DEEP NO LEARNING APPROACH FOR UNSUPERVISED CHANGE DETECTION IN HYPERSPECTRAL IMAGES

Sudipan Saha[1,*], Lukas Kondmann[1,2], Xiao Xiang Zhu[1,2]

[1] Data Science in Earth Observation, Technical University of Munich (TUM), Munich, Germany - sudipan.saha@tum.de
[2] Remote Sensing Technology Institute, German Aerospace Center, Weßling, Germany - (lukas.kondmann, xiaoxiang.zhu)@dlr.de

KEY WORDS: Change Detection, Deep Learning, Deep Image Prior, Hyperspectral Images.

ABSTRACT:

Unsupervised deep transfer-learning based change detection (CD) methods require pre-trained feature extractor that can be used to extract semantic features from the target bi-temporal scene. However, it is difficult to obtain such feature extractors for hyperspectral images. Moreover, it is not trivial to reuse the models trained with the multispectral images for the hyperspectral images due to the significant difference in number of spectral bands. While hyperspectral 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 networks can yield remarkable result in different tasks like super-resolution and surface reconstruction. Motivated by this, we make a bold proposition that untrained deep model, initialized with some weight initialization strategy can be used to extract useful semantic features from bi-temporal hyperspectral images. Thus, we couple an untrained network with Deep Change Vector Analysis (DCVA), a popular method for unsupervised CD, to propose an unsupervised CD method for hyperspectral images. We conduct experiments on two hyperspectral CD data sets, and the results demonstrate advantages of the proposed unsupervised method over other competitors.

1. INTRODUCTION

Change detection (CD) is an important application of remote sensing. It plays a crucial role in several applications including land-cover mapping, environmental monitoring, disaster management, precision agriculture, burned area monitoring, and mining activity monitoring. In the literature, most CD methods are proposed for multispectral (Saha et al., 2019) and Synthetic Aperture Radar (SAR) (Saha et al., 2020a) images. In comparison, there are only few works dedicated to the hyperspectral images that generally show lower spatial resolution, however very high spectral resolution. Hyperspectral images can provide rich information in some CD applications, e.g., monitoring of mining activity (Ehrler et al., 2011). Inspite of this, less interest in research related to hyperspectral CD is explained by the lack of labeled hyperspectral images. This scarcity is not limited to only multi-temporal hyperspectral image analysis, but extends to the hyperspectral image classification. 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 (Mou et al., 2021).

CD methods can be formulated in supervised (Zhang et al., 2018), semi-supervised (Saha et al., 2020c), and unsupervised way (Saha et al., 2019). However, unsupervised methods are preferred in the literature (Bruzzone and Prieto, 2000) due to the difficulty of collecting unlabeled bi-temporal data. Most popular paradigm for unsupervised change detection, known as change vector analysis (CVA) (Bruzzone and Prieto, 2000), applies intuitive difference operation on pre-change and post-change images. Additionally, there are unsupervised methods that rely on clustering (Celik, 2009). With the emergence of deep learning, CVA has been reformulated as deep CVA (DCVA) (Saha et al., 2019) by exploiting deep transfer learning. DCVA projects the bi-temporal images in deep feature space by using a pre-trained deep feature extractor and subsequently compares the images in the projected domain. DCVA itself does not use any training or fine-tuning of the deep model and is agnostic to how the deep model has been derived. However, DCVA depends on the availability of pre-trained feature extractor that is generally not available for hyperspectral images. Moreover, it is not trivial to reuse the networks trained for other sensors on hyperspectral images due to the huge gap of number of spectral bands and characteristics. It is even challenging to reuse a network trained on one hyperspectral sensor on images acquired from another hyperspectral sensor. There are attempts in the literature to reduce dependence on pre-trained network by exploiting self-supervised training on the target scene (Saha et al., 2020b).

The success of deep learning in the unsupervised multi-temporal analysis can be attributed to its capability to capture spatial context. While, some target scenes (e.g., urban) show high spatial complexity, this is often not the case in some other target areas (e.g., agricultural land). Besides, most hyperspectral images show coarse spatial resolution, thus obliterating the possibility to capture high spatial complexity. In contrast to multi-spectral images, hyperspectral images show complexity in the spectral domain. Motivated by this, there are works in the hyperspectral image classification that use 1D convolution (Audebert et al., 2019). While still spatial complexity 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 hyperspectral images can be captured by an untrained deep model merely initialized with a deep model initialization strategy (He et al., 2015) (Glorot and Bengio, 2010). 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 (Ulyanov

* Corresponding author

et al., 2018). One advantage of using untrained network is that they can be initialized to ingest as many number of image channels as desired, which may be helpful in hyperspectral image analysis.

Thus motivated by the possibility that an untrained deep model can capture spatio-temporal information of multitemporal hyperspectral images, we propose a CD method for hyperspectral images that uses DCVA framework along with an untrained deep model. The rest of this work is organized as follows. We briefly discuss some relevant works in Section 2. Section 3 discusses the proposed method. Section 4 presents the datasets and results. Finally we conclude this paper in Section 5.

2. RELATED WORK

Following the relevance to our work, we briefly discuss in this section about: i) hyperspectral change detection methods and ii) deep image prior.

2.1 CD in hyperspectral images

Compared to the multispectral and SAR images, deep learning based methods on hyperspectral CD are very few. Moreover, most of them are supervised (Wang et al., 2018). In (Wang et al., 2018), authors identified three unique challenges for hyperspectral CD - high dimension, limited datasets, and mixed pixel problem. To solve these problems, they proposed a pre-classification based end-to-end CD framework named GETNET based on 2-D Convolutional Neural Network (CNN). Song *et. al.* (Song et al., 2018) proposed a supervised CD method for hyperspectral images, called the recurrent three-dimensional (3D) fully convolutional network (Re3FCN), which merged a 3D fully convolutional network (FCN) and a convolutional long short-term memory (ConvLSTM). Chen and Zhou (Chen and Zhou, 2019) proposed a supervised CD method consisting of three steps: spectral dimensionality reduction, joint affinity tensor construction and binary (changed or unchanged) classification by CNN. While these works made a remarkable contribution of introducing deep learning to hyperspectral change detection, most of them agree that limited datasets is a big challenge in hyperspectral CD. Conforming to that, their works use pixels from same image for training and evaluation. However, using such large supervised networks when training and test pixels belong to same scene is not completely realizable. Molinier and Kilpi (Molinier and Kilpi, 2019) have shown that while training and test pixels come from the same scene, pixel values corresponding to testing sets are partly seen during the training phase, thus leading to overoptimistic accuracy assessment. More practical alternative is to use unsupervised approach, like the ones (Saha et al., 2019) (Saha et al., 2020a) that have been used for multispectral and SAR images.

2.2 Deep image prior

CNNs are usually trained on large labeled datasets of images. 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, this explanation has been found to be inaccurate in many instances. In (Zhang et al., 2016), authors 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. Ulyanov *et. al.* (Ulyanov et al., 2018) tried to understand 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. They showed that this simple setup provides very competitive result for several image restoration problems, e.g., inpainting, super-resolution, and denoising. This 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 photo manipulation (Bau et al., 2020) and surface reconstruction (Williams et al., 2019). Another similar line of research is random projection network (Wójcik, 2018) 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 (Wójcik, 2018) tackles this challenge by prepending the network with an input layer whose weights are initialized with a random projection matrix.

3. PROPOSED METHOD

We are interested to detect changes from a pair of co-registered hyperspectral images X_1 and X_2 consisting of B_0 channels each and acquired by the same hyperspectral sensor. We first initialize a deep model with number of input channels and number of filters of intermediate layers adjusted as per the number of channels of hyperspectral images. Subsequently this network is used to extract a set of features from both pre-change and post-change images and the difference is taken to obtain deep change hypervector. Following this, deep change hypervector is further analyzed using magnitude-based analysis to distinguish the unchanged pixels (ω_{nc}) from the changed ones (Ω_c). The proposed unsupervised hyperspectral binary CD framework is called Untrained Hyperspectral DCVA (UH-DCVA) and is shown in Figure 1.

3.1 Feature extraction

We use an untrained model for deep feature extraction. Generally in computer vision, first layer of deep models ingest input of 3 channels and project it to larger number of channels, commonly 64 (Simonyan and Zisserman, 2014). However, for our case number of input channel B_0 is generally larger than 200. We design first convolution layer such that it ingests the hyperspectral image of B_0 channels and projects it to $\beta_0 * B_0$ filters where $\beta_0 > 1$. In our experiments, we set $\beta_0 = 4$. The following convolution layer ingests input 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. No activation function is used between two convolution layers as there is no training process involved. Even though no non-linearity is achieved by simply stacking convolution layers one after another, it helps to increase the spatial receptive field of the convolution network. Moreover, it helps to increase the number of different filters in exponential order. Though theoretically speaking, since no non-linearity is used, multiple-layers network can be possibly reproduced with just 1-layer, in practice single-layer network cannot emulate systematic increment of spatial receptive field as input passes through successive layers of a multiple-layer network.

Considering the coarse spatial resolution of the hyperspectral images, we do not need to make the network as deeper as it

is standard in computer vision. In our experiments, we show that even a network with couple of untrained convolutional layers provides us satisfactory results. Though untrained, instead of using random weights, we use weights initialized with He initialization method (He et al., 2015). 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 DCVA framework. The 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, which we detail in next subsection.

3.2 Unsupervised change detection

The untrained network can be used as bi-temporal hyperspectral deep feature extractor in a DCVA framework (Saha et al., 2019) to distinguish changed pixels (Ω_c) from the unchanged ones (ω_{nc}). This is based on the assumption that such an untrained (however initialized (He et al., 2015)) network can be assumed to be a collection of morphological filters and they extract same semantic attributes for both pre-change and post-change image. Hyperspectral images X_1 and X_2 are pre-processed/normalized to have all spectral bands in range 0-1 and then separately fed through the untrained network. Deep features can be extracted from a set of L convolution layers of the network to form a deep feature hypervector (G) that is obtained as a concatenation of the deep-feature-differences of the considered layers. However, for simplicity, we assumed that features are extracted from just one layer in this work. Euclidean norm of deep change hypervector G is used to obtain the deep magnitude ρ. Being processed through same set of filters, unchanged pixels (ω_{nc}) tend to generate similar deep features and thus smaller ρ in comparison to the changed pixels (Ω_c). This is used to segregate Ω_c and ω_{nc} by using a thresholding on ρ. While any suitable thresholding method can be used, in this work we use Otsu's thresholding (Otsu, 1979).

4. RESULTS

We validate the proposed method on two bi-temporal scenes (López-Fandiño et al., 2018) (López-Fandiño et al., 2019) [1]:

1. The Santa Barbara bi-temporal scene is acquired on 2013 (Figure 2(a)) and 2014 (Figure 2(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 2(c)).

2. The Hermiston scene 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.

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

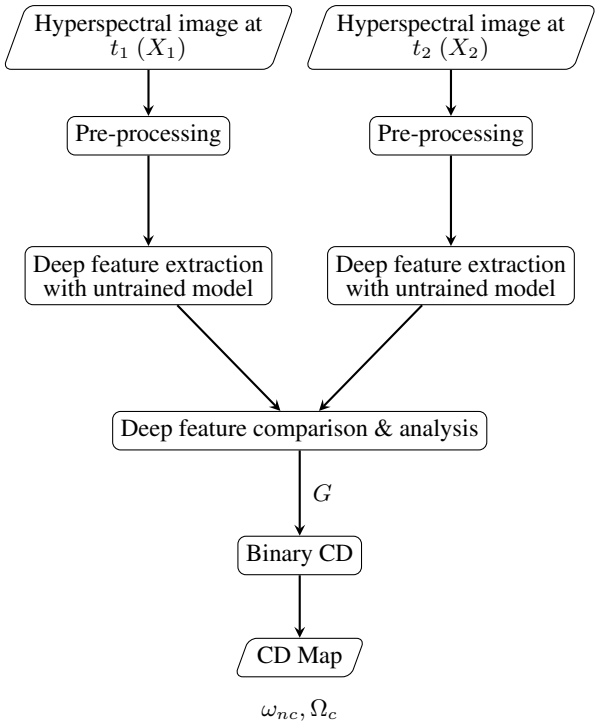

Figure 1. Proposed Untrained Hyperspectral Deep CVA (UH-DCVA) technique

We compared the proposed method to three unsupervised methods:

- Change vector analysis (CVA), the comparison to which is necessary to understand if the proposed method provides any additional benefit to the existing mechanism of pixel difference.

- Parcel change vector analysis (PCVA) that captures the spatial information as superpixel, the comparison to which is critical to understand if the benefits brought by proposed method can be merely replaced by a superpixel based analysis.

- Deep change vector analysis (DCVAPretrained) with feature extractor pre-trained on largescale computer vision dataset with VGG16/VGG19 architecture (Simonyan and Zisserman, 2014), the comparison to which is critical to understand if benefits brought by proposed method can be merely substituted by transfer learning approaches. We modulate the first layer of the network by replicating the weights as number of channels of hyperspectral images. While modulating the first layer of network is not an ideal choice, this shows the limitation of the existing methods for adapting to hyperspectral input. In absence of any suitable adaptation technique, brute force modulation of first layer is used to adapt DCVAPretrained for hyperspectral input.

  For DCVAPretrained, we have tested three different configurations: by using 1st convolutional layer of VGG16 (DCVAPretrained-1), 2nd convolutional layer of VGG16 (DCVAPretrained-2), and 1st convolutional layer of VGG19 (DCVAPretrained-3). Different combinations with both VGG16 and VGG19 are compared to ensure that the proposed method can outperform different architectures.

Method	Sensitivity	Specificity
CVA	76.92	96.69
PCVA	58.18	84.74
DCVAPretrained-1	51.24	85.88
DCVAPretrained-2	46.53	78.57
DCVAPretrained-3	50.63	86.03
Proposed (1 layer)	84.87	98.80
Proposed (2 layers)	86.32	98.81
Proposed (3 layers)	88.94	98.08

Table 1. CD results for the Santa Barbara scene

Method	Sensitivity	Specificity
CVA	92.22	97.45
PCVA	60.14	94.19
DCVAPretrained-1	61.25	76.78
DCVAPretrained-2	74.41	80.53
DCVAPretrained-3	52.31	76.92
Proposed (1 layer)	94.93	99.33
Proposed (2 layers)	96.92	98.99
Proposed (3 layers)	97.67	98.61

Table 2. CD results for the Hermiston scene

Comparison is performed in terms of sensitivity (accuracy computed over changed pixels) and specificity (accuracy computed over unchanged pixels).

The proposed method using one-layer network obtains sensitivity of 84.87% and specificity of 98.80% for the Santa Barbara scene. Using two layer network, the proposed method obtains sensitivity of 86.32% and specificity of 98.81%. While using three layer network, the proposed obtains sensitivity of 88.94% and specificity of 98.08% (shown in Figure 2(f)). Thus we observe a gradual but slow improvement in performance as the number of layers are increased (as tabulated in Table 2). The improvement is because adding more convolution layers improve the spatial receptive field of the filters and increase the complexity of the filters. However, considering the coarse resolution of the hyperspectral images, this performance saturates soon, as we observe by adding third layer specificity drops a little. We show the result obtained by CVA in Figure 2(d). The proposed method clearly outperforms CVA and PCVA. Interestingly, PCVA obtains inferior result compared to CVA (Table 2). This might be due to less effectiveness of superpixel-based representation in coarse resolution hyperspectral images. All combinations of DCVAPretrained (Figure 2(e)) obtains inferior result compared to the proposed method. Here we recall that generally first layer of deep models ingest input of 3 channels and project it to larger number of channels, commonly 64 (Simonyan and Zisserman, 2014). However, in this case DCVAPretrained ingests large number of channels (224) and projects it to 64 channels. This is reverse of what is being generally done in computer vision. This is where proposed method is particularly useful as untrained models provide us the freedom of choosing number of input and output features as we desire.

Similar result is obtained in case of Hermiston scene. The proposed method obtains superior sensitivity and specificity score in comparison to the state-of-the-art unsupervised methods. The performance of the proposed method improves as more layers are added, however specificity drops a little. Quantitative result for Hermiston scene is shown in Table 2 and visual result is omitted for sake of brevity.

Though we showed the quantitative result using a fixed threshold determination scheme, aligned with the usual practice in unsupervised change detection literature (Saha et al., 2019), in most cases the proposed method outperforms compared methods both in terms of sensitivity and specificity. This shows that slight variation in threshold would not impact the superiority of the proposed method.

5. CONCLUSION

In this work, we presented an unsupervised change detection for hyperspectral images. The proposed method uses an untrained model for feature extraction from bi-temporal hyperspectral images. As the feature extractor model is untrained, it can be initialized with as many number of input channels as desired. This is particularly convenient considering the large number of spectral bands in different hyperspectral images. Moreover, the number of filters in the subsequent layers can also be decided arbitrarily, as there is no training involved. Our experiments show that the proposed method can produce better result than the unsupervised CD methods. While the idea seems bold at first sight, similar idea has been verified before in the computer vision and machine learning literature. Our work is based on the assumption that hyperspectral images show significantly less spatial complexity. Thus the method is not applicable to very high spatial resolution hyperspectral sensor, though they are rare in practice, due to the cost of generating higher resolution in both spatial and spectral domain. The remote sensing community has observed significant rise in papers related to deep learning in the last few years. However, there have been almost no effort to identify what matters most, deep or learning or both. Our work is a first step towards that direction. In future, we will investigate whether introducing non-linearity between layers can further improve the result of the proposed method. We will also extend the proposed hyperspectral CD method by analyzing changed pixels (Ω_c) using direction-based analysis to obtain different kinds of change $\omega_{c1}, \omega_{c2}, ..., \omega_{cK}$.

ACKNOWLEDGEMENTS

The work is funded 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. Additionally, Lukas Kondmann is supported by the Helmholtz Association under the joint research school “Munich School for Data Science - MUDS”.

REFERENCES

Audebert, N., Le Saux, B., Lefèvre, S., 2019. Deep learning for classification of hyperspectral data: A comparative review. *IEEE geoscience and remote sensing magazine*, 7(2), 159–173.

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

Bruzzone, L., Prieto, D. F., 2000. Automatic analysis of the difference image for unsupervised change detection. *IEEE Transactions on Geoscience and Remote sensing*, 38(3), 1171–1182.

Celik, T., 2009. Unsupervised change detection in satellite images using principal component analysis and k-means clustering. *IEEE Geoscience and Remote Sensing Letters*, 6(4), 772–776.

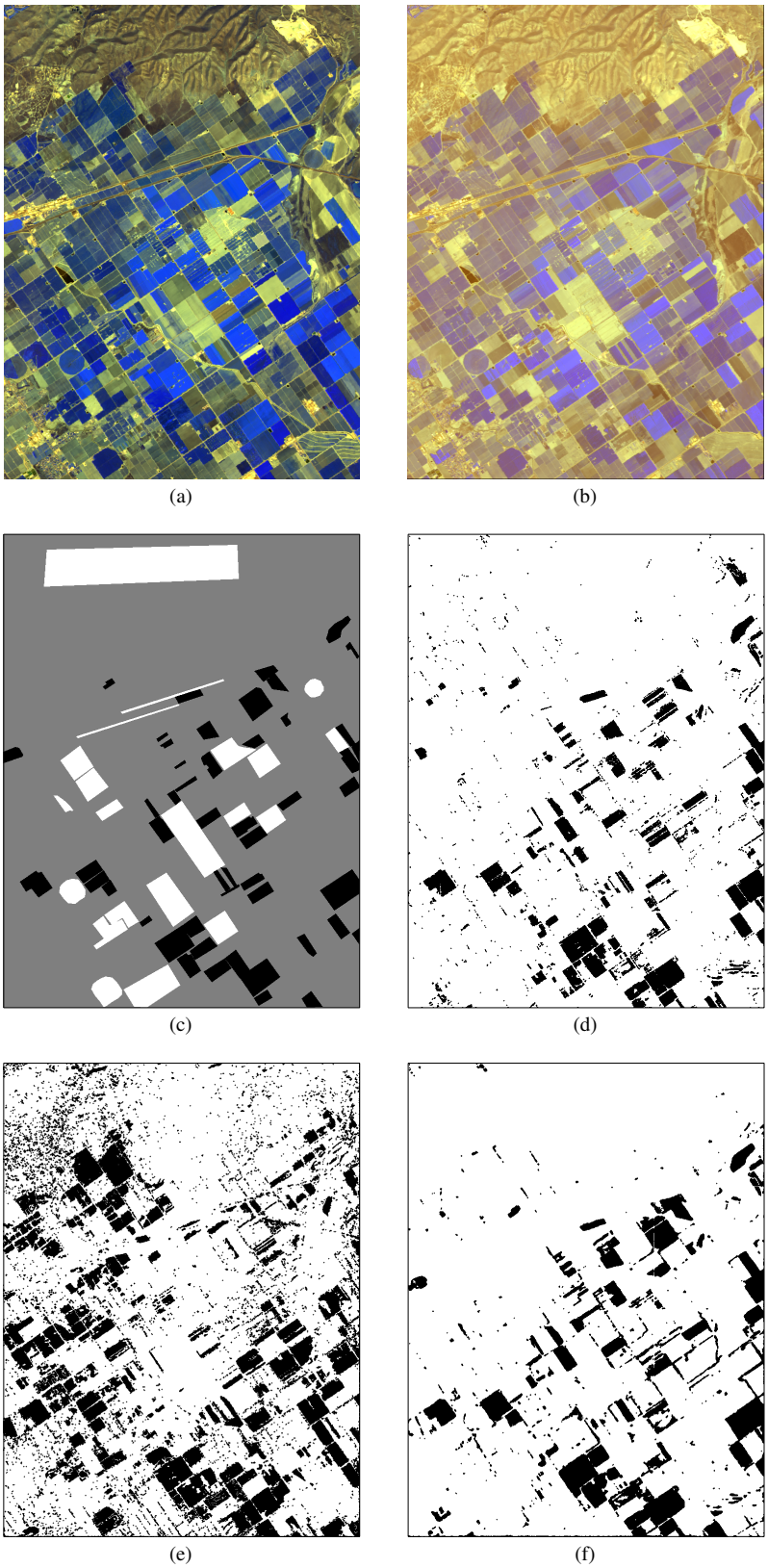

Figure 2. Santa Barbara scene, False color composition (R: band 50, G: band 20, B: band 10) images: (a) pre-change and (b) post-change, (c) reference image (white - unchanged, black - changed, gray - unknown), and CD maps: (d) CVA, (e) DCVAPretrained-1, (f) Proposed (3 layers)

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

Ehrler, C., Fischer, C., Bachmann, M., 2011. Hyperspectral Remote Sensing Applications in Mining Impact Analysis.

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

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

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

López-Fandiño, J., Heras, D. B., Argüello, F., Dalla Mura, M., 2019. GPU framework for change detection in multitemporal hyperspectral images. *International Journal of Parallel Programming*, 47(2), 272–292.

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

Mou, L., Saha, S., Hua, Y., Bovolo, F., Bruzzone, L., Zhu, X. X., 2021. Deep Reinforcement Learning for Band Selection in Hyperspectral Image Classification. *arXiv preprint arXiv:2103.08741*.

Otsu, N., 1979. A threshold selection method from gray-level histograms. *IEEE transactions on systems, man, and cybernetics*, 9(1), 62–66.

Saha, S., Bovolo, F., Bruzzone, L., 2019. Unsupervised deep change vector analysis for multiple-change detection in VHR images. *IEEE Transactions on Geoscience and Remote Sensing*, 57(6), 3677–3693.

Saha, S., Bovolo, F., Bruzzone, L., 2020a. Building Change Detection in VHR SAR Images via Unsupervised Deep Transcoding. *IEEE Transactions on Geoscience and Remote Sensing*.

Saha, S., Mou, L., Qiu, C., Zhu, X. X., Bovolo, F., Bruzzone, L., 2020b. Unsupervised Deep Joint Segmentation of Multitemporal High-Resolution Images. *IEEE Transactions on Geoscience and Remote Sensing*.

Saha, S., Mou, L., Zhu, X. X., Bovolo, F., Bruzzone, L., 2020c. Semisupervised Change Detection Using Graph Convolutional Network. *IEEE Geoscience and Remote Sensing Letters*.

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

Song, A., Choi, J., Han, Y., Kim, Y., 2018. Change detection in hyperspectral images using recurrent 3D fully convolutional networks. *Remote Sensing*, 10(11), 1827.

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

Wang, Q., Yuan, Z., Du, Q., Li, X., 2018. GETNET: A general end-to-end 2-D CNN framework for hyperspectral image change detection. *IEEE Transactions on Geoscience and Remote Sensing*, 57(1), 3–13.

Williams, F., Schneider, T., Silva, C., Zorin, D., Bruna, J., Panozzo, D., 2019. Deep geometric prior for surface reconstruction. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 10130–10139.

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

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

Zhang, Z., Vosselman, G., Gerke, M., Tuia, D., Yang, M. Y., 2018. Change detection between multimodal remote sensing data using siamese CNN. *arXiv preprint arXiv:1807.09562*.