Unsupervised Change Detection Using Convolutional-Autoencoder Multi-resolution Features

Luca Bergamasco, *Student Member, IEEE,* Sudipan Saha, *Member, IEEE,* Francesca Bovolo, *Member, IEEE* and Lorenzo Bruzzone, *Fellow, IEEE*

***Abstract*—The use of Deep Learning (DL) methods for Change Detection (CD) is currently dominated by supervised models that require a large number of labeled samples. However, these samples are difficult to acquire in the multi-temporal case. A possible alternative is leveraging methods that exploit transfer learning for CD by reusing DL models pre-trained for other tasks. However, the performance of the transfer-learning-based models decreases as much as the target images differ from the ones used for training the model. To overcome this limit, we propose an unsupervised CD method that exploits multi-resolution deep feature maps derived by a Convolutional Autoencoder (CAE). It automatically learns spatial features from the input during the training phase without requiring any labeled data. The proposed method processes the bi-temporal images to obtain and compare multi-resolution bi-temporal feature maps. These feature maps are then analyzed by a feature-selection technique to select the most discriminant ones. Furthermore, an aggregated multi-resolution difference image is computed and used for a detail-preserving multi-scale change detection. In the context of this CD approach, we propose two alternative strategies to retrieve multi-scale reliability maps. We tested the proposed method on bi-temporal multispectral images acquired by Landsat-5 and Landsat-8 representing burned areas and Sentinel-2 images representing deforested areas. Results confirm the effectiveness of the proposed CD technique.**

***Index Terms*—Convolutional Autoencoder, Unsupervised Change detection, Deep learning, Multi-temporal Analysis, Unsupervised learning, Remote Sensing.**

I. Introduction

CHANGE detection (CD) identifies changed objects by examining multi-temporal images acquired over the same geographical areas. CD is critical for many applications, such as environmental monitoring [1], [2], fire and burned area detection [3]–[6], and disaster management [7]. Many State-of-the-Art (SoA) methods detect changes by a-priori selecting the best spectral bands to focus on specific changes and/or by using ad-hoc hand-crafted features designed for a single specific scenario and a given sensor. Thus they have to be re-designed whenever any of these factors change. Deep-learning (DL) techniques may alleviate this problem since they can automatically learn features during their training phase. So, it is possible to exploit a trained DL model to perform multiple tasks by processing Remote Sensing (RS) data having characteristics similar to the ones used in the model learning [8].

Some CD methods based on Convolutional Neural Networks (CNNs) [8]–[12] have been proposed to analyze the spatial context information and automatically learn features. The Convolutional-based DL models automatically learn and extract unique spatial features [8], [10], [13] that improve the capability to accurately detect changes [12], [14]. However, most DL techniques are supervised and require a large number of multitemporal labeled training samples [10], [12] that are almost impossible to gather in many multi-temporal applications. Domain-Adaptation (DA) based methods can exploit a DL model trained with given labeled samples (source domain) to process data acquired by another sensor or in another geographical area (target domain). The DA approach can be used to fine-tune a DL model, such as VGG-16, trained using ImageNet, to process multispectral images for CD [15]. In the last years, CD methods using Generative Adversarial Networks (GANs) or adversarial models mitigate the difference between source and target domains thus allowing for multi-sensor CD [16]–[18]. Some DA CD methods transfer the knowledge between heterogeneous RS data [19], [20] or between existing labeled data and unlabeled RS images [21]. However, the DA of models pre-trained with existing labeled data or RS data acquired by sensors different from the target RS images is still a challenging open issue in the literature.

Unsupervised CD DL models can exploit transfer learning to process target RS data with pre-trained models. These methods achieve accurate results when the target images are similar to the ones used to train the model [8], [11], [22], [23]. Performance of the transfer-learning-based methods drops when the difference between target and source images increases [22]. This is a drawback as RS applications usually cover large geographic areas and are associated with wide sensor varieties. Given the lack of multi-temporal labeled datasets, it is seldom possible to obtain a pre-trained network providing informative features for the many possible cases.

An alternative way to perform CD is to exploit unsupervised deep learning methods, which do not require any labeled data for training. These methods obtained accurate results in bi-temporal CD applications [8], [13], [24]. In this context, the autoencoders are used to project the pre-change image into the domain of the post-change one. This projection reduces the impact of differences due to factors (such as atmospheric differences) not related to the changes on the ground [25]. Unsupervised neural networks can extract features to exploit in the CD task [26], [27]. However, most of these methods have limited capability to capture the multi-scale spatial features, which may improve the performance of the change detection

[28].

Multi-scale features allow achieving accurate results in the processing of high-resolution (HR) images since they can model both geometrical details and homogeneous areas. Change detection (CD) methods used many strategies to retrieve and examine multi-scale versions of the same scene, such as pyramid [29], wavelets [30], [31], and object-based methods [32], [33]. The wavelet transform can be combined with Markov Random Fields (MRFs) [34] to select the most informative representations and include the analysis of the spatial context. Liu *et al.* [35] retrieve multi-scale bi-temporal features by exploiting morphological filters. The multi-scale representations retrieved by one of these strategies (*i.e., Discrete Wavelet Transform (DWT)*) can be used to compute a change map preserving the geometrical details of the scene by considering the pixel values at the most reliable scales [30]. The labels of edges are given according to the highest resolution, whereas, in homogeneous areas, the CD exploits low-resolution representations. Most of these methods have limited capability to effectively capture multi-scale information since they depend on shallow features [30], [35]. Thus, they produce a multi-scale version of images by analyzing a limited range of characteristics (i.e., texture, spatial frequencies), but they do not capture semantic information.

To manage the spatial context information while automatically learning features during the unsupervised training, Convolutional Autoencoders (CAEs) can be used. CAEs are DL models [36] showing the characteristics of both standard Autoencoders (AEs) and Convolutional Neural Networks (CNNs). For RS tasks, where the labeled data sets are limited in comparison to other research fields, such as Computer Vision, the CAEs are often pre-trained in an unsupervised way and then fine-tuned with few labeled samples [37]–[39]. CAEs can reduce the need for labeled data since they automatically learn spatial features from the input data during the unsupervised training phase. CAEs are also more efficient in multi-scale feature extraction than other SoA methods, such as DWT. While DWT retrieves only multi-scale representations of images, CAEs capture both the visual information and semantic representation of the input image [40], [41]. In [42], the authors proposed a time-series CD method that exploits a CAE fine-tuned by a pre-trained model to transform a pre-change image into a post-change one and viceversa. The CD technique analyzes only the output images of the fine-tuned CAE without performing any explicit multi-resolution analysis. In [41], the authors trained a CAE with unlabeled samples. They use the CAE to extract multi-resolution feature maps from bi-temporal images and apply a multi-scale CD [30] to retrieve the change map. However, this method uses a static feature selection process that chooses the same number of feature maps from each layer in the network, but the layers of a CAE have a heterogeneous number of filters that can provide a varying number of informative features. Thus this feature selection may lead to the rejection of features containing relevant information about changes or the retention of non relevant ones.

To dynamically select and use all the informative spatial features retrieved by the hidden layers of a CAE, we propose an unsupervised CD method that exploits multi-resolution deep feature maps retrieved by some a-priori chosen layers of a CAE. The CAE is trained in unsupervised manner by using unlabeled samples. After the training phase, we process all the spectral bands of bi-temporal images with the trained model without any a-priori-driven band selection. We then retrieve multi-resolution bi-temporal feature maps and compare them to define multi-resolution difference feature maps. We select only the most informative difference feature maps by applying a dynamic standard-deviation-based feature selection [8], [43]. By this, we choose a dynamic number of feature maps per layer by focusing on those providing relevant information about changes. The selected feature maps are aggregated to retrieve multi-resolution difference images, that emphasize the changes. The multi-resolution difference images are analyzed by a detail-preserving multi-scale CD method inspired by [30], for which we propose two alternatives. The first one adapts [30] to optical passive sensors by calculating the standard deviation instead of the coefficient of variation to compute the reliability of maps, while the second one retrieves the reliability maps by exploiting a gradient-based method.

The paper has the following outline. Section II describes the methodology. Section III presents the experimental settings and the results. Finally, we draw our conclusion in Section IV.

II. Unsupervised change detection based on Convolutional-autoencoder feature extraction

The proposed method aims to perform change detection (CD) in bi-temporal images I_1 and I_2 acquired at time t_1 and t_2, respectively. Let us assume that a set of N unlabeled samples $X = \{X_n, n = 1, ..., N\}$ extracted from I_1 is available. The method uses X to train from scratch a convolutional-autoencoder (CAE) in an unsupervised way. Once the CAE is trained, we process I_1 and I_2 and extract bi-temporal deep feature maps of the images from the model. The feature maps are compared and fused to detect changed (ω_c) and unchanged (ω_{nc}) pixels, where ω_c includes all the relevant changes occurred in the image, while ω_{nc} represents the no change. The block scheme of the proposed method is shown in Fig. 1.

A. Unsupervised CAE training

CAEs have the property to produce an output image that is as similar as possible to the input one by unsupervised learning of spatial context features from a set of unlabeled training samples (X). The CAE contains L convolutional layers divided into an encoder and a decoder. The encoder down-samples the input images and increases the number of feature maps extracted by each layer, whereas the decoder up-samples the output of the encoder and reduces the number of feature maps. The CAE includes strided convolutional (in the encoder) and deconvolutional layers (in the decoder), Batch Normalization (BN) layers, and leaky Rectified Linear Unit (ReLU) activation functions. BN [44] layers normalize the values within a batch during its processing in the model. This normalization helps to increase the learning speed of the model and to stabilize it by reducing the overfitting, thanks

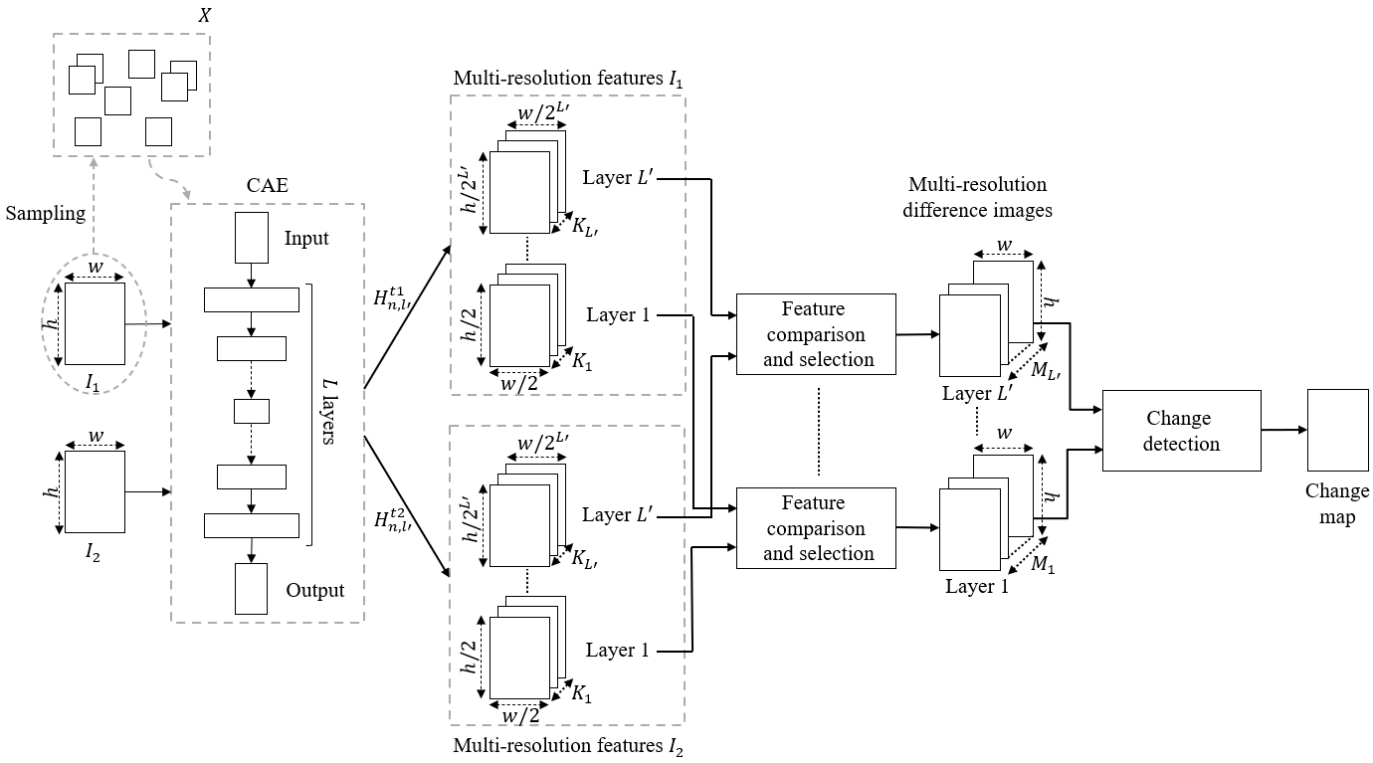


Fig. 1. Block scheme of the proposed unsupervised CD method based on CAE.

to the regularization effect of BN. Leaky ReLU [45] is an improvement of ReLU that keeps the non-linearity of ReLU but improves the handling of negative values. ReLU imposes all negative values to 0. Instead, leaky ReLU transforms them into values close to 0, according to the function $y = \alpha x$. The feature maps of a layer l, where $l = 0, ..., L$, for an input sample $X_n \in X$, where $n = 1, ..., N$, is defined by $H_{n,l} = \phi(W_{l-1,l} H_{n,l-1} + b_l)$, where $H_{n,0} = X_n$, $W_{l-1,l}$ is the weight matrix of the layer l processing the feature maps of the layer $l-1$, $H_{n,l-1}$. b_l represents the biases of layer l, and $\phi(.)$ is the leaky ReLU activation function. The training minimizes a loss function. As in [41], we use a sum-squared-error (SSE) (1) to train our model. The learning process is performed using the back-propagation method that trains the model according to the SSE evaluated between the original samples (X) and the predicted ones from CAE (X'):

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

B. Feature extraction

During the training, the CAE learns visual and semantic features from X. Since the patches included in X are sampled from I_1, it is reasonable to assume that the CAE produces spatial feature maps representing both bi-temporal images I_1 and I_2 since they were acquired over the same geographical area. To extract the feature maps, the CAE separately processes I_1 and I_2, and the bi-temporal multi-resolution feature maps are obtained from $L' = L/2$ layers. The initial layers of the encoder extract simple features (i.e., edges), as shown in previous works [8], whereas the decoder layers retrieve features providing more spatial context information about the change. Hence, we do not exploit encoder layers for CD, and we choose the L' layers composing the decoder. L' also represents the number of scale levels used during the multi-scale CD. Since from the number of CAE layers L depends the number of scale levels L', L should be carefully chosen to find the optimal trade-off between the noise reduction, the informative content of the learned features, and the loss of geometrical details. By processing the bi-temporal images, we retrieve for each layer $l' = 1, ..., L'$ feature maps representing the pre-change image I_1, $H^1_{n,l'}$, and the post-change image I_2, $H^2_{n,l'}$. The feature maps extracted by the layers of the model are in the same feature space, and therefore they are comparable. The bi-temporal feature maps extracted from L' layers produce multi-scale feature maps having heterogeneous spatial dimensions. We apply a bi-linear interpolation method to the features retrieved by the L' layers to obtain multi-scale bi-temporal feature maps having the same spatial dimensions. Corresponding $H^1_{n,l'}$ and $H^2_{n,l'}$ are compared in a multi-resolution fashion to highlight information about changes. The k-th feature maps ($k = 1, \ldots, K_{l'}$) characterizing bi-temporal images retrieved by the layer $l' = 1, ..., L'$ of the model are compared as follows:

$$DF_{n,k,l'} = (H^2_{n,l',k} - H^1_{n,l',k})^2 \tag{2}$$

where

$$l' = 1, \ldots, L'$$
$$k = 1, \ldots, K_{l'}$$

Comparing the bi-temporal feature maps, if no change occurs, the feature maps are similar, and their comparison results in small values close to 0. On the contrary, where the change occurs, the feature maps of I_2 have a different reconstruction of the objects with respect to I_1. Hence, the comparison between the feature maps of I_1 and I_2 enhances the difference due to the changes assuming values far from 0. It is worth noting that the CAE can correctly reconstruct only objects or areas of images learned by X during the training process. If I_2 has a change with structures not available in X, it will be reconstructed in an unpredictable way. However, possible unpredictable reconstructions do not decrease the capability of CD since those structures will be reconstructed in I_2 differently with respect to I_1 in any case.

Now, we have $K_{l'}$ comparisons that emphasize the change between feature maps from I_1 and I_2, respectively, where $K_{l'}$ can be in the order of hundreds. However, not all of them are informative. In each layer l', only a small amount of the difference feature maps $DF_{n,k,l'}$ carries relevant information about the change. Thus the method removes the least informative difference feature maps that provide no or small information about the change. For each layer l', we apply a feature selection (FS) inspired by [8], [43] to filter out the difference feature maps with low probability of providing relevant information about changes. Under the assumption that the presence of the change makes the standard deviation higher than the corresponding situation with no change, we choose the $M_{l'}$ difference feature maps having a standard deviation higher than the other ones since they provide the most relevant information about the change. After ordering them into descending standard deviation, we select difference feature maps until the gap between the standard deviation of two difference feature maps is greater than 10% of the maximum standard deviation value range of layer l':

$$|\sigma(DF_{n,k,l'}) - \sigma(DF_{n,k+1,l'})| > 0.1|\sigma_{max}(DF_{n,l'}) - \sigma_{min}(DF_{n,l'})| \quad (3)$$

In Fig.2, it is possible to observe an example of behavior of the standard deviation values of the difference feature maps of a single layer l' sorted in descending order. The drop in the standard deviation is expected when the change information becomes less represented. In this example, the difference between the standard deviation values of the first two difference feature maps and the third one is larger than 10% of the maximum standard deviation value range, so we assume that the first two feature maps provide the most relevant information about the change. For this reason, they are selected for the CD. This feature selection step can be computed either manually or automatically. It is worth noticing that $M_{l'}$ can be different in each layer l' and can be equal to 0 when the l' has no $DF_{n,l'}$ providing a sufficiently high $|\sigma(DF_{n,k,l'}) - \sigma(DF_{n,k+1,l'})|$. Since the selected feature maps have a high standard deviation, they are the ones with the highest probability of containing changes, and therefore they maximize the probability to detect them. This feature selection process allows analyzing all the spectral bands of the input multispectral RS images without using any a-priori band selection that is often required to maximize the performance of SoA unsupervised methods with respect to the expected kind of change. The proposed feature selection extracts the crucial information content from all the input spectral channels and makes the method independent of the kind of change. We then aggregate the $M_{l'}$ difference feature maps of a considered layer l' to compute the difference image (DI) (4) of that layer.

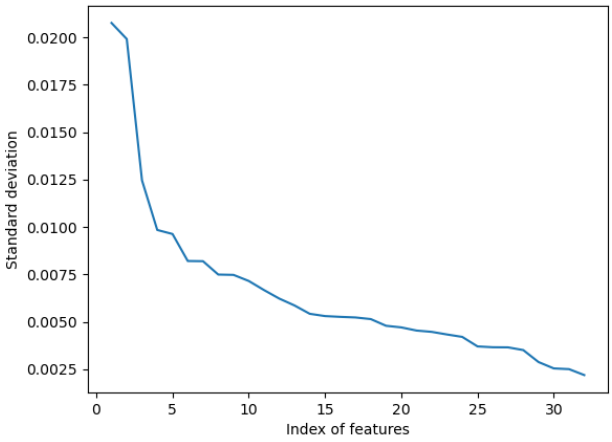


Fig. 2. Behavior of the standard deviation values of the quadratic difference between the feature maps of a layer l' sorted in descending order. In this case only the first two difference feature maps are chosen.

$$DI_{l'} = \sqrt{\sum_{k=1}^{M_{l'}} DF_{k,l'}} \quad (4)$$

where

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

C. Change detection

We compute $DI_{l'}$ for all the L' layers of the CAE having $M_{l'} > 0$ to retrieve L'_{sel} multi-resolution difference images derived from deep features, where L'_{sel} is the number of considered layers and difference images with $M_{l'} > 0$. Thus, we process the L'_{sel} multi-resolution feature maps with a detail-preserving multi-scale approach [30]. This method aims to produce a change map by applying a multi-scale analysis to the multi-resolution $DI_{l'_{sel}}$, where $l'_{sel} = 1, \ldots, L'_{sel}$, to preserve the geometrical details and homogeneous areas and handle the noise of the considered bi-temporal images. The main idea is to associate to each pixel the label ω_c, ω_{nc} of the most reliable level (i.e., the lowest resolution level in which the pixel has a homogeneous behavior). This can be identified by two strategies. In the first strategy, for each layer l'_{sel}, the local standard deviation $\sigma(DI_{l'_{sel}}(i,j))$, computed on pixels included in a moving window centered in (i,j) of $DI_{l'_{sel}}$, is compared with the global standard deviation $\sigma(DI_{l'_{sel}})$ of the DI of that layer to find the most reliable resolution level for the pixel (i,j):

$$\sigma(DI_{l'_{sel}}(i,j)) < \sigma(DI_{l'_{sel}}) \quad (5)$$

This step aims to assign the pixels of high-resolution change maps to the areas with high spatial frequencies along change borders (e.g., the object contours) and the pixels of low-resolution change maps to homogeneous change areas. This method tends to overestimate the areas with high spatial variability thus reducing the performance during the change detection. The second strategy for identifying the most reliable scale level for each pixel in (i, j) based on a gradient-based approach, such as the Canny filter [46]. The Canny filter provides better performance than other edge detection method in many scenarios with different noise conditions [47]. For each of the L'_{sel} layers, the method computes a reliable map ($RM_{l'_{sel}}$) by applying a Canny filter to $DI_{l'_{sel}}$ (6).

$$RM_{l'_{sel}} = Canny(DI_{l'_{sel}}) \tag{6}$$

Each $RM_{l'_{sel}}$ shows the areas with the highest gradient in $DI_{l'_{sel}}$. These areas are the transitions between ω_c and ω_{nc}. These transition areas are the least reliable ones. The strategy keeps the transition areas very thin, so it masks fewer changes in borders than the previous strategy. We average L'_{sel} possible combinations of the L'_{sel} DIs [30]:

$$\overline{DI_{l'_{sel}}} = \frac{1}{l'_{sel}+1} \sum_{h=0}^{l'_{sel}} DI_h, \quad l'_{sel} = 0, 1, ..., L'_{sel} \tag{7}$$

We then retrieve a change map for each resolution level l'_{sel} by applying to each $\overline{DI_{l'_{sel}}}$ a threshold $T_{l'_{sel}}$ retrieved automatically (for the experiments we used an Otsu's threshold [48]). The pixel of the final change map CM in position (i, j) is assigned to ω_c or ω_{nc}, according to the class detected in the pixel position (i, j) of the change map of the resolution level l'_{rel}, where $l'_{rel} = 0, ..., L'_{sel}$ that corresponds to the most reliable resolution level for the position (i, j) [30]:

$$CM(i,j) \in \begin{cases} \omega_{nc}, & \text{if} \overline{DI_{l'_{rel}}}(i,j) \leq T_{l'_{rel}} \\ \omega_c, & \text{if} \overline{DI_{l'_{rel}}}(i,j) > T_{l'_{rel}} \end{cases} \tag{8}$$

III. Experimental design and results

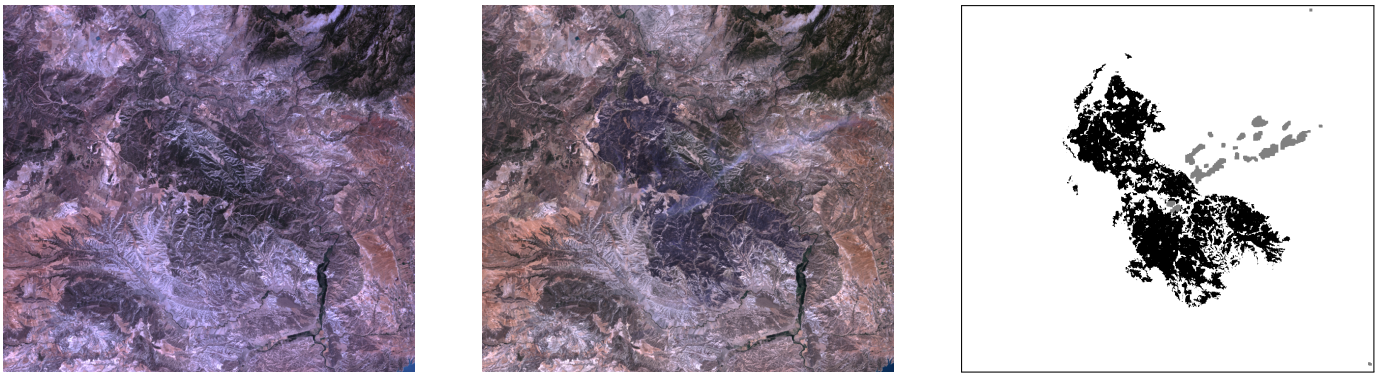
In this Section, we introduce the datasets used to test the proposed method, present the experimental design, and show and discuss the results.

A. Description of Datasets

To test the proposed method, we exploited three multispectral-image datasets. The first dataset is composed of a couple of bi-temporal images having sizes of 861×969 pixels acquired by the Landsat-8 sensor. These images include a burned area near Granada, Spain. They were acquired on June 30^{th}, 2015 (Fig. 3a) and July 16^{th}, 2015 (Fig. 3b) with a spatial resolution of $30\,\mathrm{m/pixel}$. Fig. 3c shows the reference map of the burned area [49] that points out changed (79176) and unchanged pixels (745007). Moreover, areas where reliable labels were not available are identified as "others" (10126 pixels). We exploited the six spectral bands of the Landsat-8 data having $30\,\mathrm{m}$ resolution. The image acquired on June 30^{th}, 2015 is I_1, and we randomly sampled from it an unlabeled dataset composed of patches of 64×64 pixels. We separated this dataset into an unlabeled validation set composed of 5000 patches, and an unlabeled training set X used for the unsupervised learning of CAE composed of $N = 95000$ patches. While any other patch size can be used, larger patch size may lead to fewer number of patches for model training and smaller patch size may lead to inadequate spatial context modeling. To increase N, we sample overlapped patches, so each patch shares a part of its contents with neighboring ones.

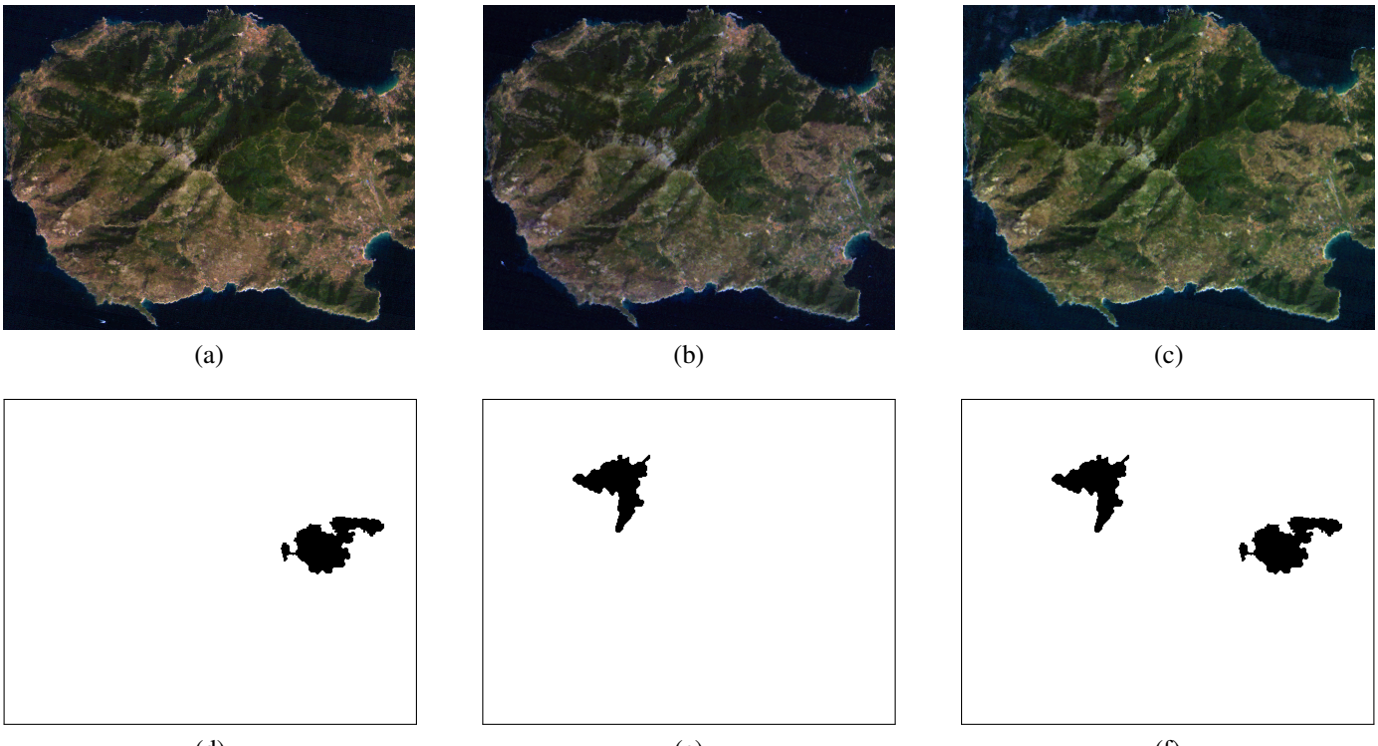
The second dataset is composed of multi-temporal images acquired by Landsat-5 sensor on the Elba island, Italy, in August 1992 (Fig. 4a), August 1994 (Fig. 4b), and September 1994 (Fig. 4c). All the images of the second dataset are acquired with a spatial resolution of $30\,\mathrm{m/pixel}$. We considered this dataset as it is a benchmark in change detection that has been used in many SoA methods [50]–[52]. We designed three scenarios. The first scenario compares the images acquired in August 1992 (I_1) and August 1994 (I_2). The problem consists in the detection of a burned area that contains some vegetation regrowth (see Fig.4d for the reference map containing 2842 changed pixels and 132122 unchanged pixels). The second scenario includes the images acquired in August 1994 (I_1) and September 1994 (I_2), where the method has to detect a burned area on an image acquired shortly after the fire (Fig.4e shows the reference map containing 2414 changed pixels and 132550 unchanged pixels). The third scenario compares the images acquired in August 1992 (I_1) and September 1994 (I_2). The method has to detect two burned areas with different characteristics, an old burned area with vegetation regrowth and a recent one, and to manage the changes that are not correlated with fires (Fig.4f shows the reference change map containing 5256 changed pixels and 129708 unchanged pixels). The images are of 326×414 pixels. For each scenario, we randomly sampled an unlabeled dataset from the pre-change image I_1 composed of patches of 64×64 pixels. In the first and third scenarios, I_1 is the image acquired on August 1992, whereas it is the one acquired on August 1994 in the second scenario. This dataset is separated into a training set used for the unsupervised training composed of $N = 14038$ patches and a validation set of 1400 patches.

The third dataset is composed by a pair of Sentinel-2 images with size 1717×1628 acquired on April 24^{th}, 2016 (Fig. 5a) and January 19^{th}, 2017 (Fig. 5b) with a spatial resolution of $10\,\mathrm{m/pixel}$. These images represent an area deforested for palm oil cultivation in Indonesia. The reference map showing the deforested area (Fig. 5c) points out the changed (73979) and unchanged (2317297) pixels. It also shows the cloud-covered areas identified as "others" (404000) that are excluded from the change detection performance evaluation. We removed the spectral bands with a spatial resolution of $60\,\mathrm{m/pixel}$ since they are sensitive to the atmospheric properties and do not provide any information for the detection of deforested areas. We interpolated the spectral bands at $20\,\mathrm{m/pixel}$ using the nearest neighbor algorithm to obtain images with a homogeneous spatial resolution of $10\,\mathrm{m/pixel}$. We randomly sampled from the image acquired on April 24^{th}, 2016 (I_1) an unlabeled dataset composed of patches of 64×64 pixels. We separated this dataset into an unlabeled training set



(a) (b) (c)

Fig. 3. Bi-temporal Landsat-8 images acquired in an area near Granada, Spain on (a) June 30^{th}, 2015, (b) July 16^{th}, 2015, and (c) the related reference map of the burned area. The white pixels represent no changes, the black pixels the changes, and the grey ones the pixels where reliable reference data are not available.



(a) (b) (c)

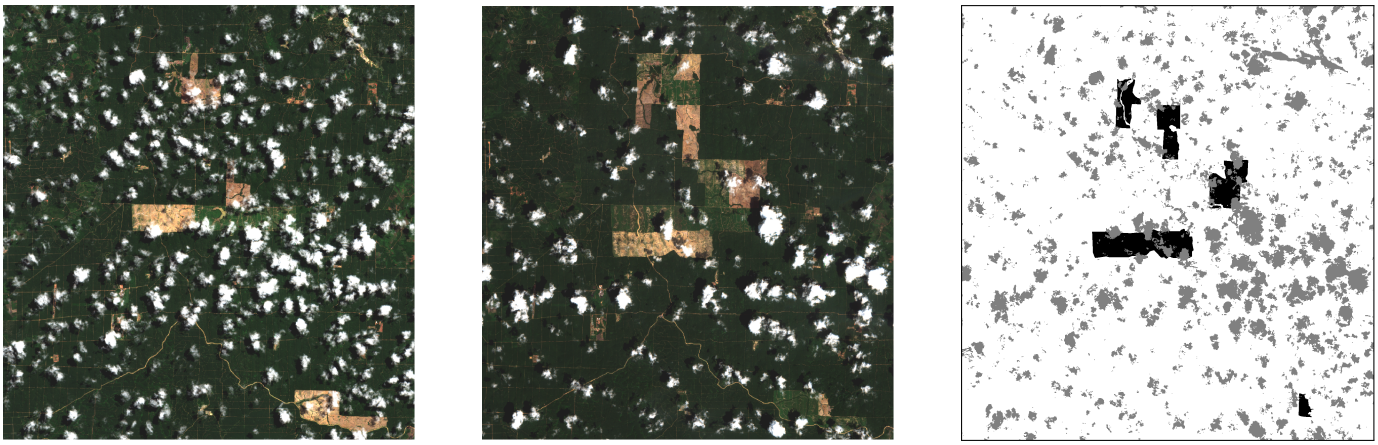
(d) (e) (f)

Fig. 4. Multi-temporal Landsat-5 images acquired on the Elba island, Italy in (a) August 1992, (b) August 1994, and (c) September 1994. Reference maps of the changes that occurred between (d) August 1992 and 1994, (e) August 1994 and September 1994, and (f) August 1992 and September 1994. The white pixels represent no changes, and the black ones the changes.

X composed of 12305 patches and a validation set composed of 1367 patches. Given the limited spatial dimensions of the images, we increased the number of samples by overlapping the patches.

B. Design of Experiments

We performed an unsupervised training of a CAE for each dataset (and scenario) with a number of epochs E. The CAEs were trained for image reconstruction, (which is a different objective from our target task (Change Detection)) thus the parameters (e.g., the number of layers) were entirely based on this task. We used the validation set to assess the quality of reconstructed images during the training by computing the validation loss. When the validation loss diverges from the training one, the model overfits the data reconstruction, and this may affect CD performance. Multi-scale features generated from the CAE after training were used for the CD task. In our experiments we used leaky ReLU activation functions with $\alpha = 0.2$. We trained the model using a batch size $bs = 100$ and a learning rate $lr = 10^{-4}$. We did several experiments on the proposed method in both datasets using different set-ups:



(a) (b) (c)

Fig. 5. Bi-temporal Sentinel-2 images acquired in Indonesia on (a) April 24^{th}, 2016, (b) January 19^{th}, 2017, and (c) the related reference map of the deforested area. The white pixels represent no changes, the black pixels the changes, and the grey ones the cloud-covered areas.

1) Experiment 1: The experiment objective was the analysis of the proposed CD method performance by varying the number of model layers. This hyperparameter needs to be fixed in any method based on CAEs or DL in general. For this test, we used two datasets acquired by different sensors, Landsat-8 and Sentinel-2, to analyze the effect of data with heterogeneous spatial and spectral resolutions and changes with different shapes. The Granada dataset acquired by Landsat-8 includes jagged changed areas providing an interesting test case to study the geometrical-details preservation and informative-feature generation. The Indonesia dataset acquired by Sentinel-2 contains changed areas with more regular borders than the other dataset, thus resulting less critical in terms of geometrical-detail preservation. To choose the optimal number of layers of the CAE in the Granada dataset and thus the number of multi-scale levels L', we compared the CD performance using a CAE with $L = 2, 4, 6, 8$ layers trained with $E = 50$. For the standard deviation reliability approach, we used a moving window size of 5×5. For the Indonesia dataset, we used the same test settings as for the Granada one, but we exploited a CAE trained for $E = 250$. The window size and the number of epochs are fixed according to the results achieved in Experiments 2 and 3, respectively. We can observe the number of CAE layers that results in a good trade-off between de-noising of input data and generation of informative feature maps.

2) Experiment 2: This experiment analyzed the CD performance using various moving window sizes in the standard deviation reliability approach. We tested window dimensions of 3×3, 5×5, and 7×7. We fixed the number of layers and E according to the results obtained in experiments 1 and 3, respectively. The number of layers and training epochs depends on the considered dataset:

- $L = 6$ and $E = 50$ for the Granada dataset;
- $L = 6$ for the Elba dataset with $E = 300$ for the first scenario, $E = 150$ for the second one, and $E = 200$ for the third one;
- $L = 4$ and $E = 250$ for the Indonesia dataset.

3) Experiment 3: The experiment objective consisted in the analysis of the CD performance by varying the number of training epochs. It is worth noting that the number of training epochs E is a hyperparameter that has to be a-priori fixed for CAE- and DL-based methods. We analyzed the performance by varying the number of epochs $E = 50, 100, ..., 300$ to train a CAE with $L = 6$ layers for the Granada and Elba datasets and with $L = 4$ for the Indonesia one (Experiment 1). We used a moving window size of 5×5 for the Granada and Indonesia datasets and 3×3 for the Elba scenarios (Experiment 2). We tested our method by considering the two strategies based on standard deviation and Canny filter [46] (Section II-C) to find the most reliable scale levels.

4) Experiment 4: The Experiment verified that the multi-scale feature concatenation and the feature selection are effective. We compared the proposed method against two limit variants. The first one used only the feature maps extracted from a single layer: the bottleneck one. The second one used all the feature maps from all L' layers without any feature selection. For this experiment, we used a CAE with $L = 6$ layers for the Granada and Elba datasets and with $L = 4$ for the Indonesia one (Experiment 1). For the Granada dataset, the method exploited a CAE trained for $E = 50$ and a moving window size of 5×5. For the Elba dataset, we used a window size of 3×3 and a CAE trained with $E = 300$ for the first scenario, $E = 150$ for the second one, and $E = 200$ for the third one. For the Indonesia dataset, the method exploited a CAE trained for $E = 250$ and the Canny-based reliability approach. The moving window size and the training epochs are fixed according to the results achieved in Experiments 2 and 3, respectively.

SoA comparison was performed against i) the adaptive Change Vector Analysis (CVA) [50], ii) the CVA performing the analysis of the contextual information with a Markov-Random-Fields (MRF) based method [51], iii) the adaptive semiparametric and context based CVA [52], and iv) a self-supervised segmentation method [22]. For i) and iii), we used NIR, SWIR1, and SWIR2 to detect burned areas, whereas,

TABLE I
FAs, MAs, TPs, OE, SENSITIVITY AND SPECIFICITY OF THE PROPOSED METHOD VS THE NUMBER OF LAYERS L IN THE GRANADA DATASET TRAINED FOR $E = 50$.

L	FA	MA	TP	OE	Sensitivity	Specificity
2	**1.21%**	29.99%	55435	3.97%	70.01%	**98.79%**
4	3.58%	23.73%	60385	5.51%	76.27%	96.42%
6	2.38%	**8.71%**	**72279**	**2.99%**	**91.29%**	97.62%
8	3.06%	8.72%	72275	3.6%	91.28%	96.94%

for the MRF-based one, we exploited NIR and SWIR2. As well-known in the literature, this is done since CVA benefits from selecting the most relevant spectral bands. The a-priori band selection favors the SoA methods with respect to the proposed one that, in turn, ingests all the features and allows us to demonstrate the effectiveness in feature selection. To evaluate the performance with respect to the expected kind of change, we considered the number of the true detected change (TPs), the false-alarm rate (FAs), the missed-alarm rate (MAs), the overall errors (OE), the sensitivity (TP/(TP+FN)) and the specificity (TN/(TN+FP)).

C. Experiment 1: Analysis of the performance varying the number of layers

Experiment 1 examined the number of CAE layers that results in a good trade-off between de-noising of input data and generation of informative feature maps. In the Granada dataset, the proposed method achieved the best trade-off between FAs and MAs among the examined cases using $L = 6$ layers (Table I). With $L = 6$ layers, the proposed method resulted in the lowest OE, and found the most of change areas by keeping the number of FAs relatively low. With $L = 2$, the method detected the smallest number of FAs, but it found fewer changed areas (i.e., 55435) than in the case with $L = 6$ (i.e., 72279) since the features learned by two layers provide less information than the ones learned by a deeper model. On the contrary, the proposed method using $L = 8$ detected a number of change areas as $L = 6$, but it detected more FAs (3.06% vs. 2.38%). A CAE with $L = 8$ provided informative feature maps at the cost of missing more geometrical details (Fig. 6d) than the one with $L = 6$ (Fig. 6c) due to the compression of the spatial information. The choice of the values of L depends on the spatial resolution of input images. The higher the spatial resolution is, the more L can increase. Very high spatial resolution images require more layers to cover a larger receptive field and capture more spatial features. In the Indonesia dataset, the CAE with $L = 4$ provided the best CD performance. $L = 2$ achieved similar results to $L = 4$ (Table II), but $L = 2$ minimized MAs. $L = 4$ minimized FAs and OE. A DL model with many layers increases its receptive fields and thus the capability to capture spatial features. This characteristic improved the performance when dealing with HR images as the Sentinel-2 Indonesia dataset. The CD performance decreased using a CAE with $L = 6$ and $L = 8$ because of the small number of samples that limits the capabilities in training the large number of parameters of the model incurring in overfitting. Thus from now on the CAE analyzing the Granada and Elba dataset is always composed of three layers for both the encoder and the decoder (Table III). The CAE that processes the Indonesia dataset is composed of two layers for encoder and decoder (Table IV).

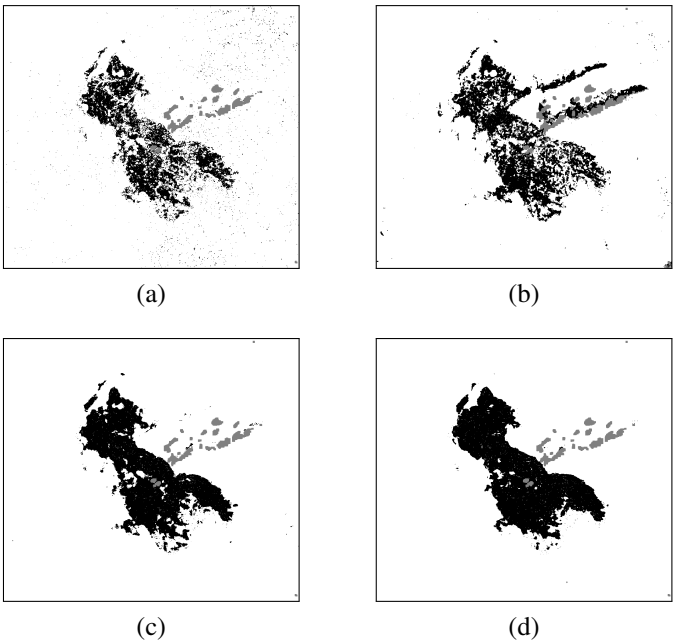


Fig. 6. Change maps computed using (a) $L' = 1$ of the CAE with $L = 2$, (b) $L' = 2$ of the CAE with $L = 4$, (c) $L' = 3$ of the CAE with $L = 6$, and (d) $L' = 4$ of the CAE with $L = 8$. All the CAEs are trained for $E = 50$ using the Granada dataset. Standard-deviation-based reliability approach was used to find the most reliable areas. (The white pixels represent no changes, the black pixels the changes, and the grey ones are no data.)

TABLE II
FAs, MAs, TPs, OE, SENSITIVITY AND SPECIFICITY OF THE PROPOSED METHOD VS THE NUMBER OF LAYERS L IN THE INDONESIA DATASET TRAINED FOR $E = 250$.

L	FA	MA	TP	OE	Sensitivity	Specificity
2	6.0%	**13.14%**	**64261**	6.21%	**86.86%**	94.01%
4	**5.3%**	19.82%	59317	**5.75%**	80.18%	**94.7%**
6	16.64%	13.38%	64084	16.54%	86.62%	83.36%
8	17.41%	14.65%	63139	17.32%	85.35%	82.59%

D. Experiment 2: Analysis of the performance varying the window size of the standard-deviation-based reliability approach

Experiment 2 studied the performance of the proposed method by varying the size of the moving window. In the Granada dataset, the window size of 3×3 maximized the detection of the changed areas and the sensitivity, whereas the size of 7×7 minimized the FAs. However, the window size of 5×5 is the optimal trade-off between the other two cases. The proposed method using a window size of 5×5 detected a similar number of changed areas with respect to the case with a window size 3×3, but fewer FAs. For this reason, we chose to use 5×5 window size for the Granada dataset experiments. The window size affects the reliability of the changed-area borders. The smaller the window size, the less reliable the local standard deviation, and the higher the FA number. The larger the window size, the lower the sensitivity

TABLE III
STRUCTURE OF THE CONVOLUTIONAL AUTOENCODER USED DURING TESTS FOR THE GRANADA AND THE ELBA DATASET.

Layer type	Kernel size	Output size
Input	-	$64 \times 64 \times 6$
Strided Conv2D	$5 \times 5 \times 32$	$32 \times 32 \times 32$
Batch Normalization	-	$32 \times 32 \times 32$
Leaky ReLU	-	$32 \times 32 \times 32$
Strided Conv2D	$5 \times 5 \times 64$	$16 \times 16 \times 64$
Batch Normalization	-	$16 \times 16 \times 64$
Leaky ReLU	-	$16 \times 16 \times 64$
Strided Conv2D	$5 \times 5 \times 128$	$8 \times 8 \times 128$
Batch Normalization	-	$8 \times 8 \times 128$
Leaky ReLU	-	$8 \times 8 \times 128$
Deconv2D	$5 \times 5 \times 64$	$16 \times 16 \times 64$
Batch Normalization	-	$16 \times 16 \times 64$
Leaky ReLU	-	$16 \times 16 \times 64$
Deconv2D	$5 \times 5 \times 32$	$32 \times 32 \times 32$
Batch Normalization	-	$32 \times 32 \times 32$
Leaky ReLU	-	$32 \times 32 \times 32$
Deconv2D	$5 \times 5 \times 6$	$64 \times 64 \times 6$

TABLE IV
STRUCTURE OF THE CONVOLUTIONAL AUTOENCODER USED DURING TESTS FOR THE INDONESIA DATASET.

Layer type	Kernel size	Output size
Input	-	$64 \times 64 \times 6$
Strided Conv2D	$5 \times 5 \times 32$	$32 \times 32 \times 32$
Batch Normalization	-	$32 \times 32 \times 32$
Leaky ReLU	-	$32 \times 32 \times 32$
Strided Conv2D	$5 \times 5 \times 64$	$16 \times 16 \times 64$
Batch Normalization	-	$16 \times 16 \times 64$
Leaky ReLU	-	$16 \times 16 \times 64$
Deconv2D	$5 \times 5 \times 32$	$32 \times 32 \times 32$
Batch Normalization	-	$32 \times 32 \times 32$
Leaky ReLU	-	$32 \times 32 \times 32$
Deconv2D	$5 \times 5 \times 6$	$64 \times 64 \times 6$

to the geometrical details. The local standard deviation using a moving window of dimension 5×5 proved to be a good trade-off between the reliability of the statistic and sensibility. However, the performance using window sizes of 3×3 and 5×5 are similar, so the choice of the window size between these values is not critical.

In the Elba dataset, we observed the same behavior for the three scenarios (Tabs. VI, VII, and VIII). The local standard deviation with a window size of 3×3 maximized the detection of the changed area with a relatively low number of FAs and OE. In this dataset, a window size larger than 3×3 is too sensitive to the transition areas between ω_c and ω_{nc}. Increasing the size reduced the FAs but increased the MAs with an underestimation of the changed areas (Fig. 7). Since the 3×3 window used to compute the local standard deviation allowed detecting more changed areas than the other window sizes with a low number of FAs, we chose to use this window size for the Elba dataset.

In the Indonesia dataset, we achieved similar CD performance with all the window size values (Table IX). The CD methods minimized the MAs using a window size of 3×3. However, it increased the number of FAs in the changed area borders. We chose using a window size of 5×5 during the tests since it minimized the OE.

TABLE V
FAS, MAS, TPS, OE, SENSITIVITY AND SPECIFICITY OF THE PROPOSED METHOD BY VARYING THE WINDOW SIZE WITH THE STANDARD-DEVIATION-BASED RELIABILITY APPROACH IN THE GRANADA DATASET. THE RESULTS WERE RETRIEVED BY USING $L' = 3$ MULTI-SCALE FEATURE MAPS EXTRACTED BY A CAE TRAINED FOR $E = 50$.

Window size	FA	MA	TP	OE	Sensitivity	Specificity
3×3	2.44%	**8.31%**	**72596**	3.01%	**91.69%**	97.56%
5×5	2.38%	8.71%	72279	2.99%	91.29%	97.62%
7×7	**2.27%**	9.78%	71430	**2.99%**	90.22%	**97.73%**

TABLE VI
FAS, MAS, TPS, OE, SENSITIVITY AND SPECIFICITY OF THE PROPOSED METHOD BY VARYING THE WINDOW SIZE IN THE STANDARD-DEVIATION-BASED RELIABILITY APPROACH. RESULTS WERE RETRIEVED WITH IMAGES BY USING $L' = 3$ MULTI-SCALE FEATURE MAPS EXTRACTED BY A CAE TRAINED FOR $E = 300$ (ELBA DATASET ACQUIRED IN AUGUST 1992 AND AUGUST 1994).

Window size	FA	MA	TP	OE	Sensitivity	Specificity
3×3	0.42%	**15.13%**	**2412**	**0.73%**	**84.87%**	99.58%
5×5	0.32%	24.84%	2136	0.84%	75.16%	99.68%
7×7	**0.25%**	39.51%	1719	1.08%	60.49%	**99.75%**

E. Experiments 3 and 4: Granada dataset

Experiment 3 showed that the method accuracy is not significantly affected by the number of epochs E. The strategy exploiting the Canny filter had a slightly higher number of OE than the one using the standard deviation (Table X). On this dataset, the method achieved the best trade-off in terms of TPs, MAs, and FAs using a model trained for $E = 50$ that was exploited to process the bi-temporal images and retrieve the final change map. The proposed method performed better than the SoA ones. It detected the changes in a more accurate way than SoA methods (Table XI). It reduced the MAs by using the standard-deviation-based strategy with respect to CVA-MRF [51] of 3.73%. The proposed method detected more MAs than the semiparametric CVA using MRF [52], but it sharply reduced the OE from 40567 of [52] to 24637. Both the proposed method and the self-supervised segmentation detected the changed areas with similar performance. However, the proposed method found fewer FAs (i.e., 2.38%) than the self-supervised segmentation (i.e., 3.9%). The strategy using the standard deviation overestimated the changes, so it detected more FAs than the strategy using the Canny filter.

Since most of the FAs were on the changed-area borders, the detection of less FAs depended on the higher accuracy of the Canny filter in identifying areas of transition between changed and unchanged regions. The standard-deviation-based strategy (Fig. 8a) is less accurate (Fig. 8b).

Quantitative results (Table XI) are confirmed by the qualitative ones (Fig. 9). We can see that the proposed method with both strategies (Fig. 9h, 9i) provided better maps than the SoA methods. The adaptive CVA (Fig. 9b), the self-supervised segmentation (Fig. 9e), and the semiparametric CVA using MRF (Fig. 9d) detected most of the changed areas, also the small ones (i.e., the lower part of the changed area), but showed many FAs, especially the semiparametric one. It is

TABLE VII
FAs, MAs, TPs, OE, Sensitivity and Specificity of the proposed method by varying the window size in the standard-deviation-based reliability approach. Results were retrieved with images by using $L' = 3$ multi-scale feature maps extracted by a CAE trained for $E = 150$ (Elba dataset acquired in August 1994 and September 1994).

Window size	FA	MA	TP	OE	Sensitivity	Specificity
3×3	0.08%	**36.7%**	**1528**	**0.74%**	**63.3%**	99.92%
5×5	0.04%	64.08%	867	1.18%	35.92%	99.96%
7×7	**0.02%**	79.25%	501	1.44%	20.75%	**99.97%**

TABLE VIII
FAs, MAs, TPs, OE, Sensitivity and Specificity of the proposed method by varying the window size in the standard-deviation-based reliability approach. Results were retrieved with images by using $L' = 3$ multi-scale feature maps extracted by a CAE trained for $E = 200$ (Elba dataset acquired in August 1992 and September 1994).

Window size	FA	MA	TP	OE	Sensitivity	Specificity
3×3	0.65%	**13.76%**	**4533**	**1.16%**	**86.24%**	99.35%
5×5	0.51%	26.5%	3863	1.52%	73.5%	99.49%
7×7	**0.39%**	38.96%	3208	1.89%	61.04%	**99.61%**

worth recalling that for the CVA-based SoA methods, an a-priori band selection was performed before the processing as a function of the expected kind of change (if all the spectral channels are used, the accuracy is degraded). The feature-selection step performed by the proposed method chose the features in an automatic and unsupervised way with no need for a-priori information and detected most of the changes with fewer FAs than CVA-based SoA methods. During the tests, we observed that the feature selection chose, for each layer, few feature maps (seven at most) providing the most relevant information about the change. For some layers, it did not select any difference feature maps since none of them had sufficiently high standard deviation values. From the qualitative viewpoint, we can notice that the use of a standard-deviation-based strategy provided a more homogeneous map than the one using the Canny filter. The latter strategy retrieved changed areas with more precise borders but created some artifacts due to the value of the standard deviation of the Gaussian kernel used in the Canny filter.

Experiment 4 demonstrated the effectiveness of the multi-scale analysis by comparing the results achieved using a detail-preserving multi-scale CD [30] with the one using single-scale feature maps. The use of multi-scale feature maps allowed the proposed method to detect almost all the changes, even the small ones (Fig. 9h), whereas the single-scale feature-map option detected only the wider changed areas and lost all the small ones. If we look at Table XI, we can observe that the FAs are slightly low with single-scale feature maps because of the smoothing effect of the deeper layers, but the MAs increase of 17.04% with respect to the multi-scale feature case. Hence the use of a multi-scale analysis allowed preserving the geometrical details. As last analysis, we observed that the proposed method using the feature selection step achieved better results than without feature selection (Table XI). It resulted in fewer FAs. This is supported by qualitative analysis, e.g., compare the bottom part of the changed area in Fig. 9h) and Fig. 9g. Hence the feature selection removed the feature maps providing less information about the change, thus

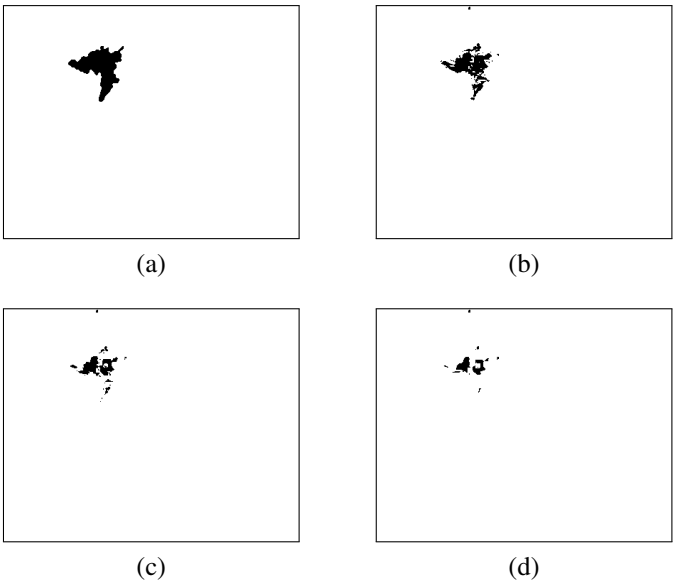


Fig. 7. Change maps retrieved by using a moving window size of (b) 3×3, (c) 5×5, and (d) 7×7 pixels in the standard-deviation-based reliability approach. The change maps are obtained by using $L' = 3$ multi-scale feature maps extracted by a CAE trained for $E = 50$ with the Elba dataset (August 1994). The change maps are retrieved by processing the images acquired in August 1994 and September 1994 and can be compared to the reference map (the white pixels represent no changes, and the black ones the changes).

TABLE IX
FAs, MAs, TPs, OE, Sensitivity and Specificity of the proposed method by varying the window size in the standard-deviation-based reliability approach. Results were retrieved with images by using $L' = 2$ multi-scale feature maps extracted by a CAE trained for $E = 250$ (Indonesia dataset).

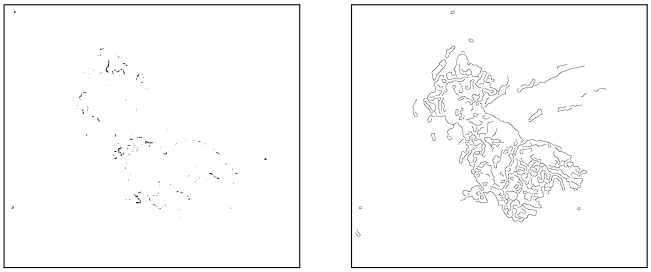
Window size	FA	MA	TP	OE	Sensitivity	Specificity
3×3	5.37%	**19.79%**	**59342**	5.81%	**80.21%**	94.63%
5×5	**5.3%**	19.82%	59317	**5.75%**	80.18%	**94.7%**
7×7	**5.3%**	19.82%	59317	**5.75%**	80.18%	**94.7%**

TABLE X
FAs, MAs, TPs, OE, Sensitivity and Specificity of the proposed method vs the number of epochs on the CAE training and by using a standard-deviation-based and Canny-filter-based approach to find the most reliable areas (Granada dataset).

E	Reliability approach	FA	MA	TP	OE	Sens.	Spec.
50	Canny	2.14%	14.16%	67962	3.3%	85.83%	97.86%
	Std. dev.	2.38%	**8.71%**	**72279**	2.99%	**91.29%**	97.62%
100	Canny	2.64%	21.52%	62138	4.45%	78.48%	97.36%
	Std. dev.	2.82%	19.61%	63650	4.43%	80.39%	97.18%
150	Canny	**1.77%**	16.54%	66081	3.19%	83.46%	**98.23%**
	Std. dev.	1.98%	11.67%	69935	**2.91%**	88.33%	98.02%
200	Canny	2.71%	19.06%	64088	4.28%	80.94%	97.29%
	Std. dev.	3.08%	15.4%	66984	4.26%	84.6%	96.92%
250	Canny	2.59%	20.36%	63055	4.3%	79.64%	97.41%
	Std. dev.	2.77%	16.51%	66102	4.09%	83.49%	97.23%
300	Canny	2.1%	15.34%	67033	3.37%	84.66%	97.9%
	Std. dev.	2.02%	13.56%	68438	3.13%	86.44%	97.98%

TABLE XI
FAs, MAs, SENSITIVITY, SPECIFICITY AND OE (IN NUMBER OF PIXELS AND PERCENTAGE) OBTAINED BY THE SOA METHODS AND THE PROPOSED ONE. WE TESTED THE PROPOSED METHOD USING THE TWO RELIABILITY APPROACHES, A SINGLE-SCALE FEATURE, AND NO FEATURE SELECTION (GRANADA DATASET).

Method	FA	MA	Sens.	Spec.	OE	
					Pixels	%
Adaptive CVA [50]	2.69%	12.49%	87.51%	97.31%	29927	3.59%
CVA + MRF [51]	2.1%	12.44%	87.56%	97.9%	25523	3.06%
Semipar. CVA + MRF [52]	5.36%	**0.77%**	**99.23%**	94.64%	40567	4.86%
Self-supervised segm. [22]	3.42%	8.4%	91.6%	96.58%	32145	3.9%
Proposed w/ std. dev.	2.38%	8.71%	91.29%	97.62%	**24637**	**2.95%**
Proposed w/ Canny	2.14%	14.16%	85.84%	97.86%	27187	3.26%
Proposed w/ single-scale feat.	**1.92%**	25.75%	74.25%	**98.08%**	34683	4.21%
Proposed no feat. sel.	3.39%	18.95%	81.05%	96.61%	40296	4.83%



(a) (b)

Fig. 8. Reliability maps retrieved using the standard-deviation-based strategy on the resolution level (a) $l'_{sel} = 1$ (the highest spatial resolution), and the Canny-filter-based strategy on (b) $l'_{sel} = 1$. These maps are derived from the feature maps retrieved by two layers of the CAE trained for $E = 50$ epochs using the Granada dataset. (The white pixels represent reliable areas, the black ones the no reliable areas at $l'_{sel} = 1$.)

improving the detection.

F. Experiments 3 and 4: Elba dataset

On this dataset, we tested the proposed method by using the three scenarios described is Section III-A. Experiment 3 evaluated the method against E in three scenarios to observe its robustness and the range of epochs in which it achieves stable performance. We used the reference maps to estimate the performance of the method during the tests. By increasing E, the accuracy of the method improved by decreasing the FA rate until a specific epoch before dropping due to the overfitting. At $E = 300, 150, 200$, the method achieved the best trade-off between OE, MAs, and FAs for the first (Tab XII), the second (Table XIII), and the third scenario (Table XIV), respectively. To test the first and the third scenario, we used a model trained for $E = 300$ and $E = 200$, respectively. For the second scenario, we used a model trained for $E = 150$. As in Section III-E, we tested our method using both the strategies to retrieve the reliability maps. As we can observe in Tabs. XII, XIII, and XIV, the method provided differences in

TABLE XII
FAs, MAs, TPs, OE, SENSITIVITY AND SPECIFICITY OF THE PROPOSED METHOD VS THE NUMBER OF EPOCHS ON THE CAE TRAINING AND BY USING A STANDARD-DEVIATION-BASED AND CANNY-FILTER-BASED APPROACH TO FIND THE MOST RELIABLE AREAS (IMAGES OF ELBA DATASET ACQUIRED IN AUGUST 1992 AND AUGUST 1994).

E	Reliability approach	FA	MA	TP	OE	Sens.	Spec.
50	Canny	0.6%	15.27%	2408	0.91%	84.72%	99.4%
	Std. dev.	0.64%	**9.82%**	**2563**	0.84%	**90.18%**	99.35%
100	Canny	0.47%	25.12%	2128	0.98%	74.88%	99.53%
	Std. dev.	0.51%	20.87%	2249	0.94%	79.13%	99.49%
150	Canny	0.56%	17.87%	2334	0.93%	82.12%	99.44%
	Std. dev.	0.6%	12.84%	2477	0.86%	87.16%	99.4%
200	Canny	0.43%	22.77%	2195	0.9%	77.23%	99.57%
	Std. dev.	0.43%	18.58%	2314	0.82%	81.42%	99.57%
250	Canny	0.42%	22.77%	2195	0.89%	77.23%	99.58%
	Std. dev.	0.48%	16.89%	2362	0.82	83.11%	99.52%
300	Canny	**0.4%**	19.18%	2297	0.8%	80.82%	**99.6%**
	Std. dev.	0.42%	15.13%	2412	**0.73%**	84.87%	99.58%

TABLE XIII
FAs, MAs, TPs, OE, SENSITIVITY AND SPECIFICITY OF THE PROPOSED METHOD VS THE NUMBER OF EPOCHS ON THE CAE TRAINING AND BY USING A STANDARD-DEVIATION-BASED AND CANNY-FILTER-BASED APPROACH TO FIND THE MOST RELIABLE AREAS (IMAGES OF ELBA DATASET ACQUIRED IN AUGUST 1994 AND SEPTEMBER 1994).

E	Reliability approach	FA	MA	TP	OE	Sens.	Spec.
50	Canny	0.02%	23.24%	1853	0.44%	76.76%	99.98%
	Std. dev.	0.02%	39.44%	1492	0.72%	60.56%	99.98%
100	Canny	0.01%	33.89%	1596	0.62%	66.11%	99.99%
	Std. dev.	0.01%	43.5%	1364	0.78%	56.5%	99.99%
150	Canny	0.02%	**23.2%**	**1854**	**0.43%**	**76.8%**	99.98%
	Std. dev.	0.08%	36.7%	1528	0.74%	63.3%	99.92%
200	Canny	0.02%	28.21%	1733	0.53%	71.79%	99.98%
	Std. dev.	0.02%	44.74%	1334	0.82%	55.26%	99.98%
250	Canny	0.04%	25.14%	1807	0.48%	74.86%	99.96%
	Std. dev.	0.02%	52.9%	1137	0.96%	47.1%	99.98%
300	Canny	0.02%	40.27%	1442	0.74%	59.73%	99.98%
	Std. dev.	**0.002%**	53.23%	1129	0.95%	46.77%	**99.99%**

terms of FAs and MAs by exploiting the two strategies used to retrieve the reliability maps. By using the standard-deviation-based strategy, it minimized the MAs, while detecting more FAs. By using the Canny-filter-based strategy, the method reduced the FAs with a slight increase of the MAs.

We can notice that the proposed method provided comparable or better results than SoA methods in all the scenarios (Tables XV, XVI, XVII). It generally detected most of the changed areas with a reduction of the OE with respect to the SoA methods. Only in the second scenario, the SoA methods achieved better performance than the proposed one. However, the SoA method performance dropped in the other two scenarios, even with an a-priori spectral-band selection. Overall, the SoA technique results in unstable performance through the scenarios, whereas the proposed method achieved stable results in all of them without the need to use any a-priori band-selection process.

The qualitative analysis of the CD maps (Fig. 10, 11, 12) confirmed the quantitative results. The SoA method results varied through different scenarios. In the second scenario (Fig. 11), the SoA methods detected the changed area slightly more accurately than the proposed one. However, in scenarios

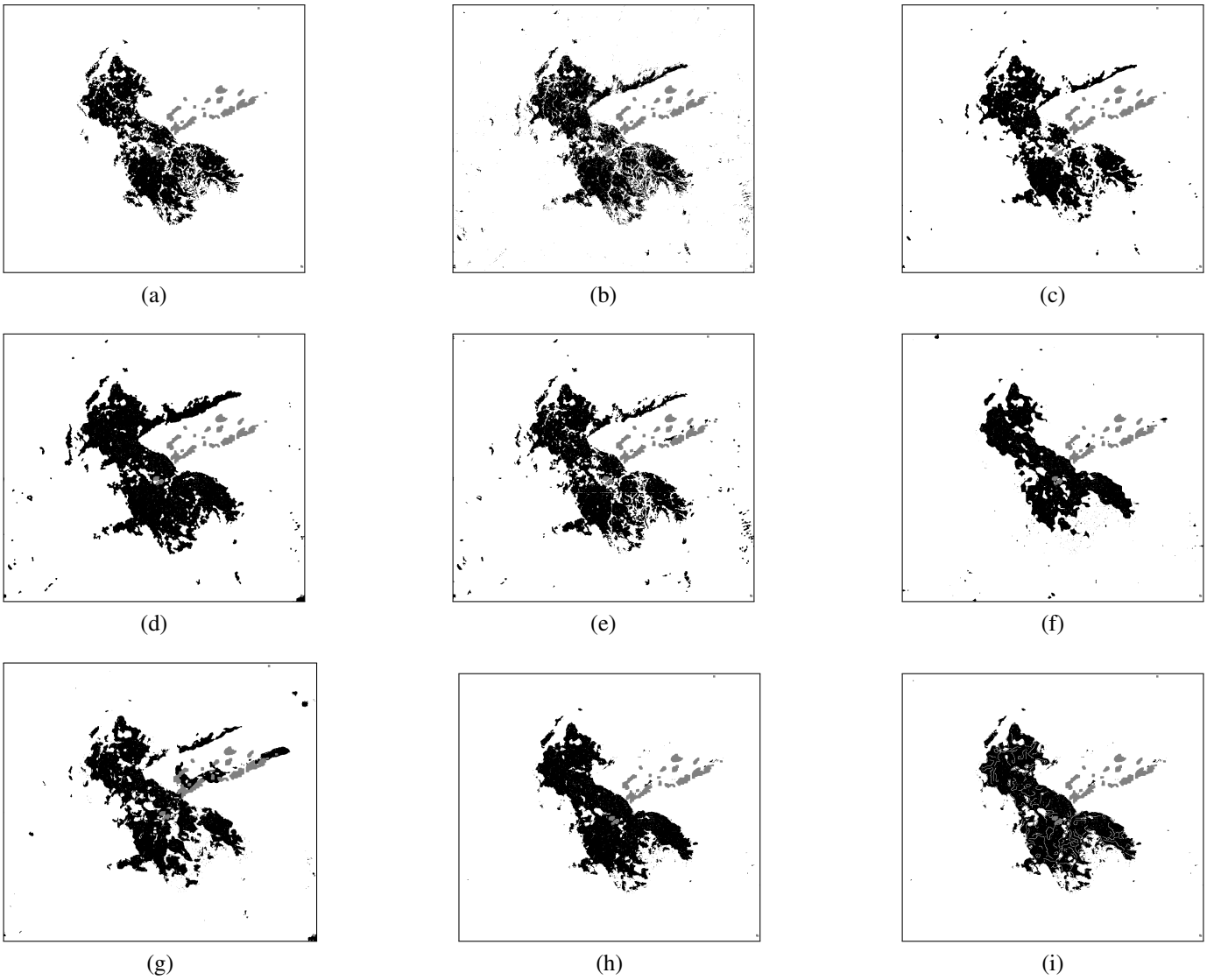


Fig. 9. Comparisons between (a) the reference map of the dataset of Granada and (b) the maps achieved by applying adaptive CVA, (c) CVA using a MRF based method, (d) semiparametric CVA using MRF, (e) the self-supervised segmentation, the proposed method exploiting (f) a single-scale feature, (g) no feature selection, (h) standard-deviation-based reliability maps, and (i) canny filter. (The white pixels represent no changes, the black pixels the changes, and the grey ones are no data)

one and three, SoA techniques incurred in higher FAs (Fig. 10b, 10c 10e) or underestimated the oldest change (Fig. 10d, 12d), which may present vegetation regrowth, and therefore has a weaker spectral contrast. Whereas the proposed method detected all the changes in every tested scenario. Through the use of the CAE and the automatic feature selection process, it handled the multiple spectral bands of the considered images by selecting the most relevant feature maps. The CVA-based SoA methods needed to select the bands before the processing. Otherwise, if all the spectral bands were used, the accuracy decreased.

In the Experiment 4, we can observe that the multi-scale approach detected the changed areas more accurately than the single-scale one. In some scenarios, the method using single-scale feature maps detected more changed areas. However, it did not preserve the geometrical details and the changed-area borders, whereas the proposed method using multi-scale features detected the changed areas with more accurate contours by keeping the spatial information and demonstrating the goodness of selected feature maps. In the second scenario, the use of all the feature maps achieved comparable results with respect to the feature-selection case. However, the feature selection allowed achieving better and more stable performance through the different scenarios and minimizing the error. Following, we provide further observation of the experimental results for each scenario:

1) Scenario 1: Self-supervised segmentation (the SoA method achieving the best performance) achieved comparable performance in the detection of changed areas with respect to the proposed method. However, the latter generally found fewer OE than self-supervised segmentation. In the qualitative results, we can observe that the proposed method accurately detected the changed area with few FAs (Fig. 10h, 10i), whereas the SoA ones underestimated it (Fig. 10d) or found

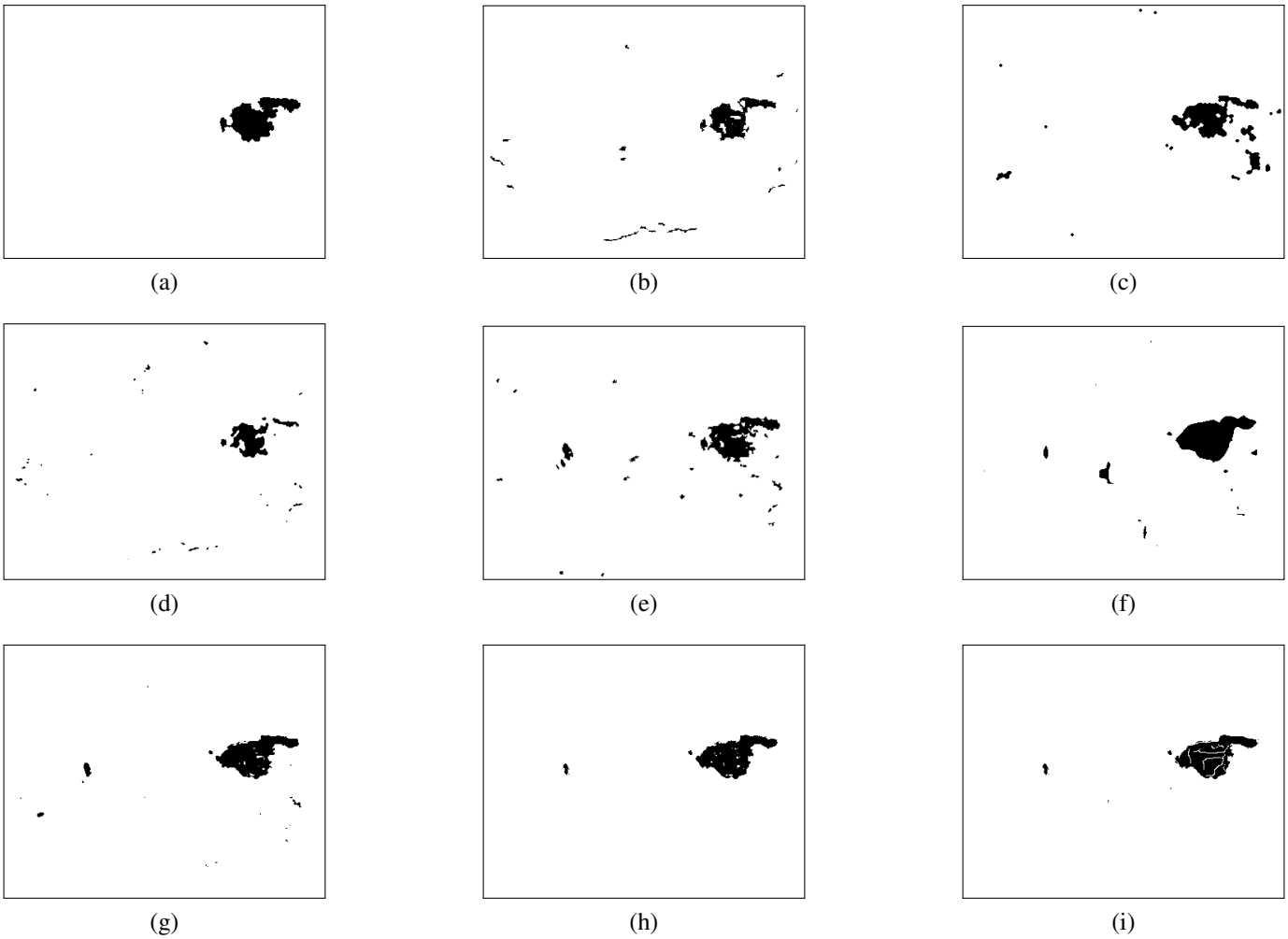


Fig. 10. Comparisons between (a) the reference map of the dataset of Elba acquired in August 1992 and August 1994, and (b) the results achieved by applying adaptive CVA, (c) CVA using a MRF based method, (d) semiparametric CVA using MRF, (e) the self-supervised segmentation, the proposed method exploiting (f) a single-scale feature, (g) no feature selection, (h) standard-deviation-based reliability maps, and (i) canny filter. (The white pixels represent no changes, the black pixels the changes, and the grey ones are no data)

TABLE XIV
FAS, MAS, TPS, OE, SENSITIVITY AND SPECIFICITY OF THE PROPOSED METHOD VS THE NUMBER OF EPOCHS ON THE CAE TRAINING AND BY USING A STANDARD-DEVIATION-BASED AND CANNY-FILTER-BASED APPROACH TO FIND THE MOST RELIABLE AREAS (IMAGES OF ELBA DATASET ACQUIRED IN AUGUST 1992 AND SEPTEMBER 1994).

E	Reliability approach	FA	MA	TP	OE	Sens.	Spec.
50	Canny	0.77%	24.51%	3968	1.7%	75.49%	99.23%
	Std. dev.	0.83%	18.68%	4274	1.53%	81.32%	99.17%
100	Canny	0.59%	23.12%	4041	1.47%	75.46%	99.48%
	Std. dev.	0.64%	18.7%	4273	1.35%	81.3%	99.36%
150	Canny	0.52%	24.54%	3966	1.46%	75.46%	99.48%
	Std. dev.	0.58%	26.88%	3843	1.6%	73.11%	99.42%
200	Canny	0.58%	19.14%	4250	1.3%	80.86%	99.42%
	Std. dev.	0.65%	**13.76%**	**4533**	**1.16%**	**86.24%**	99.35%
250	Canny	0.49%	23.21%	4036	1.37%	76.79%	99.55%
	Std. dev.	0.52%	22.43%	4077	1.37%	77.57%	99.48%
300	Canny	**0.21%**	37.9%	3264	1.68%	62.1%	**99.79%**
	Std. dev.	0.23%	36.13%	3357	1.62%	63.87%	99.77%

many errors (Fig. 10b, 10c, 10e). This showed the effectiveness of the proposed method in the detection of changes with low spectral contrast in time due to vegetation regrowth. From the Experiment 4, we can observe that the single-scale option detected more changed areas than the multi-scale one (Table XV). However, the multi-scale method preserved the geometrical details and retrieved more accurate borders (Fig. 10h). The use of feature selection minimized the OE. It resulted in less FAs (Fig. 10h) than without feature selection (Fig. 10g). This proved that the feature selection filters out no-informative feature maps.

2) Scenario 2: The proposed method achieved comparable performance than the best SoA method (i.e., the CVA using MRF). However, the proposed method found fewer FAs (Fig. 11h, 11i) than the self-supervised segmentation (Fig. 11e), and detected most of the changed area especially using the Canny-based reliability approach (Fig. 11i). The Canny filter delineated borders accurately (Fig. 13b). The standard-deviation-based strategy classified as not reliable many changes in the borders between ω_c and ω_{nc} (Fig. 13a). The results achieved in the Elba dataset confirmed that the Canny-filter strategy is more accurate in detecting borders. The proposed

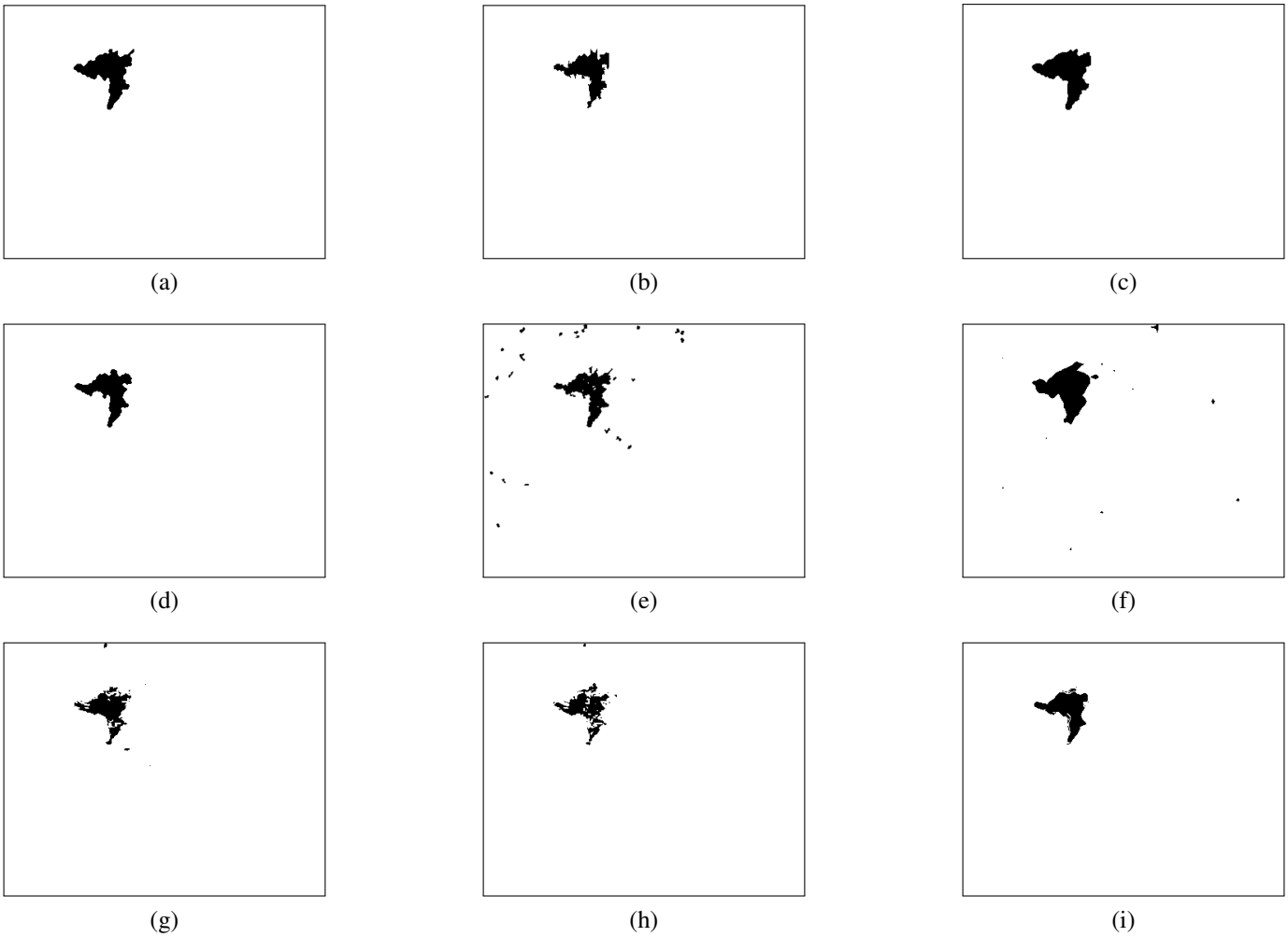


Fig. 11. Comparisons between (a) the reference map of the dataset of Elba acquired in August 1994 and September 1994, and (b) the results achieved by applying adaptive CVA, (c) CVA using a MRF based method, (d) semiparametric CVA using MRF, (e) the self-supervised segmentation, the proposed method exploiting (f) a single-scale feature, (g) no feature selection, (h) standard-deviation-based reliability maps, and (i) canny filter. (The white pixels represent no changes, the black pixels the changes, and the grey ones are no data)

method increased its sensitivity by 28.62% using single-scale feature maps with less accurate borders and geometrical details (Fig. 11f). The proposed method using multi-scale features preserved the spatial information and retrieved more accurate borders but detected slightly fewer changes (Fig. 11h) due to the standard deviation reliability approach. Without applying the feature selection, the method achieved comparable performance with respect to the feature selection option (Fig. 11g).

3) Scenario 3: Adaptive CVA detected fewer FAs than the proposed method at the cost of decreasing the sensitivity by 21.29%. The proposed method also found fewer OE than the adaptive CVA. From the qualitative results, we can observe that the proposed method accurately detected both the changed areas (Fig. 12h, 12i), and therefore the two kinds of change. Whereas the SoA methods found the most recent and thus stronger one but missed most of the older ones mitigated by vegetation regrowth (Fig. 12b, 12c). This is because the proposed method analyzed all the spectral bands. It is worth noting that the proposed method outperformed the other DL CD method (self-supervised segmentation) that had problems in the detection of the two kinds of changes and found many FAs. The proposed method using multi-scale feature maps overcame the single-scale case since it detected more changed areas with fewer FAs than the single-scale option. It also preserved better the changed-area borders, especially the ones of the most recent change (Fig. 12h), than using single-scale feature maps (Fig. 12f). The proposed method using the feature selection minimized the OE. The feature selection chose only the informative feature maps that helped the method detecting the changed areas (Fig. 12h), whereas, by using all the feature maps, some changes may not be found (Fig. 12g).

G. Experiments 3 and 4: Indonesia dataset

In Experiment 3, the proposed method using a CAE trained with the Indonesia dataset for $E = 50$ detected most of the changed areas. However, it is sensitive to changes due to environmental conditions, so it found many FAs. With the increase of E, the method detected fewer changed areas, but it decreased the FA rate of 8.19%. With $E \geq 100$, the CD results were stable (Table XVIII). The standard-deviation-

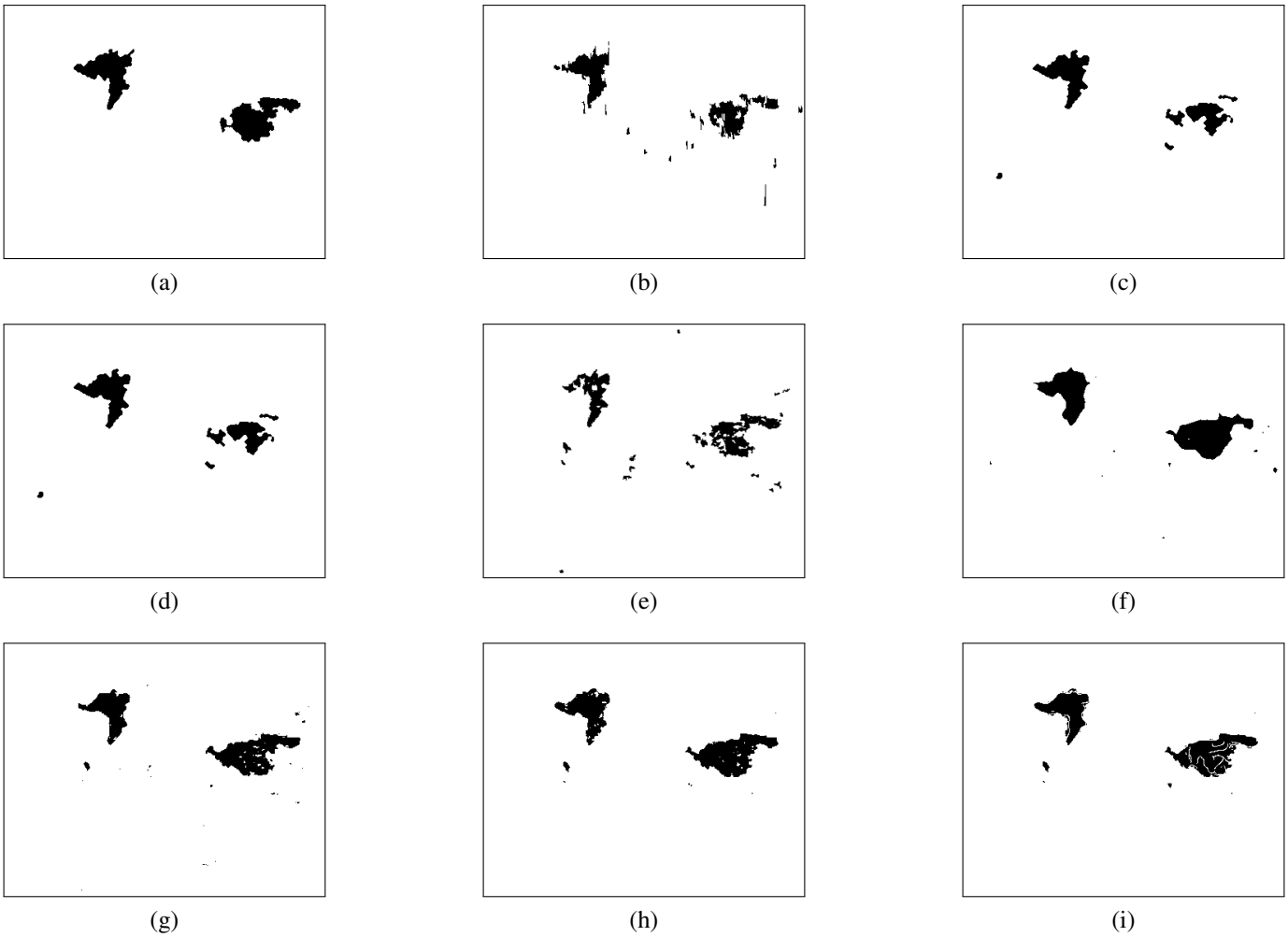


Fig. 12. Comparisons between (a) the reference map of the dataset of Elba acquired in August 1992 and September 1994, and (b) the results achieved by applying adaptive CVA, (c) CVA using a MRF based method, (d) semiparametric CVA using MRF, (e) the self-supervised segmentation, the proposed method exploiting (f) a single-scale feature, (g) no feature selection, (h) standard-deviation-based reliability maps, and (i) canny filter. (The white pixels represent no changes, the black pixels the changes, and the grey ones are no data)

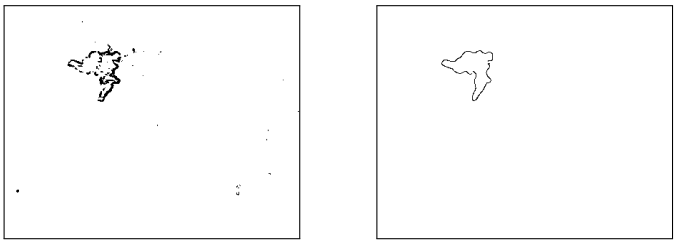


Fig. 13. Reliability maps retrieved using the standard-deviation-based strategy on the resolution level (a) $l'_{sel} = 1$ (the highest spatial resolution), and the Canny-filter-based strategy (b) $l'_{sel} = 1$. These maps derive from the feature maps retrieved by the layers of the CAE trained for $E = 150$ epochs using images acquired in August 1994 and September 1994. (The white pixels represent reliable areas, the black ones the no reliable areas at $l'_{sel} = 1$.)

based strategy detected more change areas but with more FAs, whereas the Canny-filter-based reliability approach was more conservative than the other one, as in the other datasets. It detected fewer changed areas but with less FAs, and it had slightly smaller OE than the standard-deviation strategy. The CD method achieved the best trade-off between TPs, MAs, and FAs using a model trained for $E = 250$ with the Canny-based reliability approach. For this reason, we exploited these settings to process the bi-temporal images and retrieve the change map.

The proposed method outperformed the SoA ones since it detected most of the changed areas by keeping the number of FAs relatively low (Table XIX). The semiparametric CVA using MRF [52] detected more changed areas than the proposed method using the Canny-based strategy. However, the proposed method increased the specificity by 16.31% with respect to the semiparametric CVA using MRF and incurred in the lowest number of OE (i.e., 126886). The qualitative results confirmed the quantitative ones since the SoA methods detected most of the changed areas but found many FAs due to the cloud coverage (Fig. 14b). The proposed methods using both the reliability approaches achieved good performance. It detected most of the changed areas without the FAs due to the cloud coverage (Fig. 14e, 14f). However, the SoA and

TABLE XV
FAS, MAS, SENSITIVITY, SPECIFICITY AND OE (IN NUMBER OF PIXELS AND PERCENTAGE) OBTAINED BY THE SOA METHODS AND THE PROPOSED ONE. WE TESTED THE PROPOSED METHOD USING THE TWO RELIABILITY APPROACHES, A SINGLE-SCALE FEATURE, AND NO FEATURE SELECTION (IMAGES OF ELBA DATASET ACQUIRED IN AUGUST 1992 AND AUGUST 1994).

Method	FA	MA	Sens.	Spec.	OE	
					Pixels	%
Adaptive CVA [50]	0.42%	25.72%	74.28%	99.58%	1281	0.95%
CVA + MRF [51]	1.04%	26.43%	73.57%	98.96%	2123	1.57%
Semipar. CVA + MRF [52]	**0.21%**	46.66%	53.34%	**99.79%**	1607	1.19%
Self-supervised segm. [22]	0.64%	14.32%	85.68%	99.36%	1252	0.93%
Proposed w/ std. dev.	0.42%	15.13%	84.87%	99.58%	**988**	**0.73%**
Proposed w/ Canny	0.4%	19.18%	80.82%	99.6%	1078	0.8%
Proposed w/ single-scale feat.	0.88%	**10.49%**	**89.51%**	99.12%	1457	1.08%
Proposed no feat. sel.	0.56%	14.67%	85.33%	99.44%	1158	0.86%

TABLE XVI
FAS, MAS, SENSITIVITY, SPECIFICITY AND OE (IN NUMBER OF PIXELS AND PERCENTAGE) OBTAINED BY THE SOA METHODS AND THE PROPOSED ONE. WE TESTED THE PROPOSED METHOD USING THE TWO RELIABILITY APPROACHES, A SINGLE-SCALE FEATURE, AND NO FEATURE SELECTION (IMAGES OF ELBA DATASET ACQUIRED IN AUGUST 1994 AND SEPTEMBER 1994).

Method	FA	MA	Sens.	Spec.	OE	
					Pixels	%
Adaptive CVA [50]	0.04%	20.38%	79.62%	99.96%	539	0.4%
CVA + MRF [51]	0.1%	**1.57%**	**98.43%**	99.9%	**176**	**0.13%**
Semipar. CVA + MRF [52]	**0.01%**	16.11%	83.89%	**99.99%**	397	0.29%
Self-supervised segm. [22]	0.32%	17.69%	82.31%	99.68%	853	0.63%
Proposed w/ std. dev.	0.08%	36.7%	63.3%	99.92%	998	0.74%
Proposed w/ Canny	0.02%	23.2%	76.8%	99.98%	581	0.43%
Proposed w/ single-scale feat.	0.5%	8.08%	91.92%	99.5%	861	0.64%
Proposed no feat. sel.	0.09%	28.96%	71.04%	99.91%	815	0.6%

proposed methods detected some changed areas not present in the reference map. Those are due to the vegetation regrowth not considered in the reference map that only provides information about deforestation.

Experiment 4 proved the effectiveness of the multi-scale analysis in the proposed method. Using the multi-scale analysis (Fig. 14f), the CD method is less sensitive to the changes due to the cloud coverage than the single-scale option (Fig. 14c). The latter detected most of the changed areas but was less accurate than the multi-scale approach in the identification of the related borders. These observations were confirmed by the quantitative results, where both the sensitivity and specificity of the proposed method using the multi-scale analysis increased with respect to the single-scale ones (Table XIX). Finally, the proposed method without the feature selection

TABLE XVII
FAS, MAS, SENSITIVITY, SPECIFICITY AND OE (IN NUMBER OF PIXELS AND PERCENTAGE) OBTAINED BY THE SOA METHODS AND THE PROPOSED ONE. WE TESTED THE PROPOSED METHOD USING THE TWO RELIABILITY APPROACHES, A SINGLE-SCALE FEATURE, AND NO FEATURE SELECTION (IMAGES OF ELBA DATASET ACQUIRED IN AUGUST 1992 AND SEPTEMBER 1994).

Method	FA	MA	Sens.	Spec.	OE	
					Pixels	%
Adaptive CVA [50]	**0.5%**	35.05%	64.95%	**99.5%**	2493	1.85%
CVA + MRF [51]	0.55%	38.96%	61.04%	99.45%	2759	2.04%
Semipar. CVA + MRF [52]	0.56%	38.05%	61.95%	99.44%	2723	2.02%
Self-supervised segm. [22]	0.52%	35.18%	64.82%	99.48%	2517	1.86%
Proposed w/ std. dev.	0.65%	**13.76%**	**86.24%**	99.35%	**1568**	**1.16%**
Proposed w/ Canny	0.58%	19.14%	80.86%	99.42%	1758	1.3%
Proposed w/ single-scale feat.	0.89%	14.38%	85.62%	99.11%	1914	1.42%
Proposed no feat. sel.	0.62%	18.38%	81.62%	99.38%	1769	1.31%

TABLE XVIII
FAS, MAS, TPS, OE, SENSITIVITY AND SPECIFICITY OF THE PROPOSED METHOD VS THE NUMBER OF EPOCHS ON THE CAE TRAINING AND BY USING A STANDARD-DEVIATION-BASED AND CANNY-FILTER-BASED APPROACH TO FIND THE MOST RELIABLE AREAS (INDONESIA DATASET).

E	Reliability approach	FA	MA	TP	OE	Sens.	Spec.
50	Canny	12.78%	16.35%	61885	12.89%	83.65%	87.22%
	Std. dev.	13.76%	**13.1%**	**64288**	13.74%	**86.9%**	86.24%
100	Canny	**4.59%**	26.11%	54663	**5.25%**	73.89%	**95.41%**
	Std. dev.	5.15%	22.31%	57473	5.68%	77.68%	94.85%
150	Canny	4.74%	25.82%	54875	5.39%	74.18%	95.26%
	Std. dev.	5.34%	22.2%	57552	5.86%	77.78%	94.66%
200	Canny	4.66%	23.61%	56512	**5.25%**	76.39%	95.34%
	Std. dev.	5.28%	20.38%	58901	5.75%	79.61%	94.71%
250	Canny	4.77%	21.97%	57724	5.31%	78.03%	95.23%
	Std. dev.	5.3%	19.82%	59317	5.75%	80.18%	94.7%
300	Canny	4.89%	23.32%	56729	5.46%	76.68%	95.11%
	Std. dev.	5.57%	19.79%	59336	6.01%	80.21%	94.43%

detected more changed areas than the one with feature selection. However, it incurred in more OE (i.e., 317535) than the proposed method using the feature selection (i.e., 126886). In particular, it increased the FA rate by 8.59%. Using all the features extracted by the CAE, the CD method is sensitive to the changes due to environmental conditions (Fig. 14d). The feature selection removed the feature sensitive to this kind of changes and allowed reducing the FAs (Fig. 14f).

IV. CONCLUSION

We have proposed a multi-scale Change-Detection (CD) method that exploits the spatial-context feature maps retrieved by a Convolutional Autoencoder (CAE). We tested our method using three datasets composed of bi-temporal images acquired by Landsat-5, Landsat-8 and Sentinel-2 sensors. The proposed technique can accurately detect the changes by processing all the spectral bands of the considered images with no a-priori band-selection, as required by other SoA methods to optimize results. The CAE aggregates and processes the spectral bands,

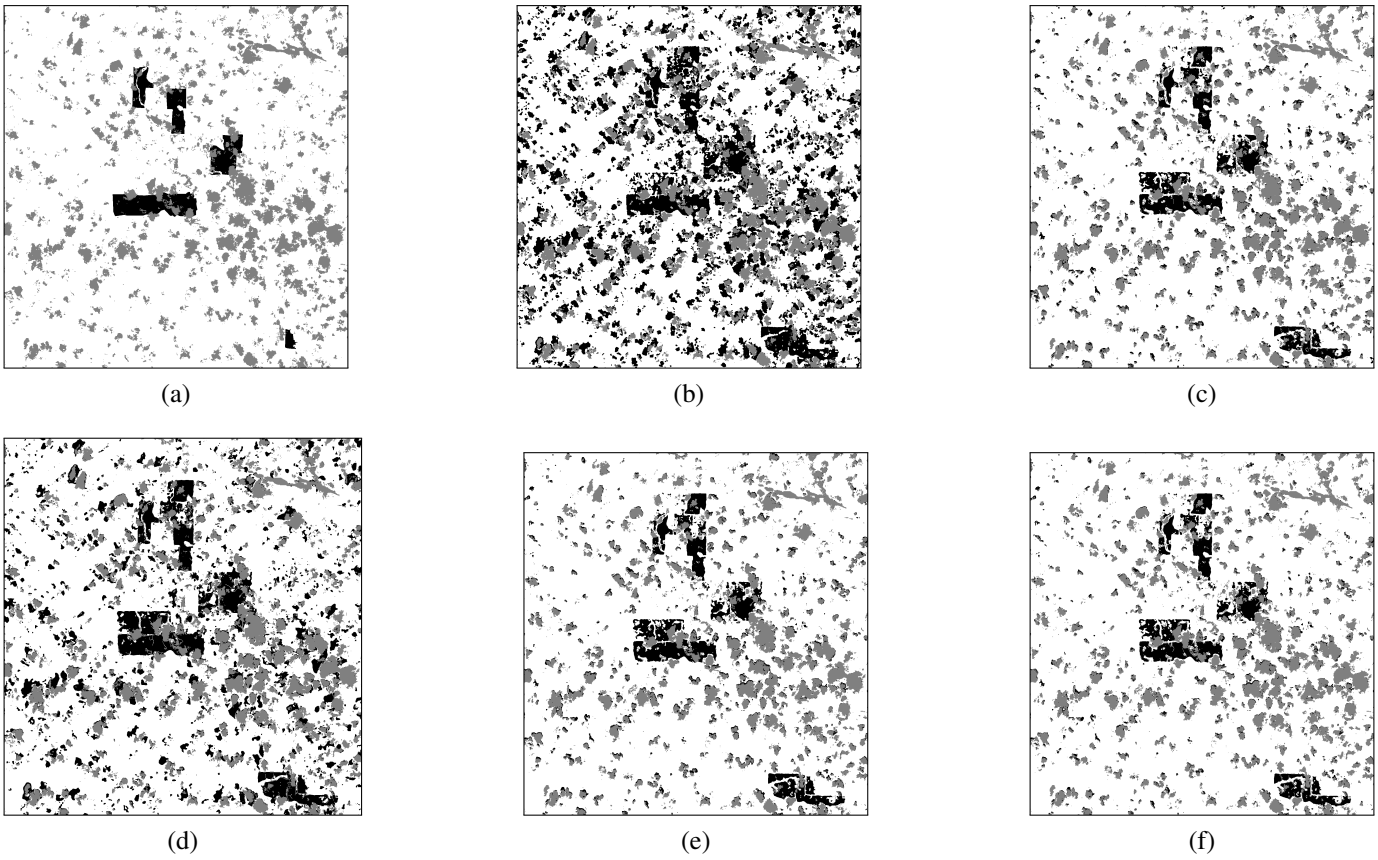


Fig. 14. Comparisons between (a) the reference map of the Indonesia dataset and the change detection maps obtained by (b) the semiparametric CVA using MRF, the proposed method exploiting (c) a single-scale feature, (d) no feature selection, (e) standard-deviation-based reliability maps, and (f) canny filter. (The white pixels represent no changes, the black pixels the changes, and the grey ones are no data)

TABLE XIX
FAS, MAS, SENSITIVITY, SPECIFICITY AND OE (IN NUMBER OF PIXELS AND PERCENTAGE) OBTAINED BY THE SOA METHODS AND THE PROPOSED ONE. WE TESTED THE PROPOSED METHOD USING THE TWO RELIABILITY APPROACHES, A SINGLE-SCALE FEATURE, AND NO FEATURE SELECTION (INDONESIA DATASET).

Method	FA	MA	Sens.	Spec.	OE	
					Pixels	%
Adaptive CVA [50]	16.22%	24.45%	75.55%	83.78%	393992	16.48%
CVA + MRF [51]	10.84%	33.18%	66.82%	89.16%	275715	11.53%
Semipar. CVA + MRF [52]	21.08%	11.7%	88.3%	78.92%	497070	20.79%
Self-supervised segm. [22]	8.11%	54.83%	45.17%	91.89%	228391	9.55%
Proposed w/ std. dev.	5.3%	19.82%	80.18%	94.7%	137488	5.75%
Proposed w/ Canny	**4.77%**	21.97%	78.03%	**95.23%**	**126886**	**5.31%**
Proposed w/ single-scale feat.	5.87%	23.83%	76.17%	94.13%	153570	6.42%
Proposed no feat. sel.	13.36%	**10.67%**	**89.33%**	86.64%	317535	13.28%

and the automatic feature selection step chooses the most relevant feature maps representing the changes. The proposed method achieved stable performance by varying the number of epochs during the training of the CAE feature extractor. It also obtained similar performance in different scenarios. This is mainly due to the capability of the CAE to automatically learn effective features during the unsupervised training. It is worth noting that supervised and transfer-learning based methods may provide better performance when labeled multi-temporal data or suitable pre-trained network are available, respectively. However, the proposed method represents an effective alternative in many applications where these requirements cannot be satisfied. It improves the performance of other SoA CD methods and does not require a-priori band selection.

In future activities, we plan to test the method on other kinds of changes (such as urban changes or floods) and data (i.e., Sentinel 1). We want to develop a multi-change detection strategy and design a hierarchical loss function for CAE training that allows preserving the spatial information of the hidden-layer feature maps.

ACKNOWLEDGMENT

The authors would like to thank the Department of Geology, Geography, and Environment of Universidad de Alcalá for providing the reference map used in our experiments on the Granada dataset. They also want to thank the reviewers for the useful comments and suggestions that allowed improving this paper.

REFERENCES

[1] K. S. Willis, "Remote sensing change detection for ecological monitoring in united states protected areas," *Biological Conservation*, vol. 182, pp. 233–242, 2015.

[2] D. A. Mengistu and A. T. Salami, "Application of remote sensing and gis inland use/land cover mapping and change detection in a part of south western nigeria," *African Journal of Environmental Science and Technology*, vol. 1, no. 5, pp. 99–109, 2007.

[3] J. G. Lyon, D. Yuan, R. S. Lunetta, and C. D. Elvidge, "A change detection experiment using vegetation indices," *Photogrammetric engineering and remote sensing*, vol. 64, no. 2, pp. 143–150, 1998.

[4] F. van den Bergh, G. Udahemuka, and B. J. van Wyk, "Potential fire detection based on kalman-driven change detection," in *2009 IEEE International Geoscience and Remote Sensing Symposium*, vol. 4. IEEE, 2009, pp. IV–77.

[5] T. K. Remmel and A. H. Perera, "Fire mapping in a northern boreal forest: assessing avhrr/ndvi methods of change detection," *Forest Ecology and Management*, vol. 152, no. 1-3, pp. 119–129, 2001.

[6] M. Zanetti, D. Marinelli, M. Bertoluzza, S. Saha, F. Bovolo, L. Bruzzone, M. L. Magliozzi, M. Zavagli, and M. Costantini, "A high resolution burned area detector for sentinel-2 and landsat-8," in *2019 10th International Workshop on the Analysis of Multitemporal Remote Sensing Images (MultiTemp)*. IEEE, 2019, pp. 1–4.

[7] S. Saha, F. Bovolo, and L. Bruzzone, "Destroyed-buildings detection from vhr sar images using deep features," in *Image and Signal Processing for Remote Sensing XXIV*, vol. 10789. International Society for Optics and Photonics, 2018, p. 107890Z.

[8] ——, "Unsupervised deep change vector analysis for multiple-change detection in vhr images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 6, pp. 3677–3693, 2019.

[9] J. Liu, M. Gong, K. Qin, and P. Zhang, "A deep convolutional coupling network for change detection based on heterogeneous optical and radar images," *IEEE transactions on neural networks and learning systems*, vol. 29, no. 3, pp. 545–559, 2016.

[10] Y. Zhan, K. Fu, M. Yan, X. Sun, H. Wang, and X. Qiu, "Change detection based on deep siamese convolutional network for optical aerial images," *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 10, pp. 1845–1849, 2017.

[11] A. M. El Amin, Q. Liu, and Y. Wang, "Convolutional neural network features based change detection in satellite images," in *First International Workshop on Pattern Recognition*, vol. 10011. International Society for Optics and Photonics, 2016, p. 100110W.

[12] L. Mou, L. Bruzzone, and X. X. Zhu, "Learning spectral-spatial-temporal features via a recurrent convolutional neural network for change detection in multispectral imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 2, pp. 924–935, 2018.

[13] P. Zhang, M. Gong, L. Su, J. Liu, and Z. Li, "Change detection based on deep feature representation and mapping transformation for multi-spatial-resolution remote sensing images," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 116, pp. 24–41, 2016.

[14] Q. Wang, Z. Yuan, Q. Du, and X. Li, "Getnet: A general end-to-end 2-d cnn framework for hyperspectral image change detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 1, pp. 3–13, 2018.

[15] S. H. Khan, X. He, F. Porikli, and M. Bennamoun, "Forest change detection in incomplete satellite images with deep neural networks," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 9, pp. 5407–5423, 2017.

[16] X. Niu, M. Gong, T. Zhan, and Y. Yang, "A conditional adversarial network for change detection in heterogeneous images," *IEEE Geoscience and Remote Sensing Letters*, vol. 16, no. 1, pp. 45–49, 2018.

[17] S. Saha, F. Bovolo, and L. Bruzzone, "Unsupervised multiple-change detection in vhr multisensor images via deep-learning based adaptation," in *IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium*. IEEE, 2019, pp. 5033–5036.

[18] B. Hou, Q. Liu, H. Wang, and Y. Wang, "From w-net to cdgan: Bitemporal change detection via deep learning techniques," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 3, pp. 1790–1802, 2019.

[19] F. Ye, W. Luo, M. Dong, H. He, and W. Min, "Sar image retrieval based on unsupervised domain adaptation and clustering," *IEEE Geoscience and Remote Sensing Letters*, vol. 16, no. 9, pp. 1482–1486, 2019.

[20] C. Kang and C. He, "Sar image classification based on the multi-layer network and transfer learning of mid-level representations," in *2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*. IEEE, 2016, pp. 1146–1149.

[21] M. Yang, L. Jiao, B. Hou, F. Liu, and S. Yang, "Selective adversarial adaptation-based cross-scene change detection framework in remote sensing images," *IEEE Transactions on Geoscience and Remote Sensing*, 2020.

[22] S. Saha, Y. T. Solano-Correa, F. Bovolo, and L. Bruzzone, "Unsupervised deep transfer learning-based change detection for hr multispectral images," *IEEE Geoscience and Remote Sensing Letters*, 2020.

[23] A. Pomente, M. Picchiani, and F. Del Frate, "Sentinel-2 change detection based on deep features," in *IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium*. IEEE, 2018, pp. 6859–6862.

[24] W. Zhao, Z. Wang, M. Gong, and J. Liu, "Discriminative feature learning for unsupervised change detection in heterogeneous images based on a coupled neural network," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 12, pp. 7066–7080, 2017.

[25] Y. Xu, S. Xiang, C. Huo, and C. Pan, "Change detection based on auto-encoder model for vhr images," in *MIPPR 2013: Pattern Recognition and Computer Vision*, vol. 8919. International Society for Optics and Photonics, 2013, p. 891902.

[26] N. Lv, C. Chen, T. Qiu, and A. K. Sangaiah, "Deep learning and superpixel feature extraction based on contractive autoencoder for change detection in sar images," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 12, pp. 5530–5538, 2018.

[27] H. Zhang, M. Gong, P. Zhang, L. Su, and J. Shi, "Feature-level change detection using deep representation and feature change analysis for multispectral imagery," *IEEE Geoscience and Remote Sensing Letters*, vol. 13, no. 11, pp. 1666–1670, 2016.

[28] F. Bovolo, "A multilevel parcel-based approach to change detection in very high resolution multitemporal images," *IEEE Geoscience and Remote Sensing Letters*, vol. 6, no. 1, pp. 33–37, 2008.

[29] F. Bovolo, L. Bruzzone, and S. Marchesi, "A multiscale technique for reducing registration noise in change detection on multitemporal vhr images," in *2007 International Workshop on the Analysis of Multitemporal Remote Sensing Images*. IEEE, 2007, pp. 1–6.

[30] F. Bovolo and L. Bruzzone, "A detail-preserving scale-driven approach to change detection in multitemporal sar images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 12, pp. 2963–2972, 2005.

[31] T. Celik, "Multiscale change detection in multitemporal satellite images," *IEEE Geoscience and Remote Sensing Letters*, vol. 6, no. 4, pp. 820–824, 2009.

[32] O. Hall and G. J. Hay, "A multiscale object-specific approach to digital change detection," *International journal of applied earth observation and geoinformation*, vol. 4, no. 4, pp. 311–327, 2003.

[33] Y. Zhang, D. Peng, and X. Huang, "Object-based change detection for vhr images based on multiscale uncertainty analysis," *IEEE Geoscience and Remote Sensing Letters*, vol. 15, no. 1, pp. 13–17, 2017.

[34] G. Moser, E. Angiati, and S. B. Serpico, "Multiscale unsupervised change detection on optical images by markov random fields and wavelets," *IEEE Geoscience and Remote Sensing Letters*, vol. 8, no. 4, pp. 725–729, 2011.

[35] S. Liu, Q. Du, X. Tong, A. Samat, L. Bruzzone, and F. Bovolo, "Multiscale morphological compressed change vector analysis for unsupervised multiple change detection," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 10, no. 9, pp. 4124–4137, 2017.

[36] J. Masci, U. Meier, D. Cireşan, and J. Schmidhuber, "Stacked convolutional auto-encoders for hierarchical feature extraction," in *International conference on artificial neural networks*. Springer, 2011, pp. 52–59.

[37] E. Othman, Y. Bazi, N. Alajlan, H. Alhichri, and F. Melgani, "Using convolutional features and a sparse autoencoder for land-use scene classification," *International Journal of Remote Sensing*, vol. 37, no. 10, pp. 2149–2167, 2016.

[38] J. Geng, J. Fan, H. Wang, X. Ma, B. Li, and F. Chen, "High-resolution sar image classification via deep convolutional autoencoders," *IEEE Geoscience and Remote Sensing Letters*, vol. 12, no. 11, pp. 2351–2355, 2015.

[39] R. Kemker, C. Salvaggio, and C. Kanan, "Algorithms for semantic segmentation of multispectral remote sensing imagery using deep learning," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 145, pp. 60–77, 2018.

[40] H. R. Kerner, K. L. Wagstaff, B. D. Bue, P. C. Gray, J. F. Bell III, and H. B. Amor, "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*, vol. 12, no. 10, pp. 3900–3918, 2019.

[41] L. Bergamasco, S. Saha, F. Bovolo, and L. Bruzzone, "Unsupervised change-detection based on convolutional-autoencoder feature extrac-

tion," in *Image and Signal Processing for Remote Sensing XXV*, vol. 11155. International Society for Optics and Photonics, 2019, p. 1115510.

[42] E. Kalinicheva, D. Ienco, J. Sublime, and M. Trocan, "Unsupervised change detection analysis in satellite image time series using deep learning combined with graph-based approaches," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 13, pp. 1450–1466, 2020.

[43] F. Bovolo and L. Bruzzone, "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*, vol. 45, no. 6, pp. 1658–1670, 2007.

[44] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *arXiv preprint arXiv:1502.03167*, 2015.

[45] A. L. Maas, A. Y. Hannun, and A. Y. Ng, "Rectifier nonlinearities improve neural network acoustic models," in *Proc. icml*, vol. 30, no. 1, 2013, p. 3.

[46] J. Canny, "A computational approach to edge detection," *IEEE Transactions on pattern analysis and machine intelligence*, no. 6, pp. 679–698, 1986.

[47] R. Maini and H. Aggarwal, "Study and comparison of various image edge detection techniques," *International journal of image processing (IJIP)*, vol. 3, no. 1, pp. 1–11, 2009.

[48] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE transactions on systems, man, and cybernetics*, vol. 9, no. 1, pp. 62–66, 1979.

[49] M. A. Belenguer-Plomer, M. A. Tanase, A. Fernandez-Carrillo, and E. Chuvieco, "Burned area detection and mapping using sentinel-1 backscatter coefficient and thermal anomalies," *Remote Sensing of Environment*, vol. 233, p. 111345, 2019.

[50] L. Bruzzone and D. F. Prieto, "An adaptive parcel-based technique for unsupervised change detection," *International Journal of Remote Sensing*, vol. 21, no. 4, pp. 817–822, 2000.

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

[52] ——, "An adaptive semiparametric and context-based approach to unsupervised change detection in multitemporal remote-sensing images," *IEEE Transactions on image processing*, vol. 11, no. 4, pp. 452–466, 2002.