



Leveraging autoencoders in change vector analysis of optical satellite images

Giuseppina Andresini¹  · Annalisa Appice^{1,2} · Daniele Iaia¹ · Donato Malerba^{1,2} · Nicolò Taggio³ · Antonello Aiello³

Received: 16 June 2021 / Revised: 18 August 2021 / Accepted: 18 August 2021 /
Published online: 23 September 2021

© The Author(s) 2021

Abstract

Various applications in remote sensing demand automatic detection of changes in optical satellite images of the same scene acquired over time. This paper investigates how to leverage autoencoders in change vector analysis, in order to better delineate possible changes in a couple of co-registered, optical satellite images. Let us consider both a primary image and a secondary image acquired over time in the same scene. First an autoencoder artificial neural network is trained on the primary image. Then the reconstruction of both images is restored via the trained autoencoder so that the spectral angle distance can be computed pixelwise on the reconstructed data vectors. Finally, a threshold algorithm is used to automatically separate the foreground changed pixels from the unchanged background. The assessment of the proposed method is performed in three couples of benchmark hyperspectral images using different criteria, such as overall accuracy, missed alarms and false alarms. In addition, the method supplies promising results in the analysis of a couple of multispectral images of the burned area in the Majella National Park (Italy).

Keywords Autoencoder · Change vector analysis · Change detection · Remote sensing · Satellite image analysis

1 Introduction

The Earth's surface is constantly changing due to anthropogenic and natural causes like the progression of desert areas, deforestation, glacier movements, fires or earthquakes (Alberti et al., 2003). Monitoring these changes over time may provide valuable information on the transformation of the Earth's environment paving the way for a better decision policy on the minimisation of the risk of disasters (Michel et al., 2012). In particular, the rapid development of multispectral (MS) and hyperspectral (HS) technology has recently unleashed the potential of change detection (CD) methods in a wide range of remote sensing applications

✉ Giuseppina Andresini
giuseppina.andresini@uniba.it

ranging from urban planning, environmental monitoring, agriculture investigation, disaster assessment and map revision (Kwan, 2019).

MS and HS sensors, mounted on space-born systems, allow a frequent revisit time of the same Earth's scene by acquiring observation data with high spectral and spatial resolution, while trying to keep constant acquisition characteristics (e.g. the same sun illumination and incidence angle if platforms are put in sun-synchronous orbit). Both technologies can reflect light in a spectrum of narrow frequencies that cover the visible, near-infrared and shortwave infrared bands (pixel spectral data). The difference between the MS and HS technology is the number of bands and how narrow the bands are. MS technology commonly refers to a small number of bands, i.e., from 3 to 10, sensed by a radiometer. HS technology could have hundreds or thousands of bands from a spectrometer. In any case, independently of the specific spectral resolution, both MS and HS sensors have definitely made available unprecedented optical information (Hoye & Fridman, 2013; Mouroulis et al., 2000) (compared to traditional RGB cameras) for learning in the Earth's observation. In general, processing MS/HS imagery data is nowadays a milestone in developing new CD methods in remote sensing.

Existing MS/HS CD methods mainly leverage machine learning (Shi et al., 2020) for comparing spectral data of each couple of images of a scene and learning patterns that delineate changes at either pixel or object level of the observed scene (Im et al., 2008). These methods are mainly classified into supervised and unsupervised methods regarding the learning paradigm they adopt (Shi et al., 2020). Supervised CD methods (e.g., Larabi et al., 2019; Seydi & Hasanlou 2017; Wu et al., 2017; Yuan et al., 2005) rely on prior information about the ground changes. Therefore, their accuracy strongly depends on the availability and quality of the ground truth that is commonly based on human intervention and tends to be generated object-wise, rather than pixel-by-pixel, since it is costly in terms of time and effort. A poor-quality ground truth map may prevent even a good supervised learning method from highlighting its quality by producing contradictory results.

Due to the limitation of supervised CD, a significant research effort is devoted to performing CD analysis in an unsupervised manner. In the unsupervised machine learning paradigm (Bruzzone & Prieto, 2000; Hussain et al., 2013), changes are commonly detected by resorting to the Change Vector Analysis (CVA) strategy that bases on a reliable measure of distance (or similarity) computed between the two images. In this strategy, a threshold is always determined to separate the changed pixels from the unchanged background.

Following this mainstream of research in CD, we propose a CVA method, named ORCHESTRA (autoEncoder-based CHange dEtection in hyper SpectRAL/ multispectral images), to analyse bi-temporal, co-registered MS/HS images of an Earth's scene, which are denoted as primary image and secondary image, respectively. The proposed method extends the traditional CVA strategy by taking advantage of autoencoder information. An autoencoder is an artificial neural network (ANN) architecture consisting of both an encoder function, mapping the input to a hidden code and a decoder, producing the reconstructed input learned by minimizing a loss function (Goodfellow et al., 2016). As the hidden code commonly reduces the size of data, autoencoders are mostly used for saving the output of the encoder function for dimensionality reduction (Ferreira et al., 2019; Hu et al., 2014; Shone et al., 2018; Wang et al., 2014; Wang et al., 2015). However, there are few recent studies that learn autoencoders, which go beyond the dimensionality reduction purpose, e.g., considering the data restored output of the decoder function for data denoising (Andresini et al., 2020; Zheng & Peng, 2018) or the loss (residual error) for the anomaly detection (An & Cho, 2015; Andresini et al., 2019; Oh & Yun, 2018; Sarafijanovic-Djukic & Davis, 2019).

Also in this study, we consider autoencoders for data restoring. In particular, we use the output of the decoder function of the primary image-specific autoencoder to restore both the primary image and the secondary image. Note that this is not directed to operate data denoising only. In principle, the autoencoder trained on the primary samples can contribute to recovering denoised samples of the pixels in the primary image, as well as denoised samples of the unchanged pixels in the secondary image, but it should see changed samples of the secondary image as anomalies, and so reconstruct them badly. So, the idea is to exploit autoencoders to disclose patterns that better delineate the pixels of the sensed scene where a change has occurred over time. We take advantage of these patterns by completing the CVA on the restored data (rather than on the original, acquired data). In particular, we compute the Spectral Angle Mapper (SAM) distance pixel-by-pixel between the restored spectral data vectors of the primary and secondary image, respectively. This distance quantifies the spectral change range at each pixel of the scene. The Otsu's algorithm (Otsu, 1972) is, finally, adopted to separate the foreground regions, where a change occurred, from the unchanged background.

We evaluate the proposed method performing the CD analysis of various, benchmark, bi-temporal, co-registered HS images collected in various urban and rural scenarios. As the change information is available on these datasets, the empirical study can verify the accuracy of the proposed CD method. In addition, we perform the evaluation of the viability of the proposed method in delineating the burnt area of bi-temporal, co-registered MS images acquired with Sentinel-2 in the area of Majella National Park (Italy).

The paper is organised as follows. The related works are presented in Section 2. The basic concepts are introduced in Section 3, while the proposed CD method is illustrated in Section 4 and the implementation is described in Section 5. The findings in the evaluation of the proposed strategy with benchmark HS data are discussed in Section 6, while the achievements in the analysis of the MS data of the burnt area in the Majella National Park (Italy) are illustrated in Section 7. Finally, Section 8 draws conclusions and proposes future developments.

2 Related work

Since obtaining a large number of labelled samples for supervised training is usually time-consuming and labour-intensive, remote sensing research devotes significant research effort for the formulation of CD methods in unsupervised machine learning.

In the unsupervised machine learning paradigm (Hussain et al., 2013; Bruzzone & Prieto, 2000), changes are commonly detected by resorting to the CVA strategy that computes a measure of similarity (or distance) between co-located pixels of a couple of images and uses a threshold-based approach to identify a distance threshold to separate changed pixels from the unchanged background. Various similarity (or distance) measures have been investigated for CVA methods (e.g., Appice et al., 2020; Falini et al., 2020; Seydi & Hasanlou, 2017; Yang & Mueller, 2007). The threshold to detect the changes is estimated by resorting to the spectral data (i.e. in a data-driven manner) (Lu et al., 2010; Najafi et al., 2017; Penglin et al., 2012) by leveraging probabilistic information extracted from the distribution of the (distance or similarity) measure among the pixels. A well-known approach commonly used for the threshold determination is Otsu's method (Otsu, 1972; Sahoo et al., 1988). In (López-Fandiño et al., 2019), Otsu's algorithms is evaluated in combination with SAM and Watershed algorithm. Alternatively, clustering algorithms are adopted (Appice

et al., 2019; Appice et al., 2020), in order to separate distances (or similarities) of changed pixels from the unchanged background.

Algebra-based methods, similarity-based methods, as well as distance-based methods belong to the threshold-based family of CD approaches. In particular, algebra-based CD methods use mathematical operations (such as image differencing or image ratio) on images taken at different times to generate a change matrix output (Ilsever & Unsalan, 2012). Similarity-based CD methods resort to the computation of a similarity measure (e.g., correlation measure) between a pair of spectral vectors (Choi et al., 2010). Distance-based CD methods are founded on a spectral distance measure (e.g. SAM, Z-score Information Divergence) computed between the spectral vectors on corresponding pixels (Choi et al., 2010). A recent study (Appice et al., 2020) adopts a spectral-spatial distance for CVA. In addition, it introduces an iterative upgrade of the traditional distance-based approach by accounting for a representation of the possible change iteratively learned through classification. Due to the lack of ground change information, classification is supervised with pseudo-labels yielded on the spectral-spatial distance information via clustering.

A different unsupervised perspective performs the change detection on image combination or transformation. In Deng et al. (2008), the Principal Component Analysis (PCA) is used to extract the difference between two images suppressing correlated information and highlighting variance in multi-temporal data. The change is identified in the second component, while the first component is assumed to be the sum of the common information. In Gao et al. (2016), PCA is used as a convolutional filter to determine the representative neighbourhood features from each pixel and generate change matrices with less noise spots. Gabor wavelets and fuzzy c-means are utilized to select interested bi-temporal pixels that have a high probability of being changed or unchanged. Then, new image patches centred at interested pixels are generated and a PCANet model—a deep learning network with its convolution filter banks chosen from PCA filters—is trained using these patches. Finally, pixels of bi-temporal images are classified by the trained PCANet model.

In Appice et al. (2019), an autoencoder architecture is trained on the pixelwise difference computed between the spectral data of two images. By assuming that the spectral differences compressed at the bottom encoder layer preserve the hidden information to separate changed pixels from the unchanged background, encoded differences are coupled to distance information and processed through clustering to separate changed pixels from the unchanged background. The data compression ability of the encoder layer of an autoencoder is also investigated in Kalinicheva et al. (2018). Each image is encoded to an equal-sized compressed feature representation and the output of the subtraction operation between the encoded images is analysed for the change detection. In Kalinicheva et al. (2019), a convolution autoencoder is trained on the patches of a time series of images. The reconstruction error of each patch is analysed to discriminate changed pixels from the background.

Supervised CD methods (Khanday, 2016) are based on the availability of ground change information (often acquired by human intervention) and use a classification framework, in which the ground truth is used to learn a classifier. The spectral, spatial or proper combination of this information is used to build a measure able to detect the change and aid in the classifier decision. ANN (Clifton, 2003; Helmy & El-Taweel, 2010) and Expectation Maximization (EM) (Ming et al., 2014) algorithms fall in this category. Although both ANN and EM are based on different concepts (basically, the former is based on nonlinear regression, the latter on a maximum likelihood with unknown parameters), they provide a binary decision like the classifiers. In Seydi and Hasanlou (2017), a supervised method is illustrated. It acquires a sample of change labels on a scene, in order to determine the optimal threshold for determining the remaining labels with a distance-based change detection approach.

In Wu et al. (2017), changes are identified by using a trained classifier to directly classify data from multiple periods (i.e., multi-date classification or direct classification) and comparing multiple classification maps (i.e. post-classification comparison). In Planinšič and Gleich (2018), a logistic regression layer is trained to perform supervised fine-tuning and classification on the autoencoder denoised representation of image time series feature extracted within tunable Q discrete wavelet transform. Finally, the transfer learning-based structure has been recently investigated to alleviate the lack of training samples and optimize the training process in a semi-supervised scenario. Transfer learning uses training in one domain to enable better results in another domain and, specifically, the lower to mid-level features learned in the original domain can be transferred as useful features in the new domain performing a fine-tuning according to a few labelled samples (Kerner et al., 2019; Larabi et al., 2019)

3 Preliminary concepts

A MS/HS sensor records reflected light in tens (MS)/hundreds (HS) of narrow frequencies covering the visible, near-infrared and shortwave infrared bands of a wavelength λ (also called spectrum). The spectrum is an M -dimensional feature vector (spectral feature vector), so that λ is spanned on M numeric spectral features (bands) $\lambda_1, \lambda_2, \dots$, and λ_M .

Let \mathbf{X} and \mathbf{Y} be two co-registered MS/HS images—digital images of an observed Earth’s scene, which are produced at different time points using the same MS/HS sensor mounted on aircraft or satellites. Note that if the searched changes concern the vegetation, images should be acquired in the same season to avoid the focus on changes related to different phenological stages of the vegetation. \mathbf{X} is denoted as the primary image, while \mathbf{Y} is denoted as the secondary image. Every image (see Fig. 1) is an hyper-cube of size $U \times V \times M$, which represents a collection of spectral vectors measured on an M -dimensional spectrum λ over a grid of $U \times V$ pixels. Every pixel (u, v) is a region of around a few square meters of the Earth’s surface, which is a function of the sensor spatial resolution. $\mathbf{X}(u, v)$ and $\mathbf{Y}(u, v)$ are one-dimensional real-valued spectrum sections of hyper-cubes \mathbf{X} and \mathbf{Y} , respectively, indexed by spatial coordinates u and v within the sensor resolution of the camera.

Every pixel of a scene for which bi-temporal MS/HS data are acquired can, in principle, be labelled according to an unknown binary target function, whose range is a finite set of two distinct labels, i.e. “changed” and “unchanged”. According to this function, a change

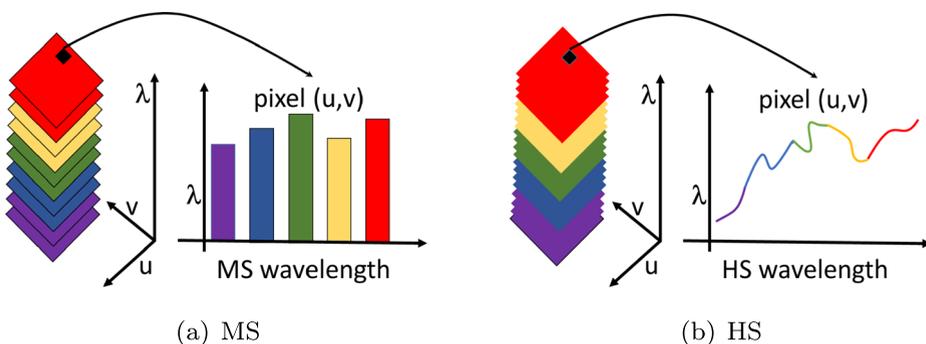


Fig. 1 MS/HS imagery data

matrix C can be associated with the bi-temporal image couple X and Y . In particular, C is a two-dimensional set of $U \times V$ change values with every value $C(u, v)$ representing the change label of the pixel indexed by the spatial coordinates u and v . A CD method takes as input images X and Y to learn C .

In this paper, an unsupervised CVA method based on autoencoder is proposed. An autoencoder is a deep learning ANN trained to attempt to copy its input to its output (Goodfellow et al., 2016). It can be viewed as being composed of two functions: an encoder f —mapping the input vector x to a hidden representation h via a deterministic mapping $h = f(x)$, parameterized by θ_f —and a decoder g —mapping back the resulting hidden representation h to a reconstructed vector in the input space $x' = g(h)$, parameterized by θ_g . The functions g and f correspond to two different ANNs combined in a single one, whose parameters $\{\theta_f, \theta_g\}$ are simultaneously learned by minimizing a loss function $\mathcal{L}(x, g(f(x))) = \mathcal{L}(x, x')$, penalising x for being dissimilar from x' such that $\mathcal{L}_{se}(x, x') = \|x - x'\|^2$.

4 The proposed methodology

In this section we describe ORCHESTRA—an unsupervised CVA method enhanced with autoencoder information. The method takes as input the bi-temporal images, X (primary image) and Y (secondary image), and learns a change matrix C . Figure 2 shows the block diagram of ORCHESTRA.

Initially, we train the autoencoder architecture $g \cdot f$ on the pixel spectral vectors acquired with the primary image X . Since the activation produced by the top layer in the decoder network g corresponds to a reconstructed feature vector in the same M -dimensional

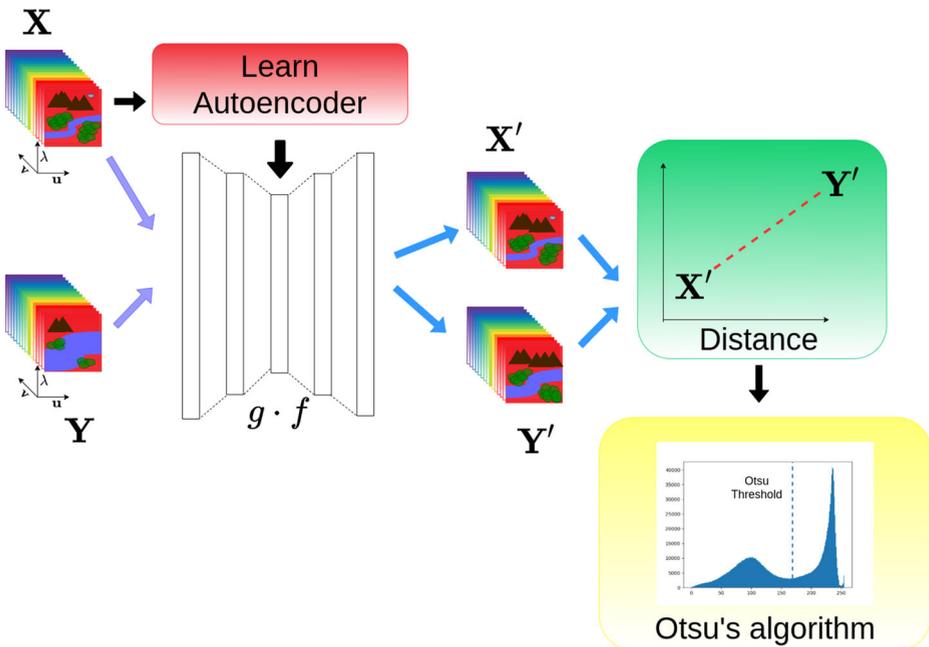


Fig. 2 ORCHESTRA methodology

spectral input space of the autoencoder, we consider this output feature vector as new learned features of the spectrum λ . The CVA of \mathbf{X} and \mathbf{Y} is then completed in this new feature space. According to these premises, $g \cdot f$ is used to restore the pixel spectral vectors of both \mathbf{X} and \mathbf{Y} and build the image reconstructions \mathbf{X}' and \mathbf{Y}' so that, for each pixel (u, v) , $\mathbf{X}'(u, v) = g(f(\mathbf{X}(u, v)))$ and $\mathbf{Y}'(u, v) = g(f(\mathbf{Y}(u, v)))$, respectively.

Some inherent remarks can be formulated on the reconstructions \mathbf{X}' and \mathbf{Y}' . As the autoencoder $g \cdot f$ is trained on the pixel spectral vectors of \mathbf{X} , we expect that \mathbf{X}' well reconstructs \mathbf{X} as it mainly performs a denoising transformation of \mathbf{X} . We also expect that $g \cdot f$ well reconstructs the spectral vectors of \mathbf{Y} associated with the unchanged pixels, while it poorly reconstructs the spectral vectors of \mathbf{Y} associated with the changed pixels. As a consequence, the spectral vector reconstructions of unchanged pixels in \mathbf{Y}' should be more similar to the corresponding spectral vector reconstructions restored in \mathbf{X}' than reconstructions associated with changed pixels. This conjecture (that is experimentally verified in Section 6) inspires the idea of computing the distance between the proposed autoencoder transformation of the original images, in order to better disentangle the differences between changed and unchanged pixels.

Then we compute pixelwise the distance between \mathbf{X}' and \mathbf{Y}' by resorting to the algorithm SAM that is commonly used in CVA methods (e.g. Appice et al., 2019, 2020; Lopez-Fandino et al., 2018). As pointed out in Seydi and Hasanlou (2017), the computation of SAM is independent of the number of spectral bands and insensitive to sunlight. Let us consider pixel (u, v) , $SAM(u, v)$ measures the angle between the bi-temporal reconstructed spectral vectors associated with (u, v) in both \mathbf{X}' and \mathbf{Y}' . This angle is computed as follows:

$$SAM(u, v) = \arccos \frac{\mathbf{X}'(u, v) \cdot \mathbf{Y}'(u, v)}{\|\mathbf{X}'(u, v)\| \|\mathbf{Y}'(u, v)\|}. \tag{1}$$

Subsequently, we perform the Otsu’s algorithm to automatically determine the upper threshold θ_{otsu} of SAM distances for separating pixels of the study scene into background (“unchanged” pixels with low SAM range) and foreground (“changed” pixels with high SAM range). In particular, we assign pixels (u, v) with $SAM(u, v)$ higher than θ_{otsu} to the label “changed”, while we assign the remaining pixels to the label “unchanged”.

The Otsu’s algorithm is an adaptive, non-parametric and unsupervised threshold algorithm introduced in Otsu (1972). It is commonly used in image binarization problems to turn a single intensity threshold that separates pixels into two classes. The threshold is determined by minimizing the intra-class intensity variance defined as a weighted sum of variances of the two classes.¹ In this paper, we assume that the SAM distances, computed pixelwise in the study scene, are represented in an histogram with L equal-width bins

(levels) denoted as $[1, \dots, L]$. Let η_i be the number of pixels at level i , so that $\sum_{i=1}^L \eta_i$ cor-

responds to the total number of pixels in the scene, i.e. $\sum_{i=1}^L \eta_i = UV$. Based upon these

premises, the probability of each level i is computed as $p_i = \frac{\eta_i}{UV}$. The Otsu’s algorithm identifies the optimal threshold level θ_{otsu} , in order to divide the pixels of the processed scene into the background class C_1 , spanned over the SAM levels $[1, 2, \dots, \theta_{otsu}]$, and the foreground class C_2 , spanned over the SAM levels $[\theta_{otsu} + 1, \dots, L]$, respectively. The

¹Minimizing the intra-class variance is equivalent to maximizing the inter-class variance, since the total variance (the sum of the intra-class variance and the inter-class variance) is constant for different partitions.

optimal θ_{otsu} is searched for minimizing the intra-class variance that is defined as a weighted sum of variances of the two classes:

$$\theta_{otsu} = \arg \min_{1 \leq \theta \leq L} \left(w_1(\theta)\sigma_1^2(\theta) + w_2(\theta)\sigma_2^2(\theta) \right), \tag{2}$$

where $\sigma_1^2(\theta)$ and $\sigma_2^2(\theta)$ are the variance computed on the two classes separated by θ . The weights $w_1(\theta)$ and $w_2(\theta)$ are the probabilities of the two classes, which are computed as follows:

$$w_1(\theta) = \sum_{i=1}^{\theta} p_i \text{ and } w_2(\theta) = \sum_{i=\theta+1}^L p_i. \tag{3}$$

Further considerations concern the fact that the direct application of the Otsu’s algorithm for change labelling will neglect the spatial arrangement of pixels. It may occasionally yield spurious assignments of pixels to classes. To avoid this issue, we may apply the principle of local auto-correlation congruence of objects (Appice et al., 2016, 2017; Du et al., 2012; Wang et al., 2015), according to which detected clusters, comprising changed objects, generally expand across contiguous areas (Appice et al., 2015). Based on this principle, we may decide to change the assignment of pixels that strongly disagree with surrounding assignments. This mainly corresponds to performing a spatial-aware correction of the change assignment defined with Otsu’s threshold. This correction assigns each pixel to the label that originally groups the majority of its neighbouring pixels (see Fig. 3).

Formally,

$$label(u, v) = \begin{cases} changed & \text{if } \#c(u, v) \geq \#u(u, v) \\ unchanged & \text{otherwise} \end{cases}, \tag{4}$$

where $\#c(u, v)$ and $\#u(u, v)$ count how many pixels, falling in neighbourhood $\epsilon(u, v)$, are labelled as “changed” and “unchanged”, respectively, with the Otsu’s threshold. The neighbourhood $\epsilon(u, v)$ is a set of pixels surrounding (u, v) in the study scene. As in (Appice et al., 2019, 2020; Appice & Malerba, 2019; Guccione et al., 2015), we consider a

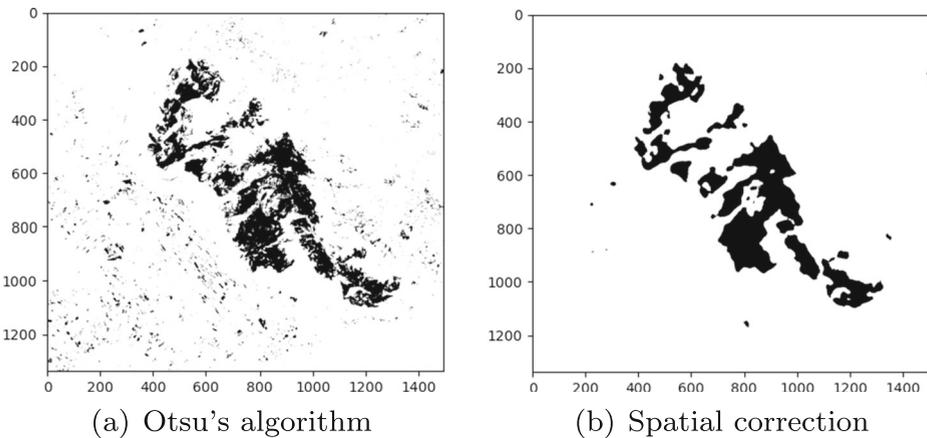


Fig. 3 Majella Park bi-temporal scene: the separation of the scene into black changed pixels and white unchanged background, which is computed via the Otsu’s algorithm (Fig. 3a); the spatial correction of the change labels assigned through clustering (Fig. 3b)

square-shaped neighbourhood. Let R be a positive, integer-valued radius, the square-shaped neighbourhood $\epsilon(u, v)$ of pixel u, v is defined as follows:

$$\epsilon(u, v) = \bigcup_{I=-R}^{+R} \bigcup_{J=-R}^{+R} \{(u + I, v + J)\}. \quad (5)$$

Finally, we analyse the time complexity of the proposed method. The time cost of the autoencoder layers is $\mathbf{O}\left(\sum_{l=1}^{d_A} n_{l-1}n_l\right)$ (İrsoy & Alpaydm, 2017), where d_A is the number of layers in the autoencoder, l is the index of a layer and n_l is the number of nodes in layer l . The time complexity of the distance computation is $\mathbf{O}(UV M)$. The time complexity of the Otsu's algorithm is $\mathbf{O}(UV)$, while the time complexity of the spatial correction operation is $\mathbf{O}(UV R^2)$. In general, the most of the time cost is spent training the autoencoder ANN.

5 Implementation details

ORCHESTRA is implemented in Python 3.8. A pre-processing step is performed to scale spectral data in the range [0, 1] and process spectral bands with values in comparable ranges.

The autoencoder is developed in Keras 2.4.3² with TensorFlow³ as the back-end. The set-up of learning rate and batch size is decided by resorting to the tree-structured Parzen estimator algorithm, as implemented in the Hyperopt library (Bergstra et al., 2013). This hyper-parameter optimization is done by using 20% of the entire training collection as a validation set. Therefore, we automatically choose the configuration of learning rate and batch size, which achieves the best validation loss in training the autoencoder. The values of learning rate and batch size explored with the tree-structured Parzen estimator, are defined as follows: learning rate varies in the range [0.00001, 0.01] and batch size ranges among 32, 64, 128, 256 and 512. The autoencoder architecture comprises 5 fully-connected (FC) layers of $128 \times 64 \times 32 \times 64 \times 128$ neurons when trained with HS data and 3 fully-connected (FC) layers of $8 \times 4 \times 8$ neurons when trained with MS data. Both architectures comprise a dropout layer to prevent overfitting. The mean squared error (*mse*) is used as the loss function. The classical rectified linear unit (*ReLU*) (Glorot et al., 2011) is selected as the activation function for each hidden layer, while for the last layer the *Linear* activation function is used. The number of epochs is set equal to 150, retaining the best models achieving the lowest loss on the validation set.

For the autoencoder architectures, both the number of layers and the number of neurons per layer are selected by taking into account the size of the spectral feature vector of each imagery dataset. In particular, the HS images are spanned on a spectral feature vector with either 224 or 242 spectral bands (see Table 1), while the MS images are spanned on a spectral feature vector with 13 spectral bands. As the MS data are simpler than the HS data, the autoencoder architecture adopted to process the MS images is simpler than the architecture to process the HS data images. On the other hand, we also account for the principle that a high number of layers may cause an increase of the computational effort that may not be rewarded with a gain in accuracy (Uzair & Jamil, 2020). With regard to the number of neurons, we follow the guidelines reported in Vanhoucke et al. (2011) and

²<https://keras.io/>

³<https://www.tensorflow.org/>

Table 1 Data scenario description: scene size (column 2), number of spectral bands (column 3), number of changed pixels in the ground truth (GT) change matrix (column 4), number of unchanged pixels in the change matrix (column 5), number of pixels with an unknown label in the change matrix

dataset	size	spectral bands	changed	unchanged	unknown
Hermiston	390 × 200	242	9986	68014	0
Santa Barbara	984 × 740	224	52134	80418	595608
Bay Area	600 × 500	224	39270	34211	26519

select the number of neurons in each hidden layer as a power of two, in order to improve the speed in the computation of the neural network. In fact, most of the computation time spent training an ANN is devoted to performing matrix multiplication. This is computed as a SIMD (single instruction, multiple data) operation in CPUs by using a batch size that is a power of 2.

Finally, the threshold-based step is performed using the implementation of Otsu's algorithm from `skimage.filters.threshold_otsu`,⁴ with the number of levels $L = 256$.

6 Experimental evaluation and discussion

In this study we consider three co-registered, bi-temporal HS datasets (see Section 6.1) acquired in both rural and urban environments. For these datasets, the ground truth change information is available to validate the accuracy of ORCHESTRA. In particular, the accuracy performance is evaluated with the Overall Accuracy (OA), the number of Missed Alarms (MA – changed pixels assigned to the unchanged background) and the number of False Alarms (FA – unchanged pixels labelled as changed). These metrics are commonly considered in remote sensing for the evaluation of change detection methods. In addition, we measure the residual error of autoencoders (mean squared error on restored HS data) on both the primary image and the secondary image to explore the ability of the autoencoder in HS data reconstruction. The results, achieved on each dataset, are discussed in Section 6.2.

6.1 HS data

We consider three public available datasets⁵—Hermiston, Santa Barbara and Bay Area. Each data set comprises a couple of co-registered HS images of a scene, as well as ground truth information of the change occurred in the sensed scene. A brief description of the datasets is reported in Table 1.

In Hermiston, the study areas cover an irrigated agricultural field. This area provides a benchmark agricultural scene, which has been frequently used in the evaluation of the accuracy of HS CD methods (e.g. Appice et al., 2019, 2020; Lopez-Fandino et al., 2017, 2018). The land-cover types are soil, irrigated fields, rivers, buildings and types of cultivated land and grassland. In this dataset, the bi-temporal HS images were acquired with the HYPERION sensor. This is a space-borne system carried on the EO-1 satellite, which includes 242 spectral bands, covering wavelengths between 400 nm and 2.5 μm . The spectral range is

⁴https://scikit-image.org/docs/dev/api/skimage.filters.html#skimage.filters.threshold_otsu

⁵<https://gitlab.citius.usc.es/hiperespectral/ChangeDetectionDataset>

divided into two intervals: the VNIR range (that includes 70 bands with wavelengths ranging from 356 to 1058 nm) and the SWIR range (that consists of 172 bands with wavelengths between 852 and 2577 nm). The spectral and spatial resolution of this sensor is about 10 nm and 30 m, respectively, over a 7.5-km strip. The Hermiston scene was monitored in the years 2004 and 2007 with the sensor over Hermiston City, Umatilla County, Oregon, USA. Each HS image of the dataset consists of 390×200 pixels acquired across 242 spectral bands.

In both Santa Barbara and Bay Area, the study areas cover an urban suburb in California. Both datasets have been already used in the evaluation of the HS CD methods illustrated in (Appice et al., 2019, 2020). In both the datasets, images were acquired by using the AVIRIS sensor. This is an optical sensor that delivers calibrated images of the upwelling spectral radiance in 224 contiguous spectral bands with wavelengths from 400 to 2500 nm. The spectral and spatial resolution of this sensor are about 10 nm and 4 m, respectively. The Santa Barbara scene was monitored in the years 2013 and 2014 with the sensor over the Santa Barbara region (California). It consists of 984×740 pixels and includes 224 spectral bands. The Bay Area scene was monitored in the years 2013 and 2015 with the sensor surrounding the city of Patterson (California). It consists of 600×500 pixels and includes 224 spectral bands.

6.2 Results

We start evaluating how the autoencoder trained on the primary image discloses knowledge that may contribute to separate changed pixels from unchanged pixels. To this aim, we explore how the autoencoder $g \cdot f$ trained on the image of the couple assigned to the primary role can accurately reconstruct the primary image, while badly reconstructing the changed pixels of the secondary image. We evaluate two configurations defined by assigning the role of the primary image to (1) the oldest image and (2) the newest image of the couple, respectively.

Table 2 reports the mean squared error (mse) computed comparing pixelwise each image to its reconstruction restored through the trained autoencoder. In both configurations, the autoencoder trained on the primary image reconstructs worse the secondary image, getting a poor restore of spectral vectors of changed pixels. This can be seen in Fig. 4a, b and c that depict the maps of the squared errors computed pixelwise on the reconstructions of the images acquired in the Hermiston dataset. The reconstructions are done with the autoencoder configuration (1) that is trained considering the oldest image acquired in 2004 as the primary image. These maps highlight that the changed area is already delineated from the poorly reconstructed pixels in the secondary image acquired on 2007. This supports our

Table 2 Autoencoder configurations: mean squared error

dataset	X	Y	$mse(\mathbf{X}, \mathbf{X}')$	$mse(\mathbf{Y}, \mathbf{Y}')$	$\frac{mse(\mathbf{Y}, \mathbf{Y}')}{mse(\mathbf{X}, \mathbf{X}')}$
Hermiston	2004	2007	0.0001501	0.0002091	1.39
	2007	2004	0.0001595	0.0002104	1.31
Santa Barbara	2013	2014	0.0000138	0.0000332	2.40
	2014	2013	0.0000066	0.0000116	1.73
Bay Area	2013	2015	0.0000057	0.0000096	1.67
	2015	2013	0.0000036	0.0000079	2.21

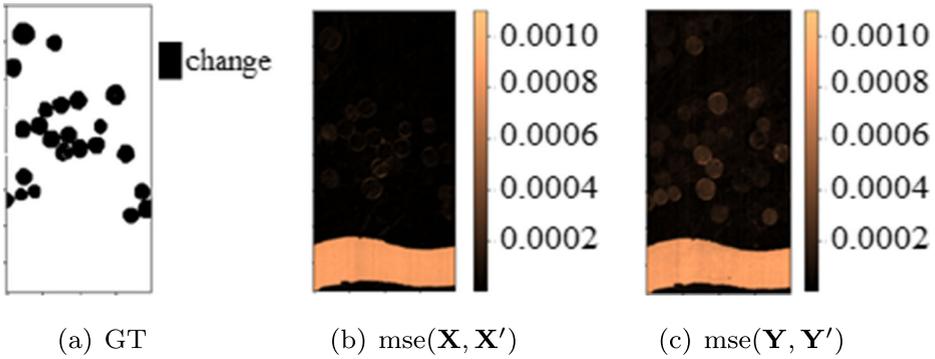


Fig. 4 Hermiston dataset: change ground truth (GT), $mse(\mathbf{X}, \mathbf{X}')$ and $mse(\mathbf{Y}, \mathbf{Y}')$ with the the image acquired on 2004 as primary image \mathbf{X} and the image acquired on 2007 as secondary image \mathbf{Y}

hypothesis that the autoencoder transformation can disclose a representation of the spectral data, which contributes to better disentangle the change.

We proceed with measuring how the autoencoder can actually improve the accuracy of the CVA strategy. Table 3 reports the accuracy metrics of both ORCHESTRA and its baseline (CVA), that is defined by implementing the basic CVA with SAM and Otsu’s algorithm on the original data (i.e. without the autoencoder architecture). Results show that both configurations of ORCHESTRA—(1) and (2)—outperform CVA. Interestingly, the highest accuracy (OA) is always achieved with the configuration of ORCHESTRA that maximizes the ratio of the mse computed on the reconstruction of the secondary image on the mse computed on the reconstruction of the primary image ($\frac{mse(\mathbf{Y}, \mathbf{Y}')}{mse(\mathbf{X}, \mathbf{X}')}$) reported in Table 2). This defines a promising criterion to automatically select the best configuration of ORCHESTRA in an unsupervised manner. Final considerations concern the spatial correction that is beneficial except for Hermiston.

Table 3 Accuracy performance (OA, FA and MA) of ORCHESTRA and CVA

method	no correction	correction			correction		
		OA	FA	MA	OA	FA	MA
Hermiston	ORCHESTRA(1)*	98.71	536	467	98.67	529	504
	ORCHESTRA(2)	98.70	518	495	98.67	506	526
	CVA	98.55	602	528	98.66	478	566
Santa Barbara	ORCHESTRA(1)*	94.55	1859	5360	97.04	438	3479
	ORCHESTRA(2)	94.08	1652	6186	96.47	356	4315
	CVA	93.51	1450	7144	95.61	262	5552
Bay Area	ORCHESTRA(1)	93.18	907	4099	95.20	36	3484
	ORCHESTRA(2)*	93.57	835	3889	95.61	38	3184
	CVA	93.10	740	4328	94.84	31	3760

(1) denotes the configuration of ORCHESTRA with the oldest image as primary image. (2) denotes the configuration of ORCHESTRA with the newest image as primary image. (*) marks the configuration of ORCHESTRA that maximises $\frac{mse(\mathbf{Y}, \mathbf{Y}')}{mse(\mathbf{X}, \mathbf{X}')}$ in Table 2. Results are reported without spatial correction, as well as with spatial correction ($R = 3$). The highest OA is in bold

Table 4 Compared competitors (OA). Results of competitors reported in the reference papers

dataset	ORCHESTRA	(Lopez-Fandino et al., 2018)	(López-Fandiño et al., 2019)	(Appice et al., 2019)	(Appice et al., 2020)
Hermiston	98.71	98.70	–	98.80	98.81
Santa Barbara	97.04	–	96.96	89.55	95.80
Bay Area	95.61	–	70.45	85.29	95.51

Finally, we analyse the accuracy of few CVA methods that have been defined in the recent literature and evaluated on the same datasets. Table 4 reports the OA results. The compared methods also use SAM and spatial information for final label assignment. In addition, Lopez-Fandino et al. (2018) and López-Fandiño et al. (2019)⁶ introduce the watershed analysis, Appice et al. (2019) resorts to autoencoder for dimensional reduction, while (Appice et al., 2020) uses an iterative combination of clustering and classification. ORCHESTRA performs closely to competitors on Hermiston. It outperforms (Appice et al., 2019, 2020 and López-Fandiño et al., 2019) on both Santa Barbara and Bay Area. On the other hand, the iterative procedure defined in Appice et al. (2020) may be considered for a future upgrade of ORCHESTRA.

Upon the completion of this comparative analysis, we perform the Friedman-Nemenyi statistical test (Demšar, 2006) on Hermiston, Santa Barbara and Bay Area. This test ranks the compared CVA methods for each dataset separately, so the best performing method is given rank of 1, the second best rank 2 and so on (Demšar, 2006). Figure 5 ranks the CVA methods according to the result of the Friedman-Nemenyi statistical test done on OA. The results of the test confirm that ORCHESTRA enables the construction of the change matrix that achieves the highest OA by having (Appice et al., 2020) as runner-up.

7 Majella national park analysis

Wildfires generate significant and complex environmental changes such as physical and chemical variations of soils, structural changes of vegetation, changes in ecological processes and ecosystem services (Meng & Zhao, 2017). Satellite MS data are traditionally exploited for monitoring burnt areas and wildfire effects. In this paper, we analyse the ability of ORCHESTRA in detecting environmental changes (e.g. physical and chemical variations of soils, structural changes of vegetation, changes in ecological processes and ecosystem services) caused by wildfires in MS images. In particular, we process two co-registered Sentinel2 L1C images acquired on both August 16, 2017 (Fig. 6a) and September 15, 2017 (Fig. 6b), in the area of the Morrone Mountain (within the Majella National Park, Italy). This area was burnt in a wildfire started on August 19, 2017 and lasted 25 days, which burnt more than 2,000 ha of an inaccessible area covered by coniferous forest and gorse. The processed MS images are composed of 1494×1338 pixels, with pixel resolution equal to 10m/pixels and MS resolution equal to 13 spectral bands (Aiello et al., 2019).

⁶The authors of López-Fandiño et al. (2019) describe a GPU framework for the CVA method described in Lopez-Fandino et al. (2018).

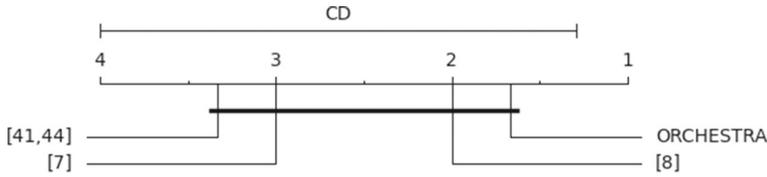


Fig. 5 Comparison based on the Friedman-Nemenyi test of OA computed on the change matrices built using ORCHESTRA and the related methods (Lopez-Fandino et al., 2018; López-Fandiño et al., 2019; Appice et al., 2019, 2020). The authors of (Lopez-Fandino et al., 2018) and (López-Fandiño et al., 2019) test the same CVA approach

We perform a preliminary analysis calculating the Normalized Burn Ratio (NBR) index on both the pre-fire and post-fire MS images. This index is commonly used to highlight burnt areas. Formally,

$$NBR = (NIR - SWIR)/(NIR + SWIR), \tag{6}$$

where the reflectance in the mid-infrared band (*SWIR*), that is sensitive to the water content of both soil and vegetation, increases after a fire. On the other hand, the near-infrared band (*NIR*) declines in reflectance after a fire due to the decrease of the phytomass chlorophyll-content. So, following the the conclusions drawn in Key and Benson (2006), we are able to assess the fire severity in a study area by measuring the difference between the NBR index calculated on both the pre-fire and a post-fire satellite images:

$$dNBR = NBR_{pre_fire} - NBR_{post_fire}. \tag{7}$$

In fact, this difference is correlated with the magnitude of changes caused by fires on the vegetation (Key & Benson, 2006). By assuming that the unburnt areas have similar spectral behaviour in two satellite images acquired before and after a fire, *dNBR* measures values around zero in unburnt areas, while it measures positive values in burnt areas. Figure 7 delineates the fire borders (red line) detected in the study area with the *dNBR* analysis conducted as described in Key and Benson (2006).

Although *dNBR* is one of the well-performing indexes in the detection of burnt areas over large fire zones with open forests and woodlands (Tran et al., 2018), it suffers from

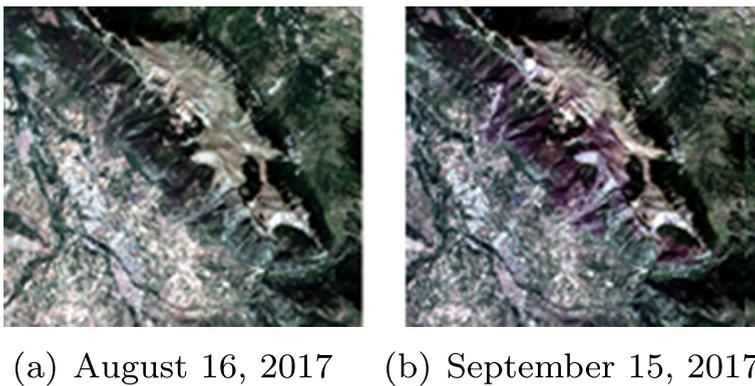


Fig. 6 Colour-based version of Sentinel2 L1C images acquired on both August 16, 2017 and September 15, 2017, in the area of the Morrone Mountain (Fig. 6b) within the Majella National Park, Italy)

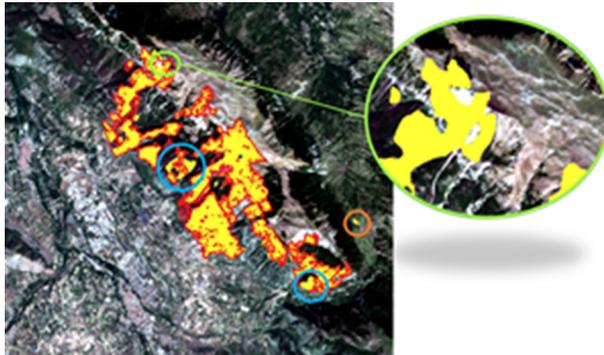


Fig. 7 Burnt area detected by *dNBR* (red line) and ORCHESTRA (yellow zone), respectively. New burnt areas detected (blu circles), the false positives areas (orange circle) and the burnt areas detected independently of clouds (green circle) by ORCHESTRA

a few limits. It is influenced by the fact that unburnt areas do not remain static over time, but they naturally undergo changes, passing from more or less dry/humid conditions over time. The parameters of the *dNBR* analysis need to be reviewed in each scenario based on several factors, e.g. seasonality of images, closeness of image acquisition to the fire event. In addition, the *dNBR* computation is sensitive to variations in the soil brightness (Epting et al., 2005a), the type of vegetation (Epting et al., 2005b) and the density of the vegetation (Lentile et al., 2009). Finally, both clouds and their shadows can worsen the scenario when the *dNBR* analysis is done on large areas.

In this study, we explore which limits of *dNBR* analysis may be overcome by performing the CVA strategy with ORCHESTRA. To this aim, we consider the configuration of ORCHESTRA that handles the pre-fire image as X and the post-fire image as Y . This configuration allows us to train the autoencoder that maximizes the ratio of the mse computed on the reconstructed images. We apply the correction with $R = 10$. In particular, we focus the attention on: (1) the correctness of the detected fire borders; (2) the ability to detect any new burnt area correctly detected, as well as the presence of false-alarm areas; (3) the robustness of the performance to possible clouds. To reduce computational effort, we use the Corine Land Cover 2018 classification and then we analyse only pixels belonging to “Forest and semi-natural areas”.

Figure 7 highlights the advantages achieved with the CVA completed with ORCHESTRA. Blu circles underline that ORCHESTRA is able to detect newly burnt areas that are undetected with the *dNBR*. The green circle is a zoom-in to show the capability of ORCHESTRA to avoid changes that are due to the presence of clouds. Finally, we note that only one polygon (orange circle) is detected as a false alarm. We can conclude that, also for this particular dataset, ORCHESTRA shows good potential to reach more effective identification of burnt areas.

8 Conclusion

This paper describes a CVA method for analyzing a couple of optical satellite images (i.e., primary image and secondary image) acquired over time on the same scene, in order to separate pixels where a change occurs from the unchanged background in the scene. In

particular, the proposed method takes advantage of autoencoders to identify spectral patterns that may aid in better disentangling changed pixels from unchanged ones. First an autoencoder is trained on the primary image and used to restore both the primary and the secondary image. Then the SAM distance is computed pixel-by-pixel between the restored images as a measure of the spectral change. Finally, the Otsu's algorithm is used on the computed distances to isolate the changed pixels, which are the pixels that measure the highest distance.

The novelty of the proposed CVA method is the specific use of an autoencoder architecture to transform the spectral data to compare, in order to enhance the spectral changes resulting in processed data. This is different from the common use of the autoencoders for data dimensionality reduction. Specifically, we base on the considerations that the autoencoder trained on the primary image should restore both the pixels of the primary image and the unchanged pixels of the secondary image accurately. Instead, it should see changed pixels of the secondary image as anomalies and reconstruct them badly. Therefore, computing a distance between restored spectral data measured at the same pixel aids in better delineating possible changes in the scene.

The experiments are performed by processing three couples of satellite HS images, collected either in a benchmark agricultural scene or in an urban scene. These experiments prove that the autoencoder component of the methodology contributes to the gain in detection accuracy. These experiments also reveal that the proposed method is able to provide competitive accuracy, compared to recent state-of-the-art CVA methods (comprising recent methods with autoencoders). In fact, with the encouraging performance of the proposed method, precise land-use and land-cover (or cropping pattern) changes may be identified. In addition, the method supplies promising results in the analysis of a couple of satellite MS images of a burnt area in the Majella National Park (Italy).

Some directions for further work are still to be explored. For example, appropriate classification algorithms may be studied to discriminate among different change types. The performances of various distance measures may be considered for the CVA. In addition, we plan to study the performance of the autoencoder-enhanced distance measures within a deep metric learning framework (e.g. Siamese network or Triplet network). Finally, we intend to investigate different autoencoder architectures, e.g. convolutional autoencoders, in the spectral data reconstruction.

Acknowledgements This work fulfills the research objectives of the PON "Ricerca e Innovazione" 2014–2020 project "CLOSE – Close to the Earth" (ARS01.00141), funded by the Italian Ministry for Universities and Research (MIUR).

Author Contributions **Giuseppina Andresini**: Conceptualization, Methodology, Validation, Software, Investigation, Writing - original draft, Writing - review & editing. **Annalisa Appice**: Conceptualization, Methodology, Validation, Writing - original draft, Writing - review & editing, Supervision, Project administration, Funding acquisition. **Daniele Iaia**: Software, Validation, Data curation. **Donato Malerba**: Conceptualization, Writing - original draft, Writing - review & editing. **Nicolò Taggio**: Validation, Writing - original draft, Writing - review & editing. **Antonello Aiello**: Validation, Writing - original draft, Writing - review & editing.

Funding Open access funding provided by Università degli Studi di Bari Aldo Moro within the CRUI-CARE Agreement.

Availability of Data and Material Public available datasets, Hermiston, Santa Barbara and Bay Area, can be downloaded at <https://gitlab.citius.usc.es/hiperespectral/ChangeDetectionDataset>. Majella National Park dataset is available from the corresponding author upon reasonable request.

Code Availability Code that support the findings of this study are available from the corresponding author upon reasonable request.

Declarations

Conflict of Interests The authors declare that they have no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Aiello, A., Ceriola, G., & Barbieri, V. (2019). Rheticus@wildfires: actionable geoinformation on burnt areas for post-fire assessment. In *12Th EARSel forest fires SIG workshop* (pp. 62–63).
- Alberti, M., Marzluff, J. M., Shulenberger, E., Bradley, G., Ryan, C., & Zumbrunnen, C. (2003). Integrating humans into ecology: opportunities and challenges for studying urban ecosystems. *Bioscience*, 53(12), 1169–1179. [https://doi.org/10.1641/0006-3568\(2003\)053\[1169:IHIEOA\]2.0.CO;2](https://doi.org/10.1641/0006-3568(2003)053[1169:IHIEOA]2.0.CO;2).
- An, J., & Cho, S. (2015). Variational autoencoder based anomaly detection using reconstruction probability.
- Andresini, G., Appice, A., Di Mauro, N., Loglisci, C., & Malerba, D. (2019). Exploiting the auto-encoder residual error for intrusion detection. In *European symposium on security and privacy workshops* (pp. 281–290).
- Andresini, G., Appice, A., Mauro, N. D., Loglisci, C., & Malerba, D. (2020). Multi-channel deep feature learning for intrusion detection. *IEEE Access*, 8, 53,346–53,359. <https://doi.org/10.1109/ACCESS.2020.2980937>.
- Appice, A., Ciampi, A., & Malerba, D. (2015). Summarizing numeric spatial data streams by trend cluster discovery. *Data Mining Knowledge Discover*, 29(1), 84–136.
- Appice, A., Di Mauro, N., Lomuscio, F., & Malerba, D. (2019). Empowering change vector analysis with autoencoding in bi-temporal hyperspectral images. In *MACLEANECMLPKDD Workshop*, (Vol. 2466 pp. 1–10). CEUR Workshop Proceedings.
- Appice, A., Guccione, P., Acciaro, E., & Malerba, D. (2020). Detecting salient regions in a bi-temporal hyperspectral scene by iterating clustering and classification. *Applied Intelligence*, 50(10), 3179–3200.
- Appice, A., Guccione, P., & Di Malerba, N. (2016). Transductive hyperspectral image classification: toward integrating spectral and relational features via an iterative ensemble system. *Machine Learning*, 103(3), 343–375. <https://doi.org/10.1007/s10994-016-5559-7>.
- Appice, A., Guccione, P., & Malerba, D. (2017). A novel spectral-spatial co-training algorithm for the transductive classification of hyperspectral imagery data. *Pattern Recognition*, 63, 229–245. <https://doi.org/10.1016/j.patcog.2016.10.010>.
- Appice, A., & Malerba, D. (2019). Segmentation-aided classification of hyperspectral data using spatial dependency of spectral bands. *ISPRS Journal of Photogrammetry and Remote Sensing*, 147, 215–231.
- Bergstra, J., Yamini, D., & Cox, D.D. (2013). Making a science of model search: hyperparameter optimization in hundreds of dimensions for vision architectures. In *ICML 2013* (pp. 115–123). Omnipress.
- Bruzzone, L., & Prieto, D. F. (2000). Automatic analysis of the difference image for unsupervised change detection. *IEEE Transactions on Geoscience and Remote sensing*, 38(2), 1171–1182.
- Choi, S., Cha, S., & Tappert, C. (2010). A survey of binary similarity and distance measures. *Journal of Systemics. Cybernetics and Informatics*, 8, 43–48.
- Clifton, C. (2003). Change detection in overhead imagery using neural networks. *Applied Intelligence*, 18(2), 215–234. <https://doi.org/10.1023/A:1021942526896>.
- Demšar, J. (2006). Statistical comparisons of classifiers over multiple data sets. *JMLR.org*, 7, 1–30.
- Deng, J. S., Wang, K., Deng, Y., & Qi, G.J. (2008). Pca-based land-use change detection and analysis using multitemporal and multisensor satellite data. *International Journal of Remote Sensing*, 29(16), 4823–4838.

- Du, Z., Jeong, Y. S., Jeong, M. K., & Kong, S.G. (2012). Multidimensional local spatial autocorrelation measure for integrating spatial and spectral information in hyperspectral image band selection. *Applied Intelligence*, 36(3), 542–552. <https://doi.org/10.1007/s10489-010-0274-8>.
- Epting, J., Verbyla, D., & Sorbel, B. (2005). Evaluation of remotely sensed indices for assessing burn severity in interior alaska using landsat tm and etm+. *Remote Sensing of Environment*, 96, 328–339.
- Epting, J., Verbyla, D., & Sorbel, B. (2005). Evaluation of remotely sensed indices for assessing burn severity in interior alaska using landsat tm and etm+. *Remote Sensing of Environment*, 96(3), 328–339. <https://doi.org/10.1016/j.rse.2005.03.002>.
- Falini, A., Tamborrino, C., Castellano, G., Mazzia, F., Mininni, R. M., Appice, A., & Malerba, D. (2020). Machine Learning, Optimization, and Data Science - 6th International Conference, LOD 2020, Siena, Italy, July 19-23, 2020, Revised Selected Papers, Part I, Lecture Notes in Computer Science. In G. Nicosia, V. K. Ojha, E. L. Malfa, G. Jansen, V. Sciacca, P. M. Pardalos, G. Giuffrida, & R. Umeton (Eds.), (Vol. 12565 pp. 113–124). Springer. https://doi.org/10.1007/978-3-030-64583-0_12.
- Ferreira, D. C., Vázquez, F. I., & Zseby, T. (2019). Extreme dimensionality reduction for network attack visualization with autoencoders. In *International joint conference on neural networks* (pp. 1–10).
- Gao, F., Dong, J., Li, B., & Xu, Q. (2016). Automatic change detection in synthetic aperture radar images based on pcanet. *IEEE Geoscience and Remote Sensing Letters*, 13(12), 1792–1796. <https://doi.org/10.1109/LGRS.2016.2611001>.
- Glorot, X., Bordes, A., & Bengio, Y. (2011). Deep sparse rectifier neural networks. In *AISTATS* (pp. 315–323). JMLR.org.
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. New York: MIT Press.
- Guccione, P., Mascolo, L., & Appice, A. (2015). Iterative hyperspectral image classification using spectral-spatial relational features. *IEEE Transactions Geoscience Remote Sensing*, 53(7), 3615–3627. <https://doi.org/10.1109/TGRS.2014.2380475>.
- Helmy, A., & El-Taweel, G. (2010). Neural network change detection model for satellite images using textural and spectral characteristics. *American Journal of Engineering and Applied Sciences*, 3(4), 604–610.
- Hoye, G., & Fridman, A. (2013). The mixel camera – keystone-free hyperspectral images. In *2013 5th workshop on hyperspectral image and signal processing: Evolution in remote sensing (WHISPERS)* (pp. 1–4). <https://doi.org/10.1109/WHISPERS.2013.8080703>.
- Hu, C., Hou, X., & Lu, Y. (2014). Improving the architecture of an autoencoder for dimension reduction. In *11th int. Conf. on ubiquitous intelligence and computing* (pp. 855–858).
- Hussain, M., Chen, D., Cheng, A., Wei, H., & Stanley, D. (2013). Change detection from remotely sensed images: From pixel-based to object-based approaches. *ISPRS Journal of Photogrammetry and Remote Sensing*, 80, 91–106.
- Ilsever, M., & Unsalan, C. (2012). *Two-dimensional change detection methods: remote sensing applications*. Berlin: Springer Science & Business Media.
- Im, J., Jensen, J. R., & Tullis, J.A. (2008). Object-based change detection using correlation image analysis and image segmentation. *International Journal of Remote Sensing*, 29(2), 399–423. <https://doi.org/10.1080/01431160601075582>.
- İrsoy, O., & Alpaydm, E. (2017). Unsupervised feature extraction with autoencoder trees. *Neurocomputing*, 258, 63–73. <https://doi.org/10.1016/j.neucom.2017.02.075>.
- Kalinicheva, E., Sublime, J., & Trocan, M. (2018). Neural network autoencoder for change detection in satellite image time series. In *ICECS 2018* (pp. 641–642). <https://doi.org/10.1109/ICECS.2018.8617850>.
- Kalinicheva, E., Sublime, J., & Trocan, M. (2019). Change detection in satellite images using reconstruction errors of joint autoencoders. In I. V. Tetko, V. Kůrková, P. Karpov, & F. Theis (Eds.) *ICANN 2019: Image Processing* (pp. 637–648). Springer International Publishing.
- Kerner, H. R., Wagstaff, K. L., Bue, B. D., Gray, P. C., Bell, J. F., & Ben Amor, H. (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. <https://doi.org/10.1109/ISTARS.2019.2936771>.
- Key, C. H., & Benson, N. C. (2006). Firemon: Fire effects monitoring and inventory system. <https://www.fs.usda.gov/treesearch/pubs/24066>.
- Khanday, W. (2016). Review of various change detection techniques for hyperspectral images. *Asian Journal of Technology and Management Research (AJTMR)*, 6(2), 39–42.
- Kwan, C. (2019). Methods and challenges using multispectral and hyperspectral images for practical change detection applications. *Information*, 10, 353. <https://doi.org/10.3390/info10110353>.
- Larabi, M., Souleyman, C., Bakhti, K., Kamel, H., & Amine, B. (2019). High-resolution optical remote sensing imagery change detection through deep transfer learning. *Journal of Applied Remote Sensing* 13. <https://doi.org/10.1117/1.JRS.13.046512>.

- Lentile, L., Smith, A., Hudak, A., Morgan, P., Bobbitt, M., Lewis, S., & Robichaud, P. (2009). Remote sensing for prediction of 1-year post-fire ecosystem condition. *18(5)* 594–608. <https://doi.org/10.1071/WF07091>.
- Lopez-Fandino, J., Garea, A. S., Heras, D. B., & Argüello, F. (2018). Stacked autoencoders for multiclass change detection in hyperspectral images. In *2018 IEEE International geoscience and remote sensing symposium, IGARSS 2018* (pp. 1906–1909). IEEE.
- Lopez-Fandino, J., Heras, D. B., Arguello, F., & Duro, R.J. (2017). Cuda multiclass change detection for remote sensing hyperspectral images using extended morphological profiles. In *2017 9th IEEE international conference on intelligent data acquisition and advanced computing systems: Technology and applications (IDAACS)*, (Vol. 1 pp. 404–409).
- López-Fandiño, J. B., Heras, D., Argüello, F., & Dalla Mura, M. (2019). Gpu framework for change detection in multitemporal hyperspectral images. *International Journal Parallel Programm, 47*, 272–292. <https://doi.org/10.1007/s10766-017-0547-5>.
- Lu, D., Mause, P., Brondizio, E., & Moran, E. (2010). Change detection techniques. *International Journal of Remote Sensing, 25*, 2365–2401.
- Meng, R., & Zhao, F. (2017). *Remote sensing of fire effects : A review for recent advances in burned area and burn severity mapping*. (pp. 261–281). US: CRC Press-Taylor & Francis Group.
- Michel, U., Thunig, H., Ehlers, M., & Reinartz, P. (2012). Rapid change detection algorithm for disaster management. *ISPRS - International Archives of the Photogrammetry. Remote Sensing and Spatial Information Sciences, I-4*, 107–111.
- Ming, H., Wenzhong, S., Zhang, H., & Chang, L. (2014). Unsupervised change detection with expectation-maximization-based level set. *IEEE Geoscience and Remote Sensing Letters, 11(1)*, 210–214. <https://doi.org/10.1109/LGRS.2013.2252879>.
- Mouroulis, P., Green, R., & Chrien, T.G. (2000). Design of pushbroom imaging spectrometers for optimum recovery of spectroscopic and spatial information. *Applied Optics, 39(13)*, 2210–2220.
- Najafi, A., Hasanlou, M., & Akbari, V. (2017). Land cover changes detection in polarimetric sar data using algebra, similarity and distance based methods. In *ISPRS - international archives of the photogrammetry remote sensing and spatial Information sciences XLII* (pp. 195–200).
- Oh, D. Y., & Yun, I. D. (2018). Residual error based anomaly detection using auto-encoder in SMD machine sound. *Sensors, 18(5)*, 1308.
- Otsu, N. (1972). A threshold selection method from gray-level histograms. *IEEE Trans. Geoscience and Remote Sensing, 9(1)*, 62–66.
- Penglin, Z., Zhiyong, L., Dan, Z., & Jiangping, C. (2012). A shape similarity based change detection approach of multi-resolution remote sensing images. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, I-7*, 263–266.
- Planinšič, P., & 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. <https://doi.org/10.1109/LGRS.2017.2786344>.
- Sahoo, P., Soltani, S., & Wong, A.C. (1988). A survey of thresholding techniques. *Computer vision. Graphics and Image Processing, 41(2)*, 233–260.
- Sarafijanovic-Djukic, N., & Davis, J. (2019). Fast distance-based anomaly detection in images using an inception-like autoencoder. In *22Nd international conference on discovery science* (pp. 493–508).
- Seydi, S. T., & Hasanlou, M. (2017). A new land-cover match-based change detection for hyperspectral imagery. *European Journal of Remote Sensing, 50(1)*, 517–533. <https://doi.org/10.1080/22797254.2017.1367963>.
- Shi, W., Zhang, M., Zhang, R., Chen, S., & Zhan, Z. (2020). Change detection based on artificial intelligence: State-of-the-art and challenges. *Remote Sensing 12(10)*. <https://doi.org/10.3390/rs12101688>.
- Shone, N., Ngoc, T. N., Phai, V. D., & Shi, Q. (2018). A deep learning approach to network intrusion detection. *IEEE transactions on emerging Topics in computational intelligence, 2(1)*, 41–50.
- Tran, B. N., Tanase, M. A., Bennett, L. T., & Aponte, C. (2018). Evaluation of spectral indices for assessing fire severity in australian temperate forests. *Remote Sensing 10(11)*. <https://doi.org/10.3390/rs10111680>.
- Uzair, M., & Jamil, N. (2020). Effects of hidden layers on the efficiency of neural networks. In *2020 IEEE 23rd international multitopic conference (INMIC)* (pp. 1–6). <https://doi.org/10.1109/INMIC50486.2020.9318195>.
- Vanhoucke, V., Senior, A., & Mao, M.Z. (2011). Improving the speed of neural networks on cpus. In *Deep learning and unsupervised feature learning workshop, NIPS 2011*.
- Wang, J., Liu, S., & Zhang, S. (2015). A novel saliency-based object segmentation method for seriously degenerated images. In *ICIA 2015* (pp. 1172–1177).

- Wang, W., Huang, Y., Wang, Y., & Wang, L. (2014). Generalized autoencoder: a neural network framework for dimensionality reduction. In *conference on computer vision and pattern recognition workshops*, pp. 496–503.
- Wang, Y., Yao, H., Zhao, S., & Zheng, Y. (2015). Dimensionality reduction strategy based on auto-encoder. In *7Th international conference on internet multimedia computing and service*, pp. 1–4.
- Wu, C., Du, B., Cui, X., & Zhang, L. (2017). A post-classification change detection method based on iterative slow feature analysis and bayesian soft fusion. *Remote Sensing of Environment*, 199, 241–255. <https://doi.org/10.1016/j.rse.2017.07.009>.
- Wu, K., Du, Q., Wang, Y., & Yang, Y. (2017). Supervised sub-pixel mapping for change detection from remotely sensed images with different resolutions *Remote Sensing* 9(3). <https://doi.org/10.3390/rs9030284>.
- Yang, Z., & Mueller, R. (2007). Spatial-spectral cross-correlation for change detection : a case study for citrus coverage change detection. In *ASPRS 2007 Annual conference*, (Vol. 2 pp. 767–777).
- Yuan, F., Sawaya, K. E., Loeffelholz, B. C., & Bauer, M.E. (2005). Land cover classification and change analysis of the twin cities (minnesota) metropolitan area by multitemporal landsat remote sensing. *Remote Sensing of Environment*, 98(2), 317–328. <https://doi.org/10.1016/j.rse.2005.08.006>.
- Zheng, J., & Peng, L. (2018). An autoencoder-based image reconstruction for electrical capacitance tomography. *IEEE Sensors Journal*, 18(13), 5464–5474.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Affiliations

Giuseppina Andresini¹  · Annalisa Appice^{1,2} · Daniele Iaia¹ · Donato Malerba^{1,2} · Nicolò Taggio³ · Antonello Aiello³

Annalisa Appice
annalisa.appice@uniba.it

Daniele Iaia
d.iaia@studenti.uniba.it

Donato Malerba
donato.malerba@uniba.it

Nicolò Taggio
taggio@planetek.it

Antonello Aiello
aiello@planetek.it

¹ Department of Informatics, Università degli Studi di Bari Aldo Moro, via Orabona, 4 - 70125 Bari, Italy

² Consorzio Interuniversitario Nazionale per l'Informatica-CINI, Bari, Italy

³ Planetek Italia Srl, Via Massaua, 12, 70132 Bari, Italy