3DCD: A NEW DATASET FOR 2D AND 3D CHANGE DETECTION USING DEEP LEARNING TECHNIQUES

V. Coletta[1,*] V. Marsocci[2], R. Ravanelli[1]

[1] Geodesy and Geomatics Division (DICEA) - Sapienza University of Rome, Rome, Italy - (virginia.coletta, roberta.ravanelli)@uniroma1.it
[2] Department of Computer, Control and Management Engineering Antonio Ruberti (DIAG) - Sapienza University of Rome, Rome, Italy - valerio.marsocci@uniroma1.it

Commission III

KEY WORDS: 3D change detection, Dataset, 2D change detection, Remote Sensing, Deep learning

ABSTRACT:

Change detection is one of the main topics in Earth Observation, due to its wide range of applications, varying from urban development monitoring to natural disaster management. Most of the recently developed change detection methodologies rely on the use of deep learning algorithms. These kinds of algorithms are generally focused on generating two-dimensional (2D) change maps, thus they are only able to detect horizontal changes in land use/land cover, not considering nor returning any information on the corresponding elevation changes. Our work proposes a step forward, creating and sharing a dataset where two optical images acquired in different epochs are provided together with both the related 2D change maps containing land use/land cover variations and the three-dimensional (3D) maps containing elevation changes. Particularly, our aim is to provide a dataset useful to address and possibly solve the change detection task in 3D. Indeed, the proposed dataset, on the one hand, can empower a further development of 2D change detection algorithms, and, on the other hand, can allow to develop algorithms able to provide 3D change detection maps from two optical images captured in different epochs, without the need to rely directly on elevation data as input. The proposed dataset is publicly available at the following link: `https://bit.ly/3wDdo41`.

1. INTRODUCTION

Change detection (CD), namely the task of identifying areas of the Earth's surface that have experienced changes by jointly analysing two or more coregistered images captured at different epochs (Bruzzone and Bovolo, 2013, Daudt et al., 2019), is one of the main topics in Earth Observation (EO). CD algorithms allow the assessment of changes that have occurred at ground level and, for this reason, are applied to various real-world problems, such as natural disaster management (Chen et al., 2018), identification of urban changes (Lyu et al., 2018, Huang et al., 2013), crops and forests management (Khan et al., 2017), etc. Nowadays, thanks to the unprecedented technological development of EO sensors, it is possible to observe more and more details at ground level and, consequently, to retrieve even more precise change maps. Several scientific studies have led to the development of algorithms capable of meeting the requirements of CD applications; however, these algorithms are generally focused on the generation of 2D change maps, thus they are only capable of assessing land use and land cover (LULC) variations. In this context, deep learning (DL) is recently overtaking traditional methods, proving to be a viable solution to tasks such as semantic segmentation, object and change detection (Zhu et al., 2017, Ma et al., 2019). In particular, several DL architectures are currently employed for solving CD tasks, such as CDNet (Alcantarilla et al., 2018) or Siamese networks (Daudt et al., 2018a). Nevertheless, the number of CD labelled remote sensing (RS) datasets openly available are still limited, due to their high production costs, requiring a lot of time and specialised knowledge in the annotation of hundreds of images (Marsocci et al., 2021). This scarce dataset availability represents thus an obstacle to the implementation and testing of new methods and the development of effective solutions to the aforementioned CD tasks (Daudt et al., 2019). Moreover, to the best of our knowledge, there are currently no labelled datasets that provide information about the elevation changes corresponding to the LULC variations observable in a pair of optical images captured in different epochs (Qin et al., 2016, Zhu et al., 2017, Ma et al., 2019). One interesting research line of DL applied to RS is moving precisely towards this direction: to infer 3D information from the smallest possible amount of 2D information, usually extracted from optical images (Mou and Zhu, 2018). Indeed, 2D CD algorithms can only detect planimetric changes such as appearing/disappearing buildings/trees, shrinking/expanding structures; these results do not suffice to fulfil the requirements of applications needing also the vertical/volumetric information, such as quantitative estimation of landslides volume, tree growth and building construction progress monitoring. 3D CD algorithms can hence offer numerous advantages.

A critical review of the last developments and applications of 3D CD using RS and close-range data is given in (Qin et al., 2016). More recently, (Okyay et al., 2019) carried out a survey of the airborne Light Detection and Ranging (LiDAR) CD methods currently employed in Earth science applications. Finally, (Shirowzhan et al., 2019, de Gélis et al., 2021) are the only available studies which have proposed a comparative analysis for 3D CD. In the first (Shirowzhan et al., 2019), five methods were compared with respect to two criteria: the ability to detect demolished and new buildings, and the capability to provide information about the magnitude of the changes. However, no quantitative results were provided, and the evaluation was carried out on a private dataset (de Gélis et al., 2021). Moreover, this study did not employ any DL algorithms. On

---

* Corresponding author

the other hand, (de Gélis et al., 2021) developed an original simulator of multi-temporal aerial LiDAR urban point clouds. The simulator was used to automatically build an annotated 3D CD dataset consisting of pairs of 3D point clouds labelled according to the synthetic changes imposed by the authors. Six different 3D CD methods were assessed either by directly coping with the 3D point clouds or using the Digital Surface Models (DSMs) generated from their rasterization (de Gélis et al., 2021). In particular, the authors compared traditional methods, such as the use of different types of thresholding and filtering algorithms on DSMs, with both a machine learning algorithm (a random forest fed with hand-crafted features) and two DL networks (consisting in a feed-forward network (FFN) and a Siamese network).

Our work, however, goes one step further by creating and sharing a dataset in which a 3D CD map, i.e. a map containing the change in elevation, is provided together with the 2D CD map and the corresponding pair of optical images. The main contribution behind the production of this dataset is to allow the development of DL algorithms that can automatically generate 3D CD maps using two aerial or satellite optical images acquired in different epochs as input, without the need of DSMs, as it will be highlighted in Section 3.

2. RELATED DATASETS

As already pointed out in the introduction (Section 1), to design and build the proposed dataset, we considered the existing literature and, in particular, the open EO CD datasets already available to the scientific community. Particularly, we analysed the main features of the open CD datasets explicitly designed for the development of DL algorithms and containing optical images, annotated with 2D change maps, and/or LiDAR point clouds, from which 3D changes can be deduced. In general, there are many datasets that contain optical images, and thus suited to perform 2D CD tasks, less that include LiDAR PCs. Concerning this issue, the annotated 3D CD dataset released by (de Gélis et al., 2021) – built through a simulator that introduces synthetic changes to LiDAR point clouds – could be an effective solution. However, at the best of our knowledge, no CD dataset containing both optical images, 2D CD maps and information about the corresponding elevation changes is currently available.

Among the 2D CD datasets designed for the development of DL algorithms, the SZTAKI Air change benchmark (Benedek and Szirányi, 2009, Benedek and Szirányi, 2008) was one of the first openly available and it is currently one of the most used in the RS community. It consists of 13 aerial image pairs, provided by the Hungarian Institute of Geodesy, Cartography and Remote Sensing or retrieved from Google Earth. The images have a spatial resolution of 1.5 m and each pixel is labelled as subjected to change or not.

The SEmantic Change detectiON Dataset (SECOND) (Yang et al., 2021) is a pixel-level annotated semantic CD dataset, which includes 4662 pairs of aerial images, with a size of 512×512 pixels, acquired from different platforms and sensors, covering three Chinese cities. It is annotated with 6 LULC classes: (i) non-vegetated area, (ii) trees, (iii) low vegetation, (iv) water, (v) buildings and (vi) playing fields.

The Onera Satellite Change Detection (OSCD) (Daudt et al., 2018b) is a dataset composed of 24 multispectral aerial image pairs acquired by Sentinel-2, manually annotated as subjected to change or not at pixel-level.

The Deeply Supervised Image Fusion Network (DSIFN)

(Zhang et al., 2020) is a DL method, proposed along with a dedicated dataset for the validation task. It is composed of 6 high resolution bi-temporal images, extracted from Google Earth. Specifically, it is made of 3600 image tile pairs for training, 340 for validation and 48 for testing. All the image tiles are characterised by a size of 512×512 pixels.

The Sentinel-2 Multitemporal Cities Pairs (S2MTCP) (Leenstra et al., 2021) consists of 1520 pairs of Sentinel-2, level 1C, images covering different urban areas around the world, with a spatial resolution of 10 m and a size of 600×600 pixels. This dataset was originally used in the paper for the self-supervised training step. The trained model was then validated on the aforementioned OSCD (Daudt et al., 2018b).

The Sun Yat-Sen University Change Detection (SYSU-CD) (Shi et al., 2021) dataset was built to validate a deeply supervised (DS) attention metric-based network (DSAMNet). It consists of 20000 pairs of 0.5 m aerial images of size 256×256 captured between 2007 and 2014 in Hong Kong. The dataset is annotated with six classes of LULC changes: (i) new urban buildings; (ii) suburban expansion; (iii) pre-construction earthworks; (iv) vegetation change; (v) road expansion; (vi) sea construction.

The S2Looking (Shen et al., 2021) is a building CD dataset, which consists of 5000 recorded bi-temporal image pairs of rural areas worldwide and more than 65,920 annotated change instances, indicating separately newly constructed and demolished buildings. The images are characterised by a size of 1024×1024 pixels, with a spatial resolution ranging from 0.5 to 0.8 m/pixel.

Finally, the dataset provided in (Lebedev et al., 2018) is a synthetic database containing 12,000 triples of synthetic images without object shift, 12,000 triples of model images with object shift and 16,000 triples of real RS image fragments.

3. DATASET DESCRIPTION

The proposed dataset covers the urban area of the city of Valladolid in Spain (Figure 1). The area of interest includes the historical and urban centre of the city and the surrounding commercial areas. The agricultural areas were not considered since no significant changes in elevation were found for these areas. Moreover, we selected and annotated only the changes affecting artificial manufacts, such as the construction and the demolition of buildings.

Figure 1. Area included in the dataset

In particular, the dataset contains 472 (i) pairs of images cropped from optical orthophotos acquired through two different aerial surveys, performed respctively in 2010 and in 2017, (ii) the corresponding LULC variation maps in raster format, i.e. the 2D CD maps, and (iii) the corresponding elevation vari-

ation maps in raster format, namely the 3D CD maps. The images contain three bands corresponding to the Red, Green and Blue channels. The main features of the data contained in the proposed dataset are described in Table 1.

To build the dataset, we started from several pairs of aerial orthophotos freely available in the website of the (Organismo Autónomo Centro Nacional de Información Geográfica, 2021), acquired in 2010 and in 2017, and covering the area of Valladolid. The original orthophotos are characterised by a Ground Sample Distance (GSD) of 0.25 m. To produce the DSMs needed for the generation of the 3D CD maps, we exploited the LiDAR data freely available in the website of (Organismo Autónomo Centro Nacional de Información Geográfica, 2021) for the same years and the same area. The DSMs were produced within QGIS (QGIS, 2022b), by rasterizing the original point clouds contained in the LAS files. The GSD of the DSMs is 1 m.

The first step of the dataset preparation was an automatic pre-processing phase, in which the images, both the optical ones and the corresponding DSMs, were cropped into smaller tiles covering a size of 200 m × 200 m. In addition, to make the GSD of the orthophotos more similar to the GSD of the DSMs, the orthophotos were downsampled, degrading their GSD from 0.25 m to 0.50 m. At the end of this operation, 472 pairs of orthophotos with a size of 400×400 pixels were produced together with the corresponding pairs of DSMs with a size of 200×200 pixels. An example is shown in Figure 2.
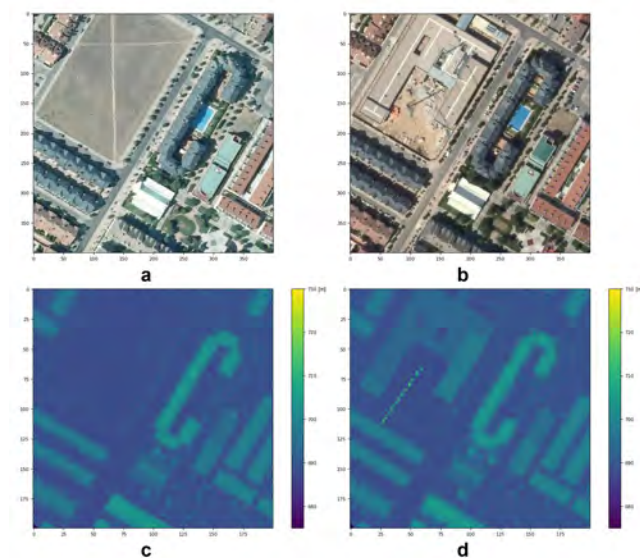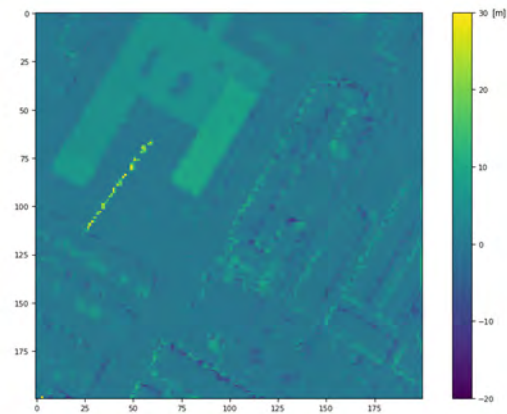

Figure 3. 3D change detection map obtained through difference of the DSMs. The colour bar is expressed in meters

only the pixels affected by a real change in elevation were considered, while the pixels in which no real change had occurred, and thus containing only noise, were set equal to zero. An example can be observed in Figure 4. A further check was carried out to assess the absence of coregistration errors, both in the orthophotos and in the DSMs.

Figure 2. Example of the data used to produce the dataset; a: Ortophoto referring to 2010; b: Ortophoto referring to 2017; c: DSM referring to 2010; d: DSM referring to 2017

After the aforementioned automatic phase of tile cropping, the raster 3D CD maps containing the elevation changes were generated through a simple difference between the DSMs (Figure 3):

$$\Delta H = H_{2017} - H_{2010} \qquad (1)$$

In particular, we considered all the elevation changes characterised by values lower than one metre in absolute value as negligible with respect to the entity of elevation variations usually affecting buildings, and for this reason they were ignored (i.e. their value was set to zero). Then, a manual control step was carried out on the resulting 3D CD maps: by means of a visual comparison with the corresponding pair of orthophotos,
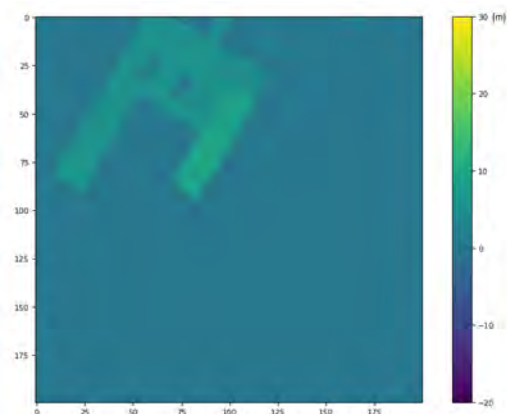

Figure 4. 3D change detection map obtained through difference of the DSMs (see Figure 3), after the manual removal of the noise. The colour bar is expressed in meters

Thus, for each pair of optical images, two CD maps were produced, focusing on the changes affecting only artificial manufacts (i.e. buildings, Figure 4). The first one is the 2D CD map, in which we annotated the pixels belonging to areas where a change in elevation occurred. These maps were constructed using the software QGis (QGIS, 2022a), taking the 400×400 orthophotos as reference and comparing them with the DSMs difference maps. Then, pixels belonging to the areas affected by a change in elevation over the years were delineated. In particular, two classes were identified: (i) no change; (ii) changes due to construction (positive elevation change) or demolition (negative elevation change) of artefacts/buildings. The 2D CD maps are characterised by the same resolution (400×400 pixels) of the orthophotos from which they derive, with a GSD of 0.50 m (Figure 5).

The second CD map is the 3D CD map (Figure 4), obtained from the difference between the DSMs as aforementioned, with a resolution of 200×200 pixels and a GSD of 1 m.
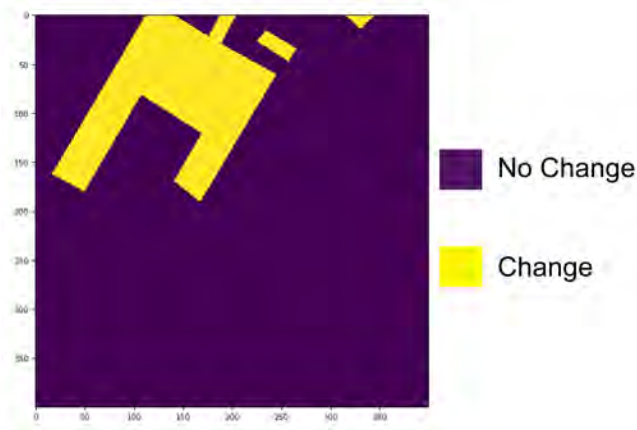
Figure 5. 2D change detection map

Once the dataset was produced, the 472 quadruplets of images (Figure 9: two ortophoto tiles – one for the 2010 and one for the 2017 –, one 2D CD map and one 3D CD map) were divided into *train*, *test* and *validation* (val) subsets to permit their direct use for benchmarking, hence avoiding reproducibility issues potentially deriving from a random split of the dataset. Specifically, the division was carried out to ensure that the percentage of pixels with and without variations was similar in all the three subsets (Table 2), assuring also that the images included in the train subset contained pixels with all the elevation variation values (ranging from -25 m to 35 m, Figure 6). In particular, the train subset contains 320 images ($\sim 68\%$), the test subset 110 images ($\sim 23\%$) and the validation subset 42 images ($\sim 9\%$). Finally, Table 2 shows the percentages, averaged over all the images contained in each subset, of the pixels affected by change and the pixels where there was no change over the years. To sum up, two full examples of the proposed dataset are shown in Figure 9.

Table 1. Main features of the data contained in the proposed dataset

Image	Number of pixels	GSD
Orthophotos from 2010	400×400	0.50 m
Orthophotos from 2017	400×400	0.50 m
2D CD maps	400×400	0.50 m
3D CD maps	200×200	1.0 m

Table 2. Percentage of pixels characterised by 'change' and 'no change' for each subset

Subset	Number of Images	Change	No change
Train	320	5%	95%
Test	110	4%	96%
Val	42	5%	95%

Figures 6, 7, 8 show the pixel elevation distribution for the three subsets.

Two different kinds of algorithms can be applied to the proposed dataset. On one hand, the well known 2D CD algorithms, such as CDNet, Fully Convolutional Network-CD (FCN-CD) (Alcantarilla et al., 2018) and Fully Convolutional Neural Network (FCNN) (Daudt et al., 2018a), could be used to provide 2D baseline benchmark metrics, with respect to other datasets and methods available in literature. Furthermore, innovative 3D CD algorithms could be developed to retrieve, automatically from the two optical images, 3D CD maps without the need of relying on elevation data as input. To conclude, in the next section (Section 4), some details about
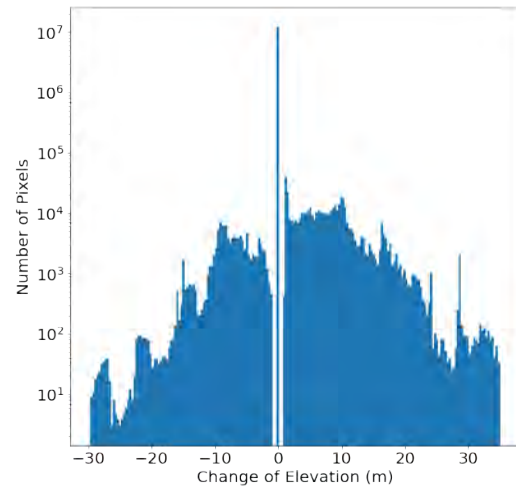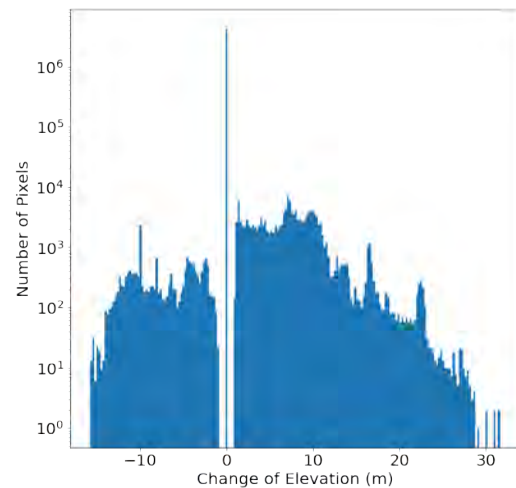

Figure 6. Pixel distribution in train
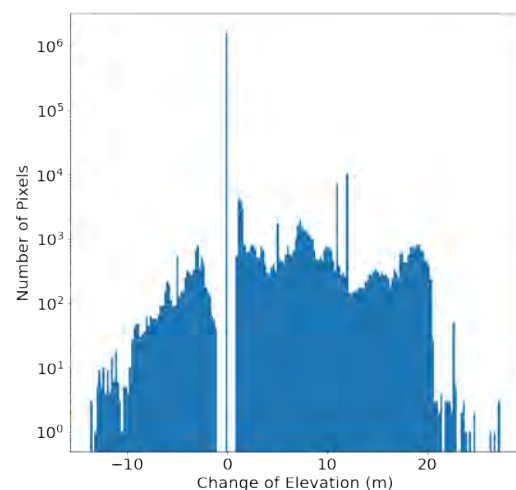

Figure 7. Pixel distribution in test

Figure 8. Pixel distribution in val

the research line we are following and the algorithms that are being developed are provided.
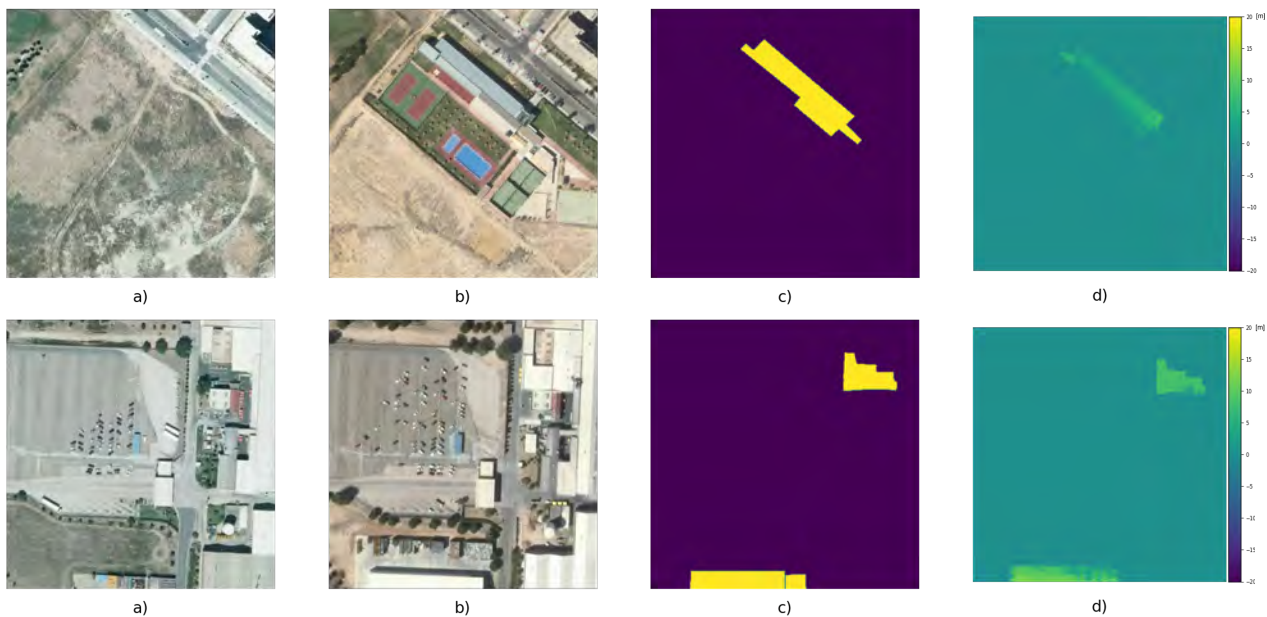
Figure 9. Two examples from the proposed dataset. a) 2010 optical image, b) 2017 optical image, c) 2D CD map, d) 3D CD map, where the colour bar is expressed in meters

4. FURTHER DEVELOPMENTS

To complete the proposed dataset, we are currently developing baseline algorithms that can solve the 3D CD task. In particular, we are considering some families of models that can approach the 3D CD task simultaneously with the 2D CD task. This strategy, in fact, would allow to output two masks containing a more complete information, useful for different RS applications, such as those reported in Section 1. In particular, we are testing models based on a Siamese U-Net network, similar to the one developed in (Alcantarilla et al., 2018). However, as previously stated, the models we are developing will differ from the above-mentioned models, as the loss will be composed of two terms: one classification term (eg.: cross entropy to solve the 2D CD task) and one regression term (eg.: mean squared error to solve the 3D CD task). Finally, an attention-based model (Vaswani et al., 2017) is under development as well, given the effectiveness of such family of models also for RS CD applications (Bandara and Patel, 2022). Moreover, we are already considering to integrate the dataset with new pairs of optical images, accompanied by the respective 2D and 3D change masks, on areas already identified and subjected to elevation variations. In conclusion, 3D CD is one of the main topics in the field of DL applied to RS, and the availability of open datasets, such as the one proposed in this work, is essential to develop and to validate algorithms able to solve this challenging task. With this contribution, we aim to show the robustness and effectiveness of the proposed dataset, emphasising its construction and validation process. Moreover, in parallel to the release of the dataset, we are developing models able to solve the 3D CD task in addition to the well studied 2D CD task. In order to support further researches on 3D CD, especially with DL methods, the dataset is publicly available at the following link: `https://bit.ly/3wDdo41`.

REFERENCES

Alcantarilla, P. F., Stent, S., Ros, G., Arroyo, R., Gherardi, R., 2018. Street-view change detection with deconvolutional networks. *Autonomous Robots*, 42(7), 1301–1322.

Bandara, W. G. C., Patel, V. M., 2022. A Transformer-Based Siamese Network for Change Detection. *CoRR*, abs/2201.01293. `https://arxiv.org/abs/2201.01293`.

Benedek, C., Szirányi, T., 2008. A mixed markov model for change detection in aerial photos with large time differences. *2008 19th International Conference on Pattern Recognition*, IEEE, 1–4.

Benedek, C., Szirányi, T., 2009. Change detection in optical aerial images by a multilayer conditional mixed Markov model. *IEEE Transactions on Geoscience and Remote Sensing*, 47(10), 3416–3430.

Bruzzone, L., Bovolo, F., 2013. A Novel Framework for the Design of Change-Detection Systems for Very-High-Resolution Remote Sensing Images. *Proceedings of the IEEE*, 101(3), 609-630.

Chen, Z., Zhang, Y., Ouyang, C., Zhang, F., Ma, J., 2018. Automated Landslides Detection for Mountain Cities Using Multi-Temporal Remote Sensing Imagery. *Sensors*, 18(3). https://www.mdpi.com/1424-8220/18/3/821.

Daudt, R. C., Le Saux, B., Boulch, A., 2018a. Fully convolutional siamese networks for change detection. *2018 25th IEEE International Conference on Image Processing (ICIP)*, IEEE, 4063–4067.

Daudt, R. C., Le Saux, B., Boulch, A., Gousseau, Y., 2018b. Urban change detection for multispectral earth observation using convolutional neural networks. *IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium*, IEEE, 2115–2118.

Daudt, R. C., Le Saux, B., Boulch, A., Gousseau, Y., 2019. Multitask learning for large-scale semantic change detection. *Computer Vision and Image Understanding*, 187, 102783.

de Gélis, I., Lefèvre, S., Corpetti, T., 2021. Change Detection in Urban Point Clouds: An Experimental Comparison with Simulated 3D Datasets. *Remote Sensing*, 13(13). `https://www.mdpi.com/2072-4292/13/13/2629`.

Huang, X., Zhang, L., Zhu, T., 2013. Building change detection from multitemporal high-resolution remotely sensed images based on a morphological building index. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 7(1), 105–115.

Khan, S. H., He, X., Porikli, F., Bennamoun, M., 2017. Forest Change Detection in Incomplete Satellite Images With Deep Neural Networks. *IEEE Transactions on Geoscience and Remote Sensing*, 55(9), 5407-5423.

Lebedev, M., Vizilter, Y. V., Vygolov, O., Knyaz, V., Rubis, A. Y., 2018. CHANGE DETECTION IN REMOTE SENSING IMAGES USING CONDITIONAL ADVERSARIAL NETWORKS. *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*, 42(2).

Leenstra, M., Marcos, D., Bovolo, F., Tuia, D., 2021. Self-supervised pre-training enhances change detection in sentinel-2 imagery. *International Conference on Pattern Recognition*, Springer, 578–590.

Lyu, H., Lu, H., Mou, L., Li, W., Wright, J., Li, X., Li, X., Zhu, X. X., Wang, J., Yu, L., Gong, P., 2018. Long-Term Annual Mapping of Four Cities on Different Continents by Applying a Deep Information Learning Method to Landsat Data. *Remote Sensing*, 10(3). `https://www.mdpi.com/2072-4292/10/3/471`.

Ma, L., Liu, Y., Zhang, X., Ye, Y., Yin, G., Johnson, B. A., 2019. Deep learning in remote sensing applications: A meta-analysis and review. *ISPRS journal of photogrammetry and remote sensing*, 152, 166–177.

Marsocci, V., Scardapane, S., Komodakis, N., 2021. MARE: Self-Supervised Multi-Attention REsu-Net for Semantic Segmentation in Remote Sensing. *Remote Sensing*, 13(16), 3275. `https://www.mdpi.com/2072-4292/13/16/3275`.

Mou, L., Zhu, X. X., 2018. IM2HEIGHT: Height estimation from single monocular imagery via fully residual convolutional-deconvolutional network. *arXiv preprint arXiv:1802.10249*.

Okyay, U., Telling, J., Glennie, C. L., Dietrich, W. E., 2019. Airborne lidar change detection: An overview of Earth sciences applications. *Earth-Science Reviews*, 198, 102929.

Organismo Autónomo Centro Nacional de Información Geográfica, 2021. Digital elevation models and maps in image format. `http://centrodedescargas.cnig.es/CentroDescargas/buscadorCatalogo.do?codFamilia=LIDAR#`.

QGIS, 2022a. Qgis documentation. `https://docs.qgis.org/3.22/en/docs/index.html`.

QGIS, 2022b. QGIS documentation: DEM from LiDAR Data. `https://docs.qgis.org/3.22/en/docs/training_manual/forestry/basic_lidar.html`.

Qin, R., Tian, J., Reinartz, P., 2016. 3D change detection–approaches and applications. *ISPRS Journal of Photogrammetry and Remote Sensing*, 122, 41–56.

Shen, L., Lu, Y., Chen, H., Wei, H., Xie, D., Yue, J., Chen, R., Lv, S., Jiang, B., 2021. S2Looking: A Satellite Side-Looking Dataset for Building Change Detection. *Remote Sensing*, 13(24), 5094.

Shi, Q., Liu, M., Li, S., Liu, X., Wang, F., Zhang, L., 2021. A deeply supervised attention metric-based network and an open aerial image dataset for remote sensing change detection. *IEEE Transactions on Geoscience and Remote Sensing*.

Shirowzhan, S., Sepasgozar, S. M., Li, H., Trinder, J., Tang, P., 2019. Comparative analysis of machine learning and point-based algorithms for detecting 3D changes in buildings over time using bi-temporal lidar data. *Automation in Construction*, 105, 102841.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., Polosukhin, I., 2017. Attention is all you need. *Advances in neural information processing systems*, 5998–6008.

Yang, K., Xia, G.-S., Liu, Z., Du, B., Yang, W., Pelillo, M., Zhang, L., 2021. Asymmetric Siamese Networks for Semantic Change Detection in Aerial Images. *IEEE Transactions on Geoscience and Remote Sensing*.

Zhang, C., Yue, P., Tapete, D., Jiang, L., Shangguan, B., Huang, L., Liu, G., 2020. A deeply supervised image fusion network for change detection in high resolution bi-temporal remote sensing images. *ISPRS Journal of Photogrammetry and Remote Sensing*, 166, 183–200.

Zhu, X. X., Tuia, D., Mou, L., Xia, G.-S., Zhang, L., Xu, F., Fraundorfer, F., 2017. Deep learning in remote sensing: A comprehensive review and list of resources. *IEEE Geoscience and Remote Sensing Magazine*, 5(4), 8–36.