POLITEHNICA UNIVERSITY OF BUCHAREST

Doctoral School of Electronics, Telecommunications and Information Technology

Decision No. 970 **from** 08-12-2022

Ph.D. THESIS SUMMARY

Eng. Mihai COCA

METODE HIBRIDE DE INTELIGENȚĂ ARTIFICIALĂ PENTRU EXTRAGEREA INFORMAȚIILOR LATENTE DIN IMAGINI DE TELEDETECȚIE: ANALIZA LA BORD A HAZARDELOR NATURALE

HYBRID ARTIFICIAL INTELLIGENCE METHODS FOR LATENT INFORMATION EXTRACTION FROM REMOTE SENSING IMAGES: ON-BOARD NATURAL HAZARD ANALYSIS

THESIS COMMITTEE

Prof. Dr. Eng. Mihai CIUC Politehnica Univ. of Bucharest	President
Prof. Dr. Eng. Mihai DATCU German Aerospace Center	PhD Supervisor
Prof. Dr. Eng. Cosmin ANCUȚI Politehnica Univ. of Timișoara	Referee
S.R. II Dr. Eng. Corina VĂDUVA Politehnica Univ. of Bucharest	Referee
Assoc. Prof. Dr. Eng. Ștefan-Adrian TOMA Ferdinand I Military Technical Academy	Referee

BUCHAREST 2022

Table of contents

Chapter 1

Introduction

1.1 Presentation of the field of the doctoral thesis

Ongoing global climate change is producing long-term impacts involving various hazards that are rapidly turning into disasters, from longer periods of drought with more frequent wildfires to massive flooding due to rapid melting of upstream snow and ice, or increased duration and intensity of tropical storms [7]. Therefore, in order to effectively minimize the impact of hazards on people and nature, automatic extraction of information from satellite imagery for early identification of a hazard and damage assessment, is essential in critical situations.

1.2 Scope of the doctoral thesis

The thesis aims to design efficient methodologies for extracting and modeling latent information from massive volumes of satellite data, with the purpose of automatically assessing and detecting disastrous events on the Earth's surface.

1.3 Content of the doctoral thesis

Chapter 2 presents an introduction to the fundamental field of remote sensing. *Chapter 3* describes a solution based on data compression via the Normalized Compression Distance (NCD) metric to detect changes produced by floods in Synthetic Aperture Radar (SAR) image time series. *Chapter 4* introduces two statistical deep learning approaches for detecting wildfire affected areas by anomalous deviations from the normal probability distribution of the background data in satellite multispectral scenes. *Chapter 5* outlines a statistical anomaly detection model developed and optimized for an Field Programmable Gate Arrays (FPGA) device, suitable for on-board integration and immediate detection of wildfire outbreak in satellite multispectral scenes. *Chapter 6* covers thesis contributions, list of published articles and future perspectives.

Chapter 2

Basics in remote sensing

The concept of remote sensing refers to any instrument used to observe the state and condition of an object, from a distance, situated on the Earth' surface. The radiant energy generated by the reflection of an incident energy, coming from the Sun or merely artificially generated, is recorded by on-board sensors, stored and transmitted to the ground station for processing image data. Sensor systems are classified according to how they detect electromagnetic energy, i.e., passive sensors which sense radiations generated by external sources, and active sensors that transmit their own energy to the target and then capture the reflection [7]. Satellite sensors are designed to capture data in different parts of the electromagnetic spectrum under different system constraints, i.e., spectral resolution, temporal resolution, spatial resolution and radiometric resolution.

The first satellite for detailed, high-resolution observation of the Earth's entire surface was Landsat 1, launched in 1972, as part of the Landsat programme, which still continues to provide reliable and uninterrupted remote sensing data, essentially as a reference for evaluating long-term changes in the Earth's terrestrial environment [37]. Sentinel-2 is an European multispectral imaging mission comprising twin satellites, flying in the same orbit, with temporal resolution of 5 days at the equator. The Multispectral Instrument (MSI) placed on-board Sentinel-2 platform samples 13 spectral bands [21]. TerraSAR-X is a German Earth observation satellite, with a main payload X-band radar sensor, which records SAR images with a resolution of up to 1 m, regardless of weather conditions, cloud cover or lack of daylight. Applications targeted by this mission include change detection, surface movement monitoring and emergency response information, among others [13].

Various materials on the Earth's surface reflect and emit energy in different parts of the spectrum [21]. The information in all spectral bands can be placed in RGB color channels to reveal different kinds of information about a scene, also there convenient combinations to extract particular information, e.g., moisture levels in vegetation by combining Near InfraRed and Short-Wave InfraRed bands in Figure 2.1, where high values mean wetter vegetation, while low values mean that plants are stressed due to insufficient moisture. Although optical imagery has tremendous potential for mapping,

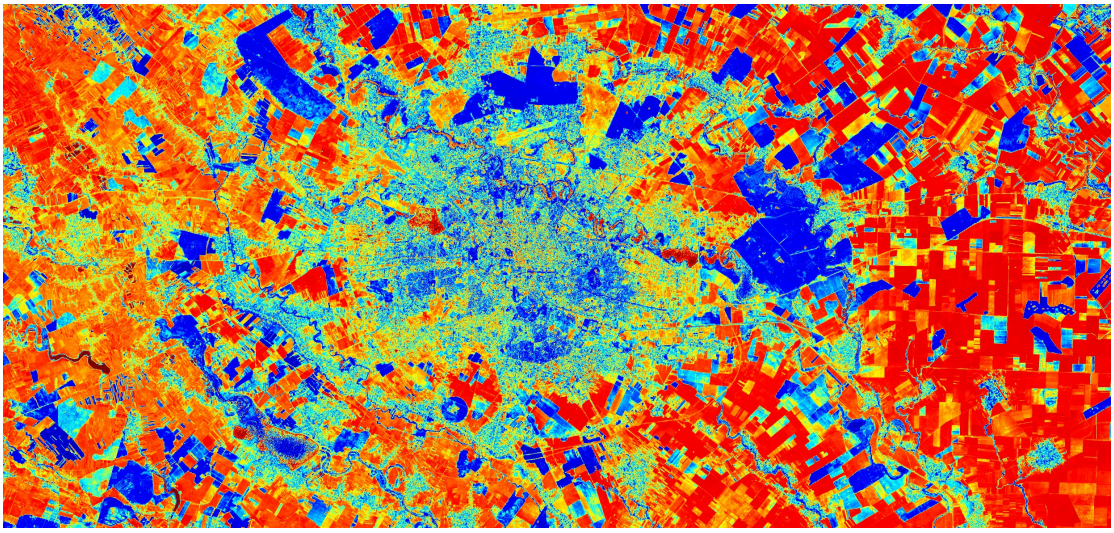


Fig. 2.1 Bucharest, Romania. Convenient combinations to display vegetation moisture index using Sentinel-2 bands.

with improved analysis capability and high image accuracy, under challenging weather conditions the adoption of SAR sensors has shown significant potential in mapping cloud-covered areas.

2.1 Overview of hazards

There are multiple types of hazards, caused naturally or by human intervention. Large scale hazards, detectable from satellites [29] are listed:

- *Floods* are the most destructive and expensive natural disasters, with the highest negative effect on people compared to other natural disasters.
- *Forest Fires* are one of the world's most devastating natural hazards, contributing to global warming, economic losses and ultimately loss of life.
- *Volcanic Eruptions* generate effects ranging from covering the landscape with volcanic ash to gases injected into the atmosphere, leading to climate change.
- *Earthquakes* strike without warning and inflict substantial damage in a short time.
- *Typhoon* is a strong tropical cyclone with large rotating thunderstorms blowing at over 100 km/h with unpredictable trajectories.
- *Landslides* occurs in mountainous regions as a result of severe storms, volcanic eruptions or earthquakes.
- *Oil Spills* are caused every year through deliberate illegal activities or accidents such as the damage of oil pipelines.

A single satellite can miss a natural disaster as it happens, but a heterogeneous constellation of multiple satellites already in orbit can drastically reduce the revisiting period and provide images from any location at any time. The International Chapter Space and Major Disasters [29] service and Copernicus Emergency Management Service (EMS) [12] provide access to satellite imagery during major disasters, which, combined with intelligent detection solutions, generates a major benefit in disaster control.

2.2 Overview of change detection

With massive amounts of data available, simultaneously with increasing computational demands, there is a need to automatically compare two satellite images and determine the locations of changes in a way that involves less time, fewer computational steps and less human interaction. Change detection (CD) solutions extract time-dependent information from targets of interest in the same geographical area by using multi-temporal images. These solutions are used in various fields, such as agricultural surveys, forest monitoring, hydrological studies or disaster assessment [3].

Time series of SAR images are widely used in CD methodologies, usually by calculating the pixel-by-pixel ratio of two consecutive SAR images [39]. The resulted difference image requires the selection of an optimal threshold value to separate the changed from unchanged pixels [54]. The high learning capacity of Deep Neural Networks (DNN) has enabled the use of stacked autoencoders (SAE) [42], or symmetric models [56] in the context of CD in SAR images. In real scenarios, there are situations where the pre-event image in the time series is unusable or the acquisition is too old, making it impossible for a classical CD algorithm to identify a possible change at ground level. Thus, changes should only be identified from the post-event image, where deviant samples are interpreted as anomalies.

2.3 Overview of anomaly detection

Anomalies are patterns in data that do not correspond to a robust defined perception of normal behavior. The definition of data normality is closely related to the domain under analysis, from *a priori* known geometric patterns on a production line to land cover classes commonly found on the land surface. Abnormality in the data can appear at the sensor level with artifacts, due to measurement variability, momentary errors or noise, or in the observed scene, semantic anomalies, due to objects deviating from the whole scene.

Classical concepts for unsupervised anomaly detection (AD), extended across multiple disciplines (e.g., statistics, medicine, engineering, natural sciences), consists of Principal Component Analysis (PCA), nearest neighbor algorithms, the One-Class Support Vector Machine (OCSVM), and Support Vector Data Description (SVDD), among

others. Because of the labor intensive and nearly impossible work of labeling data in order to characterize all anomalies in real scenarios, a dominant solution for detecting anomalies is to learn a model from normal data distribution and try to deflect anomalies.

2.3.1 Ground-based applications

Satellite data processing is mostly done on the ground, with delayed feedback in applications due to the progressive way of data collection. In remote sensing (RS) scenes, anomaly detection algorithms address a discovery task of small-scale portions that do not harmonize with the background, generally defined by the classes normally found on the Earth' surface. The effects produced by natural disasters are anomalies on the Earth's surface, with high-scale damaging potential.

Event detection in Earth science is often critical for immediately addressing negative impacts on natural resources, e.g., drought-related vegetation disturbances [48], devastating floods [10], active fire detection [43]. Wildfire is an extreme natural hazard that caused serious damages in human safety and natural ecosystems in recent years [11].

2.3.2 Space-based applications

In harsh environments, space-based missions must process huge amounts of information with limited resources. Automatic data processing near the sensor [20], immediately after acquisition, can reduce the information flow to ground stations and detect anomalous ground-level events early. Small-satellite missions embed Field-Programmable Gate Arrays (FPGAs) chips with programmatic flexibility, parallel processing and energy efficiency. System-on-chip (SoC) hybrid devices [52] include a processing unit and an FPGA on a single chip, implementing common space application on a single device.

PhiSat-1 nanosatellite mission [22] was the first Earth observation satellite with artificial intelligence (AI) on-board, used to maximize the relevant information to be downlinked by the Ground Segment [27]. The PhiSat system used an on-board Convolutional Neural Network (CNN) model to detect clouds in the images, followed by filtering and discarding unusable images and delivering only operable data [26].

On-board image processing requires reduced dimensions and limited power consumption for hardware components. FPGA device outperform Graphics Processing Unit (GPU) device in terms of power consumption with little sacrifice in performance. The FPGA dynamic reconfiguration capability, which permits dynamic adaptation of the specific problem to changing conditions, is an attractive specific feature for on-board computing.

Chapter 3

Unbiased seamless SAR image change detection based on NCD

3.1 NCD metric for change detection

This chapter proposes a methodology for detecting flood changes in satellite time series using data compression. Normalized Compression Distance (NCD) is a universal parameter-free metric successfully applied as a similarity measure for unstructured data. NCD is the computable form of Normalized Information Distance (NID) [2], which is based on Kolmogorov complexity [34]. In information theory, the Kolmogorov complexity of an object is the minimum size of a computational program that produces the same object at output. Li *et al.* [33] proposed a suitable approximation for the NID by replacing the Kolmogorov operator with the lower bound of a real compressor, e.g., bzlib, zlib. Two objects are considered close if a process can meaningfully compress one of them given the information in the other, i.e., if two slices of data are more similar, then one slice can be described more succinctly using the other [9]. For two objects x and y, the NCD metric defined as

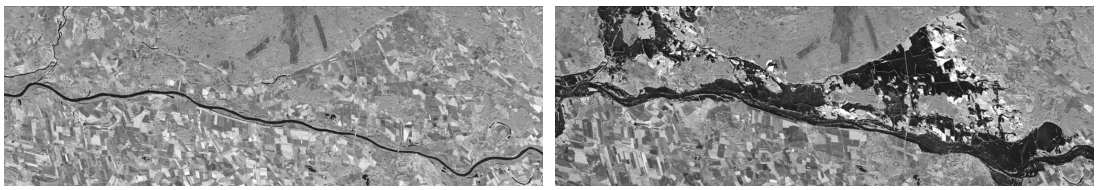
$$NCD(x,y) = \frac{C(x,y) - min(C(x),C(y))}{max(C(x),C(y))}, \tag{3.1}$$

where $C(x,y)$ is the compression dimension of the concatenation of x and y, $C(x)$ and $C(y)$ are the compressed dimension of x and y respectively. The NCD metric can be adopted to compute similarities between any two objects, exploiting the intrinsic power of data compression to fit recurrent patterns. The value of $NCD(x,y)$ is a positive number, i.e., $0 < NCD(x,y) < 1+e$, which represents how different the two objects are. The error e occurs for two reasons, imperfections in the compression technique and the expected effect of too noisy content of one object relative to the other [5].

CompLearn [8] is a library that integrates compression techniques into the process of discovering and learning patterns. This library can compute an NCD matrix for

(a) (b)

Fig. 3.1 Sendai dataset with amplitude SAR images, pre-event image (a) and post-event image (b), acquired by TerraSAR-X satellite. Sendai1024 subset in 1024x1024 pixel sub-images.



(a) (b)

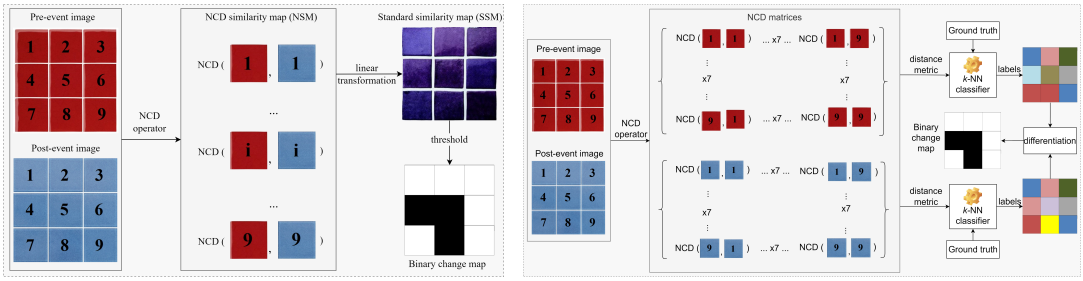
Fig. 3.2 Stendal dataset with amplitude SAR images, pre-event image (a) and the post-event image (b), acquired by TerraSAR-X satellite.

a collection of heterogeneous samples with different compressor variants. A faster approach has been proposed in [45] using a compressor optimized for running on the GPU, achieving an increase in compression speed with a small decrease in accuracy.

This chapter presents two methods that compute a difference image (DI) from the time series of two SAR images using the NCD operator (3.1) with the bzlib algorithm for image compression. Further, the DI is used to create a binary change map that highlights the flood affected areas in a scene. It is worth mentioning that all the methods proposed in this thesis work at patch level.

3.2 Study areas and Datasets

In the context of hazard assessment, time series contain a first image acquired before a natural hazard (pre-event) and a second image acquired immediately after the hazard (post-event). Two scenes containing the devastating effects of a tsunami (Figure 3.1) in a shoreline part of Sendai, Tohoku region, Japan, respectively the effects of a flood caused by the overflow of the Elbe river (Figure 3.2) in the Stendal district, Germany. Both datasets, Sendai and Stendal, were first studied in [16] using an interactive, automatic and fast processing chain tool, applicable to large and complex datasets. The input data



(a) (b)

Fig. 3.3 Methodologies for unsupervised change detection (a) and supervised change detection (b).

was in the form of collections of sub-images, i.e., patches, each generated from an initial scene that was divided into sub-images via a non-overlapping sliding quadratic window.

3.3 Unsupervised method for change detection

In the proposed unsupervised method, the NCD operator (3.1) was used to generate a similarity map of a series of pairs of samples, i.e., NCD similarity map (NSM). Next, a linear transformation was applied on NSM to fit values in the range [0, 255]. Further, the standard similarity map (SSM) was used to generate a binary change map based on an unsupervised threshold determination (Figure 3.3a).

In this approach, the NCD metric computes the degree of similarity between two patches from the same geographic location but at different times. Starting from the SSM histogram values, three algorithms were implied to estimate an optimal threshold, namely, Otsu's method [41], Kittler's method [31] and Kapur's method [30].

3.3.1 Experimental results

To test the proposed unsupervised method on the Sendai1024 subset, the dimension of patches was set to 64x64 pixels. The size 64x64 pixels is a lower bound that provides space for pattern identification for data compression purposes.

Figure 3.4 illustrates the ranges of NCD metric results for different patch sizes, e.g., 8x8, 16x16, 32x32, and 64x64, based on filtering iteration in the Sendai1024 subset. Inside the range of NCD values, between minim and maxim values, were highlighted the estimated thresholds obtained through the 3 above mentioned thresholding algorithms, applied to the SSM. The Frost algorithm [23] was used to filter the speckle effect while preserving edges in SAR images, in order to test the assumption that NCD operator is noise-resistant [5].

In Table 3.1, the best result obtained on the Sendai1024 subset on different patch sizes was given by the Kitter's threshold algorithm with a good error rate value, i.e., 4.20%, and a detection accuracy of 95.80% for 32x32 patch size, respectively the Otsu's

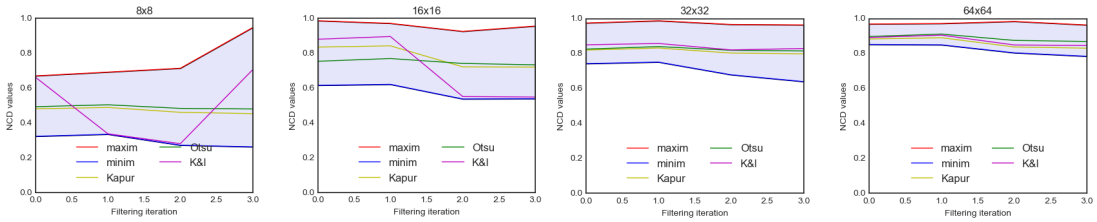


Fig. 3.4 Range of NCD values, minimum and maximum, and threshold values of Otsu, Kapur and Kittler algorithms for differentiating between changed and unchanged in Sendai1024 subset.

threshold method with a very good error rate value, i.e., 0.78%, and a very accurate detection of 99.22% for 64x64 patch size. The fact that the best results were obtained when no filtering iteration was applied validated that the NCD operator is resistant to the speckle effect.

3.4 Supervised method for change detection

In the proposed supervised method, the NCD operator (3.1) was used to generate an NCD matrix containing relational information from the entire patch collection inside a scene. The NCD matrix information as distance information was used as input in a modified version of k-NN classification algorithm to classify the samples. Further, a simple differentiation algorithm on annotations, applied to the results of the two initial scenes, produced a change map (Figure 3.3b).

3.4.1 Experimental results

The Sendai dataset has been assigned with a specific nomenclature of classes, namely 10 classes for the pre-event image and 12 classes for the post-event image. In the supervised approach, the input patch size was set to 160x160 pixels [17]. Considering a non-overlapping window of 160x160 pixels, resulted a collection of 5673 patches per image for the Sendai dataset.

A ground truth map for Sendai dataset, used in the classification evaluation was determined in [18] and manually refined in some cases. The k-NN classifier was trained at a rate of 40% of the entire Sendai sample collection. The hyperparameter k in the

Table 3.1 Results for unsupervised CD method on Sendai1024 subset.

Thresholding algorithm	**Filtering iterations**	**Patch dimension**							
		32x32				**64x64**			
		Error	**Accuracy**	**False-positive**	**T**	**Error**	**Accuracy**	**False-positive**	**T**
Kapur	0	11.04%	88.96%	10.31%	85	9.77%	90.23%	10.17%	68
Otsu		9.28%	90.72%	8.35%	91	**0.78%**	**99.22%**	**0%**	**99**
Kitter		**4.20%**	**95.80%**	**2.37%**	**118**	1.17%	98.83%	0.42%	87

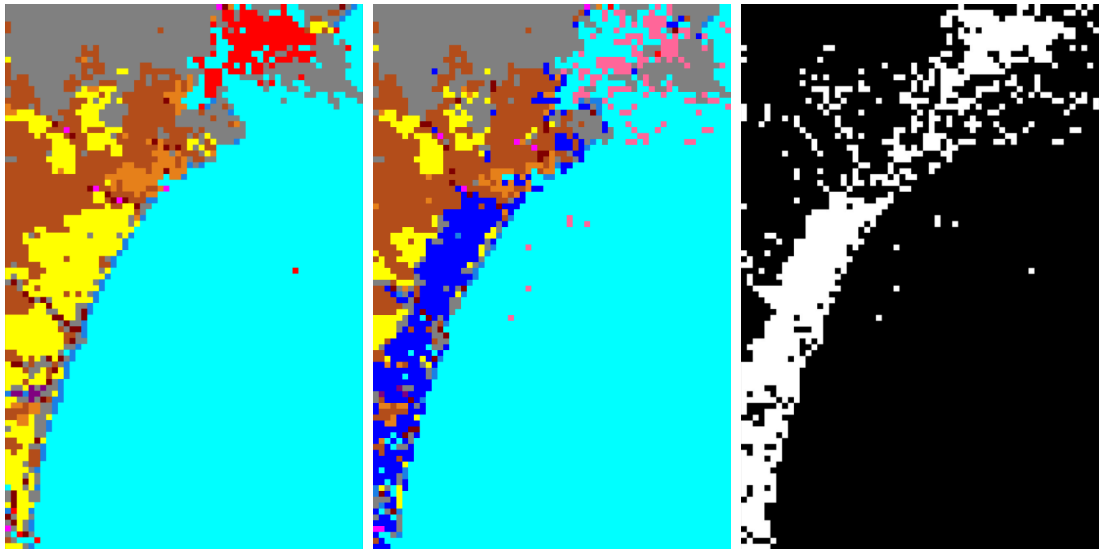


Fig. 3.5 Classification results on Sendai dataset. Pre-event image classification map (a), post-event image classification map (b), binary change map (c), and the colour legend associated with labels.

k-NN algorithm was set to 11 for pre-event classification, respectively 13 for post-event classification (Figure 3.5).

The tsunami event generated some class transitions, namely from the *Aquaculture* class to the *Debris* class, while the coastal *Agriculture* class was mostly covered by the *Flooded* area class. Resulted an overall accuracy of 95.06% for the pre-event image and 92.97% for the post-event image. In the case of the post-event image, the accuracy was poorer because the inter-class variance for flood and debris classes was small. The binary change map highlighted in Figure 3.5c was obtained by differencing the two classification maps shown in Figure 3.5a and Figure 3.5b. Finally, the resulting binary change map identified with an accuracy of 90.7% the areas that suffered changes due to natural hazard effects.

3.5 Contributions

In this chapter, two methods have been proposed that rely on data compression to extract distance information between all samples of a SAR scene. The distance information was used to compute an automatic threshold, respectively to train a k-NN instance in order to generate a binary change map.

Chapter 4

Hybrid DNN in Anomaly Detection

4.1 Self-supervised learning

Self-supervised learning is a promising example for learning convenient representations without costly labels, e.g., predictive self-supervised methods that learn beneficial image representation of the data using auto-generated labels. In remote sensing imagery, discriminative features are extracted by focusing on both object shapes and texture information [49]. This section proposes a hybrid anomaly detection method, a Deep Convolutional Neural Network (DCNN) component combined with a classical Dirichlet distribution model, which evaluates the relative anomalousness in Sentinel-2 scenes.

4.1.1 Current context

Satellite images usually incorporate many semantic variations, and lately different variants have been defined to detect contextual anomalies, multivariate statistics by maximizing the representation gap between in-domain and out-of-distribution samples [24], reconstruction-based methods [51], low-dimensional embedding followed by OCSVM in wildfire assessment [11].

4.1.2 Methodology

Anomaly detection (AD) approaches integrate two components, a statistic that has a known feedback to normal data distribution and a decision rule that inspects the statistic. An anomaly detection method based on ranking is proposed in this chapter (Figure 4.1), aiming to detect anomalies from a single observation, i.e., the post-event image.

Numerical transformations

The image preprocessing stage is done according to *PreprocessData* function defined in Algorithm 1, standardization or max-normalization. In the transformation phase, a set of discrete geometric transformations was applied on each normal sample, $F_k =$

Algorithm 1: Anomaly detection using numerical transformations

Input: F_k: a series of numerical transformations, g_θ: multinomial classification model

Output: *scores*: Anomaly scores

Data: I_{train}: scene with normal data, I_{test}: scene with normal and abnormal data

Function *PreprocessData*$(I,\ t)$**:**

- $(W,H,C) \leftarrow size(I)$
- $I \leftarrow reshape(I, W*H, C)$
- **if** $t = 1$ **then**
  - $\{mean_j\}_{j=0}^{C-1} \leftarrow \{\frac{1}{W*H}\sum_{i=0}^{W*H-1} I_{i,j}\}$
  - $\{stddev_j\}_{j=0}^{C-1} \leftarrow \{\sqrt{\frac{\sum_{i=0}^{W*H-1}(I_{i,j}-mean_j)}{W*H-1}}\}$
  - $\{Y_{i,j}\}_{i,j=0}^{W*H-1,C-1} \leftarrow \{\frac{I_{i,j}-mean_j}{stddev_j}\}$
- **else**
  - $\{Y_{i,j}\}_{i,j=0}^{W*H-1,C-1} \leftarrow \{2*(I_{i,j}/2^{16})-1\}$
- $Y \leftarrow reshape(Y, W, H, C)$
- **return** $\{NonOverlappingPatching(Y)\}$

Function *AnomalyScore*$(F_k,\ g_\theta,\ I_{train},\ I_{test})$**:**

- $X_n \leftarrow PreprocessData(I_{train}, 1)$
- $X_{an} \leftarrow PreprocessData(I_{test}, 1)$
- $X_t \leftarrow \{(F_j(x), j), x \in X_n, 0 \leq j < k\}$
- **while** *not converged* **do**
  - Train g_θ on labeled dataset X_t
- $i \leftarrow 0$
- $n \leftarrow |X_n|$
- $|scores| \leftarrow |X_{an}|$
- $scores \leftarrow 0$
- **while** $i < k$ **do**
  - $X_{i,train} \leftarrow \{softmax(g_\theta(F_i(x))), x \in X_n\}$
  - $X_{i,test} \leftarrow \{softmax(g_\theta(F_i(x))), x \in X_{an}\}$
  - $m \leftarrow 0$
  - $p \leftarrow 0$
  - **foreach** $x_j \in X_{i,train}$ **do**
    - $m \leftarrow m + x_j$
    - $p \leftarrow p + \log x_j$
  - $\bar{m} \leftarrow \frac{1}{n} m$
  - $\bar{p} \leftarrow \frac{1}{n} p$
  - $\alpha_{0,i} \leftarrow \bar{m}\frac{n(k-1)(-\Psi(1))}{n\sum \bar{m}\log(\bar{m}) - \sum \bar{m} p}$
  - **while** *not converged* **do**
    - $\bar{\alpha}_{0,i} \leftarrow \Psi^{-1}(\Psi(\sum \alpha_{0,i}) + \bar{p})$
  - $scores \leftarrow \{scores + \frac{1}{k}\sum_{j=0}^{k-1}(\bar{\alpha}_{0,i} - 1) * \log x_j, x_j \in X_{i,test}\}$
  - $i \leftarrow i + 1$
- **return** *scores*

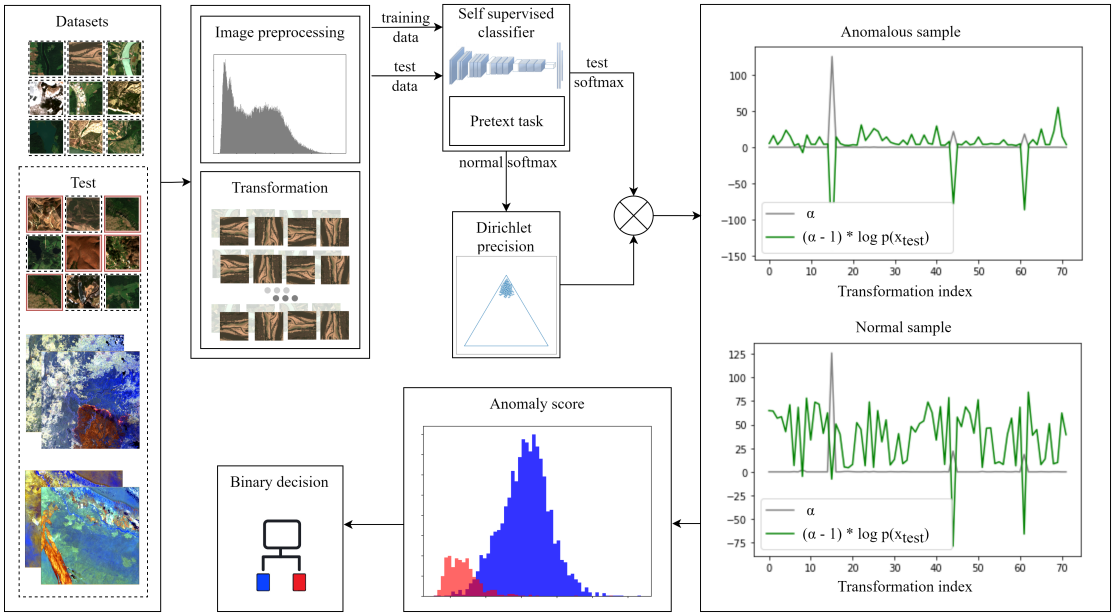


Fig. 4.1 General architecture of the anomaly detection method based on ranking. The data processing module, blocks *Image preprocessing* and *Transformation*, takes a Sentinel-2 Level-2A image as input and generates a collection of scaled and geometrical transformed non-overlapping patches. The self-supervised classifier is trained to differentiate between k geometrical transformations employed in block *Transformation*. A ranking estimation module takes a product between a softmax response vector and a Dirichlet precision computed for each transformation involved and outputs an anomaly score that is finally used in a binary decision.

$\{f,\ f(x,\ flip,\ tx,\ ty,\ rk) : X \to X^i, i \in [0,\ k-1],\ flip \in \{false,\ true\},\ (tx,\ ty) \in \mathbb{Z},\ rk \in \{0,\ 1,\ 2,\ 3\}\}$, where terms $flip$, $(tx,\ ty)$, rk parametrize common geometric transformations like flip, translation, respectively rotation. Parameter rk represents the number of counter-clockwise rotations with 90°. A transformed image is generated by applying three transformations at a time. Further, the ranking function is defined as

$$r_c(x) = \sum_{i=0}^{k-1} \log p(softmax(g_\theta(F_i(x)))|F_i), \tag{4.1}$$

where a combined log-likelihood of inference result $\log p(y|F_k)$, conditioned on each of the adapted transformation in F_k, under an estimated distribution of normal samples is implied.

Dirichlet Distribution

In the interest of its tractable analytic properties, Dirichlet distribution is preferred as a prior distribution in Bayesian learning [25]. A m-dimensional Dirichlet random variable $p = (p_1, ..., p_m)$ describes a distribution over m classes and takes values in the $(m-1)$-simplex if $p_k \geq 0, 1 \leq k \leq m$ and $\sum_{k=1}^{m} p_k = 1$. The random variable p has the following probability density (4.2) on $(m-1)$-simplex

$$Dir(p|\alpha) = \frac{\Gamma(\alpha_0)}{\prod_{k=1}^{m}\Gamma(\alpha_k)}\prod_{k=1}^{m} p_k^{\alpha_k-1}, \alpha_0 = \sum_{k=1}^{m}\alpha_k, \tag{4.2}$$

where the parameter α is an ensemble of concentration components $\alpha_k > 0$, α_0 is the precision of the distribution, p is a vector of probabilities and $\Gamma(\cdot)$ is the Gamma function. $Dir(\alpha)$ variation over a $(m-1)$-simplex of admissible values of p for a given input α, projects conditional distribution over normal samples. The parameter vector α can be approximated from a training collection of observed multinomial data, $D = \{\mathbf{p_1}, \mathbf{p_2}, ..., \mathbf{p_N}\}$ by maximizing the log-likelihood function of D, which is given by

$$G(\alpha) = N(\log\Gamma(\sum_k \alpha_k) - \sum_k \log\Gamma(\alpha_k) + \frac{1}{N}\sum_k(\alpha_k - 1)\sum_i \log p_{ik}). \tag{4.3}$$

Fixed Point Iteration method [38] provides a convergent fixed point iteration procedure for estimating parameters α_k. The initial guess for α is based on Maximum Likelihood Approximation (MLA) method [50]. The maximization of a lower bound on the likelihood, which is tight at α, determines the new prediction (4.4) as

$$\alpha_k^{new} = \Psi^{-1}(\Psi(\sum_k \alpha_k^{old}) + \frac{1}{N}\sum_i \log p_{ik}), \tag{4.4}$$

where $\Psi = \frac{d\log\Gamma(x)}{dx}$ is the digamma function which can be inverted by using a Newton-Raphson update procedure.

In Algorithm 1, line 33, is used MLA to compute the precision of the Dirichlet distribution, i.e., α_0. Digamma function $\Psi(\cdot)$ is the first derivative of the log Gamma function and it is used to converge to the maximum, the only stationary point in Dirichlet. The inverse of the digamma function is calculated numerically (lines 34-35) using five Newton-Rapshon iterations. In (4.3), first two elements can be reduced as they are invariant relative to $\mathbf{p_k}$, resulting in a rewriting of (4.1) as

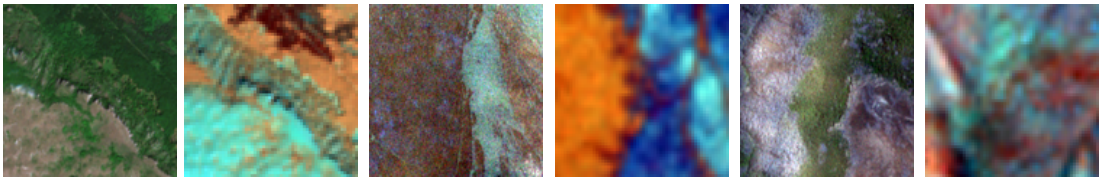
$$r_c(x) = \sum_{i=0}^{k-1}(\bar{\alpha}_i - 1)\log(softmax(g_\theta(F_i(x)))). \tag{4.5}$$

4.1.3 Study areas and Datasets

Fire is a recurring part of nature with good things and bad aspects. In 2019, numerous alarming wildfires have been triggered causing unprecedented damage.

BigEarthNet dataset

BigEarthNet [46], [47] is a benchmark archive, consisting of 590,326 Sentinel-2 patches, acquired between June 2017 and May 2018 over 10 countries of Europe. BigEarthNet is a multi-label dataset with 43 imbalanced labels. A subset of all patches containing at



(a) (b) (c) (d) (e) (f)

Fig. 4.2 Three pairs of two images each, from left to right, BigEarthNet test (first), Robore (second) and Sydney (third) datasets.

least the labels *Mixed forest, Coniferous forest, Broad-leaved forest* and *Pastures* from original BigEarthNet, containing 2570 samples, was used to create a training dataset (i.e., random 2400 samples or 93.4%) and a test dataset (i.e., the rest of 170 samples or the rest of 6.6%). In above mentioned test dataset, another 29 specific samples containing at least the labels *Mixed forest, Broad-leaved forest* and *Burned areas*, filtered out from entire BigEarthNet dataset, were added to result *BigEarthNet test* subset with 199 samples. The training dataset, i.e., *BigEarthNet train*, was used both for the training phase and as observed data to estimate the Dirichlet distribution parameters. In Figures 4.2a and 4.2b show a BigEarthNet patch in two different visualizations, containing burned areas highlighted in red-burgundy. The ground truth map was estimated from the label collection provided for each sample, where all samples that contained label *Burned areas* were considered anomalous.

Burned area datasets

Each created bellow dataset is a subarea of 72 km x 72 km generated from a Sentinel-2A product, with pre- and post-event images to generate a ground truth map through classical spectral indices, namely, the Normalized Burn Ratio (NBR), the Normalized Difference Water Index (NDWI), the Relativized Burn Ratio (RBR). Each dataset was tiled in non-overlapping patches, using a size of 120x120 pixels.

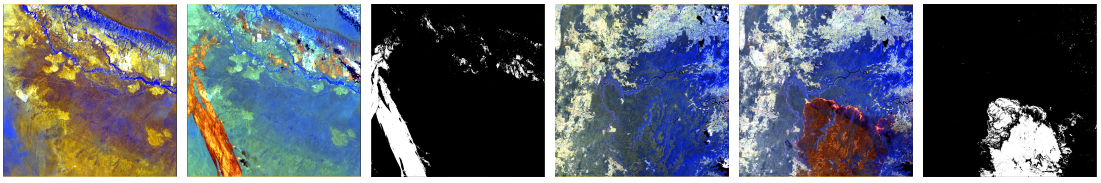
Robore

The original scene swept a south-east area of Robore town, Bolivia, South America. Figures 4.2c and 4.2d depict an anomalous sample in true-color composite (TC) and false-color (FC) composite, with the burned area highlighted in red-orange.

Sydney

The original scene swept a south-west area of Sydney town (Wollongong), New South Wales, Australia. Figures 4.2e and 4.2f exhibit an anomalous sample in TC composite and FC composite, with the burned area highlighted in light red and red-orange.

The computed RBR value range was $[-4.23, 0.93]$ (Robore), respectively $[-87.65, 0.98]$ (Sydney), and based on Otsu's thresholding method [41], resulted a threshold of 0.2596,



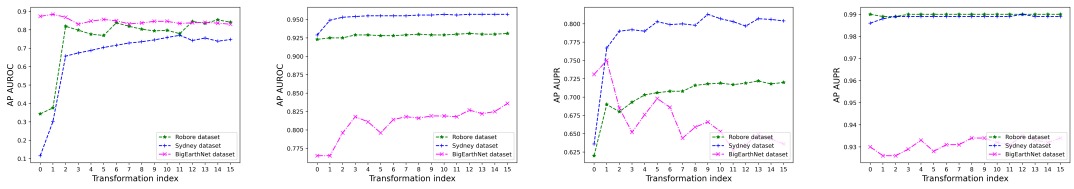
(a) (b) (c) (d) (e) (f)

Fig. 4.3 Two sets of three images each, from left to right, Robore scene (first), Sydney scene (second).

respectively 0.2725 on the RBR histogram. The binary threshold condition in the RBR histogram yields a pixel-level truth map (Figures 4.3c and 4.3f), used to achieve a patch-level truth map through a simple condition: if more than 50% pixels are anomalous, the entire patch is considered anomalous.

4.1.4 Experiments

The proposed architecture was implemented in the open-source Keras deep learning framework with a Tensorflow backend. The conducted experiments included 72 transformations, i.e., $k = 72$, consisting of translations with $\{0, \pm 30\}$ pixels (reflect mode), counter-clockwise rotations with 90° and flipping in the left/right direction. A deep 72-class classification model was trained on *BigEarthNet train* dataset to learn semantic features in a self-supervised manner. The classification model was build based on residual learning, ResNet-50 backbone [28]. ResNet-based networks are more applicable for self-supervised representation learning, preserving representations from degrading towards the end of the network [32]. The 72-class classification model g_θ was trained for 50 epochs, with batch size equal to 600 samples, in 2h and 50 minutes. Fixed point Dirichlet maximum likelihood estimation function used maximum 1000 iterations and converged when L2 norm on Dirichlet parameters' iterations was less than $1e-9$.



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

Fig. 4.4 Average precision AUROC results on standardization (a) and max-normalization (b) preprocessing methods. Average precision AUPR results on max-normalization preprocessing method when burned area is studied as the positive class (c) and negative class (d).

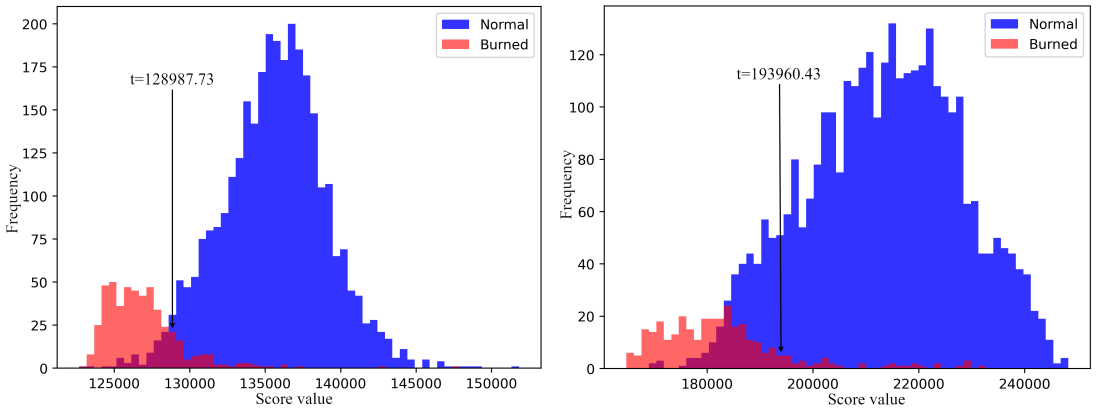


Fig. 4.5 Probability distributions of scores for Sydney (left) and Robore (right) datasets.

Results

In Figures 4.4a and 4.4b, the *BigEarthNet test* dataset performed well from the very first transformation, with small drops in performance as more transformations were included, while *Robore* and *Sydney* started with poor classification results but improved as more transformations were processed. Area Under Precision-Recall (AUPR) is implied to compute an average precision per transformation index when prior knowledge on the proportion of anomalies is available, considering the anomaly sequentially both classes, i.e, positive case, and negative case. Better performance is achieved when the anomaly is operated as negative class on all three test datasets (Figures 4.4c and 4.4d). The results are presented for the first 16 geometrical transformations.

Probability distributions of scores for *Sydney* and *Robore* are displayed in Figure 4.5 in overlay mode. Distributions of scores for *Sydney* and *Robore* datasets are bimodal and trimodal, respectively. Through computing the second derivative of probability density function, the obtained inflection points determinated an optimal threshold for each histogram, 90725.44 for *BigEarthNet test* dataset, 128987.73 for *Sydney* dataset, respectively 193960.43 for *Robore* dataset.

Table 4.1 Accuracy assessment for the testing datasets using the proposed model (DCNN-Dirichlet) and a parameter-free model (NCD)

Dataset	Method	Precision	Recall	F1-score
BigEarthNet test	NCD	**0.964**	0.714	0.781
	DCNN-Dirichlet	0.930	**0.941**	**0.935**
Robore	NCD	0.960	0.823	0.877
	DCNN-Dirichlet	**0.986**	**0.885**	**0.933**
Sydney	NCD	0.948	0.845	0.888
	DCNN-Dirichlet	**0.977**	**0.970**	**0.974**

In the Table 4.1 different metrics are presented to compare the results obtained by the proposed method and a state of the art method, parameter-free NCD method. Each test dataset was randomly split in ratio 60:40, and based on an NCD matrix computed for each test dataset, a binary nearest neighbors classifier was run on a 2-D space, where each dimension represented the average distance from normal class, respectively abnormal class.

4.2 Unsupervised learning

Anomaly detection methods in imagery focus on spectral and spatial discrimination, using statistical models [14], collaborative representation [35], or deep neural network-based anomaly detection models [6]. Autoencoders compute reconstruction errors between a restored image and the original input image. Being trained on normal background data, anomalies are expected to have large reconstruction errors, while background has small reconstruction errors [36]. In this section, is introduced a method that estimates anomalies in Sentinel-2 data based on the potential of an autoencoder reconstruction ability.

4.2.1 Methodology

The proposed methodology relied heavily on a ResNet [28] encoder-decoder architecture (Figure 4.6a). The covariance matrix of each input was computed to be used as a strong spatial feature on downstream classification. The latent data extracted from the autoencoder embedded space is ultimately used to train an OCSVM instance.

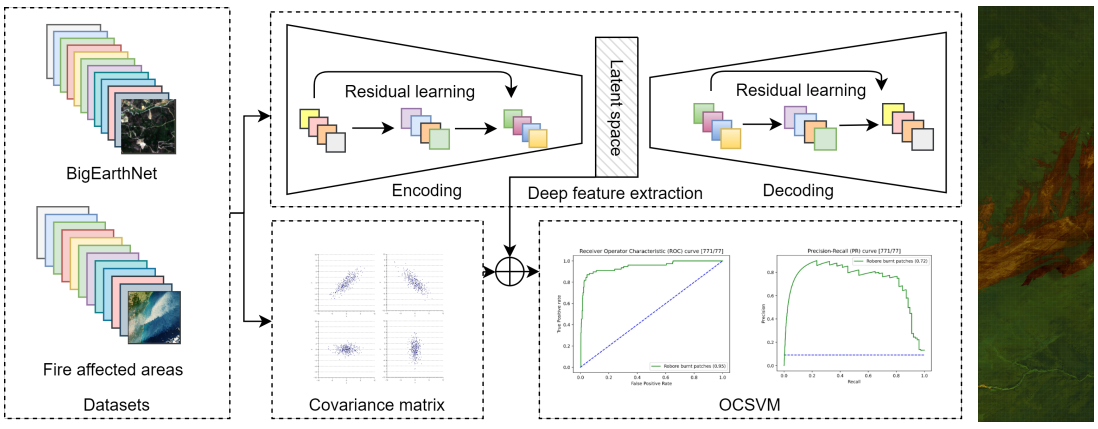


Fig. 4.6 Overview of the proposed method consisting in three modules, deep feature extraction, covariance matrix feature and OCSVM outlier detector (a). Sentinel-2 image, Robore, Bolivia, 23 August 2019 in FC representation (B12, B11, B9) (b).

Table 4.2 Evolution of AUPRC based on anomaly quantity.

Burned patches (#)	Burned patches in dataset (%)	AUPRC [0-1]
7	0.9	0.46
15-53	1.94-6.8	0.54-0.66
61	7.9	0.71
69-77	8.9-9.9	0.72

Model

An autoencoder encodes the input tensor into a lower dimensional latent representation, then decodes the latent representation back to an resemblant tensor, all with a minimum reconstruction error. The standard ResNet-18 [28] network was used with the first layer adapted to 12 input channels. For the input data, bicubic interpolation was performed for 20m and 60m resolution channels.

Datasets

BigEarthNet (4.1.3) was used to train the model. For testing, an area of 109,8 km x 30 km (Figure 4.6b) was divided into non-overlapping patches using the size 120x120 pixels.

4.2.2 Experiments

An experiment was conducted using a test dataset containing 771 normal and 77 abnormal patches, from the left half of the scene in Figure 4.6b. The test dataset started with 0.9% abnormal data and iteratively incremented with by $\sim$0.9% until all 77 available abnormal tiles were included. The OCSVM training process involved only normal data, from BigEarthNet, and the ratio of normal to abnormal areas was empirically set to 0.7.

In Table 4.2 is highlighted the evolution of AUPR Curve based on the abnormal data percentage included in test dataset. Under 2% anomalous patches in dataset, the classifier is not able to distinguish between positive and negative class points. For 7.9%, an accuracy of 0.71 is achieved.

4.3 Contributions

In this chapter, a DCNN-Dirichlet model has been proposed that detects anomalies in the post-event image. In addition, the DCNN-OCSVM model that identifies burned areas based on reconstruction ability has been presented.

Chapter 5

Meta-Recognition Anomaly Detection Hardware Accelerator on FPGA

This chapter proposes a methodology that relocates the computational effort on-board the satellite for detecting natural disaster of wildfires in multispectral imagery.

5.1 Current context

On-board real-time classification and detection methods for remote sensing imagery are based on different techniques, the computation of data covariance matrix and their inverse [15], orthogonal projection strategy [19, 4], linewise progressive anomaly detection [55], and lately AI [27].

5.2 Methodology

In this chapter, an algorithm has been proposed to extend the latent features extracted from satellite imagery through probability estimation, aiming to associate deviant samples to an anomalous class. Two stages are performed, implementation, training and testing the algorithm on a GPU cluster (Figure 5.1 (left)), followed by an optimization and deployment flow of the algorithm on an FPGA device (Figure 5.1 (right)).

In Figure 5.1 (left), a deep convolutional neural network (DCNN) classifier is trained on observed data and used as a feature extractor. Background statistical features are approximated by fitting Weibull distributions on each observed class in the training dataset, concerning a distance metric to determine a threshold for a binary decision. This methodology is related to the Meta-Recognition concept proposed in [1]. In Figure 5.1 (right), the deployment flow involved the Vitis AI Framework, to convert a Tensorflow DCNN model into a compatible format supported by the Deep Learning Processing Unit (DPU) [53] engine. After training, the frozen graph was quantized and compiled into an

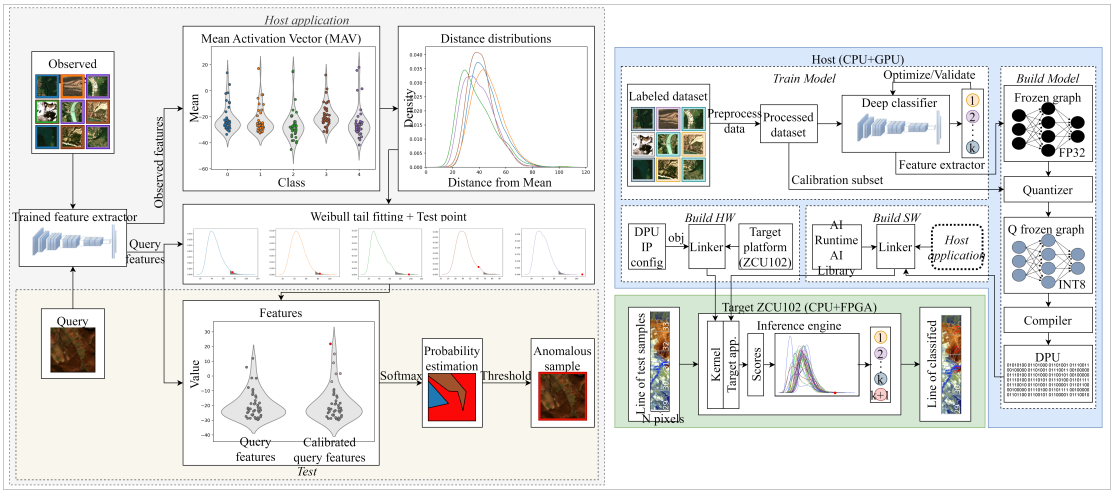


Fig. 5.1 Host application for anomaly detection (left), flowchart for the optimization and deployment of the on-board anomaly detection accelerator (right).

inference-only model that is called by the *host application* and run on a specific FPGA target board.

5.2.1 Meta-Recognition

Meta-Recognition [44] is a prediction method that uses statistical extreme value theory (EVT) for post-recognition score analysis. Basically, a statistical classifier based upon the Weibull distribution [44] is substantially better than a standard threshold test over the original score data. Recognition may be assimilated with the open set attribute of a real world classifier that rejects unseen classes at query time. For Weibull fitting, the FitHigh function is used, which is available in the libMR library [44].

Remote sensing images need a rich characterization through multi-labeling, a convolutional classifier with a Sigmoid function as the last activation layer. A multi-label feature extractor is trained on normal classes $C_i,\ 1 \leq i \leq k$ to compute for each class the corresponding MAV, $\mu_i = mean(v_{ij}),\ 1 \leq i \leq k,\ 1 \leq j \leq |C_i|$ using the activation vector v. With S_n was noted the set of indexes, in descending order, of the highest n probabilities generated by a final Sigmoid layer on activation vector v. Next, for each class C_i, a Weibull model $\rho_{C_i} = (\tau_{C_i}, \lambda_{C_i}, \kappa_{C_i}),\ 1 \leq i \leq k$ with three parameters, data shifting τ, Weibull scale λ, respectively Weibull shape κ, is computed. Parameter τ is dynamic and depends on data itself, i.e., is the smallest score (distance) on tested activation vector v, aiming to shift v in zero (5.1a). Model $\rho_{C_i}(x)$ provides meta-recognition estimated probability that determines if query sample x is anomalous or not.

In the second phase, the activations of a query sample x adjust α top activations, i.e., α top probabilities given by a Sigmoid layer, by approximating the Weibull distribution function (5.1a)

$$w_i(x) = 1 - f(i)(1 - e^{-(\frac{x-\tau_i}{\lambda_i})^{\kappa_i}}),\ 1 \leq i \leq k, \tag{5.1a}$$

$$f(i) = \begin{cases} \frac{\alpha+1-rank_i(S_\alpha(v))}{\alpha} & \text{if } i \in S_\alpha(v) \\ 0 & \text{otherwise} \end{cases}. \tag{5.1b}$$

After computing Weibull CDF on the distance between query sample x and MAVs of α top activations, revised activation vector is computed (5.2a), where operator $\circ$ is used for the scalar product between two vector.

$$\hat{v}(x) = v(x) \circ w(x), \tag{5.2a}$$

$$\hat{v}_{k+1}(x) = \sum_{i=1}^{k} v_i(x)(1 - w_i(x)). \tag{5.2b}$$

Afterwards, a pseudo-activation for the unseen class C_{k+1} is computed while preserving the total activation level constant (5.2b). Finally, the rejection is decided on the revised probabilities of normal classes $C_i,\ 1 \leq i \leq k$, with regards to the anomalous class C_{k+1} (5.3)

$$p_i(x) = \frac{e^{\hat{v}_i(x)}}{\sum_{j=1}^{k+1} e^{\hat{v}_j(x)}},\ 1 \leq i \leq k+1, \tag{5.3}$$

where x is anomalous if p_{k+1} is the highest in $p_i,\ 1 \leq i \leq k+1$ or higher than a threshold η.

5.2.2 Model quantization and compilation

The quantization step included a post-training process in which a small calibration subset of 1000 training images was used to analyze the distribution of activations and to limit the accuracy degradation. Data-Free Quantization (DFQ) algorithm [40] was used in post training process to equalize the weight ranges and to correct biases in the errors introduced during quantization.

5.3 Study areas and Datasets

5.3.1 Proposed datasets

Zamora

The Zamora scene contains disastrous wildfires that took place in Zamora, Spain during June 2022. In Figure 5.2a is shown a false-color image that renders non-visible parts of the electromagnetic spectrum.

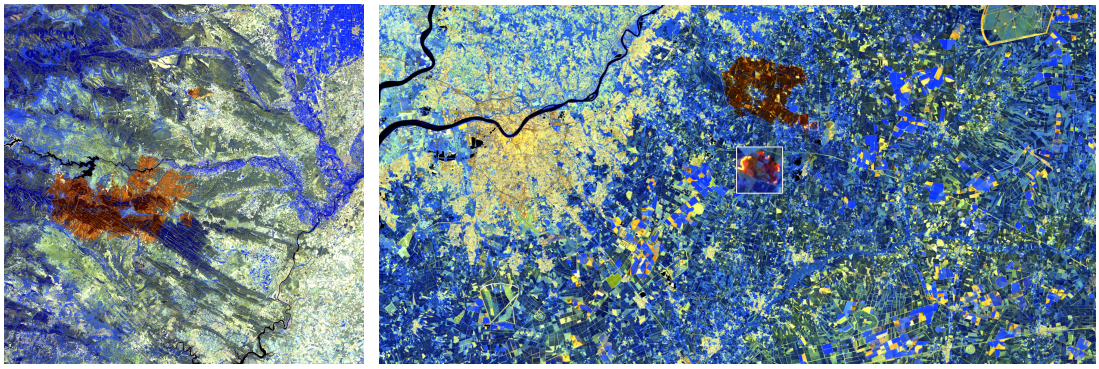


Fig. 5.2 Zamora scene (Spain, 2022) (a), Bordeaux scene (France, 2022) (b). False color image format (R: Band 12; G: Band 11; B: Band 8A). The burned areas are highlighted in brown color.

Bordeaux

The Bordeaux scene contains devastating blazes that took place in Gironde, near Bordeaux city. The composition of SWIR bands in Figure 5.2b provides a picture of the burned areas and reveals areas of ongoing fire.

5.4 Experiments

The spectral-spatial feature extractor, based on a ResNet model [28] with 50 convolutional layers, was trained on the clean version of BigEarthNet (4.1.3), i.e., samples completely covered by seasonal snow, clouds and cloud shadows were removed. Class *Burnt areas* was filtered out from the clean version of BigEarthNet, resulting in a 42 class-nomenclature [47]. The rule 60:20:20 for randomly choosing training, validation and test datasets was adopted. After training the multi-class classifier, the MAVs were calculated on the training dataset, for each class, taking into account all samples that were correctly classified. After finding the MAVs, the distance vectors were calculated with cosine distance (5.4)

$$\cos(\theta) = \frac{A \cdot B}{\|A\|_2 \|B\|_2}, \tag{5.4}$$

where A is the activation vector of an observed sample and B is the corresponding MAV of the class.

Table 5.1 Anomaly detection accelerator run time for a test collection of 60 samples

Device	# threads	FPS	Run time [s]
XCZU9EG MPSoC	4	77.813 ± 1.726	0.771 ± 0.017
	8	$\mathbf{85.677 \pm 0.954}$	$\mathbf{0.700 \pm 0.007}$
Tesla K80 GPU	multithread	19.861 ± 3.252	3.128 ± 0.744

Table 5.2 Accuracy assessment of anomaly detector on test datasets

Dataset	Device	Precision	Recall	F1-score
Zamora	Tesla K80 GPU	0.823	0.750	0.784
	XCZU9EG MPSoC	0.727	0.721	0.723
Bordeaux	Tesla K80 GPU	0.829	0.723	0.772
	XCZU9EG MPSoC	0.787	0.683	0.731

In on-board implementation, the target platform is a Xilinx Zynq UltraScale+ MPSoC ZCU102 evaluation board which combines a powerful processing system (PS) and a user-programmable logic (PL) into a powerful MPSoC, i.e., Zynq UltraScale+ XCZU9EG-2FFVB1156E MPSoC. This CPU-FPGA hybrid system was programmed on each part, the CPU with logic for input and output processing, and the FPGA with execution of the DCNN architecture. The target application is run on the CPU with Xilinx runtime application interface calls to manage runtime interection with accelerator.

A line of 60 samples is evaluated on the Zynq UltraScale+ XCZU9EG-2FFVB1156E MPSoC device in 0.7 seconds (Table 5.1). The best time performance is achieved when the anomaly detection algorithm is run on XCZU9EG MPSoC device with 8 threads on the PS side, achieving a processing speed of 81.018 ns/pixel. On the Tesla K80 GPU device, the running time is about 4.46x slower than the best time achieved on the FPGA device. Furthermore, the power consumed by the Tesla K80 GPU device is 135W compared to that of the entire Zynq UltraScale+ board which is 30W, resulting in an efficiency factor of 4.5 for the latter device.

In Table 5.2, the results of the proposed algorithm are highlighted through the Precision, Recall, and F1-score accuracy metrics. Due to quantization, there was a small decrease in F1-score metric when the algorithm is run on the embedded platform, i.e., 0.061 for Zamora dataset and 0.041 for Bordeaux dataset. The anomaly threshold η was set to 0.5. The Precision metric obtained by running the algorithm on the FPGA device was 0.727 for the Zamora dataset and 0.787 for the Bordeaux dataset, respectively.

5.5 Contributions

In this chapter, an optimized hardware architecture was designed onto an embedded MPSoC device, processing all convolutional layers in FPGA logic. The proposed accelerator achieved a processing speed of 0.7 seconds for 60 samples, which is about 4.46x faster and 4.5x more energy efficient than the GPU design.

Chapter 6

Conclusions

6.1 Original contributions

In the following, the main original contributions of the thesis are listed:

- The development of two parameter-free models, based on data compression, that extract conditional and unbiased features from bi-temporal series to determine binary change maps [C1, J1, J5];
- The development of two deep learning-based anomaly detection approaches that predict deviant samples from the normal probability distribution of the data, respectively suppress anomaly reconstruction, in multispectral images [C3, J2];
- A novel statistical anomaly detection model, developed and optimized for an FPGA device, that estimates the probability that multispectral samples belong to an anomalous class [J3, J4].

6.2 List of original publications

6.2.1 Journals

[J1] **M. Coca**, A. Anghel and M. Datcu, "Unbiased Seamless SAR Image Change Detection Based on Normalized Compression Distance," in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 12, no. 7, pp. 2088-2096, July 2019, doi: 10.1109/JSTARS.2019.2909143. **(Impact Factor:3.784 - Q2)**

[J2] **M. Coca**, I. Coca Neagoe and M. Datcu, "Hybrid DNN-Dirichlet Anomaly Detection and Ranking: Case of Burned Areas Discovery," in IEEE Transactions on Geoscience and Remote Sensing, vol. 60, pp. 1-16, 2022, Art no. 4414116, doi: 10.1109/TGRS.2022.3207311. **(Impact Factor:5.6 - Q1)**

[J3] **M. Coca** and M. Datcu, "FPGA Accelerator for Meta-Recognition Anomaly Detection: Case of Burned Area Detection," in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing (under review).

[J4] Cavallaro, Gabriele and B. Heras, Dora and Wu, Zebin and Maskey, Manil and Lopez, Sebastian and Gawron, Piotr and **Coca, Mihai** and Datcu, Mihai, "High-Performance and Disruptive Computing in Remote Sensing: HDCRS—A new Working Group of the GRSS Earth Science Informatics Technical Committee [Technical Committees]," in IEEE Geoscience and Remote Sensing Magazine, vol. 10, no. 2, pp. 329-345, June 2022, doi: 10.1109/MGRS.2022.3145478. **(Impact Factor:7.741 - Q1)**

[J5] Mihai Stoica and **Mihai Coca**, "GPU-Based Normalized Compression Distance for Satellite Images," in Journal of Military Technology, vol. 5, no. 1, pp. 49-54, July 2022, doi: 10.32754/JMT.2022.1.07.

6.2.2 Conference proceedings

[C1] **M. Coca**, A. Anghel and M. Datcu, "Normalized Compression Distance for SAR Image Change Detection," IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium, 2018, pp. 5784-5787, doi: 10.1109/IGARSS.2018.8518126.

[C2] **M. Coca**, I. Neagoe and M. Datcu, "Physically Meaningful Dictionaries for EO Crowdsourcing: A ML for Blockchain Architecture," IGARSS 2020 - 2020 IEEE International Geoscience and Remote Sensing Symposium, 2020, pp. 3688-3691, doi: 10.1109/IGARSS39084.2020.9324361.

[C3] **M. Coca** and M. Datcu, "Anomaly Detection in Post Fire Assessment," 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS, 2021, pp. 8620-8623, doi: 10.1109/IGARSS47720.2021.9554169.

6.3 Perspectives for future developments

In the following, the main future perspectives of this thesis are listed:

- The integration of the ground-based proposed methods into natural disaster management services such as the Copernicus Emergency Management Service;
- The development of anomaly detection methods using multi-sensor data fusion to avoid obstacles encountered by optical instruments;
- The improvement of the proposed method for on-board anomaly detection by increasing the utilization of the FPGA device.

References

[1] Bendale, A. and Boult, T. (2016). Towards Open Set Deep Networks. In *Computer Vision and Pattern Recognition (CVPR), 2016 IEEE Conference on*. IEEE.

[2] Bennett, C., Gacs, P., Li, M., Vitanyi, P., and Zurek, W. (1998). Information distance. *IEEE Transactions on Information Theory*, 44(4):1407–1423.

[3] Bovolo, F. and 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.

[4] Caba, J., Díaz, M., Barba, J., Guerra, R., Escolar, S., and López, S. (2022). Low-power hyperspectral anomaly detector implementation in cost-optimized fpga devices. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 15:2379–2393.

[5] Cebrian, M., Alfonseca, M., and Ortega, A. (2007). The normalized compression distance is resistant to noise. *IEEE Transactions on Information Theory*, 53(5):1895–1900.

[6] Chalapathy, R. and Chawla, S. (2019). Deep learning for anomaly detection: A survey. *arXiv:1901.03407*.

[7] Chuvieco, E. (2016). *Fundamentals of Satellite Remote Sensing: An Environmental Approach*. Number 2nd Edition in -. Boca Raton, CRC Press.

[8] Cilibrasi, R., Cruz, A. L., de Rooij, S., and Keijzer, M. (2008). Complearn. https://complearn.org/index.html. [Online; Accessed Oct. 15, 2022].

[9] Cilibrasi, R. and Vitányi, P. (2005). Clustering by compression. *IEEE Transactions on Information Theory*, 51(4):1523–1545.

[10] Coca, M., Anghel, A., and Datcu, M. (2019). Unbiased Seamless SAR Image Change Detection Based on Normalized Compression Distance. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 12(7):2088–2096.

[11] Coca, M. and Datcu, M. (2021). Anomaly Detection in Post Fire Assessment. In *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*, pages 8620–8623.

[12] Copernicus (2012). Emergency Management Service. https://emergency.copernicus.eu/index.html. [Online; Accessed Nov. 11, 2022].

[13] DLR (2020). TerraSAR-X - Germany's radar eye in space. https://www.dlr.de/content/en/articles/missions-projects/terrasar-x/terrasar-x-earth-observation-satellite.html. [Online; Accessed Nov. 12, 2022].

[14] Du, B. and Zhang, L. (2014). A discriminative metric learning based anomaly detection method. *IEEE Transactions on Geoscience and Remote Sensing*, 52(11):6844–6857.

[15] Du, Q., Tang, B., Xie, W., and Li, W. (2021). Parallel and distributed computing for anomaly detection from hyperspectral remote sensing imagery. *Proceedings of the IEEE*, 109(8):1306–1319.

[16] Dumitru, C. O., Cui, S., Faur, D., and Datcu, M. (2015). Data analytics for rapid mapping: Case study of a flooding event in germany and the tsunami in japan using very high resolution sar images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 8(1):114–129.

[17] Dumitru, C. O. and Datcu, M. (2013). Information content of very high resolution sar images: Study of feature extraction and imaging parameters. *IEEE Transactions on Geoscience and Remote Sensing*, 51(8):4591–4610.

[18] Dumitru, C. O., Schwarz, G., and Datcu, M. (2018). Sar image land cover datasets for classification benchmarking of temporal changes. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 11(5):1571–1592.

[19] Díaz, M., Guerra, R., Horstrand, P., López, S., and Sarmiento, R. (2019). A line-by-line fast anomaly detector for hyperspectral imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 57(11):8968–8982.

[20] Esposito, M. and Marchi, A. Z. (2019). In-orbit demonstration of the first hyperspectral imager for nanosatellites. In *International Conference on Space Optics—ICSO 2018*, volume 11180, pages 760–770. SPIE.

[21] European Space Agency (ESA) (2014). Resolution and Swath. https://sentinel.esa.int/web/sentinel/missions/sentinel-2/instrument-payload/resolution-and-swath. [Online; Accessed Aug. 5, 2022].

[22] European Space Agency (ESA) (2020). Phi-sat Artificial Intelligence for Earth Observation. https://www.esa.int/Applications/Observing_the_Earth/Ph-sat. [Online; Accessed Nov. 9, 2022].

[23] Frost, O. (1972). An algorithm for linearly constrained adaptive array processing. *Proceedings of the IEEE*, 60(8):926–935.

[24] Gawlikowski, J., Saha, S., Kruspe, A., and Zhu, X. X. (2022). An advanced dirichlet prior network for out-of-distribution detection in remote sensing. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–19.

[25] Geiger, D. and Heckerman, D. (1996). A characterization of the Dirichlet distribution with application to learning Bayesian networks. In *Maximum entropy and Bayesian methods*, pages 61–68. Springer, Dordrecht.

[26] Giuffrida, G., Diana, L., de Gioia, F., Benelli, G., Meoni, G., Donati, M., and Fanucci, L. (2020). Cloudscout: A deep neural network for on-board cloud detection on hyperspectral images. *Remote Sensing*, 12(14).

[27] Giuffrida, G., Fanucci, L., Meoni, G., Batič, M., Buckley, L., Dunne, A., van Dijk, C., Esposito, M., Hefele, J., Vercruyssen, N., Furano, G., Pastena, M., and Aschbacher, J. (2022). The phi-sat-1 mission: The first on-board deep neural network demonstrator for satellite earth observation. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–14.

[28] He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778.

[29] International Chapter Space and Major Disasters (March 28, 2013). Charter Brochure. https://disasterscharter.org/documents/10180/187832/charterbrochure28032013.pdf. [Online; Accessed Nov. 9, 2022].

[30] Kapur, J., Sahoo, P., and Wong, A. (1985). A new method for gray-level picture thresholding using the entropy of the histogram. *Computer Vision, Graphics, and Image Processing*, 29(3):273–285.

[31] Kittler, J. and Illingworth, J. (1986). Minimum error thresholding. *Pattern Recognition*, 19(1):41–47.

[32] Kolesnikov, A., Zhai, X., and Beyer, L. (2019). Revisiting Self-Supervised Visual Representation Learning. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1920–1929.

[33] Li, M., Chen, X., Li, X., Ma, B., and Vitanyi, P. (2004). The similarity metric. *IEEE Transactions on Information Theory*, 50(12):3250–3264.

[34] Li, M. and Vitányi, P. (1997). *An Introduction to Kolmogorov Complexity and Its Applications*. Berlin, Germany: Springer-Verlag.

[35] Li, W. and Du, Q. (2015). Collaborative representation for hyperspectral anomaly detection. *IEEE Transactions on Geoscience and Remote Sensing*, 53(3):1463–1474.

[36] Makhzani, A., Shlens, J., Jaitly, N., Goodfellow, I., and Frey, B. (2016). Adversarial autoencoders. arXiv:1511.05644.

[37] Masek, J. G., Wulder, M. A., Markham, B., McCorkel, J., Crawford, C. J., Storey, J., and Jenstrom, D. T. (2020). Landsat 9: Empowering open science and applications through continuity. *Remote Sensing of Environment*, 248:111968.

[38] Minka, T. (2000). Estimating a dirichlet distribution.

[39] Muhuri, A., Ratha, D., and Bhattacharya, A. (2017). Seasonal snow cover change detection over the indian himalayas using polarimetric sar images. *IEEE Geoscience and Remote Sensing Letters*, 14(12):2340–2344.

[40] Nagel, M., van Baalen, M., Blankevoort, T., and Welling, M. (2019). Data-free quantization through weight equalization and bias correction.

[41] Otsu, N. (1979). A Threshold Selection Method from Gray-Level Histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, 9(1):62–66.

[42] Planinšič, P. and Gleich, D. (2018). Temporal change detection in sar images using log cumulants and stacked autoencoder. *IEEE Geoscience and Remote Sensing Letters*, 15(2):297–301.

[43] Rostami, A., Shah-Hosseini, R., Asgari, S., Zarei, A., Aghdami-Nia, M., and Homayouni, S. (2022). Active fire detection from landsat-8 imagery using deep multiple kernel learning. *Remote Sensing*, 14(4).

[44] Scheirer, W. J., Rocha, A., Micheals, R. J., and Boult, T. E. (2011). Meta-recognition: The theory and practice of recognition score analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(8):1689–1695.

[45] Stoica, M. and Coca, M. (2022). GPU-Based Normalized Compression Distance for Satellite Images. *Journal of Military Technology*, 5(1):49–54.

[46] Sumbul, G., Charfuelan, M., Demir, B., and Markl, V. (2019). Bigearthnet: A Large-Scale Benchmark Archive for Remote Sensing Image Understanding. In *IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium*, pages 5901–5904.

[47] Sumbul, G., de Wall, A., Kreuziger, T., Marcelino, F., Costa, H., Benevides, P., Caetano, M., Demir, B., and Markl, V. (2021). BigEarthNet-MM: A large-scale, multimodal, multilabel benchmark archive for remote sensing image classification and retrieval [software and data sets]. *IEEE Geoscience and Remote Sensing Magazine*, 9(3):174–180.

[48] Verbesselt, J., Zeileis, A., and Herold, M. (2012). Near real-time disturbance detection using satellite image time series. *Remote Sensing of Environment*, 123:98–108.

[49] Vincenzi, S., Porrello, A., Buzzega, P., Cipriano, M., Fronte, P., Cuccu, R., Ippoliti, C., Conte, A., and Calderara, S. (2021). The color out of space: learning self-supervised representations for earth observation imagery. In *2020 25th International Conference on Pattern Recognition (ICPR)*, pages 3034–3041, Los Alamitos, CA, USA. IEEE Computer Society.

[50] Wicker, N., Muller, J., Kalathur, R. K. R., and Poch, O. (2008). A maximum likelihood approximation method for Dirichlet's parameter estimation. *Computational Statistics and Data Analysis*, 52:1315–1322.

[51] Xia, Y., Cao, X., Wen, F., Hua, G., and Sun, J. (2015). Learning Discriminative Reconstructions for Unsupervised Outlier Removal. In *2015 IEEE International Conference on Computer Vision (ICCV)*, pages 1511–1519.

[52] Xilinx (2021). Zynq ultrascale+ mpsoc data sheet: Overview (ds891), v1.9. https://docs.xilinx.com/v/u/en-US/ds891-zynq-ultrascale-plus-overview. [Online; Accessed Nov. 22, 2022].

[53] Xilinx (2022). DPUCZDX8G for Zynq UltraScale+ MPSoCs Product Guide (PG338). https://docs.xilinx.com/r/en-US/pg338-dpu/Introduction?tocId=Bd4R4bhnWgMYE6wUISXDLw. [Online; Accessed Nov. 22, 2022].

[54] Xiong, B., Chen, Q., Jiang, Y., and Kuang, G. (2012). A threshold selection method using two sar change detection measures based on the markov random field model. *IEEE Geoscience and Remote Sensing Letters*, 9(2):287–291.

[55] Zhang, L., Peng, B., Zhang, F., Wang, L., Zhang, H., Zhang, P., and Tong, Q. (2017). Fast real-time causal linewise progressive hyperspectral anomaly detection via cholesky decomposition. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 10(10):4614–4629.

[56] Zhao, W., Wang, Z., Gong, M., and 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.