Manifold Learning and Deep Generative Networks for Heterogeneous Change Detection From Hyperspectral and Synthetic Aperture Radar Images

Ignacio Masari, *Graduate Student Member, IEEE*, Gabriele Moser, *Fellow, IEEE*, and Sebastiano B. Serpico, *Life Fellow, IEEE*

***Abstract*— Unsupervised change detection (CD) stands as a critical tool for damage assessment after a natural disaster. We emphasize heterogeneous CD methods, which support the case of highly heterogeneous images at the two observation dates, providing greater flexibility than traditional homogeneous methods. This adaptability is vital for swift responses in the aftermath of natural disasters. In this framework, we address the challenging case of detecting changes between the hyperspectral and synthetic aperture radar images. This case has intrinsic difficulties, namely, the difference in the nature of the physical quantity measured, added to the great difference in dimensionality of the two imaging domains. To address these challenges, a novel method is proposed based on the integration of a manifold learning technique and deep learning networks trained to perform an image-to-image translation task. The method works in a fully unsupervised manner, further enforcing a fast implementation in real-world scenarios. From an application-oriented perspective, we focus on flooded-area mapping using the PRISMA and COSMO-SkyMed missions. The experimental validation on two datasets, a semisimulated one and a real one associated with flooding, suggests that the proposed method allows for accurate detection of flooded areas and other ground changes.**

***Index Terms*— Deep learning, dimensionality reduction, heterogeneous change detection (CD), hyperspectral images (HSIs), image-to-image translation, manifold learning, synthetic aperture radar (SAR).**

I. Introduction

THE topic of unsupervised change detection (CD) has long been addressed in environmental remote sensing and consists of identifying ground changes through the analysis of a pair of images acquired on the same location at different times without needing training sets for "change" and "no-change" or for the land-cover classes. Its applications regard urbanization, deforestation, desertification, glacier reduction, coastal alteration monitoring, and many more [1], [2], [3]. Major applications especially regard the monitoring of natural disasters, such as forest fires, oil spills, and floods [4].

Received 27 August 2024; revised 29 October 2024; accepted 6 November 2024. Date of publication 12 November 2024; date of current version 21 November 2024. This work was supported in part by the Italian Space Agency (ASI) within the Framework of the Project under Grant PRISMA-Learn-ASI no. 2022-12-U.0. *(Corresponding author: Gabriele Moser.)*

The authors are with the Department of Naval, Electric, Electronic and Telecommunications Engineering (DITEN), University of Genoa, 16145 Genoa, Italy (e-mail: gabriele.moser@unige.it).

Digital Object Identifier 10.1109/LGRS.2024.3496567

Because of the recent missions for Earth observation with hyperspectral payloads (e.g., PRISMA, EnMAP, and HySIS), frequent coverage from satellite platforms is now possible. Therefore, when a disaster occurs, pre-event hyperspectral images (HSIs) are often available on the affected geographical area. Conversely, post-event HSI acquisition is unlikely due to clouds or smoke (e.g., in the case of floods or forest fires). On the contrary, synthetic aperture radar (SAR) image acquisition is day-and-night and unaffected by clouds or smoke. Accordingly, when responding to a natural disaster, using post-event SAR with pre-event HSI is now a promising opportunity.

Historically, most unsupervised CD algorithms were conceived to take as input a homogeneous couple of images, i.e., collected by the same sensor and under similar environmental conditions [2]. In contrast, detecting changes from a pre-event HSI and a post-event SAR image is a heterogeneous CD task. This poses a more intricate problem: distinguishing changes on the ground from differences caused by the imaging modalities. Despite requiring a more complex formulation, these more adaptable techniques are particularly desirable in disaster management applications, where fast responses are critical.

Different techniques have proven effective to handle data from different domains. They can be divided into three major categories as follows: 1) image fusion [5]; 2) image-to-image (I2I) translation [6], [7]; and 3) local feature analysis [8]. We refer interested readers to [9] for a thorough state-of-the-art analysis, and we only emphasize the primary role of deep I2I translation in this context. In many I2I approaches, convolutional neural networks (CNNs) are trained to transfer the data from one domain to the other and vice versa [6] or to transfer both domains into a common one [7]. The deep I2I approach to unsupervised heterogeneous CD has proven useful for diverse types of images, including optical images acquired in different spectral ranges as well as pairs of multispectral and SAR images [10]. Different deep learning architectures have been proposed in this framework, using generative adversarial networks (GANs) [11], U-Net-based networks [12], and autoencoders (AEs) [10], among others.

This letter addresses unsupervised heterogeneous CD between an HSI and a multichannel SAR image using deep

I2I translation. To the best of our knowledge, this case of heterogeneous CD has not been addressed in the literature. In this situation, there is the additional problem of a great difference in the dimensionality of the two image domains. Consequently, even if the translation from HSI to SAR data might be feasible, the opposite is a very ill-posed problem.

To tackle this challenge, the approach proposed in this letter extends the previous method X-Net [13] by integrating it with unsupervised dimensionality reduction. X-Net combines two CNNs trained together to learn translations from the domain of each image into the other domain, with a case-specific loss function that integrates affinity matrix information to favor unsupervised CD. It was originally tested on optical, multispectral, and SAR data. It is integrated into the proposed approach because of its consistent performance throughout different domains and its high accuracy in comparison with other state-of-the-art techniques. In the proposed method, we combine X-Net with the uniform manifold approximation and projection (UMAP) [14] and with a case-specific majority voting approach.

The main novel contributions of this letter are twofold as follows: 1) exploring, for the first time, the potential of heterogeneous CD with HSI and SAR data and 2) proposing an unsupervised technique to address this problem by integrating UMAP-based dimensionality reduction, deep I2I translation, affinity matrix information, and a majority voting approach. The method is experimentally validated using PRISMA and COSMO-SkyMed second generation (CSG) imagery.

II. Methods

A. Overview of the Proposed Method

Let $\mathcal{H}$ and $\mathcal{S}$ be an HSI and a multichannel SAR image acquired over the same area at times t_0 and t_1. We assume that they are well registered. The goal of the proposed method is to detect the changes occurred in the considered area from t_0 to t_1. These changes are implicitly meant in land cover, and the spatial resolution of $\mathcal{H}$ and $\mathcal{S}$ is correspondingly assumed to be medium-to-high (e.g., 30 m). I2I translation from $\mathcal{H}$ to $\mathcal{S}$ is generally expected to work unimpaired, while translation from $\mathcal{S}$ to $\mathcal{H}$ would not be feasible, due to the aforementioned large difference in dimensionality. For this reason, a dimensionality reduction stage plays a crucial role. In the proposed approach, I2I is addressed after a transformation aimed at feature reduction in the HSI domain. In this approach, we use UMAP, a nonlinear feature reduction method from high-dimensional data. UMAP is chosen over alternatives, such as t-SNE or kernel PCA, for its ability to handle large datasets and maintain both local and global structures [15].

Let h be the number of hyperspectral channels in $\mathcal{H}$, and let $\mathcal{U}_n$ be the transformed image obtained by applying UMAP to $\mathcal{H}$ to reduce from h to n features ($n \ll h$). X-Net is applied to the heterogeneous pair $(\mathcal{U}_n, \mathcal{S})$. This allows obtaining both a pair of homogeneous SAR images ($\hat{\mathcal{S}}_n$ simulated from $\mathcal{U}_n$ at t_0 and the true $\mathcal{S}$ at t_1) and a pair of images in the domain of $\mathcal{U}_n$ (i.e., the true $\mathcal{U}_n$ image derived from $\mathcal{H}$ at t_0 and an image $\hat{\mathcal{U}}_n$ simulated from $\mathcal{S}$ at t_1). A homogeneous CD algorithm, based on image differencing, conditional random fields (CRFs) and unsupervised thresholding is then applied to the homogeneous pairs $(\mathcal{S}, \hat{\mathcal{S}}_n)$ and $(\mathcal{U}_n, \hat{\mathcal{U}}_n)$, leading to a change map $\mathcal{M}_n$. Details on the individual processing stages are described in Sections II-B and II-C.

$\mathcal{M}_n$ obviously depends on the number n of features in the transformed image $\mathcal{U}_n$, i.e., n plays the role of a hyperparameter of the method. The effect of this hyperparameter is nontrivial. While more channels generally imply preserving more spectral information, the higher the dimensionality of $\mathcal{U}_n$, the more difficult it is for the network to perform the transformation from $\mathcal{S}$ to its domain. To favor independence from n, a voting approach is integrated within the proposed method. A finite set $\mathcal{N}$ of values of n is predefined, and the aforementioned procedure, including UMAP and X-Net, is repeated for each $n \in \mathcal{N}$ separately. Given the resulting ensemble $\{\mathcal{M}_n : n \in \mathcal{N}\}$ of change maps, each pixel is labeled as "change" if it is classified as such in the majority of the individual change maps $\mathcal{M}_n$. In this respect, the number of elements of $\mathcal{N}$ is chosen to be odd.

B. Unsupervised Dimensionality Reduction

UMAP addresses unsupervised dimensionality reduction through a manifold learning approach. It starts by constructing a weighted k-nearest neighbor graph to capture local relationships, assigning weights to edges based on distances. By considering different neighborhood scales, UMAP aims at capturing both local and global structures. A low-dimensional representation is obtained by minimizing the cross entropy between probabilities derived from the input high-dimensional space and the output low-dimensional embedding. This process involves constructing a similar graph in the low-dimensional space and matching transition probabilities between neighboring points in the two spaces. Fuzzy simplicial sets are also integrated in the process to capture complex and nonlinear data structures. More details on this approach can be found in [14].

C. I2I Transformation

X-Net consists of two CNNs that are trained jointly with a loss function that incorporates an affinity-based prior in order to favor, within the translation process, the discrimination between changed and unchanged areas. Let $n \in \mathcal{N}$, and let x and y be tensors collecting the data from two colocated image patches of predefined size in $\mathcal{U}_n$ and $\mathcal{S}$, respectively. The first CNN performs the translation $\hat{y} = F_n(x)$ from the domain of $\mathcal{U}_n$ to the domain of $\mathcal{S}$, and the second CNN performs the translation $\hat{x} = G_n(y)$ from $\mathcal{S}$ to $\mathcal{U}_n$. The loss function is defined as the weighted sum of three contributing terms: the cycle-consistency loss, the weighted translation loss, and a weight decay regularization. While we provide here a brief explanation of these terms, we refer the reader to [13] for more details. The cycle-consistency loss is defined as follows:

$$\mathcal{L}_n^{\text{cyc}}(v) = \mathbb{E}_x[\delta(x, G_n(F_n(x)))] + \mathbb{E}_y[\delta(y, F_n(G_n(y)))] \quad (1)$$

where $\delta(a, b)$ indicates the squared Frobenius distance between tensors a and b, and v indicates the vector collecting all network parameters. $\mathcal{L}_n^{\text{cyc}}$ favors the desired cycle consistency in the translation, i.e., $F_n(G_n(y)) \simeq y$ and $G_n(F_n(x)) \simeq x$.

The weighted translation loss is defined as follows:

$$\mathcal{L}_n^{\mathrm{wt}}(v) = \mathbb{E}_{x,y}[\delta(x, G_n(y)|\Psi) + \delta(y, F_n(x)|\Psi)] \qquad (2)$$

where $\delta(a, b|\Psi)$ indicates the squared Frobenius distance weighted on a matrix of pixelwise weights Ψ. In this case, $F_n(x) \simeq y$ and $G_n(y) \simeq x$ are desired only for pixels that have not undergone changes, so that the translation output preserves and possibly emphasizes the discrimination between changed and unchanged areas. For this purpose, affinity matrices are used to define the weight matrix Ψ, so that its entry associated with pixel p in the patch evaluates the chances of not finding a change in p. The rationale is that the affinity matrix of a subset of pixels provides characterization of their spatial interrelations, which, when a change occurs, are expected to vary quite regardless of acquisition modality [16]. Here, we just recall that, in X-Net, affinity matrices are separately computed in the two domains in a neighborhood of each pixel, and Ψ is determined as a function of the elemen-twise distance between them. We refer to [13] for details on their computation. We also recall that the loss in (2) can be analytically proven to approximate a conditional loss $\mathbb{E}_{x,y}[\delta(x, G_n(y)) + \delta(y, F_n(x))|H_0]$, given the "no-change" hypothesis H_0 [13].

The last term of the loss function is a weight decay regularization term, aimed at preventing overfitting. This way, the total loss function can be expressed as follows:

$$\mathcal{L}_n(v) = w^{\mathrm{cyc}}\mathcal{L}_n^{\mathrm{cyc}}(v) + w^{\mathrm{wt}}\mathcal{L}_n^{\mathrm{wt}}(v) + w^{\mathrm{wd}}\|v\|_2^2. \qquad (3)$$

The training of the networks is carried through the minimization of this loss with respect to v. The positive weights w^{cyc}, w^{wt}, and w^{wd} are hyperparameters used to balance the impact of each term. The expectations in (1) and (2) are estimated through the sample means.

After training, the networks are used to generate $\hat{\mathcal{S}}_n = F_n(\mathcal{U}_n)$ and $\hat{\mathcal{U}}_n = G_n(\mathcal{S}_n)$. A homogeneous CD algorithm is then applied to the pairs $(\mathcal{S}, \hat{\mathcal{S}}_n)$ and $(\mathcal{U}_n, \hat{\mathcal{U}}_n)$. First, the distances $d_p^{\mathcal{U}}$ and $d_p^{\mathcal{S}}$ are computed on each pixel p

$$d_p^{\mathcal{U}} = \|\hat{x}_p - x_p\|_2, \qquad d_p^{\mathcal{S}} = \|\hat{y}_p - y_p\|_2 \qquad (4)$$

where x_p, y_p, $\hat{x}_p$, and $\hat{y}_p$ represent the pixel samples of $\mathcal{U}_n$, $\mathcal{S}$, $\hat{\mathcal{U}}_n$, and $\hat{\mathcal{S}}_n$, respectively. The resulting difference images are normalized and averaged, obtaining a final difference image. Finally, this last image is segmented with the fully connected CRF in [17] and thresholded with the unsupervised Otsu's method [18] to obtain the binary change map $\mathcal{M}_n$ $(n \in \mathcal{N})$.

Following the common nomenclature in the CD literature, we refer to the proposed CD method as "fully unsupervised," since it does not require a set of training pixels with change or land cover labels. However, the I2I translation component involves a supervised training where the training set is obviously made of colocated image patches in $\mathcal{U}_n$ and $\mathcal{S}$, but does not include label annotations. Similarly, we assume $\mathcal{H}$ and $\mathcal{S}$ to be coregistered. In general, the coregistration process may involve prior knowledge on the scene (e.g., in terms of interactively drawn ground control points) or may be addressed through automatic methods that do not use any training or reference set [19].

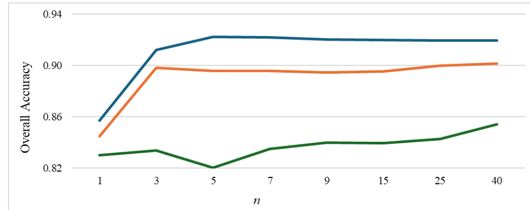

Fig. 1. OA in the Emilia-Romagna dataset after applying the voting approach with all bands until n. The proposed method, CAAE + UMAP, and SCCN + UMAP are in blue, orange, and green, respectively.

III. Experimental Results

A. Datasets and Experimental Setup

The proposed technique was tested on two datasets: a semisimulated dataset from the surroundings of Lake Bolsena, Italy, and a real case study of a flooding event in the Italian region of Emilia-Romagna.

The semisimulated dataset consisted of a pair of images acquired within a short temporal window, in an area where no natural disaster nor major appreciable change had been recorded. In this way, no major change is expected to affect the image pair. The site selected was a surrounding area of the Bolsena Lake (Italy), and the images used were a CSG image acquired on October 4, 2023 and a PRISMA image acquired on October 10, 2023. Changes were manually introduced by modifying the CSG image to simulate flooding. A precise ground truth was computed by comparing the original and the changed CSG images.

The real case study dataset consisted of a pre-event PRISMA image and a post-event CSG image. The PRISMA image was acquired on July 15, 2022, while the peak in the flood occurred on May 20, 2023, and the CSG image was acquired on May 21, 2023.

The validation of the results on the real dataset consisted in comparing the change map obtained by the proposed method with a reference change map. A flooding map on the area was provided by the global flood monitoring (GFM) system of the Copernicus Emergency Management Service.[1] However, this is a flooded-area map; hence, it is expected to locate the ground changes associated only with the flooding. On the contrary, the proposed method aims at identifying all changes occurred in the scene between July 15, 2022 and May 20, 2023. For this reason, we manually updated this ground truth by adding the changes found by visually comparing two Sentinel-2 images from July 17, 2022 and May 23, 2023.

PRISMA images include 234 spectral bands with 30-m resolution, while CSG images have two channels (HH and HV) with 5-m spatial resolution. CSG images were down-sampled to match the resolution from PRISMA. Before the I2I stage, the automatic method in [20] was applied for coregistration. This is an information-theoretic area-based registration method that does not require any ground control point as input. In the case of both datasets, the set of the numbers of channels involved in the ensemble was $\mathcal{N} = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 40\}$.

[1] https://portal.gfm.eodc.eu

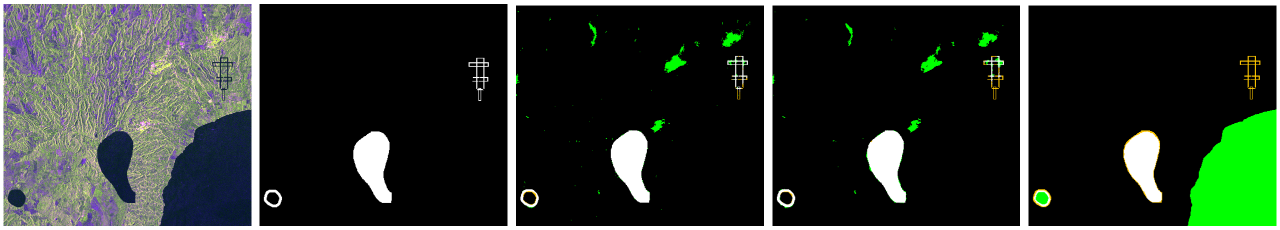

Fig. 2. Semisimulated dataset. From left to right: CSG modified image (RGB composite: |HH|, |HV|, and |HH|/|HV|); test map; overlapping of test and change maps for the proposed method, CAAE, and SCCN, respectively. Color code in overlapped images: white = true positive, black = true negative, green = false positive, and orange = false negative.

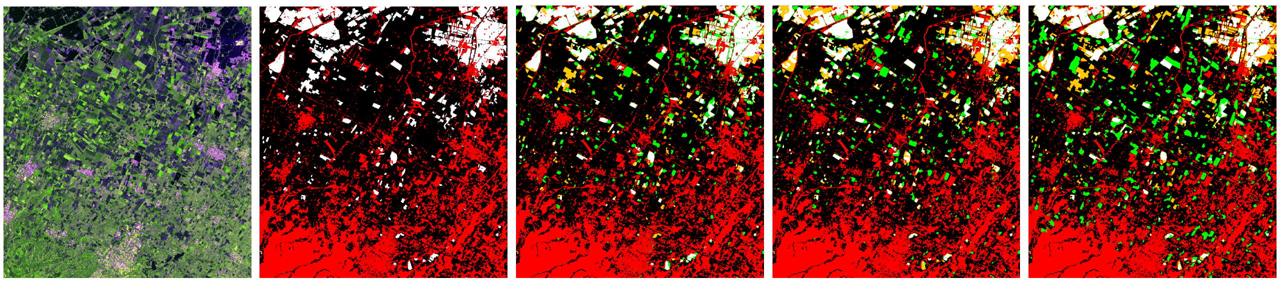

Fig. 3. Real dataset. From left to right: CSG modified image (RGB composite: |HH|, |HV|, and |HH|/|HV|); modified test map; overlapping of test and change maps for the proposed method, CAAE, and SCCN, respectively. Color code in overlapped images: white = true positive, black = true negative, green = false positive, orange = false negative, and red = no-data.

TABLE I

Accuracy Measures on the Semisimulated Dataset. FAR: False Alarm Rate, DR: Detection Rate, and OA: Overall Accuracy

Method:	Proposed	CAAE + UMAP	SCCN + UMAP	XNet + kPCA	CAAE + kPCA	SCCN + kPCA
FAR	**0.01**	**0.01**	0.15	**0.01**	**0.01**	0.15
DR	**0.98**	0.96	0.83	**0.98**	0.95	0.91
OA	**0.99**	**0.99**	0.85	**0.99**	**0.99**	0.85

TABLE II

Accuracy Measures on the Emilia-Romagna Dataset. FAR: False Alarm Rate, DR: Detection Rate, and OA: Overall Accuracy

Method:	Proposed	CAAE + UMAP	SCCN + UMAP	XNet + kPCA	CAAE + kPCA	SCCN + kPCA
FAR	0.06	0.06	0.11	**0.05**	**0.05**	0.12
DR	**0.76**	0.64	0.61	0.57	0.55	0.66
OA	**0.92**	0.90	0.85	0.90	0.90	0.85

B. Results and Experimental Comparisons

To the best of our knowledge, this work is the first to investigate heterogeneous CD with HSI and SAR data, so no directly comparable technique was available. However, in order to perform comparisons with the literature, we note that the method involves both feature reduction and I2I translation. Accordingly, comparisons were performed with previous algorithms aimed at either aspect. Regarding dimensionality reduction, we used the well-known kernel-based extension of the principal component analysis (kPCA) with the Gaussian radial basis function as kernel [21]. As for I2I translation, we compared with the symmetrical convolutional coupling network (SCCN) [7] and code aligned autoencoder (CAAE) [10] techniques.

Fig. 1 shows the behavior of the overall accuracy (OA) of the proposed and benchmark methods on the test set of the Emilia-Romagna dataset. It can be noted that the proposed method is quite robust to the choice of n. Moreover, the proposed method obtains the values of OA larger than those achieved by SCCN and CAAE for all n.

Table I shows the detection rate (DR, i.e., the recall of the "change" hypothesis), the false alarm rate (FAR, i.e., the complement of the recall of "no-change"), and the OA obtained on the semisimulated dataset. All methods obtained quite high values, with the proposed technique standing out as the top performer, with OA $\simeq$ 99% and DR $\simeq$ 98%. The combination of X-Net with kPCA yielded a very similar change map, with the same value of DR. CAAE also obtained accurate results, with OA $\simeq$ 99% and DR $\simeq$ 95% with kPCA and DR $\simeq$ 96% with UMAP. In the case of SCCN, less accurate change maps are obtained, yet with OA $\simeq$ 85%.

The experimental results for the real dataset are presented in Table II. The proposed method obtained accurate results, suggesting the effectiveness of the developed approach at addressing the CD task from input HSI and SAR data, while outperforming the compared approaches. In particular, there is a significant performance gap between this method and both CAAE and SCCN. This can be seen in OA and DR. The comparison also pointed out that, if UMAP is replaced by kPCA in the proposed technique, then a significant loss (19%) is observed in DR, thus confirming the choice of UMAP for dimensionality reduction within the developed approach.

A visual analysis of the change maps confirms the quantitative results. For the semisimulated dataset, in Fig. 2, we can notice how the proposed method and CAAE + UMAP obtained similar results for the two biggest change regions. Both methods present some false alarms in similar regions. The proposed technique visually outperformed CAAE in the detection of narrow linear changes. SCCN labeled as "change" most of the pixels associated with water surfaces in the scene, thus partially detecting the two largest change regions but also generating numerous false and missed alarms. In the case of the real dataset, in Fig. 3, we can notice that the proposed method obtained the best results. In comparison, CAAE + UMAP exhibited more missed alarms, resulting in a worst DR, while SCCN + UMAP exhibited more missed and false alarms, resulting in a lower DR and a higher FAR.

We note that, in real-world scenarios of unsupervised CD over large areas affected by disasters, such as the case study of the Emilia-Romagna region, the sheer image size can generally present a significant computational challenge, as compared with the operational requirements of rapid mapping applications. In the proposed method, the use of patch-based processing and parallelization allows this challenge to be tackled and the large data volume to be efficiently handled.

IV. Conclusion

In this letter, we investigated the potential of unsupervised heterogeneous CD with hyperspectral and multichannel SAR data, by integrating a previous I2I approach with unsupervised manifold learning and with a majority voting algorithm to tackle the large difference in the dimensionalities of the two input image domains. The change maps generated by the proposed approach, when applied both to a case study of flooded-area mapping and to a semisimulated dataset, exhibited high detection accuracy and low false alarm rates on both datasets, as well as a visually accurate detection of the flooded areas and the other changes in the scene. These results confirm the opportunities offered by the joint use of remote sensing data coming from the current hyperspectral and SAR spaceborne sensors, especially in applications to damage assessment after a disaster, and showed remarkable effectiveness of the proposed approach in addressing this task. The experiments also pointed out the accuracy improvement of the proposed approach as compared with previous deep networks for heterogeneous CD, adapted to the application with HSI. Similarly, experimental comparisons regarding the dimensionality reduction component of the proposed method confirmed the effectiveness of the choice of UMAP, as compared with kPCA, to extract a set of transformed features to be used for heterogeneous CD purposes.

Consistently with the applicative focus on supporting damage assessment after a natural disaster, the proposed approach is primarily designed to detect abrupt changes. From this perspective, a possible limitation of the method is that it is generally expected to be less effective in scenarios involving gradual changes, such as phenological shifts in vegetation. Indeed, such gradual changes may be challenging to detect for homogeneous CD algorithms as well. In addition, the method addresses the case of input medium-to-high resolution imagery and does not address the case of extremely high-resolution imagery, such as that obtained from unmanned aerial vehicles. To extend to such cases, future generalizations of the method could involve integrating it with object detection techniques to explicitly model the objects in the scene and address the detection of the related changes.

Other future extensions of the developed method may include a further integration of the dimensionality reduction and I2I translation components, by addressing also the former through a neural approach in order to learn both the embedding and translation in an end-to-end fashion.

Acknowledgment

This work was carried out using CSG and PRISMA Products, both ©ASI, and delivered under a license to use by ASI.

References

[1] X. Li and A. G.-O. Yeh, "Analyzing spatial restructuring of land use patterns in a fast growing region using remote sensing and GIS," *Landscape Urban Planning*, vol. 69, no. 4, pp. 335–354, Oct. 2004.

[2] P. de Bem, O. D. C. Junior, R. F. Guimarães, and R. T. Gomes, "Change detection of deforestation in the Brazilian Amazon using Landsat data and convolutional neural networks," *Remote Sens.*, vol. 12, no. 6, p. 901, Mar. 2020.

[3] E. Berthier, Y. Arnaud, R. Kumar, S. Ahmad, P. Wagnon, and P. Chevallier, "Remote sensing estimates of glacier mass balances in the Himachal Pradesh (Western Himalaya, India)," *Remote Sens. Environ.*, vol. 108, no. 3, pp. 327–338, Jun. 2007.

[4] Y. Afaq and A. Manocha, "Analysis on change detection techniques for remote sensing applications: A review," *Ecol. Informat.*, vol. 63, Jul. 2021, Art. no. 101310.

[5] X. Jiang, G. Li, Y. Liu, X.-P. Zhang, and Y. He, "Change detection in heterogeneous optical and SAR remote sensing images via deep homogeneous feature fusion," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 1551–1566, 2020.

[6] X. Niu, M. Gong, T. Zhan, and Y. Yang, "A conditional adversarial network for change detection in heterogeneous images," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 1, pp. 45–49, Jan. 2019.

[7] J. Liu, M. Gong, K. Qin, and P. Zhang, "A deep convolutional coupling network for change detection based on heterogeneous optical and radar images," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 3, pp. 545–559, Mar. 2018.

[8] Y. Sun, L. Lei, X. Li, H. Sun, and G. Kuang, "Nonlocal patch similarity based heterogeneous remote sensing change detection," *Pattern Recognit.*, vol. 109, Jan. 2021, Art. no. 107598.

[9] Z. Lv et al., "Land cover change detection with heterogeneous remote sensing images: Review, progress, and perspective," *Proc. IEEE*, vol. 110, no. 12, pp. 1976–1991, Dec. 2022.

[10] L. T. Luppino et al., "Code-aligned autoencoders for unsupervised change detection in multimodal remote sens. images," 2020, *arXiv:2004.07011*.

[11] X. Li, Z. Du, Y. Huang, and Z. Tan, "A deep translation (GAN) based change detection network for optical and SAR remote sensing images," *ISPRS J. Photogramm. Remote Sens.*, vol. 179, pp. 14–34, Sep. 2021.

[12] Z. Lv, H. Huang, L. Gao, J. A. Benediktsson, M. Zhao, and C. Shi, "Simple multiscale UNet for change detection with heterogeneous remote sensing images," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022.

[13] L. T. Luppino et al., "Deep image translation with an affinity-based change prior for unsupervised multimodal change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 4700422.

[14] L. McInnes, J. Healy, N. Saul, and L. Großberger, "UMAP: Uniform manifold approximation and projection," *J. Open Source Softw.*, vol. 3, no. 29, p. 861, Sep. 2018.

[15] E. Becht et al., "Dimensionality reduction for visualizing single-cell data using UMAP," *Nature Biotechnol.*, vol. 37, no. 1, pp. 38–44, Jan. 2019.

[16] M. Maire, T. Narihira, and S. X. Yu, "Affinity CNN: Learning pixel-centric pairwise relations for figure/ground embedding," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 174–182.

[17] P. Krähenbühl and V. Koltun, "Efficient inference in fully connected CRFs with Gaussian edge potentials," 2012, *arXiv:1210.5644*.

[18] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Trans. Syst. Man, Cybern.*, vol. SMC-9, no. 1, pp. 62–66, Jan. 1979.

[19] J. L. Moigne, N. S. Netanyahu, and R. D. Eastman, *Image Registration for Remote Sensing*. Cambridge, U.K.: Cambridge Univ. Press, 2011.

[20] D. Solarna, L. Maggiolo, G. Moser, and S. B. Serpico, "A tiling-based strategy for large-scale multisensor optical-SAR image registration," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2022, pp. 127–130.

[21] S. Mika, B. Schölkopf, A. Smola, K.-R. Müller, M. Scholz, and G. Rätsch, "Kernel PCA and de-noising in feature spaces," in *Proc. Adv. Neural Inf. Proc. Syst.*, vol. 11, M. Kearns, S. Solla, and D. Cohn, Eds., Cambridge, MA, USA: MIT Press, 1998, pp. 1–7.

Open Access funding provided by 'Università degli Studi di Genova' within the CRUI CARE Agreement