Change Detection in Hyperdimensional Images using Untrained Models

Sudipan Saha, *Member, IEEE,* Lukas Kondmann, Qian Song, *Member, IEEE,* and Xiao Xiang Zhu, *Fellow, IEEE*

***Abstract*—Deep transfer-learning based change detection methods are dependent on the availability of sensor-specific pre-trained feature extractors. Such feature extractors are not always available due to lack of training data, especially for hyperspectral sensors and other hyperdimensional images. Moreover models trained on easily available multispectral (RGB/RGB-NIR) images cannot be reused on such hyperdimensional images due to their irregular number of bands. While hyperdimensional images show large number of spectral bands, they generally show much less spatial complexity, thus reducing the requirement of large receptive fields of convolution filters. Recent works in the computer vision have shown that even untrained deep models can yield remarkable result in some tasks like super-resolution and surface reconstruction. This motivates us to make a bold proposition that untrained lightweight deep model, initialized with some weight initialization strategy, can be used to extract useful semantic features from bi-temporal hyperdimensional images. Based on this proposition, we design a novel change detection framework for hyperdimensional images by extracting bi-temporal features using an untrained model and further comparing the extracted features using Deep Change Vector Analysis to distinguish changed pixels from the unchanged ones. We further use the deep change hypervectors to cluster the changed pixels into different semantic groups. We conduct experiments on four change detection datasets: three hyperspectral datasets and a hyperdimensional Polarimetric Synthetic Aperture Radar dataset. The results clearly demonstrate that the proposed method is suitable for change detection in hyperdimensional remote sensing data. Code is available at https://gitlab.lrz.de/ai4eo/cd/-/tree/main/hyperdimensionalCD**

***Index Terms*—Change Detection, Deep Learning, Deep Image Prior, Hyperspectral Images, Hyperdimensional Images.**

The work is jointly supported by by the German Federal Ministry of Education and Research (BMBF) in the framework of the international future AI lab "AI4EO – Artificial Intelligence for Earth Observation: Reasoning, Uncertainties, Ethics and Beyond" (grant number: 01DD20001), the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. [ERC-2016-StG-714087], Acronym: *So2Sat*), by the Helmholtz Association through the Framework of Helmholtz AI (grant number: ZT-I-PF-5-01) - Local Unit "Munich Unit @Aeronautics, Space and Transport (MASTr)" and Helmholtz Excellent Professorship "Data Science in Earth Observation - Big Data Fusion for Urban Research"(grant number: W2-W3-100) (Corresponding author: Xiao Xiang Zhu.)

Sudipan Saha is with Department of Aerospace and Geodesy, Data Science in Earth Observation, Technical University of Munich, 85521 Ottobrunn, Germany. E-mail: sudipan.saha@tum.de

Lukas Kondmann and Qian Song are with Remote Sensing Technology Institute, German Aerospace Center (DLR), 82234 Weßling, Germany. E-mail: lukas.kondmann@dlr.de; qian.song@dlr.de

Xiao Xiang Zhu is with the Remote Sensing Technology Institute (IMF), German Aerospace Center (DLR), 82234 Weßling, Germany and with Department of Aerospace and Geodesy, Data Science in Earth Observation (SiPEO, former: Signal Processing in Earth Observation), Technical University of Munich, 85521 Ottobrunn, Germany E-mail: xiaoxiang.zhu@dlr.de

I. INTRODUCTION

Recently deep learning has attracted significant attention in Earth observation [1]. Following this trend, deep learning based methods have been developed for change detection (CD) [2], an important topic in Earth observation. Change detection plays pivotal role in several applications, including disaster management [3], [4], urban monitoring [5], and precision agriculture [6]. While change detection methods can be supervised [7], [8], [9] or semi-supervised [5], unsupervised methods are preferred in the literature [10], [2] as collecting labeled multi-temporal data is significantly challenging. Before the emergence of deep learning, change vector analysis (CVA) and its object-based variants [10], [11] were popularly used for unsupervised CD. Deep CVA (DCVA) and other transfer learning based methods [2], [12], [3] have embedded the concept of CVA in a transfer learning framework. While the transfer learning based methods do not use any training or fine-tuning of the deep model, they depend on the availability of pre-trained feature extractor that can be used to capture the semantics of the input images. In more details, such transfer learning based methods project the bi-temporal images in deep featurespace by using a pre-trained deep feature extractor and subsequently compares the images in the projected domain. Thus they perform change detection by reusing a deep model that was previously trained for some unrelated task, e.g., image classification. Most deep transfer learning based CD methods are designed for Synthetic Aperture Radar (SAR) amplitude images and multispectral images with few bands.

Remote sensing deals with a plethora of sensors showing different spatial, spectral, and temporal characteristics. In many cases, large number of bands are required to efficiently represent the information in remote sensing images. The most well known example for this are hyperspectral images that sample a broad range of electromagnetic spectrum in hundreds of spectral bands [13], [14], [15], [16], [17]. Some CD applications require rich spectral information and hyperspectral images can be very useful for such cases, e.g., monitoring of mining activity [18]. Inspite of this, less attention has been paid to develop deep transfer learning based CD methods for hyperspectral images [19], [20]. This can be attributed to the lack of labeled hyperspectral data that impedes availability of any pre-trained network. In more details, a transfer learning based hyperspectral CD method can be developed only if a pre-trained model is available for the same data, which is often unavailable for hyperspectral images. Remarkably, due to the lack of training data, some of the supervised hyperspectral image classification models are trained and tested on pixels from the same image [21]. Even if sufficient training data is collected for a particular hyperspectral sensor and geography, this model will not be straightforward applicable for another hyperspectral sensor. Currently there are a large number of hyperspectral sensors with differences in spectral coverage and number of bands. E.g., DLR Earth sensing imaging spectrometer (DESIS) have 180 bands while precursore iperspettrale della missione applicativa (PRISMA) have 237 bands [13]. Please see Table I for comparison of number of bands of different spaceborn hyperspectral sensors. Due to such differences, a model trained for one hyperspectral sensor cannot be used for transfer-learning based CD on another hyperspectral sensor. Additionally, unmanned-aerial-vehicle (UAV)-based hyperspectral imaging has become increasingly popular in various applications, such as agricultural monitoring [13]. Such UAV-based hyperspectral sensors may exhibit spectral coverage entirely different from the satellite-based ones.

In addition to hyperspectral data, another example of hyperdimensional data in remote sensing is polarimetric synthetic aperture radar (PolSAR) image. Compared with the single-polarimetric SAR data, PolSAR images contain more polarimetric information about the targets and are useful to discriminate double-bounce scatterers (such as buildings) from volume scatterers (such as forest) and surfaces using target decomposition methods [22]. Thus PolSAR data is beneficial for applications such as land classification and building extraction [23]. In practical PolSAR applications, usually the decomposed results [23], [24], [25] instead of the raw PolSAR data are used for further analysis, which constitutes a hyperdimensional (tens to over one hundred channels) data cuboid.

Models trained for multispectral (RGB/RGB-NIR) or SAR amplitude images cannot be effectively reused for feature extraction of hyperdimensional images due to their irregular number of bands. To transfer RGB-trained models on hyperdimensional images, we require to choose only three bands from hyperdimensional images, thus losing a significant amount of information. Another possible solution is to somehow modify the first layer of the pre-trained model.

Ulyanov *et. al.* [26] showed that the structure of a network is often sufficient to capture important low-level features from the images without any training. This is highly relevant for hyperdimensional images since it is challenging to transfer a model trained on RGB images to hyperdimensional images, however it is trivial to just initialize a model to ingest as many number of image channels as desired. This strategy is certainly not as good as learning complex spatial features with abundant labeled images, however good enough for change detection in hyperdimensional images. Arguably, the spatial complexity of hyperdimensional images is not high in most cases, as can be seen in Table I. This is also evident from the fact that some works in the hyperspectral image classification just use 1D convolution [27]. While spatial complexity still has an important role to play for hyperspectral multi-temporal analysis, we argue that this is not as critical as in high-resolution multispectral images. This brings forth the possibility whether complexity in low-spatial and high-spectral resolution multitemporal hyperdimensional images can be captured by an untrained deep model, merely initialized with a deep model initialization strategy [28] [29]. The likelihood of such possibility is supported by the fact that untrained models have recently shown remarkable performance in some computer vision tasks where the spatial complexity is much more critical than the hyperspectral images, e.g., deep image prior [26].

TABLE I
NUMBER OF BANDS AND GROUND SAMPLING DISTANCE (GSD) FOR SOME SPACEBORN HYPERSPECTRAL SENSORS [13]

Sensor	Bands	GSD (m)
DESIS	180	30
EnMAP	228	30
PRISMA	237	30
HISUI	185	30
HySIS	256	30
Shalom	241	10
CCRSS	328	30

We propose an unsupervised CD method for hyperdimensional images using an untrained deep model as deep feature extractor. The proposed method does not need any prior knowledge about the input or the arrangement of the spectral bands. In addition to distinguishing the changed pixels from the unchanged ones (binary CD), we also extend the method for multiple CD. The key contributions of this paper are as follows:

1) The paper shows that even an untrained model, merely initialized with a weight initialization technique [28], can be used to capture the spatio-temporal semantics, especially for hyperdimensional data where pretrained models are generally not available. Based on this, the paper proposes a change detection method which can effectively segregate changed pixels from the unchanged ones in the hyperdimensional images.
2) The paper further extends the method for multiple/multi-class CD using deep change vector obtained using untrained model to cluster the changed pixels into different groups.
3) The paper experimentally validates the proposed approach on three bi-temporal hyperspectral scenes, as well on a bi-temporal hyperdimensional PolSAR data, showing the versatility of the approach.

We organize the paper as follows. Some relevant works are discussed in Section II. Section III discusses the proposed method. Section IV presents the datasets and results related to hyperspectral images. Results related to PolSAR data are presented in Section V. Finally we conclude this paper in Section VI.

II. RELATED WORK

Following the relevance to our work, we briefly discuss in this section about: i) unsupervised CD, ii) hyperdimensional CD methods, and iii) deep image prior.

A. Unsupervised CD

Unsupervised CD methods are generally based on the concept of pixewise difference operation, i.e., change vector

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSTARS.2021.3121556, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing

analysis (CVA) [30] or clustering [31]. With the emergence of high-resolution imaging, object-based variants of CVA, e.g., Parcel Change Vector Analysis (PCVA) [11], incorporated the notion of spatial context in CVA. Morphological filters have also been employed to capture the object information [32]. Deep learning based unsupervised CD methods, e.g., DCVA [2] are based on transfer learning. DCVA incorporates CVA with pre-trained deep network based feature extraction based on the assumption that a pre-trained model is available for the target geography and sensor. In addition to optical images, transfer learning based frameworks have also shown success in SAR amplitude image analysis [3].

B. CD in hyperdimensional images

Very few deep learning based CD methods have been proposed for hyperdimensional (hyperspectral or other hyperdimensional) images [33], [34], [35]. In [33], authors identified high dimension and limited datasets as unique challenges for hyperspectral CD. Towards alleviating these challenges, they devised a pre-classification based end-to-end CD framework. Another supervised framework recurrent three-dimensional (3D) fully convolutional network (Re3FCN) was introduced by Song *et. al.* [35]. Re3FCN merges a 3D fully convolutional network (FCN) and a convolutional long short-term memory (ConvLSTM). Chen and Zhou [36] proposed a supervised CD method consisting of three steps: reduction of spectral dimension, joint affinity tensor construction, and binary (changed or unchanged) classification by CNN. While these works successfully introduce deep learning to the hyperspectral change detection, they do not present any unique solution towards circumventing the limited availability of datasets in hyperspectral multitemporal analysis. Their works use pixels from same image for training and evaluation. Using such large supervised networks when training and test pixels belong to same scene may lead to overoptimistic accuracy assessment, as shown by Molinier and Kilpi [37]. Thus it is crucial to design unsupervised/transfer-learning-based approaches, like the ones proposed for multispectral and SAR images [2] [3]. In addition to hyperspectral images, hyperdimensional CD has also been studied in the context of PolSAR images [24]. To the best of our knowledge, all deep learning based hyperdimensional CD methods are proposed in context of binary CD, without delving into multiple/multi-class CD.

C. Deep image prior

Deep models are generally trained on large labeled datasets. This makes us to believe that the excellent performance of CNNs are due to their capability to learn realistic features or data priors from the data. However, several recent works have shown that this explanation is not entirely correct. In one of such first works, [38] showed that an image classification network can overfit on the training images even when the labels are randomized. This provides us hints that the success of the deep network is possibly not always due to large amount of labeled data, rather sometimes due to the structure of the network. Further delving into this topic, Ulyanov *et. al.* [26] investigated this phenomenon in context of image generation. They showed that a large amount of the image statistics are captured by the structure of generator CNNs itself. Instead of choosing the usual paradigm of training CNNs on large dataset, they fitted CNNs on single image for image restoration problems. The network weights were randomly initialized. Their simple setup could provide remarkable result for various image restoration problems, e.g., denoising and super-resolution. This phenomenon is remarkable as it demonstrates the power of untrained network. Following this work, several other works have followed similar approach demonstrating success of untrained network for different computer vision problems, including surface reconstruction [39] and photo manipulation [40]. Another similar line of research is random projection network [41] that is proposed in the context of high-dimensional data which implies a network architecture with an input layer that has a huge number of weights, making training infeasible. Random projection network [41] tackles this challenge by prepending the network with an input layer whose weights are initialized with a random projection matrix.

III. Proposed method

Let us assume that we have a pair of coregistered hyperdimensional images X_1 and X_2 having B_0 bands, where B_0 is much larger than usual number of bands in a multispectral image. No training label or suitable pre-trained network is available to us. Our goal is two fold:

1) *Binary CD:* Distinguish the changed pixels (Ω_c) from the unchanged ones (ω_{nc}).
2) *Multiple CD:* Further cluster the changed pixels into a group of semantically meaningful groups.

To accomplish the above-mentioned goals, we initialize a deep model with number of input channels and kernels in intermediate layers modulated according to the dimension of the X_1 and X_2. This deep model, while untrained, is initialized with an appropriate weight initialization technique [28]. Following this, we use this network to extract a set of features from the bi-temporal images. Pixelwise difference is obtained as deep change vector that is thresholded to identify the changed pixels. Once changed pixels are segregated, they are further clustered based on the deep change vectors for multiple change detection. The proposed hyperdimensional CD framework is called Untrained Hyperdimensional Multiple DCVA (UHM-DCVA) and is shown in Figure 1.

A. Feature extraction

Deep models trained for multi-spectral images can ingest input images of few channels/bands, in order of three to ten [42], [43]. In contrast, hyperdimensional remote sensing images have B_0 channels that is generally larger than 200. Thus deep models trained on multi-spectral images are not suitable to ingest hyperdimensional X_1 and X_2. To overcome this challenge, we use an untrained model for deep feature extraction from X_1 and X_2. The model, being untrained, can be initialized with capacity to ingest any number of input channels and subsequently projected to any number of kernels in the successive layers.

Conforming to the dimension of X_1 and X_2, we design first convolution layer such that it ingests the hyperdimensional image of B_0 channels and projects it to $\beta_0 B_0$ kernels where $\beta_0 > 1$. We use 3×3 filters, i.e., weight of first layer is $3 \times 3 \times B_0 \times \beta_0 B_0$. In our experiments, we set $\beta_0 = 4$. The following convolution layer ingests input of dimension $\beta_0 B_0$ and projects it to $\beta_1 \beta_0 B_0$ dimension. For simplicity, we have set $\beta_1 = 1$. In this way, more layers can be added to the network. Increasing number of layers capture larger spatial receptive field. Considering the coarse spatial resolution of the most hyperdimensional images, we postulate that network need not be as deeper as it is common in multi-spectral image analysis (further validated in Section IV). Rectified Linear Unit (ReLU) function is used between successive convolution layers. Pooling operation and fully connected layers are not used. Hence the spatial size of the input is preserved through successive layers. Key structure of the network is shown in Table II and Figure 2.

Though untrained, the weights are initialized with He initialization method [28]. Their weight initialization strategy allows the initialized elements to be mutually independent and share the same distribution. Though weight initialization was initially proposed in context of obtaining efficient starting point for better training, we use it to obtain a superior feature extractor that can be subsequently used as deep feature extractor in proposed change detection framework. Note that weight initialization does not involve any training. Once initialized, the deep model is used to extract a set of features from both X_1 and X_2 separately, as detailed in Section III-B.

B. Binary CD

All bands of X_1 and X_2 are normalized to have values between 0 and 1. Untrained model is separately applied on X_1 and X_2 to extract a set of deep features for each pixel in the scene [2]. Using same model on both images ensure that two very similar inputs (pixels) are mapped to similar representation in the feature space while dissimilar pixels are mapped to dissimilar feature representation, since they are processed through same set of functions. Furthermore, a variance-based feature selection strategy is applied as in [2]. Deep features are extracted from the last layer of the network to form pixel wise deep change hypervector (G) [2] that are obtained as the deep-feature-differences of X_1 and X_2. Components of G $(g^d \ (d = 1, ..., D))$ tend to zero for unchanged pixels (ω_{nc}) while they tend to larger (positive or negative) value for the changed pixels (Ω_c). To segregate Ω_c from ω_{nc}, we compute deep magnitude ρ for each pixel as the Euclidean norm of G:

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

ρ maps the $D-$dimensional G into a $1-$dimensional index, while preserving the main properties of the changes. Unchanged pixels tend to generate smaller ρ in comparison to the changed pixels. This is used to segregate Ω_c and ω_{nc} by using a thresholding τ. While any suitable thresholding [44] method can be used, we use Otsu's thresholding [45] to compute τ. Any pixel having $\rho > \tau$ is assigned to Ω_c and to ω_{nc} otherwise.

C. Multiple CD

Changed pixels (Ω_c) are further analyzed in unsupervised way based on G to segregate different kinds of change without any a priori knowledge about the different kinds of change [2]. However, we assume an apriori knowledge about number of kinds of change (K). G is a high dimensional vector and clustering is challenging in such high-dimensional space [46]. To overcome this, we first binarize/discretize the components of G [2], [47]. Components of G are likely to be either positive or negative, and different kinds of change are likely to show different patterns on the $g^d \ (d = 1, ..., D)$, components of G. Binarization simplifies the information in G, while preserving information descriptive of clusters. G is binarized to G_{bin} with components greater than 0 set to 1 and components smaller than 0 set to 0. G_{bin} is also $D-$dimensional like G.

Assuming number of changed pixels (pixels in Ω_c) as N_c, we have N_c binary vectors of $D-$dimension each. Conversely, representing each feature as a vector, we have D vectors of N_c-dimension each. We expect pixels belonging to same kind of change to exhibit similar binary signature, while pixels belonging to different kinds of change to exhibit dissimilar binary signature. Furthermore, many features exhibit similar binary signature and thus redundant for discriminating different types of change. Out of D features, the feature which shows most similarity to other $D-1$ features can be defined as the most informative feature. Towards this, $\mathcal{R}(i,j)$ measures the correlation distance [48] between two N_c-dimensional features i and j, scaled in range 0-1 [2], where 1 represents the farthest features. $R_d \ (d \ (d = 1, ..., D))$ measures the informativeness of an individual feature:

$$R_d = -\sum_{j=1}^{D} \mathcal{R}(d,j) \tag{2}$$

In the above equation, while the term within summation computes distance of a feature from other features, coupled with the negation, R_d measures how similar is the feature d to the other $D-1$ features. The most informative feature d^* is selected by choosing the feature that maximizes R_d:

$$d^* = \arg\max_d R_d \tag{3}$$

Chosen d^* can be used to group pixels in Ω_c into two classes. Next most informative feature can be selected by following the above-mentioned process, but first discarding the most informative feature d^* and features made redundant by it. This hierarchical process allows us to select a set of informative features that are further used to cluster Ω_c into desired number of classes $\omega_{c1}, \omega_{c2}, ..., \omega_{cK}$.

IV. Validation on hyperspectral data

A. Datasets

We validate the proposed method on three publicly available bi-temporal hyperspectral scenes [49], [50] [1]:

[1]https://citius.usc.es/investigacion/datasets/hyperspectral-change-detection-dataset

TABLE II
KEY STRUCTURE OF 5-LAYER UNTRAINED FEATURE EXTRACTOR NETWORK ASSUMING NUMBER OF CHANNELS IN INPUT IMAGE IS 224. ALL CONVOLUTION LAYERS ARE FOLLOWED BY RELU ACTIVATION.

Layer number	Layer type	Input Kernel	Output Kernel	Kernel size
1	convolutional	224	896	(3,3)
2	convolutional	896	896	(3,3)
3	convolutional	896	896	(3,3)
4	convolutional	896	896	(3,3)
5	convolutional	896	896	(3,3)

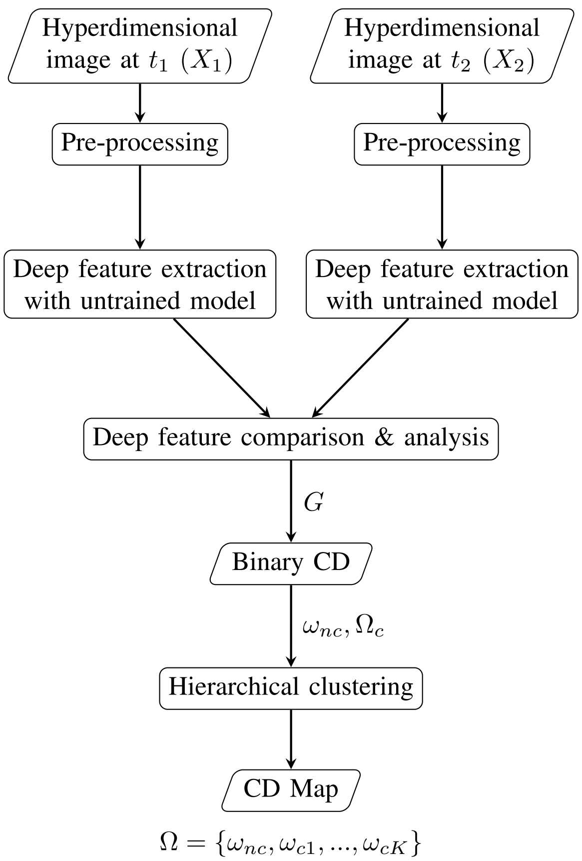


Fig. 1. Proposed Untrained Hyperdimensional Multiple Deep CVA (UHM-DCVA) technique.

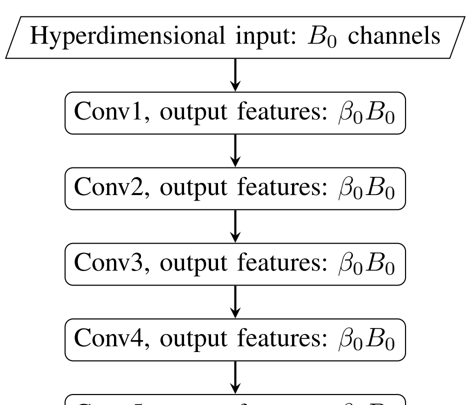


Fig. 2. The simplified network architecture considering 5 layers.

1) The Santa Barbara bi-temporal scene is acquired on 2013 (Figure 3(a)) and 2014 (Figure 3(b)) with the AVIRIS sensor (224 spectral bands) over the Santa Barbara region in California, United States. The spatial dimension of the images are 984 $\times$ 740 pixels. Reference information is known for only 132552 pixels, out of which 80418 pixels are unchanged and 52134 pixels are changed (Figure 3(c)).
2) The Bay Area bi-temporal scene is acquired on 2013 (Figure 5(a)) and 2015 (Figure 5(b)) with the AVIRIS sensor (224 spectral bands) over the area surrounding the city of Patterson (California). The spatial dimension of the images are 500 $\times$ 500 pixels. Reference information is known for only 60610 pixels, out of which 29393 pixels are unchanged and 31217 pixels are changed (Figure 5(c)).
3) The Hermiston scene (Figures 6(a) and 6(b)) is acquired on the years 2004 and 2007 with the Hyperion sensor (242 spectral bands) over the Hermiston City area in Oregon, United States. Bands B001-B007, B058-B076, and B225-242 are not calibrated, hence we exclude them from our processing. The spatial dimension of the images are 390 $\times$ 200 pixels. 68014 pixels are labeled as unchanged. Remaining pixels are changed (Figure 6(c)). The changed pixels are further grouped into 5 change types: type 1 (5558 pixels), type 2 (1331 pixels), type 3 (79 pixels), type 4 (1557 pixels), and type 5 (1461 pixels), shown in Figure 7(a).

Please note that:

1) For Santa Barbara and Bay Area scene, reference information is not known for a fraction of pixels. However, these datasets are not prepared by us and are publicly available datasets used in previous research works [49], [50]. Hence, we follow the reference maps available with those datasets.
2) We evaluate binary CD method on all three scenes, however multiple/multi-class CD method on only Hermiston scene, as multiple change reference map is available for only this scene.

B. Compared methods

We compared the proposed method to following unsupervised methods:

- Change vector analysis (CVA) using the hyperdimensional pixel values. The comparison to CVA is crucial to understand whether the proposed method provides any additional benefit over mere pixel difference.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSTARS.2021.3121556, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing

- Parcel change vector analysis (PCVA) [11] that captures the spatial information as superpixel. This comparison helps to understand whether spatio-temporal context in hyperdimensional images can be simply captured by a superpixel-based analysis.
- Spectral Angle Mapper Z-score Image Differencing (SAMZID) [19] that is designed specifically for hyperspectral CD based on spectral angle mapper and image difference. The method, as proposed in [19] originally consists of an unsupervised predictor phase and a supervised learning phase. We exclude the supervised phase and apply thresholding [45] on the change map obtained after unsupervised predictor phase. As proposed in [19], two variants are compared: $\text{SAMZID}_{\text{Sin}}$ and $\text{SAMZID}_{\text{Tan}}$.
- Autoencoding of bi-temporal Hyperspectral Images for Change Vector Analysis (AICA) [34] - a deep learning based unsupervised change detection method proposed for hyperspectral images that combines CVA with autoencoder-based training.
- Deep change vector analysis (DCVA) [2] with feature extractor pretrained on largescale computer vision dataset using VGG16/VGG19 architecture [42]. This comparison is important to understand whether a simple transfer learning approach can be used instead of the proposed method. Pre-trained VGG architecture can ingest only three channels. So we just select three optimum (RGB) channels from the hyperspectral image to feed to the network. We use three different configurations: by using 1st convolutional layer of VGG16 (DCVA3Channels-1), 2nd convolutional layer of VGG16 (DCVA3Channels-2), and 5th convolutional layer of VGG16 (DCVA3Channels-3).
- DCVA as above, however in this case we modulate the first layer of the network by replicating the weights as number of channels of hyperspectral images. In this way we can feed the unmodified entire hyperspectral images to the network. We use two different configurations: by using 1st convolutional layer of VGG16 (DCVAAllChannels-1) and 2nd convolutional layer of VGG16 (DCVAAllChannels-2).
- A variant of the proposed method using dilated convolutional layers (dilation set as 3) to understand whether the proposed method can benefit from the larger receptive field.
- A 1D variant of the proposed method using 1×1 kernels instead of 3×3 kernels. This helps us to understand whether both the spatial context/spectral information contributed to the change detection result.

  The first two compared methods are from classical CD literature. The third and fourth methods are from hyperspectral CD literature that specifically exploit properties unique to hyperspectral images. The following two methods are based on deep transfer learning. The proposed method is unsupervised, does not require any training or even any pre-trained network, thus not compared to any supervised [36] or pre-classification [33] based hyperspectral CD method. The last two methods are variant of the proposed method and are shown on the Santa Barbara scene.

C. Settings and other details

The results are reported as average of 5 runs. Comparison is performed in terms of sensitivity (accuracy in percentage computed over reference changed pixels), specificity (accuracy in percentage computed over reference unchanged pixels), and overall accuracy. In more details, given true positive (TP), true negative (TN), false positive (FP) and false negative (FN), sensitivity is TP/(TP+FN), specificity is TN/(TN+FP), and accuracy is given by (TP+TN)/(TP+TN+FP+FN), all scaled by 100. For multiple CD, kappa score is provided.

We perform a number of additional experiments on the Santa Barbara scene:

1) For the proposed method, we use a 5-layer network, however we provide a comparison of performance as number of layers is changed.
2) For the proposed method we generally use He weight initialization method [28], however its performance with respect to another weight initialization method [29] is discussed.
3) For the proposed method we use Otsu's threshold determination method [45], however its performance with few other thresholding method is shown.
4) We show variation of result as β_0 is varied.

D. Binary CD results

1) Santa Barbara: We first analyze the impact of increasing number of layers for the proposed method (Table III). We observe that both sensitivity and specificity gradually increase up to 4 layers. Sensitivity increases while specificity slightly decreases when 5 layers are used. No performance gain is observed, rather decreases for 6 layers. While adding more convolution layers improve the spatial receptive field of the filters and increase the complexity of the filters, considering the coarse resolution of the hyperspectral images this behavior saturates soon. Henceforth, we use 5 layers for all experiments related to the proposed method.

CVA obtains a sensitivity of 76.92 and specificity of 96.69 (Figure 3(d)). Remarkably, PCVA performs worser than CVA, showing that spectral and temporal complexity of hyperspectral bi-temporal images cannot be captured by mere superpixel based representation. Being designed for hyperspectral CD, $\text{SAMZID}_{\text{Sin}}$, $\text{SAMZID}_{\text{Tan}}$, and AICA outperform CVA and PCVA. DCVAAllChannels-1 and DCVAAllChannels-2 are outperformed by the DCVA3Channels-1 (Figure 3(e)) and DCVA3Channels-2. This clearly shows that structure of the network is important. VGGNet architecture, originally proposed for 3-channel input, can work satisfactorily while ingesting only 3 out of 224 spectral bands of AVIRIS sensor. However, attempting to forcefully feed the network with all bands result in decrease in performance.

The proposed method (Figure 3(f) and Table IV) clearly outperforms all the compared methods (including its dilated and 1D variant), obtaining a sensitivity 87.98, specificity of 98.57, and accuracy of 94.40. This shows the superiority of the proposed method to ingest input bi-temporal images of arbitrary dimension, which cannot be achieved with transfer learning settings (DCVAAllChannels or DCVA3Channels). Proposed model can capture the change information, which is evident from visualization of two randomly selected features (in deep-difference domain) in Figure 4. Remarkably, the proposed method's 1D variant that only captures spectral context outperforms the dilated convolution based variant. This indicates that the spectral information plays more important role on change detection than the spatial context information for the considered hyperspectal data. This also partly explains why the proposed unsupervised method outperforms transfer learning from models trained on computer vision data.

The performance of the proposed method may vary if another weight initialization strategy is used instead of He initialization method [28]. E.g., if Xavier weight initialization [29] is used, the proposed method obtains a sensitivity of 80.12% and specificity of 94.27%, which is still superior to most compared methods in Table IV.

For thresholding the Otsu's method [45] is used, as it is popular in the unsupervised CD methods [51], [2]. However any other suitable method [52], [53], [54], [55] can be used with similar result as shown in Table V for ISODATA method [52], [53] and Li's method [54].

In Section III-A, we chose β_0 as 4. In Table VI, we show variation of result with different values of β_0 that supports the choice of above-mentioned value.

2) Bay Area: The Bay Area scene shows complex urban area along with vegetation patches. As in Santa Barbara, PCVA, DCVAAllChannels-1, and DCVAAllChannels-2 do not obtain satisfactory result. CVA (Figure 5(d)), $\text{SAMZID}_{\text{Sin}}$, $\text{SAMZID}_{\text{Tan}}$, AICA, DCVA3Channels-1 (Figure 5(e)) and DCVA3Channels-2 obtain superior result in comparison to them. The proposed method outperforms all of them, in terms of sensitivity, specificity, and accuracy (Figure 5(f)). Detailed quantitative results are shown in Table VII.

3) Hermiston: The spatial complexity of Hermiston is lesser compared to the other two scenes. The changes form simple geometric pattern in this scene. Results obtained for this scene is similar to the other two scenes. Quantitative results are shown in Table VIII. The proposed method (Figure 6(f)) either outperforms or obtains comparable specificity in comparison to other methods. The proposed method outperforms CVA (Figure 6(d)), PCVA, $\text{SAMZID}_{\text{Sin}}$, $\text{SAMZID}_{\text{Tan}}$, AICA, DCVAAllChannels-1, and DCVAAllChannels-2 also in terms of sensitivity. However, DCVA3Channels-1 and DCVA3Channels-2 obtain superior sensitivity than the proposed method. This relative success of transfer learning based setup on this dataset can be attributed to the less spatial complexity of the scene.

E. Multiple CD results

Multiple CD reference map is only available for Hermiston scene. The reference map is shown in Figure 7(a). Result obtained by the proposed method, using deep features extracted using untrained model, is shown in Figure 7(c). It is evident that the proposed method is able to detect the important semantic changes. There is certainly overlap between the classes shown in blue and red. However, it is clear from Figures 6(a) and 6(b), that the blue and red classes represent similar semantic notion, making it difficult for the unsupervised multiple CD method to differentiate them.

To understand whether the proposed multiple/multi-class CD scheme benefits from using the untrained model as feature extractor, we compare it to result obtained by using original hyperspectral data (Figure 7(b)). Proposed method is visually superior than this baseline. Proposed method obtains a kappa of 0.80, in comparison to 0.72, obtained using the original hyperspectral data.

TABLE III
PERFORMANCE VARIATION OF THE PROPOSED METHOD ON THE SANTA BARBARA SCENE AS NUMBER OF LAYERS ARE VARIED. ALL RESULTS ARE REPORTED AS AVERAGE OF 5 RUNS.

Method	Sensitivity	Specificity	Accuracy
Proposed (2 layers)	83.86	98.71	92.87
Proposed (3 layers)	83.90	98.96	93.04
Proposed (4 layers)	85.86	98.97	93.81
Proposed (5 layers)	87.98	98.57	94.40
Proposed (6 layers)	84.74	98.48	93.07

TABLE IV
CD RESULTS FOR THE SANTA BARBARA SCENE. PROPOSED METHOD'S RESULT IS REPORTED AS AVERAGE OF 5 RUNS.

Method	Sensitivity	Specificity	Accuracy
CVA	76.92	96.69	88.91
PCVA	58.18	84.74	74.29
$\text{SAMZID}_{\text{Sin}}$	80.67	97.01	90.58
$\text{SAMZID}_{\text{Tan}}$	79.64	98.43	91.04
AICA	87.25	94.52	91.66
DCVA3Channels-1	78.01	93.60	87.47
DCVA3Channels-2	66.70	86.90	78.96
DCVA3Channels-3	46.93	74.33	63.56
DCVAAllChannels-1	51.24	85.88	72.26
DCVAAllChannels-2	47.56	80.74	67.69
Dilated	77.42	95.95	88.66
1D Conv	80.01	98.93	91.49
Proposed (5 layers)	87.98	98.57	94.40±0.6

V. RESULTS ON DECOMPOSED POLSAR DATA

The decomposed POLSAR bi-temporal data is a pair of 138 band real-valued data acquired using UAVSAR over an urban area in San Francisco city on September 2009, and May 2015, first presented in work by Najafi *et. al.* [24]. We use the same set of methods as for hyperspectral CD for comparison except those specifically designed for hyperspectral images (SAMZID and AICA) and DCVA3Channels-1/2 as there are no available R, G, B bands in this case. Figure 8(a) shows the reference CD map. Proposed method obtains satisfactory result (Figure 8(c)), visually significantly better than CVA (Figure 8(b)). Proposed method quantitatively outperforms all compared methods, as tabulated in Table IX.

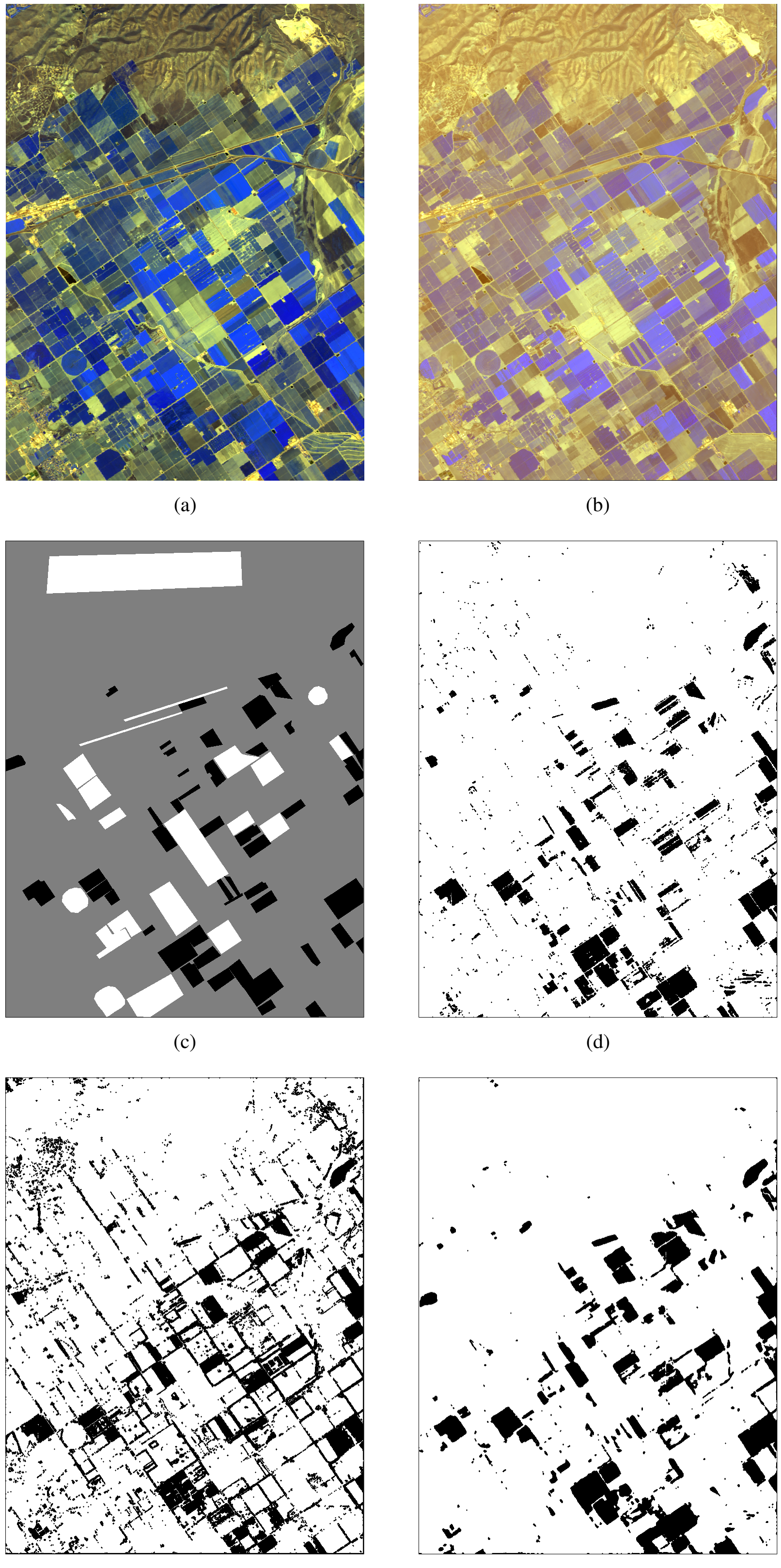
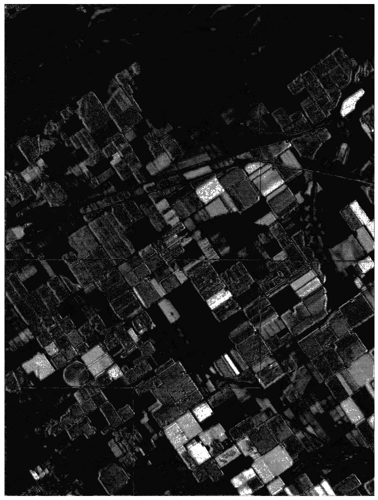


Fig. 3. Santa Barbara scene, False color composition (R: band 50, G: band 20, B: band 10) images: (a) pre-change and (b) post-change, (c) reference image (white - unchanged, black - changed, gray - unknown), and CD maps: (d) CVA, (e) DCVA3Channels-1, (f) Proposed


(a)

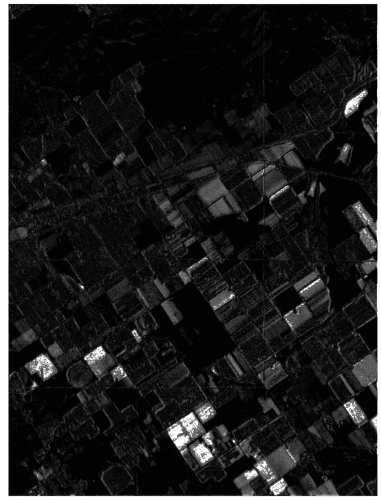

(b)

Fig. 4. Visualization of two randomly selected features, as generated by the proposed model, on the Santa Barbara scene. It is evident that the features capture the change information.

TABLE V
VARIATION OF THE RESULT FOR SANTA BARBARA SCENE AS THRESHOLD DETERMINATION SCHEME IS VARIED

Thresholding	Sensitivity	Specificity	Accuracy
Otsu	87.98	98.57	94.40
ISODATA	88.22	98.50	94.46
Li	95.71	93.23	94.20

TABLE VI
VARIATION OF THE RESULT FOR SANTA BARBARA SCENE AS β_0 IS VARIED

β_0	Sensitivity	Specificity	Accuracy
2	87.48	97.60	93.62
4	87.98	98.57	94.40
8	85.53	98.41	93.34

VI. CONCLUSION

In this work, we presented an unsupervised change detection method for hyperdimensional images. Labeled training data is scarce for hyperdimensional images and models trained on multispectral sensors cannot be directly applied on them, due to mismatch of dimension. The proposed method overcomes this problem by simply using an untrained model for feature extraction from bi-temporal hyperdimensional images. As the feature extractor model is untrained, it can be initialized with as many number of input channels as desired with appropriate weight initialization technique. Moreover, the number of filters in the subsequent layers can also be chosen in a flexible manner, as there is no training involved. Extensive experiments on four hyperdimensional datasets show the superiority of the proposed approach. The proposed approach is also capable of clustering the changed pixels into semantically meaningful groups, as shown for Hermiston dataset. While the idea seems bold and new in context of remote sensing, similar idea has been verified before in the computer vision and machine learning literature, e.g., deep image prior. The proposed approach benefits from the fact that hyperdimensional images generally

TABLE VII
CD RESULTS FOR THE BAY AREA SCENE. PROPOSED METHOD'S RESULT IS REPORTED AS AVERAGE OF 5 RUNS.

Method	Sensitivity	Specificity	Accuracy
CVA	74.44	97.54	85.64
PCVA	48.19	79.46	63.36
SAMZID$_{\text{Sin}}$	79.42	89.18	84.15
SAMZID$_{\text{Tan}}$	70.83	97.91	83.96
AICA	69.18	97.26	82.80
DCVA3Channels-1	78.27	92.47	85.16
DCVA3Channels-2	72.31	91.52	81.63
DCVA3Channels-3	48.78	58.87	53.67
DCVAAllChannels-1	40.88	64.41	52.29
DCVAAllChannels-2	47.66	78.10	62.42
Proposed (5 layers)	78.51	97.86	87.89$\pm$1

TABLE VIII
CD RESULTS FOR THE HERMISTON SCENE. PROPOSED METHOD'S RESULT IS REPORTED AS AVERAGE OF 5 RUNS.

Method	Sensitivity	Specificity	Accuracy
CVA	92.22	97.45	96.78
PCVA	60.14	94.19	89.83
SAMZID$_{\text{Sin}}$	83.83	82.96	83.07
SAMZID$_{\text{Tan}}$	81.08	83.96	83.59
AICA	64.80	99.01	94.63
DCVA3Channels-1	99.40	96.57	96.93
DCVA3Channels-2	99.44	94.58	95.20
DCVA3Channels-3	42.72	78.88	74.25
DCVAAllChannels-1	61.25	76.78	74.79
DCVAAllChannels-2	62.91	87.95	84.75
Proposed (5 layers)	95.97	98.29	97.99$\pm$0.0

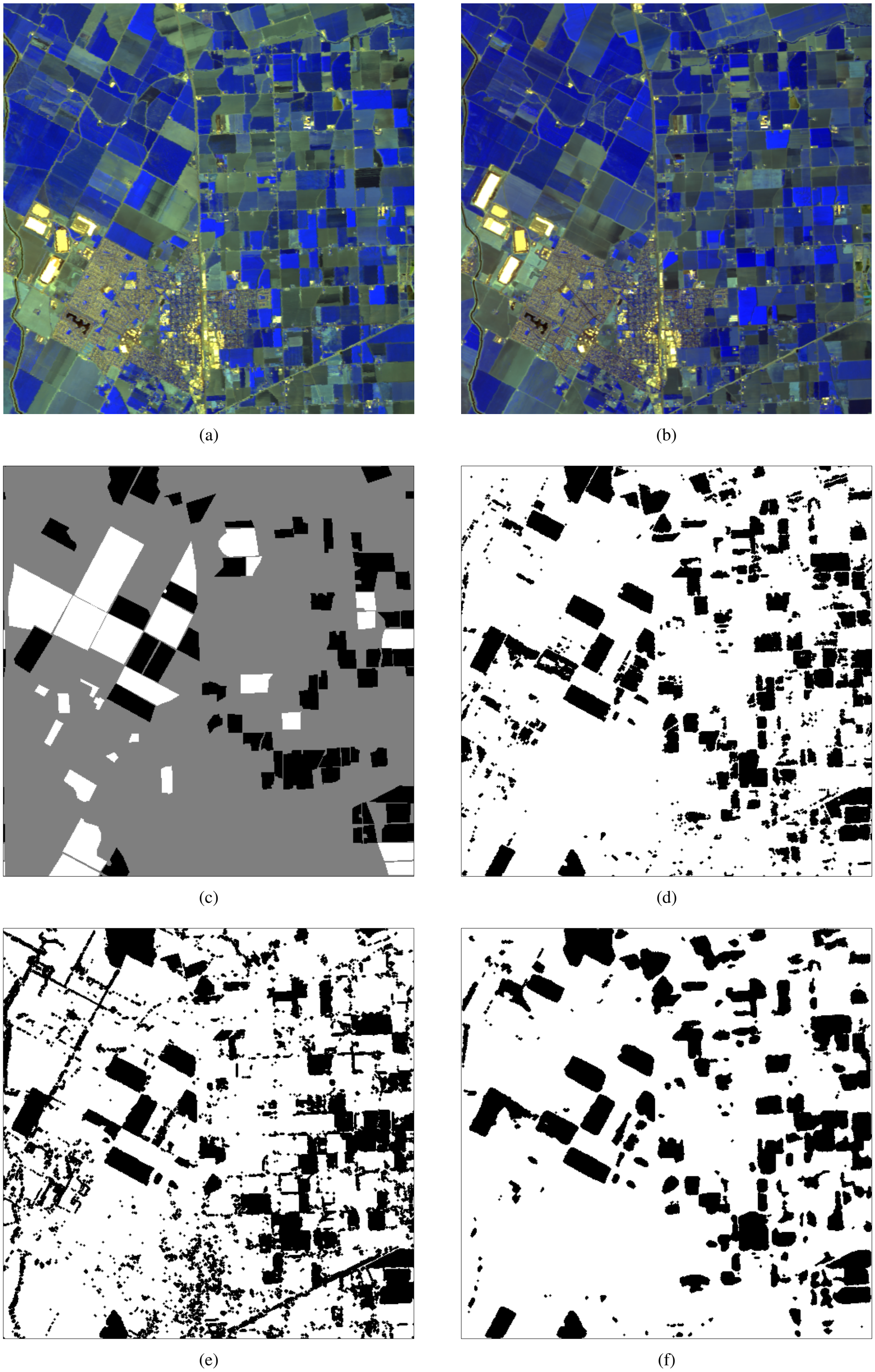


Fig. 5. Bay Area scene: (a) pre-change and (b) post-change, (c) reference image (white - unchanged, black - changed, gray - unknown), and CD maps: (d) CVA, (e) DCVA3Channels-1, (f) Proposed

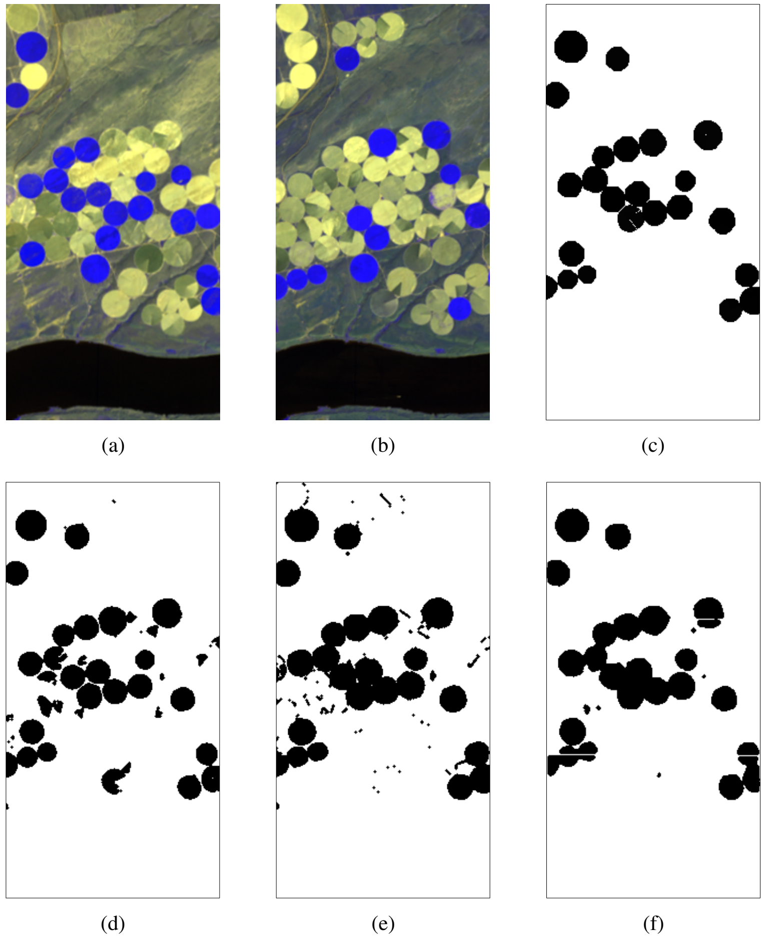


Fig. 6. Hermiston scene, False color composition (R: band 52, G: band 31, B: band 22) images: (a) pre-change and (b) post-change, Reference images: (c) binary (white - unchanged, black - changed, gray - unknown), Binary CD maps: (d) CVA, (e) DCVA3Channels-1, (f) Proposed.

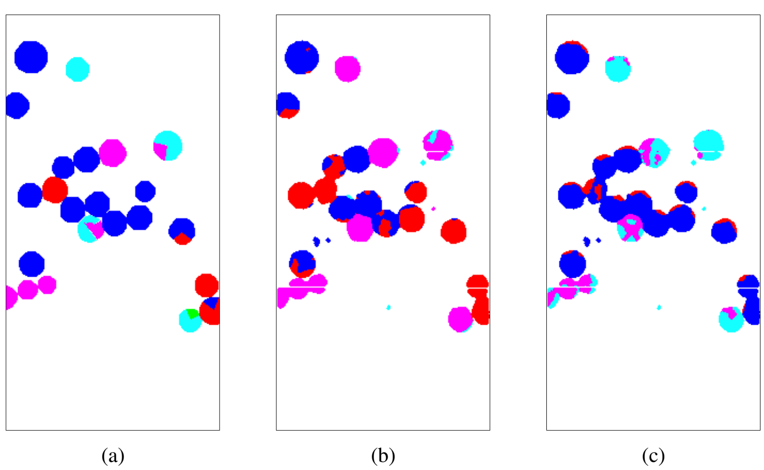
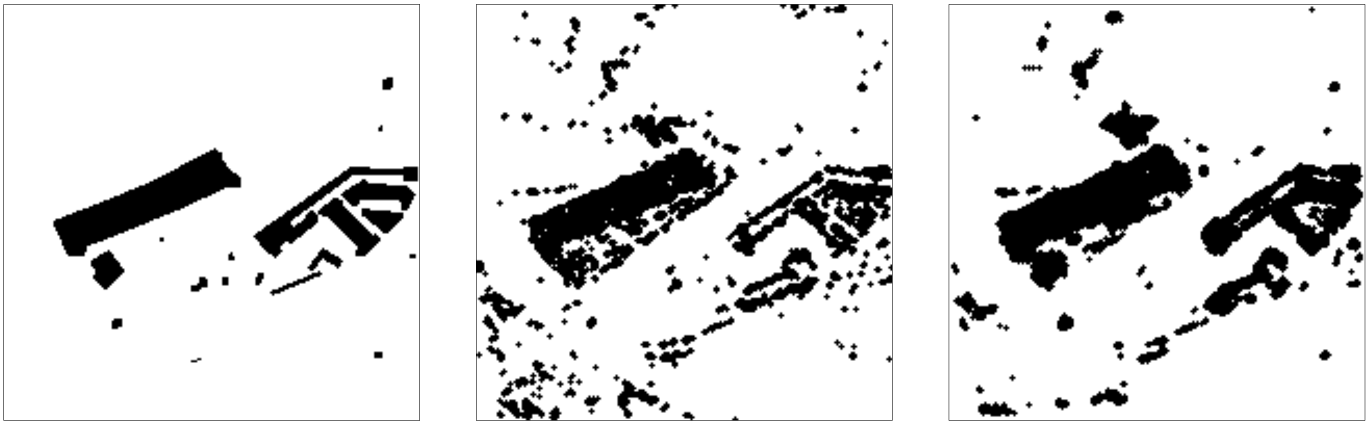


Fig. 7. Multiple CD for Hermiston scene: (a) Reference image, CD maps: (b) Using original hyperspectral pixel values and (c) Proposed.



(a) (b) (c)

Fig. 8. Decomposed POLSAR dataset (details in [24]). CD maps: (a) reference, (b) CVA, (c) Proposed.

TABLE IX
CD RESULTS FOR SAN FRANCISCO POLSAR SCENE. PROPOSED METHOD'S RESULT IS REPORTED AS AVERAGE OF 5 RUNS.

Method	Sensitivity	Specificity	Accuracy
CVA	89.78	89.35	89.39
PCVA	45.78	87.18	83.62
DCVAAllChannels-1	67.06	77.39	76.50
DCVAAllChannels-2	46.74	83.06	79.94
Proposed (5 layers)	94.51	89.63	90.05±0.9

exhibit less spatial complexity due to the cost of generating higher resolution in both spectral and spatial domain. Thus the applicability of the method to very high spatial resolution hyperdimensional sensors may not be straightforward and will be investigated in future work. Our future work will also investigate untrained models in the context of the hyperspectral image classification. As a final note, the proposed approach should not be seen as a competitor to the supervised methods, rather as a complementary to them.

REFERENCES

[1] X. X. Zhu, D. Tuia, L. Mou, G.-S. Xia, L. Zhang, F. Xu, and F. Fraundorfer, "Deep learning in remote sensing: A comprehensive review and list of resources," *IEEE Geoscience and Remote Sensing Magazine*, vol. 5, no. 4, pp. 8–36, 2017.

[2] S. Saha, F. Bovolo, and L. Bruzzone, "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.

[3] S. Saha, F. Bovolo, and L. Bruzzone, "Building change detection in vhr sar images via unsupervised deep transcoding," *IEEE Transactions on Geoscience and Remote Sensing*, 2020.

[4] L. Bergamasco, S. Saha, F. Bovolo, and L. Bruzzone, "Unsupervised change-detection based on convolutional-autoencoder feature extraction," in *Image and Signal Processing for Remote Sensing XXV*, vol. 11155. International Society for Optics and Photonics, 2019, p. 1115510.

[5] S. Saha, L. Mou, X. X. Zhu, F. Bovolo, and L. Bruzzone, "Semisupervised change detection using graph convolutional network," *IEEE Geoscience and Remote Sensing Letters*, 2020.

[6] 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.

[7] Z. Zhang, G. Vosselman, M. Gerke, D. Tuia, and M. Y. Yang, "Change detection between multimodal remote sensing data using siamese cnn," *arXiv preprint arXiv:1807.09562*, 2018.

[8] H. Chen, C. Wu, B. Du, L. Zhang, and L. Wang, "Change detection in multisource vhr images via deep siamese convolutional multiple-layers recurrent neural network," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 4, pp. 2848–2864, 2019.

[9] F. Rahman, B. Vasu, J. Van Cor, J. Kerekes, and A. Savakis, "Siamese network with multi-level features for patch-based change detection in satellite imagery," in *2018 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*. IEEE, 2018, pp. 958–962.

[10] L. Bruzzone and D. F. Prieto, "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.

[11] F. Bovolo, "A multilevel parcel-based approach to change detection in very high resolution multitemporal images," *IEEE GRSL*, vol. 6, no. 1, pp. 33–37, 2009.

[12] 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.

[13] P. Ghamisi, N. Yokoya, J. Li, W. Liao, S. Liu, J. Plaza, B. Rasti, and A. Plaza, "Advances in hyperspectral image and signal processing: A comprehensive overview of the state of the art," *IEEE Geoscience and Remote Sensing Magazine*, vol. 5, no. 4, pp. 37–78, 2017.

[14] Z. Huang, L. Fang, and S. Li, "Subpixel-pixel-superpixel guided fusion for hyperspectral anomaly detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 9, pp. 5998–6007, 2020.

[15] L. Fang, W. Zhao, N. He, and J. Zhu, "Multiscale cnns ensemble based self-learning for hyperspectral image classification," *IEEE Geoscience and Remote Sensing Letters*, vol. 17, no. 9, pp. 1593–1597, 2020.

[16] S. Liu, Q. Du, X. Tong, A. Samat, H. Pan, and X. Ma, "Band selection-based dimensionality reduction for change detection in multi-temporal hyperspectral images," *Remote Sensing*, vol. 9, no. 10, p. 1008, 2017.

[17] S. Mohla, S. Pande, B. Banerjee, and S. Chaudhuri, "Fusatnet: Dual attention based spectrospatial multimodal fusion network for hyperspectral and lidar classification," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2020, pp. 92–93.

[18] C. Ehrler, C. Fischer, and M. Bachmann, "Hyperspectral remote sensing applications in mining impact analysis." 2011.

[19] S. T. Seydi and M. Hasanlou, "A new land-cover match-based change detection for hyperspectral imagery," *European Journal of Remote Sensing*, vol. 50, no. 1, pp. 517–533, 2017.

[20] X. Tong, H. Pan, S. Liu, B. Li, X. Luo, H. Xie, and X. Xu, "A novel approach for hyperspectral change detection based on uncertain area analysis and improved transfer learning," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 13, pp. 2056–2069, 2020.

[21] L. Mou, P. Ghamisi, and X. X. Zhu, "Deep recurrent neural networks for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 7, pp. 3639–3655, 2017.

[22] S. R. Cloude and E. Pottier, "A review of target decomposition theorems in radar polarimetry," *IEEE transactions on geoscience and remote sensing*, vol. 34, no. 2, pp. 498–518, 1996.

[23] Q. Song, F. Xu, and Y.-Q. Jin, "Radar image colorization: Converting single-polarization to fully polarimetric using deep neural networks," *IEEE Access*, vol. 6, pp. 1647–1661, 2017.

[24] A. Najafi, M. Hasanlou, and V. Akbari, "Land cover changes detection in polarimetric sar data using algebra, similarity and distance based methods." *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*, vol. 42, 2017.

[25] H. Bi, J. Sun, and Z. Xu, "Unsupervised polsar image classification using discriminative clustering," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 6, pp. 3531–3544, 2017.

[26] D. Ulyanov, A. Vedaldi, and V. Lempitsky, "Deep image prior," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 9446–9454.

[27] N. Audebert, B. Le Saux, and S. Lefèvre, "Deep learning for classification of hyperspectral data: A comparative review," *IEEE geoscience and remote sensing magazine*, vol. 7, no. 2, pp. 159–173, 2019.

[28] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 1026–1034.

[29] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *Proceedings of the thirteenth international conference on artificial intelligence and statistics*, 2010, pp. 249–256.

[30] W. A. Malila, "Change vector analysis: an approach for detecting forest changes with Landsat," in *LARS symposia*, 1980, p. 385.

[31] T. Celik, "Unsupervised change detection in satellite images using principal component analysis and k-means clustering," *IEEE Geoscience and Remote Sensing Letters*, vol. 6, no. 4, pp. 772–776, 2009.

[32] N. Falco, G. Cavallaro, P. R. Marpu, and J. A. Benediktsson, "Unsupervised change detection analysis to multi-channel scenario based on morphological contextual analysis," in *2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*. IEEE, 2016, pp. 3374–3377.

[33] 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.

[34] A. Appice, N. Di Mauro, F. Lomuscio, and D. Malerba, "Empowering change vector analysis with autoencoding in bi-temporal hyperspectral images," in *CEUR Workshop Proceedings*, vol. 2466, 2019.

[35] A. Song, J. Choi, Y. Han, and Y. Kim, "Change detection in hyperspectral images using recurrent 3d fully convolutional networks," *Remote Sensing*, vol. 10, no. 11, p. 1827, 2018.

[36] Z. Chen and F. Zhou, "Multitemporal hyperspectral image change detection by joint affinity and convolutional neural networks," in *2019 10th International Workshop on the Analysis of Multitemporal Remote Sensing Images (MultiTemp)*. IEEE, 2019, pp. 1–4.

[37] M. Molinier and J. Kilpi, "Avoiding overfitting when applying spectral-spatial deep learning methods on hyperspectral images with limited labels," in *IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium*. IEEE, 2019, pp. 5049–5052.

[38] C. Zhang, S. Bengio, M. Hardt, B. Recht, and O. Vinyals, "Understanding deep learning requires rethinking generalization," *arXiv preprint arXiv:1611.03530*, 2016.

[39] F. Williams, T. Schneider, C. Silva, D. Zorin, J. Bruna, and D. Panozzo, "Deep geometric prior for surface reconstruction," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 10 130–10 139.

[40] D. Bau, H. Strobelt, W. Peebles, B. Zhou, J.-Y. Zhu, A. Torralba *et al.*, "Semantic photo manipulation with a generative image prior," *arXiv preprint arXiv:2005.07727*, 2020.

[41] P. I. Wójcik, "Random projection in deep neural networks," *arXiv preprint arXiv:1812.09489*, 2018.

[42] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[43] M. Volpi and D. Tuia, "Dense semantic labeling of subdecimeter resolution images with convolutional neural networks," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 2, pp. 881–893, 2016.

[44] M. Hasanlou and S. T. Seydi, "Sensitivity analysis on performance of different unsupervised threshold selection methods in hyperspectral change detection," in *2018 10th IAPR Workshop on Pattern Recognition in Remote Sensing (PRRS)*. IEEE, 2018, pp. 1–4.

[45] 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.

[46] J. A. Lee and M. Verleysen, *Nonlinear dimensionality reduction*. Springer Science & Business Media, 2007.

[47] D. Marinelli, F. Bovolo, and L. Bruzzone, "A novel method for unsupervised multiple change detection in hyperspectral images based on binary spectral change vectors," in *Analysis of Multitemporal Remote Sensing Images (MultiTemp), 2017 9th International Workshop on the*. IEEE, 2017, pp. 1–4.

[48] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright *et al.*, "Scipy 1.0: fundamental algorithms for scientific computing in python," *Nature methods*, vol. 17, no. 3, pp. 261–272, 2020.

[49] J. López-Fandiño, A. S. Garea, D. B. Heras, and F. Argüello, "Stacked autoencoders for multiclass change detection in hyperspectral images," in *IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium*. IEEE, 2018, pp. 1906–1909.

[50] J. López-Fandiño, D. B. Heras, F. Argüello, and M. Dalla Mura, "Gpu framework for change detection in multitemporal hyperspectral images," *International Journal of Parallel Programming*, vol. 47, no. 2, pp. 272–292, 2019.

[51] F. Thonfeld, H. Feilhauer, M. Braun, and G. Menz, "Robust change vector analysis (RCVA) for multi-sensor very high resolution optical satellite data," *International Journal of Applied Earth Observation and Geoinformation*, vol. 50, pp. 131–140, 2016.

[52] T. Ridler, S. Calvard *et al.*, "Picture thresholding using an iterative selection method," *IEEE trans syst Man Cybern*, vol. 8, no. 8, pp. 630–632, 1978.

[53] M. Sezgin and B. Sankur, "Survey over image thresholding techniques and quantitative performance evaluation," *Journal of Electronic imaging*, vol. 13, no. 1, pp. 146–165, 2004.

[54] C. H. Li and C. Lee, "Minimum cross entropy thresholding," *Pattern recognition*, vol. 26, no. 4, pp. 617–625, 1993.

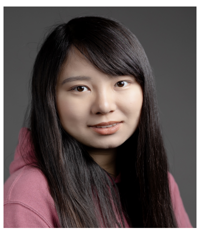
[55] M. Zanetti, F. Bovolo, and L. Bruzzone, "Rayleigh-rice mixture parameter estimation via EM algorithm for change detection in multispectral images," *IEEE Transactions on Image Processing*, vol. 24, no. 12, pp. 5004–5016, 2015.

Sudipan Saha (S'16–M'20) received the PhD degree in information and communication technologies from the University of Trento, Trento, Italy, and Fondazione Bruno Kessler, Trento, Italy in 2020. Previously, he obtained the M.Tech. degree in electrical engineering from IIT Bombay, Mumbai, India, in 2014.

He is currently a postdoctoral researcher at Technical University of Munich (TUM), Munich, Germany. Previously, he worked as an Engineer with TSMC Limited, Hsinchu, Taiwan, from 2015 to 2016. In 2019, he was a Guest Researcher with the Technical University of Munich (TUM), Munich, Germany. He is the recipient of Fondazione Bruno Kessler Best Student Award 2020. His research interests are related to multitemporal remote sensing image analysis, domain adaptation, time-series analysis, image segmentation, deep learning, image processing, and pattern recognition. Dr. Saha is a Reviewer for several international journals and served as a guest editor at Remote Sensing (MDPI) special issue on "Advanced Artificial Intelligence for Remote Sensing: Methodology and Application".

Lukas Kondmann received the bachelor degree in Economics from the Ludwig-Maximilians-University Munich in 2016 and the master degree in Social Data Science from the University of Oxford in 2019. He holds an honors degree in Technology Management (2015-2017) from the Center for Digital Technology Management in Munich and was a visiting researcher working on big data for social good at the UC Berkeley School of Information in spring 2017.

He is currently pursuing the Ph.D. degree in engineering at the Technical University of Munich and the German Aerospace Center in Munich. His research is centered around time-series analysis of multispectral remote sensing imagery with a focus on monitoring the Sustainable Development Goals (SDGs).



Qian Song Qian Song (S'16-M'20) received the B.E. degree (Hons.) from the School of Information Science and Technology, East China Normal University, Shanghai, China, in 2015, and the Ph.D. degree (Hons.) from Fudan University, Shanghai, China, in 2020. She is currently a post-doctoral fellow with the Remote Sensing Technology Institute (IMF), German Aerospace Center (DLR), Wessling, Germany.

She has been awarded as the URSI (International Union of Radio Science) Young Scientist Award in 2020. Her research interests include advanced deep learning technologies and their applications in synthetic aperture radar image interpretation.

Xiao Xiang Zhu (S'10–M'12–SM'14–F'21) received the Master (M.Sc.) degree, her doctor of engineering (Dr.-Ing.) degree and her "Habilitation" in the field of signal processing from Technical University of Munich (TUM), Munich, Germany, in 2008, 2011 and 2013, respectively.

She is currently the Professor for Data Science in Earth Observation (former: Signal Processing in Earth Observation) at Technical University of Munich (TUM) and the Head of the Department "EO Data Science" at the Remote Sensing Technology Institute, German Aerospace Center (DLR). Since 2019, Zhu is a co-coordinator of the Munich Data Science Research School (www.mu-ds.de). Since 2019 She also heads the Helmholtz Artificial Intelligence – Research Field "Aeronautics, Space and Transport". Since May 2020, she is the director of the international future AI lab "AI4EO – Artificial Intelligence for Earth Observation: Reasoning, Uncertainties, Ethics and Beyond", Munich, Germany. Since October 2020, she also serves as a co-director of the Munich Data Science Institute (MDSI), TUM. Prof. Zhu was a guest scientist or visiting professor at the Italian National Research Council (CNR-IREA), Naples, Italy, Fudan University, Shanghai, China, the University of Tokyo, Tokyo, Japan and University of California, Los Angeles, United States in 2009, 2014, 2015 and 2016, respectively. She is currently a visiting AI professor at ESA's Phi-lab. Her main research interests are remote sensing and Earth observation, signal processing, machine learning and data science, with a special application focus on global urban mapping.

Dr. Zhu is a member of young academy (Junge Akademie/Junges Kolleg) at the Berlin-Brandenburg Academy of Sciences and Humanities and the German National Academy of Sciences Leopoldina and the Bavarian Academy of Sciences and Humanities. She serves in the scientific advisory board in several research organizations, among others the German Research Center for Geosciences (GFZ) and Potsdam Institute for Climate Impact Research (PIK). She is an associate Editor of IEEE Transactions on Geoscience and Remote Sensing and serves as the area editor responsible for special issues of IEEE Signal Processing Magazine. She is a Fellow of IEEE.