A new architecture for onboard change detection based on deep learning

A NEW ARCHITECTURE FOR ONBOARD CHANGE DETECTION BASED ON DEEP LEARNING

Gabriele Inzerillo[(1)], Diego Valsesia[(1)], Enrico Magli[(1)], Aniello Fiengo[(2)]

*[(1)]**Politecnico di Torino***
Department of Electronics and Telecommunications,
Corso Duca degli Abruzzi 24, 10129, Torino, Italy.
Email: gabriele.inzerillo@polito.it, diego.valsesia@polito.it, enrico.magli@polito.it

*[(2)]**European Space Agency***
European Space Research & Technology Centre (ESTEC),
Keplerlaan 1, 2200 AG Noordwijk, The Netherlands.
Email: aniello.fiengo@ext.esa.int

ABSTRACT

Change detection (CD) from satellite imagery is a task of paramount importance for monitoring Earth, land usage and disaster management. Traditionally, CD is performed at the ground segment where the long product formation pipeline adds significant delays between the time of acquisition and the availability to the end-user. In case of time-sensitive conditions, such as natural disasters, it would be desirable to perform CD directly onboard of the spacecraft, so that alerts could be prioritized for low-latency transmission. However, onboard CD is far from trivial and requires to address a number of issues, including efficient onboard storage, image registration, etc. within the constraints of onboard resources. In this paper, we present a framework towards building an onboard CD pipeline. In particular, the essential operations required by an onboard CD pipeline are efficient storage of the images acquired for a given location during multiple revisits, their geometric registration and, only then, the actual change detection algorithm. We seek to develop a deep-learning (DL) approach that addresses all these issues with a single neural network model that is end-to-end optimized for a desired tradeoff between computational complexity, storage requirements and accuracy or resolution of change detection. This neural network has a modular architecture, conceptually consisting of an image encoder into a compact feature space for storage, an image decoder, a registration module and a change detection module. In particular, one wonders what is the optimal way of storing the image information for the purpose of best detecting change. We argue that instead of saving the images, the optimal approach is to save a compact representation generated by the neural network encoder with the objective of maximizing the downstream CD performance for a given bitrate constraint. In this coding-for-machines philosophy, compression is optimized not for the visual quality of a human observer, but rather for the algorithm performing inference tasks, such as CD, on the compressed data.

INTRODUCTION

Change detection (CD) from satellite imagery plays a crucial role in a wide range of applications including environmental monitoring, land use analysis and disaster management. The ability to rapidly and accurately detect changes in the Earth's surface can provide critical insights for decision-makers, enabling timely responses in such situations such as deforestation monitoring, urban expansion and natural disaster response (e.g., floods, earthquakes, forest fires etc...) [1, 2]. Traditional approaches to CD rely on processing data at the ground segment, which introduces significant delays due to the long pipeline between the image acquisition and the final product's delivery to the end-user. While this delay is tolerable for many applications, time-sensitive scenarios such as disaster management demand low-latency solutions, where immediate detection and alerting could make a substantial difference in response effectiveness.

To address this latency issue, performing CD directly onboard the satellite has emerged as a promising alternative [3, 4]. By enabling the spacecraft to autonomously process data and detect changes, it is possible to prioritize and transmit only critical information, reducing transmission costs and, more importantly, ensuring rapid delivery of time-sensitive alerts. However, onboard CD introduces a number of technical challenges. Satellites are constrained by limited computational power, storage, and energy resources, making it difficult to implement traditional CD pipelines onboard [5]. Key operations such as efficient storage, accurate geometric registration, and change detection must be re-engineered to fit within these constraints.

In recent years, deep learning (DL) approaches have revolutionized computer vision tasks, including CD. Several works have explored DL-based methods for CD on Earth observation data, showing promising improvements in accuracy over traditional techniques [6, 7, 8]. However, the majority of these methods are designed for ground-segment processing, where computational and storage limitations are less restrictive. The challenge of applying DL to onboard CD remains largely unexplored, with key questions surrounding the efficient storage and representation of data for downstream processing tasks.

In this work, we propose a preliminary design of a deep learning-based framework to address the challenges of onboard CD. Specifically, we focus on designing an end-to-end optimized pipeline that encompasses efficient image storage, image registration and change detection, all while minimizing computational complexity and memory usage. Inspired by the "coding-for-machines" paradigm [9], we aim to develop a neural network architecture that compresses image data not for human visual consumption, but for optimal downstream inference by the CD algorithm. This approach departs from traditional compression schemes, instead optimizing the encoding process to maximize change detection accuracy given the limited onboard resources, using a deep learning compressing architecture. Furthermore, special attention must be given to the development of a registration module, as satellite images captured onboard are neither co-registered nor orthorectified, thus, to ensure accurate change detection, a registration step is essential. Lastly, we present some preliminary experiments for the CD and Registration modules.

METHOD

In this section, we provide a general overview of the proposed DL framework, as illustrated in Fig. 1, along with a preliminary design of both the registration module and the change detection module. Although the overall compression module, which includes the blue block in Fig. 1 (i.e., block named "Feature Extractor"), has not yet been developed, we outline a possible design for it in the next subsections. The complete framework consists of several interconnected modules, which we describe in detail below.

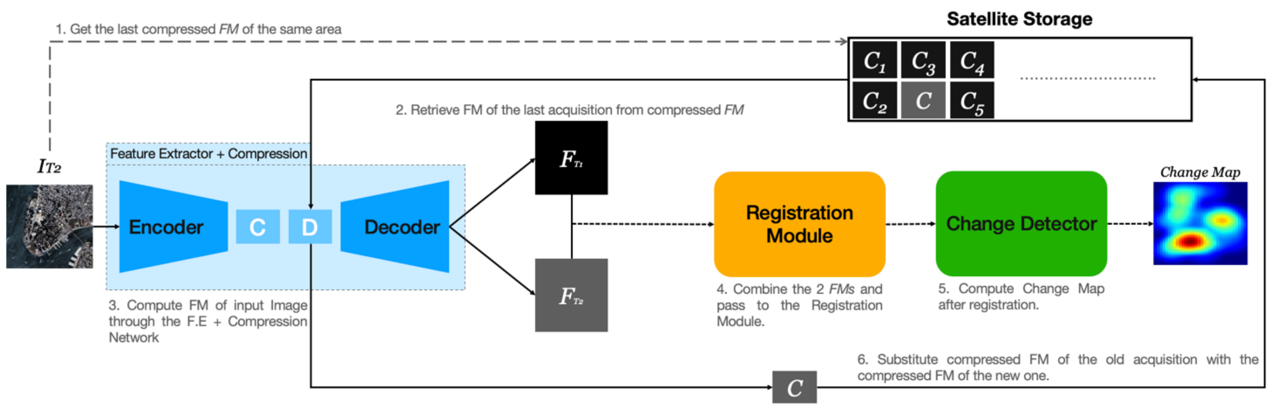

Fig. 1. Overall architecture and process steps of the proposed onboard CD framework. The satellite storage holds multiple compressed feature maps ($\boldsymbol{Cs}$) from the previous satellite revisit. When a new image is acquired ($\boldsymbol{I_{T_2}}$), the corresponding compressed feature map from the previous revisit ($\boldsymbol{F_{T_1}}$) is retrieved and decompressed. The new image is then processed through the feature extractor to generate the current feature map ($\boldsymbol{F_I}$). Both the current ($\boldsymbol{F_{T_1}}$) and previous ($\boldsymbol{F_{T_2}}$) feature maps are passed to the registration module, after which the change detection step is performed. Finally, $\boldsymbol{F_I}$ is compressed ($\boldsymbol{C}$) and stored in the satellite storage, while the old compressed feature map is discarded.

Dataset and Training Images

Since our ultimate goal is to perform CD directly onboard the satellite, we must train our architecture, including all three modules, using the raw images available onboard. We assume a Sentinel-2-like scenario, where the onboard images are non-orthorectified. Unfortunately, to the best of our knowledge, there are no publicly available datasets of non-orthorectified images. Therefore, we opted to train and test part of our framework (registration) using widely available co-registered and orthorectified Sentinel-2 image datasets [10, 11]. To approximate the geometry of non-orthorectified images, we applied geometric transformations such as affine (translation, rotation, scaling, shear) and perspective transformations. It is important to acknowledge that the differences between two images cannot be fully replicated by any transformation; however, this initial approach aims to reasonably approximate non-orthorectified images in the absence of real data.

Feature Extraction and Compression

To solve Change Detection tasks, it is crucial to keep track of past data, as we need to compare the current acquisition with previous ones. This means we must store images from past satellite revisits, however, given the limited onboard storage capacity, efficiently storing as much data as possible is essential to achieve broad coverage for our application.

A plausible straightforward approach to reduce the size of the stored data is to use a feature extractor, which essentially functions as part of an autoencoder during the inference. The feature extractor takes an image as input and generates a smaller, more compact feature map. This compressed representation should retain the same, or nearly the same, amount of information as the original image but in a more space-efficient format. By saving this smaller feature map instead of the full-resolution image, we can later reconstruct the original input during the decoding phase of the training loop. However, this approach may still be insufficient, as significantly reducing the size of the feature maps would require a high downsampling factor, leading to the loss of much of the spatial information necessary for accurate registration and change detection. To address this, one possible approach we plan to pursue is similar to that outlined in [12]. In this method, beyond feature extraction using an encoder, an additional compression step is introduced. This step is optimized end-to-end using the following loss function:

$$L_C = D + \lambda \cdot R \tag{1}$$

Here, D represents the distortion (e.g., in our training, it could be the Mean-Squared Error loss, which indicates the quality of the reconstructed image), R is the rate, and λ is a hyperparameter that can be manually adjusted to prioritize either the quality of the reconstruction or the level of data compression.

In our approach, we first focus on training the compression network independently to establish a baseline for compressing input images. This step allows us to develop an initial model for reducing data size efficiently. However, unlike traditional neural networks used for image compression, where the goal is typically to reconstruct visually accurate images, our strategy diverges after this initial step. Once the registration and change detection modules are also trained, we proceed to train the entire architecture end-to-end. In this end-to-end training phase, we replace the typical image reconstruction loss with a loss function tailored for change detection. This means that the 'distortion' is now measured in terms of the CD task performance rather than image reconstruction quality. As a result, the compressed feature maps are optimized for machine inference, specifically, to retain the most relevant information for performing accurate change detection, rather than being optimized for producing visually appealing images. This "coding-for-machines" approach ensures that the compression process prioritizes the preservation of information critical to the CD task, allowing for more efficient data storage and improved performance for change detection tasks.

Image Registration

Image registration is a critical preprocessing step in our framework, given that it operates on non-orthorectified and non-registered images. Effective registration is essential before performing change detection, as discrepancies in pixel geometry can significantly impact the accuracy of change detection.

We employ an unsupervised approach for training the image registration process. Our method begins with the dataset provided in [10], to which we apply the two types of augmentations already mentioned: affine and perspective transformations, used to generate an approximate representation of non-orthorectified images. In this setup, the augmented image serves as the target image, while the original image represents the source image. To train the registration model, depicted in Fig. 2, we utilize a multiscale neural network architecture inspired by the design proposed in [12].

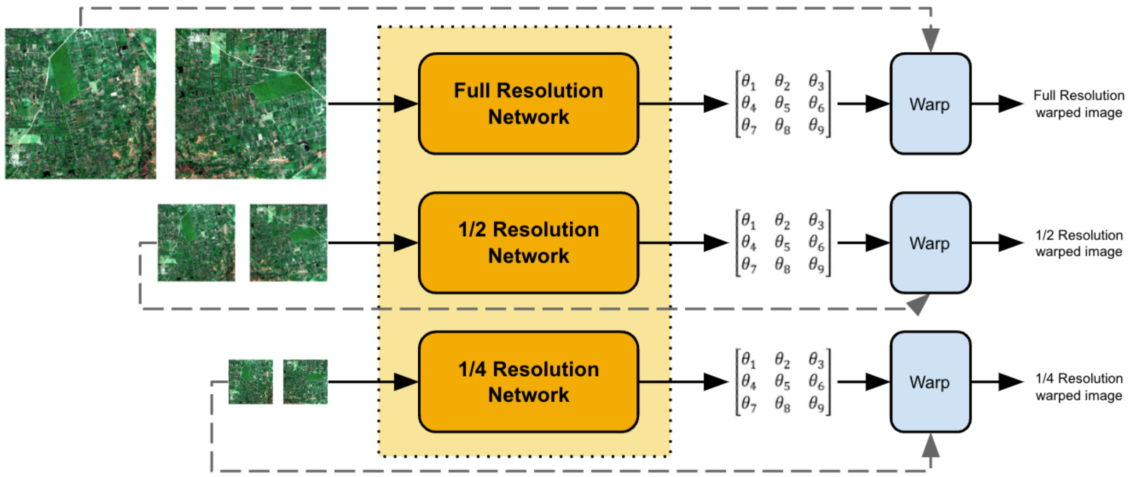

Fig. 2. Multiscale Neural Network for Image Registration. The network operates on three different resolutions of the image: full, half, and quarter resolutions. At each resolution, the network predicts the parameters of the homography matrix, which is then used to warp the source image to align with the target image. The warped images produced at each resolution are used to compute the total loss.

The network processes the original images and their downsampled versions at two additional scales (1/2 and 1/4). Each pair of source and target images, at all three scales, is input into the corresponding network, which learns to predict the corresponding homography matrices that align the source images with their respective target counterparts. The network is trained by minimizing the weighted sum of the mean squared errors (MSE) between each warped source image (transformed using the predicted homography) and its corresponding target image, facilitating accurate multiscale alignment. The total loss of the registration process is:

$$L_R = \lambda_1 \cdot L(t_1, w_1) + \lambda_{1/2} \cdot L\big(t_{1/2}, w_{1/2}\big) + \lambda_{1/4} \cdot L\big(t_{1/4}, w_{1/4}\big) \quad (2)$$

Where L is the MSE loss, t and w are the target and the warped images respectively, λ is a weight parameter and the subscript (1, ½ and ¼) indicates the resolution.

To train and evaluate our registration module, we utilized the dataset presented in [10]. However, given that this Sentinel-2 dataset was originally designed for cloud detection tasks, it includes many images with significant cloud cover, which could interfere with accurate image registration. To mitigate this issue, we carefully selected a subset of images from the dataset that contained minimal or no cloud coverage, ensuring that the registration task remained focused on aligning ground features without distortion introduced by cloud obstructions. This selection allowed us to effectively train the model while maintaining the integrity of the registration process.

Change Detection

Many sophisticated change detection architectures have been proposed in the literature [13, 14], achieving state-of-the-art performance. However, in this work, we prioritize minimizing computational complexity (in terms of FLOPS and the number of parameters), as our objective is to perform the task onboard with limited computational resources. To this end, we employ a classical U-Net architecture with early fusion and residual connections, avoiding more resource-intensive siamese networks which, while often yielding superior performance, come with computational overhead. Since we adopt a standard U-Net with early fusion, we do not delve deeply into the specifics of the architecture, as it remains consistent with well-established designs in the literature [6, 11, 15]. Early fusion concatenates the source and target images at the input stage, allowing the network to simultaneously process both images and learn feature differences directly from the combined input.

To further enhance efficiency, we plan to explore modifications to the standard U-Net architecture. These include integrating low-complexity components, such as depth-wise separable convolutions, which significantly reduce the number of parameters and computational cost compared to standard convolutions, and efficient activations and normalization layers such as the ReLU6 or Swish.

Lastly, to train our change detection module, we utilized the Onera Satellite Change Detection Dataset (OSCD) [11]. Although the dataset is relatively small, consisting of 24 pairs of multispectral images (14 for training and 10 for testing), we chose it specifically because it was created using Sentinel-2 imagery. This allows us to maintain consistency with the images used in the registration training phase, ensuring that both tasks, registration and change detection, are fine-tuned to the characteristics of the same sensor type (Sentinel-2). Using a dataset with images that differ significantly from Sentinel-2 images would introduce inconsistencies in the spatial and spectral characteristics, such as resolution, band distribution, and noise patterns, which could degrade the performance of the change detection model and of the overall architecture.

To mitigate the limitations of the dataset size, we employed data augmentation techniques, such as random rotations, flips, and small perturbations in brightness and contrast, to artificially expand the training set. This helps prevent overfitting and allows the model to generalize more effectively, even in the context of limited data. In the future, we plan to explore larger and more diverse datasets that can further enhance the robustness and accuracy of our change detection system.

EXPERIMENTAL RESULTS

In this section, we present some preliminary change detection experiments conducted using both the registration module and the change detection module to predict change maps. These experiments were performed under two distinct configurations, focusing on efficiency by utilizing only a single spectral band (i.e., Red) from the multispectral images:

1) **Co-Registered Image Pairs**: in the first configuration, we provided the registration module with image pairs from the OSCD dataset that were already co-registered (i.e., using original image pairs from OSCD dataset). This setup served as a baseline to evaluate the performance of the change detection module in ideal conditions, where geometric discrepancies between the image pairs had already been resolved.
2) **Transformed Image Pairs**: in the second configuration, we applied affine and perspective transformations to one image from each pair. The registration module was then tasked with compensating for these induced transformations. Once the registration was completed, the change detection module was used to predict the change maps. This configuration mimics real-world scenarios where images may not be perfectly aligned due to factors such as differences in acquisition geometry or atmospheric conditions.

The primary goal of these experiments was to assess the effectiveness of our registration module in correcting for geometric misalignments and its impact on the overall change detection performance. By comparing the results from both configurations, we aim to determine the importance of accurate image registration in the change detection process, particularly when dealing with non-orthorectified and unregistered images.

To evaluate the performance of the change detection module, we used standard metrics such as the F1-score and Intersection over Union (IoU), both macro-averaged across the dataset. The F1-score provides a balanced measure of precision and recall, while the IoU assesses the overlap between the predicted and ground-truth change maps. The results are presented in Table 1.

Table 1. Change Detection performance comparisons.

Configuration	**F1**	**IoU**
Co-Registered Pairs	53.72	46.04
Transformed Pairs	52.29	44.95

As we might have expected, the co-registered image pairs configuration yielded better performance in both F1-score and IoU compared to the transformed image pairs configuration. This suggests that even with successful registration, residual misalignments introduced by the transformation and subsequent compensation affect the accuracy of the change detection task. Nonetheless, the small performance gap between the two configurations highlights the robustness of the registration module in handling geometric transformations, although further improvements could be made to optimize alignment and subsequent change detection accuracy.

It is important to note that image registration is traditionally performed under the assumption that the two scenes being aligned are nearly identical. In our case, however, we face the challenge of aligning images where substantial changes may have occurred between acquisitions, such as the construction of new buildings, the appearance of new roads, or deforestation. As a result, the registration process may not be flawless, as the model may struggle to align features accurately when significant portions of the scene have changed or when new features are present. This limitation highlights a key area for future research and development. Our focus will be on improving the registration model to achieve more accurate alignment even in the presence of significant scene changes. Specifically, we aim to enhance the model's ability to identify and prioritize common features across images, rather than relying solely on predicting a homography matrix to account for geometric transformations.

Fig. 3 presents a qualitative result of our image registration process. The target image is augmented to create the non-registered (source) image, which we then warp to align with the target. For this specific example, the Peak Signal-to-Noise Ratio (PSNR) between the non-registered source image and the target image is 22.83, which improves to 26.18 after registration, indicating a significant enhancement in alignment. This improvement is also visually apparent in the second row of images.

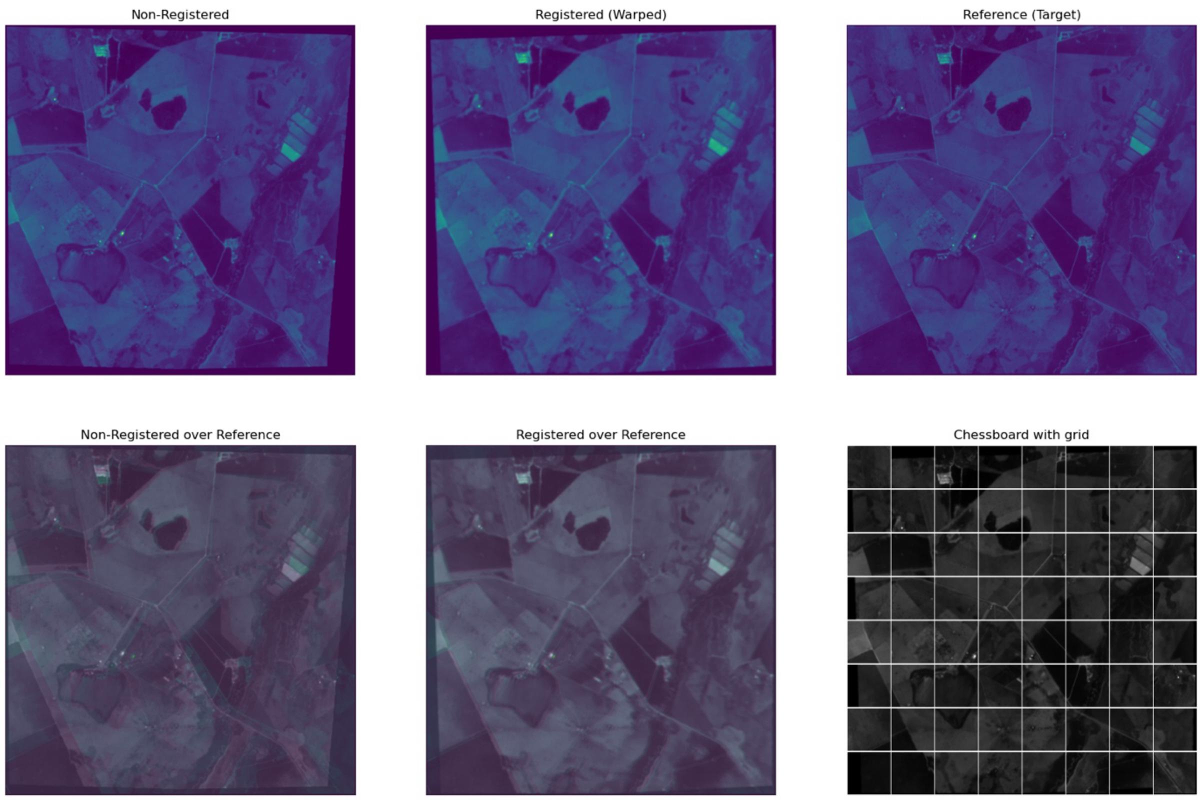

Fig. 3. Qualitative results of the image registration process. The top row displays the non-registered image (left), the registered (warped) image (center), and the reference image (right). The bottom row shows the overlay of the non-registered image on the reference (left), the overlay of the registered image on the reference (center), and a visualization using a checkerboard pattern (right) to illustrate the alignment between the registered image and the target.

CONCLUSIONS

In this paper, we have explored the potential of onboard DL for change detection in satellite imagery, focusing on the need for efficient model design to address the constraints posed by limited computational resources in spaceborne platforms. The proposed preliminary framework is designed to meet the growing demand for real-time, onboard satellite data processing, thereby reducing the need for extensive downlink bandwidth and enabling more timely responses to detected changes.

Future developments will focus on the design and implementation of the compression DL module, as well as the enhancement of the registration module to improve performance in cases where substantial changes exist between images. Additionally, we plan to conduct extensive experiments using a wider variety of datasets and operational scenarios to assess the generalizability and robustness of the approach across different conditions.

ACKNOWLEDGMENTS

This work has been received EISI support from ESA in the form of a co-funded PhD scholarship.

REFERENCES

[1] A. Singh, "Review article digital change detection techniques using remotely-sensed data." *International journal of remote sensing* 10.6 (1989): 989-1003.
[2] M. Hussain, D. Chen, A. Cheng, H. Wei and D. Stanley, "Change detection from remotely sensed images: From pixel-based to object-based approaches." *ISPRS Journal of photogrammetry and remote sensing* 80 (2013): 91-106.
[3] S. Sophiayati Yuhaniz, and T. Vladimirova. "An onboard automatic change detection system for disaster monitoring." *International Journal of Remote* Sensing 30.23 (2009): 6121-6139.
[4] V. Růžička, et al. "Unsupervised change detection of extreme events using ML On-board." *arXiv preprint arXiv:2111.02995* (2021).
[5] B. Chintalapati, et al. "Opportunities and challenges of on-board AI-based image recognition for small satellite Earth observation missions." *Advances in Space Research* (2024).
[6] R. C. Daudt, B. Le Saux and A. Boulch. "Fully convolutional siamese networks for change detection." *2018 25th IEEE international conference on image processing (ICIP)*. IEEE, (2018).
[7] D. Peng, Y. Zhang and H. Guan. "End-to-end change detection for high resolution satellite images using improved UNet++." *Remote Sensing* 11.11 (2019): 1382.
[8] H. Chen, Z. Qi and Z. Shi. "Remote sensing image change detection with transformers." *IEEE Transactions on Geoscience and Remote Sensing* 60 (2021): 1-14.
[9] J. Ballé, V. Laparra and E. P. Simoncelli, "End-to-end optimized image compression." *International Conference on Learning Representations (ICLR)*, 2017.
[10] C. Aybar, et al. "CloudSEN12, a global dataset for semantic understanding of cloud and cloud shadow in Sentinel-2." *Scientific data* 9.1 (2022): 782.
[11] R.C. Daudt, B. Le Saux, A. Boulch, and Y. Gousseau, 2018, July. "Urban Change Detection for Multispectral Earth Observation Using Convolutional Neural Networks." In *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)* 2018 (pp. 2115-2118). IEEE.
[12] Y. Ye, T. Tang, B. Zhu, C. Yang, B. Li and S. Hao, "A Multiscale Framework With Unsupervised Learning for Remote Sensing Image Registration," in *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1-15, (2022).
[13] C. Han, C. Wu, H. Guo, M. Hu, J. Li, and H. Chen, "Change guiding network: Incorporating change prior to guide change detection in remote sensing imagery." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, (2023).
[14] Chen, Keyan, et al. "Time travelling pixels: Bitemporal features integration with foundation model for remote sensing image change detection." *IGARSS 2024-2024 IEEE International Geoscience and Remote Sensing Symposium*. IEEE, (2024).
[15] R. C. Daudt, B. Le Saux, A. Boulch and Y. Gousseau, "Multitask learning for large-scale semantic change detection". *Computer Vision and Image Understanding*, *187*, 102783, (2019).