Towards Storage-Aware Onboard Change Detection

TOWARDS STORAGE-AWARE ONBOARD CHANGE DETECTION

Gabriele Inzerillo
Politecnico di Torino
Torino, Italy
gabriele.inzerillo@polito.it

Diego Valsesia
Politecnico di Torino
Torino, Italy
diego.valsesia@polito.it

Enrico Magli
Politecnico di Torino
Torino, Italy
enrico.magli@polito.it

Aniello Fiengo
European Space Agency
Noordwijk, Netherlands
aniello.fiengo@ext.esa.int

***Abstract*—Change Detection is typically performed at the ground segment of satellite systems, where images can be carefully post-processed and registered, and complexity is not an issue. We investigate a framework entirely based on neural networks to perform change detection directly onboard of satellites in order to minimize the latency of the detection pipeline. The proposed framework accounts for image compression, registration and detection of change in order to manage the limited storage of the satellite and the lack of image alignment between revisits, while minimizing the computational complexity of the design. Preliminary results on a lightweight design show good change detection performance as a function of compression rate, while managing geometric misalignment.**

***Index Terms*—Change detection, onboard processing, image compression.**

I. Introduction

Change Detection (CD) is a pivotal task in remote sensing, which seeks to identify significant evolutions of a scene from two or more images acquired at different time instants. While being the subject of intense research efforts [1, 2], most works focus on analyzing post-processed imagery after it has been downloaded to ground stations [3, 4]. This allows to have very few constraints on computational capabilities as well as access to data and side information needed for the CD process to be effective. However, performing CD at the ground segment introduces significant latency while posing many challenges for time-sensitive applications like natural disaster detection, where real-time analysis is critical.

A promising idea is to move CD directly onboard the satellite, significantly reducing latency and enabling near-real-time applications. Despite its potential, onboard processing presents challenges not addressed by current literature, especially if one wants to exploit the power of modern deep learning approaches. First, the limited computational capabilities of onboard hardware require the development of efficient CD algorithms. Additionally, CD requires storing older images as temporal reference, but onboard storage is limited and prioritized for data transmission, thus calling for efficient compression as well. Finally, a significant challenge for onboard change detection is the absence of orthorectification, a process that is not only highly complex to perform but also requires additional data, such as Digital Elevation Models (DEMs), which are typically available only on the ground, along with the co-registration of onboard images. These steps are routinely carried out at the ground segment, thanks to precise orthorectification processes with the availability of additional gelocation information, and the vast majority of the CD literature typically considers already orthorectified images. This is not the case onboard, where uncompensated geometric distortions and misalignments in images would lead to unreliable detection results.

In this paper, we take the first steps towards addressing these challenges. We propose a novel framework for onboard CD that uses a modular deep-learning architecture to tackle each of the aforementioned challenges. Importantly, this provides a single method that allows to incorporate image compression, co-registration and change detection in a single end-to-end model. In summary, our key contributions include:

- A modular deep architecture for CD comprising three sub-modules:
  - *Compression Network*: reduces onboard storage requirements by compressing images into compact latent representations.
  - *Registration Network*: aligns images through co-registration before change detection.
  - *Change Detection Network*: identifies differences between aligned images to produce the change map.
- A simulation of onboard scenarios by applying geometric transformations to orthorectified datasets, mimicking distortions and misalignments present in non-orthorectified and not co-registered images.
- An analysis of the trade-offs between compression rate and CD accuracy.

II. Background

A. Image Registration

Image registration is a critical preprocessing step in computer vision to align images representing the same scene or object, captured at different times, viewpoints, or conditions,

The activity was selected via the Open Space Innovation Platform (https://ideas.esa.int) as a Co-Sponsored Research Agreement/Study/Early Technology Development and carried out under the Discovery programme of, and funded by, the European Space Agency (contract number: 4000143591). The view expressed in this publication can in no way be taken to reflect the official opinion of the European Space Agency.

to enable meaningful comparisons. In change detection, registration geometrically aligns images so that the same pixel corresponds to the same spatial location enabling comparisons between different images.

Traditional methods rely on identifying keypoints (e.g., using SIFT [5], ORB [6]) and matching them to optimize a geometrical transformation. However, these approaches often require manual intervention or precise control points. Recent advances in deep learning have automated this process, enabling end-to-end image alignment with increased robustness and the ability to handle complex deformations [7–10].

In remote sensing, it is also common to perform orthorectification to compensate for viewing angles, but this can require side information such as digital elevation models, making it challenging for onboard processing. It is also worth noting that there is a lack of datasets of non-orthorectified imagery and the few images available are not sufficient to either investigate deep learning models or the change detection task. Widespread release of products such as Sentinel2 L1B images would be desirable to stimulate further research.

B. Change Detection

Change detection involves comparing images of the same geographical region taken at different times to detect alterations due to environmental changes, urban expansion, or natural disasters. Traditional change detection methods rely on pixel-based comparisons or feature extraction techniques, such as difference images or thresholding, but struggle with noise, misalignment and changes in pixel values brightness. Deep Learning models have significantly improved the performance through the usage of CNNs-based architecture [3] and also recent newer architectures based on transformers [11, 12]. Deep learning-based change detection methods have demonstrated state-of-the-art performance handling misalignments, noise, and complex spatial variations, making them ideal for remote sensing applications, where traditional methods fall short. Notably, Zhou *et al.* [13] demonstrated that image registration can benefit from change detection and viceversa, proposing a unified model to perform and improve both the tasks.

C. Onboard AI

The idea of performing AI-based processing onboard satellites has gained significant attention in recent years, primarily to address the challenges mentioned earlier: enabling real-time analysis, eliminating corrupted or unnecessary data, and minimizing data transfer to ground stations. These factors have driven the growing interest in Edge AI. Several works have demonstrated the feasibility of running AI-based processing onboard satellites [14–16]. Notably, Růžička *et al.* [17] developed an unsupervised model for change detection onboard satellites. However, their work did not account for image registration (assuming that images are already co-registered) or orthorectification, effectively creating a low-complexity model that does not address the geometric limitations of onboard images.

III. METHOD

In this section, we provide an overview of the proposed deep learning framework, as illustrated in Fig. 1, and present a preliminary design for the three modules which make up the entire framework. The framework follows a modular approach to the problem with clearly identifiable submodules addressing the tasks of compression, registration and change detection. The individual modules are first independently trained to solve their own respective tasks, and then jointly fine-tuned together.

A. Image Compression

The first part of the framework focuses on image compression, aiming to store compressed representations and save onboard storage, which is a critical constraint faced by onboard processing. For this step, a preliminary design based on the classic autoencoder architecture by Ballé *et al.* [18] is used. This architecture comprises four main components: 2D convolutions for downsampling, transpose convolutions for upsampling, generalized divisive normalization (GDN), and an entropy model. The entropy model, a core component of the architecture, is designed to learn the distribution of the input data, enabling the creation of a compact latent representation with a limited number of bits. While more recent compression models incorporate hyperprior mechanisms [19] to improve compression efficiency, we intentionally chose this simpler architecture to minimize complexity and computational requirements, while planning to further optimize it in future works.

The pretraining of the compression submodule follows a classical approach, where the objective is to minimize the rate-distortion Lagrangian cost. The loss function is expressed as:

$$\mathcal{L}_{\mathcal{C}} = \lambda \cdot R + D \tag{1}$$

where R represents the rate, which quantifies the number of bits required to encode the compressed representation, D is the distortion, measured as the mean squared error (MSE) between original and reconstructed image, and λ is a weighting factor that balances the tradeoff between rate and distortion. Multiple rate-distortion tradeoffs can be achieved by varying λ.

B. Image Co-Registration

The image co-registration submodule takes as input two images, $\mathbf{x}_{T_1}$ and $\mathbf{x}_{T_2}$, corresponding to time instants T_1 and T_2, after being decompressed by the image compression module. It is worth noting that due to lack of datasets with real non-orthorectified multitemporal imagery, we simulate geometric distortions by means of applying random affine and perspective transformations to the image pair. The geometric distortion model assumed by the registration submodule matches this assumption by directly estimating the parameters of a homography matrix.

We designed a lightweight registration model inspired by the work presented in [10]. The model is a hierarchical network consisting of three subnetworks, each operating on the input images at different scales: $1/4$, $1/2$, and full resolution.

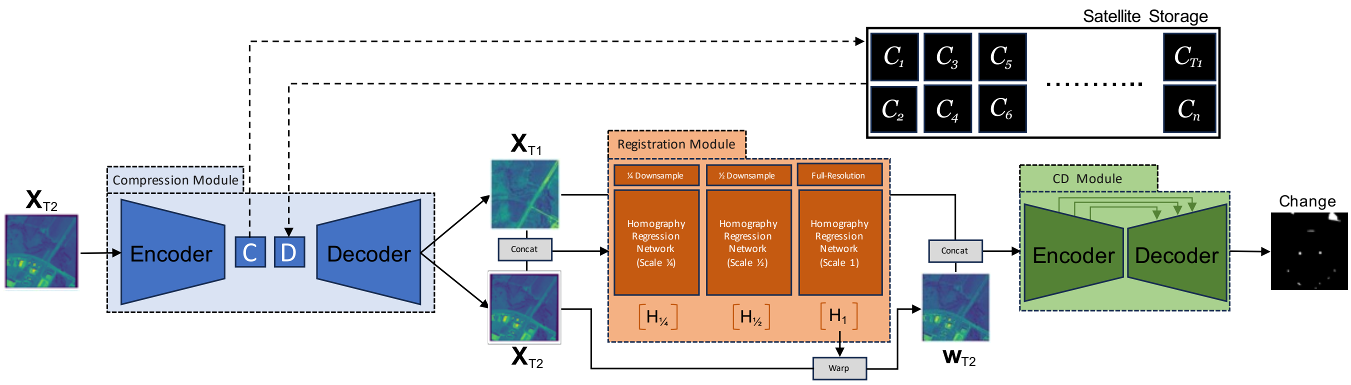

Fig. 1. The satellite storage contains multiple geotagged compressed latent representations (C_n) from previous orbits. When a new image (x_{T_2}) is acquired, the corresponding compressed latent representations from a previous revisit (C_{T_1}) is retrieved and decompressed (x_{T_1}). The newly acquired image is processed through a feature extractor to generate its feature representation. Both the reconstructed image from C_{T_1} and the current image x_{T_2} are then passed to the registration module; the registration module produces an homography matrix H_1 used to warp x_{T_2} over x_{T_1} producing the warped image w_{T_2}. Once the images are aligned, the change detection module is applied to identify differences between the two time points.

These subnetworks collaboratively estimate three homography matrices, one for each resolution, which are used to compute the total registration loss expressed as:

$$\mathcal{L}_{\mathcal{R}} = \alpha_1 \|\mathbf{x}_{T_1}^{(1)} - \mathbf{w}_{T_2}^{(1)}\|^2 + \alpha_2 \|\mathbf{x}_{T_1}^{(1/2)} - \mathbf{w}_{T_2}^{(1/2)}\|^2 + \alpha_3 \|\mathbf{x}_{T_1}^{(1/4)} - \mathbf{w}_{T_2}^{(1/4)}\|^2 \quad (2)$$

where $\mathbf{w}^{(s)}$ is the image $\mathbf{x}_{T_2}$ after warping it using the corresponding homography matrix at scale s, $\alpha_1, \alpha_2, \alpha_3$ are the weights balancing the importance of the three scales.

The homography matrix obtained from the full-resolution scale network is then applied to warp $\mathbf{x}_{T_2}$ (i.e., the image is subjected to affine and perspective transformations), aligning it with $\mathbf{x}_{T_1}$.

We experimented with two different configurations for pretraining the registration module. In the first configuration, referred to as *Solo-Reg Pretraining*, we simply minimize the registration loss presented in Equation 2. In the second configuration, named *Registered-Change Detection Pretraining* (Reg-CD from here on), we do not train the registration network in isolation. Instead, we jointly train both the registration and change detection models (which will be discussed in the next subsection) to explore how supervision on the change detection task might also improve the registration process. For the Reg-CD Pre-Training, the loss is a combination of the registration loss from Equation 2 and the change detection loss, which is defined as the negative log-likelihood between the predicted change map and the ground truth:

$$\mathcal{L} = \alpha \mathcal{L}_{CD} + (1 - \alpha)\mathcal{L}_R \quad (3)$$

being

$$\mathcal{L}_{\mathcal{CD}} = -\sum_i \left[y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i) \right] \quad (4)$$

where y_i (ground truth) and $\hat{y}_i$ (prediction) represent the true change label and the predicted probability of change at pixel i, respectively.

C. Change Detection

The network used as a submodule for change detection is an optimized version of the single-stream U-Net proposed in [3, 20]. In this version, transpose convolutions are replaced with bilinear interpolation to reduce the number of Floating Point Operations (FLOPs) and parameters, thereby decreasing the overall computational complexity.

We tested three different configurations for the change detection task:

- *Baseline*: in this configuration, we replicate the standard setting commonly used in the literature by using perfectly co-registered and orthorectified images. In this configuration, we train the module by minimizing only the loss defined in Eq. 4.
- *Solo-CD PreTrain*: in this configuration, we train the network to perform change detection starting from images that are not co-registered. We apply the aforementioned model of geometric distortion with an affine and perspective transformation. The pretrained (Solo-Reg pretraining) registration module is used to align the two images before processing them with the CD module. In this case, we only train the change detection model with Eq. 4 without updating the registration network.
- *Reg-CD PreTrain*: in this configuration, both the registration network and the change detection model are trained jointly according to Eq. 3.

IV. Experimental Results

All evaluations reported in this section are performed on the LevirCD dataset [21], one of the most widely used and popular benchmarks for change detection tasks. All images used for the experiments are three-channel RGB.

We first benchmark the performance of the registration and CD submodules when no image compression is performed in order to establish the performance of the low-complexity CD design against the state of the art [12] and evaluate the effect of image misregistration. Table I reports the results. First, the

TABLE I
CD PERFORMANCE WITHOUT COMPRESSION.

Training	IoU	F1-Score
ChangeFormer [12]	82.48	90.40
Baseline	81.62	88.95
Solo-CD	66.28	75.66
Reg-CD	73.31	82.36

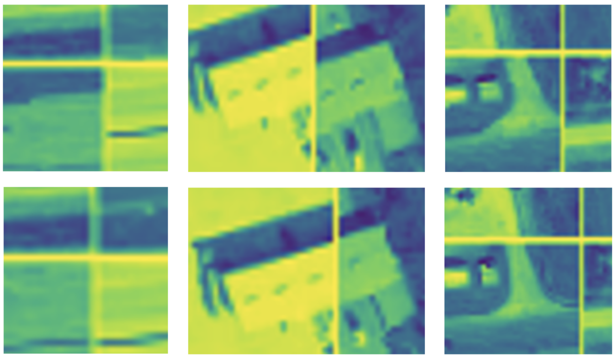

Fig. 2. Qualitative comparison of image registration using a chessboard made of alternated patches from a source image and the warped (registered) image. Top Row: registration performed by Solo-Reg pretrain; Bottom Row: registration performed by Reg-CD pretrain.

baseline configuration represents the theoretical upper limit of performance of the presented framework, as it utilizes perfectly aligned images directly from the dataset, as it is the most comparable to the results in the literature, with the exception of the use of a single band and the lower complexity. Second, the table shows that independently pretraining the change detection and registration networks (Solo-CD) proves suboptimal. Indeed, this pretraining approach leads to a registration module that often cannot compensate the full misregistration. Instead, the joint pretraining (Reg-CD) of registration and change detection allows to significantly improve performance. This can also be seen in Fig. 2 which shows how joint pretraining leads to more accurate image registration. From the performance of the Reg-CD method we can notice that misregistration and the need to correct it with the proposed lightweight approach costs about 6 percentage points of IoU or F1 score with respect to the baseline. This can be considered acceptable due to the complexity of the tasks and the design constraints around computational capabilities.

Finally, we present the results of the complete framework, including compression, in Fig. 3. Here, the performance on the CD task in terms of F1 score is benchmarked against storage requirements in terms of the rate needed by the compressed image representation. As expected, lossy compression degrades CD accuracy but the degradation seems limited until the breakdown rate of 0.5 bpp. Overall, for a practical rate around 0.5 bpp one can expect a CD F1 score equal approximately to 81.5, down for the 88.95 of the ideal baseline without misregistration and compression.

This is an interesting result showing preliminary evidence

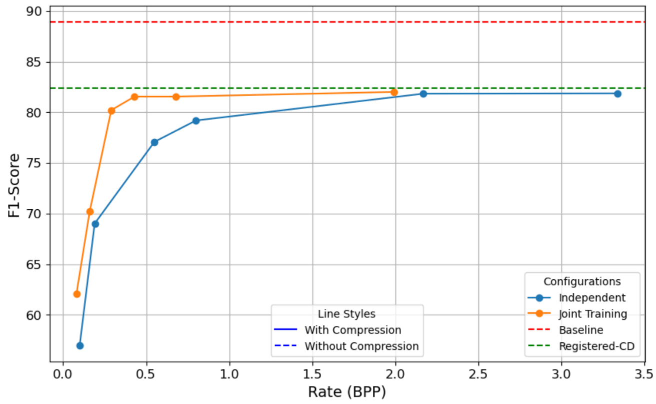

Fig. 3. Rate-F1 curve illustrating the trade-off between compression rate in bits-per-pixel (bpp) and F1-Score (change detection performance).

that an end-to-end framework entirely based on neural networks for onboard CD, accounting for both compression and image registration, can achieve good performance at a limited computational cost. Overall, the entire model presented in this paper requires approximately 550K FLOPs/pixel.

V. CONCLUSIONS

We presented a framework for onboard change detection entirely based on a modular neural network. The framework simultaneously addressed image compression, registration and change detection showing good performance on the CD task and a limited computational complexity. Further work will refine the design of the modules and introduce full end-to-end training of the entire system. We will also strive to further reduce the overall complexity of the method.

REFERENCES

[1] T. Bai, L. Wang, D. Yin, K. Sun, Y. Chen, W. Li, and D. Li, "Deep learning for change detection in remote sensing: a review," *Geo-spatial Information Science*, vol. 26, no. 3, pp. 262–288, 2023.

[2] A. Singh, "Review article digital change detection techniques using remotely-sensed data," *International journal of remote sensing*, vol. 10, no. 6, pp. 989–1003, 1989.

[3] R. C. Daudt, B. Le Saux, and A. Boulch, "Fully convolutional siamese networks for change detection," in *2018 25th IEEE international conference on image processing (ICIP)*. IEEE, 2018, pp. 4063–4067.

[4] B. Hou, Q. Liu, H. Wang, and Y. Wang, "From w-net to cdgan: Bitemporal change detection via deep learning techniques," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 3, pp. 1790–1802, 2020.

[5] D. Lowe, "Object recognition from local scale-invariant features," in *Proceedings of the Seventh IEEE International Conference on Computer Vision*, vol. 2, 1999, pp. 1150–1157 vol.2.

[6] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "Orb: An efficient alternative to sift or surf," in *2011 International Conference on Computer Vision*, 2011, pp. 2564–2571.

[7] P.-E. Sarlin, D. DeTone, T. Malisiewicz, and A. Rabinovich, "Superglue: Learning feature matching with graph neural networks," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 4938–4947.

[8] Y. Fu, Y. Lei, T. Wang, W. J. Curran, T. Liu, and X. Yang, "Deep learning in medical image registration: a review," *Physics in Medicine & Biology*, vol. 65, no. 20, p. 20TR01, 2020.

[9] S. Wang, D. Quan, X. Liang, M. Ning, Y. Guo, and L. Jiao, "A deep learning framework for remote sensing image registration," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 145, pp. 148–164, 2018.

[10] Y. Ye, T. Tang, B. Zhu, C. Yang, B. Li, and S. Hao, "A multiscale framework with unsupervised learning for remote sensing image registration," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–15, 2022.

[11] H. Chen, Z. Qi, and Z. Shi, "Remote sensing image change detection with transformers," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–14, 2021.

[12] W. G. C. Bandara and V. M. Patel, "A transformer-based siamese network for change detection," in *IGARSS 2022 - 2022 IEEE International Geoscience and Remote Sensing Symposium*, 2022, pp. 207–210.

[13] R. Zhou, D. Quan, S. Wang, C. Lv, X. Cao, J. Chanussot, Y. Li, and L. Jiao, "A unified deep learning network for remote sensing image registration and change detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 62, pp. 1–16, 2024.

[14] G. Inzerillo, D. Valsesia, and E. Magli, "Efficient onboard multi-task ai architecture based on self-supervised learning," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, pp. 1–12, 2024.

[15] G. Giuffrida, L. Fanucci, G. Meoni, M. Batič, L. Buckley, A. Dunne, C. van Dijk, M. Esposito, J. Hefele, N. Vercruyssen, G. Furano, M. Pastena, and J. Aschbacher, "The -sat-1 mission: The first on-board deep neural network demonstrator for satellite earth observation," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–14, 2022.

[16] S. Sophiayati Yuhaniz and T. Vladimirova, "An onboard automatic change detection system for disaster monitoring," *International Journal of Remote Sensing*, vol. 30, no. 23, pp. 6121–6139, 2009.

[17] V. Ržička, A. Vaughan, D. De Martini, J. Fulton, V. Salvatelli, C. Bridges, G. Mateo-Garcia, and V. Zantedeschi, "Ravæn: unsupervised change detection of extreme events using ml on-board satellites," *Scientific reports*, vol. 12, no. 1, p. 16939, 2022.

[18] J. Ballé, V. Laparra, and E. P. Simoncelli, "End-to-end optimized image compression," *arXiv preprint arXiv:1611.01704*, 2016.

[19] J. Ballé, D. Minnen, S. Singh, S. J. Hwang, and N. Johnston, "Variational image compression with a scale hyperprior," *arXiv preprint arXiv:1802.01436*, 2018.

[20] R. C. Daudt, B. Le Saux, A. Boulch, and Y. Gousseau, "Multi-task learning for large-scale semantic change detection," *Computer Vision and Image Understanding*, vol. 187, p. 102783, 2019.

[21] H. Chen and Z. Shi, "A spatial-temporal attention-based method and a new dataset for remote sensing image change detection," *Remote Sensing*, vol. 12, no. 10, p. 1662, 2020.