A 3D CNN APPROACH FOR CHANGE DETECTION IN HR SATELLITE IMAGE TIME SERIES BASED ON A PRETRAINED 2D CNN

Khatereh Meshkini [a,b,*], Francesca Bovolo [a], Lorenzo Bruzzone [b]

[a] Fondazione Bruno Kessler, Center for Digital Society, Trento, Italy; (mkhatereh, bovolo)@fbk.eu
[b] Dept. of Information Engineering and Computer Science, University of Trento, Trento, Italy; lorenzo.bruzzone@unitn.it

Commission III, WG III/1

KEY WORDS: Change Detection (CD), High Resolution (HR) Image, Deep Learning, 3D Convolutional Neural Network (CNN), Transfer Learning, Change Vector Analysis (CVA).

ABSTRACT:

Over recent decades, Change Detection (CD) has been intensively investigated due to the availability of High Resolution (HR) multi-spectral multi-temporal remote sensing images. Deep Learning (DL) based methods such as Convolutional Neural Network (CNN) have recently received increasing attention in CD problems demonstrating high potential. However, most of the CNN-based CD methods are designed for bi-temporal image analysis. Here, we propose a Three-Dimensional (3D) CNN-based CD approach that can effectively deal with HR image time series and process spatial-spectral-temporal features. The method is unsupervised and thus does not require the complex task of collecting labelled multi-temporal data. Since there are only a few pretrained 3D CNNs available that are not suitable for remote sensing CD analysis, the proposed approach starts with a pretrained 2D CNN architecture trained on remote sensing images for semantic segmentation and develops a 3D CNN architecture using a transfer learning technique to jointly deal with spatial, spectral and temporal information. A layerwise feature reduction strategy is performed to select the most informative features and a pixelwise year-based Change Vector Analysis (CVA) is employed to identify changed pixels. Experimental results on a long time series of Landsat 8 images for an area located in Saudi Arabia confirm the effectiveness of the proposed approach.

1. INTRODUCTION

Recently, the availability of High Resolution (HR) satellite images with detailed spatial, spectral and temporal information have increased the range of possible applications of the Change Detection (CD). CD has been regularly used to observe phenomena such as urbanization (Lu et al., 2011), disaster management (Stramondo et al., 2006), natural industrial disasters (Hulley et al., 2014) and Land Cover Changes (LCC) (Zanetti and Bruzzone, 2018), (Solano-Correa et al., 2020). In order to effectively exploit HR Satellite Image Time Series (SITS) in LCC detection, new challenges should be addressed in terms of data processing and algorithm development. In this context the main challenge refers to the complexity of the temporally dense SITS that requires computationally heavy processing algorithms. The first introduced methodologies in the CD context mostly focused on bi-temporal change detection like image differencing and Change Vector Analysis (CVA) (Malila, 1980) (Bovolo et al., 2012). Such approaches benefited of the use of specific features like Principal Component Analysis (PCA) (Celik, 2009), advanced statistical parameters estimation (Bruzzone and Pietro, 2000), Parcel-based (Bovolo and Bruzzone, 2005) and Markov random field (Kasetkasem and Varshney, 2002) paradigms for multi-level and multi-temporal spatial context information modelling.

In recent years, Deep Learning (DL) has become mainstream in image understanding tasks (Krizhevsky et al., 2016) (Ren et al., 2015), including remote sensing image understanding (Zhang et al., 2016). Deep learning has also been introduced for CD and it is considered as a methodology of choice for CD over the past few years (Khan et al., 2017). Deep learning-based change detection methods that have been applied to satellite image analysis are both supervised (Mou et al., 2018) and unsupervised (Louis de Jong and Sergeevna Bosman, 2019). A supervised deep learning method for CD is chosen when the labelled multi-temporal training data are available. An example of supervised change detection method is proposed in (Mou et al., 2018), where a recurrent convolutional neural network (ReCNN) architecture for extracting joint spectral-spatial-temporal features is developed. The main idea is to combine convolutional neural network (CNN) and well-established RNN for remote sensing applications. In this work, a CNN is employed to model the spectral-spatial features for a pair of multi-spectral data patches and a RNN is used to represent the temporal information in the bi-temporal satellite images. Considering the state of the art, the main issue of the supervised deep learning models in remote sensing analysis is the need of collecting and constructing ground reference data for the system-training phase which are difficult to obtain. This is even more true when we deal with long time series (more than two images).

It has been shown that a deep network trained with images of a certain domain can become useful to treat images of other domains (Volpi and Tuia, 2016). As a result, some unsupervised CD methods have been designed in this context. In (Hou et al., 2017) a CNN already pretrained on a large-scale natural image data set is used in a remote sensing context. To get better results they fine-tune the CNN-based architecture to adapt it to their optical remote sensing image. They show that deep learning-based feature extraction has better generalization capability than traditional hand-crafted features. The feature maps are produced by means of convolutional layers which result from applying multiple kernels to the original image.

Despite differences, CD methods emphasize the importance of using spatial context information, object-level information, and complex nonlinear features (Desclée et al, 2006) (Francisco et al, 2021). Moreover, a majority of existing CNN algorithms exploited 2D CNNs (Zhan et al., 2017) (El Amin et al., 2016). But a 2D CNN is unable to properly model the temporal features

* Corresponding author

since it averages and collapses the temporal information to a scalar in the convolutional layers. These methods have limited capability in capturing temporal context information, complex visual features and most of them have focused on bi-temporal CD instead of SITS (more than two images). There is still limited work on CD in image time series with more than two images as input data.

The nature of a 3D CNN with 3D kernels suits to spatio-temporal representation of the satellite images and can provide dynamic information extracted as temporal features. Recently a 3D CNN has been applied to several studies such as human action recognition (Funke et al, 2019), spatio-temporal feature learning (Li et al., 2017), (Sexton, et. al, 2013) and spatial-spectral image classification (Lifan et al., 2021). Considering the feasibility and efficiency of spatio-temporal feature representation of 3D CNN, a 3D CNN architecture is promising for SITS CD. There are rare studies that have applied 3D CNN to extract temporal information from remote sensing SITS, and in most of them temporal features are partially ignored or represented by simplistic models. In (Meshkini et al., 2021) authors developed an unsupervised deep learning-based 3D CNN method by using HR remotely sensed images in the CD context. The 3D CNN architecture was trained on a large-scale supervised video dataset for the purpose of image classification. The features were extracted and stacked from all the convolutional layers in order to generate a hyper feature map for representing the spatio-temporal information of the images. Finally, a pixel wise distance was computed to produce the change map. These 3D CNN architectures have some limitations: 1) they are usually trained targeting scene classification by using the back propagation method and the error is computed by considering the entire image, not at pixelwise level; 2) the architecture is restricted based on the fully connected layer and they can only accept a fixed size input; 3) most of the pretrained 3D CNNs are trained on RGB spectral channels, while the Near Infrared Red (NIR) channel is important for CD, especially for vegetation analysis; 4) since the architecture is not trained on remotely sensed data, the performance of the method is accurate only on the detection of very sudden abrupt changes (changes that happen in a short time with a great magnitude) and it is suitable for CD in small areas only. Thus, a shift is required in the paradigm of CNN from scene classification to pixelwise image segmentation. In (Volpi and Tuia, 2016), the authors proposed a new kind of architecture where all the learnable layers are convolutional with a series of convolutional, pooling, and activation layers followed by a series of deconvolutional and activation layers. It can accept input of any spatial dimension, produce pixelwise output for the entire image and effectively encode the spatial context information of each pixel. By exploiting the CNN-FPS network developed in (Volpi and Tuia, 2016) that is available for downloading, we design our proposed pretrained deep learning-based CD architecture that: 1) is automatic and fully unsupervised; 2) considers a multi-layer 2D CNN architecture designed for semantic segmentation which is trained on remote sensing images and can accept NIR spectral band; 3) is able to analyse the spatio-temporal features, since a transfer learning technique is developed to adapt 2D CNN to 3D by weights transformation; 4) performs a pixelwise time series feature extraction that implicitly models the spatio-temporal context information of each pixel; 5) locates the position and the time of the changes by means of CVA. The method is performed on each two adjacent years covering SITS, effectively extracts change information by representing spatio-temporal features and detects changes in space and time. Some qualitative and quantitative analysis is provided on a region in Saudi Arabia for the period 2013 to 2019 using ten images per year. The CD maps have been compared with the 3D CNN CD methodology developed in (Meshkini et al., 2021).

The rest of this paper is organized as follows. A structure for the 3D CNN CD for SITS with details is presented in section 2. Section 3 and 4 provide some information on the study area and the experimental analysis together with a discussion on the results. In section 5, the conclusion together with the discussion on scope of future research are presented.

2. PROPOSED 3D CNN APPROACH TO CHANGE DETECTION IN SITS

Figure 1 shows the block scheme of the proposed 3D CNN-based approach to CD in SITS. Let $SITS = \{SITS_1, \dots, SITS_m, \dots, SITS_M\}$ be a pre-processed time series covering M years including images acquired over the same geographical area. Let $SITS_m = \{X_1, \dots, X_n, \dots, X_N\}$ and $SITS_{m+1} = \{Y_1, \dots, Y_n, \dots, Y_N\}$ be two time series with non-uniform time sampling for the mth year and $(m+1)$th year, respectively. Let N correspond to the total number of images for each year ($N > 2$).

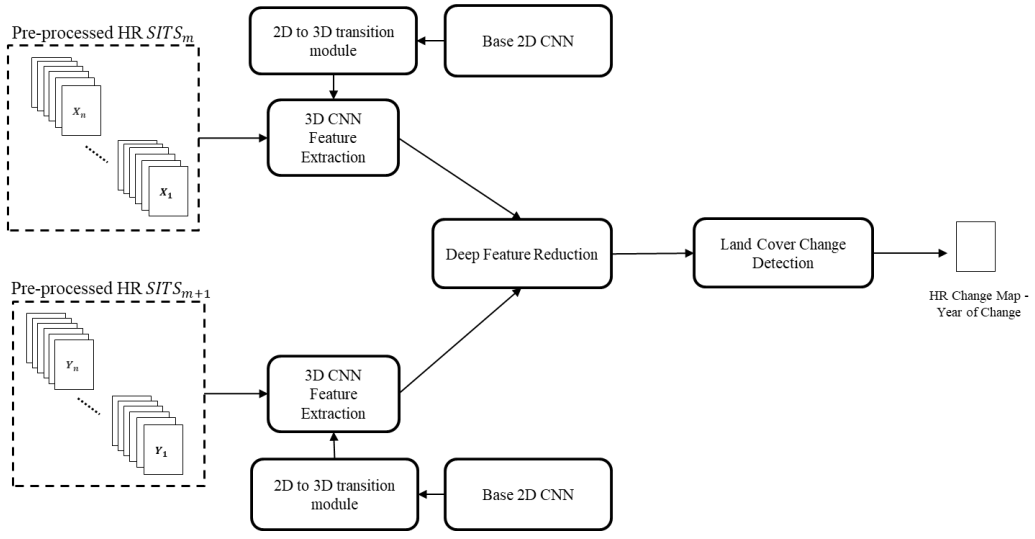

Figure 1. Block scheme of the proposed 3D CNN-based CD approach for $SITS_m$ and $SITS_{m+1}$.

Let $Band = \{b_1, b_2, \dots, b_B\}$ be the set of bands that compose images $X_n \in SITS_m$ and $Y_n \in SITS_{m+1}$ (B is the total number of bands). Given an image X_n or Y_n, the size of the input image is $B \times I \times J$ where I and J correspond to the number of rows and columns of X_n and Y_n, respectively.
The proposed method aims at detecting changes between the pre-processed $SITS_m$ and $SITS_{m+1}$ thus involving $2N$ (>>2) images. It consists of four main steps: i) 2D to 3D transition where the weights and convolutional layers of the base 2D CNN are transformed to three dimensions to generate a new 3D CNN architecture; ii) 3D CNN feature extraction where features are extracted from SITS to obtain a hyper change vector; iii) feature reduction based on a variance measure; iv) land cover change detection by a pixelwise distance calculation among $SITS_m$ and $SITS_{m+1}$. The proposed 3D CNN architecture automatically processes b_1, b_2, b_3, b_4 bands at a time and the output is a change map representing the LC changes and the year of change in $SITS$.

2.1 2D to 3D transition

The proposed approach starts from a 2D CNN trained on remote sensing images to develop a new 3D CNN architecture for feature extraction in SITS. Figure 2 represents the overview of the proposed strategy to exploit 2D weights for training a 3D CNN architecture. First, the weights of a pretrained 2D CNN are transformed to 3D to be used as the weights of the 3D CNN. Second, the 2D CNN is adapted to the 3D CNN by using 3D convolutional layers instead of 2D convolutional layers. Finally, the 3D CNN extracts the 3D features that will be passed to the feature reduction module.

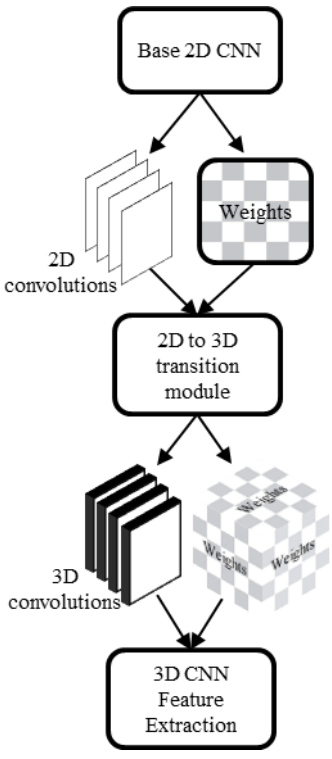

Figure 2. Structure of the proposed strategy for 2D to 3D transition.

In order to transform the convolutional layers and weights of 2D CNN to three dimensions, two techniques are considered as described in (Merino et al., 2021). Since the weights can be represented as 2D matrices, a 2D matrix can be transformed into a 3D tensor to generate the 3D weights. Let $W(x, y)$ = (R, G, B, NIR) be the matrix of weights in the position x and y of the 2D CNN, so the transformed matrix of weights can be represented as $V(x, y, z) = (R,\ G,\ B,\ NIR)$ where $x, y, z \in \mathbb{N}$ and $R,\ G,\ B,\ NIR \in \mathbb{R}$. Two strategies Extrusion and Rotation have been developed as the transformation functions from 2D to 3D.
Extrusion transforms the 2D matrix to 3D tensor by copying the R, G, B, NIR values along one axis. Given a matrix W of size $I \times J$, the transformed tensor V has in size of $I \times J \times N$ where N is the total number of images per year and $I = J$. The Extrusion can be done along the three main axes and is defined as:

$$\forall x, y, z \leq I : V(x, y, z) = W(x, y) \qquad (1)$$

Rotation transforms 2D to 3D weights by performing a 0 to 90 degrees rotation with respect to the fixed axis Z. Equation 2 represent Rotation transformation from 2D matrix to 3D tensor:

$$V(x, y, z) = W\left(x, \min\left(\left\lfloor\sqrt{y^2 + z^2}\right\rfloor, I\right)\right) \qquad (2)$$

2.2 3D CNN feature extraction

Our challenge in the development of an unsupervised method for CD in time series is obtaining a suitable pretrained CNN architecture that can properly model spatio-temporal information of SITS. Several pretrained architectures exist in literature that accept only RGB input (Nogueira et al., 2017), thus losing a large amount of information embedded in the NIR band. Among the pretrained CNN architectures, we use the best of 2D CNN architectures (CNN-FPS network) that is developed by (Volpi and Tuia, 2016) and is trained on remote sensing images for semantic segmentation however other architectures can be considered. The CNN-FPS network processes a five channel input composed of Blue, Green, Red, NIR and digital surface model (DSM). We exclude the DSM input since it is seldom available in the context of change detection and it becomes even more rare as temporal series becomes longer and denser. Removing the kind of input from the pre-trained network does not significantly impact the feature extraction performance as it is presented in (Saha et al., 2019) . The 2D convolutional network is transformed to a 3D convolutional network by considering the two transformation strategies explained in section 2.1. Extrusion and Rotation transformations have been applied separately to generate a 3D tensor from the 2D weight matrices. Moreover, the 2D convolutional layers have been transformed to 3D convolutional layers. The resulting 3D CNN architecture has a more complex network structure with the convolution and pooling operations that perform spatio-temporally to preserve the temporal information and extract relevant features from multi-temporal data. A convolutional layer l in the 3D CNN structure extracts features from the local neighborhood of feature maps in the previous layer $l - 1$. The output at position (x, y, z) denoted as f_{xyz}, is given by equation (3) (for simplification, we omit the layer notation l):

$$f_{xyz} = \sigma\left(\sum_h \sum_{n=0}^{N} \sum_{i=0}^{I} \sum_{j=0}^{J} w_{ijn}^h X_{(x+i)(y+j)(z+n)}^h + a\right) \qquad (3)$$

Where w_{ijn}^h is the 3D tensor for the (i, j, n)th value of the kernel connected to the hth feature map in the previous layer, and I and J are the height and width of the kernel, respectively. X_{xyz} represents the input activation at location (x, y, z), n is the temporal indicator with length N (the number of images per year)

and h refers to the set of SITS feature maps of the previous layer. Choosing the right layers to extract features is also important. There is a significant difference in characteristics of features depending on the layer from which they are extracted. The initial layers of the CNN capture low-level visual concepts such as edges, curves, and color patches. As we go deeper, filters capture more complex concepts by combining lower level features of the previous layers (Zeiler and Fergus, 2014). Since both high and low level features are useful to analyse HR images, a combination of them should be considered in the feature extraction step to catch information (Hariharan et al., 2015). To extract features from the 3D CNN architecture, we choose more convolutional layers than deconvolutional ones to form the hypervector since the convolutional layers learn the semantics of the image at a degraded resolution and the deconvolutional layers mainly learn to reconstruct the spatial arrangements. The first convolutional layer is excluded as it learns very primitive features that are significantly noisy. The convolutional and deconvolutional layers that have been used in this study for the 3D CNN feature extraction are shown in Table I. Considering equation (3) and Table 1, a feature map for each layer l is obtained as $f^l = \{f_1^l, \dots, f_m^l, \dots, f_M^l\}$ for each SITS after applying 3D convolutions in the spatio-temporal domain of images and accumulating the outputs of the spectral bands.

2.3 Feature reduction

The 3D CNN model that is used in the feature extraction provides a large number of features for each layer (up to 512). Not all of them carry relevant information for CD. Thus, a feature selection technique is developed in order to maintain only informative and reliable features.

Layer number	Layer type	Feature dimension
2	convolutional	64
5	convolutional	128
8	convolutional	256
10	convolutional	512
11	convolutional	512
20	deconvolutional	512
23	deconvolutional	512

Table 1. Structure of convolutional and deconvolutional layers used in the 3D CNN.

The applied feature selection strategy is the one proposed in (Bovolo and Bruzzone, 2017), which is based on the variance measurement. The feature selection is performed for each layer l of the 3D CNN architecture and the hyper feature vector for every $SITS_m \in SITS$ is obtained by concatenating the selected features from each layer $l, l = 1, \dots, L$ (L is total number of the layers).
For a given layer l, a layerwise difference vector d_l is computed by subtracting f_m^l from f_{m+1}^l. Then, a subset $d_{l'}$ of d_l is considered that contains features more sensitive to change information. The variance measurement is used as an index of sensitivity to change information. The assumption is that features containing potentially relevant change information have higher variance than those less affected by changes. Inspired by (Bovolo and Bruzzone, 2017), features are spatially divided into $S = \{S_1, \dots, S_s\}$ splits. For a given split s, feature variance ($\sigma_{l,s}^2$) is calculated for all features in d_l. Features having higher $\sigma_{l,s}^2$ values are assumed to have potentially relevant change information. Thus, features in d_l are sorted as per the descending order of $\sigma_{l,s}^2$ values. A subset d_{ls} is selected by retaining a certain percentile of sorted d_l. All the selected features $d_{l'}$ for the layer l are obtained by taking features selected on each split, i.e.,

$$d_{l'} = \bigcup_{s=1}^{S} d_{ls} \tag{4}$$

Selected features from each layer in L are concatenated to obtain change hyper feature vector F:

$$F = [d_{1'}, \dots, d_{l'}, \dots, d_{L'}] \tag{5}$$

2.4 Land cover change detection

The hyper feature vector F with dimension D created in the feature selection step represents effectively the behaviour of changed and unchanged pixels between $SITS_m$ and $SITS_{m+1}$. In order to provide a comprehensive comparison between changed and unchanged pixels the magnitude of F is calculated by considering the CVA technique proposed in (Bovolo and Bruzzone, 2007) that is extended in the context of time series CD for means of the proposed 3D CNN. The magnitude of the hyper feature for each pixel (x, y) is given by:

$$M(x,y) = \sqrt{\sum_{\sigma=1}^{D} (F_{x,y}^{\sigma})^2} \tag{6}$$

Where $F_{x,y}^{D}$ is the feature value on the Dth dimension of the positions x and y of the hyper feature space. Although by calculating the magnitude a strong compression of the hyper feature map is performed, the main information about the changes are preserved.
$M(x,y)$ is assumed to have relatively large values for the changed pixels and small values for the unchanged ones. Therefore, a thresholding strategy is implemented in order to separate them. In the literature, several thresholding techniques have been proposed such as local adaptive threshold (Kieri, 2012), (Wellner, 1993) and Otsu thresholding (Otsu, 1979). Local adaptive thresholding selects an individual threshold for each pixel based on the range of intensity values in its local neighbourhood and changes dynamically over the image. Otsu thresholding processes the image histogram and segments the objects by minimization of the variance on each of the classes. In this research, we examine both thresholding strategies to check the performance of each methods for our proposed CD approach.

3. STUDY AREA AND EXPERIMENTS

To evaluate the effectiveness of the proposed 3D CNN CD approach, a study area is selected in North-West of Saudi Arabia where several central pivot crop fields have been built. Desert to central pivot cropland is the most relevant change class and the pivot crops can be easily identified by checking the images in Figure 3.
The dataset is downloaded directly from USGS Landsat 8 Surface Reflectance Tier1 database and with cloud coverage higher than 70% are ignored. A cloud/cloud shadow mask is imposed to filter cloudy pixels in each image and the data are atmospherically corrected. For each year in the period January 1st, 2013 and December 31st, 2019, ten images of 600 × 600 pixels (18km × 18km) (see Figure 3) are selected being distributed over the seasons to guarantee variability in temporal behaviours. To proceed with the experiments, two 3D CNN architectures have been developed based on the two transformation strategies, Extrusion and Rotation. The developed 3D CNN CD algorithm has been applied six times considering each couple of adjacent years in the acquisition period (2013 to 2019) with four spectral bands (Blue, Green, Red and NIR).

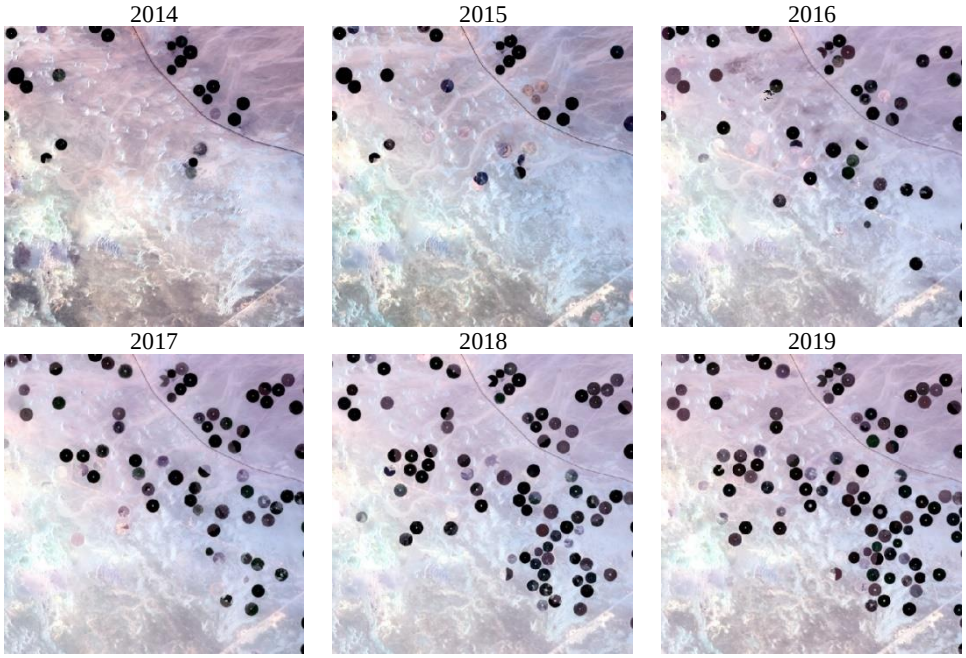

Figure 3. Examples of Saudi Arabia dataset for each year from 2013 to 2019.

The proposed method extracts features from layers $L = \{2, 5, 8, 10\}$ and generates the feature vector (f) for upcoming feature reduction. To select a subset of features being sensitive to change about 10 percent of the features are selected after arranging the feature variances in a descending order for each split. The dimension of f after feature subset selection is $d = 106$ and features have been selected with a split size of 200 × 200 pixels. Larger split size reduces the sensitivity of variance to change class thus increasing possible missed detection. Smaller split size allows to capture features that increase false alarm. Experiments were conducted to see the effect of using different thresholding strategies for discrimination of changed and unchanged pixels, the results are reported for each of the local adaptive thresholding and Otsu segmentation method separately (see Figure 4). The algorithm automatically takes as an input all the images for each couple of years, detects the changes and shows the changes for the entire processing period.

4. DISCUSSION

As it is shown in Figure 4, changes for each year have been visualized by different colors in order to make a better comparison between the different transformation and thresholding strategies. By analysing Figure 4 (a) and (b) (where the Extrusion transformation method is tested) and Figure 4 (c) and (d) (where the Rotation transformation method is tested), it is clear that both Extrusion and Rotation methods have a reliable performance and the changed areas are detected correctly. However, by looking deeply at the change maps it is revealed that Rotation transformation better models the borders of the crop fields, while Extrusion method has merged some of the crop fields together.

Figure 5 provides an RGB image of a zoom area (112 × 112 pixels) for the years from 2014 to 2018. For it a change reference map has been designed by photointerpretation

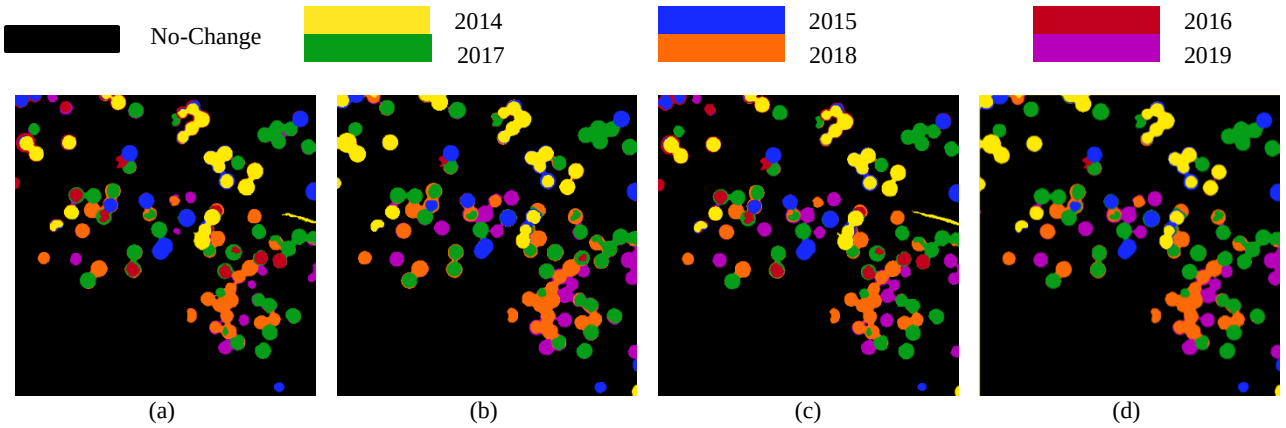

Figure 4. CD maps for the years 2013 to 2019 using: (a) Extrusion transformation and adaptive thresholding, (b) Extrusion transformation and Otsu thresholding, (c) Rotation transformation and adaptive thresholding and (d) Rotation transformation and Otsu thresholding.

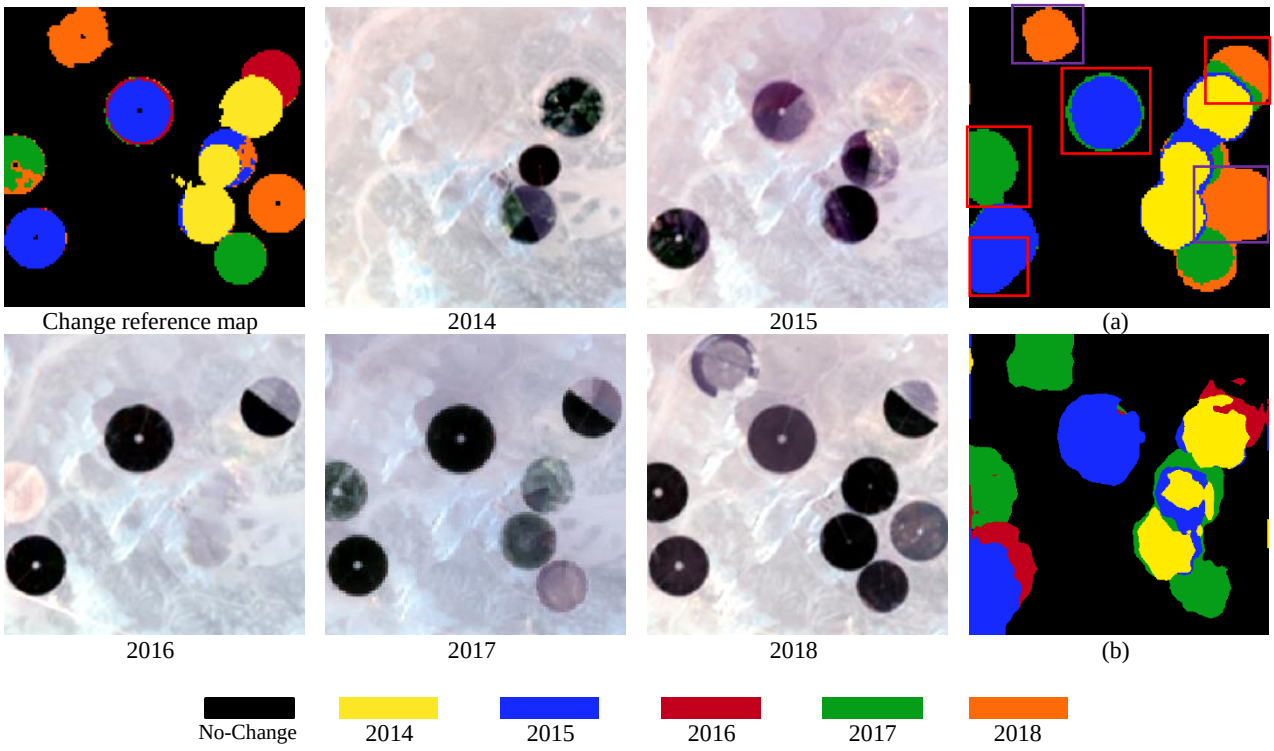

Figure 5. Change reference map, Landsat images of the years 2014 to 2018 and the CD maps using: (a) proposed approach with Rotation transformation and Otsu thresholding, (b) methodology developed in (Meshkini et al., 2021)

and prior knowledge on the evolution of the geographic area. For this area, 13 crop fields have been constructed during the processing period (2013 to 2019) resulting in 4583 changed pixels. A quantitative and qualitative performance analysis is conducted in terms of false and missed alarms in space and time. Figure 6 locates false/missed alarms showing that they mostly exist along the borders of crop fields. A detailed quantitative performance analysis is provided in Table 2. It clearly shows that the portion of false and missed alarms is considerably low in terms of pixels (less than 5% and 2%, respectively). In addition, an object-based analysis is conducted considering the 13 crop fields being changed in reference map. Three fields have been detected correctly in space, but they represent false alarms in time since the year of change is imprecise (i.e., the change is detected later or earlier than expected) (Figure 5 (a), red squares). One crop filed is a pure false alarm both in space and time (Figure 5 (a), red square bottom left corner).

In order to further prove the robustness of the proposed approach, the performance is compared with the 3D CD methodology that is presented in (Meshkini et al., 2021). In (Meshkini et al., 2021), a pretrained 3D CNN CD technique trained on a largescale video dataset is used that accepts a specific size of images with only three spectral bands (for more details refer to (Meshkini et al., 2021)). As it is clear, the proposed approach has outperformed the method in the literature in detecting the changes and defining the edges of the crop fields. Furthermore, by looking at the sample images provided in 2017 and 2018, there are two crop fields (one in the top left and another one almost right centre) that have started to be built in 2017 and continued in 2018. The reference method failed to recognize the crop field located in the right centre of area and it was unable to specify the year of the change correctly for the crop field in the top left (purple squares in Figure 5 (a)).

Year	False alarms (pixels %)	Missed alarms (pixels %)	False change fields	False change years
2014	3.97	0	0	0
2015	4.86	0.02	0	1
2016	3.67	0.35	1	0
2017	2.35	0.05	0	1
2018	2.47	1.05	0	1
Overall	14.03	1.49	1	3

Table 2. Validation result of the proposed CD method.

5. CONCLUSION

An unsupervised CD approach to the analysis of HR SITS based on a 3D CNN has been proposed. The approach uses a pretrained 2D CNN architecture for semantic segmentation to design a 3D CNN model that can deal with multi-temporal information. A 2D to 3D transformation module is implemented to transform 2D weights to 3D weights by using the Extrusion and Rotation strategies. The 2D convolutional layers have been replaced by 3D convolutional layers, the relevant features are extracted from some of the convolutional layers and are reduced by considering a variance measurement as an index of sensitivity to change information. The CD map is produced by extending the CVA strategy to SITS analysis that separates changed from unchanged pixel calculating the magnitude of the hyper feature map for each pixel. The reliability of the proposed approach is demonstrated by comparing with a 3D CNN-based CD from the literature. Quantitative analysis shows a decrease of false/missed alarms, a better capability to recognize the year of change and locate the object borders.

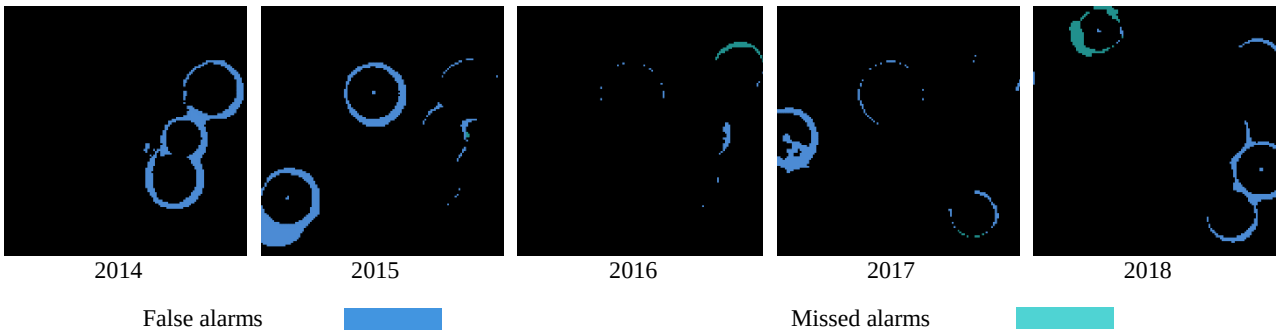

Figure 6. False/missed alarms of CD maps derived from the proposed method.

As a future work, we aim to improve the performance of the proposed method by employing a fine-tuning technique for the 3D CNN feature extraction, adding more comparisons using features extracted via 2D CNNs and performing the proposed method on the other areas with different change classes.

REFERENCES

Bovolo, F., Bruzzone, L., 2017. A split-based approach to unsupervised change detection in large-size multitemporal images: Application to tsunami-damage assessment. IEEE Trans Geosci Remote Sens 45, 1658–1670.

Bovolo, F., Bruzzone, L., 2007. A Theoretical Framework for Unsupervised Change Detection Based on Change Vector Analysis in the Polar Domain. IEEE Trans. Geosci. Remote Sens. 45, 218–236.

Bovolo, F., Bruzzone, L., 2005. A multilevel parcel-based approach to change detection in very high resolution multitemporal images. Proc. 2005 IEEE Int. Geosci. Remote Sens. Symp. IGARSS 05.

Bovolo, S., Marchesi, S., Bruzzone, L., 2012. A framework for automatic and unsupervised detection of multiple changes in multitemporal images. IEEE Trans. Geosci. Remote Sens. 2196–2212.

Bruzzone, L., Pietro, D.F., 2000. Automatic analysis of the difference image for unsupervised change detection. IEEE Trans. Geosci. Remote Sens. 1171–1182.

Celik, T., 2009. Unsupervised change detection in satellite images using principal component analysis and k-means clustering. IEEE Trans. Geosci. Remote Lett. 772–776.

Desclée et al, B., 2006. Forest change detection by statistical object-based method. Remote Sens. Environ. 1–16.

El Amin, A.M., Liu, Q., Wang, Y., 2016. Convolutional neural network features based change detection in satellite images. Proc SPIE 10011, 100110W.

Francisco et al, A.G., 2021. Deep learning-based object level change detection in overhead imagery. Proc SPIE 11729 Autom. Target Recognit. XXXI.

Funke et al, I., 2019. Using 3D convolutional neural networks to learn spatiotemporal features for automatic surgical gesture recognition in video. MICCAI 2019 LNCS 11768, 467–475.

Hariharan, B., Arbeláez, P., Girshick, R., Malik, J., 2015. Hypercolumns for object segmentation and fine-grained localization. Proc IEEE Conf Comput Vis Pattern Recognit 447–456.

Hou, B., Wang, Y., Liu, Q., 2017. Change detection based on deep features and low rank. IEEE Geosci Remote Sens Lett 14, 2418–2422.

Hulley, G., Veraverbeke, S., Hook, S., 2014. Thermal-based techniques for land cover change detection using a new dynamic MODIS multispectral emissivity product (MOD21). Remote Sens Env. 140, 755–765.

Kasetkasem, P., Varshney, K., 2002. An image change detection algorithm based on Markov random field models. IEEE Trans. Geosci. Remote Sens. 40, 1815–1823.

Khan, H.K., Xuming, H., Fatih, P., Mohammed, B., 2017. Forest Change Detection in Incomplete Satellite Images with Deep Neural Networks. IEEE Trans. Geosci. Remote Sens. Lett. 55, 5407–5423.

Kieri, A., 2012. Context dependent thresholding and filter selection for optical character recognition. Upps. Univ. Upps. Swed. Tech Rep UPTEC F12036.

Krizhevsky, A., Sutskever, I., Hinton, G., 2016. Imagenet classification with deep convolutional neural networks. Adv. Neural Inf. Process. Syst. 25, 1106–1114.

Li, Y., Zhang, H., Shen, Q., 2017. Spectral-spatial classification of hyper spectral imagery with 3D convolutional neural network. Remote Sens.

Lifan, Y., Xuan, P., Yandong, G., 2021. 3D CNN classification model for accurate diagnosis of coronavirus disease 2019 using computed tomography images. J Med. Imaging 8(S1).

Louis de Jong, K., Sergeevna Bosman, A., 2019. Unsupervised Change Detection in Satellite Images Using Convolutional Neural Networks. 2019 Int. Jt. Conf. Neural Netw. IJCNN.

Lu, D., Moran, E., Hetrick, S., 2011. Detection of impervious surface change with multitemporal Landsat images in an urban–rural frontier. ISPRS J Photogram Remote Sens 66, 298–306.

Malila, W.A., 1980. Change vector analysis: An approach for detecting forest changes with Landsat. Proc LARS Symp 385.

Merino, I., Azpiazu, J., Remazeilles, A., Sierra, B., 2021. 3D Convolutional Neural Networks Initialized from Pretrained 2D

Convolutional Neural Networks for Classification of Industrial Parts. Sensors 21, 1078.

Meshkini, K., Bovolo, F., Bruzzone, L., 2021. An Unsupervised Change Detection Approach for Dense Satellite Image Time Series Using 3D CNN. IEEE Int. Geosci. Remote Sens. Symp. IGARSS.

Mou, L., Bruzzone, L., Zhu, X.X., 2018. Learning spectral-spatial-temporal features via a recurrent convolutional neural network for change detection in multispectral imagery. IEEE Trans. Geosci. Remote Sens. 57, 924–935.

Nogueira, K., Penatti, O.A.B., dos Santos, J.A., 2017. Towards better exploiting convolutional neural networks for remote sensing scene classification. Pattern Recognit 61, 539–556.

Otsu, N., 1979. Threshold Selection Method from Gray-Level Histograms. IEEE Trans. Syst. Man Cybern. 9, 62–66.

Ren, S., He, R., Sun, J., 2015. Faster R-CNN: Towards realtime object detection with region proposal networks. Proc Adv Neural Inf Process Syst 91–99.

Saha, S., Bovolo, F., Bruzzone, L., 2019. Unsupervised Deep Change Vector Analysis for Multiple-Change Detection in VHR Images. IEEE Trans. Geosci. REMOTE Sens. 57.

Sexton, et. al, J.O., 2013. Long-term land cover dynamics by multi-temporal classification across the Landsat-5 record. Remote Sens Env. 246–257.

Solano-Correa, Y.T., Meshkini, K., Bovolo, F., Bruzzone, L., 2020. A Land-cover Driven Approach for Fitting Satellite Image Time Series in a Change Detection Context. Image Signal Process. Remote Sens. XXVI 11533.

Stramondo, S., Bignami, C., Chini, M., Pierdicca, N., Tertulliani, A., 2006. Satellite radar and optical remote sensing for earthquake damage detection: Results from different case studies. Int J Remote Sens 27, 4433–4447.

Volpi, M., Tuia, D., 2016. Dense semantic labeling of subdecimeter resolution images with convolutional neural networks. IEEE Trans Geosci Remote Sens 55, 881–893.

Wellner, P.D., 1993. Adaptive thresholding for the digitaldesk. Rank Xerox Res Cent. Camb. UK Tech Rep EPC1993-110 1–19.

Zanetti, M., Bruzzone, L., 2018. A theoretical framework for change detection based on a compound multiclass statistical model of the difference image. IEEE Trans. Geosci. Remote Sens. 56, 1129–1143.

Zeiler, M.D., Fergus, R., 2014. Visualizing and understanding convolutional networks. Proc Eur Conf Comput Vis Cham Switz. Springer 818–833.

Zhan, Y., Fu, K., Yan, M., Sun, X., Wang, H., Qiu, X., 2017. Change detection based on deep Siamese convolutional network for optical aerial images. IEEE Geosci Remote Sens Lett 14, 1845–1849.

Zhang, L., Zhang, L., Du, B., 2016. Deep learning for remote sensing data: A technical tutorial on the state of the Art. IEEE Geosci Remote Sens Mag 4, 22–40.