Leveraging Generative Deep Learning Models for Enhanced Change Detection in Heterogeneous Remote Sensing Data

1st Moslem Ouled Sghaier
OODA Technologies Inc.
Montreal, Quebec, Canada
moslem.ouled-sghaier@ooda.ca

2nd Melita Hadzagic
OODA Technologies Inc.
Montreal, Quebec, Canada
melita.hadzagic@ooda.ca

3rd Jun Ye Yu
OODA Technologies Inc.
Montreal, Quebec, Canada
jun-ye.yu@ooda.ca

4th Sofia Shton
OODA Technologies Inc.
Montreal, Quebec, Canada
shtonsofia23@gmail.com

5th Elisa Shahbazian
OODA Technologies Inc.
Montreal, Quebec, Canada
elisa.shahbazian@ooda.ca

***Abstract*—In this paper, we introduce an innovative approach for Change Detection (CD) in heterogeneous (multimodal) multitemporal remote sensing (RS) images employing deep features comparison through the utilization of two advanced deep learning models: Generative Adversarial Networks (GANs) and autoencoders. First, Deep Convolutional GANs are implemented to convert multimodal image(s) into synthetic image(s) of the same modality. Subsequently, autoencoders are trained and employed to extract compressed representations of both initial images. Finally, a change map is obtained by combining/fusing the original image with its corresponding generated change-free image resulting from the difference between the two learned compressed representations. Our proposed CD technique can accommodate CD algorithms for RS images expressing any type of change. Experimental evaluations on very high-resolution optical and Synthetic Aperture Radar (SAR) imagery validate the enhanced performance of the proposed method compared to existing state-of-the-art CD techniques in handling heterogeneous RS data.**

***Index Terms*—Heterogeneous image fusion, multimodal images, generative adversarial networks, change detection.**

I. Introduction

In Earth observation (EO) applications with numerous uses, including climate change, urban development monitoring, and disaster management, identifying regions undergoing significant changes in radiometric, textural, and/or geometric information from multisource/multimodal and multitemporal RS data is crucial for automated EO systems. However, CD from heterogeneous data, obtained from various sensors such as satellites and unmanned aerial vehicles, poses challenges due to differences in feature spaces, image domains, and class signatures. These disparities can hinder effective CD, especially in real-world scenarios characterized by cost constraints or limited availability of specific sensors during time-sensitive situations like disaster management. Combining heterogeneous sensors such as Synthetic Aperture Radar (SAR) and optical images offers significant advantages due to the complementary nature of these sensors. SAR excels in providing imagery regardless of weather conditions, including cloudy or extreme weather scenarios, thanks to its ability to penetrate clouds and work effectively during day or night. Additionally, SAR can detect surface roughness, moisture content, and terrain structure, providing valuable information not easily discernible with optical imagery. On the other hand, optical images offer high resolution and superior interpretability, capturing fine details with clarity. By fusing SAR and optical data, users can leverage the strengths of both sensors, obtaining comprehensive insights into land cover, land use, and environmental changes with enhanced accuracy and reliability.

The processing of heterogeneous images for detecting changes involves several common techniques, such as pixel-based fusion [1], [2] and kernel-based fusion [3], [4]. Although effective, these methods face challenges with handling noise and registration errors [3], necessitating meticulous features selection. Moreover, they typically focus on particular sensor types and change categories ([3]), limiting their applicability to diverse use cases. In contrast, unsupervised Bayesian methods ([5], [6]) offer potential enhancements but lack clear performance advantages over traditional methods. Recent studies have explored the utilization of Deep Learning (DL) architectures, like Convolutional Neural Networks (CNNs), Stacked Autoencoders (SAEs), and Generative Adversarial Networks (GANs) [7]–[10], to address complexities in RS data analysis and features extraction. Compared to conventional methods based on hand-crafted features extraction, these models can effectively learn high-level features from varying RS data sources. However, significant amounts of labeled training data are required to attain satisfactory outcomes [11]. Efforts to minimize training sample requirements include employing autoencoders [12] or utilizing GANs [13]–[16]. Nevertheless, none of these investigations have addressed multimodal (heterogeneous) image sets for CD.

To circumvent the complexities inherent in multimodal image fusion encountered in conventional fusion methodologies relying on directly extracted features from input raw images, while capitalizing on the complementary information offered by heterogeneous data, as well as the automation enabled by DL in features learning without supervision, and concurrently reducing the requisite training datasets, this paper introduces an innovative unsupervised approach to heterogeneous image fusion for CD. Herein, a GAN is employed to homogenize the fusion problem via image translation. Subsequently, an autoencoder is trained to acquire a compressed representation of the two images. Ultimately, a change map is derived by contrasting the original image with its corresponding change-free image generated from the discrepancy between the two learned compressed representations. The DL models are trained utilizing very high-resolution optical and SAR image datasets. A similar translation method, albeit supervised, is proposed in [17]. The proposed method is agnostic and capable of accommodating CD algorithms for RS images irrespective of the type of change undergone. Depending on the nature and type of change to be identified, optical images can be translated to SAR and vice versa, simplifying the CD problem to a straightforward comparison between two images of the same modality. The efficacy of the proposed approach is verified with real images and juxtaposed with a state-of-the-art CD technique for heterogeneous images as delineated in [18], demonstrating resilience to variations in image acquisition angle, luminosity, and scale.

The remainder of this paper is organized as follows. Section II describes our proposed method for CD, including the pixel-to-pixel image translation technique and the deep features-based CD approach. Section III presents the data, the evaluation of our method, the comparison with other methods and a discussion of the results, while Section IV presents the conclusions of this work and outlines future works.

II. Methodology

In this section, we outline the proposed methodology for CD in heterogeneous imagery, a process that is not confined to any specific type of alteration. For instance, in the context of disaster management, SAR images prove valuable for generating damage maps related to flood extent, thanks to their ability to accentuate flooded regions via their dark radiometric signature and uniform texture. Conversely, when changes stem from events like earthquakes or wildfires, optical images are preferred as they facilitate the detection of subtle alterations in pixel radiometry and the identification of damage to infrastructure.

Within our framework, we can convert an input SAR image into a synthetic optical image, or vice versa, to match the modality of a target image, depending on the nature of the desired change and the application at hand. The functional flow of our proposed method is illustrated in Figure 1. Achieving accurate conversion between images of differing modalities requires a model capable of learning and extracting the deep features characteristic of each image type. Specifically, we

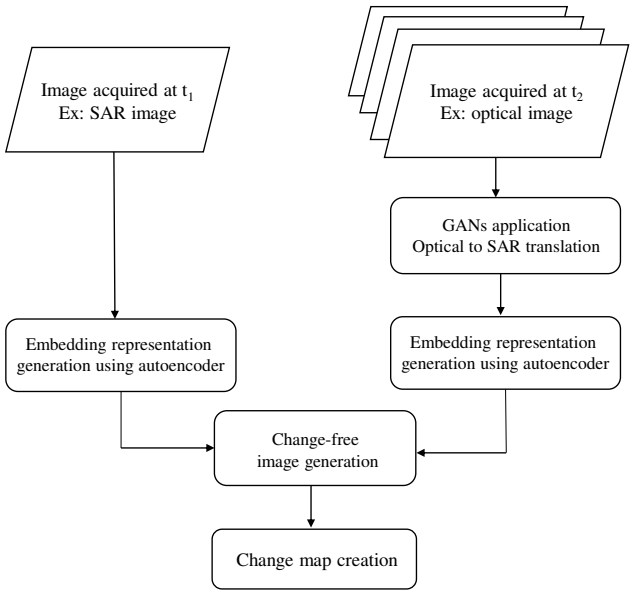


Figure 1: Flowchart of the proposed CD approach (use-case flood monitoring)

employ pixel-to-pixel conversion techniques based on GAN architectures [19], resulting in two distinct GAN models: one for translating optical to synthetic SAR images and another for SAR to synthetic optical image translation. Subsequently, an autoencoder model is deployed to generate an embedded representation from each image, allowing for comparison between the input images acquired at t_1 and t_2 based on their embedded deep features vectors. Finally, the difference between the two vectors yields the change map.

A. Synthetic RS Image Generation using GANs

The GAN used for image conversion (see Fig. 2) consists of two artificial neural networks in competition: a generative model G which tries to capture the data distribution and learn hidden relations between input features, and a discriminatory model D that aims at distinguishing between samples coming from the training dataset or the generator G. These two models are trained simultaneously, where G is trying to fool the discriminator D, while D is trying to accurately assess whether a sample comes from the training data or is produced by the generator.

We adopt the so-called *Pixel-to-Pixel* conversion using GANs, an approach that allows to train a DCNN for image-to-image conversion tasks, where a target image is generated based on a given input image. To apply this technique, the input dataset must contain pairs of images representing the same region of interest acquired by two different sensors which are matched accordingly prior to the conversion (see Section III. The generator, which is based on U-Net architecture [20], tries to generate an artificial RS image from one of the images in a given pair, while the discriminator based on samples of the training dataset decides if the input pair is real or generated (or so-called "fake") [20]. Figure 2 illustrates the proposed

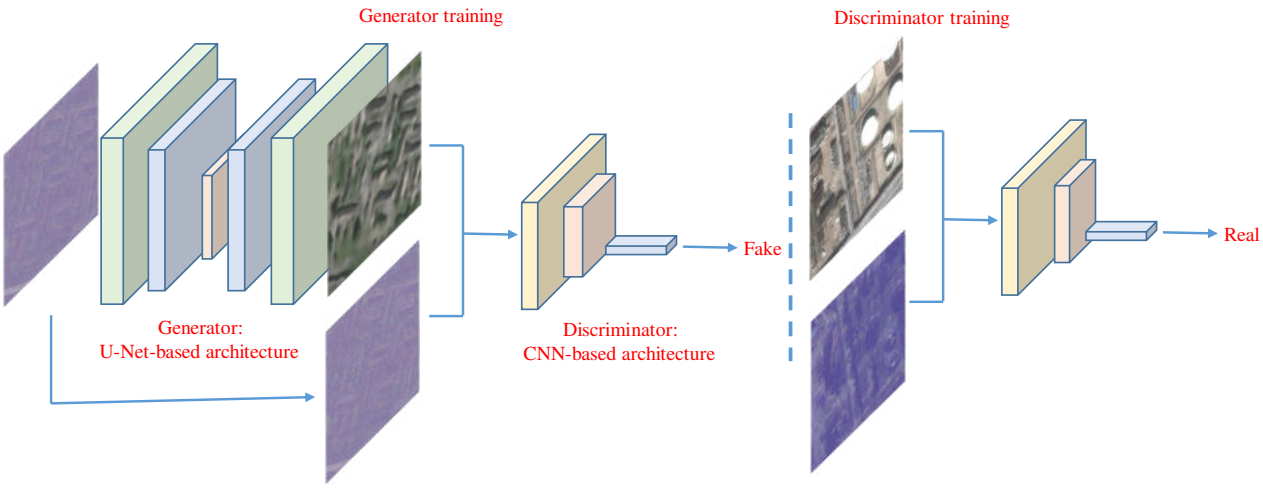


Figure 2: RS data conversion using Pixel-to-Pixel GANs architecture.

architecture design for the RS image-to-image conversion task. Based on the DL architecture described above, the two different networks were trained using the data described in Section III where the first GAN is used to generate a synthetic SAR image from an input optical image, and the second GAN is used to produce a synthetic optical image from an input SAR image, where the term "fake" refers to a synthetic/generated image.

In the design of the desired GANs architecture, the choice of the loss function is a challenging task. In discriminative DL models such as CNN [21], Recurrent Neural Network (RNN) [22] or autoencoders [23], the convergence of a model is defined as the minimization of the chosen loss function on the training dataset. However, in the case of GANs models, the convergence signals the end of the competition between the two networks. Therefore, an equilibrium between generator and discriminator loss is sought. In the literature, different loss functions are applied to train GANs, the most common being the non-saturating loss, the Least Squares [24] and Wasserstein loss [25] in larger and more recent GAN models. Given the considerable size of the neural networks architecture developed in this work and the number of images available in our dataset, the Wasserstein GAN loss is selected as a measure of the distance between predicted data distribution and the ground truth, with the aim of improving stability when training the model as well as increasing the gap between scores for real and generated images since it provides a loss function that correlates with the quality of the generated image. Hence, we define the Wasserstein loss [25] associated with the generator and the discriminator as follows:

- Critic Loss (Discriminator) = $\frac{1}{n}\sum$ critic score on synthetic images – $\frac{1}{n}\sum$ critic score on real images
- Generator Loss = - $\frac{1}{n}\sum$ average critic score on synthetic images

where the average scores are calculated across a mini-batch of n image samples.

B. Change Detection based on Deep Features Comparison

After the translation, having the two images of the same modality, any existing CD technique designed to identify changes between images of the same modality can be applied. However, the selected method has to be robust enough to be able to detect small changes caused by the imperfections in the synthetic image generation process, different acquisition angles, and image co-registration errors. To overcome this problem, we further propose to use deep features obtained from the previous step to extract relevant changes. Figure 3 shows our deep features-based CD approach in details.

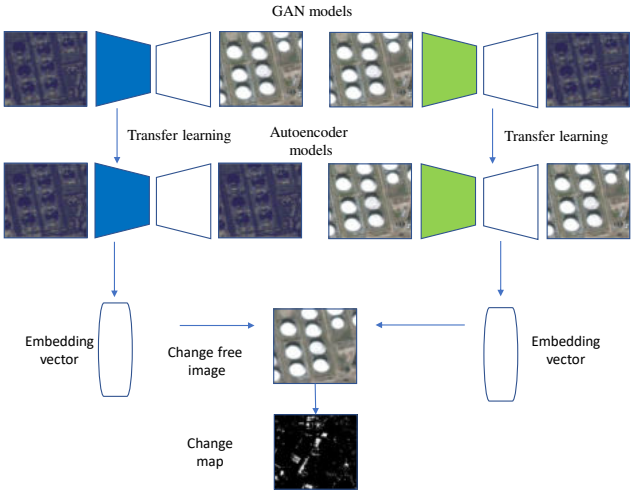


Figure 3: Deep features-based CD technique. The contracting path of the generator in the GAN model is reused as the encoder of the autoencoder model.

Let I_1 and I_2 be two input images acquired using two different sensors S_1 and S_2, respectively, and $\widehat{I}_1$ and $\widehat{I}_2$ their associated synthetic images generated using the GAN models as described in the previous section. First, an autoencoder is used to learn features specific for the given training data and

produce a compressed representation of the input image. The proposed architecture is based on the combination of the sub-model constituting the first part of the generator, also called the contracting path of the U-Net model (blue and green parts in Figure3), with a CNN-based decoder. The weights from the GAN model are used in the autoencoder training phase so that only the decoder part of the autoencoder is updated. In addition, only images of the same modality are used for training: (I_1, $\widehat{I_2}$) for the first image type and (I_2, $\widehat{I_1}$) for the second one, depending on the nature of the target change to be detected, and at each iteration the same image is given as input and output for the model. Once the model has been trained, a vector associated with the autoencoder bottleneck layer and corresponding to the compressed representation is obtained.

Let v_1 be the vector associated to image I_1 and $\widehat{v_2}$ the one corresponding to image $\widehat{I_2}$. Identifying changes between two input images amounts to computing the difference w between those two vectors, i.e., the change vector $w = \widehat{v_2} - v_1$. By subtracting the change vector w from the image acquired at t_2 and reconstructing the resulting image using the autoencoder, an image having the same radiometric and textural characteristics of that image but free from changed objects is obtained. Finally, a binary change map can be produced by subtracting the original image acquired at t_2 from the generated change-free image.

III. Experimentation

In this section, we describe the datasets used for training the proposed DL models, the pre-processing steps and results, together with the evaluation of the proposed method, the results and their discussion.

A. Datasets

SAR and optical pairs of images from the SpaceNet6 competition [26] were used to train the DL models described in Section II. This challenge aimed at developing novel approaches to automatically extract building footprints with computer vision and artificial intelligence algorithms. Available data is divided into two separate datasets:

1) The optical dataset consists of 4 image products including the panchromatic band, pan-sharpened RGB, RGB-NIR data ($0.5m$) and RGBNIR data ($2.0m$) provided by the Maxar WorldView 2 satellite. A cloud-free image strip was collected on August 31, 2019 at 10:44AM from a look angle of $18.4°$ off-nadir with an overall area of $236km^2$.
2) The SAR dataset comes from Capella Space's X-band quad-pol sensor mounted on an aircraft which mimics the space-borne sensors on Capella's future constellation of satellites. The dataset contains 204 individual image strips with two different processing levels: georeferenced magnitude and polarimetry data and Single Look Complex data captured over three days: August 4th, 23rd, and 24th 2019. Each strip features four polarizations (HH, HV, VH, and VV) of data in the X-band wavelength. Data is geo-registered, ortho-rectified and resampled with a lanczos interpolation to a spatial resolution of $0.5m \times 0.5m$ per pixel.

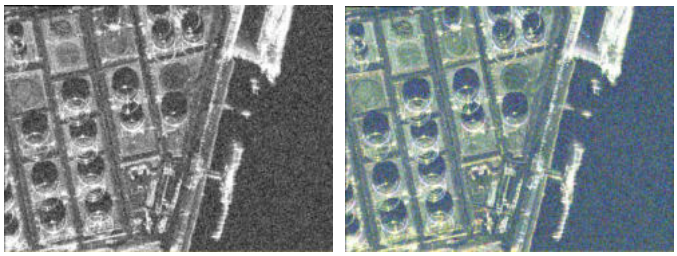
(a) SAR image. (b) Optical image.

Figure 4: A pair of SAR and optical images of the same region.

All data is tiled using $450m \times 450m$ (900 pixels $\times$ 900 pixels) tiles. The SAR images are first tiled, and then the same operation is applied to the corresponding optical image to match the SAR extent and the pixel grid. Figure 4 shows a SAR image and its corresponding optical image from the SpaceNet6 dataset.

B. Image Preprocessing

The proposed DL models are designed to process color satellite images. In the case of optical images, no additional preprocessing is needed as the raw input images already provide RGB channels for bands 1, 2 and 3, respectively. However, for SAR images, a conversion from full-polarimetric data to RGB color composition is needed. It was opted for a straightforward false color decomposition that associates the SAR HH intensity channel to the Red band, the VV intensity channel to the Green band, and the HV intensity channel to the blue band. The results of the applied decomposition to an input raw SAR image are shown in Figure 5. Other



(a) Input SAR image HH channel (Intensity). (b) Derived false color composition.

Figure 5: Comparison between the input SAR image and the derived false color image

preprocessing techniques are also applied in this step as part of data preparation, the resizing of input images to $512 \times 512 \times 3$, and normalization by dividing pixel values by 255. The data augmentation techniques such as rotation, translation, flipping and color modification are also applied.

C. Model Training and Hyperparameters Selection

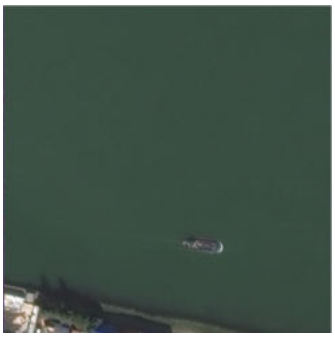
The training of models, validation and testing are performed on a local server optimized for ML/DL algorithms design and

training. The machine is equipped with an Intel(R) Xeon(R) CPU E5-2620 v4 @ 2.10GHz processor, 257GB of RAM and NVIDIA TITAN RTX 24GB GPU. Using this configuration the models were trained with a batch size of 16, 500 epochs and a learning rate of 2×10^{-4}.

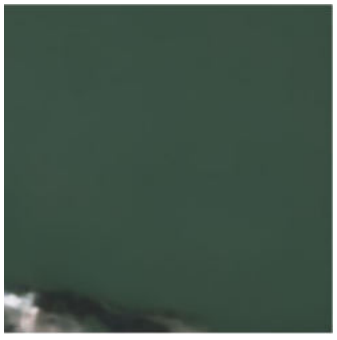
D. Results

To assess the performance of the DL model for Pixel-to-Pixel (optical or SAR) image conversion, the input SAR and optical images of three different regions of interest, each depicted in the first row of Figure 6 were considered: 1. a SAR image corresponding to an urban area (site 1), Figure 6(a), 2. a SAR image corresponding to a rural region (site 2), Figure 6(d), and 3. a SAR image corresponding to a coastal area (site 3), Figure 6(g). The second row of Figure 6 contains the associated converted/generated synthetic images from the originals (SAR or optical), i.e., Figure 6(b), Figure 6(e) and Figure 6(h), respectively. The third row contains the corresponding target optical or SAR image, depending on the conversion type. Visual inspections of the obtained results in Figure 6 show consistency of synthetic images with the ground truth when the Pixel-to-Pixel conversion algorithm is applied to three regions of interest.

(a) Input SAR image acquired at t_1.



(c) Input optical image acquired at t_2.



(b) Generated synthetic optical image associated to input SAR image.

(d) Obtained change map.

Figure 7: Results of the proposed CD method applied to heterogeneous RS images.

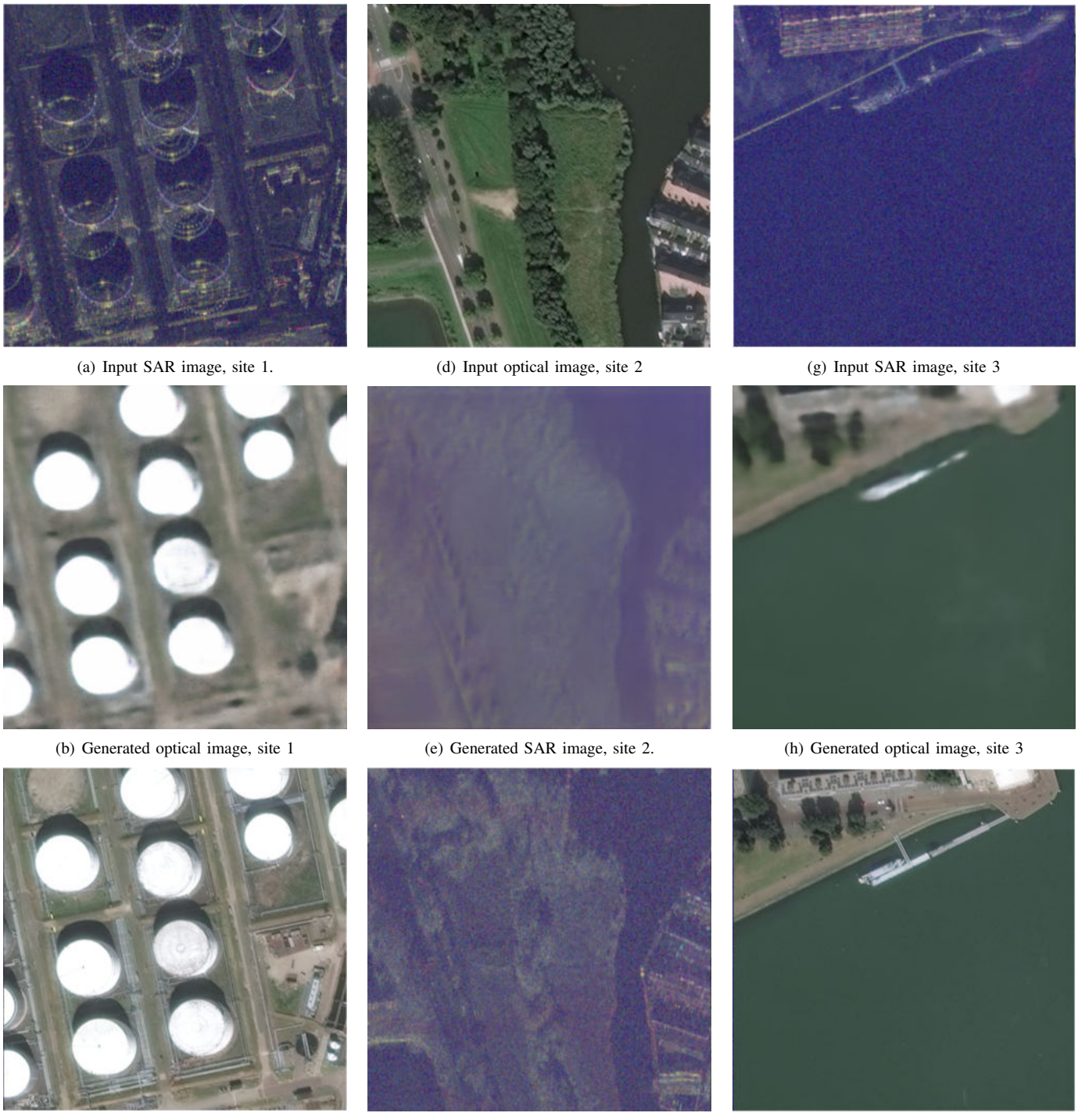
The results obtained in site 2 demonstrate some gaps in predicting the vegetation signature contained in the input SAR image. This can be explained by the complexity of simulating the trees and grass backscatter model composed of many randomly distributed elementary scatterers within a resolution cell. Despite this limitation, the vegetation areas can be well identified as they are separated from other classes present in the image such as the water surfaces and the road sides, which are also well delimited. When applied to the site 3, the model was able to correctly reproduce the sea surface radiometric and textural signature. The trees contained in the input image and their associated shades were also reproduced with a high level of detail. The overall structure of the port is present in the resulting image but some parts are missing due to their very small dimensions.

To properly train the DL models and assess the CD method performance, the datasets of multimodal images of the same region of interest (training) and the datasets of such images acquired during two different periods (CD assessment) are necessary. Similarly to other open multimodal imagery datasets, the used RS dataset SpaceNet6 providing multimodal images of the same region of interest has a very short time lapse between optical and SAR images acquisition so that almost no significant change can be identified. In order to overcome this problem, the SAR and optical image pairs were carefully analyzed by search of changes that may occur within a short period of time (2-3 days), focusing on objects and their movements (e.g. a vessel movement). Figure 7 shows the result of the proposed CD technique applied to the selected test images where Figure 7(a) depicts a SAR image acquired at t_1, Figure 7(c) depicts the associated ground truth optical image acquired at t_2 showing a vessel in the sea, Figure 7(b) shows the generated synthetic optical image obtained by applying our DL based image generator while Figure 7(d) shows the obtained binary change map.

It can be observed that the synthetic image generation algorithm was able to provide an artificial optical image that matches very well the target optical image acquired at t_2 in terms of radiometric and textural information while limiting the generation of new artifacts in the resulting image. The obtained change map demonstrates that our deep features-based CD method was able to identify the changed region between the two images and was robust in terms of neglecting the small changes caused by the process of synthetic image generation.

E. Evaluation

In this section, we present the results of a comparison of our proposed CD approach against another DL unsupervised CD technique that can be applied to heterogeneous images such as optical and SAR, and which is described in [18]. In this paper, authors introduce a new unsupervised methodology aimed at aligning the code spaces of two autoencoders based on affinity information extracted from input data. This approach is integral to a heterogeneous change detection framework designed to achieve latent space entanglement, even when input images contain changes. The method effectively mitigates the misleading impact of changes during training and demonstrates consistent performance comparable to or exceeding state-of-the-art methods across four different datasets. However, its



(a) Input SAR image, site 1. (d) Input optical image, site 2 (g) Input SAR image, site 3

(b) Generated optical image, site 1 (e) Generated SAR image, site 2. (h) Generated optical image, site 3

(c) Target optical image, site 1. (f) Target SAR image, site 2. (i) Target optical image, site 3.

Figure 6: Results of the proposed Pixel-to-Pixel image translation method applied to three different sites.

Table I: Four metrics used in CD evaluation

	I_{ref}	I
True positive (TP)	1	1
False positive (FP)	0	1
True negative (TN)	0	0
False negative (FN)	1	0

performance deteriorates when handling a limited amount of features in the input, particularly when only one channel is available in one of the images, presenting challenges in regression from one variable to many. Despite this, the methodology adeptly manages multispectral and multipolarization images, successfully mapping data across domains in a meaningful manner. It aims to align the code layers of two autoencoders so that the output of one encoder can be the input of both decoders. To ensure that the code alignment is unsupervised, the affinity matrices of the training patches are computed and compared. Then, the extracted local information is utilized to verify that the pixel pairs similar in both input domains also have a high correlation in the common latent space.

To assess the performance of the two CD techniques, the same performance metrics described in [27] are used: Let I_{Ref} and I be two binary images, where I_{Ref} denotes the reference image and I denotes the input image. A given pixel can be attributed to one of the four categories defined in Table I.

Let P be the total positive instance equal to $TP+FN$, and N be the total negative instance equal to $FP+TN$. From those four definitions, we can define the following metrics:

$$\textit{Accuracy} = \frac{TP}{TP+FP};\ \textit{Precision} = \frac{TP+TN}{P+N};\ \textit{Recall} = \frac{TP}{P} \quad (1)$$

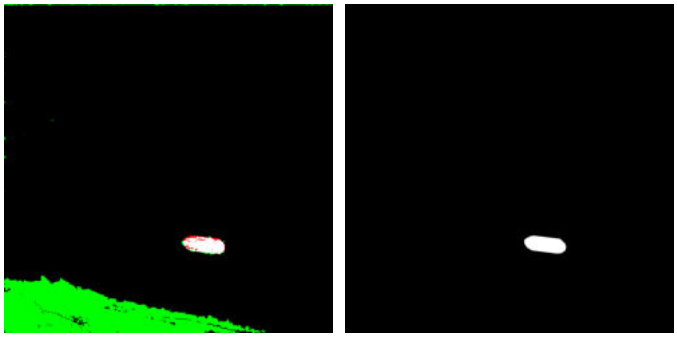
Table II presents the quantitative results of the comparison obtained using the Accuracy, Precision and Recall performance metrics. The proposed CD method compares favorably in terms of accuracy and recall, and has a similar value of precision. Both methods were applied to the boat dataset, and although both methods accurately detect the change, the method from [18] falsely marks the coastline region as changed as seen in Figure 8.

Table II: Comparison of CD methods using the accuracy, precision and recall metrics.

	Accuracy	**Precision**	**Recall**
Code-aligned autoencoders [18]	0.925	**0.92**	0.85
Our CD approach	**0.943**	0.91	**0.93**

Figure 8 shows the confusion map, with the black color marking true negatives, white marking true positives, green marking false positives, and red marking false negatives.

The obtained results demonstrate that our approach outperforms the technique proposed in [18], particularly in terms of accuracy and recall. While both approaches were devised for unsupervised change detection from heterogeneous remote sensing images using a DL-based image translation technique,



(a) Method proposed in [18]. (b) Ground truth.

Figure 8: Performance of the method from [18] on the dataset.

our method applies a comparison of deep features extracted from the latent space representation of generated synthetic images. This approach proves more robust to subtle changes, allowing for pixel-wise computation as suggested in [18]. As illustrated in Figure 4, the method described in [18] successfully detects changes in a specific region of interest within the image. However, it also produces false positives (highlighted in green in the image) by identifying other objects as changed entities.

IV. Conclusions

In this paper, a novel approach for CD from heterogeneous satellite images is presented, leveraging GANs for Pixel-to-Pixel image conversion and deep features comparison. Initially, DCNN-based adversarial networks are employed to generate synthetic remote sensing (RS) images, simplifying the CD task to a comparison between images of the same nature. Subsequently, a second DL model, utilizing autoencoders architecture, is developed to learn a compressed representation of the input image. The weights obtained from the DCNN GAN training facilitate the acceleration of this subsequent training, and the resulting deep features are juxtaposed to discern changed information and produce a change-free image. The change map is derived by subtracting the original image captured at time t_2 from its corresponding generated change-free image. Images sourced from the SpaceNet6 challenge are utilized to train the proposed Pixel-to-Pixel image translation models and evaluate the performance of the CD technique. Preliminary quantitative and qualitative analyses of the evaluation results underscore the potential of the proposed CD technique and the added value of employing generative DL models in RS image processing.

In forthcoming research, we plan to explore alternative generative models, including cycle GANs, which have the potential to enhance our approach performance and execution time by consolidating both image translation directions into a single model, thereby replacing the need for two distinct DL models. Additionally, we aim to investigate the effectiveness of progressive GANs, a reliable method for training GAN

models to produce large, high-quality images by gradually scaling up the model's size during training. Furthermore, our efforts will focus on expanding the dataset, particularly in regards to disaster-related imagery, to bolster the resilience and adaptability of the models for improved generalization across diverse scenarios.

ACKNOWLEDGMENTS

This research was funded by the Canadian Space Agency (CSA). We extend the acknowledgements to Maxar Technologies for WorldView2 optical images and Capella Space for SAR data.

REFERENCES

[1] W. Zhao, A. Li, X. Nan, Z. Zhang, and G. Lei, "Postearthquake landslides mapping from landsat-8 data for the 2015 nepal earthquake using a pixel-based change detection method," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 10, no. 5, pp. 1758–1768, 2017.

[2] Y. Han, A. Javed, S. Jung, and S. Liu, "Object-based change detection of very high resolution images by fusing pixel-based change detection results using weighted dempster–shafer theory," *Remote Sensing*, vol. 12, no. 6, 2020. [Online]. Available: https://www.mdpi.com/2072-4292/12/6/983

[3] M. O. Sghaier, M. Hadzagic, and J. Patera, "Fusion of SAR and multispectral satellite images using multiscale analysis and Dempster-Shafer theory for flood extent extraction," in *IEEE 22th International Conference on Information Fusion, FUSION 2019, Ottawa, ON, Canada, July 2-5*, 2019, pp. 1–8.

[4] N. Gupta, P. Singh, and S. Ari, "Feature fusion based unsupervised change detection in optical satellite images," in *2019 IEEE 5th International Conference for Convergence in Technology (I2CT)*, 2019, pp. 1–5.

[5] R. Touati, M. Mignotte, and M. Dahmane, "Multimodal change detection in remote sensing images using an unsupervised pixel pairwise based markov random field model," *IEEE Transactions on Image Processing*, vol. 12, no. 1, pp. 321–333, 2019.

[6] V. Ferraris, N. Dobigeon, Y. Cruz Cavalcanti, T. Oberlin, and M. Chabert, "Unsupervised change detection for multimodal remote sensing images via coupled dictionary learning and sparse coding," in *In: IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2020)*, 2020.

[7] J. Li, D. Hong, L. Gao, J. Yao, K. Zheng, B. Zhang, and J. Chanussote, "Deep learning in multimodal remote sensing data fusion: A comprehensive review," *ELSEVIER International Journal of Applied Earth Observation and Geoinformation*, vol. 112, no. 1, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S156984322200124

[8] S. N. K. B. Amit, S. Shiraishi, T. Inoshita, and Y. Aoki, "Analysis of satellite images for disaster detection," in *2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 2016, pp. 5189–5192.

[9] X. Luo, X. Li, Y. Wu, W. Hou, M. Wang, Y. Jin, and W. Xu, "Research on change detection method of high-resolution remote sensing images based on subpixel convolution," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 1447–1457, 2021.

[10] A. Gupta, S. Watson, and H. Yin, "Deep learning-based aerial image segmentation with open data for disaster impact assessment," *Neurocomputing*, vol. 439, pp. 22–33, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0925231221001429

[11] S. T. Seydi, M. Hasanlou, and M. Amani, "A new end-to-end multi-dimensional cnn framework for land cover/land use change detection in multi-source remote sensing datasets," *Remote Sensing*, vol. 12, no. 12, 2020. [Online]. Available: https://www.mdpi.com/2072-4292/12/12/2010

[12] D. B. Mesquita, R. F. d. Santos, D. G. Macharet, M. F. M. Campos, and E. R. Nascimento, "Fully convolutional siamese autoencoder for change detection in uav aerial images," *IEEE Geoscience and Remote Sensing Letters*, vol. 17, no. 8, pp. 1455–1459, 2020.

[13] J. Liu, K. Chen, G. Xu, H. Li, M. Yan, W. Diao, and X. Sun, "Semi-supervised change detection based on graphs with generative adversarial networks," in *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 2019, pp. 74–77.

[14] M. Gong, Y. Yang, T. Zhan, X. Niu, and S. Li, "A generative discriminatory classified network for change detection in multispectral imagery," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 12, no. 1, pp. 321–333, 2019.

[15] M. Gong, X. Niu, P. Zhang, and Z. Li, "Generative adversarial networks for change detection in multispectral imagery," *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 12, pp. 2310–2314, 2017.

[16] C. Ren, X. Wang, J. Gao, X. Zhou, and H. Chen, "Unsupervised change detection in satellite images with generative adversarial network," *IEEE Transactions on Geoscience and Remote Sensing*, pp. 1–15, 2020.

[17] X. Li, D. Zhengshun, Y. Huang, and Z. Tan, "A deep translation (gan) based change detection network for optical and sar remote sensing images," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 179, pp. 14–34, 09 2021.

[18] L. T. Luppino, M. Hansen, M. Kampffmeyer, F. M. Bianchi, G. Moser, R. Jenssen, and S. Anfinsen, "Code-aligned autoencoders for unsupervised change detection in multimodal remote sensing images," *IEEE Transactions on Neural Networks and Learning Systems*, vol. PP, pp. 1–13, 05 2022.

[19] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Proceedings of the 27th International Conference on Neural Information Processing Systems - Volume 2*, ser. NIPS'14. Cambridge, MA, USA: MIT Press, 2014, p. 2672–2680.

[20] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.

[21] D. Peng, L. Bruzzone, Y. Zhang, H. Guan, H. Ding, and X. Huang, "Semicdnet: A semisupervised convolutional neural network for change detection in high resolution remote-sensing images," *IEEE Transactions on Geoscience and Remote Sensing*, pp. 1–16, 2020.

[22] R. Jing, S. Liu, Z. Gong, Z. Wang, H. Guan, A. Gautam, and W. Zhao, "Object-based change detection for vhr remote sensing images based on a trisiamese-lstm," *International Journal of Remote Sensing*, vol. 41, no. 16, pp. 6209–6231, 2020. [Online]. Available: https://doi.org/10.1080/01431161.2020.1734253

[23] M. Hu, C. Wu, L. Zhang, and B. Du, "Hyperspectral anomaly change detection based on autoencoder," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 3750–3762, 2021.

[24] X. Mao, Q. Li, H. Xie, R. Y. Lau, Z. Wang, and S. P. Smolley, "Least squares generative adversarial networks," in *2017 IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 2813–2821.

[25] Y. Wei, X. Luo, L. Hu, Y. Peng, and J. Feng, "An improved unsupervised representation learning generative adversarial network for remote sensing image scene classification," *Remote Sensing Letters*, vol. 11, no. 6, pp. 598–607, 2020. [Online]. Available: https://doi.org/10.1080/2150704X.2020.1746854

[26] J. Shermeyer, D. Hogan, J. Brown, A. Van Etten, N. Weir, F. Pacifici, R. Hänsch, A. Bastidas, S. Soenen, T. Bacastow, and R. Lewis, "Spacenet 6: Multi-sensor all weather mapping dataset," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2020, pp. 768–777.

[27] M. Ouled Sghaier, I. Hammami, S. Foucher, and R. Lepage, "Flood extent mapping from time-series sar images based on texture analysis and data fusion," *Remote Sensing*, vol. 10, no. 2, 2018. [Online]. Available: https://www.mdpi.com/2072-4292/10/2/237