

Supervised Deep-Learning Models for Multispectral Change Detection

Subjects: Remote Sensing

Contributor: Eleonora Jonasova Parelius

Remote sensing is a tool of interest for a large variety of applications. It is becoming increasingly more useful with the growing amount of available remote sensing data. The large amount of data also leads to a need for improved automated analysis. Deep learning is a natural candidate for solving this need. Change detection in remote sensing is a rapidly evolving area of interest that is relevant for a number of fields. There are a large number of publications and progress, even though the challenge is far from solved. Multispectral images are common in remote sensing and well suited for change detection purposes thanks to their good balance of spatial and spectral resolution, capturing important details. Various data sets for change detection have been openly published online and a growing number of models has been introduced in the literature, both supervised, unsupervised and semi-supervised.

Keywords: change detection ; remote sensing ; optical imaging ; multispectral images ; deep learning

1. Introduction

Remote sensing (RS) denotes the acquisition of information about an object from a distance. Often, as will also be the case here, the term is used more specifically to refer to the imaging of the Earth's surface from above, such as from a satellite or an aircraft. Technological development has led to an unprecedented amount of RS imagery being available today. The information in these images is of interest for a number of fields and applications, such as cartography ^[1], agriculture ^{[2][3]}, nature conservation ^{[4][5]}, climate ^{[6][7]} and disaster monitoring ^{[8][9]}, archaeology ^{[10][11]}, law enforcement ^[12] and urban planning ^[13].

The amount of information provided by RS poses a challenge in filtering out the relevant data. Manual exploration of the images is slow and laborious, and most applications, thus, require methods for the efficient processing of RS imagery. One of the tasks common to practically all fields where RS is used is change detection (CD)—or, more accurately stated: relevant change detection.

Change detection in the context of remote sensing refers to the process of identifying differences in the structure and/or properties of objects and phenomena on Earth by analysing two or more images taken at different times ^[14]. Change detection can serve as a basis for understanding the development of various natural or human-related phenomena through time and the interactions between them.

The goal of change detection is usually to identify the pixels within the (two or more) images that correspond to changed objects on the ground. It is, however, also possible to work at the level of a scene, i.e., to identify whether the classification of the depicted scene has changed (for example, a field turning into a residential area). Some methods also seek not only to identify changed pixels (or scenes) but also to classify the type of change that has occurred, referred to as semantic change detection.

The main challenge of change detection lies in the identification of changes that are relevant for the given task. Observed changes can be divided into three categories—apparent, irrelevant and relevant changes. The first category—apparent changes—comprises the changes seen in the images that do not result from actual changes happening to the depicted objects. Instead, changes resulting from variations in imaging equipment and circumstances, such as light and atmospheric conditions, belong in this category.

The other two categories—relevant and irrelevant changes—include all the real changes that are happening to the observed objects. The boundary between relevant and irrelevant changes is entirely dependent on the application. For instance, the relevance of seasonal changes, such as snow cover or the state of the foliage on the vegetation, is determined by the specifics of the task.

While snow cover is unimportant in urban planning and might be considered an irrelevant change, it is highly relevant for assessing the state of glaciers. Similarly, the level of vegetation hydration is of no interest in cartography but is essential in the monitoring of draughts or in crop assessment. The relevance of human-made changes is also determined by the specifics of the task, e.g., the presence or absence of vehicles is irrelevant in archaeology or cartography but is important for activity monitoring.

Deep learning is a subset of machine learning. It uses neural networks consisting of a large number of layers (hence, the name deep) to learn and represent data.

With the development and growing popularity of deep-learning methods within computer vision, it is natural to also apply them to the problem of CD in remote sensing. Deep-learning models are able to represent complex and hierarchical features within the data, which makes them good candidates for RS CD.

One of the biggest challenges of change detection is the presence of apparent changes. Every pixel from the image taken at an earlier time can be changed in the image taken at a later time, without there being any changes to the objects depicted in them. This poses a challenge for the more classical algorithms (algebraic, statistic and transformation-based), which are more rigid and not able to represent complex features, thus, leading to many false positives and a need to fine-tune the detection threshold for each application. Due to the flexibility, scalability and hierarchical structure of deep-learning models, they have the potential to learn to represent data that are too complex to be described by simpler models. Each pixel can thus be considered within the context of its surrounding pixels, as higher-level features are used for decision making.

Classical approaches to multispectral image exploitation often rely on spectral indices which combine various bands in order to emphasize the desired properties. Choosing the right index or combination of bands is a task requiring expert knowledge, and, in the end, only a portion of the available information (the information contained in the selected bands) is used. Deep learning allows the use of all the available bands without the need for expert-led pre-selection. The importance and contribution of each band is simply learned by the model in the training process.

Deep-learning change-detection models can be broadly divided into two categories: fully supervised models and models that are not fully supervised, i.e., semi-supervised and unsupervised. Fully supervised methods almost always require a large amount of labelled data in order to train the network, while semi- and unsupervised methods reduce or eliminate the need for ground-truth-labelled data.

2. Supervised Deep-Learning Models for Multispectral Change Detection

Supervised change-detection methods require annotated data in order to train the network. Unlike many of the staple tasks in computer vision (such as classification or semantic segmentation), change detection receives as input two (or more) images, rather than a single one, along with a single ground truth image. The two input images can be processed in various ways, which can be roughly divided into single- and double-stream structures.

In the single-stream structure, also referred to as early fusion, the two images are joined before they are fed to the network, either by simple concatenation or other procedures, such as differencing.

The alternative—double-stream structure—is based on processing each of the images on its own before they are again joined together and compared. Double-stream architectures (depicted in **Figure 1c**) feature two identical subnetworks that run parallel to each other, with each one taking one of the images as input. This type of structure is usually referred to as Siamese, if the subnetworks share weights, and pseudo-Siamese, if they do not share weights.

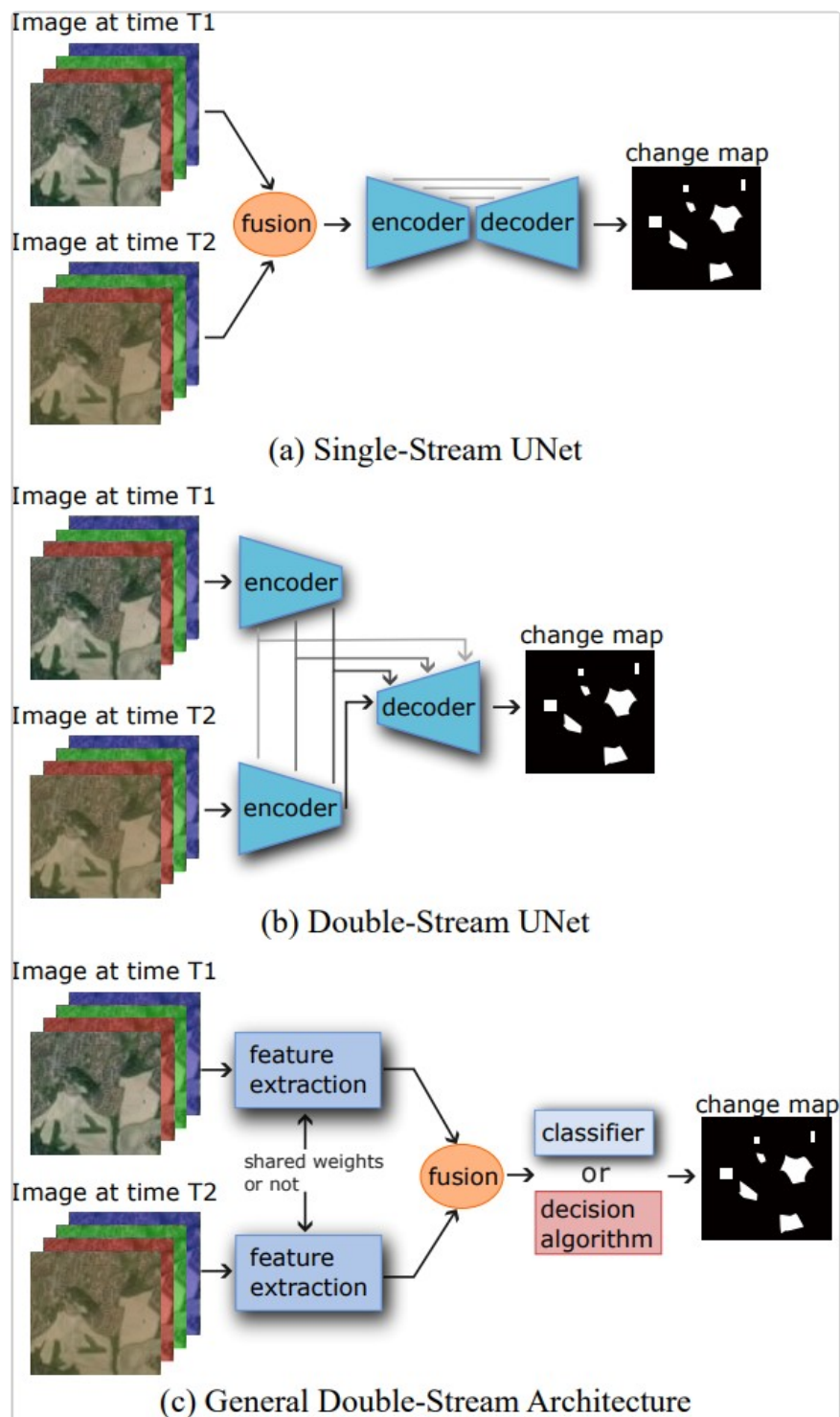


Figure 1: Structures of supervised CD models. (a) Single-Stream UNet-like network. The image data from two time points is first fused together, usually by concatenation, and then input into a UNet-like network featuring an encoder, a decoder and skip connections.

for semantic segmentation. The output of the model is a change map. (b) Double-Stream UNet. Each images is input separately into the encoder. The output of the two encoders is fused on multiple levels and fed to a single decoder, which produces a change map. (c) General Siamese Feature Extraction-based network structure. The individual images are first input into two identical Siamese (shared-weights) or pseudo-Siamese (different weights) subnetworks for feature extractions. The extracted features are then fused together, and either an automatic decision model or a machine-learning-based classifier is then used to produce a change map.

3. Semi-supervised and Unsupervised Deep-Learning Models for Multispectral Change Detection

One of the biggest challenges within deep learning for change detection is the availability of large annotated datasets. A possible way to deal with this issue lies in reducing or eliminating the need for the annotations, since unannotated data are abundant. Semi-supervised networks aim to reduce the amount of annotations needed, while unsupervised networks do not require any annotated data at all.

It can often be easier to achieve good results on end-to-end trained supervised networks compared to semi- and unsupervised networks. However, the availability, or rather unavailability, of annotated training data makes semi- and unsupervised networks very attractive.

Most unsupervised change-detection networks can be divided into two categories based on their structure. The first type of structure is depicted in **Figure 2a** and relies, in essence, on transfer learning. A double-stream architecture provides a natural way of using transfer learning. The parallel subnetworks that fulfill the role of feature extraction can be pretrained on tasks other than change detection. The feature extraction step is then followed by an automated algorithm that uses the features to make a decision about changed areas in the images without the need for additional training. The task used for pretraining is often related to the end goal of the method, such as using RS scene classification to pretrain the feature extractors.

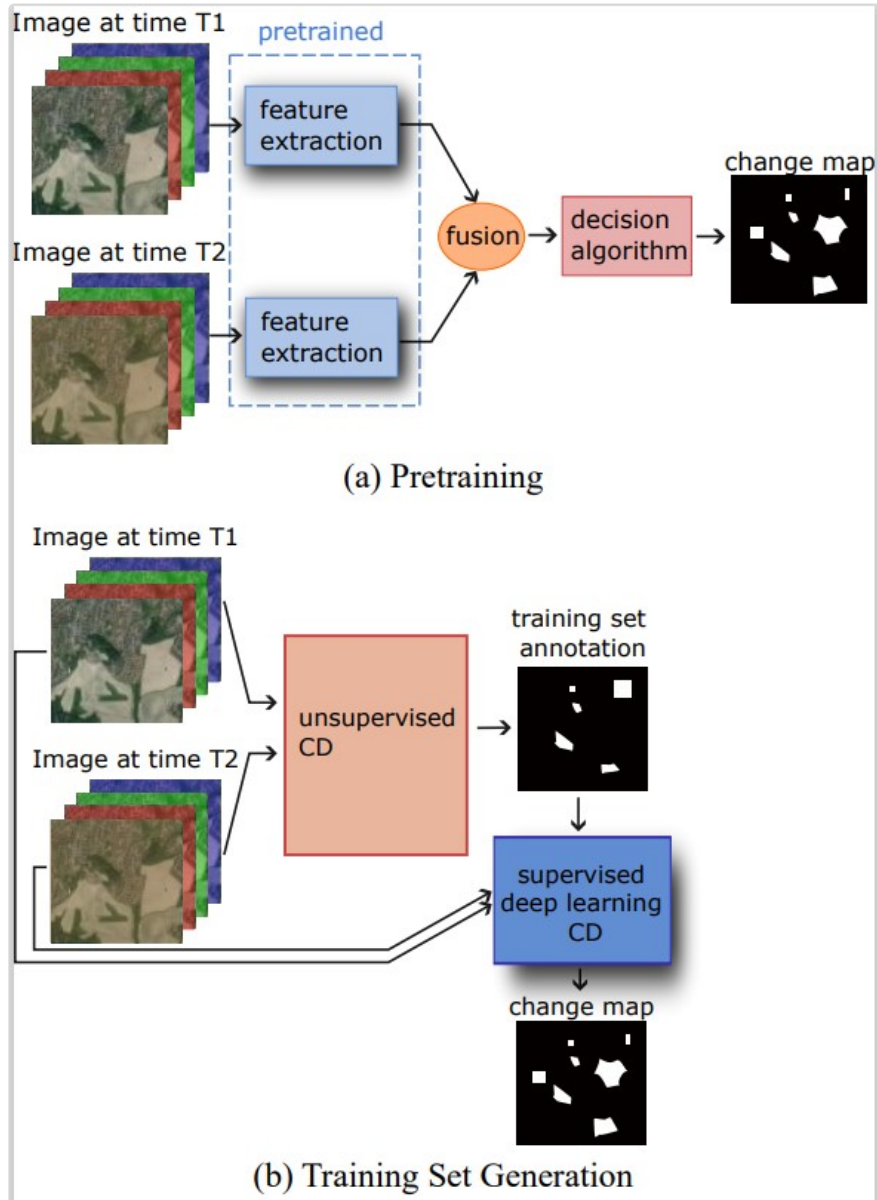


Figure 2: Two possible unsupervised network structures. (a) Network with pretrained feature extractors. Siamese feature extractors that have been pretrained on a related task are used to automatically extract image features. The features are then input into an automatic decision algorithm that compares them and outputs a change map. (b) Using an unsupervised classical change-detection method to create a labelled training dataset for a supervised deep-learning-based CD model.

Another method often used to achieve unsupervised change detection is automated training-set generation (**Figure 2b**). In this case, a method that does not require annotations is used to (often partially) annotate the data. The CVA is a commonly used algorithm for this purpose, which provides initial classifications for image pixels, dividing them into changed, unchanged and undecided categories. This auto-generated partially annotated training set is then used to train a deep-learning-based supervised classifier, which outputs a change map. In many cases, GANs are used as part of the method to generate annotations.

In the case of semi-supervised networks, the need for training data is reduced but not completely eliminated. This can be achieved through various approaches. As with fully unsupervised models, pretraining can be used. Another possibility to reduce the amount of necessary annotated data is to use a generative adversarial network.

4. Challenges and outlook

Change detection in remote sensing is a useful but demanding task with a unique set of challenges. Despite the large amount of available RS imagery, high-quality large annotated CD datasets are not simple to create. They often need to be annotated pixel-wise by hand, which is time-consuming and laborious. Furthermore, unlike most other applications of deep learning, they require two or more images, which increases the amount of data to acquire, makes the process of annotating more complex and introduces the additional need for coregistration.

The fact that the idea of "change" itself can be defined in more than one way means that there will be large differences between various annotated datasets. This leads to difficulties comparing and evaluating networks if they have not been tested on the same data. It also reduces the transferability of a network trained on one dataset to another dataset. One could argue that the concept of change is so broad that it should not be affected by (minor) differences between datasets; however, in reality, providing an annotation always implicitly chooses the category of changes to be considered as relevant, as well as their context. A network trained on one dataset will, thus, learn a particular type of changes and will not necessarily be able to recognise a type of change that it has not encountered in training.

An additional difficulty is posed by the fact that true changes are fairly rare in RS imagery. This means that the vast majority of pixels within any dataset are unchanged, and this large class imbalance requires a special approach, such as a well-chosen loss function.

The creation of a standard set of open test datasets would significantly facilitate the evaluation and comparison of various models. It is currently difficult to efficiently compare the published models, as comparisons of performance on different datasets are nearly meaningless. Ideally, models should be evaluated on several datasets, as they can exhibit significant differences.

From the presented overview of the reported models, several trends can be identified. Convolutional neural networks with a double-stream architecture are the most commonly used models for supervised change detection. The most recent ones generally include some form of an attention mechanism, which seems to improve the model's performance significantly. It is also apparent that the choice of a loss function is an important factor due to the class imbalance in the change-detection datasets. Among the most recent models are several transformer-based ones, which naturally follows from the success transformers have been having in the fields of natural language processing and computer vision.

In recent years, there has been more focus on unsupervised and semi-supervised networks to eliminate or at least reduce the need for annotated datasets. However, these models generally do not yet achieve the accuracy of their supervised counterparts. Using an unsupervised model also leads to a loss of some control over the type of change to be detected. The main method of teaching the model to distinguish between what researchers consider relevant and irrelevant, lies in the annotation of the changes as such.

Unsupervised models that rely on automatic algorithms for preliminary annotations, but also those that rely on pretraining on other types of data, give up this ability to fine-tune the types of changes to be considered relevant. This minor loss of control over the type of change is traded for the ability to train on much larger amounts of data with much less upfront effort, which is the main advantage of unsupervised models. Further development in the direction of unsupervised as well as semi-supervised models can be expected.

Deep-learning models have achieved great results in the application of change detection and, in the majority of cases, have surpassed more classical methods. This is largely due to their ability to model complex relationships between pixels within the images, thus, taking into account sufficient context to distinguish apparent and irrelevant changes from changes of interest. This is particularly important for high-resolution images and images of complex, varying landscapes.

The development of change-detection methods for remote sensing goes hand-in-hand with the development of the technology used for remote sensing. The amount of available RS imagery is increasing every year as new and improved airborne and space-borne platforms are being deployed. Satellites, in particular, are a source of large amounts of data, due to their frequent revisit times and increasingly higher spatial resolutions. They provide a means for monitoring hard-to-access areas as well as for observing changes regularly over long periods of time.

In fact, the increasing availability of satellite images with frequent revisit times and improved spatial resolution opens up the use of longer time-series, rather than focusing on two-image change detection. Time-series, by definition, include more information and present a very interesting platform for change detection, which has been thus far comparatively little explored. The facts that true changes are rare and that the revisit times of many satellites are short opens up possibilities for the development of self-supervised models, exploiting the readily available time-series of images.

Another possible avenue for future research is semantic change detection, where not only is the presence of a change detected but also the type of change is classified. Preparing ground truth including this information is more challenging than simply focusing on change/no-change.

Change detection is a complex and multifaceted problem to the point that one could say it is many problems with some common characteristics. The choice of the right approach will heavily depend on the type of data to be analysed and on the goal of the analysis, as the performance of models will somewhat differ from dataset to dataset and task to task.

While the recent years have seen an unprecedented growth in the interest for change detection in remote sensing, the challenge is far from solved.

References

1. Abdollahi, A.; Pradhan, B.; Shukla, N.; Chakraborty, S.; Alamri, A. Deep Learning Approaches Applied to Remote Sensing Datasets for Road Extraction: A State-Of-The-Art Review. *Remote Sens.* 2020, 12, 1444.
2. Hu, T.; Su, Y.; Xue, B.; Liu, J.; Zhao, X.; Fang, J.; Guo, Q. Mapping Global Forest Aboveground Biomass with Spaceborne LiDAR, Optical Imagery, and Forest Inventory Data. *Remote Sens.* 2016, 8, 565.
3. Xie, Q.; Dash, J.; Huete, A.; Jiang, A.; Yin, G.; Ding, Y.; Peng, D.; Hall, C.C.; Brown, L.; Shi, Y.; et al. Retrieval of Crop Biophysical Parameters from Sentinel-2 Remote Sensing Imagery. *Int. J. Appl. Earth Obs. Geoinf.* 2019, 80, 187–195.
4. Xu, W.; Vinã, A.; Kong, L.; Pimm, S.L.; Zhang, J.; Yang, W.; Xiao, Y.; Zhang, L.; Chen, X.; Liu, J.; et al. Reassessing the Conservation Status of the Giant Panda Using Remote Sensing. *Nat. Ecol. Evol.* 2017, 1, 1635–1638.
5. Stapleton, S.; LaRue, M.; Lecomte, N.; Atkinson, S.; Garshelis, D.; Porter, C.; Atwood, T. Polar Bears from Space: Assessing Satellite Imagery as a Tool to Track Arctic Wildlife. *PLoS ONE* 2014, 9, e101513.
6. Isikdogan, F.; Bovik, A.C.; Passalacqua, P. Surface Water Mapping by Deep Learning. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 2017, 10, 4909–4918.
7. Zhang, E.; Liu, L.; Huang, L.; Ng, K.S. An Automated, Generalized, Deep-Learning-Based Method for Delineating the Calving Fronts of Greenland Glaciers from Multi-Sensor Remote Sensing Imagery. *Remote Sens. Environ.* 2021, 254, 112265.
8. Browning, D.M.; Steele, C.M. Vegetation Index Differencing for Broad-Scale Assessment of Productivity Under Prolonged Drought and Sequential High Rainfall Conditions. *Remote Sens.* 2013, 5, 327–341.
9. Ghaffarian, S.; Kerle, N.; Pasolli, E.; Arsanjani, J.J. Post-Disaster Building Database Updating Using Automated Deep Learning: An Integration of Pre-Disaster OpenStreetMap and Multi-Temporal Satellite Data. *Remote Sens.* 2019, 11, 2427.
10. Malakhov, D.; Dyke, G.; King, C. Remote Sensing Applied to Paleontology: Exploration of Upper Cretaceous Sediments in Kazakhstan for Potential Fossil Sites. *Palaeontol. Electron.* 2009, 12, 1935–3952.
11. Emerson, C.; Bommersbach, B.; Nachman, B.; Anemone, R. An Object-Oriented Approach to Extracting Productive Fossil Localities from Remotely Sensed Imagery. *Remote Sens.* 2015, 7, 16555–16570.
12. Silván-Cárdenas, J.L.; Caccavari-Garza, A.; Quinto-Sánchez, M.E.; Madrigal-Gómez, J.M.; Coronado-Juárez, E.; Quiroz-Suarez, D. Assessing Optical Remote Sensing for Grave Detection. *Forensic Sci. Int.* 2021, 329, 111064.
13. Wellmann, T.; Lausch, A.; Andersson, E.; Knapp, S.; Cortinovis, C.; Jache, J.; Scheuer, S.; Kremer, P.; Mascarenhas, A.; Kraemer, R.; et al. Remote Sensing in Urban Planning: Contributions towards Ecologically Sound Policies? *Landsc. Urban Plan.* 2020, 204, 103921.
14. Singh, A. Digital Change Detection Techniques Using Remotely-Sensed Data. *Int. J. Remote Sens.* 1989, 10, 989–1003.

