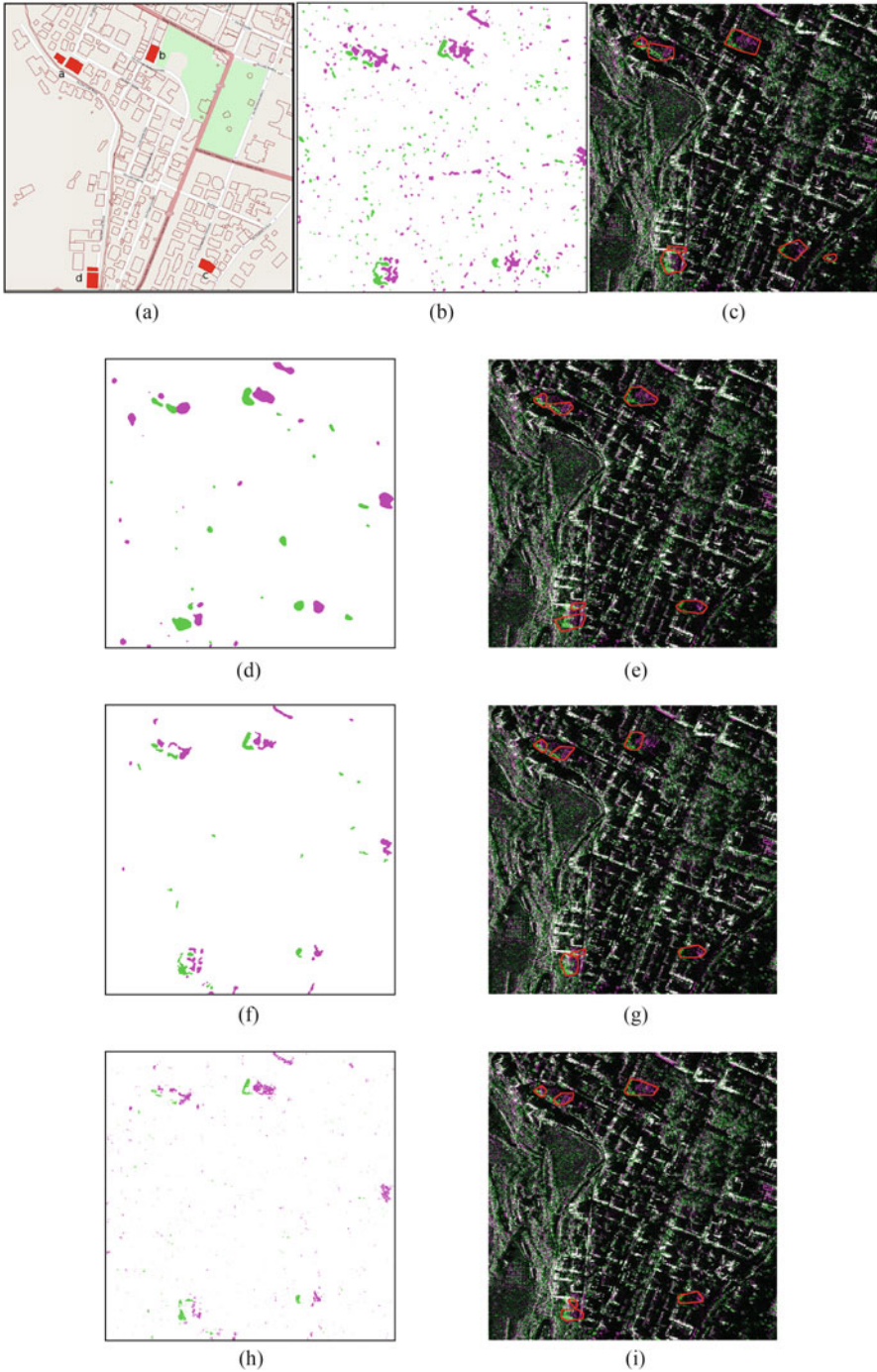


Fig. 11 Comparisons between: (a) the cadastral map of the area and the maps of increased and decreased areas (green is decreased, and magenta is decreased) and the detected buildings retrieved using (b,c) the Fuzzy-based method, (d,e) AdaBN-based method, (f,g) CycleGAN-based method, and (h,i) the presented method

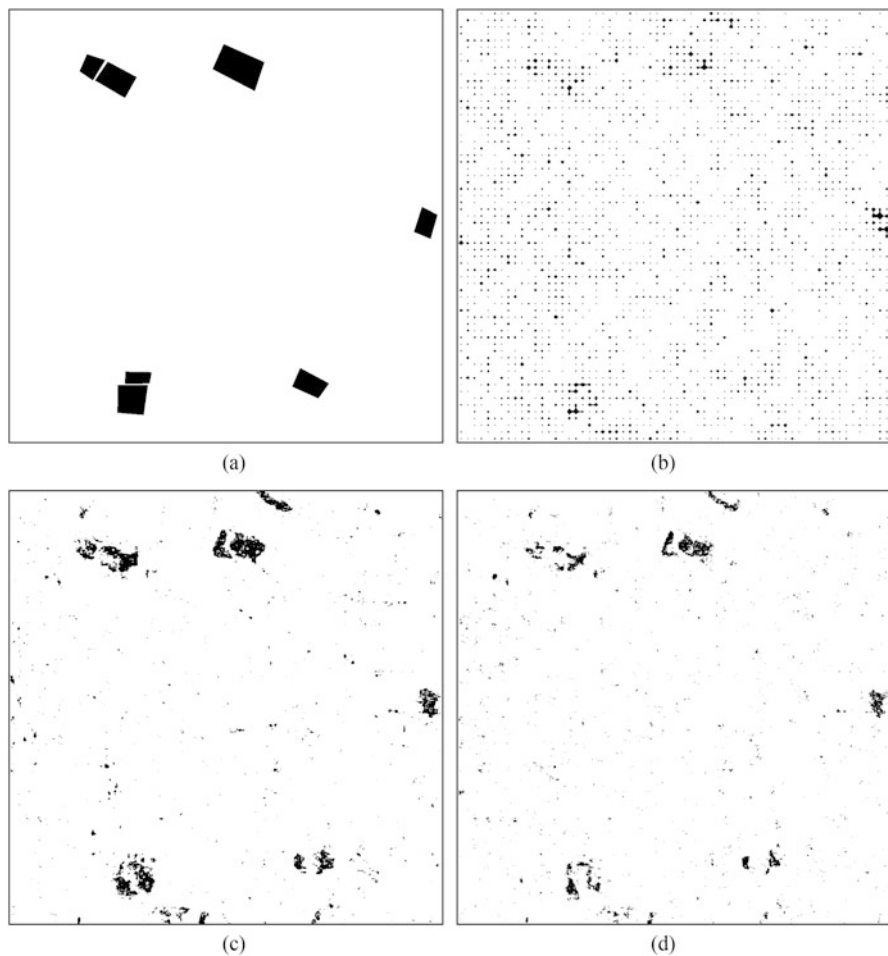


Fig. 12 Comparisons between (a) the reference map of the dataset of L'Aquila and the maps achieved by applying the presented method exploiting (b) a single-scale feature, (c) no feature selection, and (d) standard-deviation-based reliability maps (The white pixels represent no changes, and the black pixels the changes)

information in SAR images and detect relevant changes. We observed how the deep learning (DL) models automatically learn spatial context features that are used in the CD without the need to design novel features for new tasks or different SAR data. We presented a CD method using multi-scale feature maps extracted from a CAE trained in an unsupervised way. We applied this method and other SoA CD methods to two applications (i.e., flood area detection and building change detection) where the spatial context analysis and the de-noising capability are critical. The SoA and the presented methods proved the effectiveness of the features learned by DL models and multi-scale strategies in the SAR image analysis.

As it is possible to observe in this chapter, most of the CD methods process only single-sensor images and limit the analysis of the scene to a single type of change information. Image time series composed of multi-sensor multi-frequency SAR data can improve the CD performance. However, the fusion of these heterogeneous data is challenging. DL models can analyze multi-sensor data effectively. Thus one of the future steps to improve the CD method is the analysis of bi-temporal and multi-temporal heterogeneous SAR data to extract spatial and temporal information for change detection purpose.

References

1. Saha, S., Bovolo, F., Bruzzone, L. (2018, October). Destroyed-buildings detection from VHR SAR images using deep features. In *Image and Signal Processing for Remote Sensing XXIV* (Vol. 10789, p. 107890Z). International Society for Optics and Photonics.
2. Giustarini, L., Hostache, R., Matgen, P., Schumann, G. J. P., Bates, P. D., Mason, D. C. (2012). A change detection approach to flood mapping in urban areas using TerraSAR-X. *IEEE Transactions on Geoscience and Remote Sensing*, 51(4), 2417–2430.
3. Schlaffer, S., Matgen, P., Hollaus, M., Wagner, W. (2015). Flood detection from multi-temporal SAR data using harmonic analysis and change detection. *International Journal of Applied Earth Observation and Geoinformation*, 38, 15–24.
4. Li, Y., Martinis, S., Plank, S., Ludwig, R. (2018). An automatic change detection approach for rapid flood mapping in Sentinel-1 SAR data. *International Journal of Applied Earth Observation and Geoinformation*, 73, 123–135.
5. Conradsen, K., Nielsen, A. A., Schou, J., Skriver, H. (2003). 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*, 41(1), 4–19.
6. Balenzano, A., Mattia, F., Satalino, G., Davidson, M. W. (2010). Dense temporal series of C-and L-band SAR data for soil moisture retrieval over agricultural crops. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 4(2), 439–450.
7. Bovolo, F., Bruzzone, L. (2015). The time variable in data fusion: A change detection perspective. *IEEE Geoscience and Remote Sensing Magazine*, 3(3), 8–26.
8. Bazi, Y., Bruzzone, L., Melgani, F. (2005). An unsupervised approach based on the generalized Gaussian model to automatic change detection in multitemporal SAR images. *IEEE Transactions on Geoscience and Remote Sensing*, 43(4), 874–887.
9. Grover, K., Quegan, S., da Costa Freitas, C. (1999). Quantitative estimation of tropical forest cover by SAR. *IEEE Transactions on Geoscience and Remote Sensing*, 37(1), 479–490.
10. Bovolo, F., Bruzzone, L. (2005). A detail-preserving scale-driven approach to change detection in multitemporal SAR images. *IEEE Transactions on Geoscience and Remote Sensing*, 43(12), 2963–2972.
11. Inglada, J., Mercier, G. (2007). A new statistical similarity measure for change detection in multitemporal SAR images and its extension to multiscale change analysis. *IEEE Transactions on Geoscience and Remote Sensing*, 45(5), 1432–1445.
12. Mercier, G., Moser, G., Serpico, S. B. (2008). Conditional copulas for change detection in heterogeneous remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, 46(5), 1428–1441.
13. Cebrián, M., Alfonso, M., Ortega, A. (2007). The normalized compression distance is resistant to noise. *IEEE Transactions on Information Theory*, 53(5), 1895–1900.
14. Chatelain, F., Tourneret, J. Y., Inglada, J., Ferrari, A. (2007). Bivariate gamma distributions for image registration and change detection. *IEEE Transactions on Image Processing*, 16(7), 1796–1806.

15. Aiazzi, B., Alparone, L., Baronti, S., Garzelli, A., Zoppetti, C. (2013). Nonparametric change detection in multitemporal SAR images based on mean-shift clustering. *IEEE Transactions on Geoscience and Remote Sensing*, 51(4), 2022–2031.
16. Meilă, M. (2003). Comparing clusterings by the variation of information. In *Learning theory and kernel machines* (pp. 173–187). Springer, Berlin, Heidelberg.
17. Gueguen, L., Datcu, M. (2009, July). Mixed information measure: Application to change detection in earth observation. In *MultiTemp 2009: 5h Int. Workshop Analysis of Multi-temporal Remote Sensing Images*.
18. Bovolo, F., Bruzzone, L., Marchesi, S. (2007, July). A multiscale technique for reducing registration noise in change detection on multitemporal VHR images. In *2007 International Workshop on the Analysis of Multi-temporal Remote Sensing Images* (pp. 1–6). IEEE.
19. Celik, T. (2009). Multiscale change detection in multitemporal satellite images. *IEEE Geoscience and Remote Sensing Letters*, 6(4), 820–824.
20. Liu, S., Du, Q., Tong, X., Samat, A., Bruzzone, L., Bovolo, F. (2017). Multiscale morphological compressed change vector analysis for unsupervised multiple change detection. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 10(9), 4124–4137.
21. Hall, O., Hay, G. J. (2003). A multiscale object-specific approach to digital change detection. *International Journal of Applied Earth Observation and Geoinformation*, 4(4), 311–327.
22. Zhang, Y., Peng, D., Huang, X. (2017). Object-based change detection for VHR images based on multiscale uncertainty analysis. *IEEE Geoscience and Remote Sensing Letters*, 15(1), 13–17.
23. Moser, G., Angiati, E., Serpico, S. B. (2011). Multiscale unsupervised change detection on optical images by Markov random fields and wavelets. *IEEE Geoscience and Remote Sensing Letters*, 8(4), 725–729.
24. Marin, C., Bovolo, F., Bruzzone, L. (2014). Building change detection in multitemporal very high resolution SAR images. *IEEE Transactions on Geoscience and Remote Sensing*, 53(5), 2664–2682.
25. Liu, J., Gong, M., Qin, K., Zhang, P. (2016). A deep convolutional coupling network for change detection based on heterogeneous optical and radar images. *IEEE Transactions on Neural Networks and Learning Systems*, 29(3), 545–559.
26. Zhan, Y., Fu, K., Yan, M., Sun, X., Wang, H., Qiu, X. (2017). Change detection based on deep Siamese convolutional network for optical aerial images. *IEEE Geoscience and Remote Sensing Letters*, 14(10), 1845–1849.
27. El Amin, A. M., Liu, Q., Wang, Y. (2016, July). Convolutional neural network features based change detection in satellite images. In *First International Workshop on Pattern Recognition* (Vol. 10011, pp. 181–186). SPIE.
28. Mou, L., Bruzzone, L., Zhu, X. X. (2018). Learning spectral-spatial-temporal features via a recurrent convolutional neural network for change detection in multispectral imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 57(2), 924–935.
29. Saha, S., Bovolo, F., Bruzzone, L. (2019, July). Unsupervised multiple-change detection in VHR multisensor images via deep-learning based adaptation. In *IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium* (pp. 5033–5036). IEEE.
30. Zhang, P., Gong, M., Su, L., Liu, J., Li, Z. (2016). Change detection based on deep feature representation and mapping transformation for multi-spatial-resolution remote sensing images. *ISPRS Journal of Photogrammetry and Remote Sensing*, 116, 24–41.
31. Khan, S. H., He, X., Porikli, F., Bennamoun, M. (2017). Forest change detection in incomplete satellite images with deep neural networks. *IEEE Transactions on Geoscience and Remote Sensing*, 55(9), 5407–5423.
32. Niu, X., Gong, M., Zhan, T., Yang, Y. (2018). A conditional adversarial network for change detection in heterogeneous images. *IEEE Geoscience and Remote Sensing Letters*, 16(1), 45–49.
33. Saha, S., Bovolo, F., Bruzzone, L. (2019, July). Unsupervised multiple-change detection in VHR multisensor images via deep-learning based adaptation. In *IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium* (pp. 5033–5036). IEEE.

34. Hou, B., Liu, Q., Wang, H., Wang, Y. (2019). From W-Net to CDGAN: Bitemporal change detection via deep learning techniques. *IEEE Transactions on Geoscience and Remote Sensing*, 58(3), 1790–1802.
35. Saha, S., Bovolo, F., Bruzzone, L. (2020). Building change detection in VHR SAR images via unsupervised deep transcoding. *IEEE Transactions on Geoscience and Remote Sensing*, 59(3), 1917–1929.
36. Kang, C., He, C. (2016, July). SAR image classification based on the multi-layer network and transfer learning of mid-level representations. In *2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)* (pp. 1146–1149). IEEE.
37. Ye, F., Luo, W., Dong, M., He, H., Min, W. (2019). SAR image retrieval based on unsupervised domain adaptation and clustering. *IEEE Geoscience and Remote Sensing Letters*, 16(9), 1482–1486.
38. Yang, M., Jiao, L., Hou, B., Liu, F., Yang, S. (2020). Selective adversarial adaptation-based cross-scene change detection framework in remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, 59(3), 2188–2203.
39. Saha, S., Solano-Correa, Y. T., Bovolo, F., Bruzzone, L. (2020). Unsupervised deep transfer learning-based change detection for HR multispectral images. *IEEE Geoscience and Remote Sensing Letters*, 18(5), 856–860.
40. Pomente, A., Picchiani, M., Del Frate, F. (2018, July). Sentinel-2 change detection based on deep features. In *IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium* (pp. 6859–6862). IEEE.
41. Zhao, W., Wang, Z., Gong, M., Liu, J. (2017). Discriminative feature learning for unsupervised change detection in heterogeneous images based on a coupled neural network. *IEEE Transactions on Geoscience and Remote Sensing*, 55(12), 7066–7080.
42. Xu, Y., Xiang, S., Huo, C., Pan, C. (2013, October). Change detection based on auto-encoder model for VHR images. In *MIPPR 2013: Pattern Recognition and Computer Vision (Vol. 8919, p. 891902)*. SPIE.
43. Lv, N., Chen, C., Qiu, T., Sangaiah, A. K. (2018). Deep learning and superpixel feature extraction based on contractive autoencoder for change detection in SAR images. *IEEE Transactions on Industrial Informatics*, 14(12), 5530–5538.
44. Zhang, H., Gong, M., Zhang, P., Su, L., Shi, J. (2016). Feature-level change detection using deep representation and feature change analysis for multispectral imagery. *IEEE Geoscience and Remote Sensing Letters*, 13(11), 1666–1670.
45. Bovolo, F. (2008). A multilevel parcel-based approach to change detection in very high resolution multitemporal images. *IEEE Geoscience and Remote Sensing Letters*, 6(1), 33–37.
46. Masci, J., Meier, U., Cireşan, D., Schmidhuber, J. (2011, June). Stacked convolutional autoencoders for hierarchical feature extraction. In *International Conference on Artificial Neural Networks* (pp. 52–59). Springer, Berlin, Heidelberg.
47. Othman, E., Bazi, Y., Alajlan, N., Alhichri, H., Melgani, F. (2016). Using convolutional features and a sparse autoencoder for land-use scene classification. *International Journal of Remote Sensing*, 37(10), 2149–2167.
48. Geng, J., Fan, J., Wang, H., Ma, X., Li, B., Chen, F. (2015). High-resolution SAR image classification via deep convolutional autoencoders. *IEEE Geoscience and Remote Sensing Letters*, 12(11), 2351–2355.
49. Kemker, R., Salvaggio, C., Kanan, C. (2018). Algorithms for semantic segmentation of multispectral remote sensing imagery using deep learning. *ISPRS Journal of Photogrammetry and Remote Sensing*, 145, 60–77.
50. Kerner, H. R., Wagstaff, K. L., Bue, B. D., Gray, P. C., Bell, J. F., Amor, H. B. (2019). Toward generalized change detection on planetary surfaces with convolutional autoencoders and transfer learning. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 12(10), 3900–3918.
51. Bergamasco, L., Saha, S., Bovolo, F., Bruzzone, L. (2019, October). Unsupervised change-detection based on convolutional-autoencoder feature extraction. In *Image and Signal Processing for Remote Sensing XXV (Vol. 11155, p. 1115510)*. International Society for Optics and Photonics.

52. Bergamasco, L., Saha, S., Bovolo, F., Bruzzone, L. (2022). Unsupervised Change Detection Using Convolutional-Autoencoder Multi-resolution Features. *IEEE Transactions on Geoscience and Remote Sensing*.
53. Kalinicheva, E., Ienco, D., Sublime, J., Trocan, M. (2020). Unsupervised change detection analysis in satellite image time series using deep learning combined with graph-based approaches. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 13, 1450–1466.
54. Gueguen, L., Soille, P., Pesaresi, M. (2011). Change detection based on information measure. *IEEE Transactions on Geoscience and Remote Sensing*, 49(11), 4503–4515.
55. Erten, E., Reigber, A., Ferro-Famil, L., Hellwich, O. (2011). A new coherent similarity measure for temporal multichannel scene characterization. *IEEE Transactions on Geoscience and Remote Sensing*, 50(7), 2839–2851.
56. Volpi, M., Tuia, D. (2016). Dense semantic labeling of subdecimeter resolution images with convolutional neural networks. *IEEE Transactions on Geoscience and Remote Sensing*, 55(2), 881–893.
57. Li, Y., Wang, N., Shi, J., Liu, J., Hou, X. (2016). Revisiting batch normalization for practical domain adaptation. *arXiv preprint arXiv:1603.04779*.
58. Otsu, N. (1979). A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, 9(1), 62–66.
59. Zhang, W., Li, R., Zeng, T., Sun, Q., Kumar, S., Ye, J., Ji, S. (2016). Deep model based transfer and multi-task learning for biological image analysis. *IEEE Transactions on Big Data*, 6(2), 322–333.
60. Huang, Z., Pan, Z., Lei, B. (2017). Transfer learning with deep convolutional neural network for SAR target classification with limited labeled data. *Remote Sensing*, 9(9), 907.
61. Willis, K. S. (2015). Remote sensing change detection for ecological monitoring in United States protected areas. *Biological Conservation*, 182, 233–242.
62. Mengistu, D. A., Salami, A. T. (2007). Application of remote sensing and GIS inland use/land cover mapping and change detection in a part of south western Nigeria. *African Journal of Environmental Science and Technology*, 1(5), 99–109.
63. Lyon, J. G., Yuan, D., Lunetta, R. S., Elvidge, C. D. (1998). A change detection experiment using vegetation indices. *Photogrammetric Engineering and Remote Sensing*, 64(2), 143–150.
64. van den Bergh, F., Udahehuka, G., van Wyk, B. J. (2009, July). Potential fire detection based on Kalman-driven change detection. In 2009 IEEE International Geoscience and Remote Sensing Symposium (Vol. 4, pp. IV-77). IEEE.
65. Rimmel, T. K., Perera, A. H. (2001). Fire mapping in a northern boreal forest: assessing AVHRR/NDVI methods of change detection. *Forest Ecology and Management*, 152(1–3), 119–129.
66. Zanetti, M., Marinelli, D., Bertoluzza, M., Saha, S., Bovolo, F., Bruzzone, L., Costantini, M. (2019, August). A high resolution burned area detector for Sentinel-2 and Landsat-8. In 2019 10th International Workshop on the Analysis of Multitemporal Remote Sensing Images (MultiTemp) (pp. 1–4). IEEE.
67. Wang, Q., Yuan, Z., Du, Q., Li, X. (2018). GETNET: A general end-to-end 2-D CNN framework for hyperspectral image change detection. *IEEE Transactions on Geoscience and Remote Sensing*, 57(1), 3–13.
68. Bovolo, F., Bruzzone, L. (2007). A split-based approach to unsupervised change detection in large-size multitemporal images: Application to tsunami-damage assessment. *IEEE Transactions on Geoscience and Remote Sensing*, 45(6), 1658–1670.
69. Ioffe, S., Szegedy, C. (2015, June). Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *International Conference on Machine Learning* (pp. 448–456). PMLR.
70. Maas, A. L., Hannun, A. Y., Ng, A. Y. (2013, June). Rectifier nonlinearities improve neural network acoustic models. In *Proc. ICML* (Vol. 30, No. 1, p. 3).
71. Canny, J. (1986). A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6(6), 679–698.

72. Maini, R., Aggarwal, H. (2009). Study and comparison of various image edge detection techniques. *International Journal of Image Processing (IJIP)*, 3(1), 1–11.
73. Serra, P., Pons, X., Sauri, D. (2003). Post-classification change detection with data from different sensors: some accuracy considerations. *International Journal of Remote Sensing*, 24(16), 3311–3340.

Homography Augmented Momentum Contrastive Learning for SAR Image Retrieval



Seonho Park, Maciej Rysz, Kathleen M. Dipple, and Panos M. Pardalos 

1 Introduction

Recent studies have applied deep learning for synthetic aperture radar (SAR) image analysis tasks such as object detection [1], despeckling [2–4], optical data fusion [5], and terrain surface classification [6]. One important application is the task of SAR image retrieval [7–9], which aims to retrieve images from a large database that are similar to a query image. This can be further utilized to assist navigation systems when the global positioning system (GPS) is not available [10]. During the SAR image retrieval tasks, a common basic step involves extracting a compressed feature vector from a given SAR image while maintaining the semantic information. The vector is then compared with feature vectors of the SAR images in the database. This technique is sometimes called *the global descriptor approach*, where the global descriptor is a simple vector with a prescribed dimension. Thus, during testing, distances between vectors can be readily and scalably measured. Such global descriptor vectors are often extracted from images using convolutional neural networks (CNNs) [11–13]. However, the overall performance of this technique may suffer from complications such as clutter, illumination, and occlusion, which all

S. Park (✉) · P. M. Pardalos
Department of Industrial & Systems Engineering, University of Florida, Gainesville, FL, USA
e-mail: seonhopark@ufl.edu; pardalos@ise.ufl.edu

M. Rysz
Department of Information Systems & Analytics, Miami University, Oxford, OH, USA
e-mail: ryszmw@miamioh.edu

K. M. Dipple
Integrated Sensor and Navigation Services Team, Air Force Research Laboratory (AFRL/RWWI),
Eglin Air Force Base, FL, USA
e-mail: kathleen.dipple.1@us.af.mil

© The Editor(s) (if applicable) and The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

M. Rysz et al. (eds.), *Synthetic Aperture Radar (SAR) Data Applications*, Springer Optimization and Its Applications 199, https://doi.org/10.1007/978-3-031-21225-3_3

hinder the CNNs from generating accurate global descriptor vectors. To overcome these obstacles, the CNN-based *local feature approaches* have also been suggested [7–9, 14, 15]. These methods furnish the so-called *keypoints* and their corresponding local descriptors. Namely, a keypoint represents the location of interest and the local descriptor is a vector characterizing an associated keypoint. CNN-based local feature approaches generally aim to replace the traditional local feature approaches, such as SIFT [16, 17] or its variants [18]. Even though this enhances the performance of retrieving images, comparing image pairs is not scalable, and, therefore, cannot be efficiently applied over large-scale databases. To this effect, recent efforts often focus on developing a CNN-based approach that combines both global descriptor and local feature techniques by first retrieving images roughly using global descriptor, then reranking them by utilizing the local feature approach [7, 10].

We devote this chapter only to the CNN-based global descriptor method since it generally represents a principal component of the deep learning-based SAR image retrieval system. Emphasis is put on developing a contrastive learning method [19] for generating the global descriptor of a SAR image. Contrastive learning employs two neural networks and compares the global descriptors using a loss function. To prevent the networks from generating trivial descriptors, homography transformation is applied to augment the SAR images. We also demonstrate that the homography transformation can generalize SAR image deformation, thus the homography transformed SAR images are used as input data for training the networks. Contrastive learning enables the network to learn the features in a self-supervised way, which can potentially facilitate applications over large-scale dataset without involving arduous labeling processes that are usually carried out by human labor. To verify the performance of the proposed method, we conduct experiments on multiple SAR datasets containing a multitude of geographic characteristics.

This chapter is organized as follows. Related studies to our approach are discussed in Sect. 2. In Sect. 3, a method for generating global descriptor is proposed, which includes contrastive learning and homography transformation. Experimental results on public SAR dataset are reported in Sect. 4, and finally Sect. 5 provides concluding remarks.

2 Related Works

Contrastive learning [19] is a self-supervised learning method that has been actively investigated recently. It utilizes a comparison between feature representation pairs of instances in a representative space to form a loss function. Contrastive learning is founded on the idea that the feature vectors of an image and its augmented image are close to one another, whereas the feature vectors of two distinguishable images should be distant. Generally, contrastive learning allows machine learning systems not to rely on labels, but rather enables learning with *pseudo labels* that are automatically generated when comparing image feature vectors during training.

Contrastive learning can be also utilized for pretraining in a self-supervised manner in order to improve the performance before applying it for downstream tasks [20]. A primitive approach, called a *memory bank* scheme, stores whole vectors obtained at previous iterations and uses a subset of them at the current iteration [10, 21]. Since the encoder is gradually updated via backpropagation and a stochastic gradient descent algorithm, the output vectors stored in the memory bank are on occasion incompatible with those produced at the current iteration, which, in turn, leads to slow learning. Further, it may cause memory issues when involving a high number of training data instances. To circumvent this, Momentum Contrast (MoCo) [22] uses a Siamese encoder, that is, there are two distinguishable encoders with the same architectures from which one, the primary encoder, is updated with the gradient; whereas the other, the momentum encoder, is updated using a momentum rule. Also, it stores the restricted amount of the *key* vectors outputted by the momentum encoder into a queue matrix. These are then used when comparing with a *query* vector outputted by the primary encoder.

There also exist techniques to improve the performance of MoCo by adding a two-layered multilayer perceptron head, blur augmentation, and cosine learning rate schedule [23]. Recent advances of MoCo, namely MoCo v3 [20], proposed an online method for generating keys in the same batch to reduce memory usage while maintaining fast learning. It uses a form of the infoNCE [24] as a loss function that allows representative differentiations between images of given sufficient batch size (e.g., 4096). Additionally, to further improve the performance of MoCo, they adopted a vision transformation (ViT) [25] as the backbone architecture. There are various approaches to contrastive learning that differ depending on the usages of the queue matrix and the forms of the loss function. They include SimCLR [26], SEER [27], SwAV [28], SimSiam [29], and BYOL [30].

Deep neural networks (DNNs) have also been widely used for representative learning. Traditional components of image retrieval systems, such as SIFT [16, 17], RANSAC [31], have been partially replaced by DNN-based approaches. Noh et al. [7] introduced a method for generating local features using a CNN-based model, which can essentially replace traditional local descriptors such as SIFT. They also proposed a landmark dataset on which the proposed model is finetuned with the location-based image classification labels, which can be regarded as a “weakly” supervised approach. The further enhancements were obtained by applying regional aggregated selective match kernels [8], and generating both global descriptor and local features simultaneously [15]. SuperPoint [9] outputs both the keypoints and local descriptors simultaneously via a unified CNN-based backbone architecture. Its training process involves pretraining on a synthetic labeled dataset and self-supervised learning with the help of homography transformation.

Park et al. [10] proposed *deep cosine similarity neural networks* to generate a 12 normalized feature vector of a SAR image with the primary purpose of scalably comparing SAR image pairs. The proposed idea of normalizing the vector during training is also used in the training process of the encoders that is used in this work as well. Once similar images from a database are retrieved using the global descriptors, the local features are used for reranking the retrieved images. This process is

usually time-consuming since matching pairs through RANSAC [31] or Direct Linear Transformation (DLT) [32] is not scalable. Thus, replacing these matching techniques with scalable DNN-based methods is an active research area where many approaches such as SuperGlue [33] and LoFTR [34] have been proposed. They also utilize a state-of-the-art architecture, Transformer [35], as their backbone networks, which can significantly enhance the performances of the DNN-based methods for image retrieval tasks.

3 Methodology

In this section, we focus on training a deep neural network-based encoder to produce a feature vector that is used in the SAR image retrieval task. We generate a (global) feature vector, \mathbf{d} , to compress and represent a SAR image for the purpose of image retrieval. Namely, vector \mathbf{d} contains semantic information of the SAR image.

3.1 Contrastive Learning

Contrastive learning has been presented in recent literature as a means for generating feature embedding in a self-supervised manner. Broadly, it uses a Siamese encoder (twin encoders) that consists of two CNN-based encoders with the same architectures and the same sets of parameters to that are learned separately. When two inputs of both encoders are similar, it aims to produce output vectors that are likewise similar; whereas when two inputs are different, they should produce output vectors that are distinct. Contrastive learning allows the encoder to be trained in a self-supervised way, i.e., the input data does not require labels during training. Therefore, when available, one can aggregate more data for training the model without expending arduous labor that is commonly required for labeling instances.

In this work, we use MoCo-like learning [22] where the first encoder is updated via a stochastic gradient descent algorithm, and the second encoder is updated by a momentum rule. Figure 1 depicts the contrastive learning framework that we use for SAR image retrieval. As shown in the figure, two CNN-based encoders, the “primary encoder” f_{θ_q} and the “momentum encoder” f_{θ_k} , are utilized, where θ_q and θ_k are the parameter sets corresponding to the former and latter encoders, respectively.

The primary encoder f_{θ_q} uses a SAR image as an input \mathbf{x} to produce an output \mathbf{d}_q . Similarly, the momentum encoder f_{θ_k} , which has the same architecture as the primary encoder, uses a transformed SAR image $\mathbf{x}' \sim \mathcal{H}(\mathbf{x})$ that is drawn from homography transformation-based distribution $\mathcal{H}(\mathbf{x})$ to output a feature vector \mathbf{d}_k . The homograph transformation used is presented in Sect. 3.2. The primary encoder is updated using a stochastic gradient descent method where the gradients associated with the parameters are computed through backpropagation. The loss function used

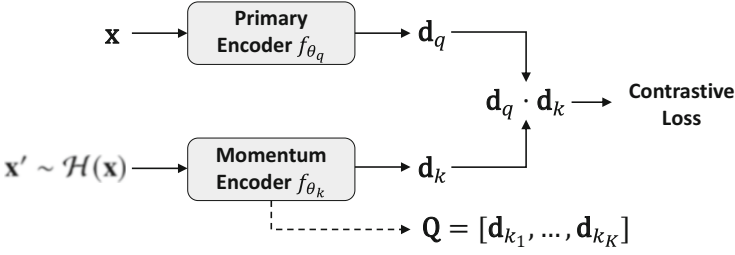


Fig. 1 Diagram of the contrastive learning scheme

for training the primary encoder is described in Sect. 3.3. The parameters θ_k in the momentum encoder are updated using the momentum rule $\theta_k \leftarrow m\theta_k + (1-m)\theta_q$, where m is a momentum parameter used to slowly and stably update parameters θ_k . Also, in the momentum encoder, the queue matrix \mathbf{Q} consists of columns of feature vectors calculated during previous iterations:

$$\mathbf{Q} = [\mathbf{d}_{k_1}, \dots, \mathbf{d}_{k_K}], \quad (1)$$

where K is a prescribed parameter representing the number of feature vectors. Matrix \mathbf{Q} is continuously enqueued using the most recently generated vectors at the current iteration while the oldest vectors are dequeued to maintain a consistent size of the matrix.

3.2 Homography Transformation

The transformed input \mathbf{x}' used by the momentum encoder should maintain the same semantic information as the original image and simultaneously generalize the images in order to prevent trivial training cases. To achieve this, we exploit homography transformation. It is well known that homography transformation can explain the translational and rotational transformations between an image pair in the same planar space. Thus, it is suitable for tasks such as one of the present interests where SAR images are taken from aircraft or satellites that are positioned at various altitudes and angles relative to a given surface area in planar space. Given a pixel location $[u, v]$ in an image, homography transformation is a linear transformation represented by a non-singular 3×3 matrix as

$$\begin{bmatrix} u' \\ v' \\ 1 \end{bmatrix} = \begin{bmatrix} H_{11} & H_{12} & H_{13} \\ H_{21} & H_{22} & H_{23} \\ H_{31} & H_{32} & H_{33} \end{bmatrix} \begin{bmatrix} u \\ v \\ 1 \end{bmatrix}, \quad \mathbf{u}' = \mathbf{H}\mathbf{u}, \quad (2)$$

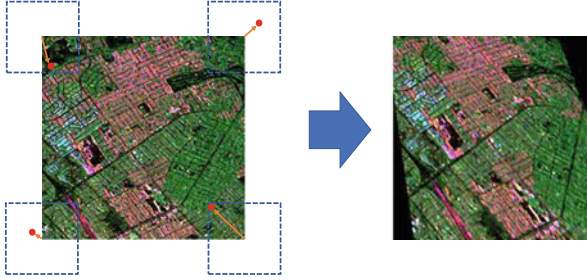


Fig. 2 An example of 4-points based homography transformation for SAR images

where H_{ij} , $\forall i, j \in \{1, 2, 3\}$ represents the element of the homography transformation matrix \mathbf{H} , and $[u', v']$ is the transformed location. Note that the homography transformation matrix is invariant with scaling, thus containing 8 degrees of freedom. To augment the image data via homography transformation, a 4-point based homography transformation is frequently utilized [36, 37]. We employ a similar approach furnished in [36] to generate the homographic transformed SAR image \mathbf{x}' , which will serve as an input to the momentum encoder.

For example, observe in Fig. 2 that the four corners of the pixel points (u_k, v_k) for $k \in \{1, 2, 3, 4\}$ (represented as red dots) are drawn at random within the small blue squares. Then, using the four corner points we construct a 4-point parameterization \mathbf{H}_{4p} that consists of the x, y displacements, $(\Delta u_k = u'_k - u_k, \Delta v_k = v'_k - v_k)$ for $k \in \{1, 2, 3, 4\}$. Note that the 4-point parameterization \mathbf{H}_{4p} is a 4×2 matrix containing 8 DOFS as the homography transformation matrix \mathbf{H} , and \mathbf{H}_{4p} can easily be converted to \mathbf{H} by the DLT algorithm [32] or the function `getPerspectiveTransform` in OpenCV. As shown in Fig. 2, after applying the homography transformation to an image, the transformed image (right image) is cropped to the same size as the original image. In what follows, with a slight abuse in notation, we denote the image distribution of the homography transformation and cropping applied to the image \mathbf{x} as $\mathcal{H}(\mathbf{x})$.

3.3 Training

The transformed image $\mathbf{x}' \sim \mathcal{H}(\mathbf{x})$ based on \mathbf{H}_{4p} is first sampled and serves as input of the momentum encoder as

$$\mathbf{d}_q = f_{\theta_q}(\mathbf{x}), \quad (3)$$

$$\mathbf{d}_k = f_{\theta_k}(\mathbf{x}'), \quad (4)$$

while the original SAR image is passed as input to the primary encoder. After l_2 -normalizing feature vectors \mathbf{d}_q and \mathbf{d}_k , a logit vector is constructed as follows:

$$\mathbf{l}_q = \left[\frac{\mathbf{d}_q^T \mathbf{d}_k}{\tau}, \frac{\mathbf{d}_q^T \mathbf{Q}}{\tau} \right]^T, \quad (5)$$

where a parameter $\tau > 0$ is called a *temperature* parameter [38, 39]. For each SAR image \mathbf{x} in the training dataset, the loss function takes the form of the infoNCE [24], i.e.,

$$L(\mathbf{x}) = -\log \frac{e^{l_{q_1}}}{\sum_{i=1}^{K+1} e^{l_{q_i}}}, \quad (6)$$

where l_{q_i} represents the i th element of the logit vector \mathbf{l}_q . Finally, for the training dataset $\{\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(N)}\}$, the overall loss function for training the primary encoder is given by

$$\min_{\theta_q} L = \frac{1}{N} \sum_{i=1}^N L(\mathbf{x}^{(i)}). \quad (7)$$

The procedure for training the primary encoder and its corresponding momentum encoder is presented in Algorithm 1. As shown in line 9 in Algorithm 1, backpropagation is conducted to calculate the gradients of the loss function with respect to the parameters θ_q . Stochastic gradient descent or its variants [40, 41] can be applied to $\text{Opt}(\theta_q, \eta, \Delta\theta_q)$ to update the parameters θ_q , where $\Delta\theta_q$ represents the gradients or their variants used for updating the parameters.

Algorithm 1 Contrastive learning of feature vector for SAR image retrieval

- 1: Input: learning rate η , minibatch size N , queue size K , momentum parameter m , temperature parameter τ
 - 2: Initialize queue matrix \mathbf{Q} , parameters of the encoders as $\theta_q = \theta_k$
 - 3: **repeat**
 - 4: Draw $\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(N)}$ samples at random from the training dataset
 - 5: Generate homography transformed images $\mathbf{x}'^{(1)}, \dots, \mathbf{x}'^{(N)}$
 - 6: Calculate $\mathbf{d}_q^{(i)} = f_{\theta_q}(\mathbf{x}^{(i)})$ for all i
 - 7: Calculate $\mathbf{d}_k^{(i)} = f_{\theta_k}(\mathbf{x}'^{(i)})$ for all i
 - 8: Calculate loss function L (Eq. 7)
 - 9: Update θ_q with $\text{Opt}(\theta_q, \eta, \Delta\theta_q)$
 - 10: Update θ_k as $\theta_k \leftarrow m\theta_k + (1 - m)\theta_q$
 - 11: dequeue old N columns from \mathbf{Q}
 - 12: enqueue new N columns to \mathbf{Q}
 - 13: **until** θ_q has converged
 - 14: Output: learned parameters θ_q
-

4 Numerical Experiments

4.1 SAR Image Data

As an experimental testbed for SAR image retrieval, we utilize the Uninhabited Aerial Vehicle Synthetic Aperture Radar (UAVSAR) images [42] from NASA. From the UAVSAR database, we use L-band and the ground projected complex cross (GRD) products of the polarimetric SAR (PolSAR). A central reason for using GRD products is that it provides information on pixel-wise geographic coordinates mapping that, in turn, leads to the precise evaluation of the performance of image retrieval. Specifically, HHHH, HVHV, VVVV products are processed to grayscale images through ASF MapReady 3.2 software, and grayscale images of HHHH, HVHV, VVVV products correspond to red, green, blue channels of the resulting PolSAR images, respectively.

Table 1 lists details of the SAR image maps used in our experiments that contain various topological characteristics including building, mountain, river, and vegetation. As shown, the SAR image maps are too big to serve as input for the encoders. Thus, we extract patches of size 600×600 with a stride of 100 pixels from the images. Since GRD formatting is arranged to locate the north to the upper side of an image map with respect to the geographic coordinates, it may contain a lot of “blank” areas (shown as the black region as in Fig. 3). Further, as the blank area does not contain any meaningful information, we eliminate this by using the local descriptor method, SAR-SIFT [18]. For each extracted patch, we generated the keypoints using SAR-SIFT and patches with at least 200 keypoints were added to the datasets in our experiments. It was deemed that a patch with less than 200 keypoints contains a significant amount of black area. The resulting patches were then resized to 224×224 pixels. Homography transformation was applied “on the fly” during training. On the resized SAR patches, i.e., 224×224 pixels, four points were drawn at random, from the four corner squares of 32×32 pixels centered at the corner points (shown as the dashed blue lines in Fig. 2) to estimate \mathbf{H}_{4p} .

For each experiment, three types of datasets consisting of the SAR patches were prepared: the query dataset, training dataset, and key dataset as shown in Fig. 4.

Table 1 PolSAR image maps from UAVSAR database

Name	Acquisition date	Size (W×H)	Region	Characteristic
Haywrld1	10/09/2018	19,779×26,236	Hayward fault, CA	Building, Mountain
Haywrld2	05/30/2019	19,758×26,206		
ykdelB1	08/28/2018	23,007×3906	Yukon-Kuskokwim delta, AK	River delta, Tundra
ykdelB2	09/17/2019	23,107×3904		
atchaf	04/01/2021	6436×7973	Atchafalaya river delta, LA	River delta, Vegetation
harvrld	08/17/2009	5981×9319	Harvard forest, MA	Forest
SanAnd	02/21/2019	43,050×6604	LA basin, CA	Building, Mountain
SRIVER	06/17/2017	16,631×21,099	Subarctic tundra area, AK	Tundra

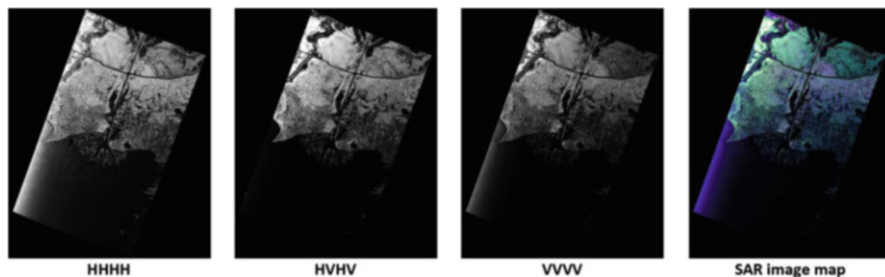
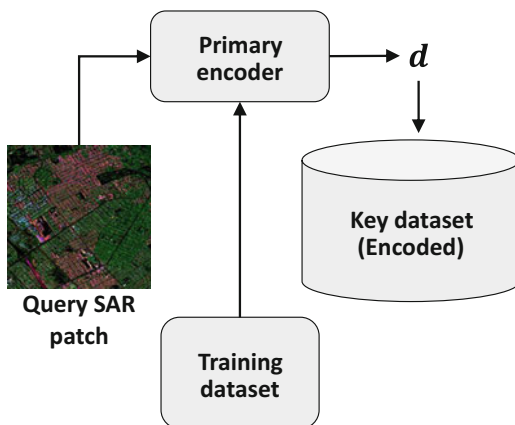


Fig. 3 SAR image map of the "atchaf" SAR data. RGB channels correspond to HHHH, HVHV, VVVV GRD products

Fig. 4 Diagram of SAR patch datasets usage



Each of the patches in the datasets has at least 1000 SAR-SIFT based keypoints. The primary encoder and momentum encoder are trained on the training dataset, whereas the SAR patches in the query dataset are used for testing. During testing, we extract the patches with similar feature vectors to the query patch from the key dataset, which is also representative of a database. We restrict the number of patches in the query dataset to 100, which are selected at random.

To verify the performance of the proposed approach, we set up four experiments that simulate multiple operational circumstances, as shown in Table 2. The query sets are HaywrD2 or ykdelB2. These were more recently acquired than the key datasets HaywrD1 and ykdelB1 (see Table 1). For experiments 1 and 3 (i.e., Exp.1 and Exp.3), we have used HaywrD1 and ykdelB1 as the training datasets, respectively. For these experiments, the training datasets were the same as the key datasets, hence the encoders already observed similar SAR image patches to the query images during training. It is, therefore, anticipated that the encoders would be able to generate more accurate feature vectors for these two experiments compared to the experiments Exp.2 and Exp.4. For experiments Exp.2 and Exp.4, the encoders had not observed the same patches during training; instead, it was trained on the training dataset consisting of patches from atchaf, harvrD, SanAnd, and SRIVER

Table 2 Details on experiment dataset. The number of SAR patches are shown in parenthesis

Exp. Num.	Exp. Name	Query set	Key set	Training set
Exp.1	Haywrd-Easy	Haywrd2 (100)	Haywrd1, ykdelB1 (20,388)	Haywrd1 (12979)
Exp.2	Haywrd-Hard	Haywrd2 (100)	Haywrd1, ykdelB1 (20,388)	atchaf, harvrd, SanAnd, SRIVER (21,555)
Exp.3	ykdelB-Easy	ykdelB2 (100)	Haywrd1, ykdelB1 (20,388)	ykdelB1 (7409)
Exp.4	ykdelB-Hard	ykdelB2 (100)	Haywrd1, ykdelB1 (20,388)	atchaf, harvrd, SanAnd, SRIVER (21,555)

SAR maps. Also, for all experiments, the Haywrd1 and ykdelB1 were used as the key dataset.

4.2 Experiment Settings

All experiments used ResNet50 or ResNet101 [43] as a backbone CNN architecture. The CNNs were pretrained on the ImageNet dataset [44]. From the output map of the *conv4* layer of ResNet, we adopted generalized mean (GeM) pooling [45] that weighs the importance of the outputs to generate the feature vector. Denote the output map of the *conv4* layer as $\mathbf{o}_{h,w}$ where h and w are indices of height and width locations of the image, respectively. We further integrated GeM pooling followed by a fully connected layer, parameterized by weights F and bias b_F , similar to the approach in [15, 46]. The feature vector \mathbf{d} that summarizes the discriminative outputs of the image is produced as an output of the GeM pooling layer as

$$\mathbf{d} = F \times \left(\frac{1}{H_{conv4} W_{conv4}} \sum_{h,w} \mathbf{o}_{h,w}^p \right)^{1/p} + b_F, \quad (8)$$

where H_{conv4} , W_{conv4} represent the height and width of the output map, and p is the parameter of the GeM pooling. In our experiments, the parameter p is set to 3. The feature vector \mathbf{d} is then l_2 normalized.

For training, we employed the Adam optimizer [40] with a learning rate of $5e-3$ that gets decreased by 0.1 after 80 epochs. The maximum number of epochs was set to 100 and the batch size N for training is 32. The temperature parameter τ was set to 0.5 without performing any hyperparameter tunings. For contrastive learning parameters, the queue size $K = 1024$ and the momentum parameter $m = 0.999$. To prevent feature vector \mathbf{d} from having an excessively high l_2 norm, prior to normalizing, we added the l_2 regularization term derived in [10] into the loss function. Specifically, this term is simply denoted as $(\|\mathbf{d}\|_2 - 1)^2$. The coefficient of the l_2 regularization was set to 0.1.

Note that when testing the image retrieval performance, since the focus is put on measuring the performance of the feature vector, we do not consider any approximate techniques for measuring distance such as query expansion [45, 47] or asymmetric quantizer distance [48].

4.3 Experiments Results

Precision and recall were used to evaluate the performance of the proposed method. The recall is the total number of the retrieved SAR patches from the database (key dataset), and the precision is defined as the ratio between the number of accurately retrieved SAR patches and the total number of retrieved SAR patches. Whenever there exist duplicate regions calculated in terms of the geographic coordinates between the query image and the retrieved image, we consider the retrieved image as the accurately retrieved image. As a primary performance measure, we have used mean average precision (mAP). The average precision is the mean value of the precision values obtained at various recall values for each query image. In the present setting, the mAP is the mean of the average precision values over 100 query images. Also, the mean precision at n (mP@ n) is used as a performance measure where n represents the recall value. We report three mean precisions; mP@1, mP@10, mP@50. For instance, it is implied that when we retrieve ten SAR patches from the database, mP@10 represents the fraction of the number of accurate images we can expect to retrieve correctly.

The experiment results are shown in Table 3. The methods vary depending on the dimension of the feature output \mathbf{d} , which is 512 or 1024, and the two considered architectures ResNet50 or ResNet101.

The ‘‘Haywrd’’ cases (Exp.1 and Exp.2) are from the Hayward fault in California, USA, and predominantly consist of human-made structures. The ‘‘ykdelB’’ cases (Exp.3 and Exp.4), on the other hand, mostly consist of natural formations. By comparing the mAP values, we observed that the proposed method more effectively retrieved similar images for the Haywrd experiment cases compared to the ykdelB cases. For both datasets, it can be seen that mAP values of the ‘‘harder’’ cases (Exp 2 and Exp. 4) are lower than those of the ‘‘easy’’ cases (Exp 1 and Exp. 3).

Although the performances are relatively lower in the hard cases, it can be observed that the proposed method works well regardless of the dimension of the feature vector and the backbone architectures, as even mP@1 results in high values (over 0.9) in all the experiments. This result strongly indicates that our method can perform remarkably well without the use of any sophisticated yet non-scalable methods such as local descriptors [7] and matching through RANSAC [31].

By way of visual results, Fig. 5 shows several retrieved SAR patches using ResNet101-1024 that outputs 1024 dimensional feature vector and its backbone is ResNet101. In most cases, the proposed method retrieves SAR images correctly

Table 3 Performance results on the UAVSAR dataset

Method	mAP	mP@1	mP@10	mP@50
<i>Exp.1 Haywrd-Easy</i>				
ResNet50-512	0.5254	1.0000	0.9870	0.8044
ResNet50-1024	0.4908	1.0000	0.9810	0.7702
ResNet101-512	0.5366	0.9700	0.9650	0.8018
ResNet101-1024	0.5314	0.9800	0.9630	0.7952
<i>Exp.2 Haywrd-Hard</i>				
ResNet50-512	0.4092	1.0000	0.9810	0.6874
ResNet50-1024	0.4287	1.0000	0.9730	0.7188
ResNet101-512	0.4223	1.0000	0.9690	0.7050
ResNet101-1024	0.4159	1.0000	0.9650	0.6976
<i>Exp.3 ykdelB-Easy</i>				
ResNet50-512	0.4708	0.9700	0.9230	0.6128
ResNet50-1024	0.4439	0.9100	0.8640	0.5760
ResNet101-512	0.4296	0.9400	0.8690	0.5644
ResNet101-1024	0.4290	0.9500	0.8630	0.5574
<i>Exp.4 ykdelB-Hard</i>				
ResNet50-512	0.4181	0.9400	0.8440	0.5750
ResNet50-1024	0.3916	0.9600	0.8340	0.5442
ResNet101-512	0.3681	0.9700	0.8680	0.5138
ResNet101-1024	0.3682	1.0000	0.8970	0.5280

from the database. In fact, only two retrieved images in Exp.4 ykdelB-Hard case were incorrect; however, these images are quite similar to a query SAR image visually.

5 Conclusions

In this chapter we have proposed a contrastive learning method for the SAR image retrieval task. A homography transformation is used for augmenting the SAR image patches and the output vector of a transformed image is compared against that of an original image. This enables the two introduced encoders to be trained in a self-supervised manner, thus no labeling process is necessary when constructing or obtaining the dataset. Our method is, therefore, especially applicable for tasks where the labeling process is time-consuming and costly. Experiments were conducted and corresponding results demonstrate that even in cases where the encoders did not encounter SAR images similar to a given query image during training, they nevertheless successfully retrieved similar SAR images from the database. This work can be utilized for applications such as image searching-based positioning or navigation, which is one of our future extensions.

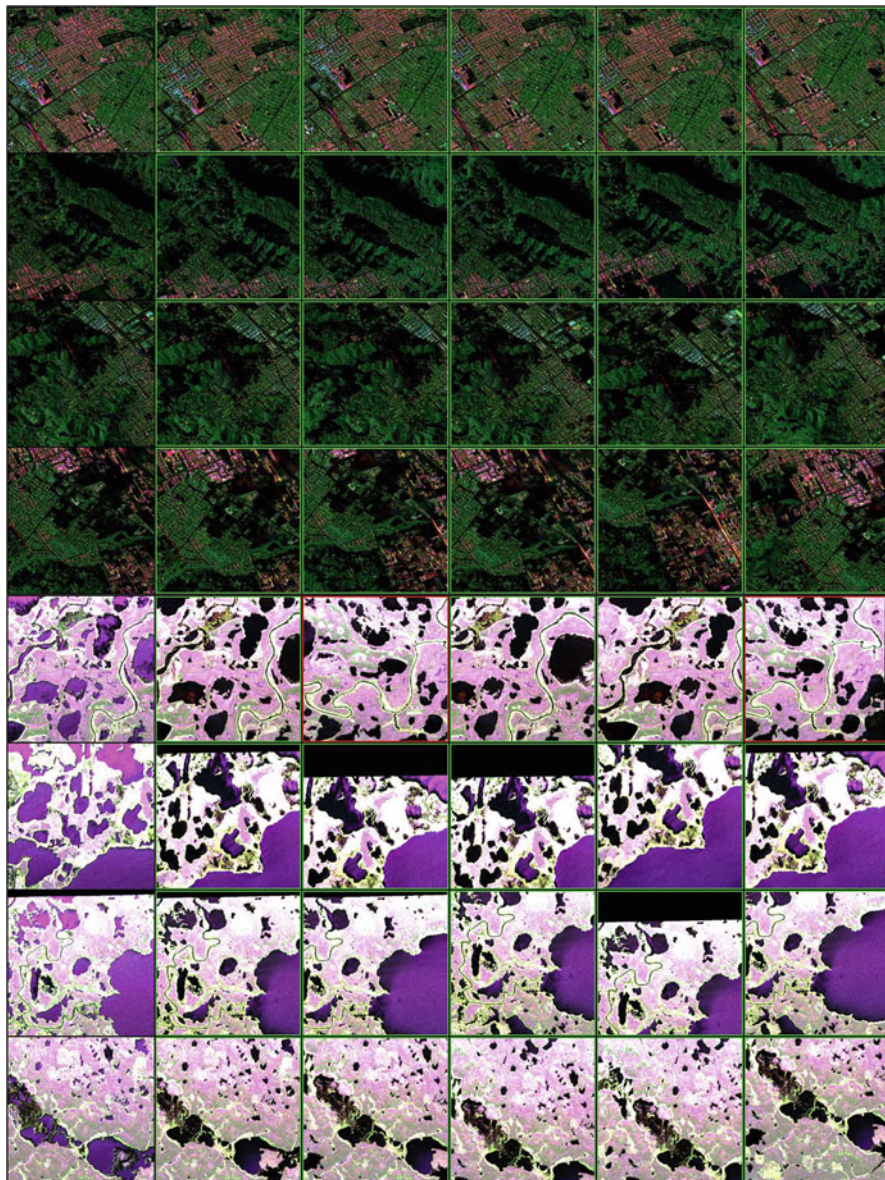


Fig. 5 Examples of retrieved SAR images. The first column is the query SAR images and the others are the retrieved ones (the second column contains the best-retrieved image and the rightmost column contains the fifth-best). The green box represents an accurately retrieved SAR image, whereas the red box presents an incorrectly retrieved one. **Top:** Exp.2 Haywrd-Hard **Bottom:** Exp.4 ykd1B-Hard

References

1. Sizhe Chen and Haipeng Wang. Sar target recognition based on deep learning. In *2014 International Conference on Data Science and Advanced Analytics (DSAA)*, pages 541–547. IEEE, 2014.
2. Xiao Tang, Lei Zhang, and Xiaoli Ding. Sar image despeckling with a multilayer perceptron neural network. *International Journal of Digital Earth*, 12(3):354–374, 2019.
3. Giovanni Chierchia, Davide Cozzolino, Giovanni Poggi, and Luisa Verdoliva. Sar image despeckling through convolutional neural networks. In *2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pages 5438–5441. IEEE, 2017.
4. Davide Cozzolino, Luisa Verdoliva, Giuseppe Scarpa, and Giovanni Poggi. Nonlocal CNN SAR image despeckling. *Remote Sensing*, 12(6):1006, 2020.
5. Michael Schmitt, Lloyd Haydn Hughes, and Xiao Xiang Zhu. The sen1-2 dataset for deep learning in SAR-optical data fusion. *arXiv preprint arXiv:1807.01569*, 2018.
6. Hemani Parikh, Samir Patel, and Vibha Patel. Classification of SAR and PolSAR images using deep learning: a review. *International Journal of Image and Data Fusion*, 11(1):1–32, 2020.
7. Hyeonwoo Noh, Andre Araujo, Jack Sim, Tobias Weyand, and Bohyung Han. Large-scale image retrieval with attentive deep local features. In *Proceedings of the IEEE international conference on computer vision*, pages 3456–3465, 2017.
8. Marvin Teichmann, Andre Araujo, Menglong Zhu, and Jack Sim. Detect-to-retrieve: Efficient regional aggregation for image search. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5109–5118, 2019.
9. Daniel DeTone, Tomasz Malisiewicz, and Andrew Rabinovich. SuperPoint: Self-supervised interest point detection and description. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 224–236, 2018.
10. Seonho Park, Maciej Rysz, Kaitlin L Fair, and Panos M Pardalos. Synthetic-aperture radar image based positioning in GPS-denied environments using deep cosine similarity neural networks. *Inverse Problems & Imaging*, 2021.
11. Albert Gordo, Jon Almazán, Jerome Revaud, and Diane Larlus. Deep image retrieval: Learning global representations for image search. In *European conference on computer vision*, pages 241–257. Springer, 2016.
12. Relja Arandjelovic, Petr Gronat, Akihiko Torii, Tomas Pajdla, and Josef Sivic. NetVLAD: CNN architecture for weakly supervised place recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5297–5307, 2016.
13. Filip Radenović, Giorgos Tolias, and Ondřej Chum. CNN image retrieval learns from bow: Unsupervised fine-tuning with hard examples. In *European conference on computer vision*, pages 3–20. Springer, 2016.
14. Kwang Moo Yi, Eduard Trulls, Vincent Lepetit, and Pascal Fua. Lift: Learned invariant feature transform. In *European conference on computer vision*, pages 467–483. Springer, 2016.
15. Bingyi Cao, André Araujo, and Jack Sim. Unifying deep local and global features for image search. In *European Conference on Computer Vision*, pages 726–743. Springer, 2020.
16. David G Lowe. Object recognition from local scale-invariant features. In *Proceedings of the seventh IEEE international conference on computer vision*, volume 2, pages 1150–1157. IEEE, 1999.
17. David G Lowe. Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60(2):91–110, 2004.
18. Flora Dellinger, Julie Delon, Yann Gousseau, Julien Michel, and Florence Tupin. Sar-sift: a sift-like algorithm for SAR images. *IEEE Transactions on Geoscience and Remote Sensing*, 53(1):453–466, 2014.
19. Raia Hadsell, Sumit Chopra, and Yann LeCun. Dimensionality reduction by learning an invariant mapping. In *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, volume 2, pages 1735–1742. IEEE, 2006.

20. Xinlei Chen, Saining Xie, and Kaiming He. An empirical study of training self-supervised visual transformers. *arXiv preprint arXiv:2104.02057*, 2021.
21. Zhirong Wu, Yuanjun Xiong, Stella X Yu, and Dahua Lin. Unsupervised feature learning via non-parametric instance discrimination. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3733–3742, 2018.
22. Kaiming He, Haoqi Fan, Yuxin Wu, Saining Xie, and Ross Girshick. Momentum contrast for unsupervised visual representation learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9729–9738, 2020.
23. Xinlei Chen, Haoqi Fan, Ross Girshick, and Kaiming He. Improved baselines with momentum contrastive learning. *arXiv preprint arXiv:2003.04297*, 2020.
24. Aaron van den Oord, Yazhe Li, and Oriol Vinyals. Representation learning with contrastive predictive coding. *arXiv preprint arXiv:1807.03748*, 2018.
25. Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020.
26. Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. In *International conference on machine learning*, pages 1597–1607. PMLR, 2020.
27. Priya Goyal, Mathilde Caron, Benjamin Lefaudeaux, Min Xu, Pengchao Wang, Vivek Pai, Mannat Singh, Vitaliy Liptchinsky, Ishan Misra, Armand Joulin, et al. Self-supervised pretraining of visual features in the wild. *arXiv preprint arXiv:2103.01988*, 2021.
28. Mathilde Caron, Ishan Misra, Julien Mairal, Priya Goyal, Piotr Bojanowski, and Armand Joulin. Unsupervised learning of visual features by contrasting cluster assignments. *arXiv preprint arXiv:2006.09882*, 2020.
29. Xinlei Chen and Kaiming He. Exploring simple Siamese representation learning. *arXiv preprint arXiv:2011.10566*, 2020.
30. Jean-Bastien Grill, Florian Strub, Florent Altché, Corentin Tallec, Pierre H Richemond, Elena Buchatskaya, Carl Doersch, Bernardo Avila Pires, Zhaohan Daniel Guo, Mohammad Gheshlaghi Azar, et al. Bootstrap your own latent: A new approach to self-supervised learning. *arXiv preprint arXiv:2006.07733*, 2020.
31. Martin A Fischler and Robert C Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395, 1981.
32. Alex M Andrew. Multiple view geometry in computer vision. *Kybernetes*, 2001.
33. Paul-Edouard Sarlin, Daniel DeTone, Tomasz Malisiewicz, and Andrew Rabinovich. Super-glue: Learning feature matching with graph neural networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4938–4947, 2020.
34. Jiaming Sun, Zehong Shen, Yuang Wang, Hujun Bao, and Xiaowei Zhou. LoFTR: Detector-free local feature matching with transformers. *arXiv preprint arXiv:2104.00680*, 2021.
35. Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. *arXiv preprint arXiv:1706.03762*, 2017.
36. Daniel DeTone, Tomasz Malisiewicz, and Andrew Rabinovich. Deep image homography estimation. *arXiv preprint arXiv:1606.03798*, 2016.
37. Ty Nguyen, Steven W Chen, Shreyas S Shivakumar, Camillo Jose Taylor, and Vijay Kumar. Unsupervised deep homography: A fast and robust homography estimation model. *IEEE Robotics and Automation Letters*, 3(3):2346–2353, 2018.
38. Geoffrey Hinton, Oriol Vinyals, and Jeff Dean. Distilling the knowledge in a neural network. *arXiv preprint arXiv:1503.02531*, 2015.
39. Gabriel Pereyra, George Tucker, Jan Chorowski, Łukasz Kaiser, and Geoffrey Hinton. Regularizing neural networks by penalizing confident output distributions. *arXiv preprint arXiv:1701.06548*, 2017.

40. Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.
41. Seonho Park, Seung Hyun Jung, and Panos M Pardalos. Combining stochastic adaptive cubic regularization with negative curvature for nonconvex optimization. *Journal of Optimization Theory and Applications*, 184(3):953–971, 2020.
42. Dataset: UAVSAR POLSAR, NASA 2021. Retrieved from ASF DAAC, 22 April 2021.
43. Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.
44. Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. ImageNet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. IEEE, 2009.
45. Filip Radenović, Giorgos Tolias, and Ondřej Chum. Fine-tuning CNN image retrieval with no human annotation. *IEEE transactions on pattern analysis and machine intelligence*, 41(7): 1655–1668, 2018.
46. Albert Gordo, Jon Almazan, Jerome Revaud, and Diane Larlus. End-to-end learning of deep visual representations for image retrieval. *International Journal of Computer Vision*, 124(2): 237–254, 2017.
47. Ondrej Chum, James Philbin, Josef Sivic, Michael Isard, and Andrew Zisserman. Total recall: Automatic query expansion with a generative feature model for object retrieval. In *2007 IEEE 11th International Conference on Computer Vision*, pages 1–8. IEEE, 2007.
48. Yue Cao, Mingsheng Long, Jianmin Wang, Han Zhu, and Qingfu Wen. Deep quantization network for efficient image retrieval. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 30, 2016.

Synthetic Aperture Radar Image Based Navigation Using Siamese Neural Networks



Alexander Semenov, Maciej Rysz, and Garrett Demeyer

1 Introduction

A well-known advantage of synthetic aperture radar (SAR) and the images they produce is their ability to precisely capture topological features without being affected by obstructive weather conditions. Consequently, SAR images have been used in a variety of applications including deep learning, target detection [4], terrain classification [11], despeckling [5, 6, 14], optical data fusion [13], among others. An important application that has been a topic of focus in the recent literature, and the broader focus of this chapter, involves using SAR images to develop navigation techniques suitable for global positioning system (GPS) denied environments. In other words, the general aim is to utilize SAR images to determine the location of a system that is navigating over a specified geographical area of interest. It is assumed that the system itself is capable of generating SAR images (e.g., via an onboard SAR) in real time during operations, and cannot rely on any other form of sensory data to navigate over the area. The navigation process under such settings usually requires retrieval of images, or image retrieval [7, 10, 15], where images stored in a large database, e.g., a database onboard an unmanned aerial vehicle's (UAVs) system, are compared to a query image and one or more of the most similar

A. Semenov (✉)

Department of Industrial & Systems Engineering, University of Florida, Gainesville, FL, USA
e-mail: asemenov@ufl.edu

M. Rysz

Department of Information Systems & Analytics, Miami University, Oxford, OH, USA
e-mail: ryszmw@miamioh.edu

G. Demeyer

Air Force Research Laboratory, Munitions Directorate, Eglin AFB, FL, USA
e-mail: garrett.demeyer.1@us.af.mil

© The Editor(s) (if applicable) and The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

M. Rysz et al. (eds.), *Synthetic Aperture Radar (SAR) Data Applications*, Springer Optimization and Its Applications 199, https://doi.org/10.1007/978-3-031-21225-3_4

images are retrieved. Since the query image represents the current location of the system, a primary goal is to find a best image in the database whose geographical location is closest to or exactly that of the query image. Once an image is selected, it is then possible to produce location coordinates.

To quantify “similarity” between SAR images in a database and a query image, global descriptor feature vectors of all images are first generated and then compared with the vector of the query image. In this regard, many studies have employed convolutional neural networks (CNNs) to generate descriptor vectors of SAR images [1, 8, 12]. There are, however, several issues that often degrade the performance of the vectors when generated via a CNN including clutter, occlusion, and illumination. One possible remedy is to generate keypoints to mark locations of interest on an image and local descriptor vectors corresponding to the keypoints [3, 7, 10, 15, 18]. In other words, the keypoint represents the location of interest and the local descriptor is a vector characterizing an associated keypoint. Nevertheless, one of the major drawbacks of the aforementioned and existing methods is their computational performance over large-scale datasets, which becomes necessary for tasks such as autonomous UAV flight operations.

This work leverages on the concept of *Siamese neural network*, which is an artificial neural network (ANN) proposed and utilized for tasks like signature verification and face recognition [2]. In contrast to classification tasks, where inputs are mapped to a fixed set of outputs, Siamese networks output similarity between inputs. Namely, its “sister” networks use the same weights for two different input vectors in order to compute the output. The output is sometimes computed in advance to serve as a baseline for comparison and performance evaluation. When used in facial recognition applications, rather than mapping each face to an individual class (as normally done in classification), Siamese networks precompute outputs for a known set of people and then retrieve the closest output. Another application of Siamese ANNs is object tracking [19], where similarity is calculated between the exemplar image (object that is tracked) and parts of a larger image within which the object is being searched. Further, vision-based localization in autonomous cars, where only visual information is available, was studied in [16]. The authors utilize deep neural networks and propose a novel loss function that takes into account the geometric distance between sequential images taken by car cameras. Similar problems of learning mapping functions that are assigning geometrically close images to close regions in Euclidean space was studied in [17]. Both articles utilized pretrained ANNs as backbones for image recognition. Learning in Siamese ANNs can be performed using contrastive loss or triplet loss.

The sections of this chapter are organized as follows. In Sect. 2 we describe the model details, including the contrastive and triplet loss functions. Section 3 furnishes a description of the data used, the model architecture details, and the computation results with corresponding discussion. The conclusions are summarized in Sect. 4

2 Model Formulation

This section introduces a deep neural network architecture that produces a global feature vector of a given SAR image that is used in the retrieval process. Let e represent a global feature vector of a SAR image that captures its semantic information. Feature vector e is constructed as an output of a function $e = f(x, \theta)$, where x is an input SAR image, and θ is a vector of parameters, where $e \in \mathbb{R}^n$, $x \in \mathbb{R}^m$, and $n \ll m$. In this case, we represent $f(x, \theta)$ as a neural network that generates the global feature vector. The resulting vectors represent “latent” features of the images, often referred to as *embeddings* in the literature. Then, the principal aim of our approach is to construct a function $f(x, \theta)$ such that images x_i and x_j from a geographically close region result in vectors e_i and e_j that are similar to one another, whereas images that are from distant regions would result in vectors that are dissimilar.

To this end, we utilize the aforementioned Siamese neural network method [2] to find the parameters θ . As a backbone, the network architecture is built based on SqueezeNet [9], which is a compact deep neural network with 10 layers and with approximately 1.2 million parameters. A primary reason for employing this architecture is to achieve computational scalability when, for example, using very large SAR image datasets is required. This is clearly a necessary consideration when performing real-time navigation tasks, one which many existing deep network architectures fail to offer due to their size. Indeed, SqueezeNet contains significantly fewer parameters than other popular ANNs such as VGG-16 (with about 138M parameters) or AlexNet (about 62M parameters) that are often used for image processing applications. SqueezeNet is a convolutional neural network and details on its architecture can be found in [9].

2.1 Contrastive Loss and Triplet Loss

Learning in Siamese ANNs can be performed using contrastive loss or triplet loss, where the former takes a lower or higher “distance” value if a sample is encoded to something similar or dissimilar, respectively; whereas the latter loss function maximized the distance from a reference input to a dissimilar input and simultaneously minimizes the same distance to a similar input.

To train a neural network embedded with a contrastive loss function, we construct two datasets: \mathcal{D}_p containing positive pairs of images, and \mathcal{D}_n containing negative pairs. In other words, positive pairs contain geographically close images, whereas negative pairs contain images that are distant. Then, the Siamese neural network is constructed from two “sister” networks $f_a(x_a, \theta)$ and $f_b(x_b, \theta)$ with shared parameters θ that are learned by minimizing the following contrastive loss function:

$$L(e_a, e_b) = \begin{cases} d(e_a, e_b), & \text{if positive image pair} \\ \max(0, m - d(e_a, e_b)), & \text{if negative image pair,} \end{cases} \quad (1)$$

where $d(\cdot)$ is the distance between two feature vectors.

Another commonly used loss function for learning Siamese ANNs is the triplet loss, which is constructed using three “sister” networks $f_a(x_a, \theta)$, $f_b(x_b, \theta)$, and $f_c(x_c, \theta)$ with shared parameters θ . Image x_a is referred to as an *anchor*, image x_b is a positive image similar to x_a , and x_c is negative image dissimilar to x_a . Then, the triplet loss is defined as:

$$L(x_a, x_b, x_c) = \max\left(\|f(x_a, \theta) - f(x_b, \theta)\|^2 - \|f(x_a, \theta) - f(x_c, \theta)\|^2 + \alpha, 0\right). \quad (2)$$

With respect to the feature vectors, minimizing loss function 2 results in increasing distance between x_a and x_c and decreasing distance between x_a and x_b . As previously, in this case, similar images contain geographically close regions while dissimilar images contain distant regions. Figure 1 illustrates the high-level architecture of the Siamese ANN with triplet loss. Note that the figure would contain two networks, or “twin” networks, in the described case when contrastive loss is used.

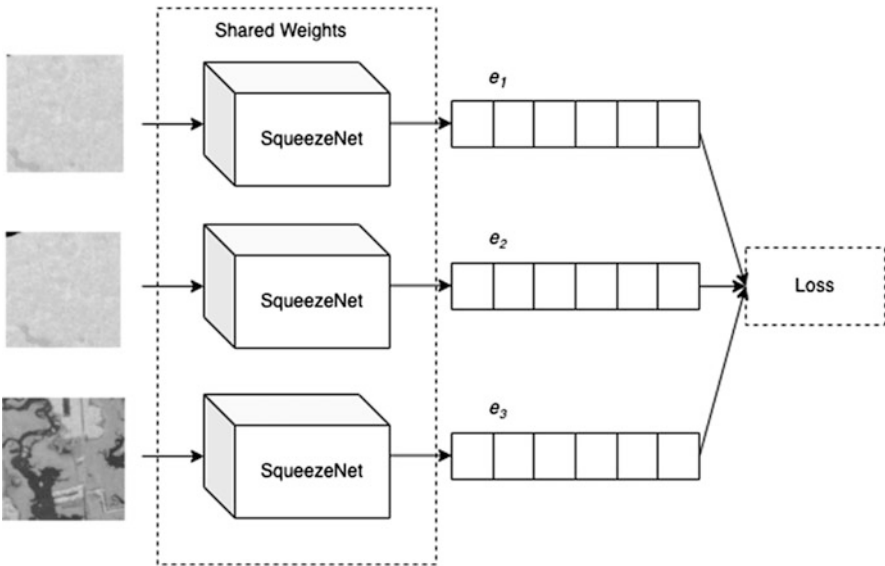


Fig. 1 Architecture of Siamese ANN with triplet loss

2.2 Dataset Generation and Retrieval

SAR data is represented as images with three channels and the geolocation (latitude and longitude) is known for each pixel. Thus, a dataset $\{\mathcal{D}\}$ can be represented as a set of tuples $\{(\mathcal{I}, l)\}$, where \mathcal{I} is an image and l is a corresponding location. In our work we use square images where l represents latitude-longitude coordinates of the center of the image. While training a Siamese ANNs, we consider two images \mathcal{I}_i and \mathcal{I}_j as “similar” if their intersection is non-empty. Otherwise, the images are considered to be “different” if their intersection is empty. This is illustrated in Fig. 1, where the images formed by the region in the dashed blue squares would be considered similar, while the image formed by the red square image would be considered different from any of the other images. To generate training data for a given geographical region (e.g., the region captured by Fig. 2) we first select a random location and then generate images that are similar to one another by shifting the selected location in random directions (up, down, left, right) multiple times using a predefined stride size while the intersections are non-empty. In order to generate different pairs we simply select two non-intersecting images at random distance from one another.

During the *retrieval* process, for given a query SAR image \mathcal{I}_q with unknown geolocation, the goal is to find its location by retrieving the geographically closest images from $D = \{(\mathcal{I}, l, e)\}$ with known coordinates. When constructing the database $\{\mathcal{D}\}$ we generate a grid of SAR images of the same size as the query image, then utilize a trained Siamese ANN to generate a global feature vector e for each of the images in $\{\mathcal{D}\}$. Finally, during retrieval we calculate $f(\mathcal{I}_q, \theta)$ and find coordinates corresponding to the top predefined number of closest vectors. To evaluate our method we retrieve coordinates of the closest vector in Euclidean space and compute geographic distance between geolocation of retrieved vector and ground truth geolocation of the query image using the Haversine formula.

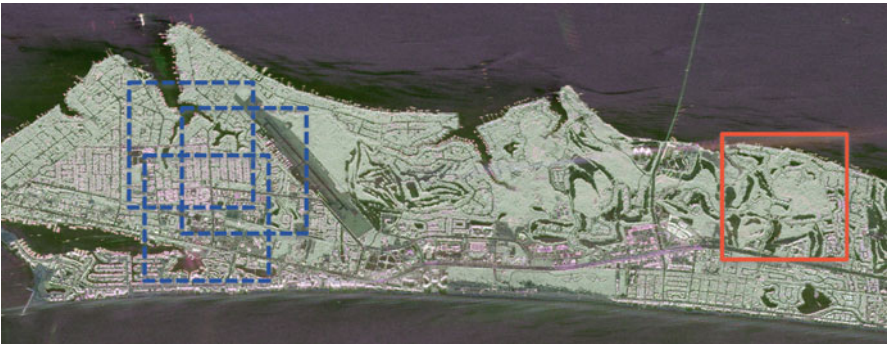


Fig. 2 Example of regional SAR image: blue squares represent similar images, red square represents a different image

3 Numerical Experiments and Results

In this section we outline the data used in our experiments, specify the architecture detail of the deployed Siamese ANN and present the corresponding computational results.

3.1 Data and Architecture Details

The data was obtained from a public SAR image data repository at NASA Jet Propulsion Laboratory.¹ The dataset contains 20,002 products from 1821 flight lines. Each product contains SAR data in multiple formats, including tiff, and metadata with geolocation. For the current experiments we use three SAR images contained within bounding box sizes $17,982 \times 11,388$, $3937 \times 26,905$, and $14,621 \times 25,956$, where, depending on the image, 10,000 pixels approximately correspond to a 40–50 km distance. The images were from various regions in Florida and California, USA, that contain both natural and man-made formations. The experiments were conducted on a computer with NVidia GeForce RTX 3090 with 24 GB GDDR6X memory.

We utilize online training dataset generation, where training inputs for the Siamese ANN (either tuples or triples) are randomly generated from the input images during each training epoch. Image patches of size 200×200 with three color channels and cut from original SAR images in tiff format are used. Two broad sets of experiments are conducted: the first involves patch images generated with an intersection by applying a stride length of 100 (Tables 1 and 2); the second involves patch images that do not contain an intersection by applying a stride length on 200 (Tables 3 and 4). Depending of the experiment configuration, the output of the neural network is an embedding vector with 128 or 256 dimensions, denoted by 128d and 256d in the tables. We implement the neural networks using Keras and utilize Adam as the optimization algorithm. The ANNs were trained for 40 epochs during all the experiments.

Table 1 Summary statistics (in km)

Architecture	Top-1			Top-5		
	Mean	Std	Median	Mean	Std	Median
Triplet loss, 128d	9.613	18.081	0.614	1.983	6.419	0.271
Triplet loss, 256d	9.223	17.801	0.554	2.129	5.774	0.265
Contrastive loss, 128d	6.595	15.590	0.452	0.786	2.377	0.246

¹ <https://uavsar.jpl.nasa.gov/>.

Table 2 Distance statistics for different cutoff values

Cutoff	Architecture											
	Triplet loss, 128d				Triplet loss, 256d				Contrastive loss, 128d			
	k	Mean	Std	Median	k	Mean	Std	Median	k	Mean	Std	Median
0.75	499	9.568	18.071	0.612	500	9.223	17.800	0.554	500	6.595	15.590	0.452
0.5	476	9.597	18.060	0.614	494	9.292	17.892	0.528	500	6.595	15.590	0.452
0.25	217	6.137	12.855	0.318	271	5.021	12.111	0.241	447	5.870	14.555	0.408
0.1	28	2.316	8.485	0.136	31	1.891	9.441	0.099	65	0.581	0.906	0.221

Table 3 Summary statistics (in km) for images with no intersection

Architecture	Top-1			Top-5		
	Mean	Std	Median	Mean	Std	Median
Triplet loss, 128d	15.747	22.963	4.886	3.185	6.510	0.718
Triplet loss, 256d	13.822	20.262	3.456	2.811	6.357	0.705
Contrastive loss, 128d	11.185	18.525	2.450	2.316	5.804	0.649

Table 4 Distance statistics for different cutoff values for images with no intersection

Cutoff	Architecture											
	Triplet loss, 128d				Triplet loss, 256d				Contrastive loss, 128d			
	k	Mean	Std	Median	k	Mean	Std	Median	k	Mean	Std	Median
0.75	481	15.685	22.928	5.009	491	13.329	19.438	3.330	500	11.185	18.525	2.450
0.5	369	14.458	21.422	4.368	439	12.575	18.499	3.113	495	11.003	18.239	2.443
0.25	104	8.929	14.935	0.815	116	5.812	12.280	0.468	315	9.477	16.609	1.713
0.1	12	0.485	0.959	0.081	6	0.305	0.419	0.126	21	0.896	0.989	0.257

For each experiment set, we compute summary statistics with respect to the distances between query images and retrieved images. Specifically, for each loss function configuration, for 500 query images we retrieve the K closest images and their coordinates from the dataset. In Tables 1 and 3 “top-1” and “top-5” represent $K = 1$ and 5 retrieved images for each query image introduced to the networks, respectively. In the case of $K = 5$, we calculate geographic distance for each of the five closest images, and return only the closest. In Tables 1, 2 and 3, an embedding distance limit, i.e., “cutoff,” is imposed to each distance in embedding space configuration. Images below the cutoff value are retrieved and the closest one is considered. The number of retrieved images for each cutoff value is denoted by k —observe that not all query images resulted in any retrieved images.

3.2 Computational Results and Discussion

The tables below provide the mean, standard deviation, and median computed distances in the described experiments. In Tables 1 and 2 it is clear that the median distances improved when the dimension of the Triplet loss was increased from 128

to 256. The same trend can be observed for the improvements in means, whose values are rather large between 1.9–9.6 km. It will be shown in the case study presented in Sect. 3.3, the large discrepancies between the mean and median values are due to a low number of outliers where the distances between query and retrieved images were large. This is clearly reflected in the standard deviation values present in the tables. When comparing the different loss function, other than the last two instances in Table 2, the contrastive loss function with 128 dimensions resulted in better median and mean values than Triplet loss, suggesting that it is more adept when deploying the proposed model. The discrepancy between these functions merits further examination and is something we reserve for future endeavors. Finally, when comparing between the “top-1” and “top-5” column blocks in Table 1, we can observe a significant improvement in performance, which suggests that there often are other regions located very closely to the query images (reflected in top-5).

Similar trends to the ones described above can be observed in Tables 3 and 4. However, when comparing the Triplet loss and contrastive loss function on a one-to-one basis in Tables 1 and 2 with the ones in Tables 3 and 4, respectively, other than for the last rows of Table 4, it is obvious that performance is degraded when the dataset consists of images that do not intersect geographically.

3.3 Case Study

For better visual exposition, we consider the experiment instance in Table 1 with “Triplet loss, 128d” corresponding to the “top-1” column block and plot several results. First, the training and validation loss evolution is illustrated in Fig. 3. The distances between embedding vectors and retrieved geographic distance are presented by the scatter plot in Fig. 4. Observe that a pronounced majority of instances are identified very closely relative to their true coordinates. However, several outliers extending approximately 80 km can be seen. Specifically, Table 1 above furnishes summary statistics of differences, where for the first set of experiments, the mean deviation is approximately 2 km, while the mean is approximately 13 km due to outlier influences. The performance is further reinforced by the histogram in Fig. 5 showing differences between the 500 retrieved images and their real coordinates, where it can be observed that over half the retrieved images were within 0–5 km from the query images. By way of computational effort, generation of embeddings for the 500 SAR images with 128d took 1.61 seconds using the specified GPU configuration, clearly demonstrating the potential for scalability of the proposed model.

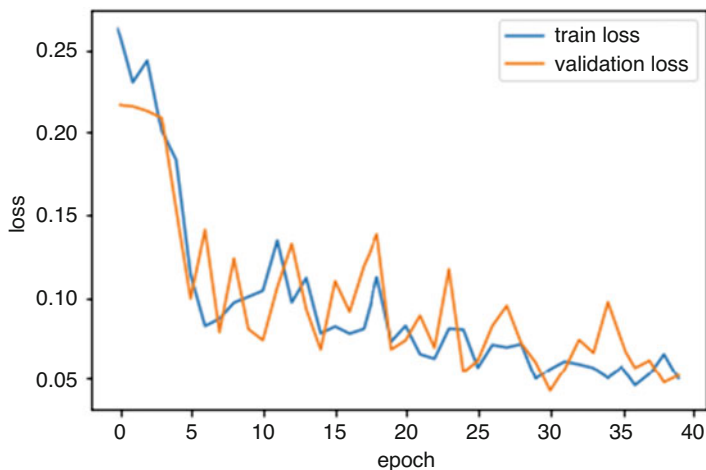


Fig. 3 Training and validation losses of Siamese network with triplet loss, 128d

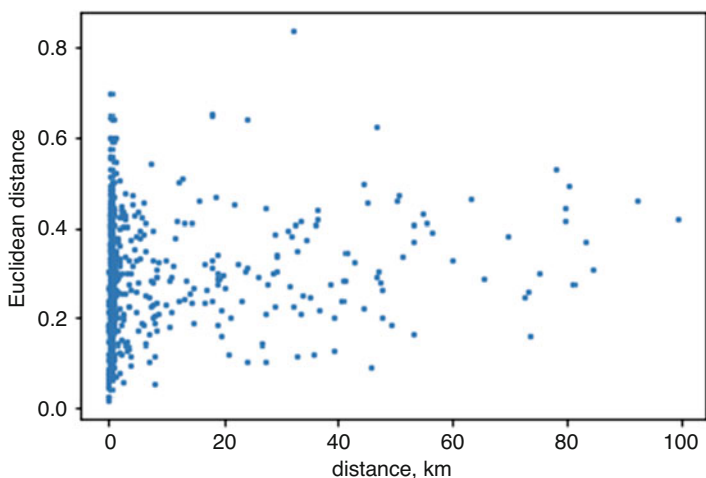


Fig. 4 Dependence between embedding space distance and geographic distance, 500 samples

4 Conclusions

In this chapter we proposed a deep learning model based on Siamese networks for navigation tasks. Comparative experiments were conducted for various learning settings using contrastive and triplet loss functions. Our results suggest that the developed model is effective and scalable. Experiments using contrastive loss and SAR datasets whose images were generated with topological overlap generally yielded superior results. Also, we observed that top-5 retrieved results contained

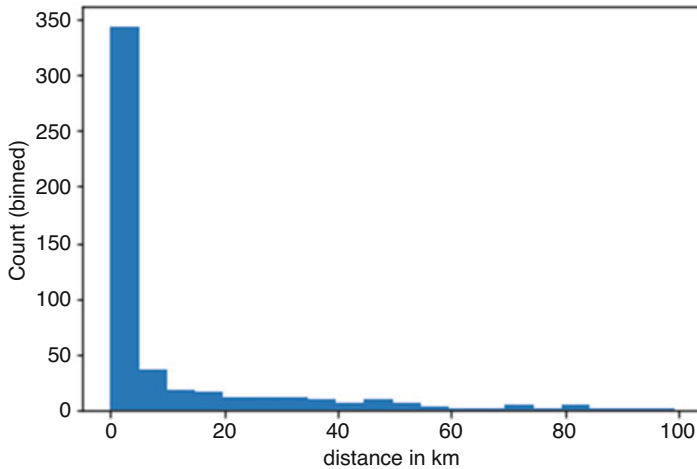


Fig. 5 Histogram of distance differences (in km)

much closer SAR regions than top-1. Potentially, it may be possible to first retrieve top- K images and subsequently rerank them using a tailored method such that the closest image out of K would be detected.

Our ongoing and future work will use a multitude of datasets representative of larger and more diverse global regions along with experimentation with very large-scale datasets. The described method is also being extended to consider zoom and other transformations, and loss functions that directly admit distances. We also seek extensions that do not rely on datasets formed by image patches generated at predefined stride lengths, but rather be able to scan the region of interest contained in the whole SAR image and thereafter find specific query locations and coordinates directly.

References

1. Arandjelovic, R., Gronat, P., Torii, A., Pajdla, T., Sivic, J.: NetVLAD: CNN architecture for weakly supervised place recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 5297–5307 (2016)
2. Bromley, J., Guyon, I., LeCun, Y., Säckinger, E., Shah, R.: Signature verification using a “Siamese” time delay neural network. In: Proceedings of the 6th International Conference on Neural Information Processing Systems, NIPS’93, p. 737-744. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA (1993)
3. Cao, B., Araujo, A., Sim, J.: Unifying deep local and global features for image search. In: European Conference on Computer Vision, pp. 726–743. Springer (2020)
4. Chen, S., Wang, H.: Sar target recognition based on deep learning. In: 2014 International Conference on Data Science and Advanced Analytics (DSAA), pp. 541–547. IEEE (2014)

5. Chierchia, G., Cozzolino, D., Poggi, G., Verdoliva, L.: Sar image despeckling through convolutional neural networks. In: 2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), pp. 5438–5441. IEEE (2017)
6. Cozzolino, D., Verdoliva, L., Scarpa, G., Poggi, G.: Nonlocal CNN SAR image despeckling. *Remote Sensing* **12**(6), 1006 (2020)
7. DeTone, D., Malisiewicz, T., Rabinovich, A.: SuperPoint: Self-supervised interest point detection and description. In: Proceedings of the IEEE conference on computer vision and pattern recognition workshops, pp. 224–236 (2018)
8. Gordo, A., Almazán, J., Revaud, J., Larlus, D.: Deep image retrieval: Learning global representations for image search. In: European conference on computer vision, pp. 241–257. Springer (2016)
9. Iandola, F.N., Moskewicz, M.W., Ashraf, K., Han, S., Dally, W.J., Keutzer, K.: SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <1mb model size. *CoRR abs/1602.07360* (2016). URL <http://arxiv.org/abs/1602.07360>
10. Noh, H., Araujo, A., Sim, J., Weyand, T., Han, B.: Large-scale image retrieval with attentive deep local features. In: Proceedings of the IEEE international conference on computer vision, pp. 3456–3465 (2017)
11. Parikh, H., Patel, S., Patel, V.: Classification of SAR and PolSAR images using deep learning: a review. *International Journal of Image and Data Fusion* **11**(1), 1–32 (2020)
12. Radenović, F., Tolias, G., Chum, O.: CNN image retrieval learns from bow: Unsupervised fine-tuning with hard examples. In: European conference on computer vision, pp. 3–20. Springer (2016)
13. Schmitt, M., Hughes, L.H., Zhu, X.X.: The sen1-2 dataset for deep learning in SAR-optical data fusion. *arXiv preprint arXiv:1807.01569* (2018)
14. Tang, X., Zhang, L., Ding, X.: Sar image despeckling with a multilayer perceptron neural network. *International Journal of Digital Earth* **12**(3), 354–374 (2019)
15. Teichmann, M., Araujo, A., Zhu, M., Sim, J.: Detect-to-retrieve: Efficient regional aggregation for image search. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 5109–5118 (2019)
16. Thoma, J., Paudel, D.P., Chhatkuli, A., Gool, L.V.: Geometrically mappable image features. *IEEE Robotics and Automation Letters* **5**(2), 2062–2069 (2020). doi:10.1109/LRA.2020.2970616
17. Thoma, J., Paudel, D.P., Gool, L.V.: Soft contrastive learning for visual localization. In: H. Larochelle, M. Ranzato, R. Hadsell, M. Balcan, H. Lin (eds.) *Advances in Neural Information Processing Systems*, vol. 33, pp. 11,119–11,130. Curran Associates, Inc. (2020). URL <https://proceedings.neurips.cc/paper/2020/file/7f2cba89a7116c7c6b0a769572d5fad9-Paper.pdf>
18. Yi, K.M., Trulls, E., Lepetit, V., Fua, P.: Lift: Learned invariant feature transform. In: European conference on computer vision, pp. 467–483. Springer (2016)
19. Zhang, Z., Peng, H.: Deeper and wider Siamese networks for real-time visual tracking. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 4591–4600 (2019)

A Comparison of Deep Neural Network Architectures in Aircraft Detection from SAR Imagery



Jin Xing, Ru Luo, Lifu Chen, Jielan Wang, Xingmin Cai, Shuo Li, Phil Blythe, Yanghanzi Zhang, and Simon Edwards

1 Introduction

Synthetic aperture radar (SAR) can offer continuous and stable Earth observation all day and all night, which has been intensively studied in various research domains [6]. With the fast development of SAR techniques, a large number of high-resolution spaceborne and airborne image datasets have been acquired, which provide exciting opportunities for SAR-based target detection. Nowadays, timely aircraft detection plays a pivotal role in traffic management and military activities, which has resulted in the design and development of a broad range of target detection algorithms [34].

The fast development of deep learning techniques in remote sensing has been reviewed by numerous scientists (e.g. [20, 46]). Generally, convolutional neural networks (CNNs) have been explored extensively due to their prominent feature extraction performance with end-to-end structures, which makes them the most widely used deep learning algorithms for man-made target detection in remote sensing [21]. When it comes to the aircraft detection using deep learning, various domain-specific solutions have been explored [4]. For example, Siyu et al. [29] combined CNN with data augmentation techniques to address the insufficient training sample challenge in automatic aircraft detection from SAR imagery, which was reported with 96.36% overall accuracy. Diao et al. [7] employed constant false alarm rate (CFAR) for aircraft pre-location to reduce the influence of background

J. Xing · S. Li · P. Blythe · Y. Zhang · S. Edwards
Newcastle University, Newcastle Upon Tyne, UK

R. Luo · L. Chen (✉) · J. Wang · X. Cai
School of Electrical & Information Engineering, Changsha University of Science & Technology,
Changsha, China
e-mail: Luoru@stu.csust.edu.cn; lifu_chen@csust.edu.cn; ningyuan@csust.edu.cn;
cc182022@stu.csust.edu.cn

© The Editor(s) (if applicable) and The Author(s), under exclusive license to
Springer Nature Switzerland AG 2022

M. Rysz et al. (eds.), *Synthetic Aperture Radar (SAR) Data Applications*, Springer
Optimization and Its Applications 199, https://doi.org/10.1007/978-3-031-21225-3_5

clutters, which was then integrated with faster region-based CNN (R-CNN) to perform aircraft detection in high-resolution SAR images. Most of these algorithms are composed of the region proposal and target detection stages, thus are named two-stage detection approaches in deep learning [8, 9]. However, its slow detection speed was considered a drawback, and new deep neural networks (DNNs) were continuously being studied for aircraft detection.

When the target detection speed is concerned, You Only Look Once (YOLO) family is frequently employed, as a typical one-stage target detection algorithm [12]. In 2015, Redmon et al. [26] proposed the YOLO algorithm to predict targets by directly performing classification regression on the input image. The average accuracy of the YOLO algorithm on Pascal VOC 2007 testing dataset reached 63.4%, which was about 10% lower compared with Faster R-CNN, but the detection speed was better than Faster R-CNN [27]. In 2017, YOLOv2 with Darknet-19 as the backbone network was proposed [25], in which the anchor mechanism based on k-means clustering calculation was used to improve the detection accuracy. Following these works, YOLOv4 [2] was invented to further improve the speed of detection. Recently, Scaled-YOLOv4 [36] and YOLOv5 [35] attracted increasing interests of researchers due to its advantages of the fast speed and satisfactory accuracy. We also note that current research is the initial stage of applying these one-stage target detection DNNs for aircraft detection, and domain-specific challenges of aircraft detection from SAR imagery are continuously being investigated [19].

Generally, there are three pivotal challenges in employing deep learning techniques for aircraft detection from SAR imagery. First, the scale heterogeneity of aircraft has made the feature extraction more challenging when using fixed receptive fields of most CNNs [44]. Second, the complex background usually incurs interference to the detection of aircraft, in which DNNs might be confused by objects with similar features (e.g. shapes and sizes) to aircraft [33]. Third, speckle noise, the multiplicative noise produced in radar echoes due to the SAR imaging process, introduces additional barriers in the aircraft detection process due to its ‘salt and pepper’ effect [3]. These challenges have been separately accommodated using different DNNs (e.g., [10]), but their effectiveness remains unclear due to the lack of integrated assessment work. Therefore, it becomes necessary to conduct comparison works of DNNs in aircraft detection to gain insights into how these three domain-specific challenges are tackled.

This book chapter aims to assess six widely used and cutting-edge DNNs for automatic aircraft detection from SAR imagery. To assure the consistence of this assessment work, the same collection of SAR images from Gaofen-3 system with 1 m resolution is selected as the testing datasets. Based on the literature review, we choose single-hot multi-box detector (SSD), Faster-RCNN, EfficientDet, YOLOv5, efficient weighted feature fusion and attention network (EWFAN) [37] and the efficient bidirectional path aggregation attention network (EBPA2N) [18]. Among these six DNNs, SSD, Faster-RCNN, EfficientDet and YOLOv5 are considered as general frameworks for object detection [43], while EWFAN and EBPA2N are specifically designed for aircraft detection with the integration of geospatial domain knowledge. This chapter provides a baseline for performance assessment in SAR-

based aircraft detection, to facilitate the integration of domain knowledge in SAR data analytics with deep learning techniques.

The rest of this book chapter is organized as follows. The testing dataset is delineated in Sect. 2. The architecture and implementation of six DNNs for aircraft detection are presented in Sect. 3. In Sect. 4, we assess their detection accuracy in three experiments and discuss their detection speed in SAR-based aircraft detection, with respect to their architecture, respectively. Finally, we conclude this book chapter and discuss future research in Sect. 5.

2 Datasets

Currently, there is no ubiquitously agreed upon standard SAR datasets for aircraft detection [23]; thus we choose our own SAR imagery datasets for better consistency in the DNN performance assessment. To evaluate the aircraft detection performance of widely used deep neural networks, 15 large-scale SAR images from Gaofen-3 system with 1 m resolution including the airports and aircraft are utilized. First, the aircraft samples are manually marked in SAR images and confirmed by SAR interpretation experts. Three large-scale SAR images of different airports are retained for independent testing to evaluate the performance of the network. Then, the dataset is generated automatically by the decomposition operation [42]. Since the number of manually annotated images is limited, we use 90-degree rotation, translation, flipping and mirroring to perform data augmentation [15]. Finally, 4396 slices with the size of 512×512 are obtained, and the ratio of training set and validation set is set to 4:1. A sample of these image tiles is shown in Fig. 1. From these image we can observe the size heterogeneity of aircraft, the complex background information, and also considerable speckle noise are present. SAR imagery datasets are more difficult to interpret compared with optical remote sensing images, thus present domain specific challenges for aircraft detection [4].

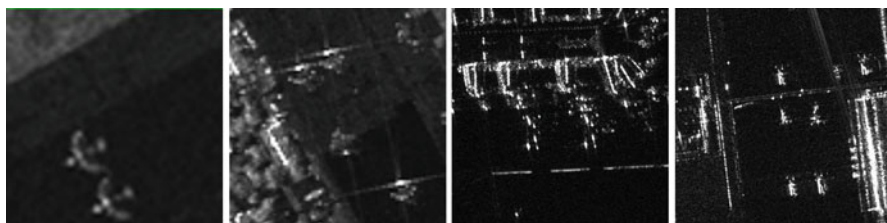


Fig. 1 Four 512×512 SAR image samples used in the performance comparison. We choose these SAR images to include the speckle noise, the heterogeneity of aircraft (e.g. sizes, shapes and positions) and the complex background to better evaluate the performance of aircraft detection

3 Methods

The automatic detection of aircrafts from SAR images is widely applied in various research fields, but there are three major challenges in aircraft detection: the speckle noise from SAR systems, the size and shape heterogeneity of aircraft and the interference of complex background [18]. For most existing target detection DNNs, SAR domain knowledge is incurred to accommodate the handling of speckle noise [14], but the other two challenges are less noticed. Therefore, we compare six DNNs, namely, SSD, Faster-RCNN, EfficientDet, YOLOv5, EWFAN and EBPA2N, to assess how they tackle these three challenges together in automatic detection of aircraft from SAR imagery. Before conducting the experiments of performance comparison, we first present these six DNNs.

3.1 SSD

SSD network is a popular convolutional neural network, which is best known for its fast detection speed with good accuracy [17]. There are many researches working in target detection favouring SSD network for object detection [8, 9, 39], since SSD is able to conduct object localization and classification in a single forward pass of the network. This DNN has also been applied to aircraft detection in remote sensing images [24], and its noticeable accuracy makes it a good candidate in our performance assessment experiments.

We illustrate the architecture of SSD in Fig. 2. SSD employs the 16-layer visual geometry group (VGG16) network as the backbone to extract features hierarchically from the given image. Six multiscale feature maps are generated by SSD for aircraft detection, with the large-size feature maps targeting small target and the small-size feature maps for large target detection. This feature extraction process is implemented as follows. Assuming the size of the input SAR image is set to 300×300 , the feature map with the dimension of 38×38 output from the Conv4_3 layer (the third convolutional layer of the four module) in VGG16 is used as the first

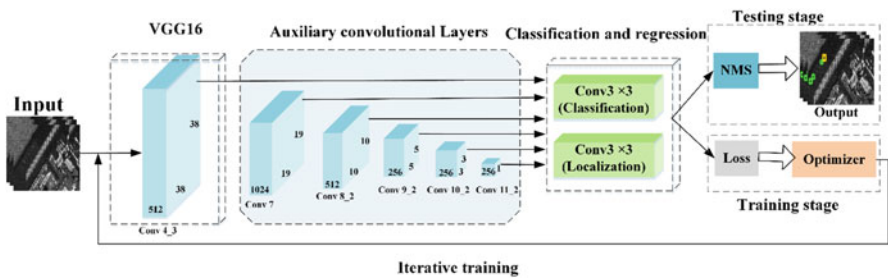


Fig. 2 SSD architecture in our experiments

feature map in our six multiscale feature map. At the same time, the fully connected layer and dropout layers in the original VGG16 are removed, and then five auxiliary convolutional layers are added after them to produce multiscale feature maps with dimensions of 19×19 , 10×10 , 5×5 , 3×3 and 1×1 , respectively. At this point, all six feature maps of different sizes are generated for subsequent classification and regression analytics. Finally, the classification of target objects and the regression of the location bounding boxes are performed via two 3×3 convolutional layers, respectively. And the detection results are output after non-maximum suppression (NMS) screening [28], to select the optimal aircraft bounding boxes.

3.2 *Faster R-CNN*

A faster R-CNN is proposed by Ren et al. [27] and is one of the most widely utilized DNN algorithms in the field of target detection. It relies on the region proposal networks (RPNs), which effectively simplifies the selection of candidate regions (i.e. target bounding boxes) in the process of detection. Faster R-CNN enables the parameter sharing among RPNs and convolutional layers, which could accelerate the speed of the algorithm training and target detection. We also note the efficiency of Faster R-CNN largely depends on the RPN, which is usually considered as a trade-off between the detection accuracy and the computation speed [11].

The architecture of Faster R-CNN includes three parts: the CNN, RPN and classification and regression modules, as shown in Fig. 3. The CNN module is designed to perform convolutional operations to the input image to extract salient feature maps. The RPN is the module used to generate candidate regions, which replaces the original selective search algorithm and can obtain a series of candidate regions on the feature map output from the convolutional network (Ren et al., *ibid.*). The steps of the RPN to generate candidate regions are delineated as follows. First, the RPN employs a sliding window with each pixel in the feature map as the centre to produce a series of anchors. Second, two 1×1 convolutional layers are incurred, one for the binary classification to determine the probability that the anchors belong to the foreground and background and one for regression to produce the coordinate offsets of the anchors, which are named as region proposals. Third, the targets are classified and regressed using the candidate region proposals, which are then normalized by sizes and the previously obtained feature maps to generate the exact positions of each targets. Finally, the detection results are offered as the combination of classification and positions.

3.3 *EfficientDet*

In 2019, Tan et al. [32] proposed a powerful one-stage detection algorithm named EfficientDet. It combines EfficientNet [31] and the weighted bidirectional feature

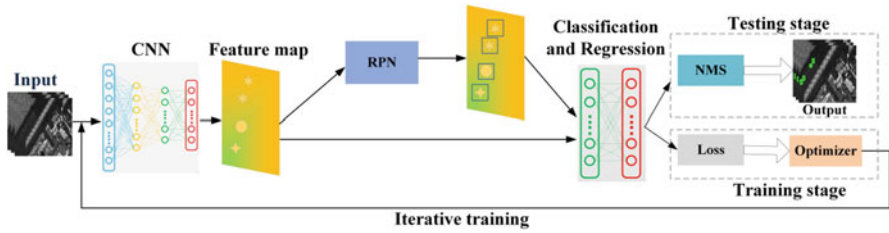


Fig. 3 Illustration of the Faster-RCNN architecture

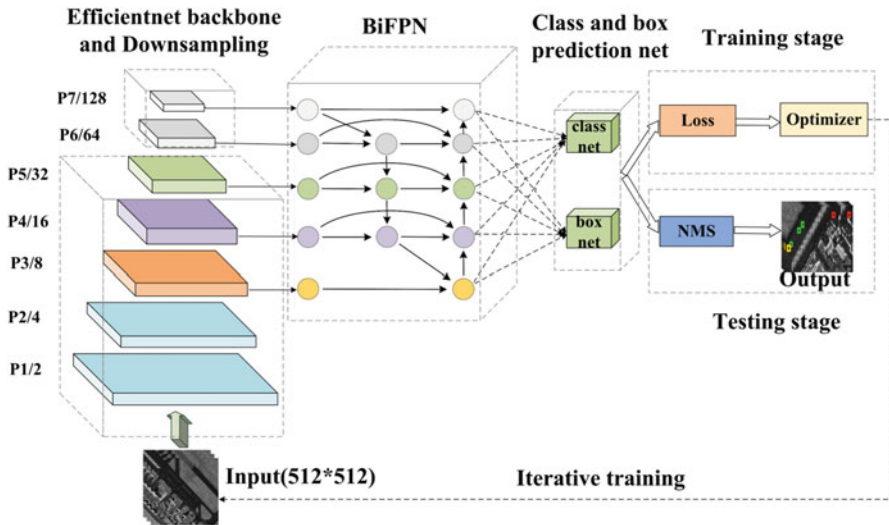


Fig. 4 Our implementation of the EfficientDet

pyramid network (BiFPN). EfficientDet has been employed for aircraft detection by Wang et al. [38] using aerial remote sensing images. EfficientDet was highly scalable according to various sizes of the input image to achieve remarkable performance, but the computation costs also increased correspondingly [30]. The balance between the detection accuracy and the computation costs has also been explored recently [13], which will attract increasing research interests in the future study of aircraft detection.

The overall architecture of EfficientDet is shown in Fig. 4. There are eight different models in EfficientDet, which are named D0 ~ D7, and we mainly use EfficientDet-D0 in this book chapter. The feature extraction backbone network is EfficientNet, and we choose the five feature maps output from its third to seventh convolutional layers for feature fusion in the BiFPN network, as a more efficient feature fusion approach based on PANet [16]. First, to simplify the network, we remove the neurons on the top-most and bottom-most branches that contribute less to the overall performance. Second, the original feature maps of the middle three

layers are separately added to the fused target feature map to enhance the performance of feature fusion. Third, the top-down and bottom-up paths are employed as another fusion module, to achieve higher-level feature fusion. Finally, when feature maps with different resolutions are fused, a weight is assigned to each feature map according to the contribution rate, allowing the network to automatically learn the importance of each feature map to enhance feature representativeness. After obtaining the five effective feature layers of the enhanced output after BiFPN, the results are predicted using a classification regression network, and the final detection results are generated using the non-maximum suppression (NMS) algorithm [22].

3.4 Efficient Weighted Feature Fusion and Attention Network (EWFAN)

The efficient weighted Feature Fusion and Attention Network (EWFAN) is proposed by Wang et al. [37], as a state-of-the-art DNN specifically designed for aircraft detection from SAR imagery. It is based on the EfficientDet-D0 framework [32] in which the adaptively spatial feature fusion (ASFF) and residual spatial attention module (RSAM) are introduced in the feature fusion process to integrate and extract high-level features. Moreover, based on complete intersection-over-union (CIoU) loss and focal loss, the complete intersection focal (CIF) loss function is proposed to measure aircraft detection accuracy in SAR image analytics.

The specific network structure of EWFAN is shown in Fig. 5. The EWFAN algorithm uses EfficientNet as the backbone network. The images are input into the backbone network and down-sampled, and then five feature maps of different resolutions are generated. These feature maps are input into weighted feature fusion and attention module (WFAM) module, which are further processed by the BiFPN module to obtain five enhanced feature maps. Among these five enhanced feature maps, three higher resolution feature maps are fed into the ASFF module, which is designed to focus more on spatial information fusion, and effectively suppress the interference of negative samples. For aircraft detection, there is less spatial information within the low-resolution feature maps due to the small target size. For the two lower-resolution feature maps accommodated by BiFPN, the introduction of RSAM (i.e. a combination of spatial attention module (SAM) [40] and residual connectivity) can effectively highlight the spatial information and improve the effectiveness of small target detection. Then, in the classification and regression steps, EWFAN generates nine prior frames with different sizes and aspect ratios on each grid of each effective feature layer, based on which the classification and regression operations are performed to obtain the prediction results. In the training process, the advanced CIF loss combines CIoU loss with the focal loss to improve the stability and accuracy of training, which also speeds up the training convergence to some extent.

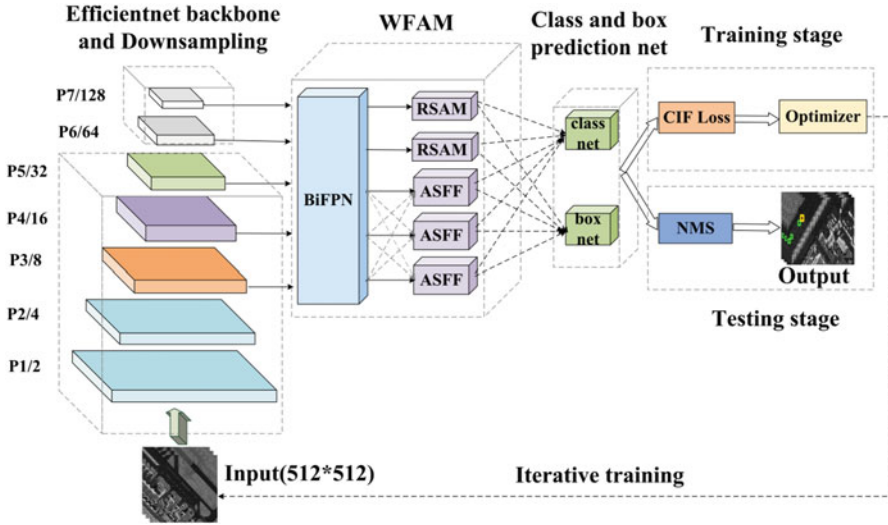


Fig. 5 The overall architecture of the efficient weighted feature fusion and attention network (EWFAN) in our implementation

3.5 YOLOv5

YOLOv5 is a lightweight DNN that presents excellent speed and precision in automatic object detection [35]. It relies on a single deep convolutional neural network (e.g. VGG16) that splits the input image into a grid of image tiles and each tile directly predicts a bounding box and object classification. Similar to EfficientNet, the YOLOv5 network also uses additional parameters of the depth and width of the network, so as to scale the network. The specific network structure is shown in Fig. 6. Firstly, the image is fed into a five-stage feature extraction backbone network with convolution blocks and down-sampling modules to extract rich image features and form high-level feature maps. Secondly, the CSP-PANet module, which is an effective combination of CSPNet and PANet, is proposed as a fusion module to integrate the semantic and detailed information of feature maps at different levels of the backbone network to enhance the feature representation capability. Finally, 1×1 convolution is applied for classification and regression to obtain the category and position of the detected target, and the detection results of the target are labelled in the form of boundary boxes. YOLOv5 not only has the advantages of YOLO including fast and low background error detection rate but also has the advantages of high model flexibility and strong universality. It can select the appropriate model for different detection tasks. Meanwhile, the positioning accuracy and recall rate of YOLOv5 need to be further improved.

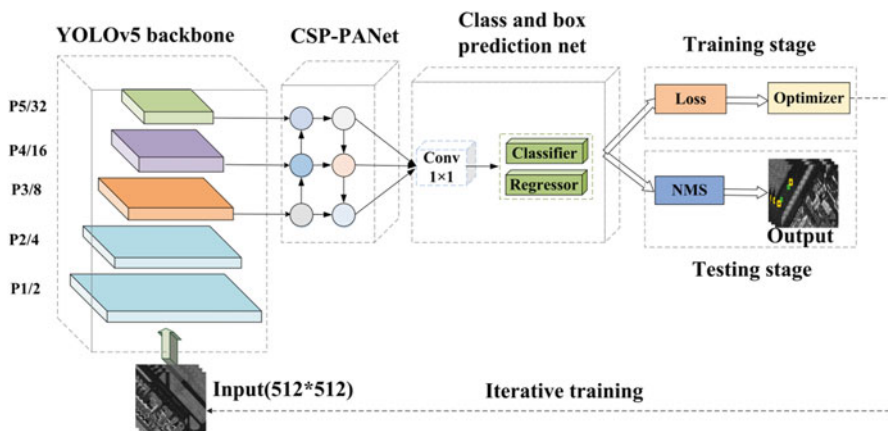


Fig. 6 The YOLOv5 Implementation

3.6 Efficient Bidirectional Path Aggregation Attention Network (EBPA2N)

Efficient bidirectional path aggregation attention network (EBPA2N) is based on the decomposition-recomposition methodology [42], with the innovative involution enhanced path aggregation (IEPA) module and effective residual shuffle attention (ERSA) module to efficiently capture the relationship among aircraft's backscattering features. Due to the combination of SAR analysis and attention mechanism, EBPA2N could better accommodate multiscale geospatial information in aircraft detection and meet the three challenges in aircraft target detection with a good detection performance. However, in some scenes with weak target imaging information and complex electromagnetic scattering characteristics, its detection performance is degraded.

The specific network structure is shown in Fig. 7. The input image goes through the YOLOv5 backbone network for feature extraction. The last three output feature maps, namely, P3, P4 and P5, are selected and input into the IEPA module and the parallel effective residual shuffle attention (PERSA) module for bidirectional path aggregation to enrich the representativeness of multiscale features. The IEPA module is used to fuse the three feature maps with different resolutions output from the backbone network to learn multiscale geospatial information. Then, three parallel ERSA modules are used to refine the multiscale features of the aircraft. In addition, a classification and bounding box prediction network is employed to generate preliminary predictions. Last but not the least, the duplicate predictions are filtered by the NMS algorithm to achieve the final aircraft detection results.

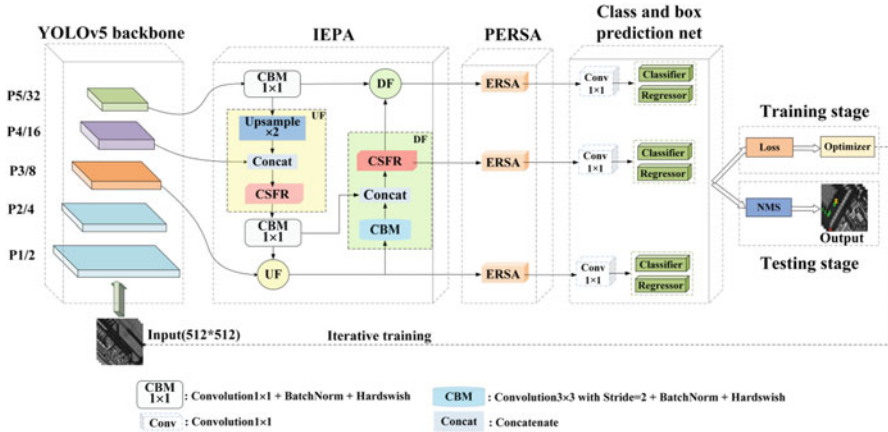


Fig. 7 The overall architecture of the efficient bidirectional path aggregation attention network (EBPA2N)

4 Results and Discussion

In this section, we compare the performance of aircraft detection using the six presented DNNs and discuss their advantages and disadvantages. Knowledge discovery from this performance comparison will offer guidance for performance improvement of existing DNNs and provide insights for future design and development of advanced DNNs for aircraft detection. We also note this book chapter provides a baseline for the performance assessment of DNNs in SAR-based aircraft detection.

4.1 Aircraft Detection in Airport I

In the first experiment, the SAR image is acquired from a small airport. As shown in Fig. 8, we could notice small buildings near the airport I, which may impact the accuracy of aircraft detection if they have similar shapes [4]. The false alarms outside the airport can be removed via geofencing, and we focus on the missed detection and false alarms inside the airport in the detailed enlargement figure. The airport is a military airport, and there are always 33 aircraft at the airport that are small and dense, but the imaging features are obvious. The missed detection rate of SSD is the highest in the experiment, with five missed detections and five false alarms. YOLOv5 has three missed detections and two false alarms. Faster-RCNN has no missed detections and only three false alarms. EfficientDet-D0, EWFAN and EBPA2N all have only one missed instance. However, it is worth noting that EfficientDet-D0 has a non-negligible false check problem. EWFAN and EBPA2N

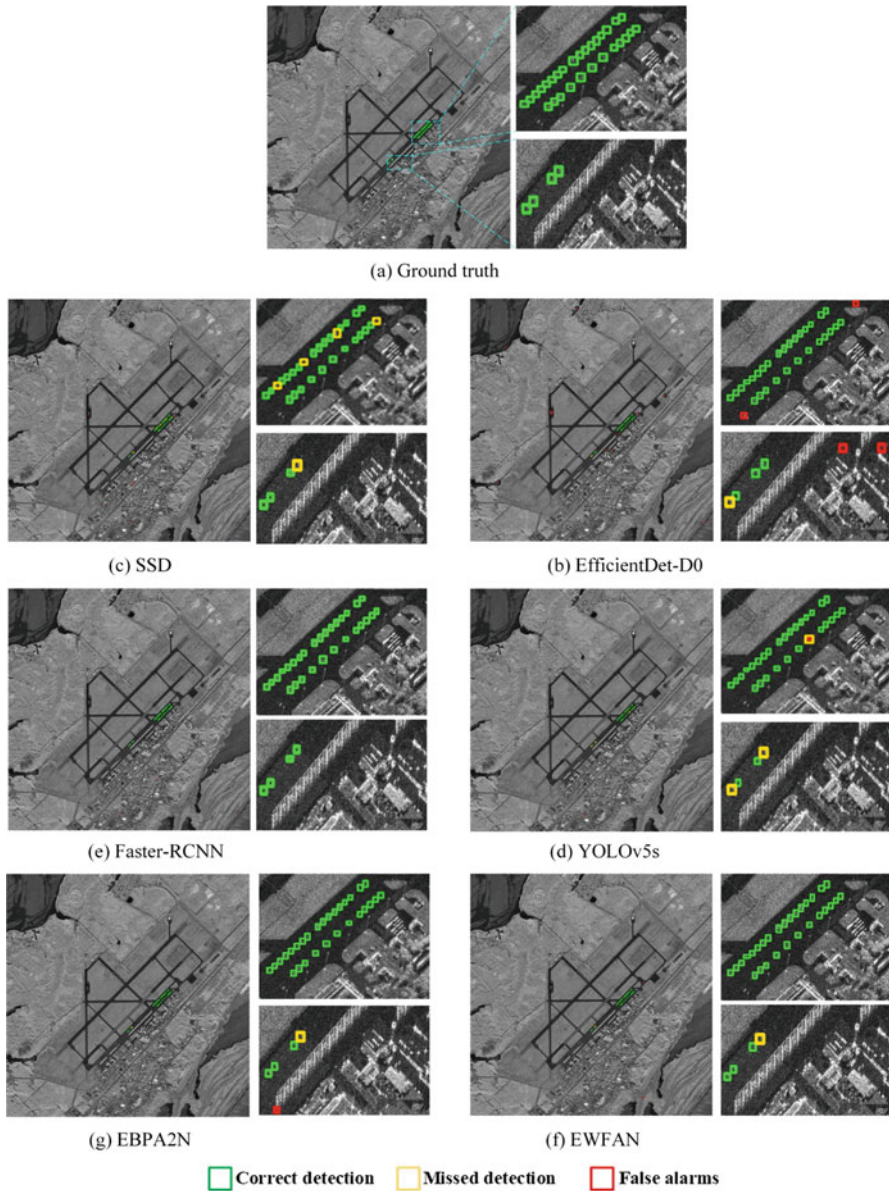


Fig. 8 The detection results of different DNNs for airport I. Red boxes and green boxes are false alarms and the correctly detected aircraft, respectively, and the yellow boxes are the missed detections

have very few false alarm results, only two and one, respectively. To summarize, the aircraft detection results clearly indicate the advantage of EWFAN and EBPA2N in the airport I.

4.2 *Aircraft Detection in Airport II*

In the second experiment, airport II is a large civilian airport (i.e. Shanghai Hongqiao Airport in China), as shown in Fig. 9. The image contains complex a background, with a considerable number of vehicles in the parking lot, dense residential areas and several traffic roads. All the background information could incur interference and false alarms. The aircraft in this airport presents multidirectional characteristics with complicated discrete scattering features, which makes detection even more difficult. From Fig. 9, the EfficientDet-D0 has some noticeable false alarm regions; YOLOv5 and Faster R-CNN have inaccurate localization (i.e. yellow missed boxes overlapped with red false alarm boxes) and more missed detections. The detection performance of EWFAN is better compared to YOLOv5 and Faster R-CNN. Furthermore, SSD and EBPA2N are the two DNNs with the best detection results obtained in this experiment, which are very close to the ground-truth images. But from the overview of this airport, SSD presents 21 false alarms, while EBPA2N only has 7 false alarms. Therefore, EBPA2N achieves the best performance in the aircraft detection at airport II.

4.3 *Aircraft Detection in Airport III*

Airport III is also a large civil airport (i.e. Beijing Capital International Airport in China) with three long runways, heavy traffic and the most complex background in our experiments. The metallic facilities in the airport (i.e. objects are shown as strong scattering points in SAR images with similar characteristics to aircraft) are very likely to perturb aircraft detection and cause false alarm. As shown in Fig. 10, these six algorithms have different false alarms and missed detection problems. From the zoomed-in view of the upper left corner of the airport, we can see that EfficientDet-D0 is less accurate in detecting small targets. The overall detection results indicate that EfficientDet-D0 has significant false alarm areas, many of which are located on the airport runway, which requires further analytics [33]. Similar problems could be found with Faster-RCNN and EWFAN, as false alarm highlighted using red boxes. While SSD has fewer missed detections, there are more red bounding boxes in the figure, which implies false aircraft detections. EBPA2N presents the best detection accuracy in this experiment, with a better balance of missed and false alarms.

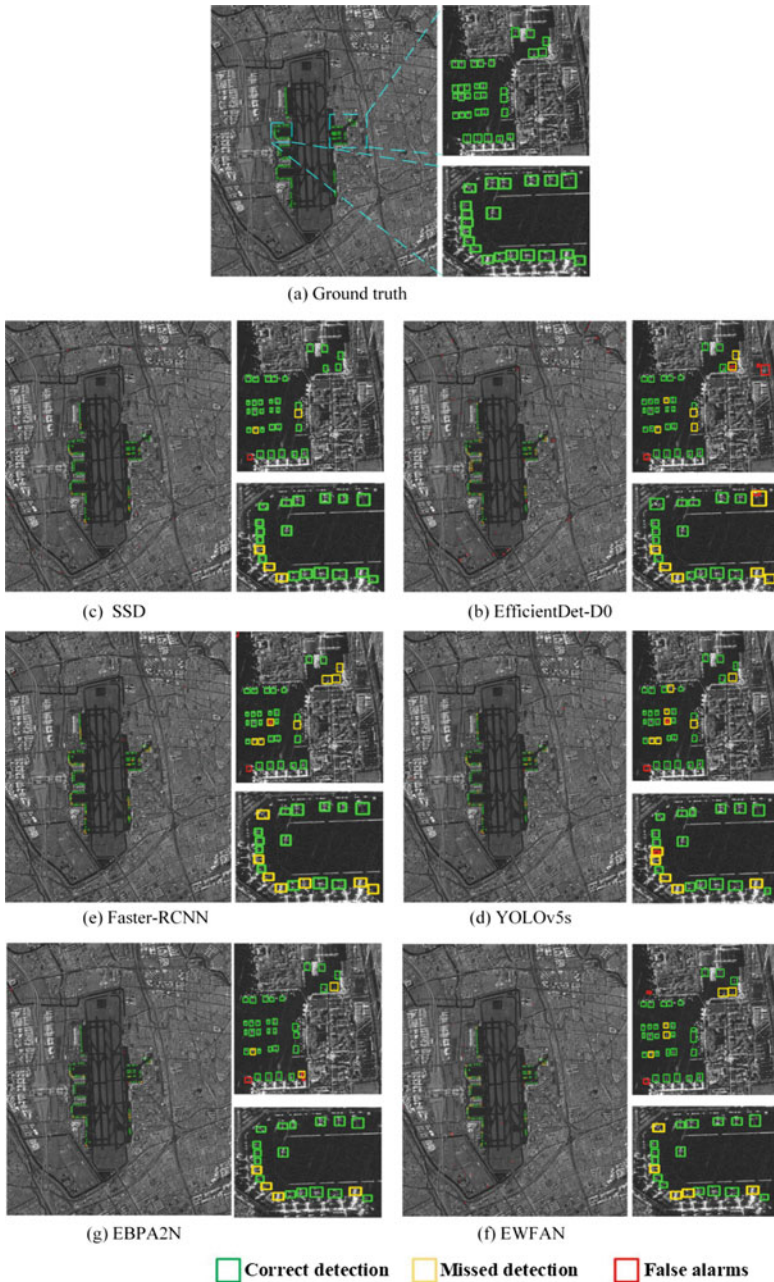


Fig. 9 The detection results of different frameworks for airport II. Red boxes and green boxes are false alarms and the correctly detected aircraft, respectively, and the yellow boxes are the missed detection

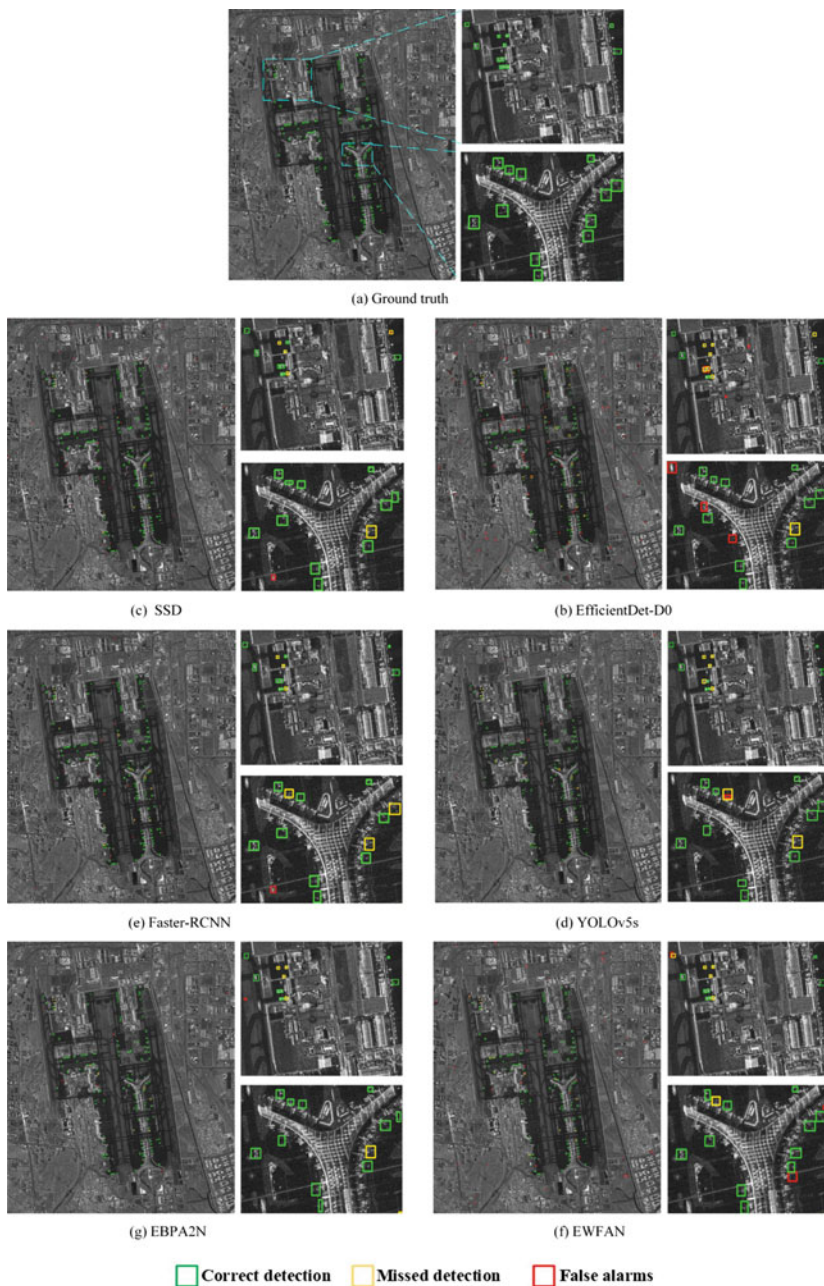


Fig. 10 The detection results of different frameworks for airport III. Red boxes and green boxes are false alarms and the correctly detected aircraft, respectively, and the yellow boxes are the missed detection

4.4 Evaluation and Analysis of Detection Performance

Table 1 shows the aircraft detection performance metrics of the different DNNs using SAR imagery acquired at the three airports described above. In Table 1, correct detection number (CDN), ground truth number (GTN), false alarm number (FAN) and the number of detected targets (DTN) denote the number of correctly detected aircraft, the number of real aircraft at the airport, the number of false alarms and the number of detected targets, respectively. We use t_1 to denote the testing time for each SAR image, and t_2 is the training time to complete 100 epochs of the same SAR dataset. From the training time and testing time of the network, we can find that SSD and Faster R-CNN required the longest computation time, which means they might not be suitable for real-time aircraft detection [19]. EfficientDet-D0 and EWFAN are in the middle with respect to computational time, while YOLOv5 and EBPA2N were the fastest. Focusing on the overall detection accuracy and false alarm rates, it can be seen that EBPA2N is the performed best, with an overall detection accuracy and overall false alarm rate of 91.90% and 4.90%, respectively.

Table 1 Detection performance of different networks

Network	Airport	DR (CDN/GTN)	FAR(FAN/DTN)	t1 (s)	t2 (h)
EfficientDet-D0	I	96.97% (32/33)	23.81% (10/42)	5.98	5.10
	II	77.50% (93/120)	34.51% (49/142)	18.05	
	III	83.22% (119/143)	46.64% (104/223)	28.03	
	<i>Overall</i>	<i>82.43%(244/296)</i>	<i>40.04% (163/407)</i>	<i>17.35</i>	
SSD	I	84.85% (28/33)	15.16% (5/33)	6.81	20.67
	II	87.50% (105/120)	16.67% (21/126)	12.25	
	III	93.01% (133/143)	24.00% (42/175)	16.53	
	<i>Overall</i>	<i>89.86% (266/296)</i>	<i>20.35% (68/334)</i>	<i>11.86</i>	
YOLOv5	I	90.91% (30/33)	6.25% (2/32)	4.80	0.69
	II	80.83% (97/120)	8.49% (9/106)	8.24	
	III	90.21% (129/143)	5.15% (7/136)	12.11	
	<i>Overall</i>	<i>86.49% (256/296)</i>	<i>6.57% (18/274)</i>	<i>8.38</i>	
Faster R-CNN	I	100% (33/33)	8.33% (3/36)	36.72	26.67
	II	72.50% (87/120)	8.42% (8/95)	68.90	
	III	89.51% (128/143)	12.33% (18/146)	92.40	
	<i>Overall</i>	<i>83.78% (248/296)</i>	<i>10.83% (30/277)</i>	<i>66.01</i>	
EWFAN	I	96.97% (32/33)	5.88 (2/34)	6.13	5.35
	II	80.83% (97/120)	15.65 (18/115)	18.49	
	III	88.11% (126/143)	20.75 (33/159)	28.80	
	<i>Overall</i>	<i>86.15% (255/296)</i>	<i>17.21% (53/308)</i>	<i>17.81</i>	
EBPA2N	I	96.97 (32/33)	3.03 (1/33)	5.01	0.882
	II	89.17 (107/120)	6.14 (7/114)	9.68	
	III	93.01 (133/143)	4.32 (6/139)	13.50	
	<i>Overall</i>	<i>91.90% (272/296)</i>	<i>4.90% (14/286)</i>	<i>9.40</i>	

By evaluating the six DNNs, we can see that the difficulty of aircraft detection continues to lie in the background complexity of the SAR scene and the heterogeneity of aircraft features (e.g. shapes, sizes and positions), making it challenging for DNNs to capture the essential features of the aircraft. From these three experiments, we can see that Faster R-CNN benefits from the two-stage detection mechanism, and the network achieves better performance in suppressing interference and false alarms. However, the network is inferior in accommodating the discrete scattering features of the aircraft. Especially in the experiment for airport II, when there are plenty of complex background objects (i.e. objects that are not aircraft) in the airport, the scattering information after imaging the aircraft with different positions and directions of parking varies greatly. We also note the detection accuracy of Faster R-CNN is poor, only 72.5%, which is the lowest among the six networks. Moreover, the network training and testing time of Faster R-CNN is the longest, which indicate the shortcoming of Faster R-CNN in tackling the scale heterogeneity and complex background information when performing real-time aircraft detection.

In aircraft detection using SAR images, EfficientDet-D0 presents noticeable false alarm problems, and the effective feature extraction of aircraft is the worst among the six networks. EWFAN is based on EfficientDet-D0, which mitigates the false alarm problem and improves the accuracy of aircraft detection to some extent, while reserving the fast speed of aircraft detection. The detection accuracy of EWFAN is close to that of SSD. Due to the use of six different resolution feature maps for multiscale prediction, SSD can better capture the multiscale features of the target and has a relatively better detection performance (i.e. the detection accuracy is 89.86%). However, the false alarm rate of SSD cannot be ignored, which is as high as 20.35%. In addition, the scattering interference from the neighbourhood features in the airport I–III scenes is gradually enhanced. As the background becomes more and more complex, the false alarm problems of EfficientDet-D0, EWFAN and SSD have increased correspondingly. This indicates that these three DNNs can achieve more efficient detection of aircraft in scenes with less scattering interference within the given neighbourhood and can obtain more satisfactory detection accuracy in complex scenes by further combining with that of geofencing information of airports [33]. Compared with EWFAN and EfficientDet-D0, SSD has the advantage of good detection speed, but the training time is longer given the same dataset.

When model deployment and computational resources are of concern, EWFAN and EfficientDet-D0 networks have a smaller number of parameters and computation intensity (i.e. the number of parameters of EfficientDet-D0 is only 3.9 M), which makes them strong candidates for fast aircraft detection. Although EWFAN adds a certain number of additional parameters based on EfficientDet-D0, it is still more ‘lightweight’ compared to SSD and suitable for deployment in the edge computing for SAR aircraft detection [45].

The network computation speed of YOLOv5 and EBPA2N have remarkable advantages and are the best among all networks. Although the false alarm rate of YOLOv5 is low, which is 6.57%, it is worth noting that its detection rate is only 86.49%, which means that there are many missed detections. And the detection performance of EBPA2N is satisfactory with less missed detections and

false alarms, which turns out to be more suitable for fast SAR-based image aircraft detection in some scenarios and tasks with requirements for real-time performance. From the three performed experiments, we found that although various advanced target detection networks have emerged in optical image processing, the results of directly migrating these networks to SAR image target detection are not satisfactory. EWFAN and EBPA2N integrate SAR domain knowledge to improve the detection performance, which has been verified in our experiments. EBPA2N network combines SAR analytical methods to accommodate the three challenges of aircraft target detection, and proposes involution enhanced path aggregation (IEPA) module to capture the relationship between aircraft discrete scattering features in a large range, so as to achieve more effective extraction of salient features of aircraft and reduce missed detection significantly. The effective residual shuffle attention (ERSA) module is proposed specifically to enhance aircraft feature extraction geospatially to deal with the interference of complex background information and speckle noise. This provides an important reference for the development of SAR special network by explicitly combining the scattering characteristics of targets in the future SAR field. Therefore, these comparison experiments confirm the necessity of integrating domain knowledge in SAR image analytics into the design and development of DNNs for SAR-based aircraft detection.

5 Conclusion and Future Works

Based on the all-day, all-weather SAR imaging systems, aircraft detection could be widely implemented in various fields (e.g. civil disaster monitoring, traffic management and urban planning). On the other hand, the speckle noise and complex background in SAR images make aircraft detection more challenging compared to optical imagery [10]. The diverse sizes, shapes and positions of aircraft target further undermines the direct application of deep learning techniques in SAR-based aircraft detection. This chapter selects six frequently utilized DNNs to compare their performance and highlights the requirement of domain knowledge integration for SAR-based aircraft detection. Following this performance assessment study, there are three important directions we plan to explore in the future research.

First, we plan to explore new methodologies of data fusion and the use of multi-source data (e.g. SAR data and optical data) to further enhance the accuracy of aircraft detection. There are very few publicly available SAR images for aircraft detection, and it is even more difficult to guarantee the quality of such imaging and ground-truth datasets. In contrast, there are relatively more optical aircraft samples, although the visual saliency of aircraft targets in obliquely projected SAR images is very different from that of centre-projected optical images, which are closer to human visual perception. Fortunately, advanced SAR and optical imagery data fusion methods have been explored using deep learning techniques

[5]. For SAR images acquired from different systems (e.g. different bands and different resolutions), the backscatter characteristics of the ground targets might be very different, which indicates the necessity of incurring additional pre-processing work [4].

Second, we will continue integrating SAR domain knowledge into the design and development of new DNNs for aircraft detection. Most of the current aircraft detection DNNs focus on combining existing deep learning techniques with imagery data analysis, which have not been effectively and explicitly integrated with the physical characteristics of SAR imaging, accommodating the electromagnetic scattering characteristics of SAR image objects and designing spatial-specific DNNs for SAR target detection. This is also a challenging but necessary topic in the future to realize the wider application of SAR imaging techniques.

Third, the investigation of interpretable SAR detection DNNs will be another focus of our future work [41]. The black box behaviour of DNNs makes their internal decision-making process difficult to understand, which significantly undermines the trustworthiness of DNNs in the domain of remote sensing applications. At this point, the combination of DNN explanation tools [1] is particularly important to help scientists understand the decision-making basis of the DNN for model development and model diagnosis. At the same time, the explanation output (e.g. heatmaps and DNN layer contribution measurements) of these explainable artificial intelligence algorithms could be utilized to measure the quality of the training and testing datasets, which will contribute to the establishment of standard baseline testing datasets for DNNs in SAR-based aircraft detection.

To summarize, this book chapter offers a baseline of performance comparison of popular DNNs in SAR-based aircraft detection. We select SAR images acquired at three different airports containing diverse aircraft and complex background information to evaluate the performance of six DNNs including the SSD, FasterRCNN, EfficientDet, YOLOv5, EWFAN and EBPA2N. The overall performance of EBPA2N is the best due to its SAR-specific attention mechanism, which could adaptively assign higher weights to aircraft regions via the analytics of geospatial neighbourhood information [18]. We also pay additional attention to the training and testing time of these six DNNs, to measure their potential in fast or real-time aircraft detection using SAR imaging systems. This book chapter highlights the importance of integrating SAR domain knowledge into the design and development of new DNNs, and this methodology could be extended to other target detection research using SAR imagery datasets.

Acknowledgement This work was supported by the National Science Foundation Grant (Nos: 41201468, 41701536, 61701047 and 41941019) of China; Fundamental Research Funds for the Central Universities, CHD (No. 300102260301/087 and 300102260404/087) and Scientific Research Fund of Hunan Provincial Education Department (18A148).

The authors also give their special thanks to the book editors, Professor Maciej Rysz and Professor Panos M. Pardalos, for their support and guidance of this book chapter.

References

1. Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., . . . & Herrera, F. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, 58, 82–115.
2. Bochkovskiy, A., Wang, C. Y., & Liao, H. Y. M. (2020). Yolov4: Optimal speed and accuracy of object detection. *arXiv preprint arXiv:2004.10934*.
3. Cai, X., Chen, L., Xing, J., Xing, X., Luo, R., Tan, S., & Wang, J. (2021). Automatic Extraction of Layover From InSAR Imagery Based on Multilayer Feature Fusion Attention Mechanism. *IEEE Geoscience and Remote Sensing Letters*, 19, 1–5.
4. Chen, L., Luo, R., Xing, J., Li, Z., Yuan, Z., & Cai, X. (2022). Geospatial transformer is what you need for aircraft detection in SAR Imagery. *IEEE Transactions on Geoscience and Remote Sensing*.
5. Chen, Y., & Bruzzone, L. (2021). Self-supervised sar-optical data fusion of sentinel-1/2 images. *IEEE Transactions on Geoscience and Remote Sensing*, 60, 1–11.
6. Cutrona, L. J. (1990). Synthetic aperture radar. *Radar handbook*, vol. 2, pp. 2333–2346.
7. Diao, W., Dou, F., Fu, K., & Sun, X. (2018, July). Aircraft detection in SAR images using saliency based location regression network. In *IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium* (pp. 2334–2337). IEEE.
8. Du, L., Li, L., Wei, D., & Mao, J. (2020a). Saliency-guided single shot multibox detector for target detection in sar images. *IEEE Transactions on Geoscience and Remote Sensing*, 58(5), 3366–3376.
9. Du, L., Zhang, R., & Wang, X. (2020b, May). Overview of two-stage object detection algorithms. In *Journal of Physics: Conference Series* (Vol. 1544, No. 1, p. 012033). IOP Publishing.
10. Guo, Q., Wang, H., & Xu, F. (2020). Scattering enhanced attention pyramid network for aircraft detection in SAR images. *IEEE Transactions on Geoscience and Remote Sensing*, 59(9), 7570–7587.
11. Han, Z., Zhang, H., Zhang, J., & Hu, X. (2017, September). Fast aircraft detection based on region locating network in large-scale remote sensing images. In *2017 IEEE International Conference on Image Processing (ICIP)* (pp. 2294–2298). IEEE.
12. Jiang, P., Ergu, D., Liu, F., Cai, Y., & Ma, B. (2022). A Review of Yolo Algorithm Developments. *Procedia Computer Science*, 199, 1066–1073.
13. Kamath, V., & Renuka, A. (2021, December). Performance Analysis of the Pretrained EfficientDet for Real-time Object Detection on Raspberry Pi. In *2021 International Conference on Circuits, Controls and Communications (CCUBE)* (pp. 1–6). IEEE.
14. Lattari, F., Gonzalez Leon, B., Asaro, F., Rucci, A., Prati, C., & Matteucci, M. (2019). Deep learning for SAR image despeckling. *Remote Sensing*, 11(13), 1532.
15. Li, M., Wen, G., Huang, X., Li, K., & Lin, S. (2021). A Lightweight Detection Model for SAR Aircraft in a Complex Environment. *Remote Sensing*, 13(24), 5020.
16. Liu, S., Qi, L., Qin, H., Shi, J., & Jia, J. (2018). Path Aggregation Network for Instance Segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 8759–8768).
17. Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C. Y., & Berg, A. C. (2016, October). Ssd: Single shot multibox detector. In *European conference on computer vision* (pp. 21–37). Springer, Cham.
18. Luo, R., Chen, L., Xing, J., Yuan, Z., Tan, S., Cai, X., & Wang, J. (2021a). A fast aircraft detection method for SAR images based on efficient bidirectional path aggregated attention network. *Remote Sensing*, 13(15), 2940.
19. Luo, R., Xing, J., Chen, L., Pan, Z., Cai, X., Li, Z., . . . & Ford, A. (2021b). Glassboxing Deep Learning to Enhance Aircraft Detection from SAR Imagery. *Remote Sensing*, 13(18), 3650.

20. Ma, L., Liu, Y., Zhang, X., Ye, Y., Yin, G., & Johnson, B. A. (2019). Deep learning in remote sensing applications: A meta-analysis and review. *ISPRS journal of photogrammetry and remote sensing*, 152, 166–177.
21. Masita, K. L., Hasan, A. N., & Shongwe, T. (2020, August). Deep learning in object detection: A review. In *2020 International Conference on Artificial Intelligence, Big Data, Computing and Data Communication Systems (icABCD)* (pp. 1–11). IEEE.
22. Neubeck, A., & Van Gool, L. (2006, August). Efficient non-maximum suppression. In *18th International Conference on Pattern Recognition (ICPR'06)* (Vol. 3, pp. 850–855). IEEE.
23. Qian, G. U. O., Haipeng, W. A. N. G., & Feng, X. U. (2020). Research progress on aircraft detection and recognition in SAR imagery. *Lei Da Xue Bao*, 9(3), 497–513.
24. Qifang, X., Guoqing, Y., & Pin, L. (2018, February). Aircraft detection of high-resolution remote sensing image based on faster r-cnn model and ssd model. In *Proceedings of the 2018 International Conference on Image and Graphics Processing* (pp. 133–137).
25. Redmon, J., & Farhadi, A. (2017). YOLO9000: better, faster, stronger. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 7263–7271).
26. Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 779–788).
27. Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster r-cnn: Towards real-time object detection with region proposal networks. *Advances in neural information processing systems*, 28.
28. Rothe, R., Guillaumin, M., & Gool, L. V. (2014, November). Non-maximum suppression for object detection by passing messages between windows. In *Asian conference on computer vision* (pp. 290–306). Springer, Cham.
29. Siyu, W., Xin, G., Hao, S., Xinwei, Z., & Xian, S. (2017). An aircraft detection method based on convolutional neural networks in high-resolution SAR images. *雷达学报*, 6(2), 195–203.
30. Song, S., Jing, J., Huang, Y., & Shi, M. (2021). EfficientDet for fabric defect detection based on edge computing. *Journal of Engineered Fibers and Fabrics*, 16, 15589250211008346.
31. Tan, M., & Le, Q. (2019, May). Efficientnet: Rethinking model scaling for convolutional neural networks. In *International conference on machine learning* (pp. 6105–6114). PMLR.
32. Tan, M. Pang, R. Le, Q. V. (2020a). EfficientDet: Scalable and Efficient Object Detection. In *Proceedings of the IEEE/CVF Computer Vision Pattern and Recognition*, Seattle, WA, USA, 13–19 June 2020, pp. 10778–10787.
33. Tan, S., Chen, L., Pan, Z., Xing, J., Li, Z., & Yuan, Z. (2020b). Geospatial contextual attention mechanism for automatic and fast airport detection in SAR imagery. *IEEE Access*, 8, 173627–173640.
34. Tan, Y., Li, Q., Li, Y., & Tian, J. (2015). Aircraft detection in high-resolution SAR images based on a gradient textural saliency map. *Sensors*, 15(9), 23071–23094.
35. Ultralytics. (2022). YOLOV5. Available online: <https://github.com/ultralytics/yolov5> (accessed on 8 May 2022).
36. Wang, C. Y., Bochkovskiy, A., & Liao, H. Y. M. (2021a). Scaled-yolov4: Scaling cross stage partial network. In *Proceedings of the IEEE/cvf conference on computer vision and pattern recognition* (pp. 13029–13038).
37. Wang, J., Xiao, H., Chen, L., Xing, J., Pan, Z., Luo, R., & Cai, X. (2021b). Integrating weighted feature fusion and the spatial attention module with convolutional neural networks for automatic aircraft detection from SAR images. *Remote Sensing*, 13(5), 910.
38. Wang, Y., Wang, T., Zhou, X., Cai, W., Liu, R., Huang, M., . . . & Zhu, Y. (2022). TransEffiDet: Aircraft Detection and Classification in Aerial Images Based on EfficientDet and Transformer. *Computational Intelligence and Neuroscience*, 2022.
39. Wang, Z., Du, L., Mao, J., Liu, B., & Yang, D. (2018). Sar target detection based on SSD with data augmentation and transfer learning. *IEEE Geoscience and Remote Sensing Letters*, pp. 1–5.
40. Woo, S., Park, J., Lee, J. Y., & Kweon, I. S. (2018). Cbam: Convolutional block attention module. In *Proceedings of the European conference on computer vision (ECCV)* (pp. 3–19).

41. Xing, J., & Sieber, R. (2021). Integrating XAI and GeoAI. *GIScience 2021*, September 27–30, 2021, Poznań, Poland.
42. Xing, J., Sieber, R., & Kalacska, M. (2014). The challenges of image segmentation in big remotely sensed imagery data. *Annals of GIS*, 20(4), 233–244.
43. Zaidi, S. S. A., Ansari, M. S., Aslam, A., Kanwal, N., Asghar, M., & Lee, B. (2022). A survey of modern deep learning based object detection models. *Digital Signal Processing*, 103514.
44. Zhao, Y., Zhao, L., Li, C., & Kuang, G. (2020). Pyramid attention dilated network for aircraft detection in SAR images. *IEEE Geoscience and Remote Sensing Letters*, 18(4), 662–666.
45. Zheng, J., Yang, T., Liu, H., Su, T., & Wan, L. (2020). Accurate detection and localization of unmanned aerial vehicle swarms-enabled mobile edge computing system. *IEEE Transactions on Industrial Informatics*, 17(7), 5059–5067.
46. Zhu, X. X., Tuia, D., Mou, L., Xia, G. S., Zhang, L., Xu, F., & Fraundorfer, F. (2017). Deep learning in remote sensing: A comprehensive review and list of resources. *IEEE Geoscience and Remote Sensing Magazine*, 5(4), 8–36.

Machine Learning Methods for SAR Interference Mitigation



Yan Huang, Lei Zhang, Jie Li, Mingliang Tao, Zhanye Chen, and Wei Hong

1 Introduction

At the beginning, the background and motivation of this chapter would be briefly introduced in this section. Specifically, the congested electromagnetic environment and the adverse impacts of interferences to synthetic aperture radar (SAR) systems would be the focus in the following.

Y. Huang (✉) · W. Hong

Southeast University, Nanjing, China

e-mail: yellowstone0636@hotmail.com; weihong@seu.edu.cn

L. Zhang

Sun Yat-san University, Guangzhou, China

e-mail: zhanglei57@mail.sysu.edu.cn

J. Li

Nanjing University of Aeronautics and Astronautics, Nanjing, China

e-mail: lijie_evelyn@nuaa.edu.cn

M. Tao

Northwestern Polytechnical University, Xi'an, China

e-mail: mltao@nwpu.edu.cn

Z. Chen

Chongqing University, Chongqing, China

e-mail: czy@cqu.edu.cn

1.1 Congested Electromagnetic Environment

With the explosive emergence of multiple electromagnetic devices, a demand for a larger bandwidth is growing, e.g., the next generation communication networks require up to 1 GHz bandwidth of additional spectrum to transmit ultra-high definition visual communications and immersive multimedia interactions. Radar systems also require a larger bandwidth due to the requirement of a high range resolution. The growing reliance on unmanned platforms ranging from underwater sensors to satellites, which make the electromagnetic environment become more congested and complicated, also acting as a push for contested access [1].

SAR is an active wideband radar system, and it can generate high-resolution radar images day and night without influences of bad weathers. Nowadays, an SAR system, which becomes one of the most popular radar systems, becomes an important tool for a wide range of scientific, commercial, and defense applications [2]. As a wideband radar system, SAR often conflicts with other systems, such as radio systems, communication systems, or other active electromagnetic devices. These devices may have strong power and dominate the same frequency bands that SAR systems work at [3]. These electromagnetic signals are regarded as the suppressive interference to radar systems. Basically, the suppressive interferences can be categorized into two classes, the narrowband interference (NBI) and the wideband interference (WBI), simply based on the frequency bandwidth of interferences.

1.2 Adverse Impacts of Interference to SAR Systems

The presence of suppressive interferences is detrimental to global and regional scientific research activities of SAR systems. Depending on the application, SAR extracts information based on the characteristics of the echo including amplitude, frequency, time delay, polarization, Doppler shift, and phase [4]. The suppressive interference is able to corrupt the desired radar measurements in a number of ways, ranging from raw data collection, image formation, and the subsequent interpretation process.

1.2.1 Data Collection Process

For the SAR raw echoes, the presence of interferences may severely reduce the signal-to-interference-and-noise ratio (SINR), submerge the response of weak scatterers, and distort the dynamic range of raw echoes. In addition, the receiver is sensitive to saturation by high-power in-band emissions, especially when aimed toward the interfering sources in the main lobe. Figure 1 compares particular range spectra of the measured SAR data with and without radio frequency interferences (RFIs). The data was recorded by an L-band airborne SAR system with bandwidth

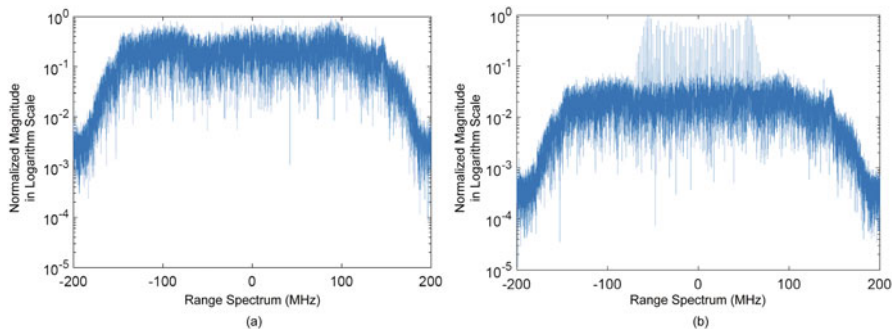


Fig. 1 Range frequency spectra of SAR echoes. (a) RFI-free frequency spectrum. (b) RFI-polluted frequency spectrum

being 300 MHz. The SAR data is linear frequency modulated, which is shown in Fig. 1a. As a special type of the NBI, the narrowband RFIs, which significantly reduce the SINR and alter the spectrum shape, occupy several narrow frequency bands in the middle of Fig. 1b.

1.2.2 Image Formation Process

An important feature of SAR systems is the capability to generate high-resolution images. Airborne SAR platforms would introduce significant trajectory deviation and motion errors due to air turbulence. Due to the absence of enough precise inertial navigation system data, it is required to estimate several critical matched filter parameters directly from the raw data, such as Doppler centroid and Doppler modulation rate [5]. However, the interference may severely affect the estimated accuracy of these parameters. In Fig. 2, the measured X-band SAR data is used to show the estimated errors of the Doppler centroid versus different input SINRs. In addition, the Doppler chirp rates are also estimated with large errors, and the SAR image quality would be severely degraded. Figure 3 illustrates the impulse response of a focused point target in real X-band SAR data with simulated interferences. As can be seen, the red dotted line denotes the interference-polluted result, and the point target is submerged by the strong interferences. The estimated Doppler centroid is obviously wrong so that the final imaging results are affected.

1.2.3 Image Interpretation Process

The ultimate goal of SAR imaging is to provide a physical understanding of the illuminated area. The existence of interferences may lead to the amplitude and phase distortion of the imaging data. Intuitively, strong interferences would generate noise-like covers on the interesting area, which leads to inaccurate spatial and radiometric

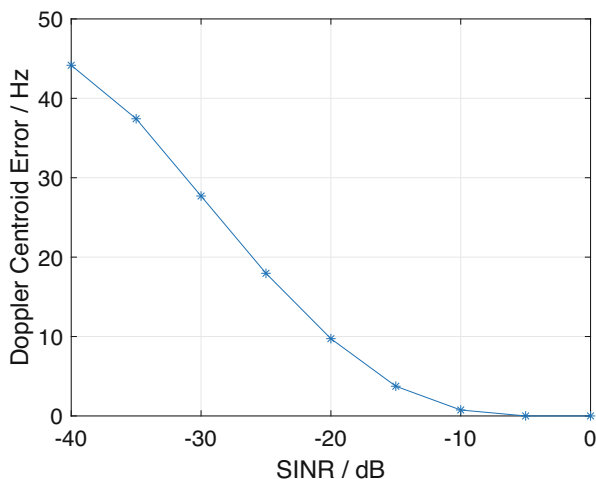


Fig. 2 Estimated error curve of the Doppler centroid versus input SINRs

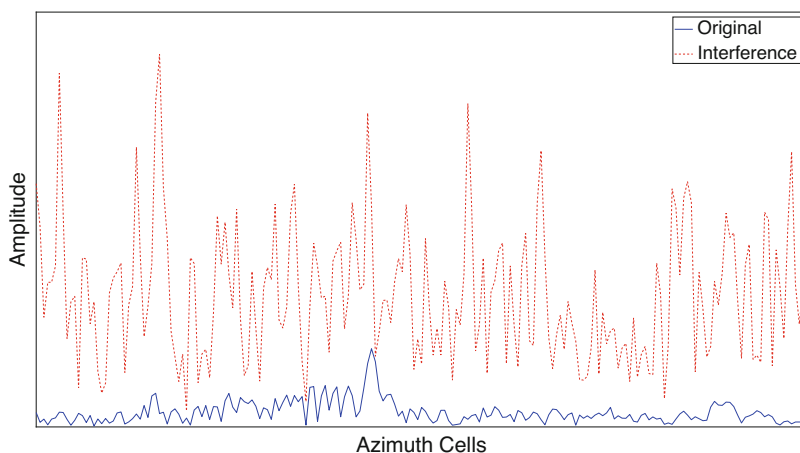


Fig. 3 The impulse response of a focused point target with or without interferences. The “Original” denotes the interference-free result. The “Interference” denotes the interference-polluted result

measurements [6]. As a result, the resulting phase distortion would de-correlate the data, producing inaccurate post-products, such as polarimetric descriptor [7], coherence [8], and retrieved biological or physical parameters [9]. Consequently, the interference hinders the subsequent image interpretation process like target detection and target recognition, and so on.

Figure 4 presents an example of National Aeronautics and Space Administration (NASA) Uninhabited Aerial Vehicle Synthetic Aperture Radar (UAV-SAR) Hawaii dataset. Obvious RFI stripes are observed, and the patterns of the artifacts are

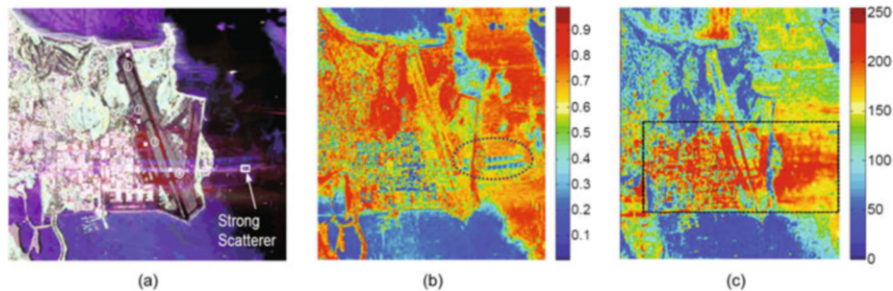


Fig. 4 Illustration of the amplitude and phase distortion affected by strong RFIs for measured NASA UAV-SAR data

very unique, which are not a duplicate of nearby scatterers or areas with strong reflectivity. The patterns are not comparable with the general radiometric artifacts such as the ambiguities, saturation effects, or processing effects.

A category of land cover classification scheme is based on the physical scattered characteristics. The presence of interference would result in estimation error of decomposition parameters and subsequently lead to wrong classification results when using these incorrect parameters. Figure 4b,c shows the entropy and anisotropy parameters obtained by Cloud-Pottier polarimetric decomposition. The regions in dashed lines highlight the anomalies introduced by interference artifacts. Obvious difference can be viewed by naked eyes compared with the nearby regions.

1.3 Notations

Before introducing the main body of this chapter, we denote tensors by Euler script letters, e.g., \mathcal{A} ; matrices are denoted by boldface capital letters, e.g., \mathbf{A} ; vectors are denoted by boldface lowercase letters, e.g., \mathbf{a} ; and scalars are denoted by the lowercase letters, e.g., a . For a 3-way tensor $\mathcal{A} \in \mathbb{C}^{N_1 \times N_2 \times N_3}$, we denote its (n_1, n_2, n_3) -th entry as $\mathcal{A}_{n_1 n_2 n_3}$ or $a_{n_1 n_2 n_3}$ and use the Matlab notation $\mathcal{A}(j, :, :)$, $\mathcal{A}(:, j, :)$, and $\mathcal{A}(:, :, j)$ to denote the j -th horizontal, lateral, and frontal slices, respectively. More often, the frontal slice $\mathcal{A}(:, :, j)$ is denoted compactly as $\mathcal{A}^{(j)}$. For $\mathcal{A} \in \mathbb{C}^{n_1 \times n_2 \times n_3}$, and we denote $\tilde{\mathcal{A}}$ as the result of discrete Fourier transformation of \mathcal{A} along the third dimension by using the Matlab “fft” command.

1.4 Summary of the Remainder of the Chapter

The rest of the contents of this chapter are organized as follows: in Sect. 2, the preliminary interference mitigation strategies are reviewed from two aspects,

interference detection and interference suppression. In Sect. 3, the machine learning methods for interference suppression are specifically analyzed with focusing on different kinds of interferences and multiple categories of machine learning methods. Finally, future trend of interference mitigation is presented in Sect. 4 to analyze the feasibility of deep learning and cognitive scheme.

2 Interference Mitigation Strategies for SAR System

2.1 Interference Detection

The interference detection problem can be modeled as a binary hypothesis test [4]. The hypothesis can be expressed as

$$H_0 : y(t) = x(t) + n(t), \quad 1 \leq t \leq T; \quad (1)$$

$$H_1 : y(t) = x(t) + i(t) + n(t), \quad 1 \leq t \leq T, \quad (2)$$

where y , x , i , and n denote received signals, real echoes, interferences, and noises in time domain, respectively. Furthermore, the variable t denotes sampled time and T is the total time. Under the null hypothesis H_0 , the instantaneous spectrum obeys Gaussian-like distribution, i.e., the amplitude obeys Rayleigh distribution, and would have a negentropy value close to 0. Under the alternate hypothesis H_1 , the instantaneous spectrum is a kind of non-Gaussian distributed with a large negentropy value. The obvious distribution difference, as well as the negentropy value difference, is the fundamental principle for correct detections. If the negentropy threshold is too high, the interference may not be completely detected, thus causing the missing alarm. On the other hand, a low negentropy threshold may lead to the false detection of some naturally occurring peaks, which is referred to as the false alarm. The detection threshold should be finely set to achieve a trade-off between detection rate and false alarm. Herein, it is calculated based on the Neyman–Pearson decision rule, which maximizes the probability of detection P_d under the constraint of the probability of false alarm P_f [10], i.e.,

$$\text{th}^* = \arg \max P_d, \quad \text{s.t.} \quad P_f \leq \alpha, \quad (3)$$

where th^* is the optimal threshold and α is the tolerable false alarm level. Typically, α is higher than 10^{-3} and commonly is in the range of 10^{-6} and 10^{-8} . Then, the optimal threshold can be determined in terms of the false alarm level [10]

$$\text{th}^* = \mu_0 + \sqrt{2}\sigma_0 \text{erf}^{-1}(1 - 2\alpha), \quad (4)$$

where μ_0 and σ_0 are the mean and the standard deviation of interference-free instantaneous spectra, and $\text{erf}^{-1}(\cdot)$ is the inverse error function. Instead of a constant threshold, the adaptive detection threshold is leveraged based on the statistics of the interference-free instantaneous spectra.

2.2 Overall Introduction of Interference Suppression Methods

In this chapter, we mainly focus on the suppressive interference for SAR systems. In the past decades, many researchers have focused on how to mitigate interferences, including NBIs and WBIs, on SAR systems effectively and efficiently. These interference suppression methods can be classified into three categories: nonparametric method, parametric method, and semi-parametric method. Actually, the semi-parametric methods, which are the focus in this chapter, belong to the group of machine learning methods.

2.2.1 Nonparametric Methods

In terms of the nonparametric methods, this kind of method inspires a basic idea for suppressing strong interferences in SAR received signals. Nonparametric methods were applied to either NBI suppression problem or WBI suppression problem via designed filters and constructed subspace projector. Zhou et al. [11] proposed an eigenvalue subspace projection (ESP) method to construct the NBI and signal subspaces via singular value decomposition (SVD) and to project the received signal onto the signal subspace for realizing NBI suppression. Lord [12] designed a notched filter, which generates zero notches at the strong NBI frequency spectrum bands. This kind of notched filter is feasible for the NBI case, while it usually suffers from wide frequency notches when dealing with WBIs. Le et al. [13] designed a least mean square (LMS) filter to adaptively mitigate NBIs for single-channel SAR systems. Unlike the NBI version of ESP method, Tao et al. [14] proposed an enhanced version of ESP method, which works well on WBI suppression problem. It utilized the short-time Fourier transform (STFT) to transform one-pulse signal into two-dimensional (2D) time-frequency domain, and the ESP method was used on the time-frequency signal.

2.2.2 Parametric Methods

Different from nonparametric methods, parametric methods have been widely used to estimate the interference parameters. Rather than WBI suppression problem, parametric methods are feasible to mitigate NBIs because the NBI is easily modeled as a sinusoidal signal. Commonly, the sinusoidal model has two parameters to estimate, i.e., amplitude and frequency. Liu et al. [15] employed the iterative adaptive

approach (IAA) to iteratively estimate the time-varying frequencies of NBIs. Huang et al. [16] proposed a gradual RELAX algorithm to estimate the parameters for each frequency peak with a reduced number of estimation iterations of the traditional RELAX algorithm. However, these methods are computationally expensive due to thousands of snapshots and hundreds of NBIs in one snapshot. In [17], a CLEAN algorithm was employed to estimate the parameters without iterations, but it is still computationally expensive when facing hundreds of interferences and thousands of azimuth samples (snapshots/pulses). With regard to the WBI suppression, the parametric method can only tackle with isolated WBI case, i.e., the WBI can be at least widely separated in 2D time-frequency domain. As mentioned above, Yang et al. [18] employed the IAA to estimate the frequencies of isolated WBIs. However, the interference environment is commonly complicated so that parametric methods may not be feasible for the practical application. Moreover, the NBIs and WBIs can be represented by these parametric models, but the real echoes do not match these models. Hence, it is not precise enough to employ these parametric models without any protection for the real echoes.

2.2.3 Semi-parametric Methods

In recent years, the machine learning methods became popular for separating or classifying different signals in a mixed environment. These methods can be classified into two mainstreams, supervised learning and unsupervised learning methods. In this chapter, unsupervised learning methods are emphasized because we do not have large amounts of interference data for supervised learning. Unsupervised learning methods, such as principal component analysis (PCA) and independent component analysis (ICA), aim to optimize an objection function under exact constraints. These optimization problems usually have several hyperparameters, which should be finely tuned for realizing good performance. Therefore, these methods are regarded as semi-parametric methods.

For the interference suppression problem, we can extract interferences as much as possible and simultaneously protect the real echo with a regularization term. Joy et al. [19] employed the inexact augmented Lagrange method (IALM) to extract low-rank NBI matrix by solving the traditional robust principal component analysis (RPCA) problem. Su et al. [20] employed the IALM in the 2D time-frequency domain for more precise extraction. Based on the above two works, in [21], the row sparse constraint is leveraged to simultaneously restrict the special kind of NBIs, i.e., radio frequency interference (RFI). Huang et al. also developed the reweighted schemes and matrix factorization techniques to deal with the drawbacks of nuclear norm minimization problem and precisely suppress the RFI [22, 23]. A block sparse Bayesian learning algorithm was proposed to mitigate RFI effectively [24]. In addition, the tensor structures are analyzed in detail to extend the 2D interference mitigation methods to higher dimensional solution space [25, 26]. The regularization term determines the performance of real echo protection, and hence, a dictionary-based regularization was proposed to constrain the sparsity of real echoes

in the image domain [27]. Furthermore, simultaneous NBI and WBI mitigation problem was fully considered via a sparse recovery method in [28]. But in practical applications, the electromagnetic environment is commonly complicated so that the interferences are densely distributed in 2D time-frequency domain. Then, the low-rank tensor recovery and tensor factorization methods were employed to deal with complicated interferences [29, 30]. In summary, utilizing the regularization term is the key advantage of semi-parametric methods, and it can protect the strong real echoes in each iteration.

3 Machine Learning Methods for Interference Suppression

As an effective tool, machine learning methods have attracted worldwide attention for the past decades on tremendous practical applications of different disciplines, such as image recognition [31], speech recognition [32], and self-driving cars [33]. Most machine learning methods actually belong to the group of semi-parametric methods except for the principal component analysis (PCA) method, so we would analyze the specific characteristics of machine learning methods in this section.

3.1 General Signal Model of Typical Interferences for SAR systems

For a classic wideband SAR system, the received echoes can be severely interfered with other electromagnetic systems that are working at the same frequency bands. Commonly, the received signal of SAR systems can be formulated as

$$\begin{aligned} y(t) &= x(t) + i(t) + n(t) \\ &= x(t) + nbi(t) + wbi(t) + n(t), \end{aligned} \quad (5)$$

where nbi and wbi denote the narrowband and wideband interferences, respectively. Since SAR systems aim to generate 2D high-resolution images, the 2D sampled signal can be formulated as

$$\begin{aligned} \mathbf{Y}(k, l) &= \mathbf{X}(k, l) + \mathbf{I}(k, l) + \mathbf{N}(k, l) \\ &= \mathbf{X}(k, l) + \mathbf{NBI}(k, l) + \mathbf{WBI}(k, l) + \mathbf{N}(k, l), \end{aligned} \quad (6)$$

where k denotes the k -th range time sample and l denotes the l -th azimuth time sample. Herein, the capital letters are corresponding to the lowercase ones in (5). Next, the NBI and WBI are specifically introduced in the following subsections.

3.1.1 Narrowband Interference

The NBI often possesses relatively narrow bandwidth than that of an SAR system. Commonly, the bandwidth ratio between the NBI and SAR signals is less than 1% [34]. This kind of interfering signal model is well studied in early researches, and typical NBI example is the narrowband RFI. In the mathematical representation, it can be modeled as a sum of complex sinusoidal tones, that is,

$$\mathbf{NBI}(k, l) = \sum_{n=1}^N A_n(l) \exp(j2\pi f_n t_k), \quad (7)$$

where $j = \sqrt{-1}$ denotes the imaginary symbol, $A_n(l)$ denotes the complex amplitude of the n -th NBI in the l -th pulse, f_n denotes the frequency of the n -th NBI, t_k denotes the k -th range sampling time, and N is the number of NBI sources.

Figure 5 illustrates the range frequency spectrum of radar echoes contaminated with NBIs. The data is recorded by an X-band airborne SAR system, transmitting a linear frequency modulated pulse with a bandwidth of 200 MHz. The strong interferences dominate a few frequency bands and appear as prominent “sin” peaks in the range frequency domain of one pulse (Fig. 5, left) and bright lines in the range frequency azimuth-time representation (Fig. 5, right), respectively. This feature makes it easier to be identified and detected.

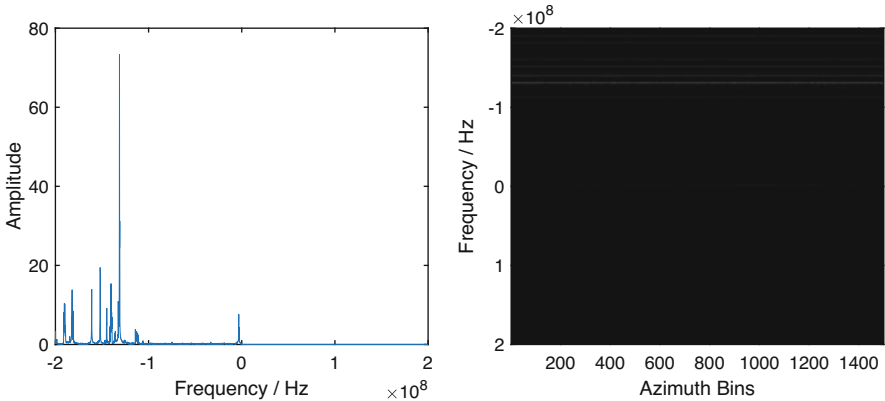


Fig. 5 Illustration of the range frequency spectrum of NBI sources

3.2 Wideband Interference

Wideband interferences are pulsed in time with various pulse repetition time, and the bandwidth is wider than the continuous narrowband. According to the modulation type, the WBIs can be modeled by two classes, i.e., the chirp modulated (CM) WBI and the sinusoidal modulated (SM) WBI. In a practical situation, the WBI may not match these two special models perfectly. However, the CM WBI and SM WBI can be considered as two special and examples of a larger family of WBI signatures whose properties are a combination of them. Mathematically, the CM WBI is modeled as

$$\mathbf{WBI}_{CM}(k, l) = \sum_{n=1}^N B_n(l) \exp\left(j2\pi f_n t_k + j\pi \gamma_n t_k^2\right), \quad (8)$$

where $B_n(l)$ denotes the amplitude of the n -th CM WBI in the l -th snapshot, and γ_n and f_n denote the chirp rate and frequency of the n -th CM WBI, respectively. Next, the SM WBI can be modeled as

$$\mathbf{WBI}_{SM}(k, l) = \sum_{n=1}^N B_n(l) \exp\{j\beta_n(l) \sin(2\pi f_n t_k + \phi_n)\}, \quad (9)$$

where $B_n(l)$ and $\beta_n(l)$ denote the amplitude and modulation index of the n -th SM WBI in the l -th snapshot, respectively, and f_n and ϕ_n denote the frequency and the initial phase of the n -th SM WBI, respectively. WBIs, which occupy a wide range of frequency bands in the range frequency spectrum, cannot be recognized intuitively in 2D time domain. Figure 6 shows a particular range frequency spectrum of radar echoes with one CM WBI and one SM WBI. This data is acquired from another X-band airborne SAR system and the sampling frequency is 60 MHz. It is shown that the target echo is contaminated with WBIs. It is hard to recognize the useful signal (i.e., the real echoes) from the corrupted signals.

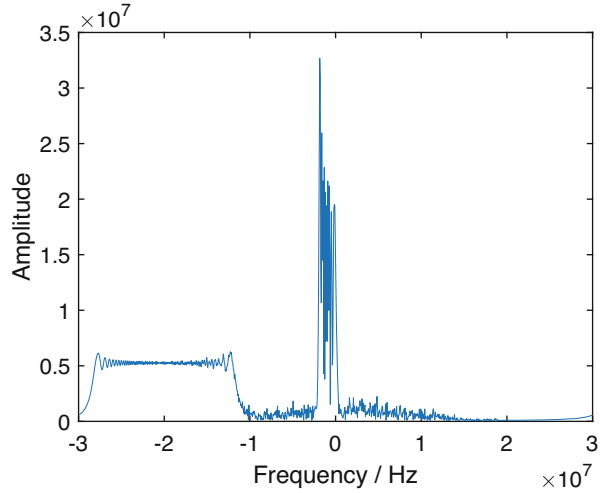
With the help of time–frequency analysis tools, such as the short-time Fourier transform (STFT) [35], the WBIs can be transformed into the time–frequency domain, which can be formulated as

$$\text{STFT}_y(t, f) = \int_{-\infty}^{\infty} y(t) h(t - \tau) e^{-j2\pi f \tau} d\tau, \quad (10)$$

where (t, f) indicates the coefficient in the time–frequency domain. Similarly, the inverse STFT (ISTFT) is performed with the synthesis window $w(t)$ as

$$y(t) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \text{STFT}_y(t', f') w(t - t') e^{j2\pi f' t'} dt' df'. \quad (11)$$

Fig. 6 Illustration of the range frequency spectrum of WBI sources



Then, the time–frequency spectra of both kinds of WBIs are shown in Fig. 7. It is easy to recognize the characteristic of the interferences along with the frequency and time axis, which demonstrates that the WBI is highly nonstationary in time or frequency domain. The time–frequency spectrum characterizes the nonstationary property obviously, in which the power of WBIs is concentrated into a few time–frequency bins in 2D time–frequency domain.

3.3 Interference Mitigation Schemes Based on Machine Learning

In this section, the machine learning techniques for solving NBI and WBI mitigation problem will be analyzed in detail. The first machine learning method emerged in previous researches is the PCA method.

3.3.1 NBI Mitigation

Principal Component Analysis

Based on the signal model in Eq. (5), the NBI can be easily observed in a few frequency bands of the frequency spectrum. The eigenvalue subspace projection (ESP) method (i.e., the PCA method) was firstly employed in [11] for NBI mitigation problem. It used the eigenvalue decomposition (EVD) method to construct the signal subspace and projected the received signal onto the signal subspace for realizing NBI suppression. Actually, the ESP method is equivalent to the classic

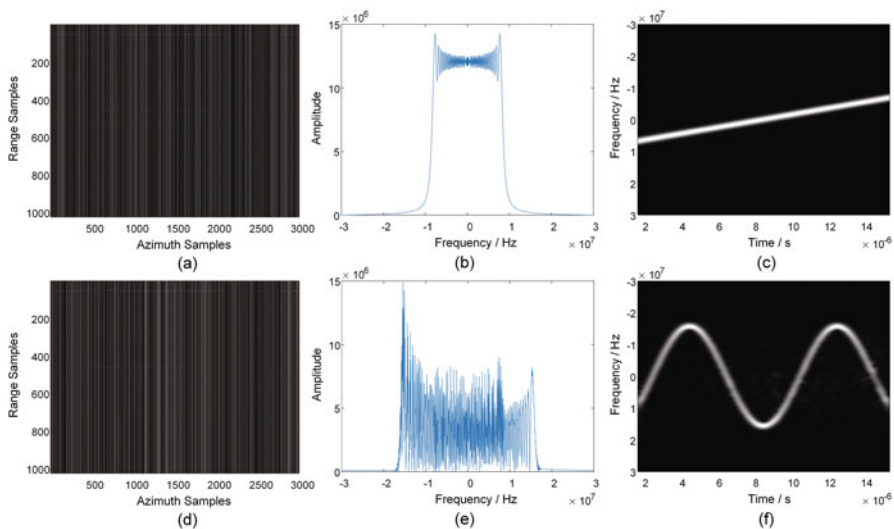


Fig. 7 Illustration of CM WBI and SM WBI in the 2D range-azimuth time domain, 1D range frequency domain, and 2D range time–frequency domain. Top row: CM WBI. Bottom row: SM WBI. Left column: raw data in 2D time domain. Middle column: range frequency spectrum. Right column: 2D range time–frequency spectrum

PCA method. The PCA method is easy to implement, and its mathematical problem can be formulated as

$$\begin{aligned} \max_{\mathbf{W}} \quad & \text{tr}(\mathbf{W}^H \mathbf{X} \mathbf{X}^H \mathbf{W}) \\ \text{s.t.} \quad & \mathbf{W}^H \mathbf{W} = \mathbf{E}, \end{aligned} \quad (12)$$

where \mathbf{E} is the identity matrix. The solution of the above PCA problem is the matrix whose columns are eigenvectors corresponding to the d largest eigenvalues when the $\mathbf{W} \in \mathbb{C}^{K \times d}$. The PCA method can be classified into the nonparametric method group. It has better performance than the notched filter, which is the most famous nonparametric method. The range frequency spectra of PCA and notched filter methods are compared in Fig. 8. As can be seen, the notched filter simply makes the dominated frequency peak area be null which does not consider the useful signal in these frequency bands. The PCA method would reduce the power of strong spectrum peaks, but it would not remove them thoroughly.

Sparse Recovery

In recent years, due to the sparsely distributed NBI in the whole frequency spectrum, the sparse recovery (SR) method, which is alternately called sparse coding in the machine learning domain, was used to mitigate NBIs by solving the following

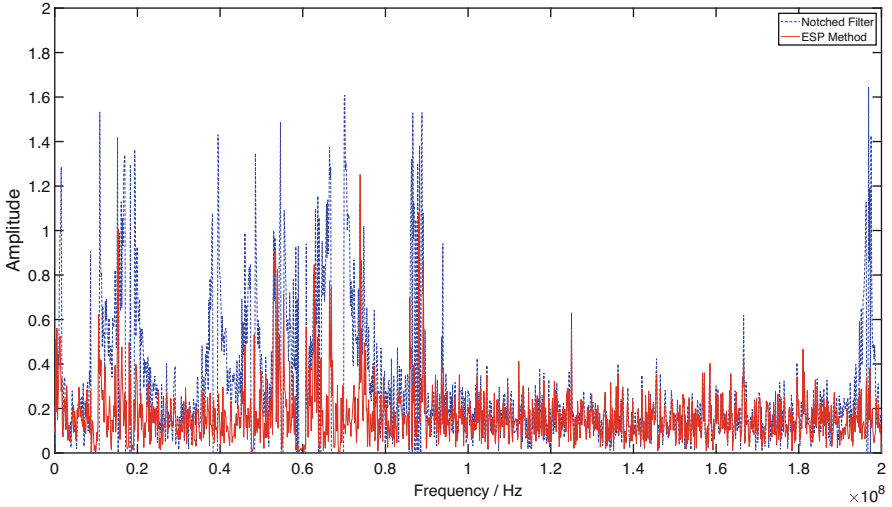


Fig. 8 Comparison of PCA method (ESP method) and notched filter method

problem [36]:

$$\{\alpha'_l, \mathbf{e}'_l\} = \arg \min_{\alpha_l, \mathbf{e}_l} \left\| \mathbf{y}(l) - \mathbf{D}^{SAR} \alpha_l - \mathbf{D}^{NBI} \mathbf{e}_l \right\|_2 + \lambda \|\alpha_l\|_1 + \tau \|\mathbf{e}_l\|_1, \quad (13)$$

where $\mathbf{y}(l)$ denotes the received signal of the l -th pulse, $\mathbf{x}(l) = \mathbf{D}^{SAR} \alpha_l$ denotes the real echoes of the l -th pulse, \mathbf{D}^{SAR} denotes the sparse coding dictionary of real SAR echoes, α_l denotes the sparse coefficients, and similarly, $\mathbf{i}(l) = \mathbf{D}^{NBI} \mathbf{e}_l$ denotes the interference of the l -th pulse, \mathbf{D}^{NBI} denotes the sparse coding dictionary of the interference, and \mathbf{e}_l denotes the sparse coefficients. Herein, the range time is omitted for simplicity. With the alternating direction method of multipliers (ADMM) [37], the above problem could be effectively solved. However, the above method applies the sparse coding method for each pulse of received signal, so the computational burden is quite expensive when performing it on the large-scale SAR data. In addition, the accuracy of the dictionary determines the final performance.

Low-Rank Recovery

Although the ESP method, i.e., the PCA method, can remove the NBI effectively, the real echoes lack of protection during the mitigation. It is quite similar to the image denoising problem when tackling with the noises. Regarding the image denoising problem, the low-rank representation methods are popular to extract the low-rank section from the corrupted images [38]. Since the interference matrix was demonstrated to be low rank, the NBI mitigation problem can be formulated as the

following problem:

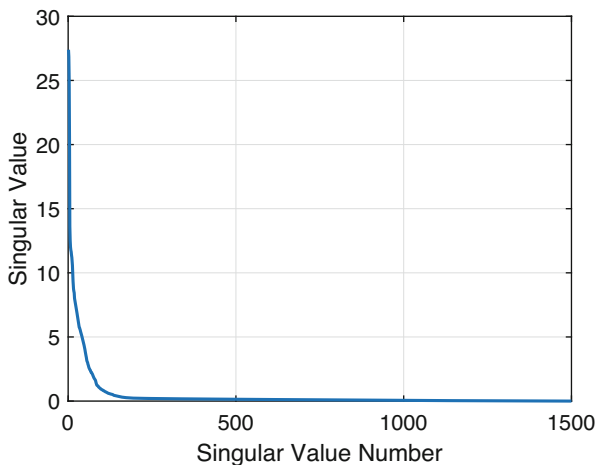
$$\begin{aligned} \min_{\mathbf{I}, \mathbf{X}} \quad & \text{rank}(\mathbf{I}) + \lambda \|\mathbf{X}\|_0 \\ \text{s.t.} \quad & \|\mathbf{Y} - \mathbf{I} - \mathbf{X}\|_F^2 < \delta, \end{aligned} \quad (14)$$

where “rank” denotes the rank operation, $\|\cdot\|_0$ is the ℓ_0 norm of matrix, and $\|\cdot\|_F$ denotes the Frobenius norm. Herein, λ and δ are hyperparameters. Basically, the above problem is NP-hard to solve. Commonly, a classic robust principal component analysis (RPCA) method [39] was employed to relax the above problem to a convex problem. The ℓ_1 regularization term was added to the original PCA problem to protect the real echoes from removing. The relaxed problem can be formulated as

$$\begin{aligned} \{\hat{\mathbf{I}}, \hat{\mathbf{X}}\} = \arg \min_{\mathbf{I}, \mathbf{X}} \quad & \|\mathbf{I}\|_* + \lambda \|\mathbf{X}\|_1 \\ \text{s.t.} \quad & \|\mathbf{Y} - \mathbf{I} - \mathbf{X}\|_F^2 \leq \delta, \end{aligned} \quad (15)$$

where $\|\cdot\|_*$ denotes the nuclear norm and $\|\cdot\|_1$ denotes the ℓ_1 norm of matrix. Herein, the nuclear norm and the ℓ_1 regularization term are convex relaxations of the rank operation and the ℓ_0 norm, respectively. The singular value curve of a measured NBI matrix is illustrated in Fig. 9. As can be seen, the NBI matrix can be approximately considered as a low-rank matrix. The measured X-band SAR data with strong NBIs is illustrated to test the performance of RPCA method, which is also named as the augmented Lagrangian method (ALM) in this case, in Fig. 10. The input SINR is -20 dB. As can be seen, the RPCA method can remove the interferences properly by using the low-rank property of NBIs and sparse regularization term for the useful signal protection.

Fig. 9 Singular value curve of the measured NBI matrix



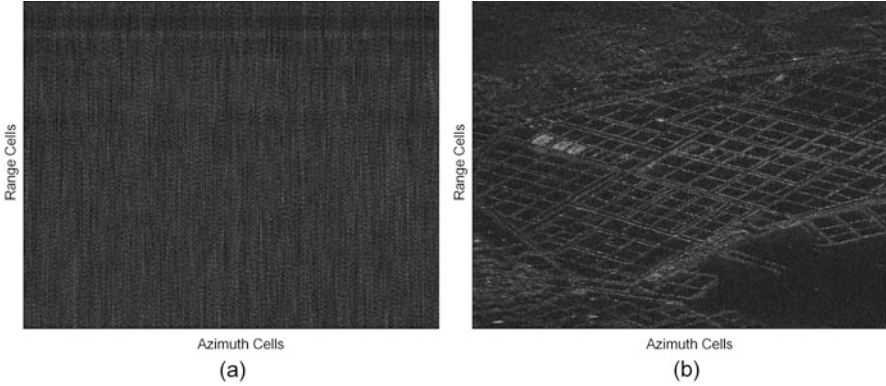


Fig. 10 Interference mitigation result of RPCA method. SINR = -20 dB. (a) NBI-polluted results. (b) Recovered result of RPCA method

Joint Sparsity and Low-Rank Recovery (JSLR)

In previous subsections, it is concluded that the NBI frequency spectrum has sparsity along the frequency axis and the NBI matrix has low-rank property in general. Figure 5 also illustrates the sparsity and low-rank property of NBIs. Then, with a simple consideration, rather than using either sparsity or low-rank property of the NBIs to recover the interference-free SAR echoes, a joint sparse and low-rank model is considered for 2D NBI mitigation problem as follows:

$$\begin{aligned} \min_{\mathbf{I}, \mathbf{X}} \quad & \text{rank}(\mathbf{I}) + \lambda_1 \|\mathbf{F}\mathbf{I}\|_0 + \alpha_1 \|\mathbf{X}\|_0 \\ \text{s.t.} \quad & \|\mathbf{Y} - \mathbf{I} - \mathbf{X}\|_F^2 < \delta, \end{aligned} \quad (16)$$

where \mathbf{F} denotes the Fourier transform matrix. Then, based on the signal model in (5), the interference matrix can be further formulated as

$$\mathbf{I} = \mathbf{D}_{NBI} \mathbf{I}', \quad (17)$$

where $\mathbf{D}_{NBI} \in \mathbb{C}^{K \times K'}$ is the overcomplete dictionary of NBI, and the matrix \mathbf{I}' is the sparse matrix with weighting coefficients of RFI based on the dictionary \mathbf{D}_{NBI} . Then, the JSLR optimization problem can be recast as

$$\begin{aligned} \min_{\mathbf{I}, \mathbf{X}} \quad & \text{rank}(\mathbf{D}_{NBI} \mathbf{I}') + \lambda_1 \|\mathbf{F} \mathbf{D}_{NBI} \mathbf{I}'\|_0 + \alpha_1 \|\mathbf{X}\|_0 \\ \text{s.t.} \quad & \|\mathbf{Y} - \mathbf{D}_{NBI} \mathbf{I}' - \mathbf{X}\|_F^2 < \delta. \end{aligned} \quad (18)$$

As the \mathbf{D}_{NBI} is the overcomplete dictionary, it is a full-rank matrix. Then, the optimization problem can be reformulated as

$$\begin{aligned} \min_{\mathbf{I}, \mathbf{X}} \quad & \text{rank}(\mathbf{I}') + \lambda \|\mathbf{I}'\|_0 + \alpha \|\mathbf{X}\|_0 \\ \text{s.t.} \quad & \|\mathbf{Y} - \mathbf{D}_{NBI} \mathbf{I}' - \mathbf{X}\|_F^2 < \delta. \end{aligned} \quad (19)$$

The above problem is the NP-hard problem, and similarly, we leverage the nuclear norm and the ℓ_1 norm to relax the rank operation and ℓ_0 norm, respectively, that is,

$$\begin{aligned} \min_{\mathbf{I}, \mathbf{X}} \quad & \|\mathbf{I}'\|_* + \lambda \|\mathbf{I}'\|_1 + \alpha \|\mathbf{X}\|_1 \\ \text{s.t.} \quad & \|\mathbf{Y} - \mathbf{D}_{NBI} \mathbf{I}' - \mathbf{X}\|_F^2 < \delta. \end{aligned} \quad (20)$$

This problem can be solved by the ADMM scheme based on the augmented Lagrangian function.

From another point of view, through a synthetic aperture period, the interferences are commonly stationary, which can be considered as a basis of the low-rank property. Recall that in the 2D range frequency and azimuth-time spectra of NBIs shown in Fig. 5, the NBIs performed like several straight lines. Actually, the reason why the NBI has both sparsity and the low-rank property is that it has relatively stationary frequencies crossing different pulses. Hence, the row sparsity, which matches the circumstance of NBIs, can be considered as a special case of joint sparsity and low-rank property. In order to formulate the row sparse problem, the optimization problem can be described as

$$\begin{aligned} \min_{\mathbf{I}, \mathbf{X}} \quad & \|\mathbf{FI}\|_0 + \beta_1 \|\mathbf{X}\|_0 \\ \text{s.t.} \quad & \|\mathbf{Y} - \mathbf{I} - \mathbf{X}\|_F^2 < \delta. \end{aligned} \quad (21)$$

With the similar formulation based on the dictionary, the row sparse problem can be recast as

$$\begin{aligned} \min_{\mathbf{I}', \mathbf{X}} \quad & \|\mathbf{FD}_{NBI} \mathbf{I}'\|_0 + \beta_1 \|\mathbf{X}\|_0 \\ \text{s.t.} \quad & \|\mathbf{Y} - \mathbf{D}_{NBI} \mathbf{I}' - \mathbf{X}\|_F^2 < \delta. \end{aligned} \quad (22)$$

And the similar relaxation is applied to the above problem, that is,

$$\begin{aligned} \min_{\mathbf{I}', \mathbf{X}} \quad & \|\mathbf{I}'\|_0 + \beta \|\mathbf{X}\|_0 \\ \text{s.t.} \quad & \|\mathbf{Y} - \mathbf{D}_{NBI} \mathbf{I}' - \mathbf{X}\|_F^2 < \delta. \end{aligned} \quad (23)$$

The row sparse problem does not involve the low-rank objective function so that the singular value decomposition (SVD) is not required when solving the row sparse problem. The computational burden can be further relieved. This problem can then be solved by alternately optimizing the interference and the real echoes. The input SINR is set as -20 dB, and then the interference mitigation results are given in Fig. 11. The interference-polluted image is the same as that in Fig. 11a. As can be

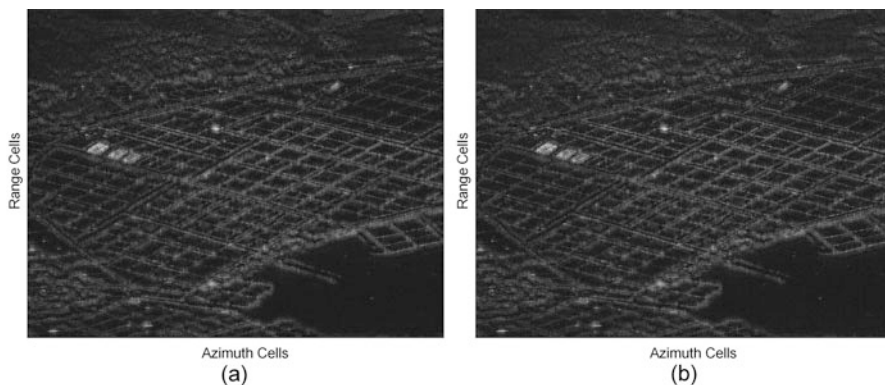


Fig. 11 Interference mitigation results of JSLR method and RS method. Input SINR is -20 dB. (a) JSLR method. Output SINR = 7.81 dB. (b) RS method. Output SINR = 8.24 dB

seen, both methods can effectively mitigate strong interferences and recover the real echoes in a high quality. Based on the workstation equipped with W-2133 CPU and 256 GB memory, the running times of RPCA method, JSLR method, and RS method are 39.36 s, 30.41 s, and 10.12 s, respectively. The RS method is the most efficient method among these methods since it avoids the SVD operation. The JSLR has heavier computational burden than the RPCA method per iteration, but it requires less iterations to realize convergence with shorter running time.

Matrix Factorization Techniques

The low-rank matrix recovery techniques, which are analyzed in previous subsections, have been widely used to mitigate NBIs in recent years. The previous signal model aims to use the nuclear norm to relax the non-convex rank operation. The convex relaxation, i.e., the nuclear norm, provides feasible solution and fast convergence. However, when solving the subproblem of nuclear norm minimization problem by the singular value thresholding (SVT) algorithm, it restricts all the singular values with a fixed threshold. This may excessively decrease the large singular values. Basically, the large singular values dominate the whole power of interferences, and the excessive decrease of large singular values may affect the recovered power of low-rank interference matrix. As can be seen in Fig. 12, the interference-free signal (marked by “Signal”) is much lower than the recovered result by using the nuclear norm minimization (marked by “ALM”), and this demonstrates that the SVT algorithm would over-reduce the large singular values of the interference matrix as well as the interference power. When mitigating the interferences from the received signal, the residual interference may be still too strong to submerge the real echoes. In addition, the nuclear norm minimization

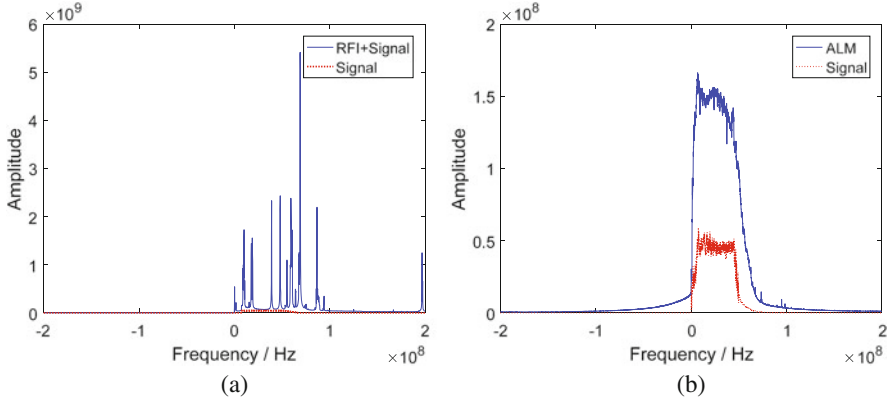


Fig. 12 Illustration of the interference mitigation results via nuclear norm minimization. **(a)** Range frequency spectra of interference-free and interference-polluted signals. **(b)** Range frequency spectra of interference-free and interference mitigation result after applying the nuclear norm minimization scheme

problem on large-scale observed matrix is time-consuming, which is a shortcoming for data processing.

Based on the non-convex low-rank recovery model in (14), a reweighted scheme is applied as follows:

$$\begin{aligned} \min_{\mathbf{I}, \mathbf{X}} \quad & \sum_{i=1}^{\min\{K, L\}} w_i \sigma_i(\mathbf{I}) + \lambda \|\mathbf{X}\|_1 \\ \text{s.t.} \quad & \|\mathbf{Y} - \mathbf{I} - \mathbf{X}\|_F^2 < \delta, \end{aligned} \quad (24)$$

where $\sigma_i(\mathbf{I})$ is the i -th singular value with its corresponding weight w_i , and the singular values should be sorted in a non-ascending order. Unlike the convex nuclear norm, the reweighted scheme is an approximation of the rank function. The above problem can be solved alternately by optimizing two subproblems as follows:

$$\min_{\mathbf{I}} \sum_{i=1}^{\min\{K, L\}} w_i \sigma_i(\mathbf{I}) + \frac{\mu}{2} \|\mathbf{Y} - \mathbf{I} - \mathbf{X}\|_F^2, \quad (25)$$

$$\min_{\mathbf{X}} \lambda \|\mathbf{X}\|_1 + \frac{\mu}{2} \|\mathbf{Y} - \mathbf{I} - \mathbf{X}\|_F^2. \quad (26)$$

The subproblem in (25) is a weighted nuclear norm minimization problem, whose closed-form solution can be formulated by the complex weighted singular value thresholding (CWSVT) algorithm [23]. The CWSVT algorithm reduces the large singular values with small thresholds while punishes the small singular values with large thresholds. Hence, it is able to relieve the influence of the SVT algorithm

when solving the nuclear norm minimization problem, but it still requires to perform SVD on the large-scale observed signal with high computational burden. The matrix factorization technique is applied to reduce the computational complexity. Suppose that $\sigma_i(\mathbf{I}) = \sigma_i(\mathbf{U})\sigma_i(\mathbf{V})$, and then the large-scale interference matrix is equivalent to the product of two small-scale matrices as follows:

$$\begin{aligned}\mathbf{I} &= \mathbf{U}_\mathbf{I} [\text{diag}(\sigma(\mathbf{I}))] \mathbf{V}_\mathbf{I}^\mathbf{H} \\ &= [\mathbf{U}_\mathbf{R} \text{diag}(\sigma(\mathbf{U}))] [\text{diag}(\sigma(\mathbf{V})) \mathbf{V}_\mathbf{I}^\mathbf{H}] = \mathbf{U}\mathbf{V}^\mathbf{H},\end{aligned}\quad (27)$$

where $\mathbf{U} \in \mathbb{C}^{K \times d}$, $\mathbf{V} \in \mathbb{C}^{L \times d}$, $r \leq d \ll \min\{K, L\}$, r is the rank of matrix \mathbf{I} , and $\sigma(\mathbf{U})$ and $\sigma(\mathbf{V})$ are the singular value vectors of \mathbf{U} and \mathbf{V} , respectively. Then, given that $w_i = w_{1i}w_{2i}$, the reweighted term in (27) can be relaxed to

$$\begin{aligned}\sum_{i=1}^d w_i \sigma_i(\mathbf{R}) &= \sum_{i=1}^d w_{1i} \sigma_i(\mathbf{U}) w_{2i} \sigma_i(\mathbf{V}) \\ &\leq \left(\sum_{i=1}^d w_{1i}^a \sigma_i^a(\mathbf{U}) \right)^{\frac{1}{a}} \left(\sum_{i=1}^d w_{2i}^b \sigma_i^b(\mathbf{V}) \right)^{\frac{1}{b}} \\ &\leq \frac{1}{a} \left(\sum_{i=1}^d w_{1i}^a \sigma_i^a(\mathbf{U}) \right) + \frac{1}{b} \left(\sum_{i=1}^d w_{2i}^b \sigma_i^b(\mathbf{V}) \right) \\ &= \frac{1}{a} \left(\sum_{i=1}^d w_{ui} \sigma_i^a(\mathbf{U}) \right) + \frac{1}{b} \left(\sum_{i=1}^d w_{vi} \sigma_i^b(\mathbf{V}) \right),\end{aligned}\quad (28)$$

where the first and the second inequalities hold due to Holder's inequality and the Jensen's inequality, respectively. Suppose that $a = b = 2$, and then the weighted nuclear norm minimization problem in (25) can be recast as

$$\min_{\mathbf{U}, \mathbf{V}} \frac{1}{2} \left(\sum_{i=1}^d w_{ui} \sigma_i^2(\mathbf{U}) \right) + \frac{1}{2} \left(\sum_{i=1}^d w_{vi} \sigma_i^2(\mathbf{V}) \right) + \frac{\mu}{2} \|\mathbf{Y} - \mathbf{U}\mathbf{V}^\mathbf{H} - \mathbf{X}\|_F^2. \quad (29)$$

The ADMM framework is applied to solve the above weighted matrix factorization problem. Note that the subproblems of optimizing \mathbf{U} and \mathbf{V} become the weighted Frobenius norm minimization problems, which could be solved similarly as the CWSVT algorithm [23]. This weighted matrix factorization model inherits the advantage of the weighted nuclear norm minimization model. Furthermore, it is able to relieve the aforementioned two shortcomings of the nuclear norm minimization model, i.e., the excessive decrease of large singular values of interference matrix and the high computational complexity.

Table 1 Computational complexities of the IALM, RMF, and MFD algorithms

Algorithms	IALM	RMF	MFD
Computational complexity	$O(KL^2 + 2KL)$	$O(KLd + Kd^2 + Ld^2 + 2KL)$	$O(KLd + Kd^2 + 2KL)$

From another perspective, the upper bound of the NBI matrix rank can be directly restrained in the constraint, and then a decomposition model can be formulated as

$$\begin{aligned} \min_{\mathbf{I}, \mathbf{X}} \quad & \lambda \|\mathbf{X}\|_1 + \frac{\mu}{2} \|\mathbf{Y} - \mathbf{I} - \mathbf{X}\|_F^2 \\ \text{s.t.} \quad & \text{rank}(\mathbf{I}) \leq r. \end{aligned} \quad (30)$$

By employing the matrix factorization technique to represent the low-rank term, the problem can be further recast as

$$\begin{aligned} \min_{\mathbf{U}, \mathbf{V}, \mathbf{X}} \quad & \lambda \|\mathbf{X}\|_1 + \frac{\mu}{2} \|\mathbf{Y} - \mathbf{UV}^H - \mathbf{X}\|_F^2 \\ \text{s.t.} \quad & \text{rank}(\mathbf{U}) = \text{rank}(\mathbf{V}) \leq r. \end{aligned} \quad (31)$$

This problem can be similarly solved via alternate optimization of the NBI matrix and the real echoes. When tackling with the matrix factorization low-rank subproblem, the QR decomposition on small-scale matrix is introduced to accelerate the algorithm.

Both the matrix factorization algorithms are more efficient than the classic ALM algorithm. Without generality, suppose that $K > L$, and then the main computational cost of the ALM algorithm is the SVD on a $K \times L$ matrix per iteration, which takes at least $O(KL^2)$ flops. The reweighted matrix factorization algorithm takes the SVD on two small-scale matrices with $O(Kd^2 + Ld^2)$ flops, and the main computational burden is the multiplication of two small matrices with $O(KLd)$ flops per iteration. The matrix factorization decomposition algorithm takes one QR decomposition with $O(Kd^2)$ flops, in addition to the multiplication of two small matrices with $O(KLd)$ flops. All of the aforementioned algorithms use an element-wise soft thresholding algorithm to update the real echoes with at least $O(2KL)$ flops per iteration. The computational complexities of the inexact ALM (IALM) algorithm, reweighted matrix factorization (RMF) algorithm, and matrix factorization decomposition (MFD) algorithm per iteration are listed in Table 1. With the same data shown in Fig. 10, the performance of RMF method and MFD method is illustrated in Fig. 13, where the details of yellow rectangles are shown in the corner. Both methods are able to mitigate the strong interferences in good performance, and both are quite fast to convergence due to matrix factorization techniques.

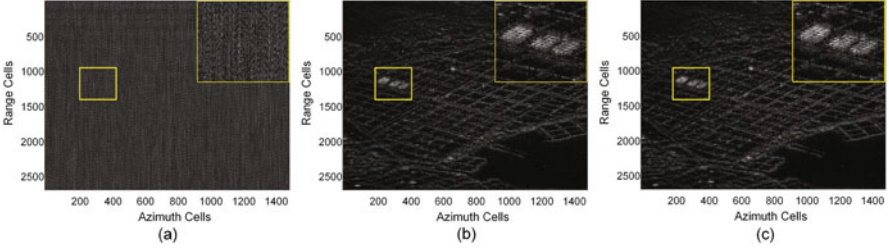


Fig. 13 Illustration of the interference mitigation results via RMF method and MFD method. Input SINR is -20 dB. (a) NBI-polluted results. (b) RMF method. Output SINR = 8.67 dB. (c) MFD method. Output SINR = 8.85 dB

Low-Rank Recovery with Image-Domain Regularization

Based on the above analysis, strong NBIs are well represented by the low-rank model and mitigated via the machine learning decomposition methods. For most of the methods, the optimization problems include the sparse regularization term to avoid real echoes being wrongly mitigated by the NBI suppression methods. In fact, the useful signals are scene-reflected SAR signals, and they are not sparse for most cases except for the water area. As one can recall, the regularization is a proper tool to overcome overfitting problem in machine learning domain. The sparse regularization term aims to keep the strong amplitudes of scatterers when extracting the low-rank interference from the received signals. If there are no sparse regularization terms in the objective function, then the low-rank signal, mixed with the strong real scatterers, would be extracted. Actually, this model is similar with the nonparametric model and parametric model, which are designed to extract the interference as much as possible without any regularizations or protections of real echoes. Therefore, to design a good regularization term is a critical problem for interference mitigation.

Basically, after applying the SAR imaging algorithms [40], such as range-Doppler (RD) algorithm and chirp-scaling (CS) algorithm, the strong scatterers in 2D image domain are sparser than those in unfocused raw data. Hence, it is better to constrain the sparsity of real echoes in image domain rather than in the raw data. Suppose that the classic SAR RD imaging operation is defined by

$$\mathbf{Z} = \mathcal{F}_a^{-1} \left(\mathbf{H}_a \odot \mathcal{F}_a \left(\mathcal{F}_r^{-1} \left(\mathbf{H}_r \odot \mathcal{F}_r \mathbf{X} \right) \right) \right) \triangleq \mathcal{F} \mathbf{X}, \quad (32)$$

where \odot denotes the Hadamard product, \mathcal{F} , \mathcal{F}_r , and \mathcal{F}_a denote imaging operation, range Fourier transform, and azimuth Fourier transform, respectively. In addition, their inverse operations are denoted by the superscript $^{-1}$. And \mathbf{H}_r and \mathbf{H}_a are range and azimuth matched filters. Vice versa, the inverse transform from focused SAR image to raw data is

$$\mathbf{X} = \mathcal{F}_r^{-1} \left(\mathbf{H}_r^* \odot \mathcal{F}_r \left(\mathcal{F}_a^{-1} \left(\mathbf{H}_a^* \odot \mathcal{F}_a \mathbf{Z} \right) \right) \right) \triangleq \mathcal{F}^* \mathbf{Z}, \quad (33)$$

where the superscript $*$ denotes the conjugate operation. Then, the low-rank representation problem with a sparse regularization in SAR image domain can be formulated as

$$\begin{aligned} \min_{\mathbf{I}, \mathbf{Z}} \quad & \|\mathbf{I}\|_* + \lambda \|\mathbf{Z}\|_1 \\ \text{s.t.} \quad & \|\mathbf{Y} - \mathbf{I} - \mathcal{F}^* \mathbf{Z}\|_{\text{F}}^2 \leq \delta. \end{aligned} \quad (34)$$

Comparing the above problem with the classic RPCA problem in (15), the sparse regularization is added to the real echoes in image domain instead of those in the raw data. To solve the above problem, it is similar to the solution of classic RPCA problem by alternately optimizing two variables. The solution of the nuclear norm minimization problem is the same, and the sparse recovery subproblem would be solved in the image domain in lieu of the raw data domain. Herein, we use the Radarsat-1 dataset [40] and simulate the strong NBIs to test the performance of image-domain regularization in Fig. 14. The input SINR is set as -20 dB, and 50 NBI artifacts are simulated to corrupt the SAR raw data. As can be seen in the enlarged details of yellow rectangles, the low-rank representation method with an image-domain regularization can better protect the strong scatterers with lower sidelobes compared with the traditional RPCA method. Both methods can recover the scene clearly with high accuracy.

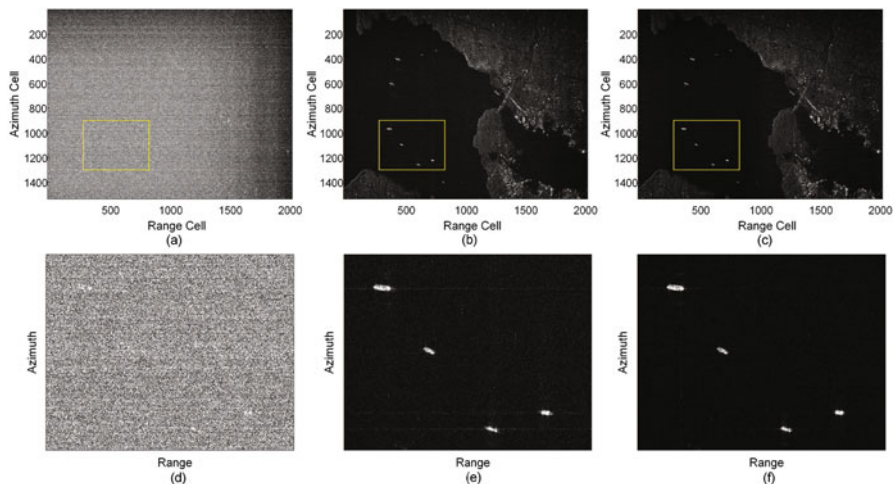


Fig. 14 Illustration of the interference mitigation results via RPCA method and the low-rank representation method with the image-domain regularization. Left column: interference-polluted images. Middle column: RPCA method. Right column: low-rank representation method with image-domain regularization

3.3.2 WBI Mitigation

Regarding the WBI, the interference's bandwidth occupies a large portion of the SAR bandwidth. Then, in range frequency domain, the WBI is not sparse in fact. Only a few researches focused on WBI mitigation problem for SAR systems [14, 28, 41]. With the help of STFT, most previous studies transform single-pulse received signal into the time–frequency domain. Herein, one WBI, such as the CM WBI and the SM WBI, can be clearly observed, as shown in Fig. 9. Then, the notched filter in time–frequency domain can be applied to remove the time–frequency bands at which strong WBIs locate. Different from the notched filter method, the machine learning framework still works for decomposing the WBIs and the real echoes. In [14], an eigen subspace filtering method, which is actually the PCA framework, was proposed to project single-pulse received signal onto the interference space. The interference space is constructed by the eigenvectors corresponding to the dominant eigenvalues. Then, remove the projected signals, and one can obtain the residual time–frequency spectrum of the useful signal. The recovered results of the ESP method and the notched filter methods are shown in Fig. 15. As can be seen, the 1D and 2D notched filters are clearly shown in Fig. 15 with obvious nulls generated at the strong WBIs' spectra. The ESP method can remove the WBIs via the eigen subspace filtering so that the WBIs' spectra would be removed thoroughly.

The WBI mitigation problem is more complicated than the NBI mitigation problem. Rather than the WBI mitigation problem, more machine learning studies focused on how to mitigate NBI and WBI simultaneously in recent years.

3.3.3 Simultaneous Mitigation of Complicated NBI and WBI

Most previous studies analyzed the simple WBI scenario, where the interferences are isolated in time–frequency domain. But in practice, the measured interferences are more complicated with different types of interferences, such as NBIs and WBIs. And the mixed interferences dominate the whole time–frequency spectra. For this case, parametric methods failed to suppress the interferences because it is hard to

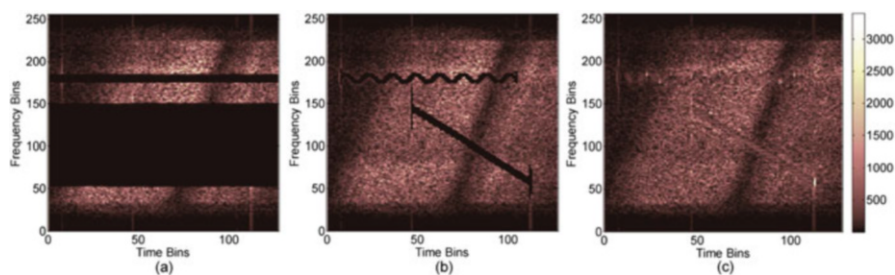


Fig. 15 Illustration of the interference mitigation results via notched filter methods and ESP method (i.e., PCA). (a) 1D notched filter. (b) 2D time–frequency notched filter. (c) ESP method

model this kind of mixed interferences. Some studies, such as [28, 41], focus on the simple interference cases that the interferences are isolated in the time–frequency domain. Machine learning methods were demonstrated to be good tools to mitigate the mixed interferences simultaneously. In practical applications, the characteristics of complicated interferences, such as the frequency, are relatively stationary across a short period of time. Basically, the STFT is a linear transform, and then the STFT coefficient vector in the l -th pulse can be formulated as

$$\mathbf{z}_l = \Phi \mathbf{i}_l, \quad (35)$$

where Φ denotes the full-rank STFT matrix [21], and \mathbf{z}_l and \mathbf{i}_l denote the tim–frequency interference vector and the interference vector in the l -th pulse, respectively. Combine multiple pulses of signals in one synthetic aperture period, and the tim–frequency spectra of interferences are

$$\mathbf{Z} = \Phi \mathbf{I}. \quad (36)$$

During the synthetic aperture time, the interference is relatively stationary so that the tim–frequency spectra would have a special low-rank property under the STFT matrix. Through the ISTFT, we have

$$\mathbf{I} = \Phi^{-1} \mathbf{Z}, \quad (37)$$

where Φ^{-1} denotes the ISTFT matrix, which is also full-rank. If \mathbf{Z} is low-rank, the original matrix \mathbf{I} is low-rank according to the fact that $\text{rank}(\mathbf{I}) \leq \min\{\text{rank}(\Phi^{-1}), \text{rank}(\mathbf{Z})\}$ and $\text{rank}(\mathbf{Z}) \leq \min\{\text{rank}(\Phi), \text{rank}(\mathbf{I})\}$. Based on this special low-rank property, the complicated interferences can be mitigated via the classic low-rank recovery methods, such as the basic RPCA model shown in (15). Furthermore, based on the extended consideration, a higher dimensional model, like a tensor structure, would provide more degrees of freedom (DOFs) and achieve better performance of interference mitigation. In the next subsections, two typical machine learning frameworks based on low-rank tensor recovery are analyzed in detail.

Multi-look Low-Rank Tensor Recovery

As analyzed above, the complicated interferences are approximately stable along a short period of time, while the real echoes are commonly not stable, especially for the heterogeneous area and the different line-of-flight views from the radar platform to targets. As can be well known that a higher dimension is able to provide more DOFs and useful information than the lower dimension. Some intrinsic properties of the interferences may not be well represented by the 1D frequency model or 2D tim–frequency model. Hence, if we cut the whole synthetic aperture time into several parts, i.e., a large, received signal matrix is divided into several small matrices, as

shown in Fig. 16. The signal model is defined as

$$\mathbf{Y} = [\mathbf{Y}_1 \ \mathbf{Y}_2 \ \cdots \ \mathbf{Y}_M] \quad (38)$$

and

$$\mathbf{Y}_m = \mathbf{X}_m + \mathbf{I}_m + \mathbf{N}_m, \quad (39)$$

where \mathbf{Y}_m , \mathbf{X}_m , \mathbf{I}_m , and \mathbf{N}_m are the received signal, real echo, interference, and background noise components of the m -th small matrices, respectively. Among these small matrices, the interferences are quite similar, while useful signal matrices would be different in a comparison. If these small matrices are stacked into a three-mode tensor with the size being $K \times L/M \times M$, the received signal is equivalently represented in a 3D range-azimuth-space (multi-look) domain.

This 3D model can better represent the potential low-rank property of the interferences and also increase the gap between interferences and real echoes so that the interferences are expected to be more easily removed. This new tensor model is

$$\begin{aligned} \mathcal{Y} &= \mathcal{X} + \mathcal{I} + \mathcal{N} \\ &= \mathcal{X} + \mathcal{NBI} + \mathcal{WBI} + \mathcal{N}, \end{aligned} \quad (40)$$

where $\mathcal{Y} \in \mathbb{C}^{K \times L/M \times M}$, $\mathcal{I} \in \mathbb{C}^{K \times L/M \times M}$, and $\mathcal{X} \in \mathbb{C}^{K \times L/M \times M}$ are the received signal tensor, the RFI tensor, and the useful signal tensor, respectively, as shown in Fig. 16. Rather than constrain the low-rank property in time-frequency spectra, the low-rank property of the interference tensor can be fully leveraged in the range-azimuth-space domain. Then, the optimization problem is formulated as

$$\begin{aligned} \min_{\mathcal{I}, \mathcal{X}} \quad & \text{trank}(\mathcal{I}) + \lambda \|\mathcal{X}\|_{\text{TR}} \\ \text{s.t.} \quad & \|\mathcal{Y} - \mathcal{I} - \mathcal{X}\|_{\text{TF}}^2 < \delta, \end{aligned} \quad (41)$$

where “trank” denotes the rank of a tensor, $\|\cdot\|_{\text{TF}}$ is the Frobenius norm of tensor, i.e., $\|\mathcal{A}\|_{\text{TF}} = \sqrt{\sum_{n_1 n_2 n_3} |\mathcal{A}_{n_1 n_2 n_3}|^2}$, and $\|\cdot\|_{\text{TR}}$ denotes the tensor regularization.

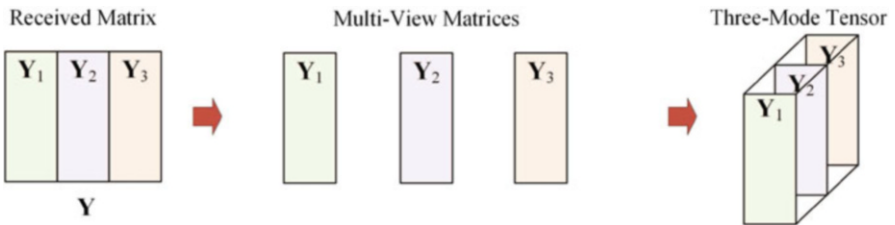


Fig. 16 Illustration of multi-look tensor model

To solve the above problem, the definition of the rank should be first confirmed. Herein, the tensor-SVD (t-SVD) algorithm is employed to relax this problem. Basically, for a tensor $\mathcal{A} \in \mathbb{C}^{N_1 \times N_2 \times N_3}$, the t-SVD is defined as

$$\mathcal{A} = \mathcal{U} * \mathcal{S} * \mathcal{V}^*, \quad (42)$$

where $\mathcal{U} \in \mathbb{C}^{N_1 \times N_1 \times N_3}$ and $\mathcal{V} \in \mathbb{C}^{N_2 \times N_2 \times N_3}$ are orthogonal, and $\mathcal{S} \in \mathbb{C}^{N_1 \times N_2 \times N_3}$ is an f-diagonal tensor, i.e., each frontal slice of \mathcal{S} is a diagonal matrix. Regards to more specific definitions and properties about t-SVD, please refer to [42] in detail. Based on the t-SVD, the optimization problem can be relaxed as

$$\begin{aligned} \min_{\mathcal{I}, \mathcal{X}} \quad & \|\mathcal{I}\|_{\mathbf{T}_*} + \lambda \|\mathcal{X}\|_{\mathbf{T}_1} \\ \text{s.t.} \quad & \|\mathcal{Y} - \mathcal{I} - \mathcal{X}\|_{\mathbf{T}_F}^2 < \delta, \end{aligned} \quad (43)$$

where $\|\cdot\|_{\mathbf{T}_*}$ is the nuclear norm of tensor, and $\|\cdot\|_{\mathbf{T}_1}$ is the ℓ_1 norm of tensor. This problem is defined as the complex tensor RPCA (CT-RPCA) problem. The tensor nuclear norm is used here, and it is defined as

$$\|\mathcal{I}\|_{\mathbf{T}_*} = \frac{1}{M} \sum_{j=1}^M \left\| \bar{\mathcal{I}}^{(j)} \right\|_*. \quad (44)$$

Basically, the block circulant matrix can be mapped to a block diagonal matrix, which can be formulated as

$$(\mathbf{F}_{N_3} \otimes \mathbf{E}_{N_1}) \cdot \text{bcirc}(\mathcal{A}) \cdot (\mathbf{F}_{N_3}^{-1} \otimes \mathbf{E}_{N_2}) = \text{bdiag}(\bar{\mathcal{A}}). \quad (45)$$

Then, the tensor nuclear norm is calculated as

$$\begin{aligned} \|\mathcal{I}\|_{\mathbf{T}_*} &= \frac{1}{M} \sum_{m=1}^M \sum_{n=1}^{\min\{K, L/M\}} \sigma_n(\bar{\mathcal{I}}) \\ &= \frac{1}{M} \sum_{j=1}^{\min\{KM, L\}} \sigma_j(\text{bdiag}(\bar{\mathcal{I}})). \end{aligned} \quad (46)$$

And further, we have

$$\begin{aligned} \|\mathcal{I}\|_{\mathbf{T}_*} &= \frac{1}{M} \sum_{j=1}^{\min\{KM, L\}} \sigma_j(\text{bdiag}(\bar{\mathcal{I}})) \\ &= \frac{1}{M} \sum_{j=1}^{\min\{KM, L\}} \sigma_j \left((\mathbf{F}_M \otimes \mathbf{E}_K) \cdot \text{bcirc}(\mathcal{I}) \cdot (\mathbf{F}_M^{-1} \otimes \mathbf{E}_{L/M}) \right) \\ &= \frac{1}{M} \sum_{j=1}^{\min\{KM, L\}} \sigma_j(\text{bcirc}(\mathcal{I})). \end{aligned} \quad (47)$$

As a result, the nuclear norm of interference tensor is equivalent to the nuclear norm of the block circulant matrix. It indicates that the tensor nuclear norm is able to constrain the rank of interference tensor in the range-azimuth-space domain. Then, the problem can be alternately optimized via two subproblems as follows:

$$\min_{\mathcal{I}} \|\mathcal{I}\|_{\mathbf{T}^*} + \frac{\mu}{2} \|\mathcal{Y} - \mathcal{I} - \mathcal{X}\|_{\mathbf{T}\mathbf{F}}^2, \quad (48)$$

$$\min_{\mathcal{X}} \lambda \|\mathcal{X}\|_{\mathbf{T}_1} + \frac{\mu}{2} \|\mathcal{Y} - \mathcal{I} - \mathcal{X}\|_{\mathbf{T}\mathbf{F}}^2. \quad (49)$$

The tensor nuclear norm minimization problem in (48) can be effectively solved via the independent frontal slice minimization analyzed in [30]. After solving the optimization problem, the \mathcal{X} tensor should be re-stacked to the size of the original matrix. Herein, the X-band airborne SAR data is provided to evaluate the performance of this multi-look tensor recovery method. Complicated NBIs and WBIs are simulated, where the input SINR is -20 dB, and they are continuously distributed in time–frequency domain. The mitigation results are shown in Fig. 17. As can be seen, the complicated interferences are well removed, and the recovered root mean square error (RMSE) shows its excellent performance. The sparse recovery method fails to be applied to this case since it is only feasible for the isolated interferences, such as the cases in Fig. 6.

Smoothing Tensor Factorization Algorithm

The analysis in the above subsection demonstrates the effectiveness of the tensor structure for complicated interference mitigation. When breaking the original large-scale matrix of received signals into several small-scale matrices, it creates extra spatial (multi-view) DOFs to explore the gap between interferences and real echoes in 3D domain, while this multi-look model loses a large amount of azimuth DOFs. This may lead to performance degradation of the interference suppression. In order

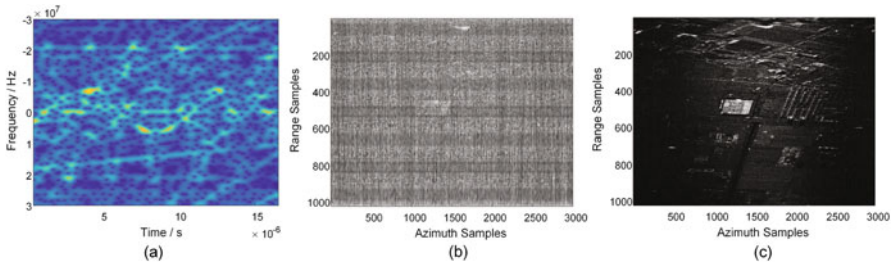


Fig. 17 Illustration of the complicated interference mitigation result via multi-look tensor recovery method. (a) Time–frequency spectrum of complicated interferences. (b) Interference-polluted image. (c) Interference mitigation result. RMSE = 0.2342

to simultaneously create a spatial dimension and try to keep as many azimuth DOFs as possible, a new tensor structure was proposed by using the azimuth smoothing technique to construct each frontal slice of the tensor, which is illustrated in Fig. 18.

As can be seen, the smoothing multi-view (SMV) model is a little bit more complicated than the multi-view model. Suppose that there are L pulses of signals in a synthetic aperture period, and $L - N + 1$ smoothing matrices are constructed based on the original signal matrix. The s -th ($1 \leq s \leq L - N + 1$) smoothing matrix can be constructed as

$$\mathbf{Y}_s = [\mathbf{y}_s \ \mathbf{y}_{s+1} \ \cdots \ \mathbf{y}_{s+N-1}]. \quad (50)$$

Benefiting from this smoothing architecture, N pulses of signals are kept in one frontal slice and the number N can be tuned from 1 to $L - 1$. Hence, based on this new SMV structure, the received signal can be formulated by a smoothing tensor in 3D range-azimuth-space domain, as shown in Fig. 18. The received signal tensor is the sum of the interference tensor, the real echo tensor, and the noise tensor. For the new SMV structure, the interference tensor is still low rank due to its relatively stationary performance during the synthetic aperture period. And the frontal slices of the real echo tensor are different because the real echoes are variant for different views or azimuth periods. Therefore, the CT-RPCA optimization problem still works for this new model.

The multi-view model divides the large-scale matrix of received signals into several small-scale matrices so that the computational burden for CT-RPCA optimization problem is acceptable. However, the SMV model aims to protect both spatial and azimuthal DOFs, and the large-scale tensor structure would involve high computational burden. In fact, the low-rank tensor recovery optimization dominates the total computational complexity, so we omit the subproblem relating to the real echo tensor. Then, in order to relieve the total computational complexity, a tensor factorization scheme is employed to break the large-scale tensor into a t-product of two small-scale tensors, that is,

$$\mathcal{I} = \mathcal{U} * \mathcal{V}, \quad (51)$$

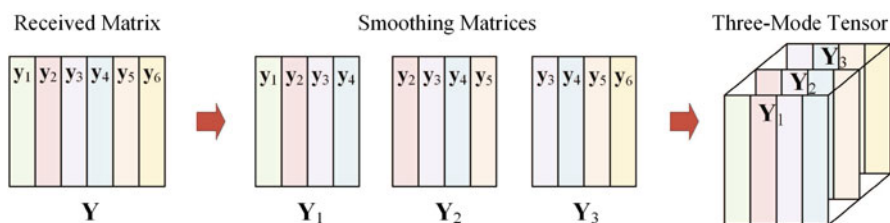


Fig. 18 Illustration of the smoothing tensor model

where $\mathcal{U} \in \mathbb{C}^{K \times \hat{r} \times (L-N+1)}$ and $\mathcal{V} \in \mathbb{C}^{\hat{r} \times N \times (L-N+1)}$. Moreover, \hat{r} is the estimated upper bound of the rank of tensor and $\hat{r} \ll K, N$. According to the property of t-product operation, one can note that

$$\bar{\mathcal{I}}^{(i)} = \bar{\mathcal{U}}^{(i)} \bar{\mathcal{V}}^{(i)}, \quad (52)$$

where $i = 1, 2, \dots, L-N+1$. Based on the definition of the block diagonal matrix, we also have

$$\text{bdiag}(\bar{\mathcal{I}}) = \text{bdiag}(\bar{\mathcal{U}}) \text{bdiag}(\bar{\mathcal{V}}). \quad (53)$$

In addition, the tensor nuclear norm is a convex relaxation of the original rank operation. The tensor nuclear norm minimization has the similar drawback as the matrix nuclear norm minimization, which may excessively decrease the large singular values of the interference matrix. As a consequence, a reweighted tensor nuclear norm is defined to approximate the true tensor rank as follows:

$$\begin{aligned} \|\mathcal{I}\|_{\text{RTNN}} &= \frac{1}{Q} \sum_{i=1}^Q \sum_{j=1}^{\min\{K, N\}} w_j \sigma_j \left(\bar{\mathcal{I}}^{(i)} \right) \\ &= \frac{1}{Q} \sum_{i=1}^Q \left\| \bar{\mathcal{I}}^{(i)} \right\|_{\text{RNN}}, \end{aligned} \quad (54)$$

where the subscript RNN denotes the reweighted nuclear norm for the matrix, i.e., $\|\mathbf{A}\|_{\text{RNN}} = \sum_j w_j \sigma_j(\mathbf{A})$. And the reweighted tensor nuclear norm can also be further derived as a constraint of the block circulant matrix:

$$\begin{aligned} \|\mathcal{I}\|_{\text{RTNN}} &= \frac{1}{Q} \sum_{j=1}^{\min\{KQ, NQ\}} w_j \sigma_j (\text{bdiag}(\bar{\mathcal{I}})) \\ &= \frac{1}{Q} \sum_{j=1}^{\min\{KQ, NQ\}} w_j \sigma_j (\text{bcirc}(\mathcal{I})). \end{aligned} \quad (55)$$

Then, based on the reweighted tensor nuclear norm and the tensor factorization technique, it can be further formulated as

$$\|\mathcal{I}\|_{\text{RTNN}} = \frac{1}{Q} \sum_{j=1}^{\min\{KQ, NQ\}} w_j \sigma_j (\text{bdiag}(\bar{\mathcal{U}}) \text{bdiag}(\bar{\mathcal{V}})). \quad (56)$$

With the relaxation introduced in (27) and (28), the reweighted tensor nuclear norm minimization problem can be relaxed as

$$\begin{aligned} \min_{\mathcal{U}, \mathcal{V}} & \frac{1}{2} \left(\|\mathcal{U}\|_{\text{RTFN}}^2 + \|\mathcal{V}\|_{\text{RTFN}}^2 \right) \\ \text{s.t.} & \quad \|\mathcal{Y} - \mathcal{U} * \mathcal{V} - \mathcal{X}\|_{\text{TF}}^2 < \delta, \end{aligned} \quad (57)$$

where the reweighted tensor Frobenius norm (RTFN) is defined as

$$\begin{aligned}\|\mathcal{U}\|_{\text{RTFN}}^2 &= \frac{1}{Q} \sum_{i=1}^Q \sum_{j=1}^{\hat{r}} w_j \sigma_j^2 \left(\bar{\mathcal{U}}^{(i)} \right) \\ &= \frac{1}{Q} \sum_{i=1}^Q \left\| \bar{\mathcal{U}}^{(i)} \right\|_{\text{RFN}}^2,\end{aligned}\quad (58a)$$

$$\begin{aligned}\|\mathcal{V}\|_{\text{RTFN}}^2 &= \frac{1}{Q} \sum_{i=1}^Q \sum_{j=1}^{\hat{r}} w_j \sigma_j^2 \left(\bar{\mathcal{V}}^{(i)} \right) \\ &= \frac{1}{Q} \sum_{i=1}^Q \left\| \bar{\mathcal{V}}^{(i)} \right\|_{\text{RFN}}^2.\end{aligned}\quad (58b)$$

Then, the large-scale low-rank reweighted tensor optimization problem is relaxed to the small-scale RFN optimization problem. And the whole optimization problem can be break into several independent minimization problems for all the frontal slices

$$\begin{aligned}\min_{\mathcal{U}, \mathcal{V}} \quad & \frac{1}{2Q} \left(\sum_{i=1}^Q \left\| \bar{\mathcal{U}}^{(i)} \right\|_{\text{RFN}}^2 + \sum_{i=1}^Q \left\| \bar{\mathcal{V}}^{(i)} \right\|_{\text{RFN}}^2 \right) \\ \text{s.t.} \quad & \|\mathcal{Y} - \mathcal{U} * \mathcal{V} - \mathcal{X}\|_{\text{TF}}^2 < \delta.\end{aligned}\quad (59)$$

This problem can be efficiently solved compared with the original problem. In the whole process, we omit the analysis of the real echo tensor optimization subproblem since it can be solved via the same solution introduced in CT-RPCA. Complicated NBIs and WBIs are simulated and added on the X-band SAR data, which is similar to the previous subsection. Herein, the tensor factorization method based on SMV model is evaluated to mitigate the complicated interferences, and the results are shown in Fig. 19. Almost all the azimuth DOFs and extra spatial DOFs are fully considered, and the mitigation performance is better than the tensor recovery method with the multi-view model.

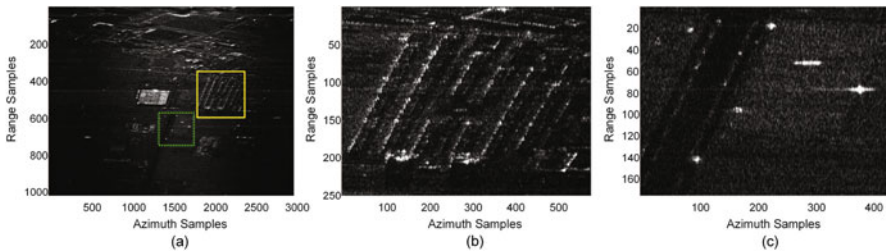


Fig. 19 Illustration of the mitigation performance of the tensor factorization method based on SMV model. (a) Mitigation result. RMSE = 0.2251. SSIM = 0.9859. (b) Enlarged area of yellow rectangle. (c) Enlarged area of green rectangle

4 Future Trend of Interference Mitigation via Deep Learning and Cognitive Scheme

In recent years, the cognitive radar system gradually becomes an intelligent means to adaptively and actively sense the complex electromagnetic environment and take careful considerations on the waveform design or system re-design work immediately. One can conclude that not all the encountered interferences can be effectively mitigated via post signal processing. The interference mitigation technology, including the aforementioned machine learning methods, cannot be regarded as a standalone restoration for the external interference problems experienced by present SAR systems. It is better that the interference environment can be greatly understood before sensor deployment, flight, or launch, and thus the corresponding mitigation methods could be developed in advance.

In addition, the deep learning tools, especially for the supervised learning algorithms, have excellent performance for nearly all the tasks in computer vision and natural language processing areas. However, one critical problem of this effective tool is the collection of as much data as possible for training and realizing excellent performance. Therefore, some pioneers attempt to produce plenty of data for training the deep neural networks before mitigating interferences, resulting in another feasible understanding for cognitive systems. In [43], an interference detection and mitigation algorithm based on deep neural networks was developed on the time–frequency spectra of SAR echoes. Both NBI and WBI can be effectively mitigated via the well-trained deep learning methods. By learning from plenty of interference-polluted data, more variety of refined signal models would be established, which is beneficial for performance improvement.

References

1. National Academies of Sciences, Engineering, and Medicine (2015) A Strategy for Active Remote Sensing Amid Increased Demand for Radio Spectrum. National Academies Press.
2. Moreira, A., Prats-Iraola, P., Younis, M., Krieger, G., Hajnsek, I., Papathanassiou, K. P. (2013) A tutorial on synthetic aperture radar. *IEEE Geoscience and remote sensing magazine* 1(1): 6–43.
3. Huang, Y., Zhao, B., Tao, M., Chen, Z., Hong, W. (2020) Review of synthetic aperture radar interference suppression. *Journal of Radar* 9(1): 86–106.
4. Tao, M., Su, J., Huang, Y., Wang, L. (2019) Mitigation of radio frequency interference in synthetic aperture radar data: current status and future trends. *Remote Sensing* 11(20): 2438.
5. Zhang, L., Qiao, Z., Xing, M., Yang, L., Bao, Z. (2012) A robust motion compensation approach for UAV SAR imagery. *IEEE transactions on geoscience and remote sensing* 50(8): 3202–3218.
6. Huang, Y., Lan, L., Zhang, L., Chen, Z., Xu, G. (2019) Simultaneous narrowband and wideband interference suppression on single-channel SAR system via low-rank recovery. *2019 IEEE International Geoscience and Remote Sensing Symposium*: 600–603.
7. Tao, M., Zhou, F., Zhang, Z. (2017) Characterization and mitigation of radio frequency interference in PolSAR data. *Radio Science* 52(11): 1405–1418.

8. Xu, H., Wu, Z., Liu, W., Li, J., Feng, Q (2015) Analysis of the effect of interference on InSAR. *IEEE Sensors Journal* 15(10): 5659–5668.
9. Spencer, M. W., Chen, C. W., Ghaemi, H., Chan, S. F., Belz, J. E (2013) RFI characterization and mitigation for the SMAP radar. *IEEE transactions on geoscience and remote sensing* 51(10): 4973–4982.
10. Richards, M. A (2014) *Fundamentals of radar signal processing*. McGraw-Hill Education.
11. Zhou, F., Wu, R., Xing, M., Bao, Z (2007) Eigensubspace-based filtering with application in narrow-band interference suppression for SAR. *IEEE Geoscience and Remote Sensing Letters* 4(1): 75–79.
12. Lord, R. (2005) Radio frequency interference suppression applied to synthetic aperture radar data. XXVIIth General Assembly of International Union of Radio Science, URSI, New Delhi, India, India, Oct. 2005.
13. Le, C. T., Hensley, S., Chapin, E (1998) Removal of RFI in wideband radars. 1998 *IEEE International Geoscience and Remote Sensing Symposium Proceedings* 4: 2032–2034.
14. Tao, M., Zhou, F., Zhang, Z (2015) Wideband interference mitigation in high-resolution airborne synthetic aperture radar data. *IEEE Transactions on Geoscience and Remote Sensing* 54(1): 74–87.
15. Liu, Z., Liao, G., Yang, Z (2013) Time variant RFI suppression for SAR using iterative adaptive approach. *IEEE Geoscience and Remote Sensing Letters* 10(6): 1424–1428.
16. Huang, X., Liang, D (1999) Gradual RELAX algorithm for RFI suppression in UWB-SAR. *Electronics Letters* 35(22): 1916–1917.
17. Huang, Y., Liao, G., Xu, J., Li, J (2018) Narrowband RFI suppression for SAR system via efficient parameter-free decomposition algorithm. *IEEE Transactions on Geoscience and Remote Sensing* 56(6): 3311–3322.
18. Yang, Z., Du, W., Liu, Z., Liao, G (2015) WBI suppression for SAR using iterative adaptive method. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 9(3): 1008–1014.
19. Joy, S., Nguyen, L. H., Tran, T. D (2016) Radio frequency interference suppression in ultra-wideband synthetic aperture radar using range-azimuth sparse and low-rank model. 2016 *IEEE Radar Conference (RadarConf)*: 1–4.
20. Su, J., Tao, H., Tao, M., Wang, L., Xie, J (2017) Narrow-band interference suppression via rpca-based signal separation in time–frequency domain. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 10(11): 5016–5025.
21. Huang, Y., Liao, G., Li, J., Xu, J (2018) Narrowband RFI suppression for SAR system via fast implementation of joint sparsity and low-rank property. *IEEE Transactions on Geoscience and Remote Sensing* 56(5): 2748–2761.
22. Huang, Y., Liao, G., Zhang, Z., Xiang, Y., Li, J., Nehorai, A (2018) Fast narrowband RFI suppression algorithms for SAR systems via matrix-factorization techniques. *IEEE Transactions on Geoscience and Remote Sensing* 57(1): 250–262.
23. Huang, Y., Liao, G., Xiang, Y., Zhang, Z., Li, J., Nehorai, A (2019) Reweighted nuclear norm and reweighted Frobenius norm minimizations for narrowband RFI suppression on SAR system. *IEEE Transactions on Geoscience and Remote Sensing* 57(8): 5949–5962.
24. Lu, X., Su, W., Yang, J., Gu, H., Zhang, H., Yu, W., Yeo, T. S (2018) Radio frequency interference suppression for SAR via block sparse Bayesian learning. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 11(12): 4835–4847.
25. Huang, Y., Lan, L., Zhang, L., Zhou, Y., Xu, G., Wen, C (2019) Narrowband Interference Suppression on Single-Channel SAR Systems via Reweighted Tensor Nuclear Norm Minimization. 2019 *IEEE International Geoscience and Remote Sensing Symposium*: 2373–2376.
26. Huang, Y., Liao, G., Zhang, L., Xiang, Y., Li, J., Nehorai, A (2019) Efficient narrowband RFI mitigation algorithms for SAR systems with reweighted tensor structures. *IEEE Transactions on Geoscience and Remote Sensing* 57(11): 9396–9409.
27. Yang, H., Chen, C., Chen, S., Xi, F., Liu, Z (2020) A dictionary-based SAR RFI suppression method via robust PCA and chirp scaling algorithm. *IEEE Geoscience and Remote Sensing Letters*.

28. Liu, H., Li, D., Zhou, Y., Truong, T. K (2018) Simultaneous radio frequency and wideband interference suppression in SAR signals via sparsity exploitation in time–frequency domain. *IEEE Transactions on Geoscience and Remote Sensing* 56(10): 5780–5793.
29. Huang, Y., Zhang, L., Li, J., Chen, Z., Yang, X (2019) Reweighted tensor factorization method for SAR narrowband and wideband interference mitigation using smoothing multiview tensor model. *IEEE Transactions on Geoscience and Remote Sensing* 58(5): 3298–3313.
30. Huang, Y., Zhang, L., Li, J., Hong, W., Nehorai, A (2019) A novel tensor technique for simultaneous narrowband and wideband interference suppression on single-channel SAR system. *IEEE Transactions on Geoscience and Remote Sensing* 57(12): 9575–9588.
31. He, K., Zhang, X., Ren, S., Sun, J (2016) Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*: 770–778.
32. Graves, A., Mohamed, A. R., Hinton, G (2013) Speech recognition with deep recurrent neural networks. In *2013 IEEE international conference on acoustics, speech and signal processing*: 6645–6649.
33. Bojarski, M., Del Testa, D., Dworakowski, D., Firner, B., Flepp, B., Goyal, P., Zieba, K (2016) End to end learning for self-driving cars. *arXiv preprint arXiv:1604.07316*.
34. Taylor, J. D (2020) *Introduction to ultra-wideband radar systems*. CRC press.
35. Stanković, L., Orović, I., Stanković, S., Amin, M (2013) Compressive sensing based separation of nonstationary and stationary signals overlapping in time-frequency. *IEEE Transactions on Signal Processing* 61(18): 4562–4572.
36. Nguyen, L. H., Tran, T. D (2016) Efficient and robust RFI extraction via sparse recovery. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 9(6): 2104–2117.
37. Wang, Y., Yin, W., Zeng, J (2019) Global convergence of ADMM in nonconvex nonsmooth optimization. *Journal of Scientific Computing* 78(1): 29–63.
38. Huang, Y., Liao, G., Xiang, Y., Zhang, L., Li, J., Nehorai, A (2019) Low-rank approximation via generalized reweighted iterative nuclear and Frobenius norms. *IEEE Transactions on Image Processing* 29: 2244–2257.
39. Candès, E. J., Li, X., Ma, Y., Wright, J (2011) Robust principal component analysis?. *Journal of the ACM (JACM)* 58(3): 1–37.
40. Cumming, I. G., Wong, F. H (2005) *Digital processing of synthetic aperture radar data*. Artech house.
41. Zhang, S., Xing, M., Guo, R., Zhang, L., Bao, Z (2011) Interference suppression algorithm for SAR based on time–frequency transform. *IEEE transactions on Geoscience and remote sensing* 49(10): 3765–3779.
42. Zhang, Z., Ely, G., Aeron, S., Hao, N., Kilmer, M (2014) Novel methods for multilinear data completion and de-noising based on tensor-SVD. In *Proceedings of the IEEE conference on computer vision and pattern recognition*: 3842–3849.
43. Fan, W., Zhou, F., Tao, M., Bai, X., Rong, P., Yang, S., Tian, T (2019) Interference mitigation for synthetic aperture radar based on deep residual network. *Remote Sensing* 11(14): 1654.

Classification of SAR Images Using Compact Convolutional Neural Networks



Mete Ahishali, Serkan Kiranyaz, and Moncef Gabbouj

1 Introduction

The number of orbiting satellites or air-borne measurement devices may be the greatest in our time; and hence, there is a growing interest in remote sensing applications. In general, the applications have been developed for both remote sensing domains: active or passive remote sensing. In the acquisition process of the active remote sensing, an electromagnetic (EM) wave is transmitted, then any difference or perturbation that can be identified between the transmitted and received (the backscattered) EM waveforms the observation of the scene. In passive remote sensing acquisition, the sensor only receives the measured wave where the source of the EM wave is natural, e.g., the sunlight in optical satellite imagery.

Synthetic Aperture Radar (SAR) imaging is an active remote sensing technique and it is extensively used in many fields with different applications including agriculture, forestry, geology, and oceanography. Since it is an active remote sensing technique, it can operate during day or night time; and depending on the sensor of choice (more specifically, the frequency of the transmitted EM wave), it can also operate under various weather circumstances where the optical satellite imagery may fail. SAR systems can be space-borne or air-borne the acquisition system is mounted on an orbiting satellite or a plane having a specified trajectory, respectively. For example, various applications using different SAR systems can be

M. Ahishali (✉) · M. Gabbouj
Faculty of Information Technology and Communication Sciences, Tampere University, Tampere, Finland
e-mail: mete.ahishali@tuni.fi; moncef.gabbouj@tuni.fi

S. Kiranyaz
Electrical Engineering Department, College of Engineering, Qatar University, Doha, Qatar
e-mail: mkiranyaz@qu.edu.qa

© The Editor(s) (if applicable) and The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

147

M. Rysz et al. (eds.), *Synthetic Aperture Radar (SAR) Data Applications*, Springer Optimization and Its Applications 199, https://doi.org/10.1007/978-3-031-21225-3_7

listed as follows: oil spill detection with ENVISAT, RADARSAT-1 [1], tsunami-related building damage investigation using TerraSAR-X [2], vegetation monitoring using Sentinel-1 [3], land use/land cover (LU/LC) classification with AIRSAR [4–6], and RADARSAT-2 [7]. A more extensive list of applications is presented in [8] using different SAR sensors. In this book chapter, more focus is drawn on the LU/LC applications. The classification of LU/LC is an interesting task and it is well motivated due to its significance especially for socioeconomic and ecological applications. For instance, the connection between vegetation intensity and socioeconomic status is studied in [9] and it is stated that socioeconomic advantage is associated with the greater vegetation intensity. Next, the studies in [10, 11] investigate the effect of different vegetation on the temperature by relating urban climate with the vegetation type. The study in [12] has performed forest biomass analysis in the Mediterranean region to provide vegetation ecosystem analysis.

Nevertheless, the LU/LC classification is a challenging problem because of the following reasons: (i) the SAR data contain speckle noise that causes the performance degradation for the traditional classification methods, (ii) the need for a significant number of features to achieve a decent classification performance, and (iii) the scale of the SAR data that is ultimately linked with (ii). The existing studies consist of unsupervised and supervised techniques. In the first group, there are proposed studies [13, 14] based on different clustering methods, but it is observed that they have certain limitations for the high-resolution SAR images having heterogeneous regions. Similarly, various superpixel segmentation approaches [15, 16] are proposed aiming to group similar pixels together utilizing some low-level characteristics such as color. An analysis is presented by Stutz et al. [17] regarding superpixel segmentation-based unsupervised approaches. Finally, the method proposed in [18] aims to reduce the effect of speckle noise on the performance using an extended version of the mean shift algorithm.

Although unsupervised SAR classification methods may be valuable in the cases where it is unfeasible to provide any labeling, it is shown that many supervised approaches [6, 19–23] provide superior classification performance over the unsupervised approaches. The traditional supervised methods proposed for the SAR classification task consist of two steps: feature extraction and classification. Considering the classification of single or dual-polarimetric SAR data, it is particularly important to extract descriptive features and propose an efficient way to ensemble them, e.g., as followed by Uhlmann and Kiranyaz [21]. In this way, it is aimed to reduce the performance degradation that can possibly occur due to the lack of full polarization in the data.

In the previous study, an acceptable performance level is achieved by Uhlmann and Kiranyaz [21] for single- and dual-polarized SAR data. Accordingly, feature vectors are collected with high dimensions (for example, greater than 200-D) and they have followed using an ensemble of classifiers over the extracted features. Since the method proposed in [21] is a traditional approach, it was a need to compute a huge number of features consisting of various EM (primary information) and image processing (secondary information) features as the traditional classifier

cannot provide reasonable classification accuracy over the direct inference on the SAR intensity data. One major drawback with these traditional classifiers is that the extraction procedure of a huge amount of these features highly increases the computational time.

In Polarimetric SAR (PolSAR) classification, more information about the target can be obtained utilizing multiple orthogonal polarizations. This brings a huge advantage as various target decomposition theorems can be used in the classification, whereas, for instance, there is only one intensity channel in the single-polarized SAR data. Overall, the usage of fully polarimetric data enriches the EM (primary) features in a traditional classification framework. Ultimately, such an increase in the amount of EM features causes a similar issue with the conventional methods as in [4, 22, 23]: the computational burden due to the usage of high-dimensional features. Nevertheless, elegant classification performance is achieved by Uhlmann and Kiranyaz [22] by integrating the high-level EM features with image processing features similar to [21].

It can be said that the aforementioned conventional approaches for both partially and fully polarimetric SAR classification have some common limitations and drawbacks. First, the classification performance highly depends on the manually selected features as they involve feature engineering. Although this limitation has been tried to be addressed, e.g., in [23–25] by collecting a massive amount features together, it is certain that first, the extraction procedure increases the computational complexity, and then the following classification task has been becoming complicated because of *the curse of dimensionality*. Eventually, the classification accuracy for some terrain types suffers from suboptimal classification performance levels using the traditional methods.

Besides the traditional classification approaches with their aforementioned limitations, methods using Convolutional Neural Networks (CNNs) [26–28] have started to become a complete *state-of-the-art* approach in many applications such as object tracking, image recognition, and segmentation. CNNs combine the feature learning and classification parts in a single learning framework and they are jointly optimized during the training procedure. It is observed in the literature that such learning-based methods tend to follow deep structures with millions of trainable parameters. The Deep Learning (DL) based methods appear in the following applications of remote sensing as investigated in [29]: interpretation of high-resolution satellite and SAR images, multimodal data fusion, hyperspectral image analysis, and 3-D large-scale city reconstruction. However, the DL approaches, in general, require a massive amount of training data, for example, in a “Big Data” scale to train their enormous number of trainable model parameters. For example, this requirement is also observed in many recent studies [30, 31] in SAR image classification. These DL-based approaches have used a great portion of the annotated data such as 75% or larger to obtain promising classification accuracies. For instance, the proposed classification approach in [30] has used nearly 80% of SAR data during the training corresponding to 28,404 and 10,817 numbers of samples from San Francisco Bay and Flevoland L-band PolSAR images, respectively. Therefore, only 8000 and 2781 samples were selected in [30] for the actual performance evaluation over San

Francisco and Flevoland areas, respectively. Comparably, the approach in [31] has employed 75% of the annotated data in the training to accurately classify Flevoland SAR image at L-band (111,520 training samples of a total of 148,520 annotated samples). The list of approaches using large proportions as the training data can be extended: 75% and 88–92% of the annotated SAR data are used in [32, 33].

On the whole, the availability of large amounts of the annotated data is unfeasible in remote sensing because of the labeling cost and inconvenience that occurred during the ground-truth formation. One can be said that after the ground-truth collection, the need for an automatic classification framework may disappear as already a large proportion of the data is labeled. Moreover, since the scales of both the data and model are large in DL-based methods, there is a requirement of a special hardware setup during the training and for the inference as well to deal with the increased computational complexity. Altogether, their low-cost operations and real-time applications are very limited considering the discussed limitations and drawbacks.

The studies in [19, 20] have proposed superior classification frameworks for SAR and PolSAR classifications based on compact and adaptive CNNs. In this way, they aim to address the aforementioned weaknesses of the traditional approaches and the limitations of the DL-based methods. First, in [19], the proposed approach is evaluated over space-borne X-band SAR images including COSMO-SkyMed single-polarized and TerraSAR-X dual-polarized intensity data. The performance comparison has been performed against the previous state-of-the-art approach in [21] and two recent deep CNNs including Xception and Inception-Resnet-v2 [34, 35] that are initially proposed for ImageNet—Large-Scale Visual Recognition Challenge [36]. Accordingly, the novel and major contributions of [19] are expressed as follows: the proposed approach can operate directly using SAR intensity data without any additional pre-processing procedure, e.g., it does not require any image processing features on contrary to the compared method in [21]. Secondly, it is shown in [19] that the proposed approach with compact CNNs outperforms the compared deep CNNs and it can achieve state-of-the-art performance levels using only a small number of training samples that can be considered as negligible amount since it corresponds to less than 0.1% of the complete SAR data. Next, thanks to the proposed compact configuration in [19], the proposed approach can achieve greater computational complexity efficiency in both training and inference; and hence it is appropriate to use the proposed classification in real-time inference. Finally, it is shown that operating with small window sizes is possible with the proposed compact models since the input size can be configured as $(2n + 1) \times (2n + 1)$ with n as small as $n = 3$. The advantage of using such small sliding window sizes is that the produced LU/LC mask would have finer details if the window size is appropriately set to a small value in the sliding window classification technique. In the later study [20], the proposed approach with the compact CNNs has been evaluated in PolSAR classification using AIRSAR L-band and RADARSAT-2 C-band PolSAR data. Correspondingly, the PolSAR classification version of the approach in [20] inherits the same previously mentioned abilities as in [19] thanks to the CNNs that combine the feature learning and classification topologies and optimize them simultaneously

during the training. Furthermore, thanks to the PolSAR data, there is a large set of additional EM features that can be computed using different Target Decompositions. These features are utilized by the compared methods in [20]. In such a challenging comparison where the competing methods including [22, 23] enjoy the enriched descriptive information, it is observed that the proposed method still produces state-of-the-art classification accuracies while it operates over only the second-order descriptors of PolSAR data. Therefore, it is shown in [19, 20] that the proposed approach with compact and adaptive CNN achieves enhanced performance levels in both partially and fully polarimetric SAR classification tasks using low-level EM information and only an insignificant number of training samples.

The rest of the chapter is organized as follows: SAR information extraction is detailed in Sect. 2. Then, the proposed methodology will be detailed in Sect. 3. The experimental results will be presented in Sect. 4 along with a brief explanation regarding the evaluated SAR and PolSAR study areas and the computational complexity analysis. Finally, Sect. 5 concludes the chapter with concluding remarks and potential research directions in the future.

2 Background and Related Work

In this section, the background knowledge will be provided regarding the SAR data processing and PolSAR information extraction for the classification. Next, prior work in this domain will be briefly reviewed including traditional and deep CNN-based classification approaches.

2.1 SAR Data Processing

The observed target is defined by the complex backscattering matrix \mathbf{S} in a SAR acquisition system. In the case of having horizontally and vertically linear polarizations, the backscattering matrix would be $\mathbf{S} \in \mathbb{C}^{2 \times 2}$ with the corresponding elements as,

$$\begin{bmatrix} E_h^r \\ E_v^r \end{bmatrix} = \mathbf{S} \begin{bmatrix} E_h^t \\ E_v^t \end{bmatrix}, \quad \text{where} \quad \mathbf{S} = \begin{bmatrix} S_{hh} & S_{hv} \\ S_{vh} & S_{vv} \end{bmatrix}, \quad (1)$$

where E_h^t , E_v^t are the transmitted and E_h^r , E_v^r are the received electric fields. Hence, assuming a monostatic system configuration where it holds that $S_{hv} = S_{vh}$ (using reciprocity theorem), the target in PolSAR data is represented by five distinct parameters: (i) three absolutes consisting of co-polarized intensities including $|S_{hh}|$, $|S_{vv}|$ and one of the cross-polarized intensities as $|S_{hv}|$ or $|S_{vh}|$ and (ii) two relative phases with ϕ_{hv-hh} and ϕ_{vv-hh} . Correspondingly, it follows

$$\begin{bmatrix} E_h^r \\ E_v^r \end{bmatrix} = \begin{bmatrix} |S_{hh}| e^{j\phi_{hh}} & |S_{hv}| e^{j\phi_{hv}} \\ |S_{vh}| e^{j\phi_{vh}} & |S_{vv}| e^{j\phi_{vv}} \end{bmatrix} \begin{bmatrix} E_h^t \\ E_v^t \end{bmatrix}. \quad (2)$$

Additionally, the total scattering power is used as another target descriptor in the classification algorithms. The total scattering power, the so-called Span, can be computed by $\text{Span}(\mathbf{S}) = \text{Tr}(\mathbf{S}\mathbf{S}^{*T}) = |S_{hh}|^2 + |S_{vh}|^2 + |S_{hv}|^2 + |S_{vv}|^2$. Note the fact that the matrix \mathbf{S} has the complete information of the observed target; however, in practice, the availability of these backscattering parameters is determined by the actual polarization of the utilized acquisition system.

2.2 PolSAR Information Extraction

The scattering mechanisms of various terrain types can be better characterized using the PolSAR data. For example, this better characterization advantage is studied by Lee et al. [37] and the scattering mechanisms for different target categories can be summarized as follows: human-made objects including vehicles and buildings tend to have specular and double bounce scatterings, open-areas (e.g., consisting of lands without structures) have surface scattering, and bushes and trees demonstrate volume scattering.

One can utilize the second-order representations of the backscattering \mathbf{S} matrix to exploit the scattering information. Because of the random scattering and speckle noise, the PolSAR data are obtained through multilooking; and accordingly, the second-order target descriptors consisting of average coherency \mathbf{T} and covariance \mathbf{C} matrices are computed by averaging over n number of looks as follows,

$$\mathbf{T} = \frac{1}{n} \sum_{i=1}^n \mathbf{k}_i \mathbf{k}_i^{*T} \quad \text{and} \quad \mathbf{C} = \frac{1}{n} \sum_{i=1}^n \mathbf{\Omega}_i \mathbf{\Omega}_i^{*T}, \quad (3)$$

where \mathbf{k}_i and $\mathbf{\Omega}_i$ are the Pauli-based scattering and lexicographic basis for the i th look, respectively:

$$\begin{aligned} \mathbf{k}_i &= [S_{hh} + S_{vv}, S_{hh} - S_{vv}, 2S_{hv}]^T / \sqrt{2} \\ \mathbf{\Omega}_i &= [S_{hh}, \sqrt{2}S_{hv}, S_{vv}]^T. \end{aligned} \quad (4)$$

One intuitively can say that both the coherency and covariance matrices have the equivalent information regarding the polarimetric description since their linear transformation from one to the other is possible and they are both Hermitian positive definite 3×3 matrices.

In general, the collected features in PolSAR classification can be grouped in two categories: (i) the obvious features from the backscattering matrix elements

and the second-order descriptors including the coherency \mathbf{T} and covariance \mathbf{C} matrices. Apparently, the first group contains the aforementioned low-level EM features from (2) and (3). On the other hand, the second group (ii) contains the computed features using different TD theorems including the coherent and incoherent decompositions.

2.2.1 Coherent Target Decompositions

In the coherent TDs, the matrix \mathbf{S} is tried to be represented as the sum of coherent scattering responses such that

$$\mathbf{S} = \sum_{i=1}^k \alpha_i \mathbf{S}_i. \quad (5)$$

Basically, even though the \mathbf{S} matrix has the complete information regarding the observation, in PolSAR classification, more discriminative features can be obtained with the computed α_i coefficients using alternative representations. The following TDs can be included in this group: Pauli decomposition, Krogager decomposition [38], and Cameron decomposition [39].

For example, in Pauli decomposition, \mathbf{S} is represented by the Pauli basis matrices as follows:

$$\mathbf{S} = \begin{bmatrix} S_{hh} & S_{hv} \\ S_{vh} & S_{vv} \end{bmatrix} = \alpha_1 \mathbf{S}_1 + \alpha_2 \mathbf{S}_2 + \alpha_3 \mathbf{S}_3, \quad (6)$$

where $\mathbf{S}_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$, $\mathbf{S}_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$, and $\mathbf{S}_3 = \frac{1}{\sqrt{2}} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ with the corresponding representations coefficients as $\alpha_1 = (S_{hh} + S_{vv})/\sqrt{2}$, $\alpha_2 = (S_{hh} - S_{vv})/\sqrt{2}$, and $\alpha_3 = \sqrt{2}S_{hv}$. One can note that the concatenation of these coefficients form the Pauli-based scattering vector in (4). Then, Pauli-based features are obtained by the squared-amplitudes: $|\alpha_1|^2$, $|\alpha_2|^2$, and $|\alpha_3|^2$. These polarimetric components are also used to illustrate a PolSAR image, an pseudo-colored image can be obtained by assigning these polarimetric components to RGB components. As another example to coherent TDs, the Krogager decomposition tries to represent distinct scattering mechanisms independent from the incident angle as follows,

$$\mathbf{S} = e^{j\phi} \left(e^{j\phi_s} k_s \mathbf{S}_s + k_D \mathbf{S}_{D(\theta)} + k_H \mathbf{S}_{H(\theta)} \right), \quad (7)$$

where for an incident θ angle, three PolSAR coherent components are obtained as k_s , k_D , and k_H by decomposing \mathbf{S} using sphere, diplane, and helix bases.

2.2.2 Incoherent Target Decompositions

The incoherent TDs try to exploit the distributed scatters by decomposing the second-order target descriptors including \mathbf{T} and \mathbf{C} matrices; some of the incoherent TDs are $H/\alpha/A$ or eigenvector-eigenvalue (Cloude-Pottier) decomposition [40], Freeman decomposition [41], and Huynen decomposition [42].

For example, in Cloude-Pottier decomposition, the eigenanalysis is performed over the coherency matrix \mathbf{T} as follows:

$$\mathbf{T} = \mathbf{U}_3 \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix} \mathbf{U}_3^{*T} \quad (8)$$

with the following orthonormal eigenvectors:

$$\mathbf{U}_3 = \begin{bmatrix} \cos \alpha_1 & \cos \alpha_2 & \cos \alpha_3 \\ \sin \alpha_1 \cos \beta_1 e^{i\delta_1} & \sin \alpha_2 \cos \beta_2 e^{i\delta_2} & \sin \alpha_3 \cos \beta_3 e^{i\delta_3} \\ \sin \alpha_1 \sin \beta_1 e^{i\gamma_1} & \sin \alpha_2 \sin \beta_2 e^{i\gamma_2} & \sin \alpha_3 \sin \beta_3 e^{i\gamma_3} \end{bmatrix}. \quad (9)$$

Cloude-Pottier's decomposition provides several polarimetric parameters including the following average set of four angles: $f = [\bar{\alpha} \ \bar{\beta} \ \bar{\delta} \ \bar{\gamma}]$, the anisotropy A , and the entropy H . They are defined as follows for real eigenvalues $\lambda_1 > \lambda_2 > \lambda_3$:

$$\begin{aligned} \bar{\alpha} &= \sum_{i=1}^3 p_i \alpha_i, & \bar{\beta} &= \sum_{i=1}^3 p_i \beta_i, & \bar{\delta} &= \sum_{i=1}^3 p_i \delta_i, & \bar{\gamma} &= \sum_{i=1}^3 p_i \gamma_i, \\ A &= \frac{p_2 - p_3}{p_2 + p_3}, & H &= - \sum_{i=1}^3 p_i \log_3 p_i, \end{aligned} \quad (10)$$

where $p_i = \frac{\lambda_i}{\sum_{i=1}^3 \lambda_i}$. Here, the randomness is defined by the entropy H parameter. Accordingly, for a randomly observed target, it is $H = 1$ with equal eigenvalues. On the other hand, as stated in [43], for the cases where the entropy is small enough, the first eigenvector is sufficient to describe the observed target. In practice, one is likely to describe the partial target whose scattering is between isotropic and random (i.e., $H = 0$ and $H = 1$, respectively). It is investigated in [44] that such partial scatterings can be described by the averaged $\bar{\alpha}$ angle in (10) as single-bounce, dipole, and double bounce scatterings correspond to $\bar{\alpha} \approx 0$, $\pi/4$, and $\pi/2$, respectively.

2.3 *Prior Work*

In the following, existing studies on SAR and PolSAR classification will be discussed.

2.3.1 **Classification of Partially Polarized SAR Data**

In a classification framework, one can directly use the backscattering elements in (2). However, SAR data may not have such full polarization and one can have single or dual-polarized SAR data; in this case, the information is limited regarding the observed target. Such degradation decreases the classification performance as it is observed in many studies [45–48]. Therefore, the classification approaches for partially polarized data tend to improve the performance using an excessive number of features. Among the traditional ML-based approaches, the study in [21] proposes to integrate different EM features with the texture and color features. One can say that the previous approaches have used only pixel-wise information in the classification and neglected the correlation within small neighborhoods, whereas the same method in [21] has also utilized region-based feature extraction procedure and they have obtained an elegant classification accuracy using only less than 0.1% of the whole SAR data in the training. The collected features in [21] include the elements of \mathbf{S} and the following additional image processing features: (i) texture features: the Edge Histogram Descriptor (EHD) [49], Local Binary Pattern (LBP) [50], Gray-Level Co-occurrence Matrix (GLCM) [51], and Gabor wavelets [52] and (ii) color features: MPEG-7 Color Structure Descriptor (CSD) [53], MPEG-7 Dominant Color Descriptor (DCD) [49], and hue-saturation-value color histogram [53]. Accordingly, to compute the abovementioned image processing features over the single- and dual-polarized SAR data, several different methodologies are followed to obtain pseudo-colored RGB and gray-level images in [21]. Correspondingly, color features for dual-polarized SAR data are extracted over two RGB images produced using the available magnitudes of two backscattering elements of \mathbf{S} such that assigning $|S_{vh}|$, $|S_{vh}| - |S_{vv}|$, $|S_{vv}|$ and $|S_{vv}|$, $|S_{vv}| - |S_{vh}|$, $|S_{vh}|$ to R, G, and B channels, respectively. However, there is only one intensity SAR data available for the single-polarized case. In this case, the pseudo-colored image is produced by assigning intensity values to corresponding pre-determined Hue, Saturation, and Intensity (HSI) values as followed in [54]. Next, the texture features are obtained over the gray-level images produced by calculating the total scattering power $\text{Span}(\mathbf{S})$ for the dual-polarized data and directly using the available backscattering element for the single-polarized data. Then, an ensemble of traditional classifiers is built to perform classification over the extracted features and they are trained together to maximize the classification performance.

2.3.2 PolSAR Data Classification

There are various traditional approaches for PolSAR classification such as [22, 25, 55, 56] that are based on, for example, Support Vector Machines (SVMs) and Random Forest (RF). In general, utilized PolSAR features in these methods can be grouped into two categories; in the first category, features collected from the backscattering \mathbf{S} matrix and the second-order descriptors including \mathbf{T} and \mathbf{C} matrices, whereas in the second category, the extracted features are based on the different TDs. Hence, contrary to the previously discussed approaches proposed for the partially polarized SAR data, the PolSAR classification approaches additionally utilize different combinations of the high-level EM features that are extracted using various TDs.

The DL methods for PolSAR classification constitute the new trend using deep CNNs. For example, in [30], it is proposed to collect the real-valued diagonal elements and the magnitudes of the complex non-diagonal elements of the matrix \mathbf{T} in order to form a 6-dimensional real-valued input for a deep CNN classifier. The study in [32] has used a considerably deep model consisting of 11 layers. On the other hand, the proposed deep CNN in [32] takes only the 3-dimensional input that is formed by the Pauli decomposition $[\alpha_1, \alpha_2, \alpha_3] = [S_{hh} + S_{vv}, S_{hh} - S_{vv}, 2S_{hv}]^T / \sqrt{2}$. Finally, another study in [31] has proposed using a deep CNN consisting of two separate branches. The two branches are trained jointly over separate inputs: the first branch's input is similar to 6-dimensional input in [30] and the second branch is trained using 3-channel pseudo-colored RGB images. In the end, they form a single-stage classification framework for the inference.

Nevertheless, in order to achieve adequate classification accuracies, the aforementioned DL-based methods demand a great amount of training data which undermines the main purpose of using such automatic classification frameworks. Furthermore, considering the remaining limited amount of test data implies that the deep CNN-based methods have not been evaluated over a significant proportion of the PolSAR data; and hence, one can say that the reliability issues may occur on the achieved classification accuracies by deep CNNs. To this end, the method in [22] has achieved satisfactory classification accuracies using only a small number of training samples (e.g., less than 0.1% of the PolSAR data) and with the reduced computational time complexity compared to DL-based approaches. Accordingly, the method in [22] uses similar color and texture feature extraction techniques as followed in [21]. In addition, the EM features are enriched thanks to the coherent and incoherent TDs. Finally, an ensemble of traditional classifiers, which is also similar to [21], is trained in [22] over the extracted features.

3 Methodology

The classification methodologies proposed in [19, 20] are illustrated in Fig. 1. For the illustration purposes in the figure, the pseudo-colored image for partially polarized SAR data is created by the previously explained HSI color-space transformation using the available intensity at X-band over Po Delta in Italy; and the PolSAR image of San Francisco at C-band (SFBay_C) is created by assigning the diagonal elements of the \mathbf{T} matrix: T_{33} , T_{22} , and T_{11} to R, G, and B channels, respectively. In the proposed classification framework, pixel-wise classification is performed using an $N \times N$ sliding window over the EM channels. Accordingly, the replaced $N \times N$ window at the center of the i th pixel is fed to the adaptive and compact 2D-CNN and its corresponding class prediction determines the predicted label of the i th pixel. This procedure is performed for each pixel in the image and the final produced LU/LC mask is obtained as illustrated in Fig. 1. Obviously, the number of weight kernels connecting the input layer to the first hidden convolutional layer is equal to the number of input EM channels. In the study of [19], it has varied between one to four for partially polarized SAR data, whereas it has varied between three to six for fully polarimetric SAR data. Next, another hyper-parameter in the proposed approach [19, 20] is the size of the sliding window operator. In general, if the classifier is based on deep CNNs, the size (N) is needed to be kept high. On the other hand, in the proposed approach using small window sizes is possible, e.g., 5×5 or 7×7 . The advantage of using such small sizes will be discussed in Sect. 4. In the following, the adaptive CNN implementation will be detailed.

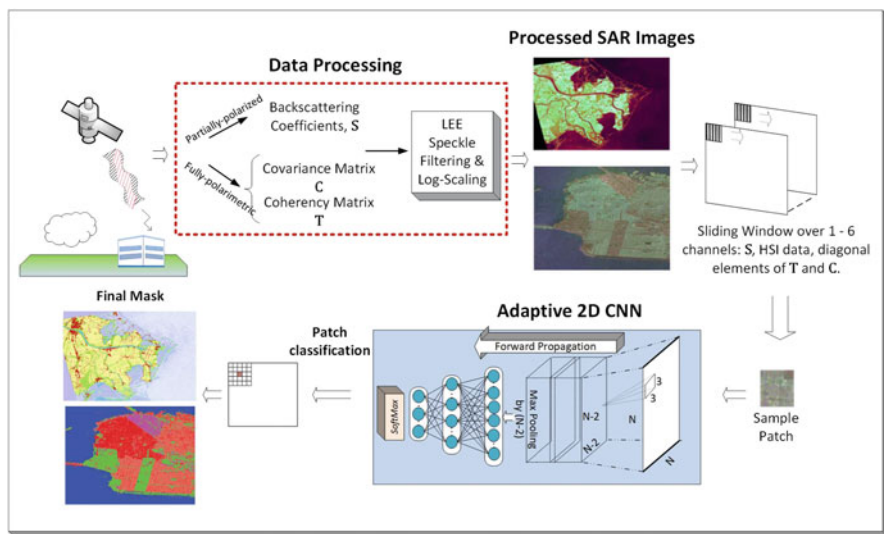


Fig. 1 The proposed SAR classification framework in [19, 20] using compact CNNs based on sliding window pixel-wise classification

3.1 Adaptive CNN Implementation

Adaptive CNNs have several novel modifications in order to obtain an adaptive structure in the network and simplify the neuron representation. To begin with, the network architecture consists of only two different layer types besides the input and output layers: (i) *CNN* layers encapsulating the conventional convolutional and sub-sampling layers and (ii). Multi-Layer Perceptron (MLP) or fully connected layers. In this way, every neuron of the CNN layers can perform convolution and down-sampling operations. Accordingly, the intermediate output of a CNN neuron is sub-sampled to get the final output of that neuron. Then, all of the final computed output maps of the corresponding layer are convolved with their weight kernels and they are further summed to form the inputs for the neurons in the next layer. In the modified CNN layers, the CNN parameters are independent of the image dimension of the input layer. The intermediate output of the k th neuron at layer l , y_k^l , is sub-sampled to produce the final output s_k^l of that particular neuron. The final output maps for each neuron are convolved with their corresponding weight kernels and the results are summed to produce the input map of the next layer neuron as follows:

$$\mathbf{x}_k^l = \sum_{i=1}^{N_{l-1}} \text{conv2D}(\mathbf{w}_{i,k}^{l-1}, \mathbf{s}_i^{l-1}, \text{'NoZeroPad'}) + b_k^l. \quad (11)$$

Thanks to the adaptivity, there are introduced two properties with the proposed approach [19, 20]. First, the user can select different sliding window sizes N and there is no need to perform further modification in the network structure. This is especially important for the SAR classification since the window size is generally dependent on the SAR data (e.g., resolution of the data) and the designed experimental setup. Second, the number of CNN layers can be set to any number independent of the size of the input (N). This is possible in the implementation since the sub-sampling pooling size is adjusted automatically based on the output feature map size of the last CNN layer, i.e., just before the first fully connected layer. In this way, the feature maps are sub-sampled accordingly to produce scalar output values for the input of the first fully connected layer. For instance, let the last hidden convolutional layer has the feature map dimension of 8×8 . Then, the feature maps are sub-sampled with a pooling size of 8. Also note the fact that besides the sub-sampling, the feature map dimensions are reduced because of the convolution operation without zero-padding in the proposed CNN configuration. Therefore, after convolution with the kernel sizes of (K_x, K_y) , the feature map size would be decreased by $(K_x - 1, K_y - 1)$ where the width and height are K_x and K_y for the convolution kernels.

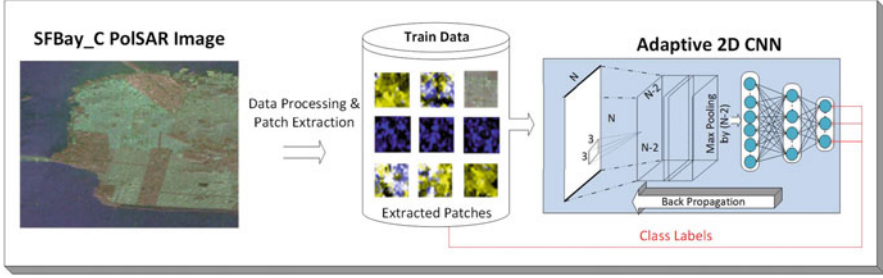


Fig. 2 The training technique for the proposed adaptive 2D CNNs including the data processing, patch extraction, and the BP iteration to update the network parameters

3.2 Back-Propagation for Adaptive CNNs

The Back-Propagation (BP) procedure is illustrated in Fig. 2 for the proposed adaptive CNNs. Given a classification problem with N_L classes, first, using 1-of- N_L encoding procedure, the labels are converted to the target class vectors. Next, for each training pixel, patches are cropped by placing the $N \times N$ window at the center. Let the corresponding true/target and predicted class vectors be $[t_1, t_2, \dots, t_{N_L}]$ and $[y_1, y_2, \dots, y_{N_L}]$, respectively. The BP is initiated by computing the derivative of the classification error at the output layer with respect to its training parameters including the weights and biases. Accordingly, the mean-squared error is computed for the classification error:

$$\mathcal{L} = \mathcal{L}(y_1^L, y_2^L, \dots, y_{N_L}^L) = \sum_{i=1}^{N_L} (y_i^L - t_i)^2. \quad (12)$$

The BP formulation for the MLP layers is already well-studied. For the CNN (convolutional) layers, the BP formulation consists of four different parts including inter-BP among CNN layers, intra-BP within a CNN neuron (that includes convolution and sub-sampling operations), BP from the first MLP layer to the last convolutional layer, and finally the computations of the kernel (weights) and bias sensitivities. The readers are referred to [21] for the complete formulation of these four steps of one BP iteration.

4 Experimental Results

In this section, the experimental setup is first presented together with the utilized benchmark SAR data in the studies of [19] and [20], respectively. Then, the classification performance of the proposed approach of [19, 20] using the adaptive and compact CNNs will be compared against the methods in [21, 22], and several

other recent methods. The quantitative analysis to evaluate the classification performance will be performed by comparing the class-specific and overall classification accuracies. In SAR classification, one can say that the visual evaluation has also significance along with the quantitative analysis. Hence, the visual evaluation will be carried out by inspecting the final produced segmentation masks (i.e., qualitative analysis) by the methods for each study site. Next, the sensitivity of the methods will be investigated against the two hyper-parameters: the number of input channels and the sliding window size N . There are also overall classification accuracy comparisons between the proposed approach and two recently proposed deep CNNs: Xception [34] and Inception-ResNet-v2 [35]. Finally, this section will be concluded by another sensitivity analysis regarding the number of layers and neurons to show that the compact CNN architectures improve the classification performance compared to their deep counterparts.

4.1 Benchmark SAR Data

The experimental evaluations are performed using four different study sites and six SAR images. The details for each benchmark data are presented in Table 1; correspondingly, there are two partially polarized and four fully polarized SAR images. The single-polarized PDelta_X image covers the Po Delta region located in the Northeast of Italy. The Dresden_X image has dual-polarization and mainly covers the Dresden area in the Southeast of Germany. From the four PolSAR images, two of them are acquired by an air-borne system, whereas the rest of the images are all space-borne. The first set of the PolSAR images cover the San Francisco Bay area in California, USA. The second set consists of the Flevoland area of the Netherlands. In general, these images have been widely used as benchmark data in PolSAR classification approaches. The image dimensions, the number of training samples, and the ground-truth data (GTD) size are presented in Table 2.

Table 1 The benchmark SAR images that are utilized in the experiments consisting of partially polarized and fully polarized SAR data

Study site	System and band	Abbreviation	Date	Incident angle	Mode
Po delta	COSMO-SkyMed, X-band	PDelta_X	Sep 2007	30°	Single
Dresden	TerraSAR-X, X-band	Dresden_X	Feb 2008	41–42°	Dual
SF Bay	AIRSAR, L-band	SFBay_L	1988	10–60°	Quad
SF Bay	RADARSAT-2, C-band	SFBay_C	Apr 2008	30°	Quad
Flevoland	AIRSAR, L-band	Flevo_L	Aug 1989	40–50°	Quad
Flevoland	RADARSAT-2, C-band	Flevo_C	Apr 2008	30°	Quad

Table 2 Dimension, the number of training samples per class, the total number of samples in the ground-truth data (GTD), and the number of classes are presented for each SAR data

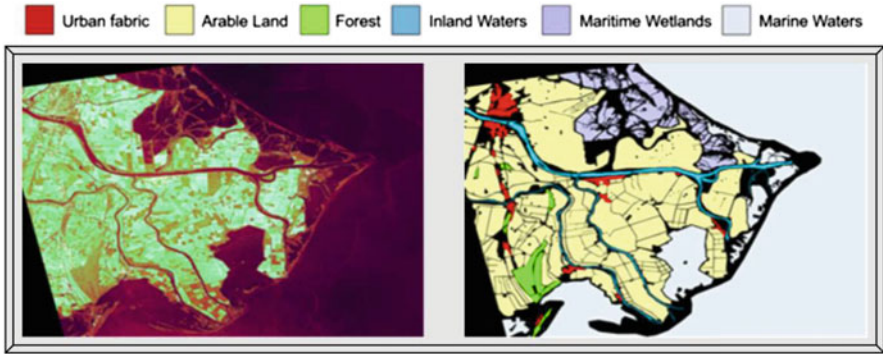
Data	Dimension	Train samples per class	Total GTD	# classes
Po Delta	4642×3156	2000	612,000	6
Dresden	2209×3577	1000	606,000	6
SF Bay	900×1024	~ 292	123,459	5
SF Bay	1426×1876	500	252,500	5
Flevoland	750×1024	120–480	209,979	15
Flevoland	1639×2393	500	202,000	4

4.1.1 Po Delta, COSMO-SkyMed, X-band (PDelta_X)

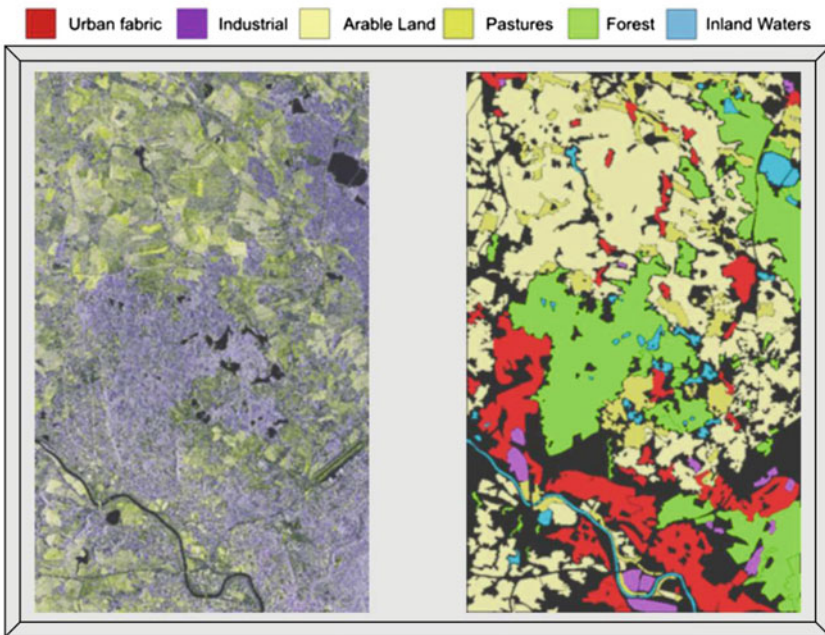
The benchmark PDelta_X single-polarized SAR data have mainly the natural terrain types and also several different types of water classes. The image has only one polarization: HH in Strip Map HImage mode. Originally, it has 16716×18308 image dimension with a 3-m resolution, but because of the computational burden, the PDelta_X is downscaled by 3.6×5.8 in the approach of [21]. Thus, in the proposed approach of [19], the same downscaling factor is used to downscale the PDelta_X image to make a fair comparison possible using the same GTD with the previous studies. The GTD is collected by visually inspecting the optical image data and utilizing [57]. Accordingly, the data includes soil-vegetation classes and several different water-based terrains as the natural target, and the human-made targets with different structures are grouped into a single category. The selected six classes are urban fabric, arable land, forest, inland waters, maritime wetlands, and marine waters. The used GTD are the same as the one in [21]. By assigning each class label to a particular RGB value, the constructed terrain map is presented in Fig. 3 along with the produced pseudo-colored PDelta_X image. The same sample points are selected as in [21] to make an appropriate comparison with the competing method. The selected training samples correspond to 1–2% of the total GTD size (2000 sample pixels for each class) and 0.08% of the whole data.

4.1.2 Dresden, TerraSAR-X, X-band (Dresden_X)

The Dresden_X has 4419×7154 pixels with an approximate 4×4 square meters pixel resolution. Due to the aforementioned reason for the PDelta_X, the Dresden_X SAR image is also downscaled by 2×2 . It is dual-polarized SAR data with VH/VV and it is acquired in Strip Map mode and radiometrically enhanced (RE) Multi-look ground range detected (MGD) having the effective number of looks as 6.6. Accordingly, the coordinates are projected to the ground range in the MGD mode and each pixel is represented by the magnitudes (S_{vh} and S_{vv} elements in (2)) where the phase information is lost. Expectedly, the RE and MGD modes provide speckle noise reduction. Similarly, the same GTD of Dresden_X in [21] are also used in this study. The GTD are manually collected as previously discussed by utilizing [57].



(a) PDelta_X SAR image.



(b) Dresden_X SAR image.

Fig. 3 The pseudo-colored image is produced for the single-polarized PDelta_X SAR data and its corresponding GTD are shown with the terrain class labels in (a). In (b), the GTD are shown for the dual-polarized Dresden_X SAR data and its pseudo-colored image is produced by assigning $|S_{vv}|$, $|S_{vv}| - |S_{vh}|$, and $|S_{vv}|$ to R, G, B, channels, respectively

Accordingly, there are labeled six classes including human-made terrains: urban fabric and industrial and natural terrains: arable land, pastures, forest, and inland waters. The GTD are presented in Fig. 3 together with the pseudo-colored image. For the Dresden_X, the chosen train/test ratio is 0.01 which corresponds to 1000 sample pixels for the training and 100 000 pixels for the testing per class as in [21].

4.1.3 San Francisco Bay, AIRSAR, L-band (SFBay_L)

The benchmark SFBay_L PolSAR data have mostly urban terrain types and several natural classes. The data are fully polarized and have a 900×1024 dimension with a 10×10 pixel resolution in meters. The GTD for SFBay_L are collected by selecting regions and sample points by investigating the pseudo-colored image and aerial photographs from the TerraServer Web site [58]. The collected GTD are identical to the ones that are used in various previous studies such as [4, 23, 59]. The chosen five classes are water, urban, forest, bare soil, and natural vegetation including, e.g., woodland and scrub. Similar class types are followed by the studies in [60–63]. The selected training and testing regions are illustrated in Fig. 4. Correspondingly, only 1–2% of the GTD ($\sim 0.1\%$ of the overall data) are used in the training as followed in [22].

4.1.4 San Francisco Bay, RADARSAT-2, C-band (SFBay_C)

The SFBay_C PolSAR data have similar terrain types with SFBay_L, but the latter is air-borne whereas the SFBay_C data are space-borne with a 10×5 square-meter resolution. In the experiments, the same sub-region with a size of $\sim 1400 \times 1800$ and the GTD are used by many other approaches such as [22, 60, 64–66]. Similarly, 1:100 proportion is used to create the train:test splits corresponding to 500:50 000 number of samples from each class.

In the SFBay_C, there are three major terrain types: human-made, water, and vegetation. In fact, the SFBay_C PolSAR data are important in the performance evaluation especially for the human-made objects since the human-made terrain traces are further grouped within three distinct categories based on their inclusion with the natural terrain including high-density urban, low-density urban, and developed class types. One can say that the class overlap between these three categories is expected, but it is observed in Fig. 4 that there are visible differences between them. For example, the high-density urban and low-density urban classes mainly include the samples from the urban areas, where the areas are better congested in the high-density urban class with better concentrated human-made structures, e.g., compared to the low-density urban terrain type. On the other hand, the samples from the developed class belong to the relatively sparse human-made objects combined with the vegetation compared to other urban class categories. Thus, these three classes are considered distinct classes and they should be separable by the classifier. Considering that the provided GTD have some hierarchical structure within the urban class types, the classification of SFBay_C may be considered as a hierarchical classification problem. Nevertheless, as specified by Uhlmann and Kiranyaz [22], the GTD accuracy is not a hundred percent guaranteed as other human-made classes can also have some sample pixels belonging to trees and different plant types, for example, from the gardens of houses.

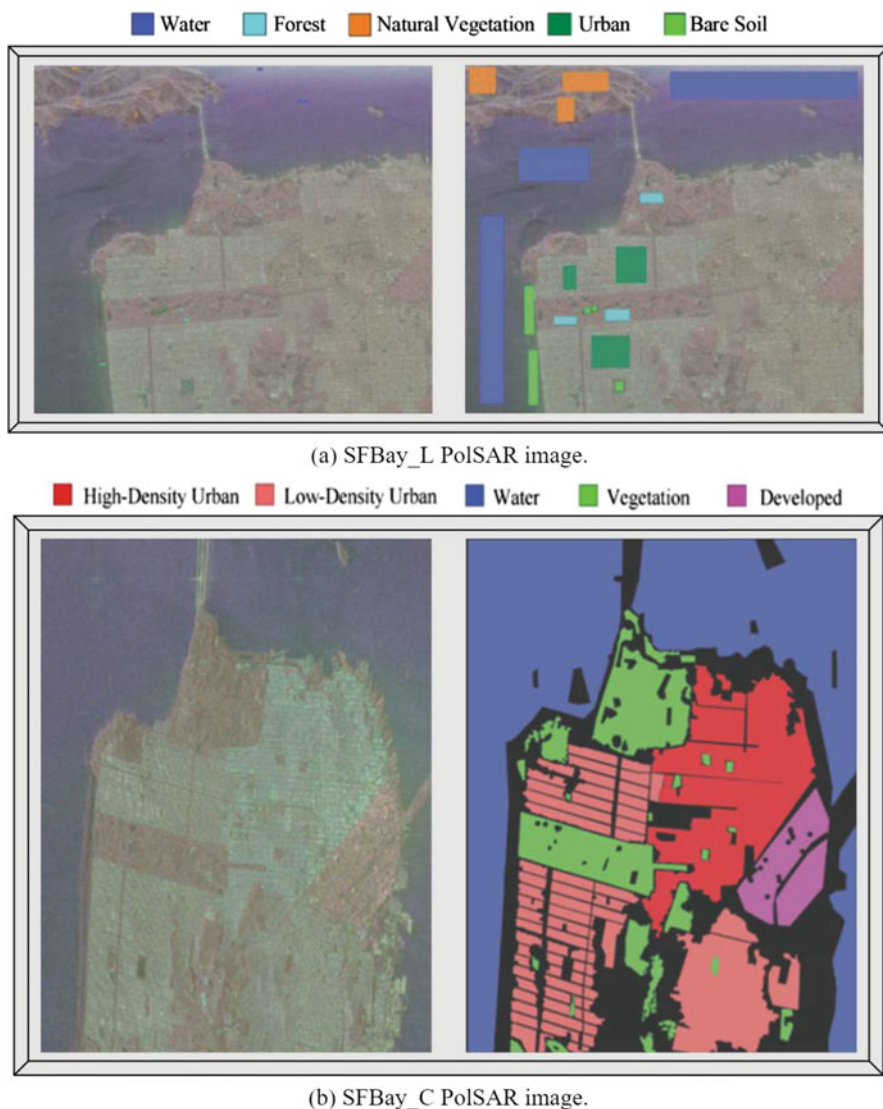


Fig. 4 The pseudo-colored images are produced for the PolSAR SFBay_L and SFBay_C images with their corresponding GTD in (a) and (b), respectively

4.1.5 Flevoland, AIRSAR, L-band (Flevo_L)

The Flevo_L PolSAR data were acquired in the mid-August of 1989 within the MAESTRO-1 Campaign. The data have 12×6 square-meter resolution with the dimension of 750×1024 . The Flevoland scene, in general, is used in many land and crop classification applications since it consists of a large vegetation field, different

soil types, and a small area of water and human-made structures. In particular, the Flevo_L fully polarimetric data have been widely used and the scene has a well-established corresponding GTD including 15 classes obtained by Yu et al. [67]: water, peas, rapeseed, lucerne, barley, beet, potatoes, building, forest, grass, stem beans, bare soil, and three different types of wheat. The GTD information is provided in Fig. 5. As the number of classes is larger compared to other benchmark study sites used in this study, the number of training samples for Flevo_L is varied between 120 and 480 per class to evaluate the training size effect on the classification performance.

4.1.6 Flevoland, RADARSAT-2, C-band (Flevo_C)

The Flevo_C PolSAR data that were acquired in April 2008, have 10×5 square meters resolution. Even though this study site has a very similar terrain cover with the Flevo_L, the number of classes is significantly smaller and it has a larger emphasis on the human-made terrain types. The selected sub-region has the dimension of $\sim 1600 \times 2400$ pixels and it is illustrated in Fig. 5 along with the corresponding GTD. The presented GTD are identical in the study of [22] and proposed by Yang et al. [68] including four classes water, urban, forest, and cropland. In the experiments with Flevo_C, the 1:100 ratio is chosen according to [22] for the train:test splits.

4.2 Experimental Setup

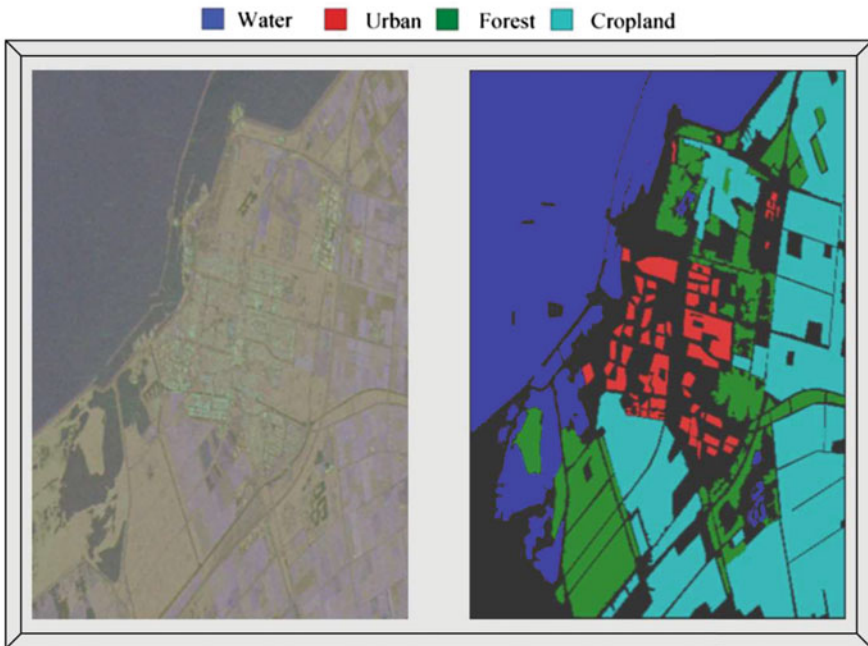
Among the partially polarized SAR images, only a single-polarized PDelta_X image is speckle filtered to reduce the speckle noise since the Dresden_X data are already acquired in the RE MGD mode. However, all four PolSAR images have been filtered using the polarimetric speckle filtering approach in [69] with a 5×5 window size even though they have been multi-look averaged over a four look data. Such filtering procedure for the study sides provides an appropriate classification performance comparison with the existing methods in [21–23].

The hyper-parameters are selected based on the validation split collected by using 50% of the training samples for the competing methods including traditional classifiers with SVM and CNBC approaches and the proposed approach [19, 20] based on the adaptive and compact CNNs. The search-space for the SVM hyper-parameters are presented for each feature combination as follows: (i) the kernel function among linear, third order polynomial, radial basis function (RBF), and sigmoid, (ii) the gamma parameter (except for the linear kernel function), $\gamma = 1/2^n$ where $n = \{1, 2, 3, 4\}$, and iii. the regularization parameter, $C = 1/2^n$ with $n = \{0, 1, 2, 3\}$.

The implementation of the proposed adaptive CNNs is performed using C++ over MS Visual Studio 2015 in 64 bit. The multithreading is achieved using Intel®



(a) Flevo_L PolSAR image.



(b) Flevo_C PolSAR image.

Fig. 5 The pseudo-colored images are produced for the PolSAR Flevo_L and Flevo_C images with their corresponding GTD in (a) and (b), respectively

OpenMP API. On the other hand, the implementation of the compact CNNs is based on a non-GPU implementation. The SVM and CNBC classifiers are also implemented using C++ but with MS Visual Studio 2013 in 32 bit. The experimental

evaluations of the proposed approach and the traditional classifiers have been carried out on a computer having i7-4790 CPU with 3.6 GHz and 16 Gb system memory. The implementation of the deep CNNs: Xception and Inception-ResNet-v2 are performed with TensorFlow [70] and Keras [71] on Python using a Nvidia® TITAN-X GPU card.

The followed network configurations in the proposed approaches of [19, 20] are as follows: there is only a single convolutional layer having 20 weight kernels (neurons) with $K_x = K_y = 3$ kernel dimensions and the sub-sampling factor is adjusted automatically based on the input size as discussed before. Finally, there is only one hidden MLP layer having 10 neurons connecting the pooled features to the output layer. The activation functions of the layers are set to the hyperbolic tangent function.

The neurons of the input layer have been fed by the patch (cropped by the sliding window) of the particular EM channel as illustrated in Fig. 1. Hence, the number of input neurons (weight kernels) is determined by the number of channels. As previously discussed, for the partially polarized data, the number of channels is changed between 1 and 4 depending on the data. In a one-channel CNN setup, the available SAR intensity is directly used to classify single-polarized data. In the second setup, the HSI channels are included and a four-channel setup is designed. It is observed in Sect. 4 that the addition of HSI channels increases the classification accuracy achieved by the one-channel setup. Next, a two-channel setup is used to classify dual-polarized SAR data by using two available intensities as the input of the CNN layer. Next, to classify PolSAR data, there are three different setups followed by Ahishali et al. [20] using the proposed approach: (i) the CNN takes only the diagonal elements of the \mathbf{T} matrix as the input, (ii) $\text{Span}(\mathbf{S})$ is included in the input layer with a four-channel setup, and (iii) a six-channel input is obtained by using the diagonal elements of the \mathbf{T} and \mathbf{C} matrices.

Overall, thanks to the followed compact architecture of the proposed CNN configuration, the over-fitting does not occur even though the training data size is limited. Thus, as the only early stopping procedure during the training, the maximum numbers of epochs (training iterations) are presented in Table 3 selected for each study site. It is observed that especially for SFBay_L data, the compact CNNs converge faster within only 40 epochs and achieve greater than 99% classification accuracy which will be shown in Sect. 4.3. In the training, dynamic learning rate adaptation is followed for the adaptive CNNs: the learning rate ϵ is initially set to 0.05, then it is changed after each epoch based on the computed MSE over the training samples. Accordingly, it is decreased by 30% in the next iteration if the train MSE is larger than the previous iteration; otherwise, it is increased by 5%.

In the training procedures of the Xception and Inception-ResNet-v2, the networks are first trained from the randomly initialized weights using the stochastic gradient descent and the following training parameters: momentum, batch size, and the initial learning rate ϵ are set to 0.9, 32, and 0.045, respectively. Then, ϵ for the next iterations is set to 0.94 times the previous iteration's ϵ in every second epoch. Using this experimental setup, the Xception has converged after 400 epochs,

Table 3 The maximum number (#) of iterations is presented for the early stopping procedure for the training phase of each SAR data

Data	# of iterations
PDelta_X	200
Dresden_X	200
SFBay_L	40
SFBay_C	400
Flevo_L	600
Flevo_C	400

whereas Inception-ResNet-v2 has required 160 epochs. Additionally, the pre-trained versions of the models that are trained over the ImageNet dataset are investigated for the SAR classification. Accordingly, in the proposed transfer-learning procedure, all layers except the MLP and the first convolutional layers of the Xception and Image-Resnet-v2 are kept frozen and the unfrozen layers are trained only for 25 epochs using the same training parameters. Next, all layers are fine-tuned for another 75 epochs using a smaller ϵ as 0.001 and the same decaying factor as 0.94 in every second epoch.

4.3 Results

The classification performance evaluations are carried out for the proposed approaches in [19, 20] over each study site including both partially polarized and fully polarized SAR data.

4.3.1 Results on PDelta_X

The sliding window size, N , is varied from 5×5 to 27×27 in order to explore the effect of N on the classification performance and on the quality of the final produced segmentation mask that is produced after classifying each single-pixel of the PDelta_X. Accordingly, the overall classification accuracies are presented in Table 4 obtained by the varied N and the number of input channels. It is observed that using only the backscattering $|S_{hh}|$ coefficient, the accuracy achieved by the proposed adaptive 2D CNN approach is superior to the previous state-of-the-art approach in [21] by a significant accuracy gap ($>10\%$) even though the compared approach utilizes large dimensional feature vector (greater than 200-D) including color and texture features in addition to $|S_{hh}|$. A fairer comparison can be made by comparing both approaches using the same information as the input data, i.e., only $|S_{hh}|$ coefficient. In this case, the accuracy gap surpasses 40% between the proposed approach [19] and the competing one.

Further performance improvement is possible by increasing the number of channels in the proposed approach. For example, in a four-channel setup using

Table 4 Overall classification accuracies are presented obtained by the proposed approach [19] over the PDelta_X using one and four-channel inputs and varied window sizes. The best accuracies are highlighted in bold

PDelta_X	One-channel	Four-channel
Window size	$ S_{hh} $	$ S_{hh} + \text{HSI channels}$
5×5	0.7098	0.7080
7×7	0.7482	0.7501
9×9	0.7698	0.7668
11×11	0.7890	0.7838
13×13	0.8075	0.8037
15×15	0.8147	0.8167
17×17	0.8276	0.8300
19×19	0.8387	0.8442
21×21	0.8404	0.8537
23×23	0.8480	0.8539
25×25	0.8487	0.8632
27×27	0.8533	0.8615

Table 5 The confusion matrix is presented over PDelta_X obtained by the proposed approach [19] using the best setup that provides the highest overall accuracy: four-channel ($|S_{hh}|$, Hue, Sat., and Int.) and $N = 25$. The correctly classified numbers of samples are highlighted in bold per class

		Predicted						
		Urban fab.	In. waters	Forest	Mart. wet.	Mar. waters	Ar. land	Total
True	Urban fab.	92,264	607	1322	54	0	5753	100,000
	In. waters	931	85,308	3824	6781	1210	1946	100,000
	Forest	934	2581	90,507	909	186	4883	100,000
	Mart. wet.	166	6153	1157	80,683	11,744	97	100,000
	Mar. waters	48	2196	166	17,502	80,067	21	100,000
	Ar. land	4680	1055	4875	253	52	89,085	100,000
	Total	99,023	97,900	101,851	106,182	93,259	101,785	600,000

the additional HSI components, the overall accuracy achieved by $N = 25$ is increased by about 1%. It is also observed that 25×25 window size provides the best classification accuracy and the classification performance does not improve by further increasing the window size after $N = 25$ for the four-channel setup. Nevertheless, a decent classification accuracy is always achievable with $>75\%$ by the adaptive CNNs using small N values such as $N \geq 9$. Additionally, the confusion matrix is provided obtained by the proposed approach in Table 5. Accordingly, the most confused classes are marine waters and maritime wetlands. In fact, these two classes are poorly distinguishable even by the human eye since they have comparable characteristics.

The classification accuracies of classes are shown in Fig. 6 to provide a more detailed comparison. Correspondingly, the classification performances of the classes

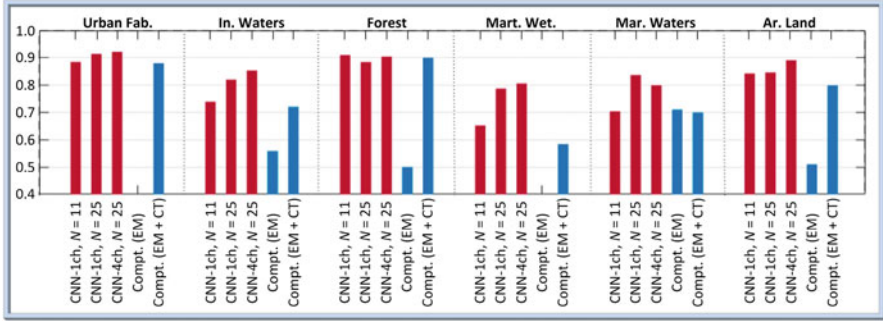


Fig. 6 The class-specific accuracies are presented by the proposed adaptive CNNs and the competing method [22] over the PDelta_X. The proposed approach in [19] uses a one-channel setup ($|S_{hh}|$) with $N = 11$ and 25 and a four-channel setup ($|S_{hh}|$ and HSI components) with $N = 25$, whereas the competing method uses $|S_{hh}|$ as features (EM) and in the second setup, color and texture features included (EM + CT) in the 208-D feature vector

are all improved with the proposed approach in [19]. Especially, substantial performance gaps occur for some class types such as inland waters, maritime wetland, marine waters, and arable land; considering the accuracy improvement of over 20% for the wetland class. Hence, it is shown that as discussed in the previous argument, manually extracted features provide poor discrimination power for some classes, whereas the adaptive proposed CNN is able to “learn-to-extract” features and always provide $>80\%$ accuracy for each terrain class type. Overall, one can say that the competing method in [21] has the reliability issue considering the poor classification performance for some classes.

The computed final segmentation masks and the overlaid regions over the corresponding GTD are shown in Fig. 7 for PDelta_X. In the quantitative analysis based on Table 4, it is shown that the larger N values provide better classification performance in terms of the overall accuracy. On the other hand, it is also revealed in Fig. 7 that the trade-off is valid between the quantitatively better results and the finer segmentation mask with a small N . For instance, even though the optimal window size is determined as 25×25 for the four-channel setup, the blocking artifacts occur in the computed masks. Nonetheless, a detailed segmentation mask can be obtained using $N = 11$ window size. The overlaid regions shown in Fig. 7 for the proposed approach can be compared with Fig. 8 having the overlaid regions that are obtained by the competing method. It is observed in Fig. 8 that the forest class samples are mostly misclassified as the urban fabric and the arable land samples as the urban fabric and forest classes in the competing method. In general, it is demonstrated that the classification performance of the classes is improved using the adaptive CNNs. This is further confirmed in the zoomed locations of Fig. 9, e.g., especially for the urban fabric, arable land, and forest class types and it is seen that the segmentation error (noise) has been greatly reduced.

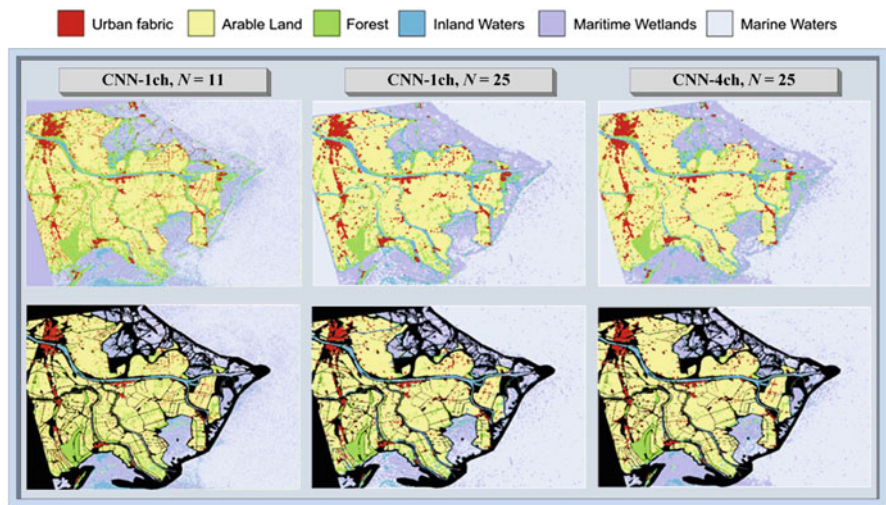


Fig. 7 The final segmentation masks are shown (first row) obtained by the proposed approach [19] over PDelta_X using one and four-channel input setups and different sliding window sizes (N). Their corresponding overlaid regions on the GTD are shown in the second row



Fig. 8 The overlaid regions of the segmentation masks are shown obtained by the competing method in [21] over PDelta_X. There are three different feature setups: (i) color features (C), (ii) color and texture features (CT), and (iii) EM and CT features

4.3.2 Results on Dresden_X

Similarly, there are six classes in the Dresden_X data. However, it has more human-made objects than the PDelta_X. In Table 6, overall classification accuracies are provided with varied N from 5 to 27. Accordingly, the best accuracy is achieved as 81.33% using 21×21 window size and utilizing only $|S_{vh}|$ and $|S_{vv}|$ components with the two-channel setup. The advantage of using small N is valid for the Dresden_X in quantitative analysis as well: the accuracy tends to decrease for $N > 21$. The competing method by Uhlmann and Kiranyaz [21] can achieve comparable accuracy only if the classifier uses color and texture features in addition to the backscattering components by a 209-D composite feature vector. In the

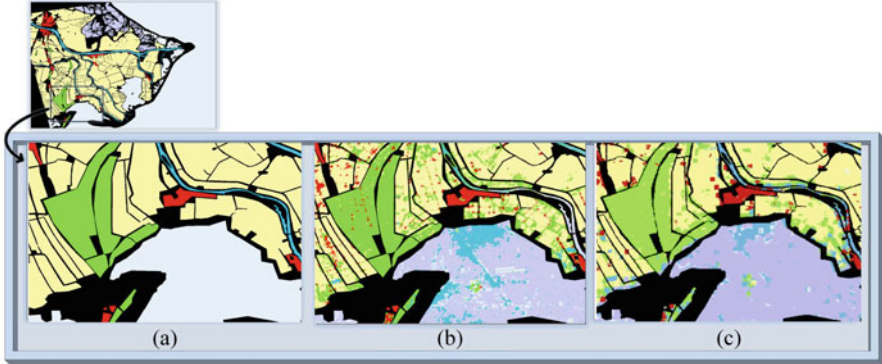


Fig. 9 The zoomed regions are for the PDelta_X: the GTD (a), overlaid regions obtained by the competing method (b) in [21], and the proposed method (c)

Table 6 Overall classification accuracies are presented obtained by the proposed approach [19] over the Dresden_X using one and four-channel inputs and varied window sizes. The best accuracies are highlighted in bold

Dresden_X	Two-channel
Window size	$ S_{vh} $ and $ S_{vv} $
5×5	0.7059
7×7	0.7509
9×9	0.7654
11×11	0.7797
13×13	0.7898
15×15	0.7980
17×17	0.8007
19×19	0.8105
21×21	0.8133
23×23	0.8029
25×25	0.8092
27×27	0.8062

case of when both approaches use identical information including only $|S_{vh}|$ and $|S_{vv}|$ components, there is a substantial accuracy gap larger than 30% between the proposed and competing approaches.

For the class-specific performance analysis, the confusion matrix is presented in Table 7. It shows that the pastures class is confused by the arable land and the urban fabric class is mostly confused by the industrial class. The latter was expected since the multi-label classification can be considered for the Dresden_X data as previously discussed. Similarly, the classification performance of each class is bar plotted in Fig. 10. The accuracy obtained by the proposed two-channel input is compared with the best accuracy levels achieved by the competing method [21] in which a 209-D feature vector is used for the classification. Based on Fig. 10, the proposed approach [19] produces better or similar accuracy except for the inland water class, but the

Table 7 The confusion matrix is presented over the Dresden_X obtained by the proposed approach of [19] using the best setup that provides the highest overall accuracy: two-channel ($|S_{vh}|$ and $|S_{vv}|$) and $N = 21$. The correctly classified numbers of samples are highlighted in bold per class

		Predicted						
		Urban Fab.	Industrial	In. waters	Forest	Pastures	Ar. land	Total
True	Urban Fab.	73,409	18,980	169	3323	1775	2344	100,000
	Industrial	16,492	78,870	172	1003	415	3048	100,000
	In. waters	1474	1192	93,012	1955	2182	185	100,000
	Forest	3081	1189	855	90,712	2193	1970	100,000
	Pastures	3961	1199	977	3863	73,175	16,825	100,000
	Ar. land	2895	1035	113	1337	15,802	78,818	100,000
	Total	101,312	102,465	95,298	102,193	95,542	103,190	600,000

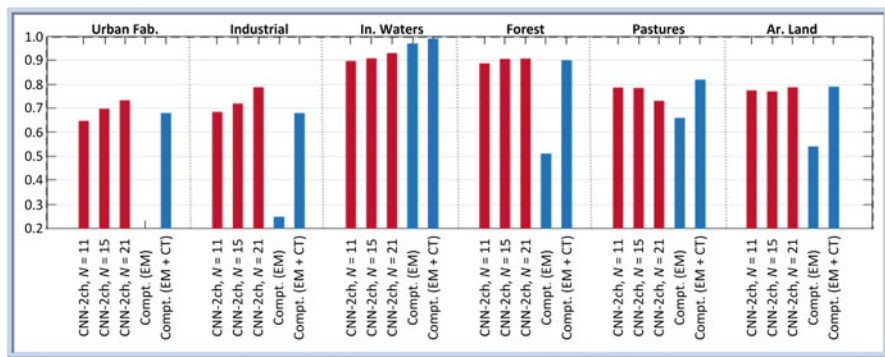


Fig. 10 The class-specific accuracies are presented by the proposed adaptive CNNs and the competing method [22] over the Dresden_X. The proposed approach in [19] uses two-channel setup ($|S_{vh}|$ and $|S_{vv}|$) with $N = 11, 15,$ and $21,$ whereas the competing method uses $|S_{vh}|$ and $|S_{vv}|$ as features (EM) and in the second setup, color and texture features included (EM + CT) in the 209-D feature vector

accuracy difference is considerable (larger than 50% for the industrial and urban classes) when they use only $|S_{vh}|$ and $|S_{vv}|$ components.

The final computed segmentation masks by the adaptive CNNs are shown in Fig. 11 together with the overlaid regions on the GTD. These overlaid masks can be compared with the ones produced by the competing method in Fig. 12 and the zoomed regions in Fig. 13. There is an existing classification noise in the competing method’s segmentation masks. In the proposed approach, the noise level is especially reduced for the arable land, forest, and forest classes.

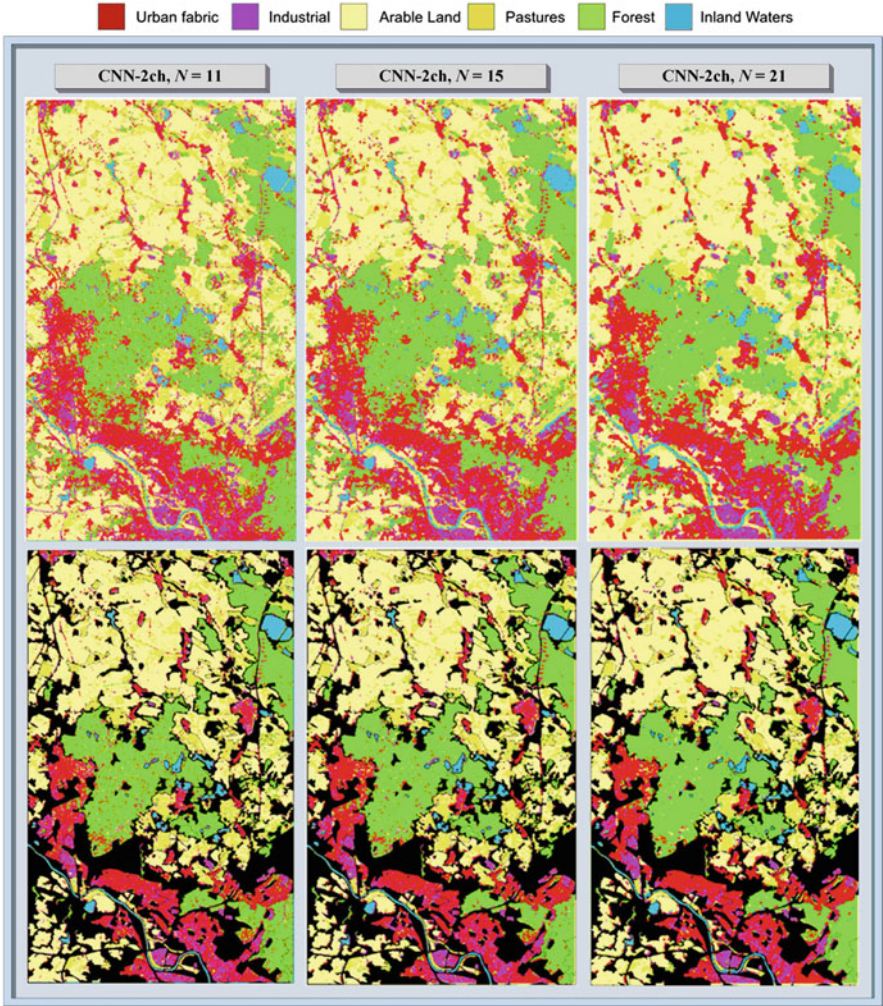


Fig. 11 The final segmentation masks are shown (first row) obtained by the proposed approach [19] over the Dresden_X using the two-channel input setup and different sliding window sizes (N). Their corresponding overlaid regions on the GTD are shown in the second row

4.3.3 Results on SFBay_L

The comparisons over the SFBay_L are performed between the proposed approach in [20] and the following competing methods: SVM and CNBC. Accordingly, the overall accuracy and individual class accuracy improvements are reported. The used features are presented in Table 8 for the competing methods. For the competing methods, the feature vectors are gradually concatenated and the final composite

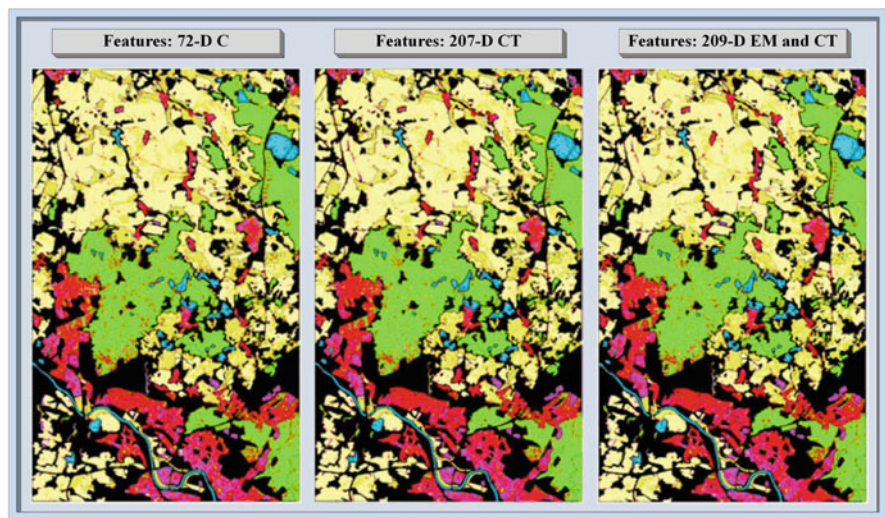


Fig. 12 The overlaid regions of the segmentation masks are shown obtained by the competing method in [21] over the Dresden_X. There are three different feature setups: (i) color features (C), (ii) color and texture features (CT), and (iii) EM and CT features

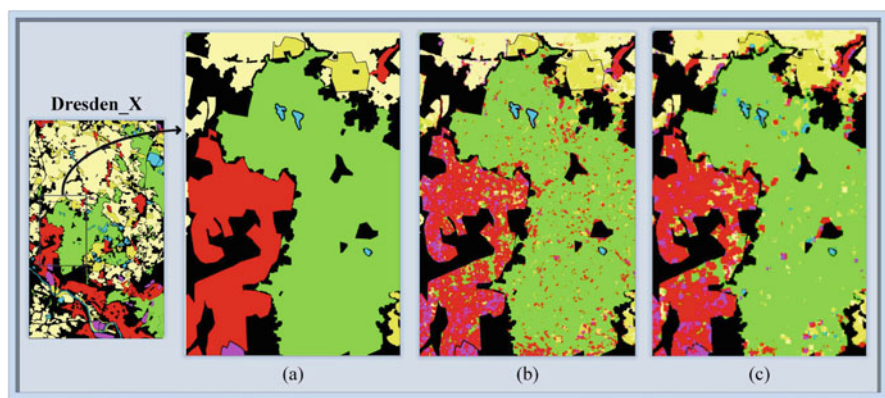


Fig. 13 The zoomed regions are shown for the Dresden_X: the GTD (a), overlaid regions obtained by the competing method (b) in [21], and the proposed method (c)

vector is used for the CNBC and SVM methods. In the CNBC classifier, each binary classifier has a 4-layer MLP classifier that has an *In-16-8-Out* neuron configuration.

The classification accuracies are presented over the SFBay_L in Table 9 for the competing methods using different combinations of features and in Table 10 using the proposed adaptive CNNs. Comparing the overall accuracies, the proposed approach with [20] outperforms both the competing methods even though it uses only the three-channel input setup using diagonal elements of \mathbf{T} . Note the fact that

Table 8 Different features are presented with their dimensions for the competing methods. The collected features include the second-order descriptors and polarimetric elements computed by different Target Decompositions (TDs) theorems

Abbreviation	Features	Dimension
FV ₁	T and C Matrices	12
FV ₂	Span, $H/A/\alpha$ [40]	7
FV ₃	Eigenanalysis—Eigenvalues	3
FV ₄	Correlation Coefficients	6
FV ₅	Touzi [72]	4
FV ₆	Krogager [38]	3
FV ₇	Freeman [41]	3
FV ₈	Huynen [42]	3
FV ₉	VanZyl [73]	3
FV ₁₀	Yamaguchi [74]	4

Table 9 Overall classification accuracies are presented obtained by the competing methods over the SFBay_L with the increasing number of features

Features	Dimension	CNBC	SVM
FV ₁	12	0.9583	0.9563
FV ₂ + (FV ₁)	19	0.9723	0.9734
FV ₃ + (FV ₁ + FV ₂)	22	0.9763	0.9746
FV ₄ + (FV ₁ + FV ₂ + FV ₃)	28	0.9786	0.9791
FV ₅ + (FV ₁ + ... + FV ₃ + FV ₄)	32	0.9790	0.9798
FV ₆ + (FV ₁ + ... + FV ₄ + FV ₅)	35	0.9759	0.9802
FV ₇ + (FV ₁ + ... + FV ₅ + FV ₆)	38	0.9806	0.9800
FV ₈ + (FV ₁ + ... + FV ₆ + FV ₇)	41	0.9796	0.9801
FV ₉ + (FV ₁ + ... + FV ₇ + FV ₈)	44	0.9803	0.9798
FV ₁₀ + (FV ₁ + ... + FV ₈ + FV ₉)	48	0.9807	0.9792

the competing methods utilize 4–12 times more features including various TDs and other low-level EM features. In general, the proposed approach has obtained 1.32% higher accuracy than the CNBC and 1.37% higher than the SVM classifier. The best accuracy levels are achieved by the CNBC and SVM using 48-D and 35-D features, respectively, whereas the proposed approach utilizes a six-channel input at most. One can similarly compare the accuracy levels obtained by using only 12-D FV₁. Accordingly, the accuracy improvement is approximately 3% by the proposed approach using a three-times less number of features.

The optimal window size N for SFBay_L seems to be between 19 and 23 based on the number of input channels. On the other hand, any $N \geq 9$ results with >98.2% accuracy if the number of input channels is selected appropriately. Therefore, the proposed approach in general yields better accuracy for the SFBay_L. For the visual evaluation, the final computed segmentation mask and the samples used for the training are shown in Fig. 14 for the SFBay_L. Even though the best accuracy is

Table 10 Overall classification accuracies are presented obtained by the proposed approach [20] over the SFBay_L using three, four, and six-channel inputs and varied window sizes. The best accuracies are highlighted in bold

SFBay_L	Three-channel	Four-channel	Six-channel
Window size	T_{11}, T_{22}, T_{33}	$T_{11}, T_{22}, T_{33}, \text{Span}$	$T_{11}, T_{22}, T_{33}, C_{11}, C_{22}, C_{33}$
7×7	0.9748	0.9726	0.9751
9×9	0.9846	0.9825	0.9764
11×11	0.9888	0.9812	0.9843
13×13	0.9807	0.9864	0.9900
15×15	0.9847	0.9865	0.9918
17×17	0.9888	0.9889	0.9893
19×19	0.9884	0.9915	0.9936
21×21	0.9807	0.9939	0.9934
23×23	0.9911	0.9908	0.9901
25×25	0.9817	0.9777	0.9855
31×31	0.9700	0.9630	0.9515

Table 11 The confusion matrix is presented for the SFBay_L obtained by the proposed approach [20] using the best setup that provides the highest overall accuracy: four-channel (T_{11}, T_{22}, T_{33} , and Span(S)) and $N = 21$. The correctly classified number of samples is highlighted in bold per class

		Predicted					
		Water	Urban	Forest	Bare soil	Nat. vegetation	Total
True	Water	78,621	0	0	0	27	78,648
	Urban	0	17,940	38	4	0	17,982
	Forest	0	313	4202	0	20	4535
	Bare soil	78	71	68	6956	0	7173
	Nat. vegetation	0	0	128	0	13,531	13,659
	Total	78,699	18,324	4436	6960	13,578	121,997

achieved by using a 21×21 window size and four-channel input setup, the coarse visual resolution in the computed segmentation mask suffers from illustrating the high details. In this case, a 7×7 window size can be chosen to achieve a more detailed segmentation mask.

The confusion matrix of the proposed method is presented in Table 11 over the SFBay_L and the class-specific classification accuracies are provided in Fig. 15. It is observed that for some classes such as water and urban, the accuracy levels are comparable, but for example, the significant accuracy gaps appear for some terrain types (for the forest, it is larger than 20%). Hence, as previously discussed, the competing methods fail to provide satisfactory performance valid for each terrain type. Basically, they are unable to provide a decent discrimination capability for some classes due to the manually extracted features.

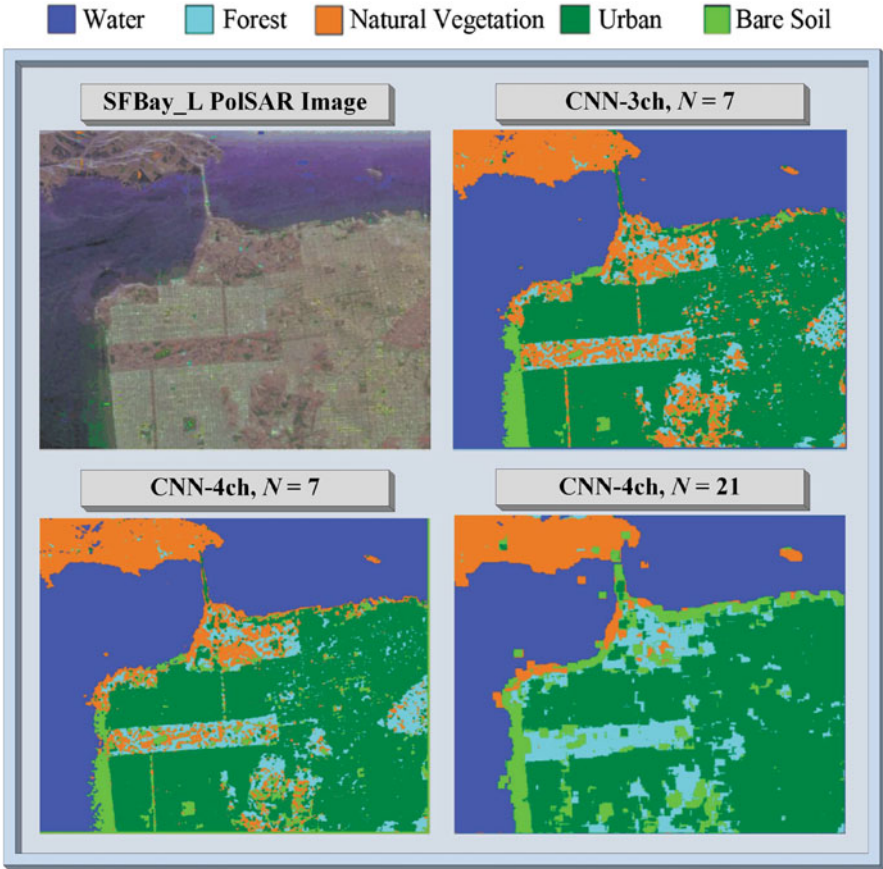


Fig. 14 The final segmentation masks are shown obtained by the proposed approach [20] over the SFBay_L using three and four-channel input setups and different sliding window sizes (N)

4.3.4 Results on SFBay_C

In this study site, the sliding window size N is varied between $N = 5$ and $N = 19$. The results are presented in Table 12 using 3–6 input channels. It is observed that the proposed approach [20] with the three-channel setup obtains considerably better overall classification accuracy (larger than 10%) than the competing method [22] using EM features with the 46-D feature vector. When the competing method uses 187-D features including EM, color, and texture features, the accuracy gap is still greater than 3% between the proposed approach and the competing method even though the proposed approach utilizes only a 4-D feature vector. Hence, this is an important achievement considering the time complexity burden due to computing such high-dimensional feature vectors.

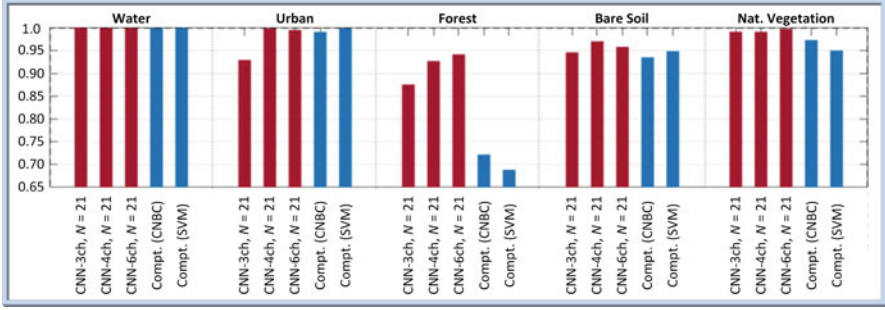


Fig. 15 The class-specific accuracies are presented by the proposed adaptive CNNs and the competing methods with CNBC and SVM over the SFBay_L. The proposed approach in [20] uses three (T_{11} , T_{22} , and T_{33}), four (T_{11} , T_{22} , T_{33} and Span(S)), and six-channel (T_{11} , T_{22} , T_{33} , C_{11} , C_{22} , and C_{33}) setups with $N = 21$, whereas the competing methods use all EM features in Table 8

Table 12 Overall classification accuracies are presented obtained by the proposed approach [20] over the SFBay_C using three, four, and six-channel inputs and varied window sizes. The best accuracies are highlighted in bold

SFBay_C	Three-channel	Four-channel	Six-channel
Window size	T_{11}, T_{22}, T_{33}	$T_{11}, T_{22}, T_{33}, \text{Span}$	$T_{11}, T_{22}, T_{33}, C_{11}, C_{22}, C_{33}$
5×5	0.8417	0.8415	0.8324
7×7	0.883	0.8794	0.8687
9×9	0.9119	0.9136	0.9017
11×11	0.9310	0.9292	0.9220
13×13	0.9380	0.9389	0.9312
15×15	0.9440	0.9439	0.9395
17×17	0.9496	0.9466	0.9417
19×19	0.9496	0.9532	0.9452

To investigate the class-specific classification performance, the confusion matrix is presented in Table 13 for the proposed approach [20] and the obtained accuracies per class is shown in Fig. 16 over the SFBay_C. As observed previously for the other study sites, there are improvements for some classes such as high urban and low urban classes. However, considering the water class, the accuracy is not further improved by the proposed approach for both SFBay_L and SFBay_C datasets. This is indeed expected as the water classification accuracy already approaches one in the confusion matrices. Thus, there is limited room for improvement as provided in the confusion matrices. One important observation is that the method in [22] can achieve larger than 80% classification accuracy if they use all the computed features. Naturally, the feature extraction procedure increases the computational complexity and the use of ensemble classifier renders the real-time application of the method.

For qualitative analysis, the final segmentation masks are shown in Fig. 17 for the SFBay_C. Accordingly, although the optimal window size is $N = 19$, the coarse resolution appears on the masks with large window sizes as previously observed.

Table 13 The confusion matrix is presented for the SFBay_C obtained by the proposed approach in [20] with the best setup that provides the highest accuracy: four-channel (T_{11} , T_{22} , T_{33} , and Span(S)) and $N = 19$. The correctly classified number of samples is highlighted in bold per class

		Predicted					
		Water	Developed	High. Urban	Low. Urban	Vegetation	Total
True	Water	49,545	1	108	5459	287	50,000
	Developed	0	47,873	100	523	1504	50,000
	High. Urban	0	412	46,901	2228	459	50,000
	Low. Urban	219	124	1578	47,131	948	50,000
	Vegetation	183	941	437	1579	46,860	50,000
	Total	49,947	49,351	49,351	49,124	50,058	250,000

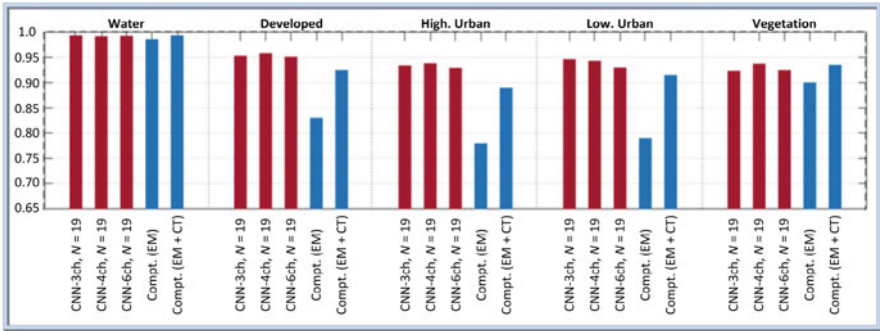


Fig. 16 The class-specific accuracies are presented by the proposed adaptive CNNs and the competing method [22] over the SFBay_C. The proposed approach [20] uses three (T_{11} , T_{22} , and T_{33}), four (T_{11} , T_{22} , T_{33} , and Span(S)), and six-channel (T_{11} , T_{22} , T_{33} , C_{11} , C_{22} , and C_{33}) setups with $N = 19$, whereas the competing methods use EM features from Table 8 and in the second setup, color and texture features included (EM + CT) in the 187-D feature vector

4.3.5 Results on Flevo_L

The Flevo_L PolSAR data have 15 classes and naturally, the classification task is more challenging than the previous benchmark datasets used in this study. Thus, to evaluate the performance of the proposed approach in [20], the number of training samples is varied for the Flevo_L and its effect is explored. In this manner, the number of selected samples is changed between 120 and 480, and additionally, an adaptive (Adp.) mode is followed by randomly choosing 2% of samples for each class in the GTD for the training. The obtained overall accuracies are presented in Table 14. It is observed that the best accuracy is achieved as 92.49% using the six-channel setup and $N = 9$ with 480 training samples. Note that the largest used number of training samples per class (480) is still approximately 3.4% of the total GTD. One can also use $\sim 2\%$ of the GTD for the training in the Adp. mode and can achieve 92.33% classification accuracy yielding an insignificant performance drop.

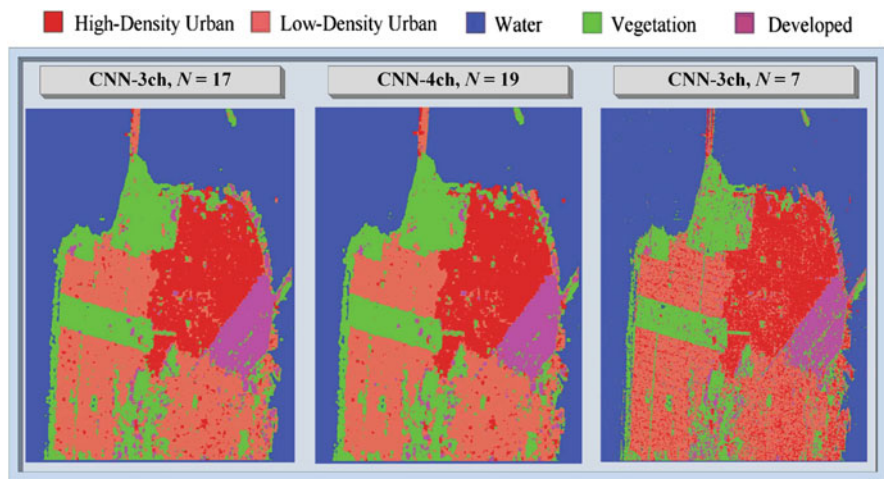


Fig. 17 The final segmentation masks are shown obtained by the proposed approach in [20] over the SFBay_C using three and four-channel input setups and different sliding window sizes (N)

In the competing method [22], achieved the best accuracy using FV_1 is $<83\%$ with a similar training size in the proposed approach. When the EM, color, and texture features (46-D, 60-D, 81-D, respectively) are utilized in [22], the best accuracy is achieved as 91.5% by a large ensemble of RF classifiers. Therefore, it is shown that the proposed approach exceeds this performance level utilizing only 6-D features and with a compact CNN.

The qualitative analysis by visually inspecting the final segmentation masks further reveals the classification performance enhancements by the proposed approach. For instance, considering the overlaid regions of the segmentation masks on the GTD in Figs. 18 and 19, although the competing method in [22] uses a 187-D feature vector, the segmentation masks produced by the proposed approach have nor or less noise, e.g., for water, barley, and rapeseed classes. The classification performance of individual classes can be further analyzed by the confusion matrix presented in Table 15 obtained with the proposed approach for the Flevo_L.

In the previous discussion about the effect of different N values, it was mentioned that high overall classification accuracies are achieved by large N values and as a trade-off, the finer details are removed in the segmentation masks. On the other hand, for the Flevo_L, it is observed that the best accuracy is obtained by using small N values. This is observed in the Flevo_L specifically because the study area consists of mostly heterogeneous regions with a high number of terrain classes. Consequently, the optimal sliding window size N is generally determined by the SAR data.

Table 14 Overall classification accuracies are obtained by the proposed approach [20] over the Flevo_L using three, four, and six-channel inputs and varied window sizes. The best accuracies are highlighted in bold

Flevo_L	Three-channel	Four-channel	Six-channel	
Window size	T_{11}, T_{22}, T_{33}	$T_{11}, T_{22}, T_{33}, \text{Span}$	$T_{11}, T_{22}, T_{33}, C_{11}, C_{22}, C_{33}$	Samples
5×5	0.8633	0.8594	0.9014	120
7×7	0.8806	0.8910	0.910	120
7×7	0.8946	0.8955	0.9191	240
7×7	0.8994	0.9006	0.9186	360
7×7	0.9022	0.9036	0.9227	480
7×7	0.9040	0.9018	0.9233	Adp.
9×9	0.8968	0.8837	0.9073	120
9×9	0.8968	0.8842	0.9133	240
9×9	0.8971	0.9014	0.9235	360
9×9	0.90	0.9043	0.9249	480
9×9	0.9033	0.9030	0.9203	Adp.
11×11	0.8765	0.870	0.8896	120
11×11	0.8915	0.8923	0.9105	Adp.
21×21	0.7964	0.7992	0.8133	120
21×21	0.8350	0.8368	0.8496	Adp.
31×31	0.7640	0.7245	0.7795	120
31×31	0.8465	0.8167	0.8524	Adp.

4.3.6 Results on Flevo_C

The overall classification accuracies are presented in Table 16 for the proposed approach in [20] over the Flevo_C PolSAR data. The best accuracy is achieved by the three-channel input setup and $N = 15$ as 96.35% which is higher than the accuracies obtained in [22] using 46-D EM features and 187-D EM, color, and texture features as 91.19% and 95.68%, respectively.

The previous statements are valid for this study site as well such as improved classification performance with the minimum computational complexity. The confusion matrix obtained by the proposed approach is provided in Table 17. Additionally, the overlaid regions on the GTD are shown in Figs. 20 and 21 for the proposed approach and the competing method in [22]. Accordingly, the segmentation mask obtained by the competing method using EM features is poorly segmented for some classes, e.g., cropland is mostly misclassified as urban and forest. Overall, the segmentation mask of the proposed approach is better discriminative than the competing method even though they use 187-D features.

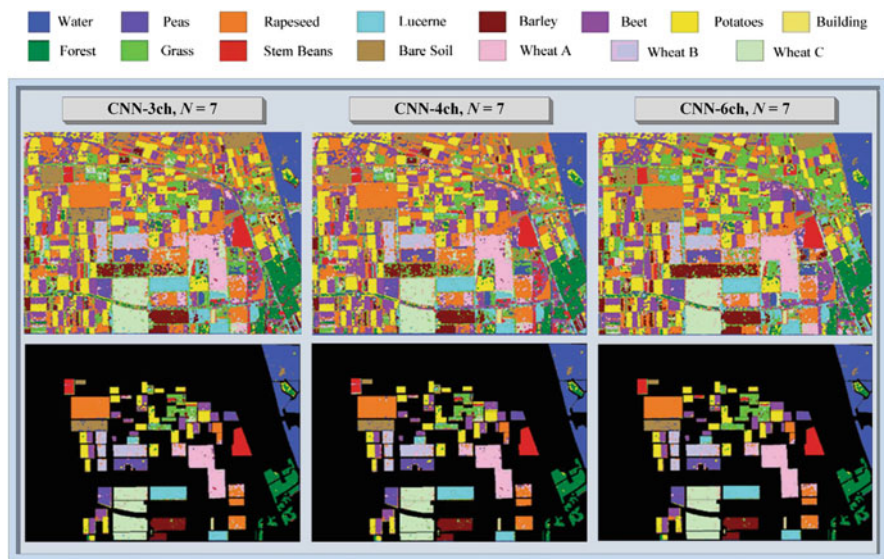


Fig. 18 The final segmentation masks are shown (first row) obtained by the proposed approach [20] over the Flevo_L using three, four, and six-channel input setups with $N = 7$. Their corresponding overlaid regions on the GTD are shown in the second row

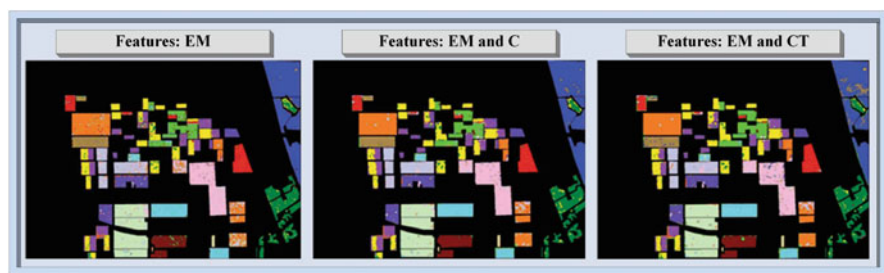


Fig. 19 The overlaid regions of the segmentation masks are shown obtained by the competing method in [22] over the Flevo_L. There are three different feature setups: (i) EM, (ii) EM and color (C) features, and (iii) EM and all image processing features including color and texture (CT)

4.3.7 Deep Versus Compact CNNs

Two deep CNNs: Xception and Inception-ResNet-v2 are evaluated against the proposed approach in [19]. In the previous discussion, it was stated that such deep CNN models are not well-designed to operate using small sliding window sizes for the pixel-wise classification. The Xception model requires at least 71×71 and the Inception-ResNet-v2 requires 75×75 input dimensions due to their significant number of convolutional layers. In this manner, the cropped input patch for each window is up-sampled using bilinear interpolation to feed the networks.

Table 15 The confusion matrix is presented for the Flevo_L obtained by the proposed approach [20] with the best setup that provides the highest accuracy: six-channel (T_{11} , T_{22} , T_{33} , C_{11} , C_{22} , and C_{33}) and $N = 7$. The order of the classes is as follows: water, forest, lucerne, grass, rapeseed, beet, potatoes, peas, stem, beans, bare soil, wheat A, wheat B, wheat C, barley, and building. The correctly classified number of samples is highlighted in bold per class

		Predicted														Total
True		0	57	0	1	45	18	19	426	1	9	2	4	7	28,672	
75	14,565	0	27	0	117	723	0	13	0	4	0	1	0	19	15,544	
272	0	9790	745	7	0	1	0	3	3	0	88	57	11	0	10,977	
0	0	320	8896	0	173	171	1	207	3	0	9	133	238	50	10,201	
77	0	3	80	19,147	232	3	167	54	58	1208	248	144	1	0	21,422	
4	54	4	272	460	12,740	535	89	235	15	9	0	13	65	64	14,559	
79	786	26	218	87	733	18,376	175	217	12	48	1	128	15	23	20,924	
0	8	0	16	57	358	171	9221	17	1	340	0	0	0	0	10,189	
0	100	1	222	3	100	51	1	7780	8	16	2	7	0	12	8303	
349	0	0	37	0	0	0	0	6	5765	0	0	2	32	0	6191	
0	0	0	17	358	37	0	93	17	0	16,220	372	178	0	0	17,292	
0	0	0	0	639	5	0	0	0	0	409	9266	100	0	0	10,419	
0	0	37	9	109	10	0	0	6	0	214	123	21,077	0	0	21,585	
73	0	13	172	0	0	0	0	0	4	0	0	0	6962	0	7224	
0	2	0	8	0	0	0	0	0	0	0	0	0	0	511	521	
Total	28,993	15,534	10,194	10,776	20,867	14,506	9765	8574	6295	18,469	10,118	21,842	7328	686	204,023	

Table 16 Overall classification accuracies are obtained by the proposed approach in [20] over the Flevo_C using three, four, and six-channel inputs and varied window sizes. The best accuracies are highlighted in bold

Flevo_C	Three-channel	Four-channel	Six-channel
Window size	T_{11}, T_{22}, T_{33}	$T_{11}, T_{22}, T_{33}, \text{Span}$	$T_{11}, T_{22}, T_{33}, C_{11}, C_{22}, C_{33}$
5×5	0.9078	0.9034	0.9115
7×7	0.9357	0.9353	0.9367
9×9	0.9543	0.9539	0.9552
11×11	0.9575	0.9571	0.9537
13×13	0.9630	0.9596	0.9561
15×15	0.9635	0.9631	0.9614

Table 17 The confusion matrix is presented over the Flevo_C obtained by the proposed approach [20] with the best setup that provides the highest accuracy: three-channel (T_{11} , T_{22} , and T_{33}) and $N = 15$. The correctly classified numbers of samples are highlighted in bold per class

		Predicted				
		Water	Urban	Forest	Cropland	Total
True	Water	49,616	287	27	70	50,000
	Urban	13	48,489	701	797	50,000
	Forest	44	1539	46,664	1753	50,000
	Cropland	89	632	1346	47,933	50,000
	Total	49,762	50,947	48,738	50,553	200,000

It is a fact that the deep CNNs are better designed for the high-resolution SAR images since the correlation within the neighborhood starts to decrease with a large N on small-resolution data. Hence, given the image dimensions in Table 2, the deep CNNs are evaluated over the PDelta_X and Dresden_X SAR data. The obtained overall classification accuracies are presented in Tables 18 and 19 for the PDelta_X and Dresden_X, respectively. Expectedly, due to the limited size of the training data, the deep CNNs lack the classification performance that is achieved by the proposed method. Moreover, it is observed that if a detailed segmentation mask is desired with the deep CNNs, 5×5 window size results with a poor classification accuracy as seen in Table 19: there is $>20\%$ accuracy gap between the proposed and deep CNN approaches over the Dresden_X data. Nevertheless, the Inception-ResNet-v2 model produces a slightly improved accuracy than the proposed approach for $N = 27$ over the Dresden_X. However, as discussed, the practical usage of such large N values is limited considering the problem with obtaining a high-detailed segmentation mask.

4.3.8 Sensitivity Analysis on Hyper-Parameters

There are several hyper-parameters of the proposed adaptive CNNs. The effect of the hyper-parameter N is already discussed in the results. The remaining parameters include the number of hidden layers and neurons per layer. In the experiments,

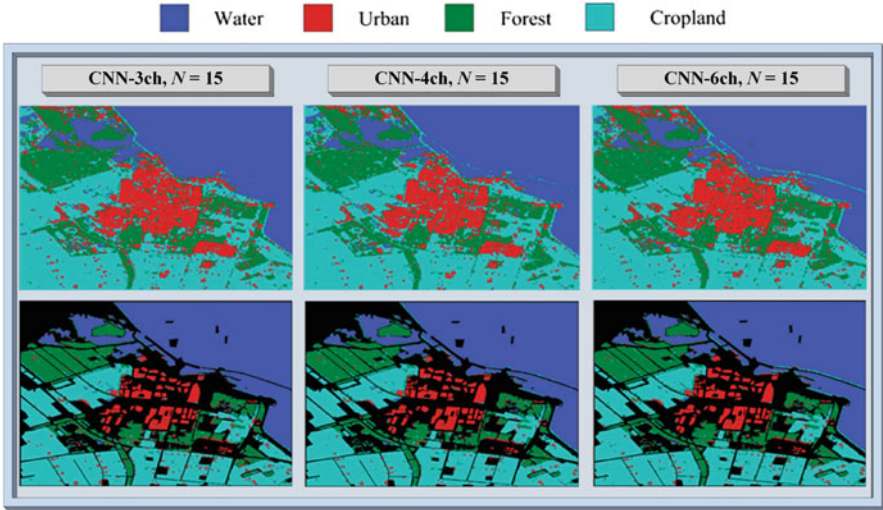


Fig. 20 The final segmentation masks are shown (first row) obtained by the proposed approach [20] over the Flevo_C using three, four, and six-channel input setups and with $N = 15$. Their corresponding overlaid regions on the GTD are shown in the second row

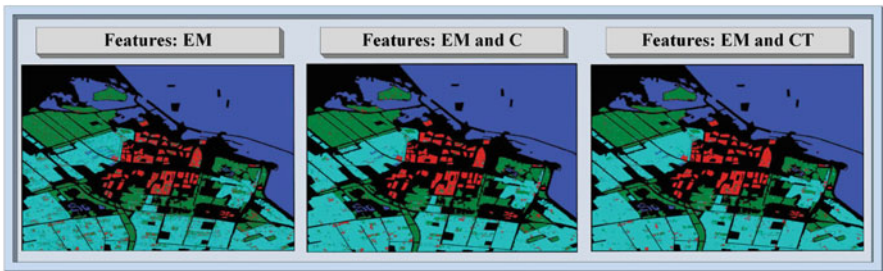


Fig. 21 The overlaid regions of the segmentation masks are shown obtained by the competing method in [22] over the Flevo_C. There are three different feature setups: (i) EM, (ii) EM and color (C) features, and (iii) EM and all image processing features including color and texture (CT)

a compact CNN was used with a single hidden CNN layer having 20 neurons (kernel filters) and with a hidden MLP layer having only 10 neurons. Recall the fact that the number of neurons in the input is determined by the number of EM channels, whereas the number of neurons in the output layer is equal to the number of classes. Hence, to analyze the sensitivity of the proposed approaches in [19, 20] on the network parameters, the number of hidden layers and neurons are varied considerably. To this end, the multipliers m and n are defined to control the number of hidden CNN layers and the number of neurons in hidden layers, respectively. The default setup is defined with $m = n = 1$, for instance, if $m = 2$ and $n = 4$, then the network would have the following [In-80-80-40-Out] configuration: two

Table 18 Overall classification accuracies are presented obtained by the proposed approach [19] and deep CNNs including Xception and Inception-ResNet-v2 over the PDelta_x using four-channel input and varied window sizes. Xception* and Inception-ResNet-v2* indicate that the training started with the ImageNet weights (except for the first convolutional and MLP layers). The best accuracies are highlighted in bold

Window size	Xception	Xception*	Inception ResNet-v2	Inception ResNet-v2*	Proposed
5×5	0.6688	0.6963	0.6862	0.6928	0.7080
11×11	0.7563	0.7736	0.7608	0.7656	0.7838
17×17	0.7896	0.7943	0.8032	0.8121	0.830
25×25	0.8445	0.8435	0.8555	0.8447	0.8632

Table 19 Overall classification accuracies are presented obtained by the proposed approach [19] and deep CNNs including Xception and Inception-ResNet-v2 over the Dresden_x using two-channel input and varied window sizes. Xception* and Inception-ResNet-v2* indicate that the training started with the ImageNet weights (except for the first convolutional and MLP layers). The best accuracies are highlighted in bold

Window size	Xception	Xception*	Inception ResNet-v2	Inception ResNet-v2*	Proposed
5×5	0.4637	0.5004	0.4657	0.4754	0.7059
11×11	0.6481	0.6836	0.6556	0.6488	0.7797
17×17	0.7342	0.7596	0.7441	0.7459	0.8007
21×21	0.7706	0.7783	0.7767	0.7796	0.8133
27×27	0.7960	0.8068	0.8116	0.8064	0.8062

hidden CNN layers having $4 \times 20 = 80$ neurons and one hidden MLP layer with $4 \times 10 = 40$ neurons. On the other hand, the N values and number of input channels are set to the optimal values according to the achieved highest overall classification accuracy for each SAR data.

For different network configurations, the classification accuracies are presented in Table 20. The overall accuracy slightly varies in a margin of around $\pm 3\%$. Therefore, it is shown that the proposed approach is robust to hyper-parameter values in general. Furthermore, it is observed that the classification performance degrades when the number of hidden layers increased; and the best accuracy is obtained by using single hidden CNN and MLP layers. This outcome was expected because of the over-fitting phenomenon which reduces the generalization capability of the network. For instance, using $m = 1, n = 8$ and $m = 2, n = 8$ network configurations, there has been no convergence at all for some study sites within 1000 epochs and 5 different BP runs.

5 Conclusions and Future Work

The novel approaches proposed in [19, 20] based on adaptive and compact CNNs provide an accurate and fast classification of SAR images including the partially

Table 20 Overall classification accuracies are presented for the proposed approach in [19, 20] using different network configurations. The multiplier m and n control the number of hidden CNN layers and the number of hidden neurons, respectively, where $m = n = 1$ represents the initial network architecture. The number of input channels and sliding window size N are set to the best values for each study site. Convergence has not occurred for the entries with “–” for five different BP runs with 1000 epochs. The best accuracies are highlighted in bold

(a) Partially polarized data			(b) PolSAR data				
	PDelta_X	Dresden_X		SFBay_L	SFBay_C	Flevo_L	Flevo_C
$m = 1, n = 1$	0.8487	0.8133	$m = 1, n = 1$	0.9939	0.9532	0.9233	0.9635
$m = 1, n = 2$	0.8493	0.8176	$m = 1, n = 2$	0.9931	0.9560	0.9305	0.9641
$m = 1, n = 4$	0.8683	0.8371	$m = 1, n = 4$	0.9862	0.9498	0.8944	0.9687
$m = 2, n = 1$	0.8015	0.7772	$m = 1, n = 8$	0.7366	0.4214	–	–
$m = 2, n = 2$	0.7945	0.7747	$m = 2, n = 1$	0.9753	0.9371	0.8967	0.9503
$m = 2, n = 4$	0.7908	0.7766	$m = 2, n = 2$	0.9615	0.9318	0.8611	0.9482
			$m = 2, n = 4$	0.8245	0.9360	0.7898	0.9463
			$m = 2, n = 8$	0.6019	–	–	–

polarized and fully polarized SAR data. Thanks to the compact network configuration, the introduced capabilities can be summarized as follows, first, it is possible to operate and generalize well on the data when the training/annotated samples are limited such as in the case of only 0.1% of the whole scene are annotated. In this way, the labeling cost caused by the human intervention is greatly reduced compared to the existing deep CNN methods that require a significant amount of training data. Moreover, the reduced computational time complexity allows real-time implementation of the proposed approaches in [19, 20]. Next, contrary to the deep CNNs, it is possible to use a small sliding window size N for the pixel-wise classification. This brings the following advantages: (i) the overall classification accuracy tends to decrease when N increased above a certain value since the correlation within the patch reduces and (ii) the details in the final computed LU/LC segmentation mask are inversely proportional to N .

There are few existing previous methods such as [21, 22] providing fine-detailed masks and operating with small training data sizes. Accordingly, they extract image processing features consisting of color and texture and an ensemble of classifiers is utilized to classify the SAR data. It is also shown that using one compact classifier and small dimensional input channels, superior classification accuracy can be obtained by the proposed approach. Hence, the conventional feature extraction procedure can totally be eliminated in this approach. Overall, the experimental evaluations prove that the proposed approach improves the reliability of the inter-class classification and removes the segmentation noise in the computed masks.

In future work, it is planned to have more performance comparisons using different datasets. For example, the SAR data used in this work have X, L, and C frequency bands. It would be an interesting task to benchmark the proposed approach over new bands and also explore different fusion techniques for multi-frequency classification. In [6], a dual-band 1D-CNN is proposed to combine L

and C bands for the classification. Although, it seems feasible, the computational complexity will be much higher if the same topology is followed for the 2D-CNNs. Recently, tensorization frameworks have been used for the data fusion as proposed in [75]. These include the future work directions considering the multifrequency PolSAR classification using the proposed approach. One can also investigate the combination of some of the color features with EM features as the alternative set of input channels to the CNN classifier. Finally, in the experiments, only the real-valued EM channels are used in the classification. On the other hand, it is a fact that the observed target is represented with complex values in the acquired SAR data. Such issues these issues and further investigations will be addressed in the future.

References

1. C. Brekke and A. H. S. Solberg, "Oil spill detection by satellite remote sensing," *Remote Sensing of Environment*, 2005.
2. Y. Endo, B. Adriano, E. Mas, and S. Koshimura, "New Insights into Multiclass Damage Classification of Tsunami-Induced Building Damage from SAR Images," *Remote Sensing*, vol. 10, no. 12, 2018. [Online]. Available: <http://www.mdpi.com/2072-4292/10/12/2059>
3. P.-L. Frison, B. Fruneau, S. Kmiha, K. Soudani, E. Dufrêne, T. Le Toan, T. Koleček, L. Villard, E. Mougin, and J.-P. Rudant, "Potential of Sentinel-1 Data for Monitoring Temperate Mixed Forest Phenology," *Remote Sensing*, vol. 10, no. 12, 2018. [Online]. Available: <http://www.mdpi.com/2072-4292/10/12/2049>
4. T. Ince, M. Ahishali, and S. Kiranyaz, "Comparison of polarimetric SAR features for terrain classification using incremental training," in *Progress in Electromagnetics Research Symposium*, St. Petersburg, 2017, pp. 3258–3262.
5. M. Ahishali, T. Ince, S. Kiranyaz, and M. Gabbouj, "Performance Comparison of Learned vs. Engineered Features for Polarimetric SAR Terrain Classification," in *Progress in Electromagnetics Research Symposium (PIERS)*, 2019.
6. M. Ahishali, S. Kiranyaz, T. Ince, and M. Gabbouj, "Multifrequency PolSAR image classification using dual-band 1d convolutional neural networks," in *2020 Mediterranean and Middle-East Geoscience and Remote Sensing Symposium (M2GARSS)*. IEEE, 2020, pp. 73–76.
7. Z. Qi, A. G.-O. Yeh, X. Li, and Z. Lin, "A novel algorithm for land use and land cover classification using RADARSAT-2 polarimetric SAR data," *Remote Sensing of Environment*, 2012.
8. K. Ouchi, "Recent trend and advance of synthetic aperture radar with selected topics," 2013.
9. J. Mennis, "Socioeconomic-Vegetation Relationships in Urban, Residential Land," *Photogrammetric Engineering & Remote Sensing*, vol. 11, pp. 911–921, 2006.
10. P. Jonsson, "Vegetation as an urban climate control in the subtropical city of Gaborone, Botswana," *International Journal of Climatology*, 2004.
11. X.-L. Chen, H.-M. Zhao, P.-X. Li, and Z.-Y. Yin, "Remote sensing image-based analysis of the relationship between urban heat island and land use/cover changes," *Remote Sensing of Environment*, vol. 104, no. 2, pp. 133–146, 2006. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0034425706001787>
12. E. Santi, S. Paloscia, S. Pettinato, G. Fontanelli, M. Mura, C. Zolli, F. Maselli, M. Chiesi, L. Bottai, and G. Chirici, "The potential of multifrequency SAR images for estimating forest biomass in Mediterranean areas," *Remote Sensing of Environment*, 2017.
13. A. M. Braga, R. C. Marques, F. A. Rodrigues, and F. N. Medeiros, "A median regularized level set for hierarchical segmentation of SAR images," *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 7, pp. 1171–1175, 2017.

14. R. Jin, J. Yin, W. Zhou, and J. Yang, "Level Set Segmentation Algorithm for High-Resolution Polarimetric SAR Images Based on a Heterogeneous Clutter Model," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 10, no. 10, pp. 4565–4579, 2017.
15. J. Cousty, G. Bertrand, L. Najman, and M. Couprie, "Watershed cuts: Thinnings, shortest path forests, and topological watersheds," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, pp. 925–939, 2010.
16. M. Ciecholewski, "River channel segmentation in polarimetric SAR images: Watershed transform combined with average contrast maximisation," *Expert Systems with Applications*, vol. 82, pp. 196–215, 2017.
17. D. Stutz, A. Hermans, and B. Leibe, "Superpixels: An evaluation of the state-of-the-art," *Computer Vision and Image Understanding*, vol. 166, pp. 1–27, 2018. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1077314217300589>
18. F. Lang, J. Yang, S. Yan, and F. Qin, "Superpixel Segmentation of Polarimetric Synthetic Aperture Radar (SAR) Images Based on Generalized Mean Shift," *Remote Sensing*, vol. 10, no. 10, p. 1592, 2018.
19. M. Ahishali, S. Kiranyaz, T. Ince, and M. Gabbouj, "Dual and single polarized SAR image classification using compact convolutional neural networks," *Remote Sensing*, vol. 11, no. 11, p. 1340, 2019.
20. —, "Classification of polarimetric SAR images using compact convolutional neural networks," *GIScience & Remote Sensing*, vol. 58, no. 1, pp. 28–47, 2021.
21. S. Uhlmann and S. Kiranyaz, "Classification of dual- and single polarized SAR images by incorporating visual features," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 90, pp. 10–22, 2014.
22. —, "Integrating color features in polarimetric SAR image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 4, pp. 2197–2216, 2014.
23. S. Kiranyaz, T. Ince, S. Uhlmann, and M. Gabbouj, "Collective Network of Binary Classifier Framework for Polarimetric SAR Image Classification: An Evolutionary Approach," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 42, no. 4, pp. 1169–1186, 2012.
24. S. Uhlmann, S. Kiranyaz, T. Ince, and M. Gabbouj, "Dynamic and data-driven classification for polarimetric SAR images," in *Proceedings of SPIE - The International Society for Optical Engineering*, vol. 8180, 2011.
25. W. Yang, T. Zou, D. Dai, and H. Sun, "Polarimetric SAR image classification using multifeatures combination and extremely randomized clustering forests," *EURASIP Journal on Advances in Signal Processing*, vol. 2010, pp. 1–12, 2010.
26. Y. Lecun, Y. Bengio, and G. Hinton, "Deep learning," pp. 436–444, 2015.
27. M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L. C. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 4510–4520.
28. M. Tan and Q. Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," in *International Conference on Machine Learning*, 2019, pp. 6105–6114.
29. X. X. Zhu, D. Tuia, L. Mou, G. S. Xia, L. Zhang, F. Xu, and F. Fraundorfer, "Deep Learning in Remote Sensing: A Comprehensive Review and List of Resources," *IEEE Geoscience and Remote Sensing Magazine*, vol. 5, no. 4, pp. 8–36, 2017.
30. Y. Zhou, H. Wang, F. Xu, and Y. Q. Jin, "Polarimetric SAR Image Classification Using Deep Convolutional Neural Networks," *IEEE Geoscience and Remote Sensing Letters*, vol. 13, no. 12, pp. 1935–1939, 2016. [Online]. Available: <http://ieeexplore.ieee.org/document/7762055/>
31. F. Gao, T. Huang, J. Wang, J. Sun, A. Hussain, and E. Yang, "Dual-Branch Deep Convolution Neural Network for Polarimetric SAR Image Classification," *Applied Sciences*, vol. 7, no. 5, p. 447, 2017. [Online]. Available: <http://www.mdpi.com/2076-3417/7/5/447>
32. S. H. Wang, J. Sun, P. Phillips, G. Zhao, and Y. D. Zhang, "Polarimetric synthetic aperture radar image segmentation by convolutional neural network using graphical processing units," *Journal of Real-Time Image Processing*, vol. 15, no. 3, pp. 631–642, 2018.

33. W. Lin, X. Liao, J. Deng, and Y. Liu, "Land cover classification of radarsat-2 SAR data using convolutional neural network," *Wuhan University Journal of Natural Sciences*, vol. 21, no. 2, pp. 151–158, 2016.
34. F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, 2017.
35. C. Szegedy, S. Ioffe, and V. Vanhoucke, "Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning," in *AAAI*, 2017.
36. O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, "ImageNet Large Scale Visual Recognition Challenge," *International Journal of Computer Vision*, 2015.
37. 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.
38. E. Krogager, "New Decomposition of the Radar Target Scattering Matrix," *Electronics Letters*, vol. 26, no. 18, pp. 1525–1527, 1990.
39. W. L. Cameron and L. K. Leung, "Feature motivated polarization scattering matrix decomposition," in *IEEE International Conference on Radar*, 1990, pp. 549–557. [Online]. Available: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=201088>
40. S. R. Cloude and E. Pottier, "An entropy based classification scheme for land applications of polarimetric SAR," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 35, no. 1, pp. 68–78, 1997.
41. A. Freeman and S. L. Durden, "A three-component scattering model for polarimetric SAR data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 36, no. 3, pp. 963–973, 1998.
42. J. R. Huynen, "Phenomenological theory of radar targets," 1970.
43. A. Marino, *A New Target Detector Based on Geometrical Perturbation Filters for Polarimetric Synthetic Aperture Radar (POL-SAR)*. Springer, 2012.
44. S. Uhlmann, "Advanced Techniques for Classification of Polarimetric Synthetic Aperture Radar Data," Ph.D. dissertation, Tampere University of Technology, 2014.
45. J. S. Lee, M. R. Grunes, and E. Pottier, "Quantitative comparison of classification capability: Fully polarimetric versus dual and single-polarization SAR," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 39, no. 11, pp. 2343–2351, 2001.
46. A. Lonnqvist, Y. Rauste, M. Molinier, and T. Hame, "Polarimetric SAR Data in Land Cover Mapping in Boreal Zone," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 10, pp. 3652–3662, 2010. [Online]. Available: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5475228>
47. V. Turkar, R. Deo, Y. S. Rao, S. Mohan, and A. Das, "Classification accuracy of multi-frequency and multi-polarization SAR images for various land covers," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 5, no. 3, pp. 936–941, 2012.
48. H. Skriver, "Crop classification by multitemporal C- and L-band single- and dual-polarization and fully polarimetric SAR," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, no. 6, pp. 2138–2149, 2012.
49. B. S. Manjunath, J. R. Ohm, V. V. Vasudevan, and A. Yamada, "Color and texture descriptors," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 11, no. 6, pp. 703–715, 2001.
50. M. Pietikäinen, T. Ojala, and Z. Xu, "Rotation-invariant texture classification using feature distributions," *Pattern Recognition*, vol. 33, no. 1, pp. 43–52, 2000.
51. R. M. Haralick, I. Dinstein, and K. Shanmugam, "Textural Features for Image Classification," *IEEE Transactions on Systems, Man and Cybernetics*, vol. SMC-3, no. 6, pp. 610–621, 1973.
52. B. S. Manjunath, P. Wu, S. Newsam, and H. Shin, "A texture descriptor for browsing and similarity retrieval," *J. Signal Process., Image Commun.*, vol. 16, no. 1–2, pp. 33–43, 2000.
53. M. J. Swain and D. H. Ballard, "Color indexing," *International Journal of Computer Vision*, vol. 7, no. 1, pp. 11–32, 1991.

54. X. Zhou, C. Zhang, and S. Li, "A perceptive uniform pseudo-color coding method of SAR images," in *CIE International Conference of Radar Proceedings*, 2007.
55. L. Chen, W. Yang, Y. Liu, and H. Sun, "Feature evaluation and selection for polarimetric SAR image classification," in *International Conference on Signal Processing Proceedings, ICSP*, 2010, pp. 2202–2205.
56. G. Gigli, R. Sabry, and G. Lampropoulos, "Classifier combination and feature selection methods for polarimetric SAR classification," *Multisensor, Multisource Information Fusion: Architectures, Algorithms, and Applications 2007*, vol. 6571, p. 65710B, 2007.
57. "Corine Land Cover," 2006. [Online]. Available: <http://sia.eionet.europa.eu/CLC2006/>
58. "U.S. Geological Survey Images." [Online]. Available: <http://terraserverusa.com>
59. S. Uhlmann, S. Kiranyaz, M. Gabbouj, and T. Ince, "Polarimetric SAR images classification using collective network of binary classifiers," in *2011 Joint Urban Remote Sensing Event*, Munich, 2011, pp. 245–248.
60. X. Liu, L. Jiao, X. Tang, Q. Sun, and D. Zhang, "Polarimetric convolutional network for PolSAR image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 5, pp. 3040–3054, 2018.
61. X. Zhang, J. Xia, X. Tan, X. Zhou, and T. Wang, "PolSAR Image Classification via Learned Superpixels and QCNN Integrating Color Features," *Remote Sensing*, vol. 11, no. 15, p. 1831, 2019.
62. X. Huang, X. Nie, H. Qiao, and B. Zhang, "Supervised PolSAR image classification by combining multiple features," in *2019 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2019, pp. 634–638.
63. J. J. Yin, J. Yang, and Y. Yamaguchi, "A new method for polarimetric SAR image classification," in *2009 2nd Asian-Pacific Conference on Synthetic Aperture Radar*. IEEE, 2009, pp. 733–737.
64. B. Ren, B. Hou, J. Chanussot, and L. Jiao, "PolSAR Feature Extraction Via Tensor Embedding Framework for Land Cover Classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 4, pp. 2337–2351, 2019.
65. Y. Chen, Y. Li, L. Jiao, C. Peng, X. Zhang, and R. Shang, "Adversarial reconstruction-classification networks for PolSAR image classification," *Remote Sensing*, vol. 11, no. 4, p. 415, 2019.
66. B. Ren, B. Hou, Z. Wen, W. Xie, and L. Jiao, "PolSAR image classification via multimodal sparse representation-based feature fusion," *International Journal of Remote Sensing*, vol. 39, no. 22, pp. 7861–7880, 2018.
67. P. Yu, A. K. Qin, and D. A. Clausi, "Unsupervised polarimetric SAR image segmentation and classification using region growing with edge penalty," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, no. 4, pp. 1302–1317, 2012.
68. W. Yang, D. Dai, J. Wu, and C. He, "Weakly supervised polarimetric SAR image classification multi-model Markov aspect model," *100 Years ISPRS Advancing Remote Sensing Science, Pt 2*, 2010.
69. J. S. Lee, "Polarimetric SAR speckle filtering and its implication for classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 37, no. 5 II, pp. 2363–2373, 1999.
70. M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. J. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Józefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. G. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. A. Tucker, V. Vanhoucke, V. Vasudevan, F. B. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng, "TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems," *CoRR*, vol. abs/1603.0, 2016. [Online]. Available: <http://arxiv.org/abs/1603.04467>
71. Chollet François, "Keras: The Python Deep Learning library," 2015.
72. R. Touzi, "Target scattering decomposition in terms of roll-invariant target parameters," *IEEE Transactions on geoscience and remote sensing*, vol. 45, no. 1, pp. 73–84, 2006.

73. J. J. van Zyl, "Application of Cloude's target decomposition theorem to polarimetric imaging radar data," in *Radar polarimetry*, vol. 1748. International Society for Optics and Photonics, 1993, pp. 184–191.
74. Y. Yamaguchi, T. Moriyama, M. Ishido, and H. Yamada, "Four-component scattering model for polarimetric SAR image decomposition," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 8, pp. 1699–1706, 2005.
75. S. De, D. Ratha, D. Ratha, A. Bhattacharya, and S. Chaudhuri, "Tensorization of multifrequency PolSAR data for classification using an autoencoder network," *IEEE Geoscience and Remote Sensing Letters*, vol. 15, no. 4, pp. 542–546, 2018.

Multi-Frequency Polarimetric SAR Data Analysis for Crop Type Classification Using Random Forest



Siddharth Hariharan, Dipankar Mandal, Siddhesh Tirodkar, Vineet Kumar, and Avik Bhattacharya

1 Introduction

Classification of crops is an efficient way for managing agricultural areas and monitoring yield. One of the successful ways of doing this is the use of Synthetic Aperture Radar (SAR) data as recognized widely in the last two decades. The potential of SAR for crop classification is significant since radar backscattering is sensitive to the dielectric properties of the vegetation and the soil, plant geometry, and surface roughness. Additionally SAR systems can operate efficiently under all weather conditions making them ideal for crop classification. Crop classification

S. Hariharan

Computer Engineering Department, TPCT's Terna Engineering College, Mumbai, India
e-mail: siddharthkalpagam@ternaengg.ac.in

D. Mandal (✉)

Kansas State University, Manhattan, KS, USA
e-mail: dmandal@ksu.edu

S. Tirodkar

Geophysical and Multiphase Flows Laboratory, IDP in Climate Studies, Indian Institute of Technology Bombay, Mumbai, India
e-mail: siddhesh.t@iitb.ac.in

V. Kumar

Department of Water Resources, Delft University of Technology, Delft, The Netherlands
e-mail: V.Kumar-1@tudelft.nl

A. Bhattacharya

Microwave Remote Sensing Lab, Centre of Studies in Resources Engineering, Indian Institute of Technology Bombay, Mumbai, India
e-mail: avikb@csre.iitb.ac.in

using SAR data has produced admissible accuracies in literature [4, 13, 15, 25, 29–31, 43, 48, 56].

Multi-frequency polarimetric SAR analysis of crops is even more useful since the different depths of penetration of the EM wave at various frequencies give unique information about the crop structure, vegetation water content, and biomass [44]. Multi-frequency data, available by airborne SAR systems like the AIRSAR, indicate interesting crop classification results [19, 24, 34] due to differences in multi-frequency polarimetric parameters associated with different plant geometries. Rao et al. [50] indicated monotonic decrease in co-pol phase difference (ϕ_{HH-VV}) over corn fields at multi-frequency (P-, L-, C-band) polarimetric AIRSAR data. At higher frequencies, radar backscatter return is more correlated with heads and fruiting part, while they are better correlated with wet biomass and foliar area at lower frequencies [27].

With the availability of multi-frequency SAR data from space borne platforms (multi-sensor), crop classification studies have received much attention from remote sensing community [47, 48]. These studies indicated that high biomass crops (e.g., corn) were well classified using the low frequency data, while higher frequency data were needed to accurately classify low biomass crops. New generation SAR system promises better availability of multi-frequency data from Copernicus Sentinel-1 program (C-band), RADARSAT Constellation Mission (C-band), SAOCOM-1A/B (L-band), TerraSAR-X (X-band), and upcoming ROSE-L, Biomass, and NASA-ISRO SAR (NISAR L- and S-band) missions. Nevertheless, diversity in frequency is still only attainable by integrating data from multiple platforms and enabling enhanced crop characterization capabilities by the synergy among these cross platforms [6, 17, 44, 52].

The major features used in crop classification experiments are confined to backscatter intensities at different polarization channels (HH, VV, and HV or VH) [14, 41, 46, 53]. Additional information about the physical nature of the crops can be obtained by generating the target decomposition polarimetric parameters from fully polarimetric SAR data. Moreover, the ratios of individual backscatter coefficients convey additional target scattering information [28, 45]. In addition to the individual backscattering coefficients, polarimetric target decomposition parameters obtained from both model based and eigenvalue-eigenvector based polarimetric parameters [9, 60, 65] can better characterize different crops.

Identifying relevant and important polarimetric parameters is an integral part of machine learning studies. Many machine learning techniques employ parameter selection to form subsets resulting in dimensionality reduction. Some of the machine learning techniques used for crop classification using SAR data are neural network classifier [8], maximum likelihood classifier [33], Wishart classification [33, 55], Support Vector Machines (SVM) [39], decision-tree classifier [47].

Among different classifiers, the Random Forest (RF) is gaining much attention for land cover classification over agricultural areas [10, 11, 13, 22, 36, 38, 42, 58, 64]. RF also provides a parameter subset as a part of its classification technique. RF additionally handles the diverse dynamic ranges of the polarimetric parameters and it does not require parameter scaling or normalization. This provides an added

advantage over other techniques since the polarimetric parameters used in crop classification have varied ranges. In this work, we use polarimetric target decomposition parameters obtained from both model based and eigenvalue-eigenvector based polarimetric parameters in addition to the individual backscattering coefficient parameters in a RF classifier.

The rest of this chapter is organized in the following order: Sect. 2 briefly describes the study area and datasets used. Section 3 explains in detail the methodology of multi-frequency crop classification used in this study. Section 4 discusses the results and their subsequent observations and interpretations in depth; and finally, this chapter is succinctly summarized and concluded in Sect. 5.

2 Study Area and Dataset

We conducted the study over the international agricultural super-site at Flevoland area in The Netherlands (Fig. 1). The test site is bounded between $52.266605^{\circ}\text{N}$, 5.648201°E (upper left coordinate) and $52.326725^{\circ}\text{N}$, 5.441733°E (lower right). The terrain is flat and lies ± 3 m below the mean sea level. This region is dominated by agricultural crops and nominal field sizes are ≈ 80 ha. The major crops grown in area includes wheat, barley, potato, sugarbeet, and maize. Secondary crops include rapeseed, pea, onion, steam bean, and grass.

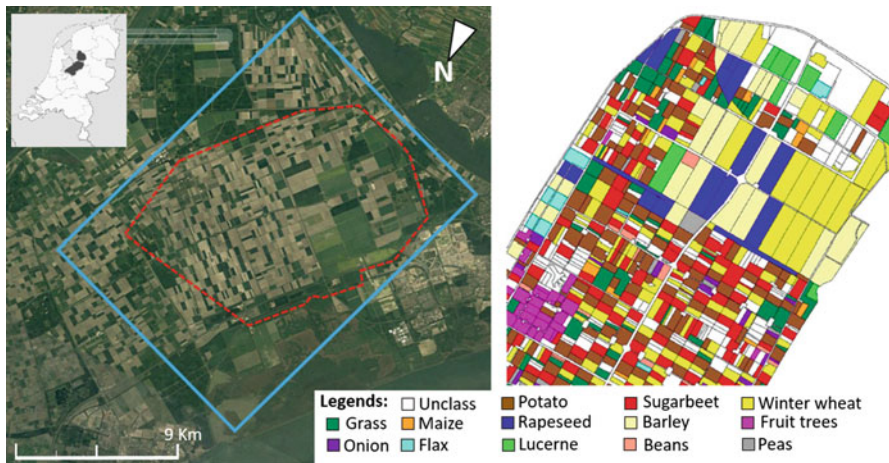


Fig. 1 The Flevoland test site location overlaid on Landsat-5 optical image. The extent of AIRSAR acquisition is presented in cyan rectangular box, and test site in red dashed line. A reference crop map indicates different agricultural crop parcels managed during the JPL-SAR experiment 1991

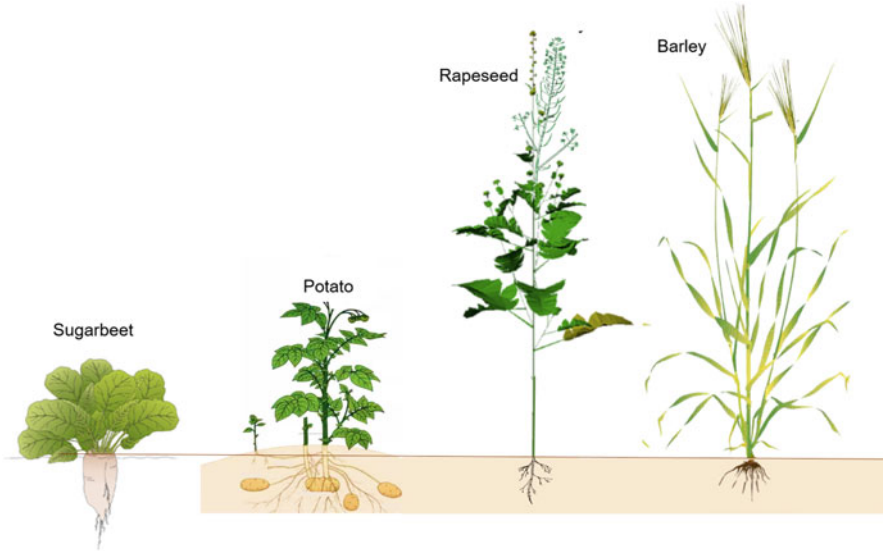


Fig. 2 Plant morphological structures for short-stem broad-leaf (sugarbeet and potato), and long-stem (rapeseed and barley) crops

For the analysis, we group these crops in two types: short-stem broad-leaf (SSBL), and long-stem (LS) (Fig. 2). In the long-stem category, barley and rapeseed were analyzed, while in the SSBL category, sugarbeet and potatoes were included. During the JPL-SAR experiment 1991, the Flevoland test site had 406, 317, 101, and 13 fields of potato, sugar beet, barley, and rapeseed, respectively [20, 54, 62].

We utilized the AIRSAR datasets [3] acquired during the campaign as part of the JPL-SAR experiment 1991. The acquisition of AIRSAR data (Process ID: cm3253, Flight-line: flevoland116-1.91109) coincided with the agricultural growing season in June. The acquisition of AIRSAR data was in multi-frequency (C-band: 5.7 cm, L-band: 25 cm, and P-band: 68 cm) and full-polarimetric mode. The nominal pixel spacing in range and azimuth was $6.66 \text{ m} \times 12.15 \text{ m}$. The AIRSAR data is provided in compressed Stokes product format.¹ We generate 3×3 covariance matrices for individual frequencies, i.e., C-, L-, and P-band from these products of AIRSAR data using PolSARPro toolbox. Other polarimetric features are subsequently generated from the elements of 3×3 covariance matrices. An overview of Pauli RGB images of these multi-frequency datasets over the Flevoland area is shown in Fig. 3.

¹ <https://airsar.asf.alaska.edu/data/cm/cm3253/>.

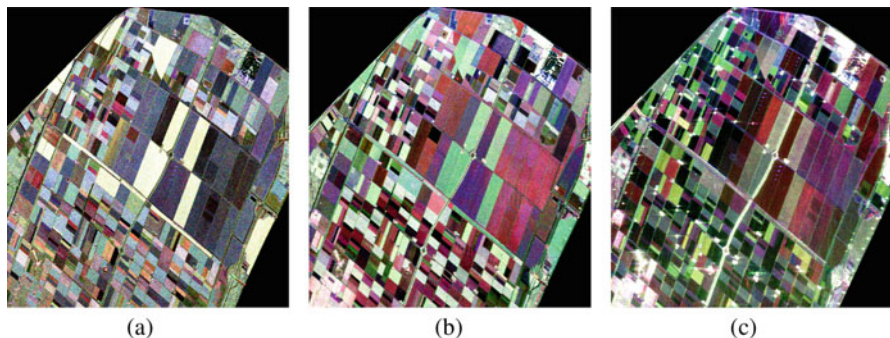


Fig. 3 Multi-frequency Pauli RGB images of the study area (Red= $|S_{HH} - S_{VV}|$, Green= $2|S_{HV}|$, and Blue= $|S_{HH} + S_{VV}|$). These three colors are the magnitudes of the scattering matrix elements when they are expressed in the Pauli basis. (a) C-band. (b) L-band. (c) P-band

PolSAR Data Preprocessing Guidelines

SAR data processing guidelines in PolSARPro: https://github.com/dipankar05/springer-multifrequencySAR-crop/blob/main/PolSARpro_features_guide.pdf

In the H-V basis, we generated co-polarized phase (ϕ_{HH-VV}), co-polar coherence amplitude (ρ_{HHVV}), co- and cross-pol ratios ($\sigma_{HH}^0/\sigma_{VV}^0$, $\sigma_{HV}^0/\sigma_{HH}^0$, and $\sigma_{HV}^0/\sigma_{VV}^0$), and diagonal elements of the 3×3 covariance matrix. Subsequently, the covariance matrix in H-V basis is transformed to the circular R-L basis to obtain the $\sigma_{RR}^0/\sigma_{LL}^0$. Apart from these non-decomposition (ND) parameters, we derive 14 features from the target decomposition parameters [9, 60, 65], as presented in Table 1.

The training datasets for crop classification are generated using the reference map (Fig. 1) provided during the campaign and aptly used in literature [20, 62]. From these reference map, we generated training samples by drawing region of interest, which we kept at $\approx 15\%$ sampling rate in this research. This selection of sampling rate is taken considering the stability in classification accuracies above 15%.

3 Methodology

3.1 Random Forest

Random Forests (RFs) are an ensemble learning technique for classification and regression which is constructed by several decision trees that are trained and their results are combined through a voting process by the majority of the individual decision trees [7]. The multiple decision trees of the RF are trained on a boot-

Table 1 SAR polarimetric decomposition (14) and non-decomposition (9) parameters used in this study

Decomposition parameters	Description
Touzi [60]	Touzi symmetric scattering type magnitude ($\alpha_{s_1}, \alpha_{s_2}$)
	Touzi symmetric scattering type phase (ϕ_{s_1}, ϕ_{s_2})
	Kennaugh-Huynen target helicity (τ_{m_1}, τ_{m_2})
Yamaguchi 4-component [65]	Odd-bounce scattering power (P_s), Double-bounce scattering power (P_d)
	Volume scattering power (P_v), Helix scattering power (P_c)
Cloude-Pottier [9]	Entropy (H), Anisotropy (A), Average target scattering mechanism (α); $Span = \sum_{i=1}^3 \lambda_i$
Non-decomposition parameters	
Co-polarized phase	ϕ_{HH-VV}
Co-polar coherence amplitude	$ \rho_{HHVV} $
Co-polarized and Cross-polarized Ratio	Co-polarized: $\sigma_{HH}^0/\sigma_{VV}^0$; Cross-polarized: $\sigma_{HV}^0/\sigma_{HH}^0$ and $\sigma_{HV}^0/\sigma_{VV}^0$ and $\sigma_{RR}^0/\sigma_{RL}^0$
Diagonal elements of the 3×3 covariance matrix	C_{11}, C_{22} and C_{33}

strapped sample of the original training data. At each node of every decision tree, one among a randomly selected subset of input parameters is chosen as the best split and subsequently used for node splitting [37]. Each tree uses only a portion of the input samples (typically two-third) for the training while the remaining roughly one-third (referred to as Out-Of-Bag (OOB)) of the samples are used to validate the accuracy of the prediction. In general, RF increases the diversity among the decision trees by randomly resampling the data with replacement and by randomly changing the parameter subsets for node splitting at each node of every decision tree.

3.2 Parameter Importance Evaluation

Parameter importance evaluation helps in identifying the most relevant parameters out of the total set for classification by ranking them in descending order of their importance. In RF, for every decision tree the misclassification rate is calculated from the OOB observations. The parameter whose importance is to be evaluated, is randomly permuted for the OOB observations, and then the modified OOB values are passed down the tree to get new predictions. This difference in the misclassification rate between the modified and the original OOB observations averaged over all trees, is the parameter importance measure which is used in this study [59]. This difference in classification accuracy before and after random permutation of the parameter whose importance is to be determined is the Mean

Decrease Accuracy (MDA) [49] measure used in this study. RF based parameter selection using this technique was studied in [2, 12, 21]. It is necessary to note that in this study the original MDA scores were normalized (the highest MDA score was set to 100 and the others were scaled accordingly) for the sake of comparison.

3.3 Partial Probability Plot

The RF is capable of identifying important parameters and generating partial dependence plots [18, 23] which may be used to establish relationships between the parameters and the predicted classes. The partial dependence plots provide a unique way to visualize the marginal effect of a parameter on the classification using RF. The partial dependence function is given as in (1) [18],

$$\tilde{f}(x) = \frac{1}{n} \sum_{i=1}^n f(x, x_{i,c}) \quad (1)$$

where x is the parameter for which partial dependence is sought, and $x_{i,c}$ is the other parameters in the data.

$$f(x) = \log p_k(x) - \sum_j \log p_j(X)/K \quad (2)$$

The logits (i.e., log of fraction of votes) is the predicted classification function as given in (2). Here K is the number of classes and p_j is the proportion of votes for class j .

The partial dependency plots produced with probability distribution based on scaled margin distances are the partial probability plot used in this study. The partial probability plot provides a visual representation of the probability of occurrence of a class for each parameter over its entire dynamic range [5, 49]. By partialling out the average effect of all other parameters, we can analyze the influence of a given parameter on the probability of occurrence of the predicted class.

In this study, the partial probability plot of polarimetric parameters was useful for crop characterization and separation. For polarimetric parameters with diverse ranges, the partial probability plot helped to identify an optimal dynamic range [22] in which the probability of occurrence of “crop” class was ≥ 0.8 . Identification of this range can be useful for crop characterization and separation. Mainly, the partial probability plots help to study the underlying physical scattering mechanisms associated with crops through their diverse optimal dynamic ranges.

3.4 Processing Steps for Parameter Selection and Classification Using RF

The schematic overview of the workflow used in this study is shown in Fig. 4.

The processing steps for multi-frequency crop classification in this study are as follows:

- Polarimetric target decomposition and non-decomposition parameters were generated from the coherency matrix and covariance matrix $\langle [T] \rangle$ and $\langle [C] \rangle$, which resulted in total 23 parameters.
- RF was created using 1000 decision trees and 23 parameters. It was decided to use 1000 decision trees since Breiman [7] suggested that as many trees as possible can be used in the RF ensemble since they do not overfit.
- Parameter selection and classification performed for each band (C-, L- & P-band) individually.
- The top 10 parameters with the highest MDA scores were chosen as the parameter subset in this study for multi-frequency crop classification.
- RF parameter ranking in co-ordination with partial probability plots were used to analyze separability and mixing among crop classes.

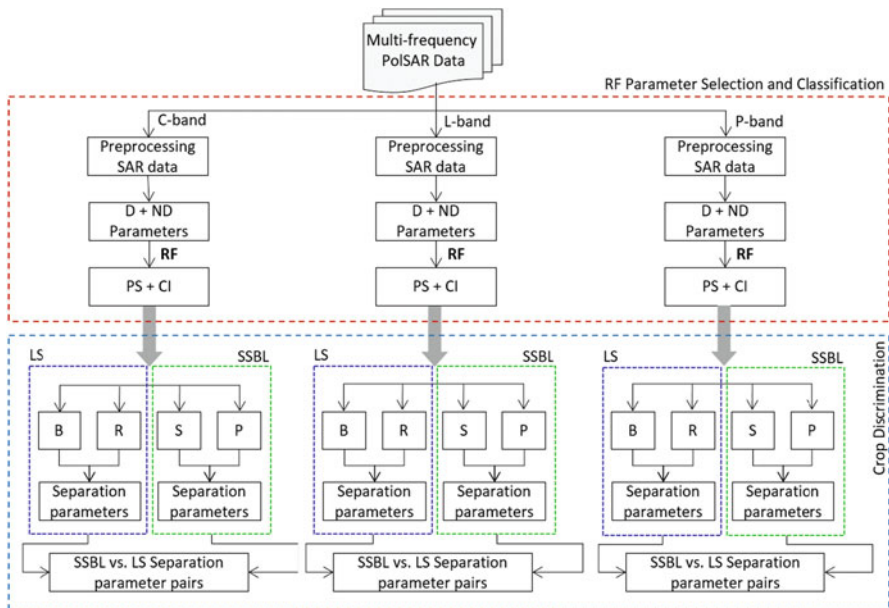


Fig. 4 Schematic workflow for multi-frequency crop classification and crop separability analysis using RF. D: Decomposition parameters, ND: Non-decomposition parameters, PS: Parameter Selection, CI: Classification, SSBL: Short-stem broad-leaf crop, LS: Long-stem crop, B: Barley, R: Rapeseed, S: Sugarbeet, P: Potato

- In addition, we analyze separability of crop classes by measuring dissimilarity between partial probability plot curves using the Frèchet distance [1, 16]. The Frèchet distance closer to 0 indicates similarity between curves while closer to 1.0 indicates distinct curves.

RF Classification Code

RF Classification code along with partial probability plotting and Frèchet distance for R: <https://github.com/dipankar05/springer-multifrequencySAR-crop/tree/main/Codes>

4 Results and Discussion

Classification over the Flevoland area with the help of 23 polarimetric parameters was conducted. Out of this, only the top 10 were selected and subsequently used for classification with RF. These 10 parameters were used since it was observed that the Overall Accuracy (OA) does not change significantly beyond these parameters as shown in Fig. 5.

The number of training samples for each crop is given in Table 2. For ease of analysis, number of training and testing points of only 4 selected crops (2 in SSBL and 2 in LS categories) have been included out of the total 12 crop classes in the Flevoland study area.

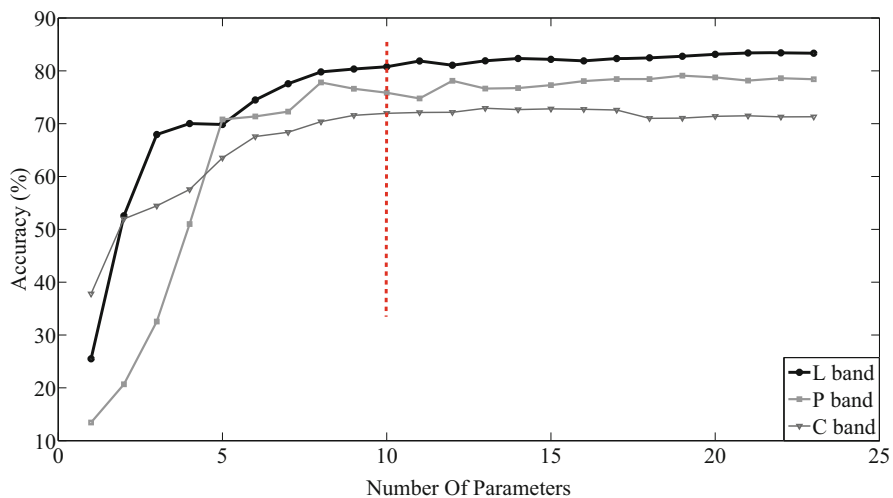


Fig. 5 Overall RF classification accuracy for number of parameters at multiple frequencies

Table 2 Flevoland: Number of training and test samples.

Crop	#Train	#Test1	#Test2
Barley	1100	600	592
Rapeseed	1115	563	571
Sugarbeet	2808	1342	1303
Potato	2357	1108	1119
Winter Wheat	1029	615	720
Lucerne	1516	636	688
Flax	1340	789	902
Beans	706	464	393
Fruit trees	1451	725	743
Grass	2281	1026	1251
Peas	855	439	409
Unclass	3009	1504	1314

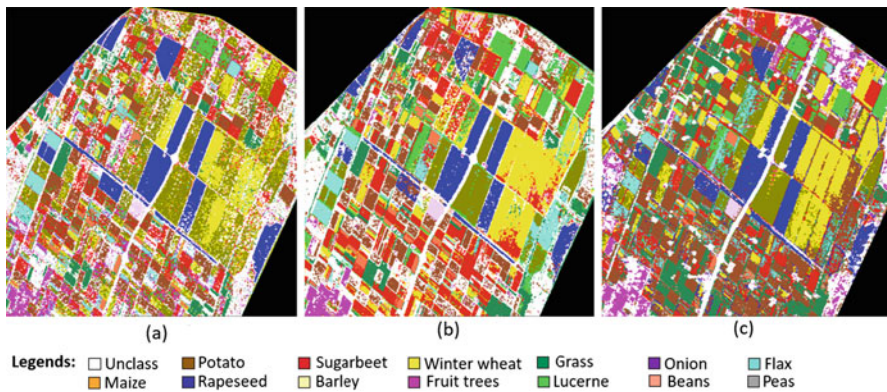


Fig. 6 RF classified images for C-, L-, and P-band over the Flevoland area. (a) C-band. (b) L-band. (c) P-band

The multi-frequency RF classified images are shown in Fig. 6. The overall classification accuracy for the multi-frequency crops are given in Table 3. In the following subsections we analyze the LS (barley and rapeseed) and SSBL (sugarbeet and potato) crops.

The number of test samples for each crop for the two sets is given in Table 2. The independence between two sets of randomly selected test samples was measured using the Wilcoxon signed-rank test [26, 63] using the sample median. The Wilcoxon signed-rank test evaluated the independence between these two sets of test samples to be $>90\%$ for all three bands (90.84%, 95.41%, and 96.36% for C-, L-, and P-band, respectively).

Table 3 RF Overall Accuracy using top 10 parameters

Class	C	L	P
Barley	66.81	91.73	86.18
Beans	78.45	86.85	80.39
Flax	92.65	99.75	90.75
Grass	67.84	43.18	63.06
Lucerne	47.64	99.69	83.49
Peas	71.99	97.77	80.52
Potatoes	84.32	95.42	94.57
Rapeseed	91.45	81.54	73.42
Winter wheat	77.03	73.13	44.27
Fruit trees	57.33	62.29	84.57
Sugarbeets	46.49	86.22	66.89
Unclass	22.39	63.27	98.53
Overall user accuracy (%)	71.95	80.77	75.85
Kappa	0.68	0.78	0.73

Table 4 RF classification accuracy for the two independent test samples

Band	Test1 (%)	Test2 (%)
C	71.31	68.08
L	83.33	83.30
P	78.41	77.14

! Attention

The Wilcoxon signed-rank test [26, 63] was used in this study since it can be performed without assuming underlying distribution of the samples [40]; and we know that the polarimetric parameters in general for SAR data seldom follow normal distribution. The parameter median was calculated instead of mean since the mean of the parameter can be misleading when outliers are present in the data [35] which is possible for these kind of parameters.

The multi-frequency RF classification accuracy for the two test samples is given in Table 4. In addition, we provided the normalized MDA scores of top 10 parameters at C-, L-, and P-band for individual crops in Table 5, 6, 7, and 8.

4.1 Separation Among Long-Stem (LS) Crops

Stems of barley are long in length with thin diameter, and canopy consists of narrow leaves, while rapeseed plants have ramified stems with secondary and tertiary stems and pods. The RF parameter selection helped us to identify the important parameters for barley and rapeseed classification at different frequencies. The optimal dynamic range of these parameters were analyzed and the parameters which were used

Table 5 Normalized MDA scores of the top 10 parameters for barley classification

C-band	MDA scores	L-band	MDA scores	P-band	MDA scores
P_v	100	$\sigma_{HV}^0/\sigma_{HH}^0$	100	$\sigma_{HH}^0/\sigma_{VV}^0$	100
P_s	65.00	P_v	98.44	P_v	97.07
P_d	47.76	H	89.32	P_d	81.91
H	39.77	$\sigma_{HH}^0/\sigma_{VV}^0$	87.99	A	75.65
$ \rho_{HHVV} $	38.45	P_d	83.26	$\sigma_{HV}^0/\sigma_{VV}^0$	71.40
$\sigma_{HH}^0/\sigma_{VV}^0$	34.33	α_{s_1}	72.97	P_s	65.33
P_c	32.52	Φ_{s_2}	69.29	Φ_{s_1}	62.44
α	29.71	P_s	68.78	$\sigma_{RR}^0/\sigma_{RL}^0$	60.91
Φ_{s_2}	27.92	$ \rho_{HHVV} $	64.97	$\sigma_{HV}^0/\sigma_{HH}^0$	60.24
$\sigma_{HV}^0/\sigma_{HH}^0$	27.67	α_{s_2}	64.81	$SPAN$	56.81

Table 6 Normalized MDA scores of the top 10 parameters for rapeseed classification

C-band	MDA scores	L-band	MDA scores	P-band	MDA scores
P_v	100	P_c	100	P_v	100
$\sigma_{HV}^0/\sigma_{HH}^0$	53.07	$\sigma_{HV}^0/\sigma_{VV}^0$	74.41	P_c	83.93
$SPAN$	41.30	P_v	64.36	$\sigma_{HH}^0/\sigma_{VV}^0$	73.31
H	33.64	P_d	53.59	A	53.17
P_c	31.94	$\sigma_{HV}^0/\sigma_{HH}^0$	53.22	P_d	48.88
$\sigma_{HH}^0/\sigma_{VV}^0$	31.78	τ_{m_1}	48.50	Φ_{s_2}	46.70
$\sigma_{RR}^0/\sigma_{RL}^0$	31.49	Φ_{s_2}	47.58	Φ_{s_1}	45.71
α	30.94	Φ_{s_1}	44.35	$ \rho_{HHVV} $	43.95
$\sigma_{HV}^0/\sigma_{VV}^0$	30.84	P_s	43.10	$\sigma_{HV}^0/\sigma_{HH}^0$	42.04
C_{22}	30.26	$\sigma_{HH}^0/\sigma_{VV}^0$	40.86	$\sigma_{HV}^0/\sigma_{VV}^0$	40.68

Table 7 Normalized MDA scores of the top 10 parameters for sugarbeet classification

C-band	MDA scores	L-band	MDA scores	P-band	MDA scores
P_v	100	$\sigma_{HV}^0/\sigma_{HH}^0$	100	τ_{m_2}	100
P_d	60.12	P_d	85.83	$\sigma_{HV}^0/\sigma_{HH}^0$	83.60
A	53.91	$\sigma_{HH}^0/\sigma_{VV}^0$	84.46	P_v	81.26
Φ_{s_2}	52.75	P_v	84.46	τ_{m_1}	72.49
τ_{m_1}	52.13	H	70.99	$\sigma_{HH}^0/\sigma_{VV}^0$	72.31
P_c	50.83	P_s	61.95	$\sigma_{HV}^0/\sigma_{VV}^0$	71.26
P_s	49.95	$ \rho_{HHVV} $	58.80	P_d	67.16
$\sigma_{HH}^0/\sigma_{VV}^0$	47.94	Φ_{s_2}	58.75	A	67.09
$\sigma_{HV}^0/\sigma_{HH}^0$	46.84	$\sigma_{HV}^0/\sigma_{VV}^0$	57.42	α_{s_2}	63.76
$\sigma_{HV}^0/\sigma_{VV}^0$	38.88	α	56.23	P_s	61.14

to discriminate barley from rapeseed were identified to be P_v and P_c at C-band. The partial probability plots of P_v for barley and rapeseed given in Fig. 7a and b, respectively, show separation between them.

Table 8 Normalized MDA scores of the top 10 parameters for potato classification

C-band	MDA scores	L-band	MDA scores	P-band	MDA scores
$ \rho_{HHVV} $	100	P_v	100	P_v	100
P_d	98.30	$\sigma_{HV}^0/\sigma_{VV}^0$	68.62	P_d	87.10
P_v	97.06	P_d	64.48	Φ_{s_2}	85.08
Φ_{s_2}	91.74	$\sigma_{HV}^0/\sigma_{HH}^0$	58.99	Φ_{s_1}	82.84
A	73.55	$\sigma_{HH}^0/\sigma_{VV}^0$	55.63	A	73.82
$\sigma_{HH}^0/\sigma_{VV}^0$	68.39	Φ_{s_1}	54.30	$\sigma_{HV}^0/\sigma_{VV}^0$	73.66
τ_{m_1}	57.96	Φ_{s_2}	53.57	$\sigma_{HH}^0/\sigma_{VV}^0$	73.54
P_c	55.55	P_s	52.91	$\sigma_{HV}^0/\sigma_{HH}^0$	71.59
$\sigma_{HV}^0/\sigma_{HH}^0$	52.90	τ_{m_1}	48.79	P_s	66.75
P_s	49.34	α	46.14	τ_{m_2}	63.48

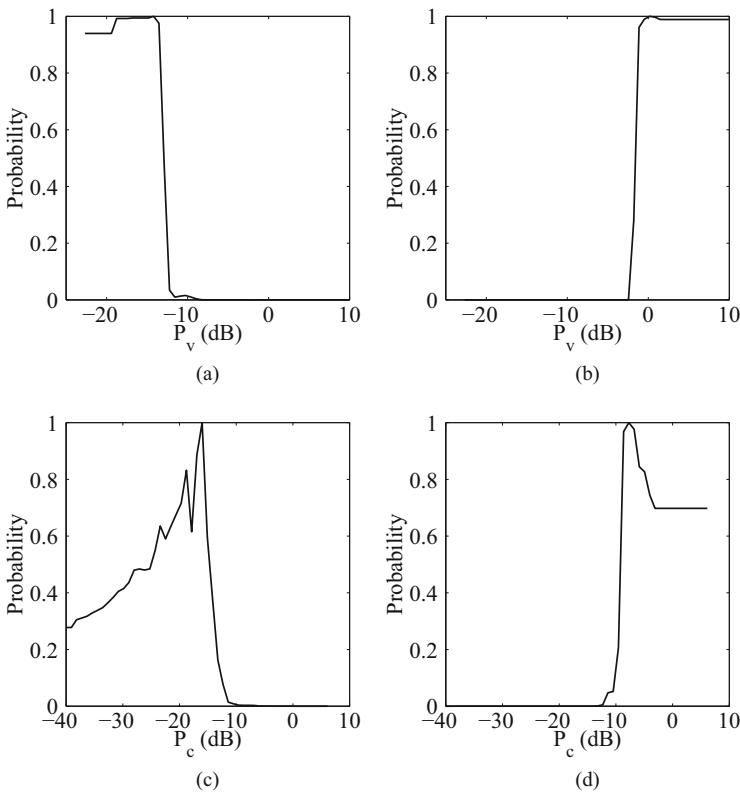


Fig. 7 Separation between barley and rapeseed using partial probability plots. (a) Barley C-band. (b) Rapeseed C-band. (c) Barley C-band. (d) Rapeseed C-band

The presence of secondary stems in rapeseed also gives rise to complex multiple or helical scattering thereby contributing to high P_c as compared to barley. The partial

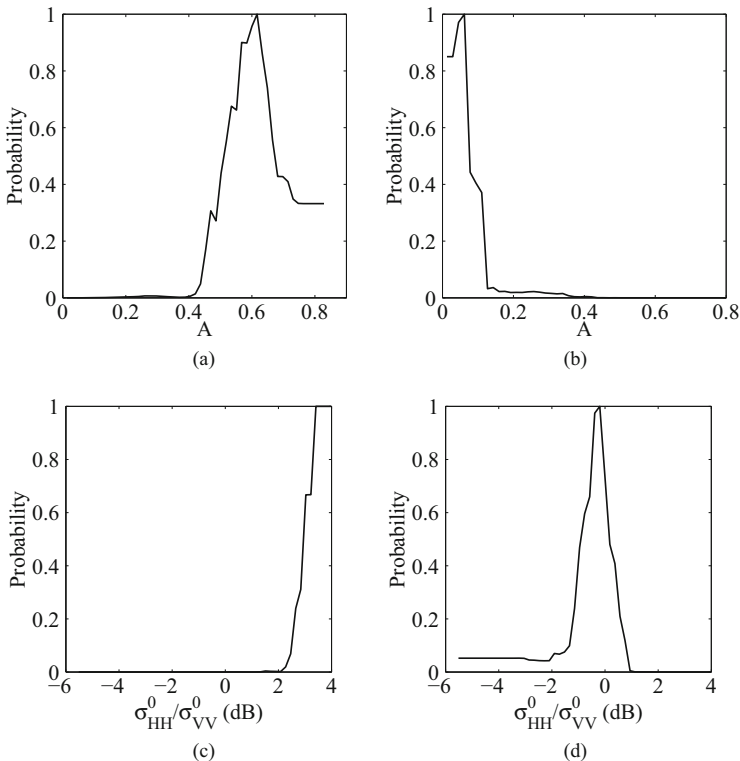


Fig. 8 separation between sugarbeet and potato using partial probability plots. (a) Sugarbeet L-band. (b) Potato L-band. (c) Sugarbeet L-band. (d) Potato L-band

probability plots of P_c for barley and rapeseed given in Fig. 7c and d, respectively, show separation between them.

Further, the Frèchet distance between the partial probability plots of P_v at C-band between barley and rapeseed was found to be 1.0, indicating high dissimilarity between the curves. Furthermore, the Frèchet distance between the partial probability plots of P_c at C-band between barley and rapeseed was found to be 0.99.

4.2 Separation Among Short-Stem Broad-Leaf (SSBL) Crops

The RF parameter selection and the optimal dynamic range evaluation using the partial probability plot was useful to discriminate sugarbeet from potatoes both being short-stem broad-leaf category. It was observed that sugarbeet and potatoes can be discriminated using A , and $\sigma_{HH}^0/\sigma_{VV}^0$ for L-band as shown in Fig. 8.

From the partial probability plots, the $\sigma_{HH}^0/\sigma_{VV}^0$ peak was observed to be more for sugarbeet as compared to potatoes at L-band as shown in Fig. 8c and d, respectively. Similar observation was reported by Skriver et al. [57]. This may be due to the relatively smooth surface scattering from sugarbeet compared to the rough surface scattering from potato. The anisotropy A for sugarbeet was reported to be higher than potatoes in L-band [51] which was also observed in our study as shown in Fig. 8a and b, respectively. Anisotropy can be effectively characterized for random scatterers for which $H \geq 0.7$ [32]. In our study for L-band, sugarbeet and potatoes were both observed to have $H \geq 0.7$. However, there was a difference in their anisotropy. Anisotropy is high when there is a big difference between the second and third scattering mechanisms. $A \approx 0$ implies that the second and the third dominant scattering mechanisms are almost the same. So in case of potato it seems that the second and third scattering mechanisms were equally dominant while for sugarbeet the third scattering mechanism (mostly noise) was non-existent as compared to the second dominant scattering mechanism.

The Frèchet distance between the partial probability plots of A at L-band for sugarbeet and potato was found to be 0.9. Additionally the Frèchet distance between the partial probability plots of $\sigma_{HH}^0/\sigma_{VV}^0$ at L-band for sugarbeet and potato was found to be 1.

4.3 Separation Between SSBL and LS Crops

Short-stem broad-leaf (SSBL) crops can be realized as canopy consisting of disc like scatterers. Unlike SSBL, the long-stem (LS) crops have predominantly cylindrical scatterers [20, 61]. It is important to identify polarimetric parameters which separate these two crop types. In this study we have determined a pair of polarimetric parameters based on the highest difference of MDA scores from the parameters selected by RF which best separate SSBL from LS crops. We first calculated the normalized MDA score difference between the same pair of parameters for two different crop types. The pair having the highest normalized MDA score difference between the two crops was selected for separation among the crop types. A few pairs of polarimetric parameters were thus identified which successfully separate SSBL from LS crops.

From Fig. 9a, it can be seen that for L-band, rapeseed (LS crop) and sugarbeet (SSBL crop) can be separated using $\sigma_{HH}^0/\sigma_{VV}^0$ and P_v and thus misclassification is avoided between them. The separation is about 4.0 dB for $\sigma_{HH}^0/\sigma_{VV}^0$ and about 10 dB for P_v . From Fig. 9b, it can be seen that for L-band, barley (LS crop) and potato (SSBL crop) can be separated successfully using the polarimetric parameter pair of $\sigma_{HV}^0/\sigma_{HH}^0$ and P_v . The separation is about 4.0 dB for $\sigma_{HV}^0/\sigma_{HH}^0$ and about 10.1 dB for P_v . As can be seen from Tables 5 and 7, the parameters $\sigma_{HH}^0/\sigma_{VV}^0$, and P_v are part of the top 10 parameters required for L-band classification of rapeseed and sugarbeet, respectively, thus validating the RF parameter ranking and

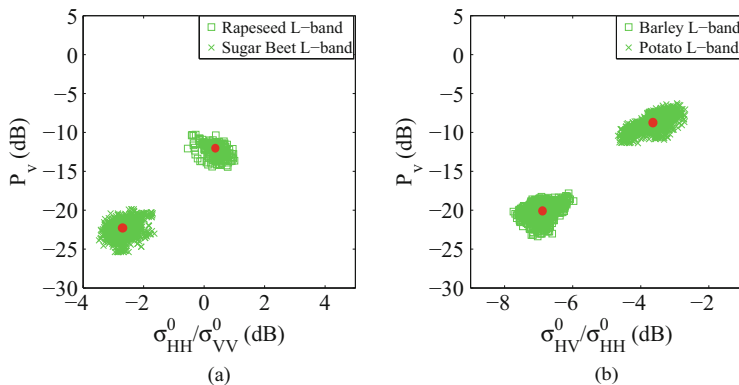


Fig. 9 Separation between small stem and broad-leaf crops at L-band. (a) Rapeseed-Sugarbeet. (b) Barley-Potato

its contribution in crop classification. Also $\sigma_{HV}^0/\sigma_{HH}^0$ and P_v are part of the top 10 parameters required for L-band classification of barley and sugarbeet, respectively, as seen from Tables 6 and 8.

From Fig. 10a, it can be seen that for P-band, rapeseed (LS crop) and sugarbeet (SSBL crop) can be separated using $\sigma_{HV}^0/\sigma_{HH}^0$ and P_d . The separation is about 2.5 dB for $\sigma_{HV}^0/\sigma_{HH}^0$ and about 6.0 dB for P_d . From Fig. 10b, it can be seen that for C-band, rapeseed (LS crop) and potato (BBSL crop) can be separated successfully using the polarimetric parameter pair of $\sigma_{HV}^0/\sigma_{HH}^0$ and P_c . The separation is about 2.5 dB for $\sigma_{HV}^0/\sigma_{HH}^0$ and about 4.0 dB for P_c . As can be seen from Tables 5 and 7, the parameters $\sigma_{HV}^0/\sigma_{VV}^0$, and P_d are part of the top 10 parameters required for P-band classification of rapeseed and sugarbeet, respectively, thus validating the RF parameter ranking and its contribution in crop classification. In fact they are amongst the top 5 ranked parameters for both crops. Also $\sigma_{HV}^0/\sigma_{HH}^0$ and P_c are part of the top 10 parameters required for C-band classification of rapeseed and potato, respectively, as seen from Tables 5 and 8.

4.4 Analyzing the Mixing Among Crop Classes

The RF classification accuracies were used to correlate the parameter ranking with the underlying physical scattering mechanism related to crop targets. RF based partial probability plots were used to study mixing among crop classes, since these plots give the marginal effect of the parameter on the classification accuracy. Analyzing partial probability plots to study mixing among crop classes helps us validate the RF parameter ranking. This is because the amount of mixing among crop classes is related to the more dominant parameters having similar partial

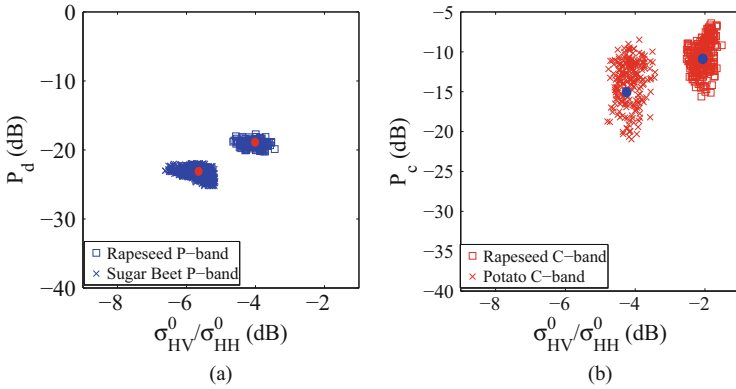


Fig. 10 Separation between small stem and broad-leaf crops at (a) P- and (b) C-band

Table 9 Confusion matrix for the four crop classes at C-band (%)

C	Barley	Rapeseed	Potato	Sugarbeet
Barley	84.54	0	0	1.27
Rapeseed	0	91.99	0	0
Potato	0	0	78.84	0
Sugarbeet	2.88	0.1	6.28	67.71

Table 10 Confusion matrix for the four crop classes at P-band (%)

P	Barley	Rapeseed	Potato	Sugar_beet
Barley	100	0	0	0
Rapeseed	0	76.91	0	0.28
Potato	0	18.34	99.63	0
Sugar_beet	0	0	0	59.21

probability plots. Distinct partial probability plots imply low to no mixing among crop classes.

In the confusion matrix (Table 9), it was observed that 6.28% of potato was misclassified as sugarbeet at C-band. The odd-bounce scattering power P_s (ranked 7 and 10 for sugarbeet and potato, respectively) had similar partial probability plots (Frèchet distance = 0.18) shown in Fig. 11a and b, respectively. This may be due to the fact that both potato and sugarbeet, being short-stem broad-leaf crops exhibit single bounce scattering from the wide leaf of the crop. Again from Table 10, in P-band, high mixing of 18.34% was observed among rapeseed and potato classes. The volume scattering power P_v is dominant for crops with ramified stems and was observed to be Rank 1 for both rapeseed and potato, respectively. P_v was observed to have relatively similar partial probability plots (Frèchet distance = 0.32) as shown in Fig. 12a and b, respectively.

In contrast to C and P-band, significant mixing was not observed among the four crops in L-band. This means the relevant parameters chosen by RF all have distinct partial probability plots at L-band. It was observed that barley and rapeseed crop classes did not have any intermixing among each other at all frequencies. Hence,

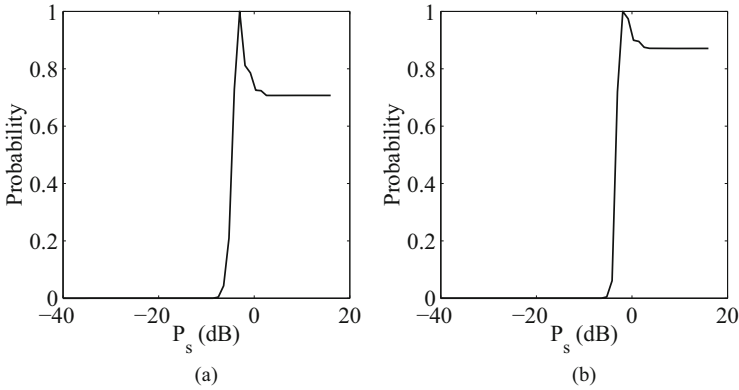


Fig. 11 Mixing among sugarbeet and potato classes at C-band. (a) Sugarbeet. (b) Potato

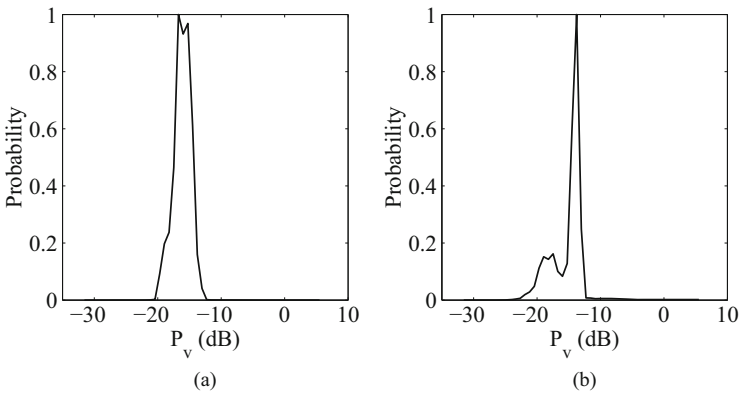


Fig. 12 Mixing among rapeseed and potato classes at P-band. (a) Rapeseed. (b) Potato

it is interesting to note that the top parameters selected by RF have distinct partial probability plots for barley and rapeseed at all frequencies. The partial probability plots of $\sigma_{HV}^0/\sigma_{HH}^0$ (ranked 1 and 5 for barley and rapeseed, respectively) at L-band is shown in Fig. 13a and b. The Frèchet distance between them was observed to be 0.97 indicating distinct plots. It is to be noted that the confusion matrix given in Tables 9 and 10 do not sum up to 100%, since only classification accuracies of 4 selected crops have been included out of the total 12 crop classes in the Flevoland study area.

Some notable insights from the above study:

- When top ranked parameters of different crops have similar partial probability plots, the crop classes can be easily mixed.
- Mixing among crop classes reduces as the rankings get lowered for parameters with similar partial probability plots. For instance, 18.34% mixing was observed

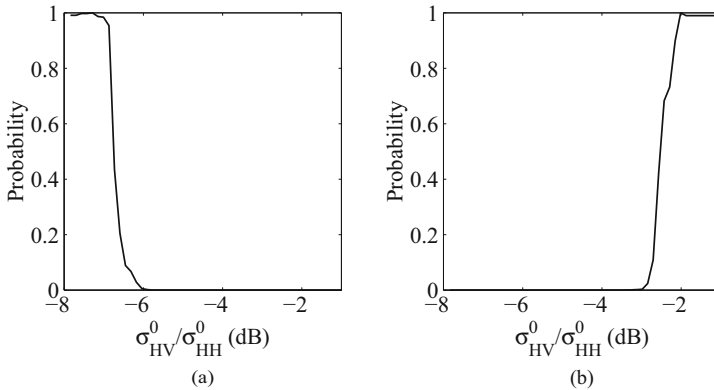


Fig. 13 Distinct partial probability plots of selected parameters for barley and rapeseed (L-band). (a) Barley. (b) Rapeseed

among crops classes for top ranked parameters (rank 1–3) while only 6.28% mixing for bottom ranked parameters (rank 7–10).

- Crop classes which do not mix, their top ranked parameters mostly have distinct partial probability plots.

5 Summary

In this study we utilized polarimetric target decomposition and non-decomposition parameters for crop analysis. It was observed that the model-based decomposition powers along with the ratio of the backscattering coefficients were important for crop classification. Moreover the Eigenvalue/Eigenvector based decomposition parameters were useful for critical analysis of crops in some cases. It was observed that similar crop types have different scattering properties which was evident from the partial probability plots of the important polarimetric parameters. The parameter selection by RF and the evaluated normalized MDA scores for multi-frequency data was thus useful for crop analysis. Separation between Long-stem and short-stem broad-leaf crops at different frequencies was made possible using a pair of polarimetric parameters having the highest normalized MDA score difference between the crop types. It was seen that the crops which were separated physically by the polarimetric parameter ranges were also the ones having the highest difference between their normalized MDA scores. This was helpful in validating the RF parameter ranking at multiple frequencies.

This study can be extended by incorporating parameters from multiple sources, for example: Leaf Area Index (LAI), Normalized Difference Vegetation Index (NDVI), soil moisture, temperature, etc., in addition to polarimetric parameters. The ranking of these parameters from multiple sources can be useful for diverse

crop analysis studies. The partial probability plots which evaluate optimal dynamic ranges can be useful for further analysis like crop yield, annual crop growth monitoring, etc. Multi-temporal analysis in addition to multi-frequency can be of an added advantage for crop studies. Evaluation of optimal dynamic ranges for crop parameters over the entire growth stage for multi-temporal datasets will be very useful for agriculture studies like crop planning and harvest.

Acknowledgments The authors would like to thank the NASA/JPL for providing AIRSAR data products. Authors acknowledge the GEO-AWS Earth Observation Cloud Credits Program, which supported the computation on AWS cloud platform through the project: “AWS4AgriSAR-Crop inventory mapping from SAR data on cloud computing platform.”

References

1. Alt H, Godau M (1995) Computing the Fréchet distance between two polygonal curves. *International Journal of Computational Geometry & Applications* 5(01n02):75–91
2. Archer KJ, Kimes RV (2008) Empirical characterization of random forest variable importance measures. *Computational Statistics & Data Analysis* 52(4):2249–2260
3. ASF (Retrieved from ASF DAAC 25 December 2020) Dataset: AIRSAR, NASA 1991. <https://asf.alaska.edu/>
4. Bargiel D (2017) A new method for crop classification combining time series of radar images and crop phenology information. *Remote sensing of environment* 198:369–383
5. Baruch-Mordo S, Evans JS, Severson JP, Naugle DE, Maestas JD, Kiesecker JM, Falkowski MJ, Hagen CA, Reese KP (2013) Saving sage-grouse from the trees: A proactive solution to reducing a key threat to a candidate species. *Biological Conservation* 167(0):233–241, <https://doi.org/10.1016/j.biocon.2013.08.017>, <http://www.sciencedirect.com/science/article/pii/S0006320713002917>
6. Blaes X, Vanhalle L, Defourny P (2005) Efficiency of crop identification based on optical and SAR image time series. *Remote sensing of environment* 96(3):352–365
7. Breiman L (2001) Random forests. *Machine learning* 45(1):5–32
8. Chen K, Huang W, Tsay D, Amar F (1996) Classification of multifrequency polarimetric SAR imagery using a dynamic learning neural network. *Geoscience and Remote Sensing, IEEE Transactions on* 34(3):814–820
9. Cloude S, Pottier E (1997) An entropy based classification scheme for land applications of polarimetric SAR. *Geoscience and Remote Sensing, IEEE Transactions on* 35(1):68–78, <https://doi.org/10.1109/36.551935>
10. Deschamps B, McNairn H, Shang J, Jiao X (2012) Towards operational radar-only crop type classification: comparison of a traditional decision tree with a random forest classifier. *Canadian Journal of Remote Sensing* 38(1):60–68
11. Dey S, Mandal D, Robertson LD, Banerjee B, Kumar V, McNairn H, Bhattacharya A, Rao Y (2020) In-season crop classification using elements of the Kennaugh matrix derived from polarimetric radarsat-2 SAR data. *International Journal of Applied Earth Observation and Geoinformation* 88:102059
12. Díaz-Uriarte R, De Andres SA (2006) Gene selection and classification of microarray data using random forest. *BMC bioinformatics* 7(1):3
13. Dingle Robertson L, M Davidson A, McNairn H, Hosseini M, Mitchell S, de Abelleira D, Verón S, Le Maire G, Plannells M, Valero S, et al. (2020) C-band synthetic aperture radar (sar) imagery for the classification of diverse cropping systems. *International Journal of Remote Sensing* 41(24):9628–9649

14. Ferrazzoli P, Guerriero L, Schiavon G (1999) Experimental and model investigation on radar classification capability. *Geoscience and Remote Sensing, IEEE Transactions on* 37(2):960–968
15. Foody G, McCulloch M, Yates W (1994) Crop classification from c-band polarimetric radar data. *International Journal of Remote Sensing* 15(14):2871–2885
16. Fréchet MM (1906) Sur quelques points du calcul fonctionnel. *Rendiconti del Circolo Matematico di Palermo (1884–1940)* 22(1):1–72
17. Freeman A, Villasenor J, Klein J, Hoogeboom P, Groot J (1994) On the use of multi-frequency and polarimetric radar backscatter features for classification of agricultural crops. *International Journal of Remote Sensing* 15(9):1799–1812
18. Friedman JH (2001) Greedy function approximation: A gradient boosting machine. *The Annals of Statistics* 29(5):pp. 1189–1232, <http://www.jstor.org/stable/2699986>
19. Gonzalez-Sampedro M, Le Toan T, Davidson M, Moreno J (2002) Assessment of crop discrimination using multi-site databases. *EUROPEAN SPACE AGENCY-PUBLICATIONS-ESA SP 475:63–68*
20. González Sanpedro M, et al. (2008) Optical and radar remote sensing applied to agricultural areas in Europe. *Universitat de València*
21. Hariharan S, Tirodkar S, De S, Bhattacharya A (2014) Variable importance and random forest classification using radarsat-2 PolSAR data. In: *Geoscience and Remote Sensing Symposium (IGARSS), 2014 IEEE International*, pp 1210–1213, <https://doi.org/10.1109/IGARSS.2014.6946649>
22. Hariharan S, Mandal D, Tirodkar S, Kumar V, Bhattacharya A, Lopez-Sanchez JM (2018) A novel phenology based feature subset selection technique using random forest for multitemporal PolSAR crop classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 11(11):4244–4258
23. Hastie T, Tibshirani R, Friedman J, Franklin J (2005) The elements of statistical learning: data mining, inference and prediction. *The Mathematical Intelligencer* 27(2):83–85
24. Hoekman DH, Vissers MA (2003) A new polarimetric classification approach evaluated for agricultural crops. *Geoscience and Remote Sensing, IEEE Transactions on* 41(12):2881–2889
25. Hoekman DH, Vissers MA, Tran TN (2011) Unsupervised full-polarimetric SAR data segmentation as a tool for classification of agricultural areas. *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of* 4(2):402–411
26. Hollander M, Wolfe DA, Chicken E (2013) *Nonparametric statistical methods*. John Wiley & Sons
27. Inoue Y, Kurosu T, Maeno H, Uratsuka S, Kozu T, Dabrowska-Zielinska K, Qi J (2002) Season-long daily measurements of multifrequency (Ka, Ku, X, C, and L) and full-polarization backscatter signatures over paddy rice field and their relationship with biological variables. *Remote Sensing of Environment* 81(2–3):194–204
28. Jia M, Tong L, Zhang Y, Chen Y (2013) Multitemporal radar backscattering measurement of wheat fields using multifrequency (l, s, c, and x) and full-polarization. *Radio Science* 48(5):471–481
29. Jiao X, Kovacs JM, Shang J, McNairn H, Walters D, Ma B, Geng X (2014) Object-oriented crop mapping and monitoring using multi-temporal polarimetric radarsat-2 data. *ISPRS Journal of Photogrammetry and Remote Sensing* 96:38–46
30. Kumar V, Rao YS, Bhattacharya A, Cloude SR (2019) Classification assessment of real versus simulated compact and quad-pol modes of alos-2. *IEEE Geoscience and Remote Sensing Letters* 16(9):1497–1501
31. Kussul N, Mykola L, Shelestov A, Skakun S (2018) Crop inventory at regional scale in Ukraine: developing in season and end of season crop maps with multi-temporal optical and SAR satellite imagery. *European Journal of Remote Sensing* 51(1):627–636
32. Lee JS, Pottier E (2009) *Polarimetric radar imaging: from basics to applications*. CRC Press
33. Lee JS, Grunes MR, Pottier E (2001) Quantitative comparison of classification capability: fully polarimetric versus dual and single-polarization SAR. *Geoscience and Remote Sensing, IEEE Transactions on* 39(11):2343–2351

34. Lemoine G, De Grandi G, Sieber A (1994) Polarimetric contrast classification of agricultural fields using maestro 1 AIRSAR data. *International journal of remote sensing* 15(14):2851–2869
35. Leys C, Ley C, Klein O, Bernard P, Licata L (2013) Detecting outliers: do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology* 49(4):764–766
36. Li H, Zhang C, Zhang S, Atkinson PM (2020) Crop classification from full-year fully-polarimetric L-band UAVSAR time-series using the random forest algorithm. *International Journal of Applied Earth Observation and Geoinformation* 87:102032
37. Liaw A, Wiener M (2002) Classification and regression by randomForest. *R news* 2(3):18–22
38. Loosvelt L, Peters J, Skriver H, De Baets B, Verhoest NE (2012) Impact of reducing polarimetric SAR input on the uncertainty of crop classifications based on the random forests algorithm. *Geoscience and Remote Sensing, IEEE Transactions on* 50(10):4185–4200
39. Löw F, Michel U, Dech S, Conrad C (2013) Impact of feature selection on the accuracy and spatial uncertainty of per-field crop classification using support vector machines. *ISPRS Journal of Photogrammetry and Remote Sensing* 85:102–119
40. Lumley T, Diehr P, Emerson S, Chen L (2002) The importance of the normality assumption in large public health data sets. *Annual review of public health* 23(1):151–169
41. Macelloni G, Paloscia S, Pampaloni P, Marliani F, Gai M (2001) The relationship between the backscattering coefficient and the biomass of narrow and broad leaf crops. *Geoscience and Remote Sensing, IEEE Transactions on* 39(4):873–884
42. Mahdianpari M, Mohammadimanesh F, McNairn H, Davidson A, Rezaee M, Salehi B, Homayouni S (2019) Mid-season crop classification using dual-, compact-, and full-polarization in preparation for the Radarsat constellation mission (RCM). *Remote Sensing* 11(13):1582
43. Mandal D, Kumar V, Rao YS (2020) An assessment of temporal RADARSAT-2 SAR data for crop classification using KPCA based support vector machine. *Geocarto International* pp 1–13
44. Mandal D, Bhattacharya A, Rao YS (2021) Radar Remote Sensing for Crop Biophysical Parameter Estimation. Springer
45. McNairn H, Brisco B (2004) The application of c-band polarimetric SAR for agriculture: a review. *Canadian Journal of Remote Sensing* 30(3):525–542
46. McNairn H, Duguay C, Brisco B, Pultz T (2002) The effect of soil and crop residue characteristics on polarimetric radar response. *Remote sensing of environment* 80(2):308–320
47. McNairn H, Shang J, Champagne C, Jiao X (2009) TerraSAR-x and radarsat-2 for crop classification and acreage estimation. In: *Geoscience and Remote Sensing Symposium, 2009 IEEE International, IGARSS 2009, IEEE, vol 2*, pp II–898
48. McNairn H, Shang J, Jiao X, Champagne C (2009) The contribution of ALOS PALSAR multipolarization and polarimetric data to crop classification. *Geoscience and Remote Sensing, IEEE Transactions on* 47(12):3981–3992
49. R Core Team (2013) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, <http://www.R-project.org/>, ISBN 3-900051-07-0
50. Rao K, Rao Y, Wang J (1995) Frequency dependence of polarization phase difference. *International Journal of Remote Sensing* 16(18):3605–3617
51. Riedel T, Liebeskind P, Schmillius C (2002) Seasonal and diurnal changes of polarimetric parameters from crops derived by the Cloude decomposition theorem at L-band. In: *Geoscience and Remote Sensing Symposium, 2002. IGARSS'02. 2002 IEEE International, IEEE, vol 5*, pp 2714–2716
52. Robertson LD, Davidson A, McNairn H, Hosseini M, Mitchell S (2019) Assessment of multi-frequency SAR for crop type classification and mapping. In: *IGARSS 2019–2019 IEEE International Geoscience and Remote Sensing Symposium, IEEE*, pp 489–492
53. Saich P, Borgeaud M (2000) Interpreting ERS SAR signatures of agricultural crops in Flevoland, 1993–1996. *Geoscience and Remote Sensing, IEEE Transactions on* 38(2):651–657
54. Schotten C, Van Rooy W, Janssen L (1995) Assessment of the capabilities of multi-temporal ers-1 SAR data to discriminate between agricultural crops. *International Journal of Remote Sensing* 16(14):2619–2637

55. Skriver H (2012) Crop classification by multitemporal c-and l-band single-and dual-polarization and fully polarimetric SAR. *Geoscience and Remote Sensing, IEEE Transactions on* 50(6):2138–2149
56. Skriver H, Svendsen MT, Nielsen F, Thomsen A (1999) Crop classification by polarimetric SAR. In: *Geoscience and Remote Sensing Symposium, 1999. IGARSS'99 Proceedings. IEEE 1999 International, IEEE, vol 4, pp 2333–2335*
57. Skriver H, Svendsen MT, Thomsen AG (1999) Multitemporal c-and l-band polarimetric signatures of crops. *Geoscience and Remote Sensing, IEEE Transactions on* 37(5):2413–2429
58. Sonobe R, Tani H, Wang X, Kobayashi N, Shimamura H (2014) Random forest classification of crop type using multi-temporal TerraSAR-x dual-polarimetric data. *Remote Sensing Letters* 5(2):157–164
59. Strobl C, Zeileis A (2008) Danger: High power!? Exploring the statistical properties of a test for random forest variable importance. <http://nbn-resolving.de/urn/resolver.pl?urn=nbn:de:bvb:19-epub-2111-8>
60. Touzi R (2007) Target scattering decomposition in terms of roll-invariant target parameters. *Geoscience and Remote Sensing, IEEE Transactions on* 45(1):73–84, <https://doi.org/10.1109/TGRS.2006.886176>
61. Ulaby FT, Dobson MC (1989) *Handbook of radar scattering statistics for terrain*. ARTECH HOUSE, 685 CANTON STREET, NORWOOD, MA 02062(USA), 1989, 500
62. Vissers M, van der Sanden J (1992) Groundtruth collection for the JPL-SAR and ERS-1 campaign in Flevoland and the Veluwe (NL) 1991. Tech. Rep. BCRS 92–26, Netherlands Remote Sensing Board
63. Wilcoxon F (1945) Individual comparisons by ranking methods. *Biometrics bulletin* pp 80–83
64. Xie Q, Wang J, Liao C, Shang J, Lopez-Sanchez JM, Fu H, Liu X (2019) On the use of Neumann decomposition for crop classification using multi-temporal radarsat-2 polarimetric SAR data. *Remote Sensing* 11(7):776
65. Yamaguchi Y, Moriyama T, Ishido M, Yamada H (2005) Four-component scattering model for polarimetric SAR image decomposition. *Geoscience and Remote Sensing, IEEE Transactions on* 43(8):1699–1706, <https://doi.org/10.1109/TGRS.2005.852084>

Automatic Determination of Different Soil Types via Several Machine Learning Algorithms Employing Radarsat-2 SAR Image Polarization Coefficients



Emrullah Acar and Mehmet Sirac Ozerdem

1 Introduction

It is striking at first glance that the soils in the various pits to be opened on any piece of land have different appearances from each other: color, depth, ease of processing; the change of many features such as loose or tight fit causes this difference. As a result, the plant-growing functions of these soils are diverse. If it is taken into account that the fertility of the soil is also impressed by external factors, then it can be smoothly understood how difficult it would be to study and comprehend the soils whose properties vary according to all these factors individually. For this reason, as in the branches of science just like botany, zoology, etc., the systematization and classification of the soils emerges as a necessity. Thus, lands that are similar to each other in terms of their characteristics in terms of providing ease of study and understanding can be gathered under certain groups and classified. It is difficult to find a measure or a feature that can be taken as a basis in the classification of soils and can be applied in a general way because the soil was formed under the influence of various factors over thousands of years and gained a character according to these factors. However, the practitioner is more concerned with the general characteristics that affect fertility of the soil and whether the soil can be cultivated easily [1].

The determination of the soil types over the large agricultural lands by conventional methods (including laboratory analyses and soil sampling) is troublesome and expensive [2, 3]. However, the up-to-date information acquired from remote sensing

E. Acar (✉)
Batman University, Batman, Turkey
e-mail: emrullah.acar@batman.edu.tr

M. S. Ozerdem
Dicle University, Diyarbakır, Turkey
e-mail: sozerdem@dicle.edu.tr

© The Editor(s) (if applicable) and The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

M. Rysz et al. (eds.), *Synthetic Aperture Radar (SAR) Data Applications*, Springer Optimization and Its Applications 199, https://doi.org/10.1007/978-3-031-21225-3_9

219

(RS) technologies can perform that in a shorter time with less cost than conventional methods. Moreover, the easy update of this technology ensures a major advantage even in difficult geographical regions [4, 5].

RS data has gone out in the last decades as hopeful data resources in order to determine digital soil types and mapping in all scales. This data contains soil information such as spectral reflectance, and it has large spatial coverage. Thus, it allows mapping of inaccessible areas as well as it generates comprehensive and consistent data in both space and time. RS data proposes possibilities of supplementing or reducing conventional soil sampling in soil surveys [6]. On the basis of these advantages, many studies have been implemented by employing RS data in digital mapping of soil [7].

The microwave zone of the electromagnetic spectrum is employed for determining soil parameters since this zone is precision to changes in soil content. Among the all types of microwave sensors, Synthetic Aperture Radar (SAR) systems (Radarsat-1, Sentinel-1, Radarsat-2, Envisat-ASAR, ERS-1/2 etc.) have a major potential in determining soil parameters such as soil types [8]. Therefore, many studies, which are related with SAR data, have been researched to determine soil parameters. A SAR system includes single-polarized, dual-polarized, and quadrature-polarized (hh, hv, vh, and vv) levels. Here, the quadrature-polarized (polarimetric) SAR can produce data that is more detailed by multiple polarizations and penetrate the soil surfaces [9]. Therefore, polarimetric SAR can generate appropriate high-resolution images for soil monitoring, and it can be employed for agricultural monitoring, determining soil types, and digital soil mapping when used together with machine learning techniques.

A number of studies have been conducted in the literature employing RS data and machine learning models, and some of these studies are summarized in order. Forkuor et al. [10] have proposed digital mapping of soil properties utilizing RS data as well as multiple linear regression and machine learning models. Zeraatpisheh et al. [11] have recommended a research in order to predict soil properties (calcium carbonate equivalent, soil organic carbon, and clay content by using digital soil mapping and machine learning approaches in Borujen region, Iran). Hoa et al. [12] have proposed a work for mapping soil salinity intrusion in Vietnam by employing the Sentinel-1 SAR data and many machine learning models (Support Vector Regression, Multilayer Perceptron Neural Networks, Radial Basis Function Neural Networks, Gaussian Processes, and Random Forests). Moreover, different researches have been successfully implemented by employing multispectral optical data and hyper-spectral data, which are based on the correlation between several indices information derived from soil reflectance spectra and spectrum bands for mapping soil salinity [13–17]. Saadat et al. [18] have presented a numerical approach in order to classify landform based on Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) images and a 10-m resolution digital elevation model (DEM). Ehsani et al. [19] have proposed an approach for characterization of landscape elements through the combination of remotely sensed spectral data and morphometric parameters via Self-Organizing Map (SOM) and Artificial Neural Networks (ANNs). Vibhute et al. [20] have recommended

a research for classification of five various soil types by employing Hyperion hyperspectral satellite data and Support Vector Machine approach. Scull et al. [21] have proposed an approach in order to predict soil type in a desert landscape with the aid of classification tree analysis. Zhai et al. [22] have suggested a study to determine an accurate and efficient classification of soil texture with the aid of ANN and remote sensing data.

The main objective of this chapter is to determine two soil types (Clayey and Clayey+Loamy) automatically over the agricultural fields by employing various machine learning algorithms (K-NN, ELM, and NB) and four different Radarsat-2 SAR polarization coefficients (vv, vh, hv, and hh). The major contributions of the proposed study can be listed as follows:

1. Different machine-learning-based pipeline is proposed for automatic determination of soil types over the bare and vegetated fields.
2. In the proposed system, four different Radarsat-2 SAR polarization coefficients are employed as feature vectors, which have not been used before for determining soil types.
3. With this proposed system, the determination of the areas of plant and tree species growing in different soil types will become automatic. Thus, optimum use of water resources will be ensured by providing controlled irrigation even in drought.

The rest of this chapter is organized as follows. In the material and method section, the pilot area, employed dataset, and different machine learning techniques are explained. In the results and discussion section, the classification results obtained from machine learning approaches are given and the obtained results are evaluated. Finally, in the conclusion section, the recommended study is summed up.

2 Materials

2.1 Pilot Area

The pilot area contains two separate agricultural fields, which cover approximately 4 and 16 km² of frame in the Dicle University, Turkey (40° 04' - 40° 26' E, 37° 46' - 38° 04' N). The mean slope and height of the study area are 3.05% and 650 m; yearly maximum mean temperature is 8.2 and 34.5 °C in the winter and summer, respectively. Furthermore, yearly mean temperature and rainfall are 23.8 °C and 496 mm, respectively. The position of the pilot area is shown in Fig. 1. Here, the locations of local soil measurements are indicated by red dots [4].

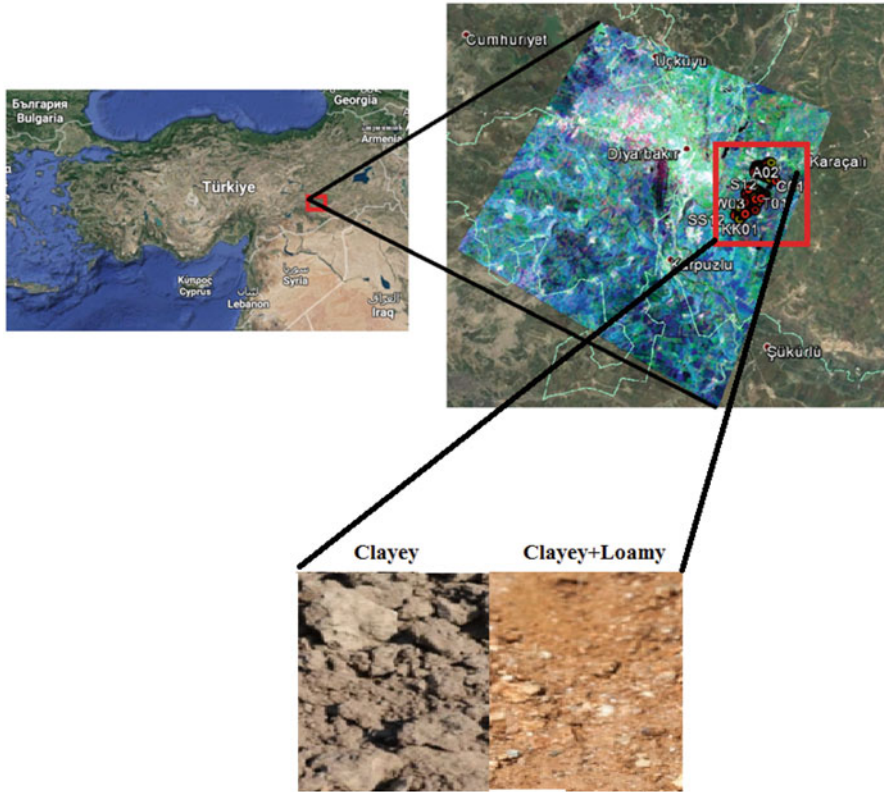


Fig. 1 The position of the pilot area. Here, the locations of local soil type samples are indicated in the red square

2.2 *Measurement of Local Soil Type Samples*

Local soil measurements were carried out over pilot area at the same time with the transition of Radarsat-2 in March 3, 2016. The pilot areas were divided into 0.1×0.1 -km grids, and local soil samples were received from each grid with 3–5 cm depth. Approximately, 156 local soil samples (32 clayey, 20 Loamy, 104 Clayey+Loamy) were gathered as given in Table 1, and the locations of soil samples were enrolled with the aid of GPS instrument. Each soil sample was then analyzed at the Agricultural Test and Analysis Laboratory of Dicle University (DUPTAM) in order to determine soil types.

Table 1 Number of local soil samples for each soil type

Soil type	Number of soil samples
Clayey	30
Clayey+Loamy	126

Table 2 The main features of the attained Radarsat-2 SAR image

Beam mode	Date	Transition	Time
Q13	03 March 2016	Descending	03:25:47

2.3 Acquisition of Radarsat-2 SAR Image Data

In this stage, a full polarimetric Radarsat-2 SAR image data, which belongs to 3 March 2016, was acquired. The obtained SAR data is in the single-look complex data format, and it keeps phase, resolution, and amplitude information. Moreover, it has 5.83 m spatial resolution and 30×30 km area. The main features of the obtained Radarsat-2 SAR data are given in Table 2.

2.4 Radarsat-2 SAR Image Preprocessing

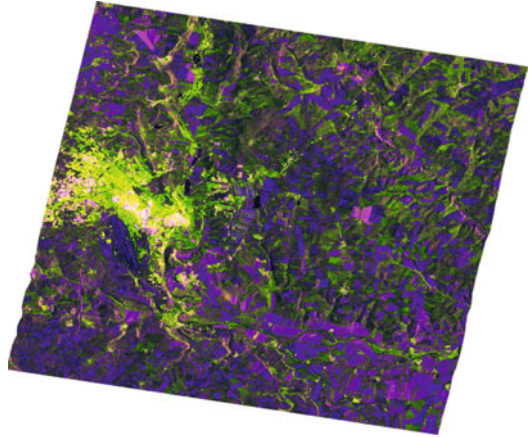
This phase was implemented in the following steps, respectively. First of all, Sentinel-1 Toolbox (S1TBX) was run for reading of Radarsat-2 SAR data, which was calibrated in order to rectify this data radiometrically. After that, Refined Lee Filter was employed to reduce the quantity of the speckle noise and blurring on the Radarsat-2 SAR data. Geometric distortions were fixed by utilizing SRTM-3 digital elevation model, and the filtered SAR data was geocoded for terrain correction. Subsequently, Universal Transverse Mercator with WGS84 was selected as the default map projection, and the positions of the local soil samples were recorded by GPS instrument. These coordinates were then turned into SHP file format with the aid of QGIS software and imported to the preprocessed Radarsat-2 SAR image [23]. Finally, the preprocessed Radarsat-2 SAR image was converted into RGB format as shown in Fig. 2.

3 Methods

3.1 Feature Extraction

In this stage, each soil type sample was represented by 3×3 pixel groups on the Radarsat-2 image in order to form the feature vectors. The SAR polarization coefficients for each soil type were then calculated by taking the average of the

Fig. 2 The preprocessed RGB Radarsat-2 SAR image



pixel groups. This process was repeated for four different polarizations, which are vertical–vertical (vv), vertical–horizontal (vh), horizontal–vertical (hv), and horizontal–horizontal (hh) of the Radarsat-2 SAR data. As a result, feature vectors with four polarimetric SAR coefficients (vv, vh, hv, and hh) obtained from four different polarizations were created for each soil type sample.

3.2 *K-Nearest Neighbor (K-NN) Algorithm*

K-NN is the one of the most effective algorithms in classification. It is a supervised machine learning method that is based on the distance calculation [24]. The K-Nearest Neighbor algorithm offers a simple result; however, it is used for convenient outcome in many classification conditions [25]. In this method, the following steps are applied, respectively:

- First, the parameter k , which is the number of nearest neighbors to a given point, is assigned. For instance, allow k to be equal to two and thus situated; classification will be revealed with respect to the two closest neighbors.
- The distance of the new data, which will be added to the sample dataset, with respect to the existing data, is computed one by one.
- The K-Nearest Neighbor of the related distances is taken into account. It is appointed to the class of k neighbors with respect to the feature values.
- The chosen class is then taken into account as the class of the observation values, which is hoped to be predicted. Namely, the new data is labeled [24].

In this algorithm, there are ten distance calculation methods that are Chebyshev, Cosine, City Block, Hamming, Euclidean, Correlation, Mahalanobis, Jaccard Minkowski, and Seucclidean [26]. In addition, the most commonly preferred distance

measure for calculating is the Euclidean Distance, and this parameter is given in Eq. 1 [27, 28].

$$d(x, y) = \sqrt{\sum_{j=1}^N (x_j - y_j)^2}. \quad (1)$$

Here, x and y symbolize data points to be classified and the learning points, respectively.

3.3 *Extreme Learning Machine (ELM) Algorithm*

ELM is one of the most vigorous machine learning techniques for feed-forward neural networks that contain a single hidden layer. It can be employed for regression and classification objectives. In ELM approach, setting of bias parameters and input weights is not essential in contrast to the other usual neural network applications. For this reason, these parameters can be chosen randomly, and the output weights can be determined analytically. This state authorizes ELM to become fast and easy in the data processing [29] and makes a great generalization performance for a single feed-forward neural network [30]. In contrast to gradient-based learning algorithms, ELM has much superiority, such as the potential of attaining the minimal training error, using single hidden layer and actuating by non-differentiable activation functions. Moreover, the observation numbers are more than neuron numbers in the ELM hidden layer [31]. The ELM output (y) can be figured out thanks to Eq. 2 [30].

$$y_t = \sum_{j=1} \alpha_{j,t} A \left(\sum_{i=1}^k w_{i,j} x_i + b_j \right). \quad (2)$$

In this place, $\alpha \{j, t\}$, b_j , m , n , x_i , $w_{i,j}$, and $A(\cdot)$ state the output weights, the biases of the neurons, the neuron numbers of the hidden layer, the neuron numbers of input layer, the input, the input weights, and the activation function, respectively [30].

3.4 *Naive Bayes (NB) Algorithm*

NB algorithm is one of the simplest machine learning algorithms. It is based on Bayes theorem, and it can be easily applied in many areas for classification [32].

$$P(h/X) = P(X/h) \times P(h)/P(X). \quad (3)$$

In Eq. 3, X and h represent the feature vector and the probability that a feature vector belongs to a class such as T , respectively. In addition, $P(h/X)$ expresses the subsequent probability.

$$f(x) = \begin{cases} \text{if } -x, x < 0 \\ \text{if } +x, x \geq 0 \end{cases}. \quad (4)$$

The algorithm of the Naive Bayes classifier is applied with the following steps:

- By employing Eq. 4, it is assumed that the class label of each x in the dataset is certain. X is a vector consisting of m features and is expressed as $X = (x_1, x_2, \dots, x_m)$.
- Next, suppose that you have n classes denoted by C_1, C_2, \dots, C_n . Naive Bayes algorithm tries to find the value with maximum $P(C_i/X)$ consecutive probability among all classes in order to find out whether a vector X belongs to class C_i . This is demonstrated using Bayes theorem and Eq. 5.

$$P(C_i/X) = (P(X/C_i) \times P(C_i))/P(X). \quad (5)$$

- Since the $P(X)$ value is the same for all types, only the expression $P(X/C_i) * P(C_i)$ should be the highest.
- The term $P(C_i)$ is defined as the ratio of the number of elements in the class C_i to the total number of elements.
- The term $P(X/C_i)$ is expressed by Eq. 6 when it is assumed that X is a feature vector consisting of n values.

$$P(H/C_i) = \prod_{k=1}^n P(X_k/C_i). \quad (6)$$

- As a result, the Naive Bayes classifier assigns the class C_i with the maximum value of $P(X/C_i) * P(C_i)$ as the class of vector X [33].

3.5 Performance Metrics

In this phase, some statistical metrics (sensitivity, specificity, precision, recall, F1-score, and accuracy) were employed in order to determine the overall performance of the proposed system. The mathematical expression of each metric is presented in Eqs. 7, 8, 9, 10 and 11, respectively.

$$Sensitivity = TP/(TP + FN) \quad (7)$$

$$\textit{Specificity} = TN / (FP + TN) \quad (8)$$

$$\textit{Precision} = TP / (TP + FP) \quad (9)$$

$$\textit{F1 - Score} = TP / (TP + 1/2 * (FP + FN)) \quad (10)$$

$$\textit{Accuracy} = (TP + TN) / (TP + TN + FP + FN). \quad (11)$$

Here, TP , TN , FP , and FN denote the number of true positives, true negatives, false positives, and false negatives, respectively [34].

4 Results and Discussion

In this section, the following operations were carried out respectively. First, the positions of the local soil type samples were imported to the preprocessed Radarsat-2 SAR image with the help of QGIS software. Four different Radarsat-2 SAR polarization coefficients (vv , vh , hv , hh) were then computed in order to form feature vectors for each sample. Finally, different machine learning algorithms (K-NN, ELM, and NB) were utilized to determine two soil types (Clayey and Clayey+Loamy) automatically on the agricultural fields.

4.1 Determination of Soil Types via Polarimetric SAR Coefficients and K-NN

At this stage, leave-one-out cross-validation approach together with different K-Nearest Neighbor numbers ($k = 1, 2, 3, 4, 5, 6, 7, 8, 9, 10$) was performed for defining optimal parameters of the k . In the leave-one-out cross-validation approach, $n-1$ of the n -dimensional dataset is chosen as the training set and the remaining one as the test set. This process is repeated n times to ensure that each data in the dataset is employed as a test set without being included in the training set. Then, the performance of the whole system is calculated by taking the average of the test performance values obtained n times. After that, the testing accuracy for all k numbers is presented in Fig. 3, and the performance metrics of the K-NN classifier are given in Table 3. Considering Fig. 3 and Table 3, it can be said that the maximum performance of the proposed system was computed as 98.4% sensitivity, 60.0% specificity, 91.2% precision, 94.6% F1-Score, and 91.1% accuracy, in the $k = 8$ scenario. Additionally, a receiver operating characteristic (ROC) curve, which is a graphical plot that exemplifies the diagnostic capability of a binary classifier system, is presented in Fig. 4.

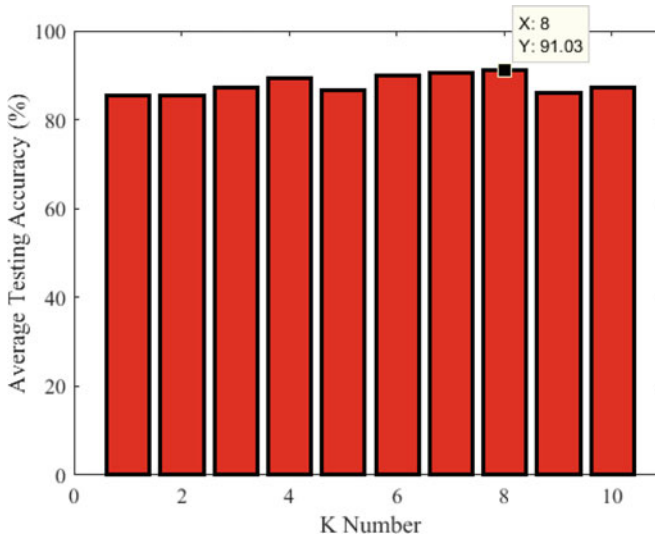


Fig. 3 Testing accuracy for all k numbers

Table 3 Performance metrics of K-NN classifier

Statistical metrics	Value(%)
Sensitivity	98.4
Specificity	60.0
Precision	91.2
F1-score	94.6
Accuracy	91.1

4.2 Determination of Soil Types via Polarimetric SAR Coefficients and ELM

In this part, leave-one-out cross-validation approach with 5 activation functions (1- sigmoid, 2- sinusoidal, 3- radial basis, 4- hard limit, and 5- triangular basis) and 5 different hidden neuron numbers (1,2,3,4,5) was implemented in order to determine ideal parameters of the proposed system. The accuracy for all activation functions (AF) and hidden neuron numbers (HNN) scenarios are illustrated in Fig. 5, and the performance metrics of the ELM classifier are presented in Table 4, respectively.

Evaluating Fig. 5 and Table 4, the maximum testing performance of the whole system was calculated as 100% sensitivity, 3.3% specificity, 81.3% precision, 89.7% F1-Score, and 82.0% accuracy in the HNN = 3 and AF = Sinusoidal structure. Moreover, the ROC curve of the proposed ELM classifier is presented in Fig. 6.

Fig. 4 The ROC curve of the proposed KNN system

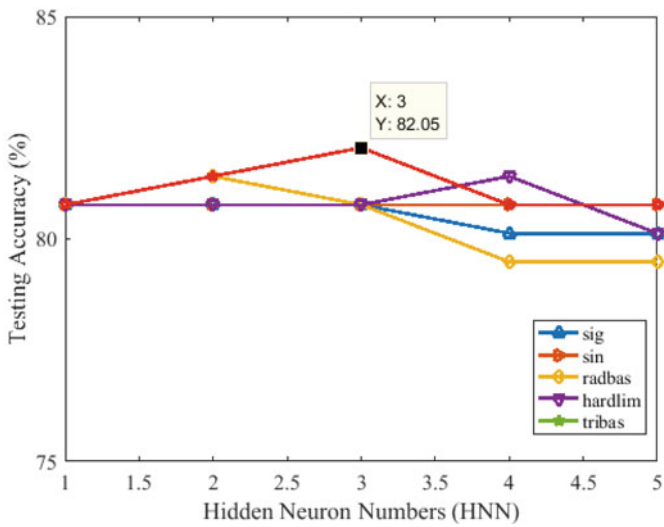
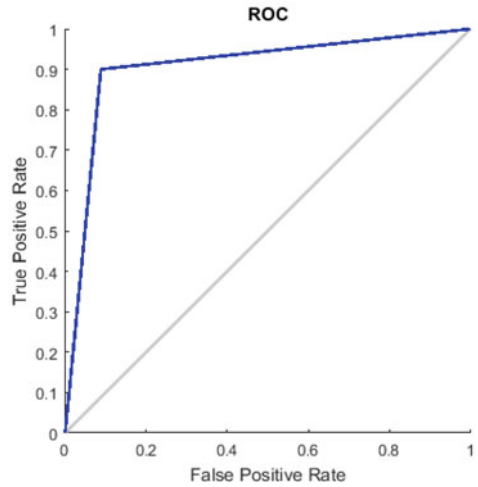


Fig. 5 Testing accuracy for all AF and NHN scenarios

Table 4 Performance metrics of ELM classifier

Statistical metrics	Value(%)
Sensitivity	100
Specificity	3.3
Precision	81.3
F1-score	89.7
Accuracy	82.0

Fig. 6 The ROC curve of the proposed ELM classifier

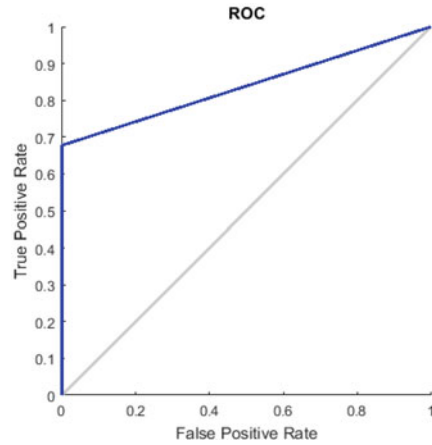


Table 5 Optimal parameters for NB classifier

Model parameters	Definition
Distribution name	Normal (Gaussian) distribution
Kernel smoothing density support	Unbounded
Prior probabilities	[0.81, 0.19]
Cost	[0 1, 1 0]
Score transform	None

Table 6 Performance metrics of NB classifier

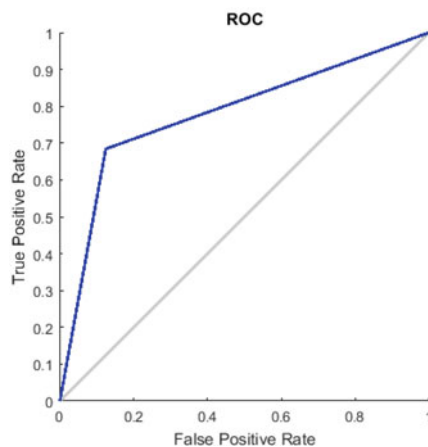
Statistical metrics	Value (%)
Sensitivity	95.2
Specificity	43.3
Precision	87.6
F1-score	91.2
Accuracy	85.2

4.3 Determination of Soil Types via Polarimetric SAR Coefficients and NB

In this phase, leave-one-out cross-validation was implemented in order to compute the overall performance of the proposed system with the optimal model parameters as tabulated in Table 5. Moreover, the performance metrics of the NB classifier together with ROC curve are presented in Table 6 and Fig. 7, respectively.

When Fig. 7 and Table 6 are analyzed, the maximum testing performance of the proposed system was computed as 95.2% sensitivity, 43.3% specificity, 87.6% precision, 91.2% F1-Score, and 85.2% accuracy.

Fig. 7 The ROC curve of the proposed NB classifier



5 Conclusion

The controlled use of water resources, which are decreasing due to drought and may be depleted in the future, is extremely important for the future of humanity and continuation of production. One of the important areas where water resources are used is agricultural land. Automatic determination of soil types over agricultural lands plays a key role in predicting which plant can grow on these lands and how often water can be used there. In this context, the recommended methodology offers a solution for the determination of soil types.

In the proposed work, the following steps were enforced, respectively. In the first step, the geographic locations of the soil type samples were transferred to the preprocessed SAR image by employing QGIS software. In the second step, four Radarsat-2 SAR image polarization coefficients (vv, vh, hv, and hh) were then calculated to constitute feature vectors for each sample. In the last step, several machine learning algorithms (K-NN, ELM, and NB) were utilized in order to determine different soil types (Clayey and Clayey+Loamy) automatically over the agricultural fields.

During the classification stage, various statistical metrics were employed to calculate the performance of the proposed system. In addition, the ROC curve of each classification model was drawn to get information about the performance value of the system visually. As a result, while the highest accuracy value was reached in the KNN classifier among the proposed systems, the lowest precision value with the highest sensitivity value was also reached in the ELM classifier.

In the continuation of this study, automatic detection of soil types is planned by utilizing different active and passive remote sensing image indices and deep learning techniques. Thanks to this system to be developed, optimum use of water resources in agricultural lands will be achieved.

Acknowledgments The authors would like to thank European Space Agency for Sentinel-1 Toolbox and Dicle University Science and Technology Application and Research Center (DUBTAM) for soil sample measurements. This work was supported by TUBITAK 1001 (No. 114E543) research project.

References

1. Çepel, N. (1960). Toprak türü ve bunun arazide el muayenesi ile tayini. *Journal of the Faculty of Forestry Istanbul University*, 10(2), 102–109.
2. Dobos, E., Montanarella, L., Nègre, T. and Micheli, E. (2001). A regional scale soil mapping approach using integrated AVHRR and DEM data. *International Journal of Applied Earth Observation and Geoinformation*, 3(1), 30–42.
3. Mulder, V. L., De Bruin, S., Schaeppman, M. E. and Mayr, T. R. (2011). The use of remote sensing in soil and terrain mapping—A review. *Geoderma*, 162(1–2), 1–19.
4. Özerdem, M. S., Acar, E. and Ekinci, R. (2017). Soil moisture estimation over vegetated agricultural areas: Tigris Basin, Turkey from Radarsat-2 data by polarimetric decomposition models and a generalized regression neural network. *Remote Sensing*, 9(4), 395.
5. Acar, H., Özerdem, M. S. and Acar, E. (2020). Soil Moisture Inversion Via Semiempirical and Machine Learning Methods With Full-Polarization Radarsat-2 and Polarimetric Target Decomposition Data: A Comparative Study. *IEEE Access*, 8, 197896–197907.
6. Malone, B. P., Jha, S. K., Minasny, B., and McBratney, A. B. (2016). Comparing regression-based digital soil mapping and multiple-point geostatistics for the spatial extrapolation of soil data. *Geoderma*, 262, 243–253.
7. Mulder, V. L., De Bruin, S., Schaeppman, M. E., and Mayr, T. R. (2011). The use of remote sensing in soil and terrain mapping—A review. *Geoderma*, 162(1–2), 1–19.
8. Moran, M. S., Peters-Lidard, C. D., Watts, J. M., and McElroy, S. (2004). Estimating soil moisture at the watershed scale with satellite-based radar and land surface models. *Canadian Journal of Remote Sensing*, 30(5), 805–826.
9. Oliver, C., and Quegan, S. (2004). Understanding synthetic aperture radar images. SciTech Publishing.
10. Forkuor, G., Hounkpatin, O. K., Welp, G., and Thiel, M. (2017). High resolution mapping of soil properties using remote sensing variables in south-western Burkina Faso: a comparison of machine learning and multiple linear regression models. *PloS one*, 12(1), e0170478.
11. Zeraatpisheh, M., Ayoubi, S., Jafari, A., Tajik, S., and Finke, P. (2019). Digital mapping of soil properties using multiple machine learning in a semi-arid region, central Iran. *Geoderma*, 338, 445–452.
12. Hoa, P. V., Giang, N. V., Binh, N. A., Hai, L. V. H., Pham, T. D., Hasanlou, M., and Tien Bui, D. (2019). Soil salinity mapping using SAR sentinel-1 data and advanced machine learning algorithms: A case study at Ben Tre Province of the Mekong River Delta (Vietnam). *Remote Sensing*, 11(2), 128.
13. Allbed, A., Kumar, L., and Sinha, P. (2014). Mapping and modelling spatial variation in soil salinity in the Al Hasa Oasis based on remote sensing indicators and regression techniques. *Remote Sensing*, 6(2), 1137–1157.
14. Nicolas, H., and Walter, C. (2006). Detecting salinity hazards within a semiarid context by means of combining soil and remote-sensing data. *Geoderma*, 134(1–2), 217–230.
15. Sidike, A., Zhao, S., and Wen, Y. (2014). Estimating soil salinity in Pingluo County of China using QuickBird data and soil reflectance spectra. *International Journal of Applied Earth Observation and Geoinformation*, 26, 156–175.

16. Mashimbye, Z. E., Cho, M. A., Nell, J. P., De Clercq, W. P., Van Niekerk, A., and Turner, D. P. (2012). Model-based integrated methods for quantitative estimation of soil salinity from hyperspectral remote sensing data: A case study of selected South African soils. *Pedosphere*, 22(5), 640–649.
17. El Harti, A., Lhissou, R., Chokmani, K., Ouzemou, J. E., Hassouna, M., Bachaoui, E. M., and El Ghmari, A. (2016). Spatiotemporal monitoring of soil salinization in irrigated Tadla Plain (Morocco) using satellite spectral indices. *International Journal of Applied Earth Observation and Geoinformation*, 50, 64–73.
18. Saadat, H., Bonnell, R., Sharifi, F., Mehuys, G., Namdar, M., and Ale-Ebrahim, S. (2008). Landform classification from a digital elevation model and satellite imagery. *Geomorphology*, 100(3–4), 453–464.
19. Ehsani, A. H., and Quiel, F. (2009). A semi-automatic method for analysis of landscape elements using Shuttle Radar Topography Mission and Landsat ETM+ data. *Computers and Geosciences*, 35(2), 373–389.
20. Vibhute, A. D., Kale, K. V., Dhumal, R. K., and Mehrotra, S. C. (2015, December). Soil type classification and mapping using hyperspectral remote sensing data. In 2015 International Conference on Man and Machine Interfacing (MAMI) (pp. 1–4). IEEE.
21. Scull, P., Franklin, J., and Chadwick, O. A. (2005). The application of classification tree analysis to soil type prediction in a desert landscape. *Ecological modelling*, 181(1), 1–15.
22. Zhai, Y., Thomasson, J. A., Boggess III, J. E., and Sui, R. (2006). Soil texture classification with artificial neural networks operating on remote sensing data. *Computers and Electronics in Agriculture*, 54(2), 53–68.
23. Acar, E., and Özerdem, M. S. (2020). On a yearly basis prediction of soil water content utilizing SAR data: a machine learning and feature selection approach. *Turkish Journal of Electrical Engineering and Computer Sciences*, 28(4), 2316–2330.
24. Bozkurt, M. R., Uçar, M. K., Bozkurt, F., and Bilgin, C. (2019). In obstructive sleep apnea patients, automatic determination of respiratory arrests by photoplethysmography signal and heart rate variability. *Australasian physical and engineering sciences in medicine*, 42(4), 959–979.
25. S. Kundu and S. Roy, “K-Nearest Neighbour (KNN) Approach using SAT Based Technique for Rectilinear Steiner Tree Construction,” 2017, <https://doi.org/10.1109/ISED.2017.8303915>.
26. Öreñç, S. (2019). Kronik obstrüktif akciğer hastalığı teşhisi için makine öğrenmesi temelli yeni bir yöntem geliştirilmesi (Master’s thesis, Sakarya Üniversitesi).
27. Brown, J. F., Wardlow, B. D., Tadesse, T., Hayes, M. J., and Reed, B. C. (2008). The Vegetation Drought Response Index (VegDRI): A new integrated approach for monitoring drought stress in vegetation. *GIScience and Remote Sensing*, 45(1), 16–46.
28. Emrullah, A. (2016). Extraction of texture features from local iris areas by GLCM and Iris recognition system based on KNN. *European Journal of Technique (EJT)*, 6(1), 44–52.
29. Ertuğrul, Ö. F., and Tağluk, M. E. (2017). A fast feature selection approach based on extreme learning machine and coefficient of variation. *Turkish Journal of Electrical Engineering and Computer Sciences*, 25(4), 3409–3420.
30. Huang, G. B., Zhu, Q. Y., and Siew, C. K. (2006). Extreme learning machine: theory and applications. *Neurocomputing*, 70(1–3), 489–501.
31. Li, M., and Wang, D. (2017). Insights into randomized algorithms for neural networks: Practical issues and common pitfalls. *Information Sciences*, 382, 170–178.
32. Pawlak, Z. (2003). A rough set view on Bayes’ theorem. *International Journal of Intelligent Systems*, 18(5), 487–498.
33. McCallum, A., and Nigam, K. (1998, July). A comparison of event models for naive bayes text classification. In *AAAI-98 Workshop on Learning for Text Categorization (Vol. 752, No. 1, pp. 41–48)*.
34. Hlaváč, V. (2016). Classifier performance evaluation. Czech Technical University.

Ocean and Coastal Area Information Retrieval Using SAR Polarimetry



Andrea Buono, Emanuele Ferrentino, Yu Li, and Carina Regina de Macedo

List of Abbreviations

ALOS	Advanced Land Observing Satellite
AOI	Angle of Incidence
ASAR	Advanced SAR
ASI	Italian Space Agency
BSA	Back Scattering Alignment
CFAR	Constant False Alarm Rate
CNN	Convolutional Neural Network
CONAE	National Commission of Space Activities
CP	Compact Polarimetric
CPD	Co-polarized Phase Difference
CSA	Canadian Space Agency
CSG	CSK Second Generation

A. Buono (✉)
University of Naples Parthenope, Naples, Italy
e-mail: andrea.buono@uniparthenope.it

E. Ferrentino
National Institute of Geophysics and Volcanology, Rome, Italy
e-mail: emanuele.ferrentino@ingv.it

Y. Li
Beijing University of Technology, Beijing, China
e-mail: yuli@bjut.edu.cn

C. R. de Macedo
University of the Littoral Côte d'Opale, UMR 8187 - LOG - Laboratoire d'Océanologie et de
Géosciences, Wimereux, France

CNRS, UMR 8187, Lille, France

University of Lille, Lille, France
e-mail: carina.macedo@fc.up.pt

CSK	Cosmo Sky-Med
CTLR	Circular Transmit Linear Receive
CVCNN	Complex-Valued CNN
dB	Decibel
DBN	Deep Belief Network
DLR	German Aerospace Center
DOP	Degree of Polarization
DP	Dual Polarimetric
ENVISAT	ENVIRONMENTAL SATellite
ESA	European Space Agency
FP	Full Polarimetric
HH	Horizontal transmit horizontal receive
HV	Horizontal transmit vertical receive
ISRO	Indian Space Research Organization
JAXA	Japanese Aerospace eXPLORATION Agency
NESZ	Noise Equivalent Sigma Zero
NPH	Normalized Pedestal Height
NRCS	Normalized Radar Cross Section
PALSAR	Phased Array L-band SAR
PD	Polarization Difference
PolSAR	Polarimetric SAR
PR	Polarization Ratio
PSD	Power Spectral Density
PTSM	Polarimetric TSM
RCM	Radarsat Constellation Mission
RVCNN	Real-Valued CNN
SAE	Sparse Auto-Encoder
SAOCOM	Microwave Observing Argentinian Satellite
SAR	Synthetic Aperture Radar
SGD	Stochastic Gradient Descent
SIR	Spaceborne Imaging Radar
SNR	Signal-to-Noise Ratio
SPM	Small Perturbation Method
SVM	Support Vector Machine
TSM	Two-Scale Model
UAVSAR	Unmanned Aerial Vehicle SAR
VV	Vertical transmit vertical receive

1 Radar Polarimetry

This section presents a general overview of the theoretical background needed to model the sea surface polarimetric observables. We focus on the general formulation of the polarimetric scattering process that describes the electromagnetic interaction between the wave transmitted by a microwave sensor and a target to be observed on the Earth's surface.

In this section, we first introduce the mathematical formalism used to describe polarimetric observables in Sect. 1.1. Then, in Sect. 1.2, an overview of the operational and planned polSAR systems and the main characteristics of the polarimetric imaging modes is provided. As a reference, information about multi-polarization

models to describe sea surface scattering is given in Sect. 1.3. Finally, in Sect. 1.4, meaningful experimental showcases are presented to demonstrate some of the most representative polarimetric properties of sea surface scattering.

1.1 Polarimetric Scattering Descriptors

In this subsection, a general overview of the polarimetric scattering descriptors of the electromagnetic interaction between the transmitted wave and the observed surface is presented.

Considering a monostatic radar sensor, a microwave pulse transmitted by a radiation source interacts with the observed surface, and it is scattered back to be received by the original transmitting antenna. The received electromagnetic pulse brings information about the scene that can be accessed taking into account the transmitted and received waves. According to the Jones formalism, in the far zone of the scatterer, the transformation of the incident wave into the scattered one is given in Eq. (1), where the transformation is ruled by the scattering matrix \mathbf{S} [34]:

$$\mathbf{E}^s = \frac{e^{-jkr}}{r} \mathbf{S} \mathbf{E}^i, \quad (1)$$

where the ratio e^{-jkr}/r is the spherical wave factor with j being the imaginary unit, k the electromagnetic wave number, and r the distance between the SAR antenna and the center of the scene. \mathbf{E}^s and \mathbf{E}^i are the complex Jones vectors describing, respectively, the scattered and incident waves.

Considering the monostatic backscattering case, i.e., the transmitting and receiving antennas are in the same location, and under the backscatter alignment convention (BSA), the scattering matrix \mathbf{S} is a 2×2 complex-valued matrix also known as Sinclair matrix and, adopting the linear horizontal (h)–vertical (v) polarization basis and considering the propagation through a reciprocal medium, can be given by

$$\mathbf{S} = \begin{pmatrix} S_{hh} & S_{hv} \\ S_{hv} & S_{vv} \end{pmatrix} = \begin{pmatrix} |S_{hh}| e^{j\varphi_{hh}} & |S_{hv}| e^{j\varphi_{hv}} \\ |S_{hv}| e^{j\varphi_{hv}} & |S_{vv}| e^{j\varphi_{vv}} \end{pmatrix} = e^{j\varphi_{hh}} \begin{pmatrix} |S_{hh}| & |S_{hv}| e^{j\varphi_x} \\ |S_{hv}| e^{j\varphi_x} & |S_{vv}| e^{j\varphi_c} \end{pmatrix}, \quad (2)$$

where S_{pq} , with $p, q \in \{h, v\}$, is the complex scattering amplitude, while $S_{vh} = S_{hv}$ results from the reciprocity assumption. The diagonal and off-diagonal terms of the scattering matrix \mathbf{S} call for, respectively, the same (termed co-polar) and orthogonal (cross-polar) polarization for both incident and scattered waves. $|S_{pq}|$ are the modulus of the scattering amplitude, and φ_c and φ_x are the relative phase between, respectively, the co-polarized and the cross-polarized channels:

$$\varphi_c = \varphi_{hh} - \varphi_{vv} \quad , \quad \varphi_x = \varphi_{hv} - \varphi_{hh}. \quad (3)$$

The phase differences between co-polarized and cross-polarized channels are often indicated as CPD and XPD, respectively. The Jones formalism shown in Eq. (1) represents a first-order coherent scattering model (i.e., a completely polarized and deterministic scattering) and does not allow describing the phenomena of depolarization that may arise from the scattering of random and distributed scenes. The second-order descriptors are used to deal with the polarimetric scattering from the distributed and depolarizing scenes based on both an incoherent and coherent approaches. The latter is based on the coherency \mathbf{T} and covariance \mathbf{C} matrix, while the former on the Stokes formalism where the 4×4 Muller matrix \mathbf{M} connects the partially polarized scattered wave (\mathbf{g}^s) to the fully polarized incident wave (\mathbf{g}^i) [34]:

$$\mathbf{g}^s = (kr)^{-2} \langle \mathbf{M} \rangle \mathbf{g}^i, \quad (4)$$

where $\langle \cdot \rangle$ denotes ensemble average, the matrix \mathbf{M} is real and never symmetric, its elements are ensemble averages of combinations of scattering amplitudes, and more details can be found in [33, 34]. The Stokes vector \mathbf{g} represents the polarization properties of an electromagnetic plane wave based on non-coherent power measurements. The Stokes vector, different from the Jones formalism, can describe partially polarized waves. In Eq. (4), since the incidence wave is deterministic, no ensemble average is made in \mathbf{g}^i . According to the Stokes formalism, the degree of polarization (DOP) of a partially polarized target can be evaluated:

$$\text{DOP} = \frac{\sqrt{g_1^2 + g_2^2 + g_3^2}}{g_0}. \quad (5)$$

The DOP is a basis-invariant parameter ranging between 0 and 1 that measures the amount of polarized scattering/component of the target/electromagnetic wave.

One of the most powerful tools that characterize polSAR is the polarization synthesis, i.e., the synthesis of the power given the polarization properties of any couple of transmitting/receiving antenna once a complete polarimetric measure has been performed in an orthogonal basis. The normalized radar cross section (NRCS) can be evaluated based on the Kennaugh matrix, \mathbf{K} , for any possible combination of transmitting/receiving antenna polarizations, assuming that the antennas match in both load and polarization:

$$\sigma^0 = \frac{4\pi}{2k^2} \langle \mathbf{g}^r \rangle^T \langle \mathbf{K} \rangle \mathbf{g}^t, \quad (6)$$

where the superscript T is the transpose operator, and \mathbf{g}^t and \mathbf{g}^r are, respectively, the transmitted and received polarizations described using the Stokes formalism. Considering the BSA convention, \mathbf{K} is given by Guissard [34]

$$\mathbf{K} = \text{diag}(1, 1, 1, -1)\mathbf{M}. \quad (7)$$

Once polarization synthesis is performed, by varying the transmitting/receiving polarization couple, the NRCS can be plotted versus both the tilting Φ and the ellipticity χ angles, i.e., for any polarization. Hence, the co-polarized (transmitting and receiving polarizations call for the same Stokes vector) and cross-polarized (transmitting and receiving polarizations call for the orthogonal Stokes vectors) signatures of the target are obtained. Once the former is normalized with respect to the total backscattered power, namely the SPAN, the normalized pedestal height (NPH) can be derived, which is the lowest NRCS in the normalized co-polarized signature, while varying the tilting and ellipticity angles (i.e., the polarization). The NPH describes the amount of unpolarized energy with respect to the total received power.

The Mueller matrix is in a one-to-one mapping with the coherency matrix \mathbf{T} [16, 48]. The latter completely describes the polarimetric scattering properties of a generic distributed and depolarizing scene with the advantage of, unlike the Mueller matrix, being Hermitian and semi-definite positive matrices. The latter properties allow the decomposition of the \mathbf{T} and \mathbf{C} matrices in elementary scattering mechanisms providing a physical interpretation of the scene scattering processes. The coherency/covariance matrices are both ensemble averages of combinations of scattering amplitudes. Considering the backscattering case and assuming reciprocity, the coherency matrix \mathbf{T} and the covariance matrix \mathbf{C} can be expressed, respectively, as [16]

$$\begin{aligned} \mathbf{T} &= \langle \mathbf{k} \mathbf{k}^\dagger \rangle \\ &= \frac{1}{2} \begin{pmatrix} \langle |S_{hh} + S_{vv}|^2 \rangle & \langle (S_{hh} + S_{vv})(S_{hh} - S_{vv})^* \rangle & 2 \langle (S_{hh} + S_{vv}) S_{hv}^* \rangle \\ \langle (S_{hh} - S_{vv})(S_{hh} + S_{vv})^* \rangle & \langle |S_{hh} - S_{vv}|^2 \rangle & 2 \langle (S_{hh} - S_{vv}) S_{hv}^* \rangle \\ 2 \langle S_{hv}(S_{hh} + S_{vv})^* \rangle & 2 \langle S_{hv}(S_{hh} - S_{vv})^* \rangle & 4 \langle |S_{hv}|^2 \rangle \end{pmatrix} \end{aligned} \quad (8)$$

and

$$\mathbf{C} = \langle \mathbf{k} \mathbf{k}^\dagger \rangle = \begin{pmatrix} \langle |S_{hh}|^2 \rangle & \sqrt{2} \langle S_{hh} S_{hv}^* \rangle & \langle S_{hh} S_{vv}^* \rangle \\ \sqrt{2} \langle S_{hv} S_{hh}^* \rangle & 2 \langle |S_{hv}|^2 \rangle & \sqrt{2} \langle S_{hv} S_{vv}^* \rangle \\ \langle S_{vv} S_{hh}^* \rangle & \sqrt{2} \langle S_{vv} S_{hv}^* \rangle & \langle |S_{vv}|^2 \rangle \end{pmatrix}, \quad (9)$$

where \mathbf{k} is the target scattering vector projected into the Pauli (lexicographic) basis [16]. The \mathbf{T} matrix can be uniquely diagonalized as follows:

$$\mathbf{T} = \mathbf{U} \mathbf{D} \mathbf{U}^{-1} = \sum_{i=1}^3 \lambda_i \mathbf{u}_i \cdot \mathbf{u}_i^\dagger = \mathbf{T}_1 + \mathbf{T}_2 + \mathbf{T}_3. \quad (10)$$

Equation (10) describes the decomposition of the \mathbf{T} matrix into the sum of three elementary scattering mechanisms (i.e., surface scattering from a plane flat structure, double-bounce scattering from a dihedral structure, and volume scattering from a randomly oriented cloud of dipoles) described by the eigenvectors \mathbf{u}_i in

which the power contribution for each mechanism is given by the eigenvalues λ_i [16, 48]. Since \mathbf{T} matrix satisfies Hermitian symmetry and is semi-definite positive, it is characterized by real non-negative eigenvalues that satisfy the following relationship:

$$\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq 0. \quad (11)$$

\mathbf{T} and \mathbf{C} matrices share the same eigenvalues, and their eigenvectors are linked to each other by a con-similarity transformation [16, 48]:

$$\mathbf{C} = \mathbf{F}^{-1} \mathbf{T} \mathbf{F} \quad , \quad \mathbf{F} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 0 & 1 \\ 1 & 0 & -1 \\ 0 & \sqrt{2} & 0 \end{pmatrix}. \quad (12)$$

The eigen-decomposition of the coherency matrix results in meaningful basis-invariant synthetic parameters that are strictly related to the scattering properties of the observed target. The polarimetric entropy, H , is given by

$$H = - \sum_{i=1}^3 p_i \log_3(p_i) \quad , \quad p_i = \frac{\lambda_i}{\sum_{j=1}^3 \lambda_j}, \quad (13)$$

while the mean scattering angle $\bar{\alpha}$ is defined by

$$\bar{\alpha} = \sum_{i=1}^3 \frac{\lambda_i}{\sum_{j=1}^3 \lambda_j} \cos^{-1}(|\mathbf{u}_i(1)|). \quad (14)$$

The entropy H is a basis-invariant measure, bounded between 0 and 1, of the randomness of polarimetric scattering mechanisms that characterize the observed target. $H = 0$ means deterministic scattering, while $H = 1$ means completely unpolarized scattering. The mean scattering angle $\bar{\alpha}$, ranging between 0 and 90°, represents the average scattering mechanism of the target. $\bar{\alpha} = 0^\circ$, 45° , and 90° stand for surface, volume, and double-bounce scattering, respectively.

The eigenvalues of the coherency/covariance matrices can also be used to express the NPH:

$$\text{NPH} = \frac{\lambda_3}{\lambda_1}. \quad (15)$$

1.2 PolSAR Imaging Modes

In this subsection, a brief overview of polarimetric SAR missions is presented.

In the last decades, SAR satellites were launched acquiring information from the Earth surface at different microwave wavelengths and exploiting as well different polarimetric and imaging modes. Since the polarimetric SAR provides reliable, detailed, and valuable information on the physical properties and processes that rule the observed scene, spaceborne polarimetric missions were launched recently or are planned to be launched in the next years (e.g., the BIOMASS mission that consists of a P-band SAR that will focus on the global distribution of forest biomass). An overview of the main polSAR spaceborne missions is provided in Table 1.

According to the polarimetric information content, the SAR can be classified as full polarimetric (FP), dual polarimetric (DP), and compact polarimetric (CP). The FP SAR transmits and receives radiation on an orthogonal linear polarization basis, providing the complete scattering matrix information on the observed scene. Operational SAR missions are continuously acquiring information at different frequencies, e.g., the C-band CSA (Canadian Space Agency), Radarsat-2, the L-band JAXA (Japanese Aerospace Exploration Agency), Alos (Advanced Land Observing Satellite), PalSAR-2 (Phased Array Type L-band SAR), and the X-band ASI (Italian Space Agency), CSG (Constellation of Small Satellites for the Mediterranean Basin Observation). The planned biomass mission will be equipped with a FP SAR. An important limitation related to the FP SARs is related to its limited area coverage (a swath smaller than 70 km) that impacts directly on its use for operational monitoring services.

When compared to the FP SAR systems, the DP ones provide less polarimetric information, i.e., they transmit a single linear polarization and receive usually both amplitude and phase (coherently) in the corresponding orthogonal basis, providing a single row/column of the scattering matrix, see Eq. (2). However, as advantage, the DP SAR missions offer doubled area coverage when compared with the FP systems. The operational DP SARs we can mention are the X-band ASI CSK, calling for an incoherent DP imaging mode (i.e., alternating bursts are transmitted/received, so no phase link is measured between the two polarimetric channels), the X-band DLR (German Aerospace Center), TerraSAR-X, and C-band ESA (European Space Agency), Sentinel-1.

A single circular or slant linear polarization is transmitted by the CP SAR architectures and is then received according to the linear orthogonal basis. The CP SAR system only measures the wave coherency matrix associated to the received electromagnetic wave, limiting the amount of scattering information that can be extracted when compared to FP missions. Even though the information received by the CP system is biased and/or dependent on the considered CP mode, CP SAR data have been successfully exploited for coastal areas observation [77]. CP SAR missions are a trade-off solution between the area coverage and the amount of polarimetric information. CP SAR satellites have been launched recently, for example, the L-band CONAE (National Space Activities Commission) SAOCOM

Table 1 Operational and planned spaceborne polarimetric missions. The information listed refers to the finest spatial resolution polarimetric mode

Mission/SAR	Agency	Launch	Mode	Frequency band	NESZ (dB)	Spatial resolution (slant rg. x az.) (m)	Area coverage (rg. x az.) (km)
Radarsat-2	CSA	2007	DP, FP	C	-36.5 ± 3	5.2×7.6	25×25
CSK	ASI	2007	DP	X	≤ -22	15×15	30×30
TerraSAR-X	DLR	2007	DP, FP	X	-19 --- -26	1.2×2.2	10×5
RISAT-1	ISRO	2012	DP, CP, FP	C	≤ -19	4×9	25×25
Alos PalSAR-2	JAXA	2014	DP, CP, FP	L	-28	10×10	30×30
Sentinel-1	ESA	2014	DP	C	≤ -22	5×5	80×80
SAOCOM	CONAE	2018	DP, CP, FP	L	≤ -25	10×10	20×20
RCM	CSA	2019	DP, CP, FP	C	≤ -24	9×9	20×20
CSG	ASI	2019	DP, FP	X	≤ -19	3×3	40×15
BIOMASS	ESA	2023	DP, CP, FP	0.4, P	≤ -24	$\leq 50 \times 50$	2×60

(Argentine Microwaves Observation Satellite), and the C-band CSA RCM (Radarsat Constellation Mission). Another operational mission that has an CP SAR is the C-band ISRO (Indian Space Research Organization), RISAT-1 (Radar Imaging SATellite).

To acquire data from the Earth's surface, different operational modes can be used by the SAR systems. The most fundamental and simplest mode is the stripmap where the radar antenna is fixed to one swath and draws a strip on the ground during the time in which the platform moves. The illuminated area in the Earth's surface is limited in the range size and theoretically unlimited in the along-track (azimuth) direction [22]. Taking the TerraSAR-X system as example, when working on stripmap mode, its standard scene size is 30×50 km (range \times azimuth) and its spatial resolution is 1.2×3.3 m (slant range \times azimuth) for single polarization and, respectively, 15×50 km and 1.2×6.6 m for dual polarization.

In case a wider swath is needed, the antenna can be operated on scanSAR mode where, to illuminate different range subswaths, the antenna beam elevation is periodically switched. In this case, the azimuth resolution is degraded compared to stripmap mode [22]. The standard scene size for the TerraSAR-X system operating on scanSAR mode is 100×150 km, and the spatial resolution is 1.2×18.5 m. An improvement on the azimuth resolution can be achieved using the spotlight mode at the expense of azimuth coverage. The radar antenna beam operating on spotlight mode is steered during the acquisition time from forward to backward, pointing always in the same area on the ground [22]. The spotlight mode of the TerraSAR-X system acquires a standard scene size of 10×10 km with a spatial resolution of 1.2×1.7 m for single polarization and 1.2×3.4 for dual polarization.

1.3 Sea Surface Polarimetric Scattering

This subsection deals with multi-polarization models to describe/predict the backscattering from the sea surface at microwaves.

Over decades, the problem of scattering of electromagnetic waves from natural rough surfaces has been investigated. The rough surface scattering problem plays an important role for the radar remote sensing and its application on extracting information on the observed scene, being of paramount importance for a broad range of operational applications such as sea wind retrieval, soil moisture estimation, sea ice, and oil slick observation [21, 36, 38]. To obtain closed-form solutions is not a trivial task, and therefore, approximation approaches to deal with limiting scattering cases were proposed, i.e., the high-frequency and low-frequency approaches. The analytical high-frequency approach based on the Kirchhoff-tangent approximation is valid for very rough slopes, and it has a good performance in modeling quasi-specular scattering, while it lacks polarization sensitivity. Other very commonly used general analytical approach is the small perturbation model (SPM) based on the low-frequency approximation for small vertical variations. The SPM yields for proper polarization sensitivity (considering the regime where the model is valid);

however, it lacks in accounting specular scattering, multiple scattering effects, and long-scale features in the surface spectrum [84]. Furthermore, the Kirchhoff-tangent approximation and the SPM cannot well model natural rough surfaces since these surfaces have different scales. In this context, in order to overcome the problem of the natural sea surface being a very complicated composite surface, the two-scale approximation assumes that the small roughness rides on top of large fluctuations. The latter model has a good compromise between interpretation, practical implementation, and accuracy issues [24, 35].

First to introduce the three scattering models will be presented in this subsection, the Bragg scattering will be described as well as some general properties of the ocean backscattering. Considering a range of angles of incidence between 20 and 60°, under low-to-moderate wind conditions and the absence of long waves (validity range of the Bragg scattering regime), the SAR sea surface backscattering is primarily due to the Bragg scattering. The incident radiation is backscattered by the wind-generated waves (i.e., capillary or short-wave length waves) of the sea surface, following the relation: $\lambda_B = \lambda_r/2 \sin(\theta)$, which says that sea surface roughness scale, λ_B , is comparable with the radar wavelength λ_r . Considering all frequencies, the ocean backscattering decreases with increasing incidence angle while increasing when wind speed increases. The VV-polarized return is higher than the HH one, while the cross-polarized (HV or VH polarization) NRCS is much lower than the co-polarized ones, often being below the noise floor of the SAR sensor [37].

In [36], a scattering model was proposed, which is an extension of the conventional Bragg scattering model, since it includes a roughness-induced rotation symmetric disturbance. According to [36], the great advantages of the model are the ability to describe processes that reduce the degree of polarization of the electromagnetic wave (i.e., depolarization effects) and the capability to describe cross-polarized backscattering. In this framework, it is the possible extension of the Bragg scattering theory to a range of natural surfaces.

Under intermediate incidence angles and low-to-moderate sea-state conditions, the X-Bragg coherency matrix \mathbf{T}_X can be predicted introducing a roughness disturbance by rotating the Bragg coherency matrix about an angle β in the plane perpendicular to the scattering plane [36]:

$$\mathbf{T}_X = \begin{pmatrix} C_1 & C_2 \text{sinc}(2\beta_1) & 0 \\ C_2^* \text{sinc}(2\beta_1) & C_3 (1 + \text{sinc}(4\beta_1)) & 0 \\ 0 & 0 & C_3 (1 - \text{sinc}(4\beta_1)) \end{pmatrix}. \quad (16)$$

The angle β is an uniformly distributed random variable in the range 0–90°, and its distribution width, β_1 , corresponds to the amount of roughness of the sea surface [36].

$$p(\beta) = \begin{cases} \frac{1}{2\beta_1} & |\beta| \leq \beta_1 \\ 0 & \text{elsewhere} \end{cases}. \quad (17)$$

C_1 , C_2 , and C_3 are combinations of the complex Bragg scattering coefficients:

$$\begin{cases} C_1 = |R_h + R_v|^2 \\ C_2 = (R_h + R_v)(R_h^* - R_v^*) \\ C_3 = \frac{1}{2}|R_h - R_v|^2 \end{cases} \quad (18)$$

R_h and R_v are, respectively, the Bragg scattering coefficients perpendicular (V) and parallel (H) to the incidence plane and depend on the local incidence angle, θ , and the relative electric permittivity, ϵ . More details can be found in [36].

Please note that the X-Bragg scattering model does not implement a high depolarizing condition. Thus, this model is not valid for dealing with high depolarizing targets [10].

The electromagnetic scattering model named polarimetric two-scale model (PTSM) [21, 38], as the original TSM, accounts for depolarization effects. However, the former different from the latter one has the advantage to provide closed-form expressions of the elements of the covariance matrix holding large-scale surface slopes. Compared with the X-Bragg model, the PTSM brings improvement when it removes the assumption of a uniform incidence plane rotation β and no variation in the incidence angle. The PTSM was first developed to retrieval soil moisture, and it was expanded to deal with the sea surface scattering purposes in [73, 74].

Considering the PTSM, the ocean can be modeled as being composed of large-scale roughness with slightly roughened, tilted facets whose slope is the same as a smoothed surface at the center of the roughened facet. The small-scale roughness $\delta(x, y)$ is considered as a zero-mean stochastic process with height standard deviation small when compared to the electromagnetic wavelength. Considering $\delta(x, y)$ as a band-limited process, the power spectral density is [32, 38]

$$W(k) = \frac{S_0}{k^{(2+2H_t)}} = s^2 \frac{S_{0n}}{k^{(2+2H_t)}} = s^2 W_n(k), \quad (19)$$

where $k = \sqrt{k_x^2 + k_y^2}$, and k_x , k_y are the Fourier mates of x (azimuth) and y (range), respectively. Via the dimensional facet-size-dependent constant S_{0n} , S_0 is proportional to the roughness variance s^2 [32]. W_n is the normalized power spectral density, and $0 < H_t < 1$ is the Hurst coefficient related to the fractal dimension D by $D = 3 - H_t$. More details about values of D and H_t for the sea surface can be found in [49, 82].

If we consider a sensor illuminating an area at a global incidence angle θ and a field scattered by a single tilted rough facet, the full expression of NRCS can be obtained as [38]

$$\left\{ \begin{array}{l} \sigma_{hh}^0 = \frac{4}{\pi} k^4 \cos^4 \theta_i s^2 W_n (2k \sin \theta_i) \\ \quad \times |F_h(\theta_i) \cos^2 \beta_s + F_v(\theta_i) \sin^2 \beta_s|^2 \\ \\ \sigma_{vv}^0 = \frac{4}{\pi} k^4 \cos^4 \theta_i s^2 W_n (2k \sin \theta_i) \\ \quad \times |F_v(\theta_i) \cos^2 \beta_s + F_h(\theta_i) \sin^2 \beta_s|^2 \\ \\ \sigma_{vh}^0 = \sigma_{hv}^0 = \frac{4}{\pi} k^4 \cos^4 \theta_i s^2 W_n (2k \sin \theta_i) \\ \quad \times |[F_v(\theta_i) - F_h(\theta_i)] \sin \beta_s \cos \beta_s|^2 \end{array} \right. , \quad (20)$$

where F_h and F_v are the Bragg scattering coefficients for horizontal and vertical polarizations, respectively [38]. θ_i is the local incidence angle, and the angle β_s is the rotation of the local incidence plane around the look direction \hat{k} related to the facet slopes. Equation (20) does not hold at near-grazing angles (i.e., $\theta_i \cong \pi/2$) and k values smaller than about $2\pi/L$ (i.e., $\theta_i < \lambda/2L$, where L is the facet linear size).

Considering that the large-scale roughness height variations are larger than the incident radiation wavelength and the facet size is larger than the small-scale roughness correlation length, the returns from different facets are uncorrelated. Within this context, the NRCS from the entire surface can be expressed by averaging that of a single facet over β_s and θ_i . Via a Taylor series expansion, the NRCS expressions of an entire resolution cell are given as follows [38]:

$$\left\{ \begin{array}{l} \langle \sigma_{hh}^0 \rangle_{a,b} = \frac{4}{\pi} \left[C_{0,0}^{hh} + \left(C_{2,0}^{hh} + 2 \frac{\Re(C_{0,0}^{hv}) - C_{0,0}^{hh}}{\sin^2 \theta} + C_{0,2}^{hh} \right) \sigma^2 \right] \\ \langle \sigma_{vv}^0 \rangle_{a,b} = \frac{4}{\pi} \left[C_{0,0}^{vv} + \left(C_{2,0}^{vv} + 2 \frac{\Re(C_{0,0}^{hv}) - C_{0,0}^{vv}}{\sin^2 \theta} + C_{0,2}^{vv} \right) \sigma^2 \right] \\ \langle \sigma_{hv}^0 \rangle_{a,b} = \frac{4}{\pi} \left(C_{0,0}^{hh} + C_{0,0}^{vv} - 2\Re(C_{0,0}^{hv}) \right) \frac{\sigma^2}{\sin^2 \theta} \end{array} \right. , \quad (21)$$

where $C_{k,n=k}^{pq}$ are series expansion coefficients of the function $(k \cos(\theta_i))^4 W F_p F_q^*$ [38]:

$$C_{k,n-k}^{pq} = \frac{1}{n!} \binom{n}{k} \left. \frac{\partial^n \left(W k^4 \cos^4 \theta_i F_p F_q^* \right)}{\partial a^k \partial b^{n-k}} \right|_{a=b=0} . \quad (22)$$

Even though the good compromise between interpretation, practical implementation, and accuracy issues [35] associated to the composite model combining the Bragg scattering mechanisms (contribution associated to capillary and short-gravity waves) and local-tilting effects associated to long waves, it is still difficult to obtain a consistent description of the sea surface NRCS over different polarization states and wind and wave conditions as well as over a large range of frequencies and incidence angles [41, 72]. Discrepancies between model and measurement are more significant when considering the sea surface under the influence of surface current straining. In

the literature, the authors have been suggested that that the contribution of surface breaking waves can be considered summing up with the polarized Bragg scattering mechanisms to describe the co-polarized sea surface backscattering, reducing the discrepancies [41, 42].

Other studies analyzed sea surface polarimetric scattering under more complex environmental conditions, i.e., high sea states, the presence of breaking waves and ice caps, internal waves, etc. When dealing with breaking waves, some studies propose that, at intermediate incidence angles, the scattering mechanisms associated to near-breaking events and intermediate-scale breaking waves are characterized as non-polarized, and, therefore, contributing the same for both co-polarized channels [43, 44]. Within this context, according to Kudryavtsev et al. [44], the contribution of breaking wave can be estimated from co-polarized SAR measurements. The model assumes that the sea surface can be described as the sum of a polarized two-scale Bragg sea surface scattering contribution, $\sigma_{qq,b}^0$, and a NP scattering from breaking waves, σ_{wb}^0 [43]:

$$\sigma_{qq}^0 = \sigma_{qq,b}^0 + \sigma_{wb}^0, \quad (23)$$

where q stands for horizontal or vertical polarization. The term σ_{wb}^0 can be solved taking into account the co-polarization difference (PD) and the two-scale Bragg scattering polarization ratio (PR) [43]:

$$\sigma_{wb}^0 = \sigma_{vv}^0 - \frac{\sigma_{vv}^0 - \sigma_{hh}^0}{1 - \sigma_{hh,b}^0 / \sigma_{vv,b}^0} = \sigma_{vv}^0 - \frac{PD}{1 - PR}, \quad (24)$$

where PR is mainly ruled by the local geometry and tilting effects [41, 43]. More details about PR can be found in Kudryavtsev et al. [43, 44].

1.4 Experimental Showcases

In this subsection, some key polarimetric characteristics of the sea surface scattering are shown by means of an experimental showcase.

The polarimetric SAR scene consists of a FP L-band Alos PalSAR-1 image collected over the Tosashimizu coast (Pacific Ocean, Japan) on April 21, 2011. The spatial resolution is 30×10 m (range \times azimuth), while the incidence angle at mid-range is about 24° . An excerpt of the SAR scene is shown as an RGB image in Fig. 1a, where red, green, and blue colors refer to the modulus of HH-, VV-, and HV-polarized scattering amplitudes, respectively. The dominant yellowish color witnesses that most of the backscattered signal is from the co-polarized channels. To quantitatively show this property, an along-range transect is selected covering a distance of almost 7 km. The behavior of the HH-, VV-, and HV-polarized NRCSS along the transect is shown—in gray, black, and purple lines—in Fig. 1b, where a

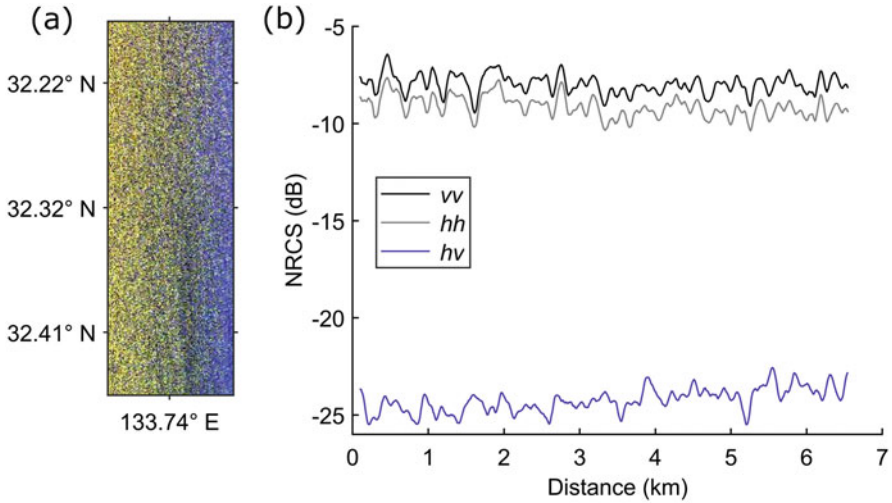


Fig. 1 (a) False color RGB ($R = |S_{hh}|$, $G = |S_{vv}|$, $B = |S_{hv}|$) image of a L-band Alos PalSAR-1 SAR scene excerpt collected over sea surface; (b) NRCS values (in dB scale) evaluated along a range transect where black, gray, and blue plots refer to vv , hh , and hv polarizations. A smoothing factor of 9 is used to improve visualization

dB scale is used and a 9-pixel-long smoothing window is applied for visualization purposes. It is clear how, over sea surface, co-polarized backscattering dominates over the cross-polarized one, i.e., about 15-dB difference, on average. In addition, it can be noted that the intensity of the VV-polarized backscattering is slightly larger (within 2 dB) than the HH-polarized one. Those properties all come from the peculiar characteristics of the Bragg/tilted-Bragg scattering ruling over sea surface under low-to-moderate wind conditions and in the incidence angle range from about 20 to 60°.

The correlation properties that characterize sea surface polarimetric backscattering can be analyzed using a second-order descriptor, i.e., the 3×3 covariance matrix. To analyze the correlation between the co-polarized channels, the complex-valued element C_{13} must be considered since its phase represents the phase difference between co-polarized channels, i.e., the so-called CPD φ_c . The corresponding probability density function (pdf) is shown in Fig. 2a, where it can be noted that it follows a Gaussian distribution whose width is quite narrow, i.e., less than 10°, witnessing that a large degree of correlation is in place between the HH- and VV-polarized backscattering channels. When dealing the correlation between co- and cross-polarized channels, sea surface satisfies the reflection symmetry property with respect to the line of sight that results in those backscattering channels being uncorrelated. The amplitude of the complex-valued element C_{12} , which is related to the correlation between the HH- and HV-polarized backscattering channels, is shown—in dB scale—as a false color image in Fig. 2b. By visually inspecting Fig. 2b, it can be observed that over sea surface reflection symmetry applies

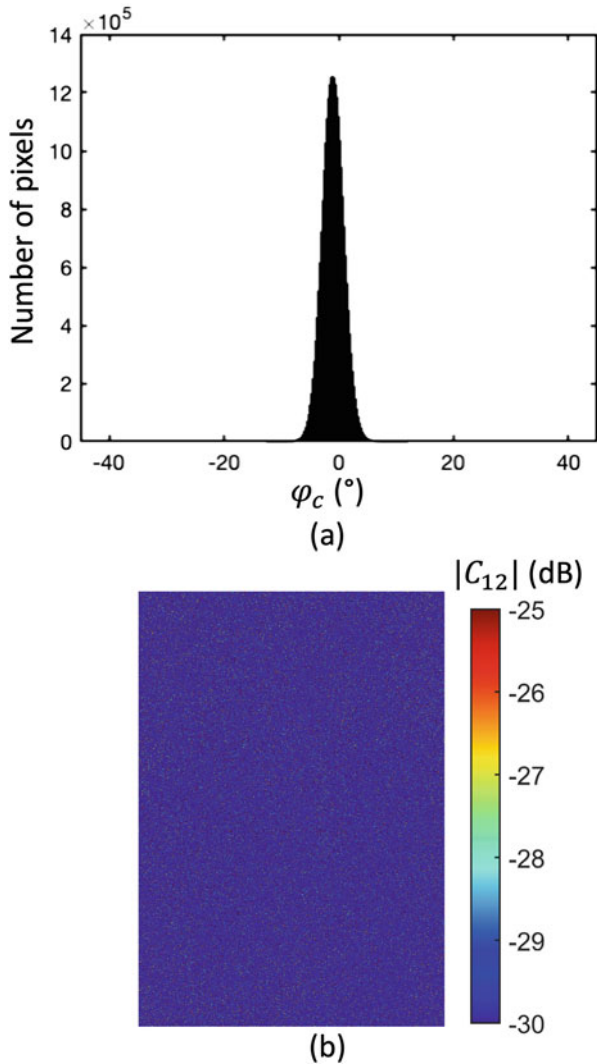


Fig. 2 Behavior of the covariance matrix estimated over sea surface relevant to the Alos PalSAR-1 SAR scene shown in Fig. 1: (a) Probability density function relevant to the phase (evaluated in degrees) of C_{13} ; (b) false color image of the amplitude (in dB scale) of C_{12}

everywhere, i.e., the correlation between HH- and HV-polarized backscattering channels calls for extremely low values (on average, about -34 dB).

To show that the ocean calls for a dominant scattering, i.e., the Bragg/tilted Bragg surface scattering mechanism, the three real and non-negative eigenvalues of the covariance matrix are evaluated, see Fig. 3, where they are normalized with respect to the total backscattered power (namely, the SPAN). As expected, sea surface is

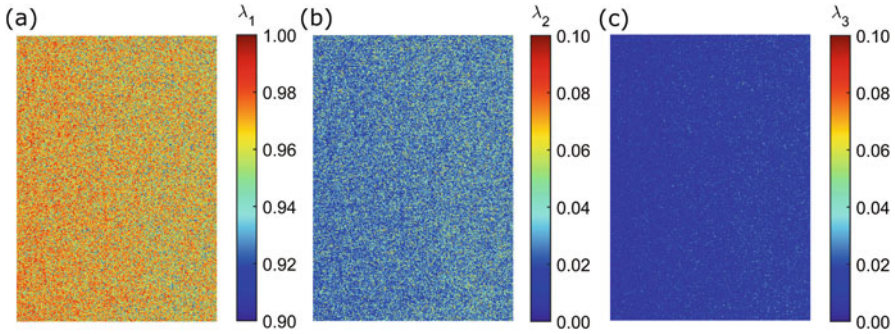


Fig. 3 False color images of the normalized eigenvalues of the coherency matrix evaluated over sea surface relevant to the Alos PalSAR-1 SAR scene shown in Fig. 1: (a) λ_1 ; (b) λ_2 and (c) λ_3

characterized by λ_1 values larger than 0.9, while λ_2 and λ_3 values lower than 0.1, witnessing that a single scattering mechanism dominates. On average, $\lambda_1 = 0.96$, while the secondary eigenvalues λ_2 and λ_3 are equal to 0.03 and 0.01, respectively.

2 SAR Polarimetry for Sea Oil Spill Observation

In this section, the capability of polSAR satellite measurements to observe oil spills at sea is presented.

Oil pollution has become one of the most frequent and catastrophic marine accidents. There are many sources of marine oil spill pollution, including the natural leakage of hydrocarbons from seabed oil and gas reservoirs, the discharge of industrial wastewater and domestic sewage, the exploitation of offshore oil resources, the blowout accidents of drilling platforms, the rupture of oil pipelines, the leakage and illegal sewage discharge oil tankers, etc. Oil spill mostly occurs in offshore waters, and it will cause huge damage to the marine environment and ecological resources. Crude oil contains a large number of toxic compounds and heavy metals. Once enters the marine ecological cycle, they will first affect the health and safety of low-grade marine plants, then fish, higher mammals, and human beings through the food chain. Oil spill pollution will not only affect the marine traffic, but also cause huge losses to the marine salt industry, offshore water power generation, seawater desalination, and marine aquaculture. To summarize, it will seriously threaten the people's health and economic development of the coastal region.

Remote sensing plays a key role in the early warning, response, and damage assessment of marine oil spill [12]. Compared with optical sensors, SAR has stronger capability of all-day and all-weather observation, which has demonstrated its advantage in operational oil spill observation, especially during adverse weather conditions when oil spill accidents frequently happen. SAR platforms for marine

oil spill observation can be mainly divided into spaceborne and airborne platforms. Spaceborne platforms have the advantage of large coverage and are relatively cost-effective, while airborne platforms have a higher signal-to-noise ratio and are more flexible in repeat observations during emergency responses and can better acquire the drifting and emulsification process of the oil spill.

In ancient Greece, the inhibitory effect of oil film on sea surface fluctuation was recorded in the literature [4]. In ancient navigation, experienced sailors spilled oil to the sea with wind and waves and use the attenuation characteristics of oil film to sea waves to prevent ships from overturning. Italian scientist Maragoni [13] explained this phenomenon theoretically for the first time: substances with different viscosity coefficients on the liquid surface will produce elastic resistance, so as to attenuate the amplitude of surface fluctuation. Therefore, the attenuation of sea oil film is called Marangoni attenuation (damping). Synthetic aperture radar observes backscattering caused by sea surface fluctuations of the sea surface. The oil film on the sea surface will diffuse and form a film with different viscosity coefficients, which attenuate the short-gravity wave and capillary wave, reduce the roughness of the sea surface, weaken the SAR backscattering, and form a dark area in graytones intensity SAR images, as shown in Fig. 4. Therefore, the detectability of sea oil film is closely related to the surface wind field of the sea. If the wind speed is too low, the sea surface will not fluctuate, resulting in extreme low backscattering; and if the wind speed is too high, the oil spill on the sea surface will be dispersed and drifted quickly, making it difficult to be detected. Therefore, the ideal wind speed for SAR sea surface oil spill detection is usually required to be 3–14 m/s [13].

This section is organized as follows: in Sect. 2.1, an overview on the use of *polSAR* imagery to observe sea oil slicks is presented; in Sect. 2.2, the most relevant *polSAR* approaches to monitor oil spills at sea are critically reviewed; in Sect. 2.3, an experimental showcase of marine oil spill observation methods using conventional classifiers and convolutional neural networks is demonstrated.

2.1 Overview

Marine oil spills are observed as dark spots on the sea surface. However, many other natural phenomena, such as low wind area, biogenic oil film, rain cell, upwelling, internal wave, atmospheric wave, etc., can also form similar strip or patch-like dark areas, may result in a false alarm. These phenomena are referred as “look-alikes.” Therefore, distinguishing oil film between look-alikes has become the key problem of SAR marine oil spill observation. Early SAR oil spill observation mainly relied on single-polarization SAR images. The oil film and look-alikes are classified based on the gray level, texture, and shape information. Therefore, these methods are usually composed of three steps: dark spot detection, feature extraction, and classification [13]. However, oil spill identification with single-polarization SAR data usually requires prior knowledge of the oil film and auxiliary information such as sea surface wind speed. The state of oil spill on the sea surface is influenced by the

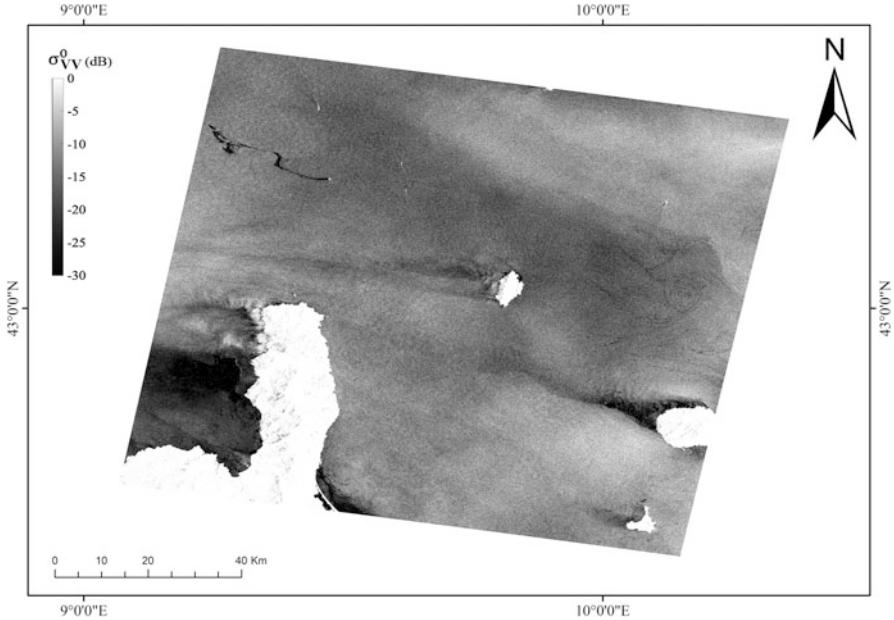


Fig. 4 An excerpt of the VV-polarized NRCS graytone image, shown in dB scale, collected by the C-band Sentinel-1 SAR on October 8, 2018, over a coastal region in the Mediterranean Sea affected by a certified ship-borne oil spill (see the low backscattering slick in the northwestern area). Note that the oil slick is about 30 km long. Several low backscattering regions of different origins (sheltered regions, low wind areas) can also be observed

sea state, which is sometimes very complex. The shape of oil film on the sea surface is related to the oil type and movement of oil source during leakage. Therefore, oil spill observation based on single-polarization SAR data is difficult to implement, and the accuracy cannot be guaranteed. In recent years, oil spill detection based on polSAR data has become a hot research topic. As introduced before, polSAR can obtain detailed backscattering properties of the target and provide more sufficient information for oil spill identification. Study demonstrated that the sea surface covered by mineral oil has distinct scattering mechanisms from the sea surface covered by biological oil film or open water: the Bragg scattering of the sea surface covered by mineral oil is weakened, while the non-Bragg scattering mechanism is enhanced, resulting in obvious depolarization effect. On the other side, the sea surface covered by biological oil film or open water is still dominated by Bragg scattering, with a backscattering of high degree of polarization [6, 56].

Polarization decomposition parameters were considered first for marine oil spill observation. By eigenvalue decomposition of the polarization coherency matrix, polarimetric entropy H and average polarization angle $\bar{\alpha}$ and anisotropy A can be extracted. These parameters are widely used in polSAR-based image analysis. The sea surface covered by biogenic film or open water is dominated by Bragg scattering,

resulting in a small H and $\bar{\alpha}$, while the Bragg scattering mechanism of the mineral oil-covered area is suppressed, making the polarization scattering mechanism more complex, and therefore, H and $\bar{\alpha}$ increase. In [59], the effectiveness of these polarization decomposition features in sea surface oil film detection was confirmed based on SIR-C/X SAR data, while a polarization constant false alarm rate (CFAR) filter for sea surface oil film detection based on these features [57]. In [58], it was first reported that the pedestal height can reflect the degree of depolarization of ground objects. The pedestal height of open water is lower than that of the oil-covered region, which can be used to effectively distinguish mineral oil and its look-alikes. The phase information in polSAR data is closely related to the polarization characteristics of ground objects. In [67], it was first proposed that the co-polarized phase difference (CPD) can be used to effectively distinguish oil spill and biogenic film. The analysis shows that the phase correlation of co-polarized backscattering signals on the sea surface of open water or covered by biogenic slick is high, leading to a phase difference close to 0, while the existence of mineral oil reduces the phase correlation of co-polarized channels and expands the distribution of phase difference. Therefore, the variance of CPD can be used as a powerful feature for oil film classification. In [79], C-band Radarsat-2 and X-band TerraSAR-X data in the North Sea oil spill experiment were analyzed, and the features including H and $\bar{\alpha}$, CPD standard deviation, amplitudes of co-polarized channels, correlation coefficient, and other features were compared on their capability in distinguishing between mineral oil and biogenic oil films.

The NRCS of different polarization channels can be used to classify marine oil films. In the range of medium incidence angles (approximately 20–60°), the radar backscattering of sea surface can be modeled by Bragg scattering. Theoretically, the Bragg scattering on the sea surface is related to the electromagnetic wave number, radar incidence angle, Fresnel coefficient, and sea surface two-dimensional spectrum. The ratio of different polarization backscattering cross sections is only a function of local incidence angle, sea surface slope, and equivalent dielectric coefficient. In [53], it studied the L-, C- and X-band images obtained in the SIR-C/X oil film field experiment, which confirmed this fact. In [63], the tilted Bragg scattering model was used to describe the backscattering of sea surface and oil-water mixture, and Unmanned Aerial Vehicle SAR (UAVSAR) images obtained during the oil spill accident in the Gulf of Mexico were used to retrieve the oil-water mixing ratio. They estimated that the oil-water mixing ratio of the region near the accident site was between 65 and 90%. In [66], the boundary perturbation method was considered to analyze the characteristics of different NRCS of oil-free and biogenic film covered sea surface. The experiments using actual SAR data show that the TSM and Marangoni damping model can better describe the sea surface backscattering cross section under the tilted modulation of large-scale waves. When the range of incidence angles is small, the NRCS of different polarization channels is dominated by the specular scattering component. With the increase of incidence angle, the specular scattering component decreases rapidly, and the radar backscattering cross section is gradually dominated by the contribution of Bragg scattering component. In this case, the small perturbation method (SPM) can be

used for modeling. In [89], an improved backscatter model was proposed based on polarimetric scattering mechanism for marine oil spill observation, which is more in line with the characteristics of oil film covered sea surface, and improved the accuracy of the retrieval of equivalent dielectric constant and other parameters. The NRCS-based analysis methods only rely on the amplitude information of polSAR image; therefore, they can be used in polSAR sensors with incoherent polarimetric imaging modes such as ENVISAT ASAR and CSK. In the dual-polarization mode of these platforms, the signals of different polarization modes are transmitted and received alternately. However, the application of sea surface SPM is greatly limited by the incident angle. At small incidence angles, the specular scattering is too strong, while at large incidence angles the backscattering is too weak. Moreover, the Fresnel coefficient will saturate with the change of sea surface equivalent dielectric constant [36], which affects the accurate retrieval of oil–water mixing ratio. In addition, only when the oil film thickness reaches the order of the skin depth of electromagnetic wave (approximately 1 and 4 mm at C-band and L-band, respectively), its change will have an observable effect on the radar NRCS [85]. However, usually the thickness of oil film is far from the above conditions. In addition, methods based on the NRCS ratio do not make full use of the depolarization effect of sea oil film, which greatly limits its ability to distinguish between mineral oil and its look-alikes.

2.2 PolSAR for Marine Oil Spill Observation

2.2.1 Feature Extraction to Monitor Sea Oil Spills

In this paragraph, some of the most widely used features derived from polSAR imagery to observe marine oil pollution and their expected behavior are summarized. Most of them have been already introduced in Sect. 1.1 and will be interpreted in terms of slick-free and oil slick-covered sea surface, while other features will be first introduced that are specifically proposed for sea oil slick observation purposes. They are as follows:

- Polarimetric entropy
- Degree of polarization
- Ellipticity angle
- Normalized pedestal height
- CPD standard deviation
- Conformity coefficient
- Correlation coefficient
- Coherence coefficient

The intensity of co-polarized channels is frequently used in single-polarization SAR-based oil spill detection algorithms, and therefore, it can be taken as a reference. The VV-polarized NRCS is usually considered due to its higher signal-to-noise ratio (SNR) if compared to the HH-polarized one [54, 55].

The polarimetric entropy, H , assumes values close to 0 over clean sea surface, since the almost completely polarized Bragg scattering dominates, while over oil slick-covered areas, the more random scattering mechanism due to depolarization effects results in larger entropy values, i.e., close to 1. However, for weak-damping slicks as biogenic films, although the backscattering power is still lower than the sea background due to their damping properties, the main scattering mechanism is still Bragg-like [10], i.e., entropy values similar to that of clean sea surface apply.

The degree of polarization, DOP, can be derived from the Stokes vectors of any coherent SAR imaging mode including DP, CP, and FP architectures [77]. When dealing with clean sea surface and weak-damping slicks, the significantly polarized Bragg scattering mechanism results in large DOP values, i.e., close to 1. When sea surface is covered by mineral oil, the latter induces remarkable depolarization effects, and therefore, lower DOP values (approaching 0) are observed.

The ellipticity angle, χ , describes the polarization status of a monochromatic plane electromagnetic wave [17, 77]:

$$\sin(2\chi) = -\frac{g_3}{mg_0}, \quad (25)$$

where g_0 and g_3 represent the first (i.e., the total backscattered power) and the fourth (i.e., the one related to circular polarization) elements of the Stokes vector \mathbf{g} , while m is the degree of polarization of the wave. For slick-free sea surface, where Bragg scattering rules, χ is negative, while for oil-covered sea surface, since a more random scattering mechanism is in place, χ is positive [62]. Therefore, χ can be used as a straightforward binary descriptor to distinguish slick-free from oil-covered sea surface [62, 67].

The normalized pedestal height, namely NPH, represents the amount of unpolarized backscattering energy. Hence, for clean sea surface, the almost completely polarized Bragg scattering mechanism results in NPH values close to 0, while for the oil-covered sea surface, much larger NPH values (approaching 1) are expected due to the non-Bragg scattering that reflects the depolarization induced by the mineral oil.

The standard deviation of the CPD, evaluated from φ_c using a sliding window, is an unbiased estimator of the correlation between co-polarized backscattering channels. Over slick-free sea surface, the correlation between co-polarized channels is high, and therefore, a narrow CPD distribution, i.e., a low CPD standard deviation value, is expected [60]. This also applies over weak-damping surfactants. When dealing with mineral oil slicks, the depolarization they induce in scattering mechanism results in a remarkable reduction of the correlation between HH and VV backscattering channels. Accordingly, the pdf of the CPD broadens resulting in a larger standard deviation of φ_c [11, 70].

The conformity coefficient, μ , was first used for soil moisture estimation from CP SAR imagery purposes. When a FP SAR measurement is available, its proxy is defined as [93]:

$$\mu \cong \frac{2(\operatorname{Re}(S_{hh}S_{vv}^*) - |S_{hv}|^2)}{|S_{hh}|^2 + 2|S_{hv}|^2 + |S_{vv}|^2}. \quad (26)$$

The conformity coefficient μ evaluates whether surface scattering is dominant among all the elementary scattering mechanisms. Over a slick-free sea surface, Bragg scattering results in a very small cross-polarized backscattering power while calling for high correlation between co-polarized channels, i.e., $\operatorname{Re}(S_{hh}S_{vv}^*) > |S_{hv}|^2$; hence, μ is positive. Over oil-covered sea surface, as non-Bragg surface scattering is in place, the co-polarized correlation is lower, while the cross-polarized backscattering component keeps almost the same, i.e., it is very likely to have $\operatorname{Re}(S_{hh}S_{vv}^*) < |S_{hv}|^2$, thus resulting in negative μ values. Considering weak-damping slicks, since Bragg scattering is still dominant, positive μ values are expected. Under this rationale, conformity coefficients can be used to effectively distinguish crude oil from biogenic slicks without any need of external thresholding methods.

The correlation and coherency coefficients can be derived from the coherence matrix as follows [78]:

$$\rho_c = \left| \frac{\langle S_{hh}S_{vv}^* \rangle}{\langle S_{hh}^2 \rangle \langle S_{vv}^2 \rangle} \right|, \quad C_c = \frac{|\langle T_{12} \rangle|}{\sqrt{\langle T_{11} \rangle \langle T_{22} \rangle}}. \quad (27)$$

They both range between 0 and 1. Over a slick-free sea surface, the co-polarized channels are highly correlated, so they are expected to be very close to 1, while over oil-covered sea surface, a much lower co-polarized correlation is expected; thus they are both approaching 0.

The general behavior of the above introduced set of polarimetric over slick-free and slick-covered sea surface is summarized in Table 2.

Table 2 Main polSAR features used for sea oil spill monitoring. Note that the VV-polarized NRCS is also listed as a reference

PolSAR feature	Sea surface	Mineral oil—strong damping	Biogenic slick—weak damping
H	Lowest	High	Low
DOP	High	Low	High
χ	Negative	Positive	Negative
NPH	Lowest	High	Low
σ_{φ_c}	Lowest	High	Low
μ	Positive	Negative	Positive
ρ_c	Highest	Low	High
C_c	Highest	Low	High
σ_{vv}^0	High	Lowest	High

2.2.2 CP SAR Architectures

The FP spaceborne SAR system alternately transmits horizontal and vertical polarization signals. Therefore, the pulse repetition rate of linear frequency modulation signal is twice that of single-polarization SAR systems, resulting in halved swath width, which lead to range ambiguity effect and increased system power requirements [14]. In addition, the large system complexity and data volume also increase the cost of FP SAR systems. In order to overcome such issues, the CP architectures were proposed. The CP SAR systems can obtain part of the polarization characteristics of the observed targets without reducing the width of the swath [75]. At present, it has achieved promising results in land use classification, biomass estimation, soil moisture retrieval, and several marine applications [1, 47]. For the application of maritime monitoring, the revisit time is a very important technical index, so CP SAR has become a hot research field for marine oil spill observation [61]. The commonly used CP SAR imaging modes in oil spill detection mainly include $\pi/2$, also known as circular polarization transmitting, linear polarization receiving (CTLR), or hybrid polarization mode, and $\pi/4$ or slant linear modes. These two modes transmit circular polarized or 45° linear polarized signals, respectively, and receive horizontal and vertical polarization signals simultaneously. The target scattering vectors \mathbf{k} of those CP SAR sensors are

$$\mathbf{k}_{\frac{\pi}{2}} = \frac{1}{\sqrt{2}} \begin{pmatrix} S_{hh} - jS_{vv} \\ S_{vh} - jS_{vv} \end{pmatrix}, \quad \mathbf{k}_{\frac{\pi}{4}} = \frac{1}{\sqrt{2}} \begin{pmatrix} S_{hh} + S_{hv} \\ S_{vh} + S_{vv} \end{pmatrix}. \quad (28)$$

One way for processing CP SAR data is to reconstruct the pseudo-quad-polarization covariance matrix from the compact polarization scattering vector by using iterative algorithms [65, 81] and then use feature extraction methods for FP SAR data. In [86], a CP SAR image reconstruction algorithm based on polarization decomposition was proposed and applied to ship detection. In [18], an empirical model was exploited to estimate the constant parameter N in the range of incidence angles to improve the reconstruction accuracy. The advantage of feature extraction methods based on pseudo-quad-polarization reconstruction is that the analysis methods for fully polarized modes can be directly used. However, for these methods, assumptions of backscattering characteristics are required, which do not always hold for the sea surface. Therefore, sometimes there is a large deviation between the reconstructed pseudo-quad-polarization covariance matrix and the real data. In [52], the backscattering characteristics of the sea surface were analyzed through the statistical analysis of FP UAVSAR data and put forward an improved hypothetical equation, which obtained a better reconstruction performance.

Another way is to analyze the scattering vector of CP SAR data and extract the features directly. The Stokes vector of the radar signal can be obtained from the CP scattering vector, so as to further calculate the degree of polarization m and relative phase δ , wave polarization entropy H_w , ellipticity angle χ , and average polarization angle $\bar{\alpha}_w$. In [50], it was found that the sign of δ can distinguish different sea surface scattering mechanisms: for sea surface region δ is close to

90° , and for oil film covered area δ is negative, making it a binary classification index. Through the image analysis of mineral oil naturally leaked from Radarsat-2 sea surface, in [51], it was also found that compared with the clean sea surface, the m value of the oil film covered area is significantly reduced, indicating that the depolarization effect is obvious and the ellipticity angle is high, and the opposite sign of χ indicates that the scattering mechanism is no longer Bragg scattering. Conformity coefficient μ was first used for soil moisture estimation based on CP SAR. It can effectively distinguish single surface scattering, double-bounce scattering, and volume scattering [93]. The μ extracted from FP and $\pi/4$ CP SAR images have been proven to have a very good ability to distinguish between mineral oil film and biogenic look-alikes. It is positive on the open water and negative on the mineral oil film, so can be used as a logic classifier for sea surface oil spill detection. Based on the extended Bragg scattering model, in [88], a new method was proposed to extract features from the Stokes matrix of CP SAR, and its performance in distinguishing sea oil spill from biogenic slicks and low-wind-speed area was confirmed through experiments. In [91], various features extracted from $\pi/2$ CP SAR mode were analyzed by using the quad-pol reconstruction and direct feature extraction, respectively. It was found that the two kinds of feature extraction methods have their own advantages over each other, and the marine oil spill classification performance of CP SAR is close to the FP SAR mode. In [90], performance of features extracted from DP, CP, and FP SAR imaging modes was investigated. It was found that the classification accuracy will not always increase with the number of features, indicating that there is a large amount of complementary information between polarimetric features, which highlights the importance of feature selection and optimization. In [45], actual $\pi/2$ CP mode RISAT-1 SAR data were used for the first time during an oil spill experiment carried out on Norwegian waters.

Comparative studies have been made on CP SAR modes. The $\pi/2$ CP SAR mode has the advantages of convenient polarization calibration, polarization channel power balance, not dependent on the direction of ground objects and not easily affected by the Faraday rotation effect of the ionosphere [81]. However, it is difficult to transmit an ideal circular polarized signal in engineering practice. The $\pi/4$ CP SAR mode is relatively easier to implement, but there is a 3 dB loss of received power due to the mismatch between the transmitting and receiving polarimetric channels. In [8, 9], the variability of CP SAR features under different incident angles was investigated through experiments on Alos-1, Alos-2, and Radarsat-2 spaceborne SAR images. The ability of CP SAR modes in distinguishing oil film and weak-damping look-alikes was verified, and the differences between polarimetric characteristics obtained under various CP and FP SAR systems were discussed.

2.2.3 Challenges and Research Trend

Controversy still exists on the polarimetric scattering mechanism of marine oil spill. There are beliefs that the increase of non-Bragg scattering component the

SAR image of an oil spill area is probably not mainly caused by the sea surface physical process, but thermal noise of the radar system [40]. In [26], the effects of both additive and multiplicative noise were also analyzed on L-, C- and X-band SAR data. The depolarization effect observed over oil spills in spaceborne SAR imagery was mainly attributed to the additive noise of the SAR sensor. It was also claimed that the non-Bragg scattering occurring over slick-covered sea areas is likely due to a misinterpretation of SAR images collected at a too low SNR [19, 26]. However, it has to be noted that the polarimetric backscattering mechanism of oil film is related to many factors including wind speed and the amount of oil leakage. In [10], a sensitivity analysis on the standard deviation of the CPD for marine oil spill observation was undertaken and confirmed that noise plays a role in broadening the distribution of CPD, especially at large incidence angles. The authors claimed that the depolarization is both induced by the noise and oil film. At lower incidence angle, the scene induced depolarization is dominant, while at larger incidence angle, the noise floor plays a more important role. In other words, together with the depolarization inherently introduced by oil, an additional depolarization contribution is due to noisy oil samples. Nonetheless, from the aspect of oil spill detection, this is a good point since the heavier the oil depolarization, the larger the separability with the polarized sea scattering is. On the other side, it makes the classification/characterization/oil parameter retrieval made on “noisy” oil samples unreliable.

The retrieval of the detailed properties of oil slicks and their evolving/drifted under marine environment has been a hot research topic, which received growing attention in recent years. In [74], the use of PTSM is proposed to retrieve the dielectric parameters of oil slick from the polSAR imagery. In [25], FP and CTRL SAR features were analyzed to observe evolving oil spills. They also developed and explored new quantitative and semi-automated methods for analyzing oil slick evolution using a time series of L-band polSAR images with short repeat time [27]. In [39], an analysis on newly formed sea ice distinction near the oil platform in the Pechora Sea was performed using Radarsat-2 polSAR observations. These studies demonstrated the valuable role played by polarimetric information.

With the increase of available polSAR data, deep-learning-based methods have shown great potential in improving the accuracy of marine oil spill classification. As data-driven pattern recognition methods, deep-learning-based algorithms can better exploit the semantic and contextual information within high-resolution SAR images without the need of prior knowledge. Chen et al. [15] used stacked auto-encoder (SAE) and deep belief neural network (DBN) to extract and optimize polSAR features. An oil spill detection method exploiting convolutional neural network and image stretching based on superpixel was proposed in [27], where the effectiveness of the approach was successfully demonstrated on Sentinel-1 DP SAR data. In [46], different parameters were exploited with sensitivity to the dielectric constant and ocean wave damping properties and used CNNs for learning nonlinear features, shapes, and textural and statistical patterns, in order to obtain significant classification accuracy. In [80], a novel oil spill identification method based on multi-layer deep feature extraction by CNN was proposed. These studies show that

compared with traditional supervised learning methods, the deep-learning-based methods with unsupervised pre-training can improve the accuracy of sea surface oil film detection, especially when the training samples are limited.

2.3 Experimental Showcase

2.3.1 SAR Polarimetry for Sea Oil Spill Observation: Conventional Classifiers

This showcase is addressed by means of a C-band FP SAR scene collected at C-band from the Radarsat-2 mission. The image was obtained in the North Sea area near Norway, which was obtained from an oil-on-water field experiment [79]. We selected a sub-region with 2000×2000 pixels from the original SAR image that contains clean sea surface and three types of oil films, including mineral oil film, emulsified oil film, and biogenic film. Since the emulsified oil represents an intermediate behavior between crude oil film and clean sea surface, whose polarimetric features are not typical, the classification of emulsified oil film is not considered in this chapter. Figure 5 shows the Pauli RGB image and artificial sampling labels.

In this section, support vector machine (SVM) is used as a representative of classic supervised classifier for its high performance in remote sensing applications. SVM relies on the maximization of the classification margin based on the principle of structural risk minimization. Its good generalization ability is obtained by

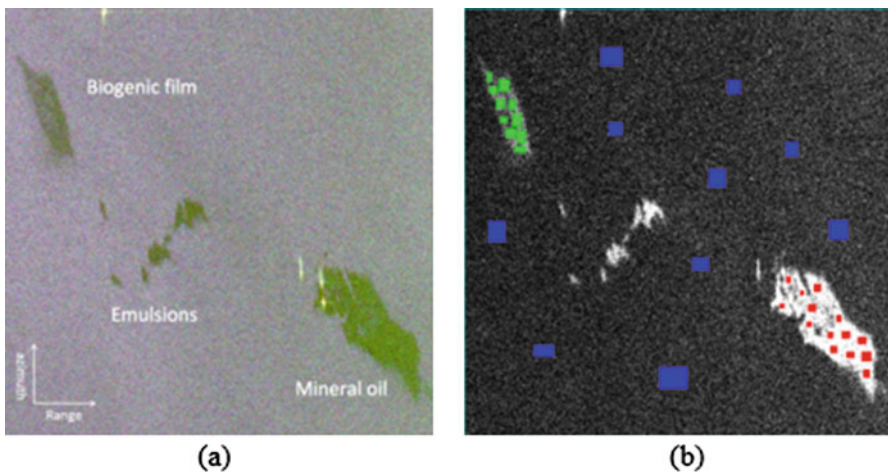


Fig. 5 Experimental showcase: (a) false color Pauli RGB image and (b) VV-polarized NRCS graytone image with sampling labels (blue: seawater; red: mineral oil; green: biogenic film)

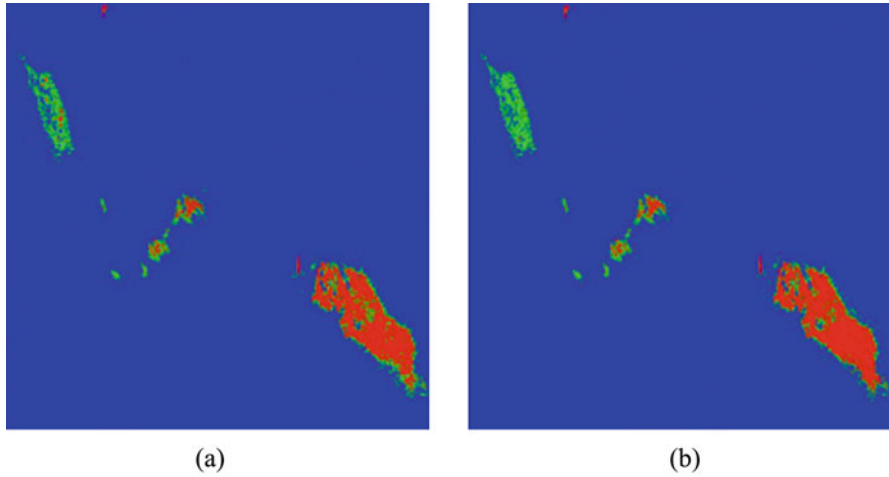


Fig. 6 Classification result relevant to: (a) the first and (b) the second feature sets

Table 3 Confusion matrix derived from the first feature set. OA = 91.94%

Ground truth (pixels)			
	Mineral oil	Biogenic film	Sea
Mineral oil	12,862	846	24
Biogenic slick	1106	13,324	1039
Sea	20	4945	64,867

Table 4 Confusion matrix derived from the second feature set. OA = 95.60%

Ground truth (pixels)			
	Mineral oil	Biogenic film	Sea
Mineral oil	13,564	728	17
Biogenic film	413	16,182	983
Sea	11	2205	64,930

constructing a lower Vapnik–Chervonenkis dimension function set in a high-dimensional space.

Pixel-level samples of mineral oil, biogenic film, and clean seawater were selected in the region of interest. Two feature sets are generated in the experiment. The first feature set consists of the 9 independent real-valued elements of the coherency matrix, while the second feature set consists—in addition to the first feature set—of the 10 polarimetric features listed in Table 2. The multi-layer perceptron kernel function is used as the kernel function of the SVM. The classification maps are shown in Fig. 6, while the corresponding confusion matrices are listed in Tables 3 and 4, respectively.

By analyzing the results listed in Tables 3 and 4, it can be seen that the classification accuracy is improved by 3.7% when adding the pre-defined polarimetric

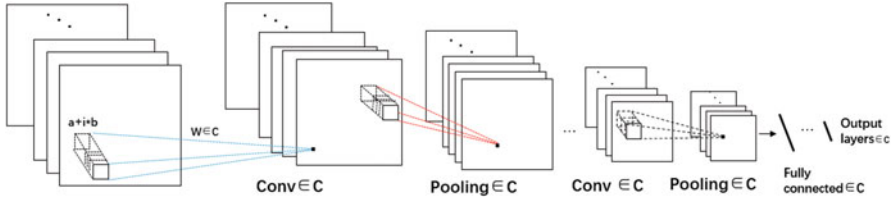


Fig. 7 Network structure of CVCNN[83]

features. The significant improvement in accuracy is mainly due to the reduced mis-classification rate between biogenic film and mineral oil, manifesting that the artificially defined polSAR features based on scattering mechanism carry key information for distinguishing mineral oil spill and its biogenic look-alikes.

2.3.2 SAR Polarimetry for Sea Oil Spill Observation: Convolutional Neural Network Classifiers

In the fields of image classification and target detection, deep learning models represented by the convolutional neural network have shown superior performance. Several studies were proposed to classify polSAR images using CNN, but since these methods take real-valued data as the input, the phase information in polSAR images could not be fully utilized. Therefore, [83] presented a complex-valued convolutional neural network (CVCNN) with complex-valued data as its input, which is of great significance for processing images containing complex-valued information. Trabelsi et al. [83] applied CVCNN to the classification of ground objects in polSAR images and achieved good results. In this chapter, the CVCNN is adjusted for marine oil spill detection tasks, and its performance is compared with real-valued convolutional neural network (RVCNN) based on different feature sets as the input. The network structure of CVCNN [83] used in this chapter is shown in Fig. 7, including input layer, convolutional layers, pooling layers, fully connected layer, and output layer. Different to the general CNNs, the weight and bias of CVCNN, including the input and output of all layers, are fully defined in the complex domain.

In CVCNN, the model input is a complex number, then its real part A and imaginary part B are expressed as logically different real numbers, and complex operation is simulated internally using real number algorithms. In the convolutional layer, complex convolution operation can be expressed as

$$\mathbf{W} \times \mathbf{h} = (\mathbf{A} \times x - \mathbf{B}^* y) + j (\mathbf{B}^* x + \mathbf{A} \times y) \quad (29)$$

$$\begin{bmatrix} \text{Re}(\mathbf{W} \times \mathbf{h}) \\ \text{Im}(\mathbf{W} \times \mathbf{h}) \end{bmatrix} = \begin{bmatrix} \mathbf{A} & -\mathbf{B} \\ \mathbf{B} & \mathbf{A} \end{bmatrix}, \quad (30)$$

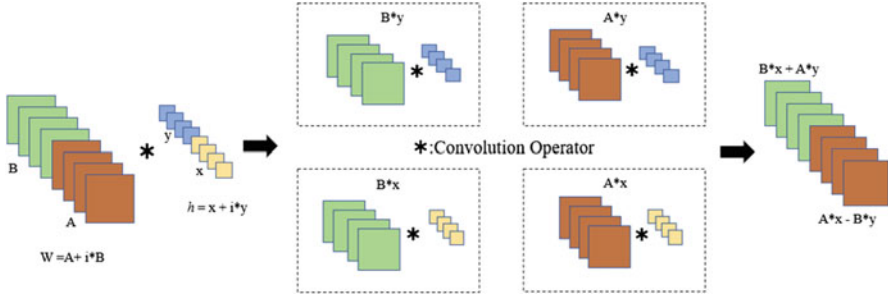


Fig. 8 Complex convolutional operation process

where \mathbf{h} is the input of complex-valued vector network, i.e., $\mathbf{h} = x + jy$; \mathbf{W} is the weight complex matrix, i.e., $\mathbf{W} = \mathbf{A} + j\mathbf{B}$. $\text{Re}(\cdot)$ and $\text{Im}(\cdot)$ mean real and imaginary parts, respectively. For a more intuitive representation, the complex convolutional operation process is shown in Fig. 8.

Accordingly, the ComplexReLU (CReLU) function is used as the complex activation function of CVCNN, whose operation process is as follows:

The weight and bias are initialized by means of random initialization. The complex backpropagation algorithm adopts stochastic gradient descent (SGD) to optimize the real part and imaginary part.

Figure 9 shows the structure of the proposed CVCNN and RVCNN for comparison. It should be noted that all the off-diagonal elements of the coherency matrix are complex-valued data, which contain important phase information. Therefore, T_{11} , T_{12} , T_{13} , T_{22} , T_{23} , and T_{33} are input into the classification network as six channels. While for RVCNN, the input is real-valued T_{11} , T_{22} , T_{33} , and the real part and imaginary part of T_{12} , T_{13} , and T_{23} , so nine channels are taken as the input. To avoid the interference by the network layer structure, the network layer parameters of the two networks are kept consistent. Finally, the network outputs the classification result of three types of targets, namely mineral oil slick, biogenic oil film, and seawater.

The same two feature sets are also applied in the oil spill classification experiments based on real-valued and complex-valued convolutional neural networks, respectively. For CNN, each pixel in the image is represented by a local patch defined by a neighborhood window. In this chapter, a 12×12 sliding window was used to obtain the data input of the CNN. Therefore, the classification model captures not only the polarimetric characteristics but also the spatial and texture patterns surrounding the center pixel that to be classified.

For the first feature set, T_{11} , T_{12} , T_{13} , T_{22} , T_{23} , and T_{33} are directly input into the CVCNN as 6 channels, while the real-valued T_{11} , T_{22} , T_{33} , and the real part and imaginary part of T_{12} , T_{13} , and T_{23} are input into the RVCNN as 9 channels, respectively. Then for the second feature set, the previously introduced polarimetric features are also input into RVCNN and CVCNN as supplementary feature dimensions.

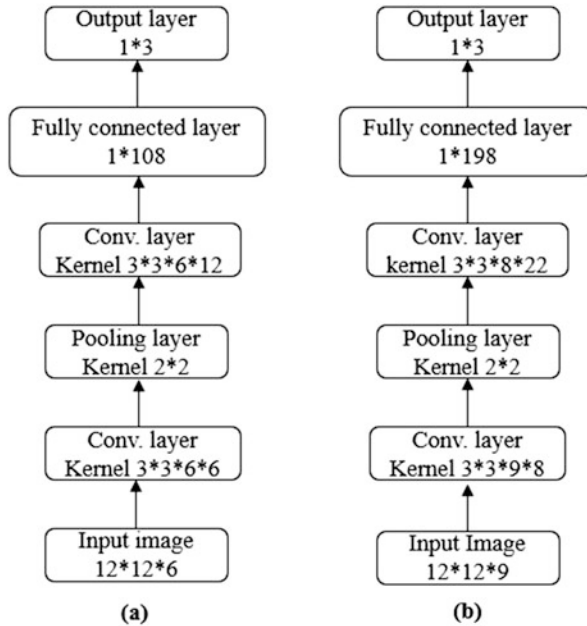


Fig. 9 Flowchart of: (a) CVCNN and (b) RVCNN

It can be observed from Fig. 10 and Tables 5 and 6 that, based on the elements only on T matrix, CVCNN has better classification performance than RVCNN by 0.3494%. The main reason is that the proposed RVCNN can better take advantage of the phase information contained in the polSAR data, which greatly helps to distinguish mineral oil and biogenic films. As shown in Fig. 11 and Tables 7 and 8, when the pre-defined polSAR features are introduced, the oil spill classification accuracy derived by RVCNN- and CVCNN-based methods improved by 0.9032% and 0.6736%, respectively, resulting in the classification accuracy derived by CVCNN still slightly higher than RVCNN. The results demonstrated that these polarimetric SAR features have relatively larger help to RVCNN-based model, by providing key polarimetric information hidden in the phase information of the complex backscattering coefficients. The experimental results demonstrated that the introduced SAR features provide key polarimetric information for improving the performance of oil spill classification. It is preliminarily shown that the CVCNN has the overall best performance for its ability of extracting both special and polarimetric information from polSAR data. Theoretically, deep-learning-based methods have higher potential given a larger number of training samples. Its advantage on dealing with complex function fitting problems may provide greater help in oil spill observation tasks under more complex environments and various oil type conditions.

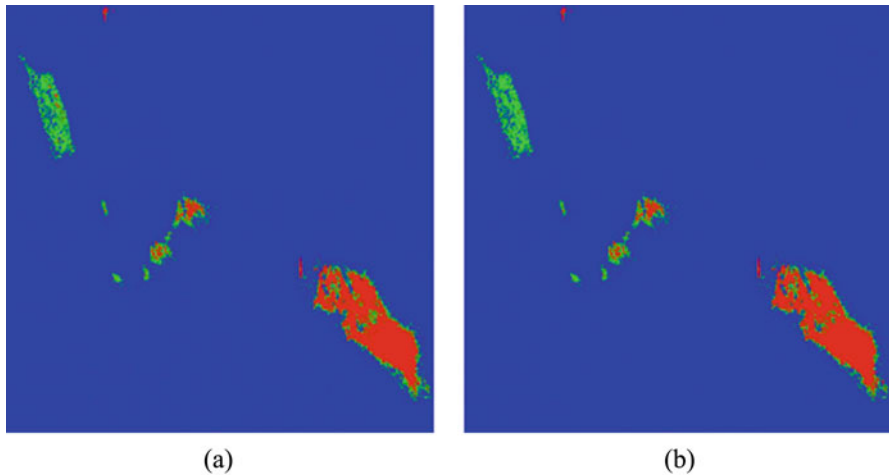


Fig. 10 Classification results obtained from the first feature set using (a) RVCNN and (b) CVCNN

Table 5 Confusion matrix derived from the first feature set by RVCNN. OA = 95.97%

Ground truth (pixels)			
	Mineral oil	Biogenic film	Sea
Mineral oil	13,602	426	9
Biogenic film	373	16,940	1422
Sea	13	1749	64,499

Table 6 Confusion matrix derived from the first feature set by CVCNN. OA = 96.32%

Ground truth (pixels)			
	Mineral oil	Biogenic film	Sea
Mineral oil	13,774	264	11
Biogenic film	197	17,219	1525
Sea	17	1632	64,394

3 SAR Polarimetry for Shoreline Monitoring

In this section, the capability of polSAR satellite measurements to monitor shorelines and to support coastal area management is provided.

Coastal areas are, worldwide, economic and natural resources of extraordinary value that, being often fragile and dynamic environments mostly largely urbanized, are particularly vulnerable to natural and anthropogenic hazards. Although in 2011 it was estimated that the world coastline length is around 1 million kilometers [5], this value tends to rapidly change over time due to natural phenomena, e.g., sea-level rise, erosion and sedimentation, and human-induced processes, e.g., urbanization and deforestation. In addition to the coastal area vulnerability issue, the fast growth

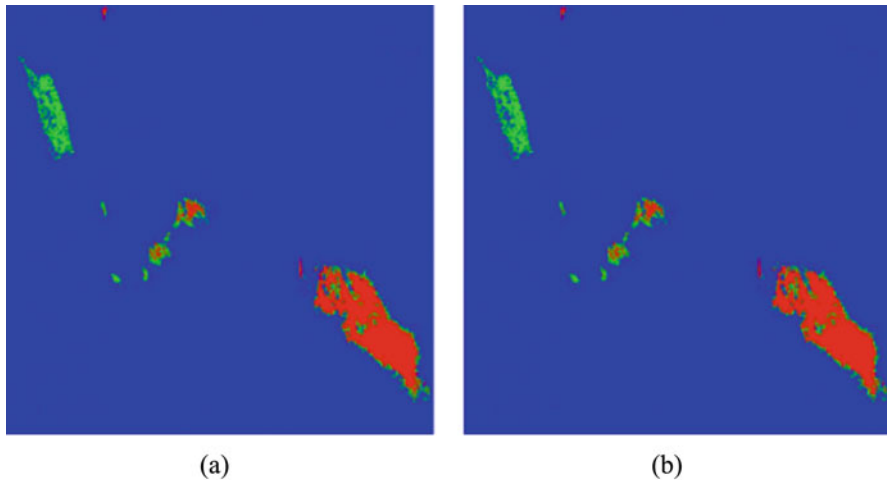


Fig. 11 Classification results obtained from the second feature set using (a) RVCNN and (b) CVCNN

Table 7 Confusion matrix derived from the second feature set by RVCNN. OA = 96.87%

Ground truth (pixels)			
	Mineral oil	Biogenic film	Sea
Mineral oil	13,729	198	32
Biogenic film	238	17,153	845
Sea	21	1764	65,053

Table 8 Confusion matrix derived from the second feature set by CVCNN. OA = 96.99%

Ground truth (pixels)			
	Mineral oil	Biogenic film	Sea
Mineral oil	13,811	197	18
Biogenic film	161	17,241	910
Sea	16	1677	65,002

of the coastal population density and the increase in economic assets and critical infrastructures in coastal areas pose a serious threat to human society. Hence, accurate and systematic observation of the coasts over time and, therefore, methods to predict the coastal evolution play a fundamental role in coastal zone management. Indeed, effective operational service for coastal areas monitoring is a key topic for local authorities that face the aforementioned threats for the stability of land and the safety of people they are responsible for. Within this context, microwave remote sensing plays a fundamental role in coastal area monitoring. In particular, SAR sensors, due to all-day and almost all-weather acquisitions, together with a wide area coverage and a fine spatial resolution, can be very useful for coastal area monitoring purposes [71].

This section is organized as follows: in Sect. 3.1, the most up-to-date advancements in coastline extraction from polSAR imagery is presented; the methodology to process polSAR imagery to extract the coastline is detailed in Sect. 3.2, while experimental results are showcased and discussed in Sect. 3.3.

3.1 *State-of-the-Art*

In this section, an overview of the studies that deal with shoreline monitoring using polSAR measurement is presented. In [87], the enhancement of the land/sea contrast is undertaken using polarimetric methods. Experimental results, undertaken on C-band Convair airborne polSAR data, show that the polarimetric combination outperforms single-polarization ones. The radar frequency dependence is discussed in [64], where results showed that higher frequency (C- or X-band) provides the best coastline localization. In [3], an analysis of the polarimetric channels with respect to the angle of incidence is undertaken. Experimental results show that, at low AOI ($<30^\circ$), the cross-polarized channel performs better than the co-polarized ones, while at higher AOI, no polarization dependence is exhibited. Co-polarized DP CSK SAR data, collected using the incoherent PingPong Stripmap mode, are successfully exploited to extract coastlines by Nunziata et al. [68]. Coastline extraction in an intertidal flat area is addressed in [23], where experimental results, obtained using CSK DP SAR scenes, pointed out that the extraction accuracy decrease in the case of water within the intertidal flat. In [7], a multi-polarization analysis of coastline extraction is undertaken using X-band single-polarization CSK SAR data. Results show that the performance of cross-polarized channels depends on the incidence angle while showing a small sensitivity to sea-state conditions. Conversely, co-polarized channels show a remarkable sensitivity to sea-state conditions. In addition, it is also pointed out that sandy coasts are badly detected due to very limited sand/sea contrast. In [69], the DP metric, based on the correlation between co- and cross-polarized channels, is exploited to enhance the contrast between sea and land. Results, obtained processing DP C—and X-band SAR data, demonstrate the soundness of the proposed approach for coastline extraction purposes. The FP information is exploited in [29]. In this chapter, the surface and the volume component obtained from the Freeman–Durden decomposition are used to enhance the discrimination between the sea and a challenging scenario that includes sandy beaches. Experimental results show that the surface component provides the best performance in terms of accuracy in detecting the sandy beach. In [20], shoreline rotation has been analyzed to provide a better understanding of the morphodynamic processes of natural embayed beaches. In [30], a two-year time series of multi-polarization Sentinel-1 SAR imagery is exploited to analyze the changes in the water-covered area of the Monte Cotugno (Italy) reservoir. Experimental results, verified using independent in situ measurements, demonstrate, first, that Sentinel-1 time series can be successfully used to support the smart water management of reservoirs, and second, multi-polarization feature outperforms SP ones in terms of

accuracy of the extracted waterline profile. In [31], the joint use of non-local speckle filtering and multi-polarization features is applied on a very challenging scenario that includes different habitats such as wetland, salt marshes, sand dunes, sand banks, mudflats, and intertidal flats. Experimental results, undertaken on C-band FP Radarsat-2 SAR imagery, show that the joint combination of non-local speckle filters and dual-polarimetric information provides the best accuracy.

3.2 Methodology

In this section, the methodology developed to address coastline extraction, depicted in the block diagram of Fig. 12, is discussed. Coastline extraction is basically based on two steps. The first step relies on the enhancement of the separation between sea and land, and it is addressed by exploiting the polarimetric information. Following the block diagram of Fig. 12, first, a pre-processing of the polSAR imagery is undertaken that includes calibration, spatial multilooking to reduce the speckle noise using a window size $N \times N$, spatial geocoding, and subset generation that includes the region of interest. Then, a multi-polarization feature is introduced that was found to improve the degree of scattering separability between land and sea [69]. This parameter, labeled as r , consists of average product between co- and cross-polarized backscattering amplitudes:

$$r = \langle |S_{xx}| |S_{xy}| \rangle. \quad (31)$$

Accordingly, low r values are expected over sea surface due to negligible cross-polarized backscattering, while larger r values are expected over land—depending on coastal morphology, e.g., sand, rocks, vegetation, urban, ice—due to the significant contribution of both co- and cross-polarized backscattering.

To generate a binary image where land and sea are clearly distinguished, a CFAR algorithm is used to obtain a global threshold. CFAR is an adaptive algorithm used in radar systems to detect target returns against a background of clutter [76].

When dealing with r , since it describes the scattering from a first-order Bragg scattering surface, it is expected to be Rayleigh distributed over the sea surface. Hence, according to [69], the relationship between the detection threshold th and the probability of false alarm P_{fa} is given by

$$P_{fa} = \int_{th}^{\infty} \frac{r}{\sigma^2} e^{-\frac{r^2}{2\sigma^2}} dr \quad th = \sigma \sqrt{-2 \ln(P_{fa})}, \quad (32)$$

where σ is the standard deviation of the Rayleigh distribution. According to Eq. (32), for a given P_{fa} , a global threshold th can be obtained. In this test case, a P_{fa} equal to 10^{-6} is used.

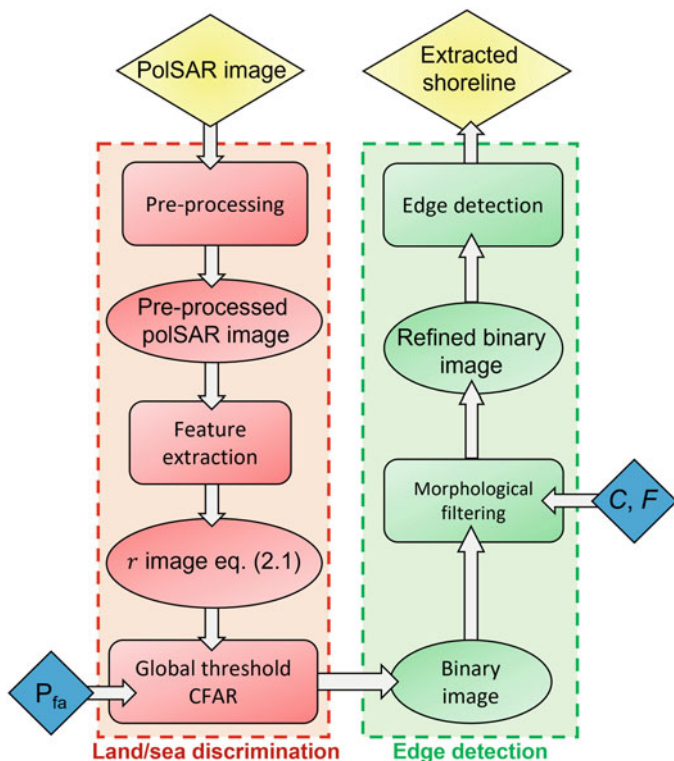


Fig. 12 Block diagram of shoreline extraction

The second step relies on the extraction of continuous coastlines [29], see the edge detection block in Fig. 12. First, the global threshold th obtained from the CFAR approach is used to generate a binary image that separates land from the sea according to the decision rule $r \geq th$.

Then, to refine the binary image by removing artifacts and filling holes, morphological filtering is addressed. The artifacts are removed using an image processing operation called area opening that removes all connected pixels that have fewer than C pixels, while the holes are filled using an image filtering that replicates the pixels inside a hole of $C \times C$ pixels. Finally, to extract the one-pixel continuous coastline from the binary output, the conventional Sobel edge detector is used [2], which is an image processing technique for finding the boundaries of objects within images detecting discontinuities in brightness. Once edges are extracted from the binary output, the one-pixel continuous coastline is obtained.

3.3 Experimental Showcase

In this section, a showcase is discussed to show the benefits of polarimetric methods in monitoring coastal areas. The SAR data set consists of two C-band DP (VV+VH) Interferometric Wide (IW) ESA Sentinel-1 collected over the Calabria coast, in Italy. The area includes the towns of Cetraro Marina and Marina di Belmonte, two municipalities of Cosenza, an Italian town of the region Calabria, on July 11, 2016 and July 3, 2021 in descending mode with an AOI around 46° . Wind conditions, estimated from the cross-polarized channel according to [92], call for low-to-moderate wind (5.1 m/s and 4.7 m/s, respectively). A square pixel whose spacing is 14 m is considered. The false color SAR data collected on July 11, 2016 is ground-projected and shown as an image in Fig. 13a, where red, green, and blue channels stand for VV, VH, and VV/VH power ratios. To reduce the speckle noise in the SAR image, a boxcar filter with a window size 9×9 is applied. The RGB speckle filtered image is shown in Fig. 13b. The metric r (31) is evaluated using $N = 9$, and the output is shown in false color in Fig. 13c. The r image clearly shows that land and sea are well separated, with sea surface exhibiting a very homogeneous behavior in terms of r values. To clearly separate land from the sea, a binary image is generated where the global threshold is obtained using the CFAR approach described in Sect. 3.2 with a probability of false alarm, $P_{fa} = 10^{-6}$. To refine the image, i.e., to fill in the hole and to filter out the isolated pixels, morphological filtering is applied, with $F = 100$ and $C = 9$. The resulting binary image is shown in Fig. 13d, where sea and land are clearly separated. To extract the coastline in a simple and effective way, an edge detection approach based on the Sobel operator is applied on the binary image of Fig. 13d. The extracted coastline, superimposed on the SAR image, is shown in red in Fig. 13e. It can be noted that the coastline extracted well fits the SAR image coastal profile. The same processing flowchart is applied on the SAR scene relevant to July 3, 2021.

To analyze the changes that occurred on the coast from 2016 to 2021, two areas are considered (see the yellow and green boxes of Fig. 13e). Results are shown in Fig. 14, where the waterlines extracted are superimposed to the VV-polarized NRCS SAR image. The yellow and blue lines refer to the coast in 2016 and 2021, respectively, while the white line refers to an area where the overlapping occurs. The first area, enclosed in the yellow box of Fig. 13e, refers to the coast of Cetraro Marina, an area strongly affected by coastal erosion as reported in the European Atlas of the Seas [28]. By visually inspecting Fig. 14a, it can be noted that the extracted coastline rarely calls for white color, witnessing a non-overlapping between coastlines extracted in 2016 and 2021. A comparison between the yellow and blue lines shows a remarkable loss of coastal area, indicating significant coastal erosion occurred in this time frame (2016–2021). The changes in terms of the coastal area between 2016 and 2021 are estimated from the related binary imagery considering the actual pixel spacing. A net erosion of $11.2 \times 10^4 \text{ m}^2$ is estimated, i.e., five years resulted in a degradation loss of an area that is about 21 times larger than a regular American football pitch.

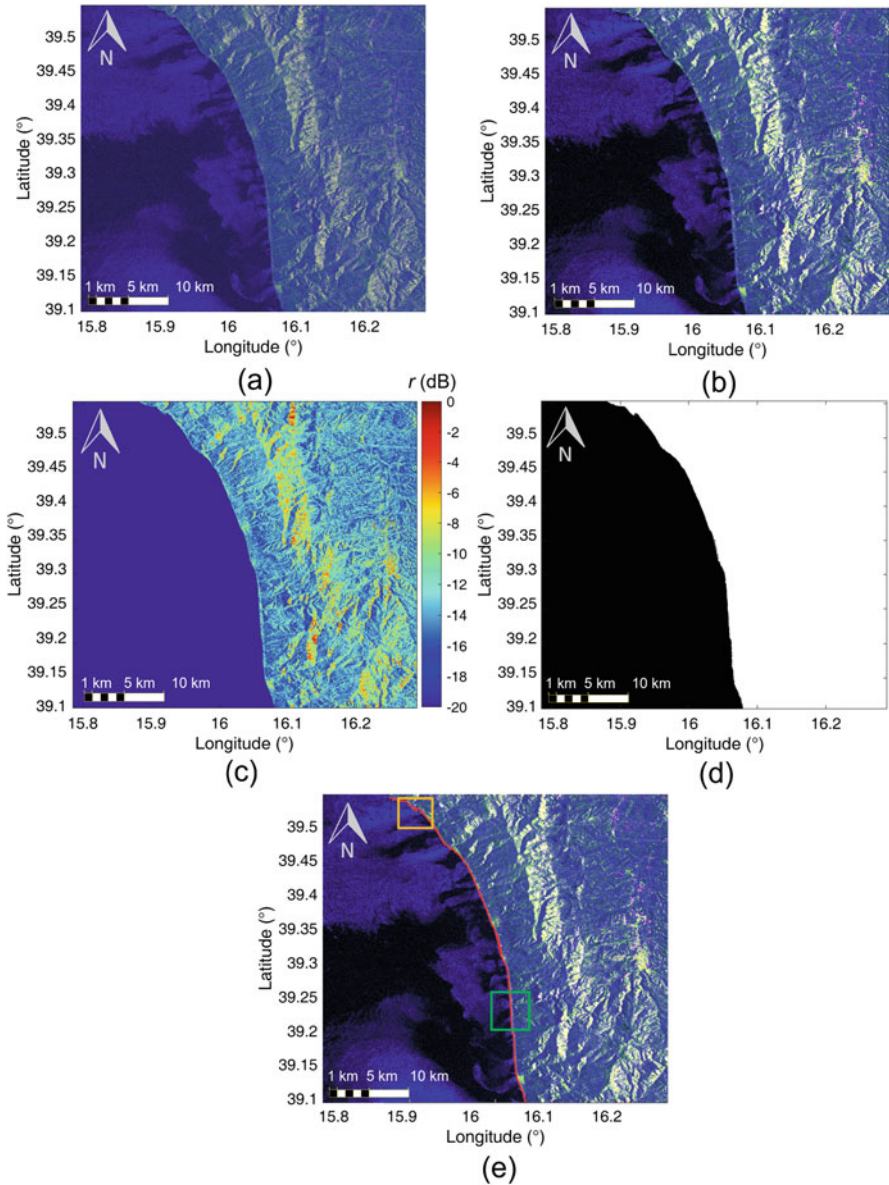


Fig. 13 Excerpt of Sentinel-1 SAR imagery collected over the coast of Calabria, Italy, on 11 July 2016. (a) RGB color composite (R: VV, G: VH, B: VV/VH power ratio) SAR imagery; (b) boxcar filtered RGB image using a window size 9×9 ; (c) r image in false color; (d) refined binary image obtained after the CFAR and the morphological filtering; (e) coastline extracted (in red) superimposed on RGB SAR image

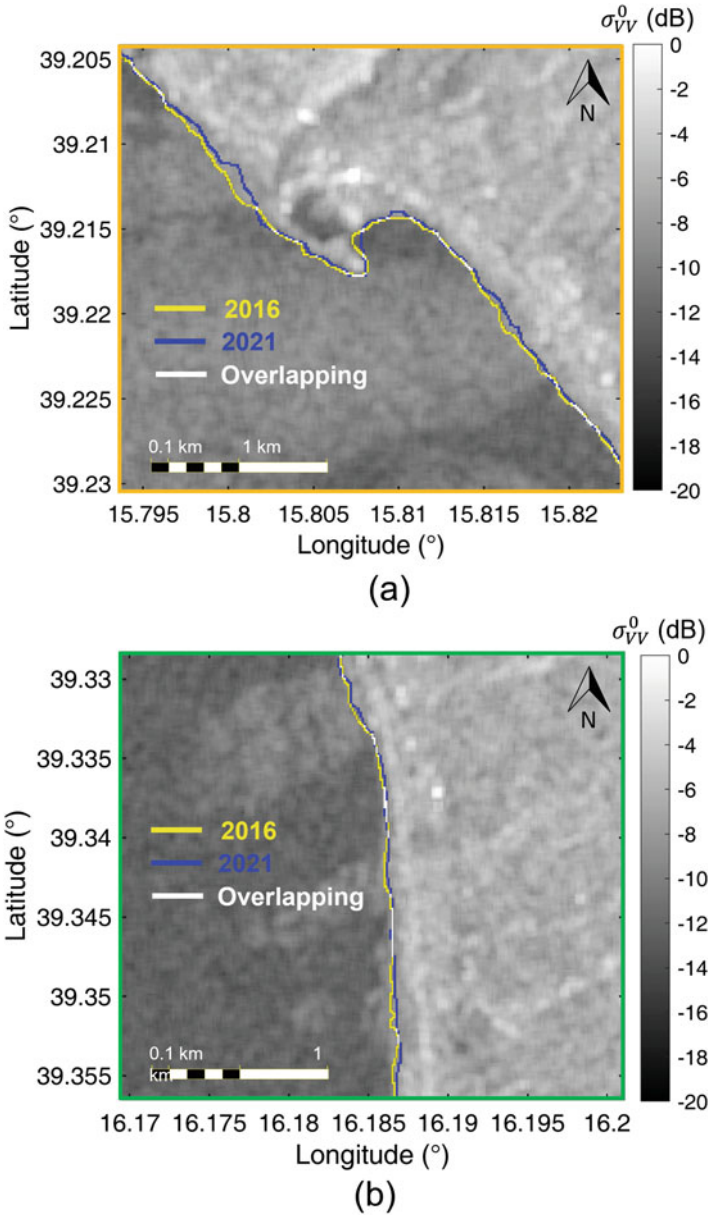


Fig. 14 Coastline extracted related to the SAR scene of: (a) Cetraro Marina (see yellow box of Fig. 13e); and (b) Marina di Belmonte (see green box of Fig. 13e). The coastlines, superimposed on the corresponding VV-polarized NRCS SAR image, are coded as yellow, blue, and white lines for 2016, 2021, and 2016–2021 overlapping, respectively

The second area, enclosed in the green box of Fig. 13e, refers to the coast of Marina di Belmonte, which is less affected by coastal erosion according to [28]. By visually inspecting Fig. 14b, it can be noted that the 2016 and 2021 coastlines are very close to each other, indicating that no significant changes occurred. A net erosion of $4.6 \times 10^4 \text{ m}^2$ is estimated, i.e., about 40% less than the first test area.

4 Conclusions

In this chapter, the added value provided by polSAR satellite measurements in the framework of the monitoring of oceans and coastal areas is presented.

Section 1 deals with basic and advanced concepts of radar polarimetry that lie at the basis of the exploitation of polSAR data. Introduction to modeling of polarimetric sea surface scattering is also theoretically provided along with a meaningful experimental showcase.

Section 2 offers an overview of polarimetric approaches to observe sea oil spills from polSAR imagery. Benefits and drawbacks are critically pointed out. A though experimental showcase is also presented, which is based on the classification of ocean slicks using DL methods.

Section 3 provides the most up-to-date information on the use of polSAR measurements to extract coastal profiles and to monitor their changes over time. An experimental showcase is also discussed to highlight the potential of polSAR imagery to observe coastal areas affected by significant erosion processes.

The key message this chapter would convey to the reader is that a large set of polSAR measurements is nowadays available that can be successfully exploited to produce reliable and effective added-valued products and to develop advanced geophysical parameter estimation algorithms when proper electromagnetic models are considered.

Acknowledgments This chapter was written with the support of the ESA-NRCSS project Dragon 5 Monitoring harsh coastal environments and ocean surveillance using radar remote sensing sensors (ID 57979), the National Key R & D Program of China (2021YFA0715101), the Strategic Priority Research Program of Chinese Academy of Sciences (XDB 41020104), and Scientific Research Project of Beijing Educational Committee (KM202110005024).

The authors would thank the Alaska Satellite Facility Distributed Active Archive Center supported by the National Aeronautics and Space Administration for providing the Alos-PALSAR data. They would also thank ESA to provide free of charge the Sentinel-1 SAR data through the Copernicus Scientific Hub.

References

1. Ainsworth TL, Kelly JP, Lee JS (2009) Classification comparisons between dual-pol, compact polarimetric and quad-pol SAR imagery. *ISPRS Journal of Photogrammetry & Remote Sensing* 64(5):464–471

2. Bachofer F, Qu  n  herv   G, Zwiener T, Maerker M, Hochschild V (2016) Comparative analysis of edge detection techniques for SAR images. *European Journal of Remote Sensing* 49(1):205–224
3. Baghdadi N, Pedreros R, Lenotre N, Dewez T, Paganini M (2007) Impact of polarization and incidence of the ASAR sensor on coastline mapping: example of Gabon. *International Journal of Remote Sensing* 28(17):3841 – 3849
4. Bamler R (2000) Principles of synthetic aperture radar. *Surveys in Geophysics*
5. Bird E (2011) *Coastal Geomorphology: An Introduction*. Wiley
6. Brown CE, Fingas MF, Hawkins R (2003) Synthetic aperture radar sensors: Viable for marine oil spill response?
7. Buono A, Nunziata F, Mascolo L, Migliaccio M (2014) A multipolarization analysis of coastline extraction using X-band COSMO-SkyMed SAR data. *IEEE Journal of Selected Topics on Applied Earth Observation and Remote Sensing* 7(7):2811–2820
8. Buono A, Nunziata F, Migliaccio M (2016) Analysis of full and compact polarimetric SAR features over the sea surface. *IEEE Geoscience and Remote Sensing Letters* 13(10):1527–1531
9. Buono A, Nunziata F, Migliaccio M, Li X (2016) Polarimetric analysis of compact-polarimetry SAR architectures for sea oil slick observation. *IEEE Transactions on Geoscience & Remote Sensing* 54(10):5862–5874
10. Buono A, Nunziata F, de Macedo CR, Velotto D, Migliaccio M (2018) A sensitivity analysis of the standard deviation of the copolarized phase difference for sea oil slick observation. *IEEE Transactions on Geoscience and Remote Sensing* 57(4):2022–2030
11. Buono A, De Macedo CR, Nunziata F, Velotto D, Migliaccio M (2019) Analysis on the effects of SAR imaging parameters and environmental conditions on the standard deviation of the co-polarized phase difference measured over sea surface. *Remote Sensing* 11(1)
12. Buono A, Li Y, Paes RL (2021) Remote sensing of the oceans: Blue economy and marine pollution. *Remote Sensing* 13(8)
13. Carlo, Marangoni (1871) Sul principio della viscosita' superficiale dei liquidi stabilito dal sig. J. Plateau. *Il Nuovo Cimento Series 2*
14. Chen, Jie, Quegan, Shaun (2011) Calibration of spaceborne CTLR compact polarimetric low-frequency SAR using mixed radar calibrators. *IEEE Transactions on Geoscience & Remote Sensing* 49(7):2712–2723
15. Chen G, Li Y, Sun G, Zhang Y (2017) Application of deep networks to oil spill detection using polarimetric synthetic aperture radar images. *Applied Sciences* 7(10):968
16. Cloude S (2009) *Polarisation: applications in remote sensing*. OUP Oxford
17. Cloude SR, Goodenough DG, Chen H (2012) Compact decomposition theory. *IEEE Geoscience and Remote Sensing Letters* 9(1):28–32
18. Collins, M, J, Denbina, M, Atteia, G (2013) On the reconstruction of quad-pol SAR data from compact polarimetry data for ocean target detection. *IEEE Transactions on Geoscience & Remote Sensing* 51(1):591–600
19. Corcione V, Buono A, Nunziata F, Migliaccio M (2021) A sensitivity analysis on the spectral signatures of low-backscattering sea areas in Sentinel-1 SAR images. *Remote Sensing* 13(6)
20. Di Luccio D, Benassai G, Di Paola G, Mucerino L, Buono A, Roskopf CM, Nunziata F, Migliaccio M, Urciuoli A, Montella R (2019) Shoreline rotation analysis of embayed beaches by means of in situ and remote surveys. *Sustainability* 11(3)
21. Di Martino G, Iodice A, Natale A, Riccio D (2016) Polarimetric two-scale two-component model for the retrieval of soil moisture under moderate vegetation via L-band SAR data. *IEEE Transactions on Geoscience and Remote Sensing* 54(4):2470–2491
22. Di Martino G, Iodice A, Monti-Guarnieri A (2020) Acquisition modes. In: Di Martino G, Iodice A (eds) *Maritime Surveillance with Synthetic Aperture Radar*, SciTech Publishing
23. Ding X, Nunziata F, Li X, Migliaccio M (2015) Performance analysis and validation of waterline extraction approaches using single- and dual-polarimetric SAR data. *IEEE Journal of Selected Topics on Applied Earth Observation and Remote Sensing* 8(3):1019 – 1027
24. Elfouhaily TM, Gu  rin CA (2004) A critical survey of approximate scattering wave theories from random rough surfaces. *Waves in random media* 14(4):R1

25. Espeseth MM, Skrunes S, Jones CE, Brekke C, Holt B, Doulgeris AP (2017) Analysis of evolving oil spills in full-polarimetric and hybrid-polarity SAR. *IEEE Transactions on Geoscience and Remote Sensing* 55(7):4190–4210
26. Espeseth MM, Brekke C, Jones CE, Holt B, Freeman A (2020) The impact of system noise in polarimetric SAR imagery on oil spill observations. *IEEE Transactions on Geoscience and Remote Sensing* 58(6):4194–4214
27. Espeseth MM, Jones CE, Holt B, Brekke C, Skrunes S (2020) Oil spill response-oriented information products derived from a rapid-repeat time-series of SAR images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* PP(99):1–1
28. European Commission (2022) European Atlas of the Seas. URL https://ec.europa.eu/maritimeaffairs/atlas/maritime_atlas/
29. Ferrentino E, Nunziata F, Migliaccio M (2017) Full-polarimetric SAR measurements for coastline extraction and coastal area classification. *International Journal of Remote Sensing* 38(23):7405–7421
30. Ferrentino E, Nunziata F, Buono A, Urciuoli A, Migliaccio M (2020) Multipolarization time series of Sentinel-1 SAR imagery to analyze variations of reservoirs water body. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 13:840 – 846
31. Ferrentino E, Buono A, Nunziata F, Marino A, Migliaccio M (2021) On the use of multipolarization satellite SAR data for coastline extraction in harsh coastal environments: The case of Solway Firth. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 14:249–257
32. Franceschetti G, Riccio D, et al. (2002) Fractal models for scattering from natural surfaces. In: *Scattering*, Elsevier, pp 467–485
33. Gil JJ (2000) Characteristic properties of Mueller matrices. *Journal of the Optical Society of America* 17(2):328–334
34. Guissard A (1994) Mueller and Kennaugh matrices in radar polarimetry. *IEEE Transactions on Geoscience and Remote Sensing* 32(3):590–597
35. Guissard A, Sobieski P, Baufays C (1992) A unified approach to bistatic scattering for active and passive remote sensing of rough ocean surfaces. *Trends of Geophysical Research* 1:43–68
36. Hajnsek I, Pottier E, Cloude SR (2003) Inversion of surface parameters from polarimetric SAR. *IEEE Transactions on Geoscience and Remote Sensing* 41(4):727–744
37. Holt B (2004) SAR imaging of the ocean surface. *Synthetic Aperture Radar Marine User's Manual* (25–80)
38. Iodice A, Natale A, Riccio D (2011) Retrieval of soil surface parameters via a polarimetric two-scale model. *IEEE Transactions on Geoscience and Remote Sensing* 49(7):2531–2547
39. Ivonin D, Ivanov AY, Johansson AM, Brekke C (2021) Newly-formed sea ice distinction near the oil platform Prirazlomnaya in the Pechora Sea using polarimetric Radarsat-2 SAR observations. In: *EUSAR 2021*
40. Jie C, Quegan S (2010) Improved estimators of Faraday rotation in spaceborne polarimetric SAR data. *IEEE Geoscience & Remote Sensing Letters* 7(4):846–850
41. Kudryavtsev V, Hauser D, Caudal G, Chapron B (2003) A semiempirical model of the normalized radar cross-section of the sea surface 1. Background model. *Journal of Geophysical Research: Oceans* 108(C3):FET-2
42. Kudryavtsev V, Akimov D, Johannessen J, Chapron B (2005) On radar imaging of current features: 1. Model and comparison with observations. *Journal of Geophysical Research: Oceans* 110(C7)
43. Kudryavtsev VN, Chapron B, Myasoedov AG, Collard F, Johannessen JA (2012) On dual co-polarized SAR measurements of the ocean surface. *IEEE Geoscience and Remote Sensing Letters* 10(4):761–765
44. Kudryavtsev VN, Fan S, Zhang B, Mouche AA, Chapron B (2019) On quad-polarized SAR measurements of the ocean surface. *IEEE Transactions on Geoscience and Remote Sensing* 57(11):8362–8370
45. Kumar LJV, Kishore JK, Rao PK (2014) Decomposition methods for detection of oil spills based on RISAT-1 SAR

46. Laurentiis LD, Jones CE, Holt B, Schiavon G, Frate FD (2020) Deep learning for mineral and biogenic oil slick classification with airborne synthetic aperture radar data. *IEEE Transactions on Geoscience and Remote Sensing* 59(10):8455–8469
47. Lavalle M, Pottier E, Solimini D, Desnos YL (2009) Compact polarimetric SAR interferometry: PALSAR observations and associated reconstruction algorithms. ESA Special Publication
48. Lee JS, Pottier E (2017) Polarimetric radar imaging: from basics to applications. CRC press
49. Li BL, El-Shaarawi AH, Piegorsch WW (2002) Fractal dimensions. *Encyclopedia of Environmental Metrics* 2:821–825
50. Li H, Perrie W, He Y, Lehner S (2013) Target detection on the ocean with the relative phase of compact polarimetry SAR. *IEEE Transactions on Geoscience and Remote Sensing* 51(6 Part1):3299–3305
51. Li H, Perrie W, He Y, Wu J, Luo X (2015) Analysis of the polarimetric SAR scattering properties of oil-covered waters. *IEEE Journal of Selected Topics in Applied Earth Observations & Remote Sensing* 8(8):3751–3759
52. Li Y, Zhang Y, Chen J, Zhang H (2014) Improved compact polarimetric SAR quad-pol reconstruction algorithm for oil spill detection. *IEEE Geoscience and Remote Sensing Letters* 11(6):1139–1142
53. Melshelmer C, Alpers W, Gade M (1998) Investigation of multifrequency/multipolarization radar signatures of rain cells, derived from SIR-C/X-SAR data. In: *International Geoscience & Remote Sensing Symposium*
54. Meng T, Chen KS, Yang X, Nunziata F, Xie D, Buono A (2022) Simulation and analysis of bistatic radar scattering from oil-covered sea surface. *IEEE Transactions on Geoscience and Remote Sensing* 60:1–15
55. Meng T, Yang X, Chen KS, Nunziata F, Xie D, Buono A (2022) Radar backscattering over sea surface oil emulsions: Simulation and observation. *IEEE Transactions on Geoscience and Remote Sensing* 60:1–14
56. Migliaccio M, Nunziata F (2014) On the exploitation of polarimetric SAR data to map damping properties of the Deepwater Horizon oil spill. *International Journal of Remote Sensing* 35(10):3499–3519
57. Migliaccio M, Tranfaglia M (2005) A study on the use of SAR polarimetric data to observe oil spills. In: *IEEE Oceans*
58. Migliaccio M, Ferrara G, Gambardella A, Nunziata F (2006) A new stochastic model for oil spill observation by means of single-look SAR data. In: *IEEE US/EU Baltic International Symposium*, pp 24–29
59. Migliaccio M, Gambardella A, Tranfaglia M (2007) SAR polarimetry to observe oil spills. *IEEE Transactions on Geoscience & Remote Sensing* 45:506–511
60. Migliaccio M, Nunziata F, Gambardella A (2009) On the co-polarized phase difference for oil spill observation. *International Journal of Remote Sensing* 30(6):1587–1602
61. Migliaccio M, Nunziata F, Montuori A, Brown C (2011) Marine added-value products using Radarsat-2 fine quad-polarization. *Canadian Journal of Remote Sensing* 37(5):443–451
62. Migliaccio M, Nunziata F, Li X (2015) Sea oil slick observation using hybrid-polarity SAR architecture. *IEEE Journal of Oceanic Engineering* 40(2):426–440
63. Minchew B, Jones CE, Holt B (2012) Polarimetric analysis of backscatter from the Deepwater Horizon oil spill using L-band synthetic aperture radar. *IEEE Transactions on Geoscience & Remote Sensing* 50(10):3812–3830
64. Moon WM, Staples G, Kim D, Park S, Park K (2010) Radarsat-2 and coastal applications: Surface wind, waterline, and intertidal flat roughness. *Proceedings of the IEEE* 98(5):800 – 815
65. Nord ME, Ainsworth TL, Lee JS, Stacy NJS (2008) Comparison of compact polarimetric synthetic aperture radar modes. *IEEE Transactions on Geoscience and Remote Sensing* 47(1):174–188
66. Nunziata F, Sobieski P, Migliaccio M (2009) The two-scale BPM scattering model for sea biogenic slicks contrast. *IEEE Transactions on Geoscience & Remote Sensing* 47(7):1949–1956

67. Nunziata F, Migliaccio M, Gambardella A (2011) Pedestal height for sea oil slick observation. *IET Radar Sonar & Navigation* 5(2):103–110
68. Nunziata F, Migliaccio M, Li X, Ding X (2014) Coastline extraction using dual-polarimetric COSMO-SkyMed PingPong mode SAR data. *IEEE Geoscience and Remote Sensing Letters* 11(1):104 – 108
69. Nunziata F, Buono A, Migliaccio M, Benassai G (2016) Dual-polarimetric C- and X-band SAR data for coastline extraction. *IEEE Journal of Selected Topics on Applied Earth Observation and Remote Sensing* 9(11):4921 – 4928
70. Nunziata F, de Macedo CR, Buono A, Velotto D, Migliaccio M (2019) On the analysis of a time series of X–band TerraSAR–X SAR imagery over oil seepages. *International Journal of Remote Sensing* 40(9):3623–3646
71. Nunziata F, Li X, Marino A, Shao W, Portabella M, Yang X, Buono A (2021) Microwave satellite measurements for coastal area and extreme weather monitoring. *Remote Sensing* 13(16)
72. Plant W (1990) Bragg scattering of electromagnetic waves from the air/sea interface, surface waves and fluxes, vol. II. *Remote Sensing*, GL Geernaert and WL Plant, Eds, Kluwer Academic 10:978–94
73. Quigley C, Brekke C, Eltoft T (2020) Comparison between dielectric inversion results from synthetic aperture radar co-and quad-polarimetric data via a polarimetric two-scale model. *IEEE Transactions on Geoscience and Remote Sensing* 60:1–18
74. Quigley C, Brekke C, Eltoft T (2020) Retrieval of marine surface slick dielectric properties from Radarsat-2 data via a polarimetric two-scale model. *IEEE Transactions on Geoscience and Remote Sensing* 58(7):5162–5178
75. Raney RK (2007) Hybrid-polarity SAR architecture. In: *IEEE International Conference on Geoscience & Remote Sensing Symposium*
76. Scharf L, Demeure C (1991) *Statistical Signal Processing: Detection, Estimation, and Time Series Analysis*. Addison-Wesley series in electrical and computer engineering, Addison-Wesley Publishing Company
77. Shirvany R, Chabert M, Tournet JY (2012) Ship and oil-spill detection using the degree of polarization in linear and hybrid/compact dual-pol SAR. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 5(3):885–892
78. Skrunes S, Brekke C, Eltoft T (2014) Characterization of marine surface slicks by Radarsat-2 multipolarization features. *IEEE Transactions on Geoscience & Remote Sensing* 52(9):5302–5319
79. Skrunes S, Brekke C, Eltoft T, Kudryavtsev V (2014) Comparing near-coincident C- and X-band SAR acquisitions of marine oil spills. *IEEE Transactions on Geoscience & Remote Sensing* 53(4):1958–1975
80. Song D, Zhen Z, Wang B, Li X, Gao L, Wang N, Xie T, Zhang T (2020) A novel marine oil spillage identification scheme based on convolution neural network feature extraction from fully polarimetric SAR imagery. *IEEE Access* 8:59801–59820
81. Souyris JC, Stacy N, Ainsworth T, Lee JS, Dubois-Fernandez P (2014) SAR compact polarimetry (CP) for Earth observation and planetology: Concept and challenges
82. Stiazznie M, Agnon Y, Shemer L (1991) Fractal dimensions of random water surfaces. *Physica D: Nonlinear Phenomena* 47(3):341–352
83. Trabelsi C, Bilaniuk O, Zhang Y, Serdyuk D, Subramanian S, Santos JF, Mehri S, Rostamzadeh N, Bengio Y, Pal CJ (2017) Deep complex networks
84. Ulaby FT, Long DG, Blackwell WJ, Elachi C, Fung AK, Ruf C, Sarabandi K, Zebker HA, Van Zyl J (2014) *Microwave radar and radiometric remote sensing*, vol 4. University of Michigan Press Ann Arbor, MI, USA
85. Wang W, Fei L, Peng W, Wang J (2010) Oil spill detection from polarimetric SAR image. In: *IEEE 10th International Conference on Signal Processing Proceedings*
86. Yang J, Zhang X (2011) On the ship detection performance with compact polarimetry. *IEEE National Radar Conference - Proceedings*

87. Yerey M, Beaudoin A, Beaudoin J, Walter G (2001) Global shoreline mapping from an airborne polarimetric SAR: Assessment for Radarsat-2 polarimetric modes p 66
88. Yin J, Yang J, Zhou ZS, Song J (2015) The extended Bragg scattering model-based method for ship and oil-spill observation using compact polarimetric SAR. *IEEE Journal of Selected Topics in Applied Earth Observations & Remote Sensing* 8(8):3760–3772
89. Yu L, Zhang Y, Jie C, Migliaccio M, Buono A (2016) Model-based sea surface scattering analysis for the DWH oil spill accident case. In: *In Proceedings of IEEE International Geoscience and Remote Sensing Symposium*
90. Yuanzhi Z, Yu L, Liang X, Jinyu T (2017) Comparison of oil spill classifications using fully and compact polarimetric SAR images. *Applied Sciences* 7(2):193
91. Zhang, Yuanzhi, Li, Yu, Lin, Hui, Chen, Jie (2015) Comparisons of circular transmit and linear receive compact polarimetric SAR features for oil slicks discrimination. *Journal of Sensors*
92. Zhang B, Perrie W (2012) Cross-polarized synthetic aperture radar: A new potential measurement technique for hurricanes. *Bulletin of the American Meteorological Society* 93(4):531 – 541
93. Zhang B, Perrie W, Li X, Pichel WG (2011) Mapping sea surface oil slicks using Radarsat-2 quad-polarization SAR image. *Geophysical Research Letters* 38(10)