Change Detection through Intra-urban Classification of Very High-Resolution Satellite Images

Mariana de Andrade Dias Alves

Instituto Superior Técnico, Universidade de Lisboa, Portugal

December 2022

Abstract

Satellite images offer big cities the ability to easily acquire their entire territory at a good resolution (70 cm for the Pleiades satellites), with high frequency updates that allow the automatic detection of certain changes. Following the "Zéro Artificialisation Nette" biodiversity plan proposed by the Ministry of Ecological Transition in France, municipal governments are searching for solutions to measure soil artificialisation yearly and accurately. In this work, a tool is proposed to help monitor land use and understand the artificialisation phenomena in metropolises, based on the processing of Pleiades images. The tool receives as inputs two very high-resolution satellite images of a region at times T1 and T2 and produces a change map, which indicates, at a pixel-level (0.5 m), surfaces which have been artificialised or naturalised between T1 and T2. It is composed of three core modules: a preprocessing module, responsible for converting the input images in a standardised comparable format (through calibration and corrections); an image segmentation module, responsible for segmenting the input images into semantically meaningful objects at a higher level of abstraction (through deep learning segmentation); and a change detection module, responsible for performing a change detection analysis between the feature maps. It is shown that the proposed solution can generate artificial maps which correctly identify artificialised and naturalised regions, with a higher spatial resolution (0.5m) than the current benchmark solution: CORINE Landcover Dataset. The solution is also applicable in other urban topics and is currently being tested in CNES projects in Toulouse, Paris and Senegal.

Keywords: soil artificialisation, change detection, very high-resolution satellite images, computer vision, deep learning segmentation

1. Introduction

Biodiversity is undergoing massive and rapid erosion. Among the main causes is the artificialisation of soils[1]. According to cadastral data, as of 2019, 3.5 million hectares in France have been artificialised [11] and, at the national level, between 20000 and 30000 hectares of NAF soils are consumed.

As an effort to curb this trend, the French Ministry of Ecological Transition presented the "Zéro Artificialisation Nette" (ZAN) biodiversity plan [11], with the goal of reaching "zero net artificialisation" by 2050. The plan measures net artificialisation in terms of a balance of artificialised surfaces and naturalised surfaces) and by surface area [not volume]. In accordance to the most recent version of the law, the ZAN goal is tracked through the annual publication of an inventory of the consumption of natural spaces. Currently, the two main providers of this type of data are the Teruti-lucas Survey used by the Ministry of Agriculture and Food and the CORINE Landcover database (CLC+) used by the Ministry of Ecological Transition [10]. The survey is based on on-site polling and the database is based on the automatic analysis of Sentinel satellite images. The main drawbacks of the Teruti-Lucas Survey are the manual resources needed and the time it takes to get the results. The main limitation of the CLC+ database is the geometric accuracy of its satellite data of 10 m, which does not allow for the detection of dispersed rural areas and roads, and, consequently, results in an underestimation of artificialisation.

There is a clear need for a solution, which can provide governments with data at a higher geometric resolution in a narrower time interval.

Due to the technological improvements of both acquisition sensor technology and data processing

[1] According to the French Ministry of Ecological Transition, Soil Artificialisation consists in transforming natural, agricultural or forest soils through development operations that can lead to partial or total loss of permeability, in order to assign them to urban or transport functions (housing, activities, shops, infrastructures, public facilities, etc.) [11].

algorithms of the last decade, there are currently available multitemporal and multispectral satellite images with very high spatial resolution, i.e. which are acquired with a metric to sub-metric spatial resolution in the panchromatic channel, by passive sensors.

The Pleiades constellation, developed at the Centre National d'Études Spatiales (CNES), is composed of two optical Earth-imaging satellites, capable of acquiring stereo and multispectral satellite images with a Very High spatial Resolution (VHR) [2].

The constellation can revisit any place on earth within a day, providing data on entire territories in a very short window of time. The optical sensors aboard the satellite capture 5 spectral bands, namely, a panchromatic band (0.47–0.83 μm) with a high geometric resolution of 0.7 m/pixel (much better than the 10 m resolution of Sentinel 2), and a blue (0.43–0.55 μm), green (0.50–0.62 μm), red (0.59-0.71 μm) and near infrared (0.74–0.94 μm) bands with a lower spatial resolution of 2.8 m/pixel [13]. Finally, Pleiades satellites allow for the acquisition of stereo images, which provide height information and allow for the generation of digital height models, which indicate, the height of objects, such as buildings and trees, relative to the surface of the earth [2].

The exploration of such images could provide the necessary tools for metropolises to efficiently measure artificialisation in France.

Artificialisation detection through the comparison of satellite images can be inserted into the research domain of **Remote Sensing Image Change Detection**, which refers to the use of remote sensing images and related data, in the same region at different periods, to, through image processing and analysis, identify and extract significant changes, and generate change images, which reflect accurate ground object change information.

Despite the potential of VHR satellite images in artificialisation detection, challenges associated with these types of data must be taken into consideration when performing change detection. Firstly, spectral variability is high. Buildings, for example, have complex appearances due to pipelines, chimneys, etc; as a result, spectral characteristics in VHR images are significantly heterogeneous. High spectral variability within geographic objects of interest increases within-class variance, resulting in an increase in the uncertainty of image interpretation methods [26]. Additionally, differences in the conditions of acquisition of the images to be compared, namely, time of acquisition (season of the year), sensor parameters, view angles (satellite orbit), etc. also contribute to the heterogeneity between images. Differences in the acquisition view angle, for example, can induce misalignments due to the impact of topography, small changes in relief of the terrain or the presence of buildings [27]. Finally, shadows cast by terrain, buildings, and trees, are a significant issue in larger areas for change detection [26].

Several systems have been proposed to tackle such issues and perform change detection using VHR satellite images.

In the unsupervised domain, Bovolo et al.[3] proposed a method, where a multi-level extension of the Change Vector Analysis method coupled with manual tresholding is applied to features, generated from segmentation rules based on spectral information to the input images, in order to generate a change map. Although simple to implement, spectral-based segmentation methods identify all spectral changes, including, for example, weather changes in vegetation foliage, which are not of interest for our application.

In the deep learning realm, Song et al. [28] proposed a method, where a neural network is trained to generate change maps from input images, by using pixels and labels with a low level of uncertainty, obtained by feeding the outputs of five traditional unsupervised methods for change detection into an uncertainty algorithm. Preprocessing of the input images are necessary before applying the method. The main drawback from the method is that the performance of the method is limited by the performance of the unsupervised change detection methods chosen.

Regarding supervised solutions, Saha et al. [24] proposed a method, which extracts deep features with the highest variability between two input images from a pre-trained CNN [30] and applies a multi-level extension of the CVA coupled with an automatic thresholding method on those features to generate change maps. The pre-trained CNN performs segmentation for classes "impervious surfaces", "building", "low vegetation", "tree", "car" and "clutter/background". Although showing promising results, it is important to note that the CNN was trained using manually anotated datasets from two regions of Germany, which are not representative of urban areas around the world. Additionally, as in any supervised approach, the segmentation accuracy is dependent on the quality of the labelled data.

In the present work, a change detection tool is proposed to detect soil artificialisation in multitemporal very-high resolution ($<$ 1m spatial resolution) stereo satellite images. Similarly to [24], the tool exploits a UNET [20] semantic segmentation to gen-

erate features useful to detect changes of interest. Differently from [24], the tool includes a labelling method which generates labelled data from images. This means the tool can be tailored or generalised to any region in the world, as long as input data is available. In order to mitigate errors originated by the high spectral variability of the input images, altitude information extracted from the satellite images is paired with the input spectral information. The the tool addresses shadow-induced errors through filtering and the grabcut algorithm [21] is exploited to mitigate classification errors from the deep learning method.

As a satellite imagery based solution, the proposed tool does not require the resources or time associated with the Teruti-Lucas Survey. By processing satellite images with a sub-metric resolution, the tool allows for a finer final product resolution than the CLC+ geographic database.

2. Background

The present work proposes an object-oriented change detection method, with image preprocessing operations.

2.1. Image preprocessing

Traditional image preprocessing techniques include any regularization operations to mitigate errors or increase quality of the images, which are performed on satellite images before information is exploited. Three common preprocessing operations, used in this work are calibration, pan-sharpening and the generation of digital height models.

2.1.1 Calibration

Raw images captured by optical satellites are characterized by pixel values, called digital numbers (DN), which quantify the energy recorded by the satellite optical detectors, influenced by acquisition conditions [17]. For this reason, DNs are not physically interpretable or comparable. Image calibrations or corrections aim at normalizing those pixel values, in order for them to be spectrally and geometrically independent from acquisition conditions, such as light and atmospheric conditions, the sensor internal parameters, the satellite's frame of reference, the earth's curvature, etc.

Radiometric Calibration is the process of converting digital numbers into a physical unit, which is spectrally independent of the images' acquisition conditions [5].

Calibrated values are called **surface reflectivity** (equivalent to surface reflectance or top of canopy (TOC) reflectivity), which is a ratio denoting the fraction of light that is reflected by the underlying surface in the given spectral range. As such, its values lie in the range [0,1] [5]. Depending on the available data regarding acquisition conditions, different atmospherical models can be used to obtain surface reflectivity. In the present work, the 6S radiative transfer model [29, 25] is used.

Rectification is a geometric correction, which removes the effects of image perspective (tilt) and relief (terrain) for the purpose of creating a planimetrically correct image, projected in a standard frame of reference, from which it can be adequately compared with other images [8]. If the frame of reference is a coordinate reference system, this process, named **orthorectification** (orthographic = map), also includes the assignment of cartographic coordinates (latitude, longitude) to image data (line, column). Orthorectification is extensively detailed in literature. In this work, the python library ORFEO Toolbox[6], developed at CNES, is used to perform orthorectification.

2.1.2 Pan-sharpening

A very common remote sensing process, called pan-sharpening, is to fuse a panchromatic image with high spatial resolution with a multi-spectral one with a low spatial resolution, so as to get an image combining the spatial resolution of the panchromatic image with the spectral richness of the multi-spectral image [5].

Several pan-sharpening methods have been proposed. The Relative Component Substitution (RCS) method [8] is used in the present work.

2.1.3 Digital Elevation Models

Models of the Earth's surface can be generated using stereo or tri-stereo images, which are images of a scene taken from different points of view. Different elevation models exist depending on the reference location from which altitude is measured. For the proposed application, a digital height model (DHM) is explored.

A DHM represents the height of all objects in Earth's surface relative to the bare ground surface [15] and corresponds to the difference between a digital surface model (DSM) and a digital terrain model (DTM).

A DSM represents the height of the Earth's surface including trees, buildings, and any other surface objects. The height can be measured relative to an Earth ellipsoid or an Earth geoid[18]. Several approaches have been proposed to calculate DSMs from satellite images. In the current work, the library CNES Algorithms to Reconstruct Surface (CARS) [32], developed at CNES, is used.

A DTM is a height representation of the bare ground (bare earth) topographic surface of the Earth excluding trees, buildings, and any other surface objects. The height can be measured relative to

an Earth ellipsoid or an Earth geoid [18]. Although multiple approaches have been proposed for the generation of DTM's, in this work, the Bulldozer library [15], developed at CNES, is used. Bulldozer relies on the cloth simulation principle, firstly proposed in [33], where a filter performs an operation similar to dropping a cloth on top of an inverted DSM. The cloth will cover the buildings without plunging into them, thus revealing the shape of the bare ground.

2.2. Object-oriented change detection

Object-oriented change detection methods process images through the analysis of image segments and not image pixels. Images segments are obtained through semantic segmentation. Firstly, an image is divided into several regions with homogeneity of shape and spectral properties; and secondly, each of these regions is associated with a semantic labelling. After segmentation, traditional change detection methods are carried out [31].

In this work, three types of semantic segmentation are used, namely, spectral-based, deep learning and graph-based segmentation.

2.2.1 Spectral-based segmentation

Spectral-based segmentation refers to the use of spectral indices and thresholding to extract objects or features of interest from a raster image.

Spectral indices are mathematical equations, which combine spectral reflectance from two or more spectral bands in order to highlight pixels showing the relative abundance or lack of a land-cover type of interest in an image. In the present work, indices on vegetation [22], bare soil [12], shadows [23] and water [19] are explored.

Segmentation maps can then be obtained by thresholding any spectral index, manually or through automatic algorithms. In the present work, the multi-otsu method [16] will be explored.

2.2.2 Deep learning segmentation

Deep Learning Segmentation refers to the use of deep neural networks to segment an image into objects or features of interest. Several network architecture have been proposed for the problem of supervised segmentation. One of the most renown architectures, which is explored in this work, is the U-Net.

The U-Net [20] is a fully convolutional neural network (FCN) that performs a compression, followed by a decompression of an image. The spatial compression step allows to extract feature information from the input image. The decompression step allows to generate pixel-wise segmentation from the information calculated during the compression. As in any FCN, convolutional layers are responsible for identifying pattern features in data.

2.2.3 Graph-based segmentation

GrabCut is an image segmentation method based on graph cuts, proposed by Rother et al.[21]. Its output is a binary map which separates background pixels from foreground pixels.

The idea is to represent an image as a graph, such that each pixel is a node connected to it's 4 or 8 surrounding neighbors (the surrounding pixels). The edges connecting pixel nodes have weights which reflect inter-pixel color similarities. There is a virtual "source" node denoting foreground, and a virtual "sink" node denoting background. The source and sink nodes are both connected to every single pixel-node. The weights associated with the edges connecting the "sink" and "source" nodes to the pixels reflect the probability of the pixel to match a color distribution of the background or foreground.

The goal is then to solve a mincut algorithm, with a cost function equal to the sum of all weights of the edges that are cut. Pixels connected to the source node become foreground and those connected to the sink node become background. This iterative process stops when the classification converges.

2.2.4 Change detection

The change detection method employed in the present work is image differencing.

Image Differencing is one of the simplest unsupervised change detection methods. A **difference image**, containing change information, is generated by finding the difference between each corresponding pixel in the two images, band by band [1]. For this technique to work, the two images must first be aligned so that corresponding points coincide, and their pixel values must be made compatible, either by careful calibration, or by post-processing (using color mapping).

3. Proposed Approach

The proposed framework for the *Zéro Artificialisation Nette* tool is presented in figure 1.

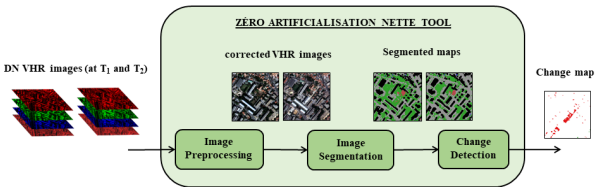


Figure 1: ZAN Tool architecture.

The tool receives as input two very high-resolution (VHR) satellite images of a region A at

times T1 and T2 and produces a change map, which indicates at a pixel-level (0.5 m) surfaces which have been artificialised or naturalised between times T1 and T2.

To achieve its goal, the tool is composed of three core modules: a preprocessing module, responsible for converting the input images in a standardised comparable format; an image segmentation module, responsible for segmenting the input images into semantically meaningful objects at a higher level of abstraction; and a change detection module, responsible for performing a change detection analysis between the feature maps. Depending on data availability, the tool is capable of generating labelled training datasets from any image.

The Image Preprocessing Module takes as input a Pleiades stereo or tri-stereo image bundle, in sensor geometry and digital number measurements, and generates a pansharpened image in top of canopy reflectivity pixel values, merged with a digital height model, both orthorectified into a reference coordinate system, provided as a parameter. The following operations, described in [8, 4, 32, 15], are performed:

- a radiometric calibration, where digital numbers are converted into top of canopy reflectivity pixel values of the nadir multi-spectral image;
- a pansharpening operation, where the nadir multi-spectral calibrated image is merged with the corresponding panchromatic image;
- an orthorectification, where the pansharpened image in sensor geometry is converted into a reference ground geographic geometry;
- a digital height model generation, from the stereo panchromatic images;
- a co-registration operation, where the digital height model is projected into the reference geometry of the pansharpened image;
- a concatenation operation, which generates the output raster image in GeoTIFF format, with five bands: red, green ,blue ,nir and height.

The Image Segmentation Module takes as input an orthorectified calibrated image composed of five bands (red, green, blue, near-infrared and height) and generates as output an artificial map composed of one band, classifying pixels as non-artificial and artificial. To achieve its goal, the module is divided into two steps: an initial classification step, in which the input image is used to generate a segmentation map of five classes, namely, vegetation, bare soil, building, road and water; and a binarisation step, in which the five classes are converted into two classes, namely, artificial and non-artificial.

The classification task is a multi-label problem solved using a supervised learning approach.

The network architecture considered is a UNET [20], with 5 levels of compression and 16 convolutional filters in the first and last layers. The activations functions of the hidden and output layers are the ReLU function and the sigmoid function, respectively, which are adapted to the multi-label problem. The loss function chosen for learning is the binary cross-entropy calculated as follows

$$L = -\sum_{j=0}^{M}\sum_{k=0}^{N} y_{jk}\log(\hat{y}_{jk}) + (1-y_{jk})\log(1-\hat{y}_{jk}) \tag{1}$$

where M is the number of pixels per window, N is the number of classes, y is the prediction result and $\hat{y}$ is the corresponding label (expected result).

As in any supervised approach, representative labelled data is necessary to train the model. A labelling method is used to generate binary masks of vegetation, bare soil, building, road and water classes, by exploiting spectral information from the input images, Openstreetmap [7] and Urban Atlas [14] open source urban planning datasets and Sentinel-2 satellite images[2].

Supervised prediction through the above mentioned methods generates a probability map with five bands, each containing the probabilities of pixels belonging to each class. Segmented maps are then generated through algorithm 1, where $MO(p, n)$ is the application of a multi-otsu method [16] to a probability map p to generate n thresholds and $GC(R, a, b, c, d)$ is the application of the grabcut algorithm [21] to a color image R, initialised with binary masks a, b, c and d corresponding to 'background', 'probably background', 'probably foreground' and 'foreground', respectively, to generate binary segmented masks.

As opposed to simply applying the argmax() function to the probability map, algorithm 1 improves upon the results of the deep learning approach, by using them as inputs in a graph-based binary segmentation method: grabcut. The grabcut algorithm typically generates a more conservative segmentation result i.e. smaller and more cohesive objects than the argmax() function. It is therefore ideal for the segmentation of objects with simple shaped borders, but is not as effective otherwise. Therefore, algorithm 1 applies the

[2]Refer to thesis for further information.

Algorithm 1 Urban Classification Algorithm

Input:

R, the color image to be segmented
p, the probability map

procedure URBAN CLASSIFICATION(R, p)
  for each $p \in \{p_{building}, p_{road}, p_{water}\}$ **do**
    $\tau_{min}, \tau_{middle}, \tau_{max} \leftarrow MO(p_k, 4)$
    $a \leftarrow p < \tau_{min}$
    $b \leftarrow \tau_{min} < p < \tau_{middle}$
    $c \leftarrow \tau_{middle} < p < \tau_{max}$
    $d \leftarrow \tau_{max} < p$
    $g \leftarrow GC(R, a, b, c, d)$
    $p \leftarrow p \cdot g$
  end for
  $p_{soil} \leftarrow p_{soil} \cdot \neg(g_{building} \vee g_{road} \vee g_{water})$
  $p_{all} \leftarrow [p_{vegetation}, p_{soil}, p_{building}, p_{road}, p_{water}]$
  $c_{urban} \leftarrow \underset{k}{\text{argmax}}\, p_{all}(k)$
end procedure

Output:

c_{urban}, the segmented urban map

grabcut algorithm to classes 'building', 'road' and 'water' and uses the results to exclude pixels from belonging to the 'building', 'road' and 'water' classes. Additionally, it prioritizes the 'vegetation' class probability results (as it has been shown to be the class with better performance).

Once the classification task is performed, binarisation of the segmented 5-class urban map is performed by classifying all pixels belonging to the classes vegetation, bare soil and water as non-artificial and all pixels belonging to the classes building and road as artificial.

The Change Detection Module takes as input two artificial maps, with classes artificial and non-artificial, and generates an artificialisation map, with classes artificialised, unchanged or naturalised. To achieve this goal, the following image differencing is performed, pixel by pixel,

$$c_{change} = c^{T1}_{artificial} - c^{T2}_{artificial} \tag{2}$$

where $c^{T1}_{artificial}$ and $c^{T2}_{artificial}$ are the classifications (1 or 2) given to the pixel in the input images of time $T1$ and $T2$ respectively.

Additionally, in order to mitigate classification errors induced by the presence of shadows, the following filtering operation is performed

$$c_{change} = c_{change} \cdot \neg(s^{T1} \oplus s^{T2}) \tag{3}$$

where s^{T2} are the shadow masks of image $T1$ and image $T2$. The shadow masks for both images are obtained by calculating the shadow indices [23] of each input image and thresholding the result, such that all pixels with a shadow index inferior to τ_{shadow} are considered to be shadows. The threshold can be obtained through automatic methods, but in the current implementation, the threshold was empirically defined at 0.12. The defined filtering operation is based on the assumption that *if a shadow appears in a given region in one image, but does not appear in the other image and a change appears in the corresponding region of the change map, that change was probably detected due to the misclassification of the shadow.*

Finally, a morphological opening with a radius of 3 pixels (1.5m) is performed on the change map, in order to remove objects which are not of interest for the considered application (< 10 m^2).

4. Results

4.1. Problem description

The proposed solution is tested using a pair of Pleiades stereo images of the Montpellier region, around October of 2018 and July of 2020.

The training dataset is composed of two labelled images of Montpellier and Toulouse, in October of 2018 and 2017 respectively, generated from the proposed labelling method.

4.1.1 Ground-truth image

The ground-truth artificialisation dataset (used for validation) is the result of filtering public data on demolition and construction permits of Montpellier, provided in the SITADEL database [9] between 2017 and 2020. In total, it contains 89 artificialised parcels and 26 naturalised parcels. Although the information provided by public authorities is overall reliable, the definition of change used to generate the ground-truth can be questioned: a building razed following a demolition permit can be considered as naturalised in the validation dataset even if the soil is not yet permeable. Moreover, a large parcel with a small construction will be considered as completely artificialised even if the change was minimal when compared to the size of the parcel.

4.1.2 Conversion to parcel level

As the results obtained from the system are at pixel-level, a conversion into parcel-level is necessary before performing the validation of the system. The conversion into parcel-level is performed as follows:

$$c_{change}(p) = \text{argmax}\,(n_{art}(p), \tau_p \cdot N, n_{nat}(p))) - 1 \tag{4}$$

where $c_{change}(p)$ is the class of a parcel (unchanged, artificialised or naturalised), N is the number of pixels in the parcel, $n_{art}(p) = \sum_{k=1}^{N}(c_{change}(k) = -1)$ is the number of pixels in the parcel classified

as artificialised, $n_{nat}(p) = \sum_{k=1}^{N}(c_{change}(k) = -1)$ is the number of pixels in the parcel classified as naturalised and $\tau_p = 0.1$ is the percentage of pixels in a parcel above which the parcel is considered changed.

4.1.3 Metrics

The tool is evaluated at two different stages: firstly during the training of the classifier network and secondly at the output stage, through the validation process described above.

At the network training stage of validation, five metrics are tracked, namely, the loss and accuracy of the training and validation datasets, and the precision, recall and confusion matrix of each class. The model chosen for prediction corresponded to the model with the lowest validation loss (in a total of 400 epochs).

At the output stage of validation, three metrics are evaluated, namely, the precision, recall and normalised confusion matrix of each class (by row).

4.2. Segmentation analysis

The segmentation approach is consolidated through the comparison of its results with other approaches.

4.2.1 Label sensitivity study

To validate the decision of segmenting the input images into a multi-label of five classes, three opposing segmentation problems are studied:

- a mono-label problem with five classes : vegetation, bare soil, building, road and water;
- a mono-label problem with two classes : non-artificial and artificial;
- a multi-label problem with two classes : non-artificial and artificial.

The labels used for the 2-class problems are generated by performing a binarisation operation on the corresponding 5-class label.

The loss and output layer activation function chosen for the mono-label problems are the **weighted categorical cross-entropy** and **softmax** functions. The loss is given by:

$$L = -\sum_{j=0}^{M}\sum_{i=0}^{N} y_{ji}\log(\frac{\hat{y}_{ji}}{\sum_{u=0}^{N}\hat{y}_{ju}}) \qquad (5)$$

where M is the number of pixels per window, N is the number of classes, y is the prediction result and $\hat{y}$ is the corresponding label (expected result). These are the default functions for mono-label problems.

Performance is evaluated at the artificial map, with classes natural and artificial, stage. In order for the class assignment phase, i.e. the operations which transform the probability map into an artificial map, to minimally influence the analysis on label sensitivity, the same method is used in all considered problems to select the dominating class: $c_{urban} = argmax(p_1, ..., p_N)$, where N is the number of classes considered, c_{urban} is the urban segmentation map and p_k is the probability map of class k.

Figure 2 shows a close-up of all the maps generated.

(a) Mono-label with 2 classes.

(b) Mono-label with 5 classes.

(c) Multi-label with 2 classes.

(d) Multi-label with 5 classes.

Figure 2: Close-ups of mono-label and multi-label artificial maps (red and green correspond to the artificial and non-artificial classes respectively) superimposed over the prediction images (in RGB composition).

It is possible to observe that the borders of objects, such as buildings, are more clearly defined in the multi-label maps as opposed to the mono-label maps. Additionally, the mono approaches seem to not be unable to identify roads.

In order to quantitatively evaluate the aforementioned results, a 2-class mono-label of the evaluated regions of both Montpellier images is used as a ground-truth. The label is generated through the same method as the 2-class mono-label training label. There are several issues with this approach. Firstly, all the inaccuracies generated from the labelling method which are present in the training label will also be present in the validation label. Secondly, the label is generated using the same method as one of the approaches' training label. Consequently, if training were to be 'perfect', that is, if the model were to learn to perfectly replicate the labelling method's output, the 2-class mono-label would automatically generate the best metrics, even

if it was not the most accurate label. However, because no additional ground-truth data is available and the 2-class mono-label labelling method directly generates images with the same format as the artificial maps, the decision is made to compare the artificial maps with the label.

Table 1 presents the accuracy of the model and the recall and precision of both classes for the considered approaches.

Table 1: Performance of artificial map generation methods.

Mode	Mono-Label		Multi-Label	
Nb of Classes	2	5	2	5
Accuracy	0.84	0.86	0.96	0.96
Non-artificial Recall	90.8%	93.0%	98.3%	**98.6%**
Non-artificial Precision	87.9%	88.0%	**96.5%**	96.2%
Artificial Recall	66.8%	66.5%	**90.5%**	89.8%
Artificial Precision	73.2%	78.2%	95.5%	**96.0%**

Overall, the multi-label approaches have superior results. Although the 2-class multi-label approach presents the best results, because the difference is not significant and a five class segmentation allows for several applications to be explored, it is possible to validate the choice of a multi-label approach with five classes.

4.2.2 Method sensitivity study

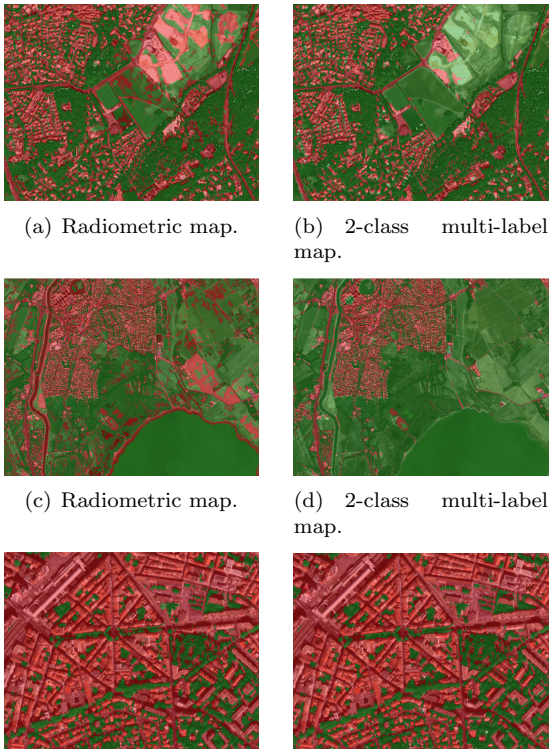
The choice of object segmentation through deep learning was made based on its research-proven efficacy, as well as the availability of label information for the considered application.

To evaluate the relevancy of the proposed segmentation method, which performs supervised learning, using labelled data, generated from a combination of radiometric masks and open source datasets, an alternative radiometric classification approach is tested.

The radiometric method generates an artificial binary map of classes non-artificial or artificial, by combining the label masks of three classes: vegetation, bare soil and water to define the class non-artificial. Each label is generated through the thresholding of a different radiometric index. The vegetation label is obtained by the same process as the proposed approach's vegetation label i.e. NDVI and NIR thresholding. The bare soil label is also generated through NDVI and NIR thresholding. Finally, the water label is generated through the thresholding of the SWM Index [19], using Sentinel-2 input images and an empirically defined minimum threshold.

The artificial map generated from the radiometric method is compared with the one generated from the best approach of the last analysis (2-class multi-label approach). Because no ground-truth is available to quantitatively evaluate both methods, only a qualitative analysis is performed.

In figure 3 several close-ups of the maps, which overall exemplify the benefit of the deep learning approach, are presented.



(a) Radiometric map. (b) 2-class multi-label map. (c) Radiometric map. (d) 2-class multi-label map. (e) Radiometric map. (f) 2-class multi-label map.

Figure 3: Close-ups of the 2-class multi-label and radiometric artificial maps (red and green correspond to the artificial and non-artificial classes respectively) superimposed over the images.

In the first close-up, it is possible to see how the neural network was able to detect a patch of bare soil, while the radiometric index classified it as artificial. In the second close-up, it is possible to see how the canal and the sand were both classified as artificial in the radiometric map, but identified as non-artificial in the deep learning map. Finally, the third close-up shows an example where the radiometric map is as successful as deep learning in identifying non-artificial regions. In this particular

case, the neural network only needed the radiometric information to perform a prediction.

Overall, there is a clear benefit in a machine learning approach if information additional to radiometric information is provided for a training image, even if the number of classes is kept as two. The information provided by osm in the label, as well as by the DHM generated by the stereo images, allows for the neural network to be able to more clearly distinguish artificial from non-artificial.

4.3. Output analysis

Performance metrics for the output of the proposed system are presented in table 2.

Table 2: Metrics (in %) of Change Classes for proposed solution.

Approach	Precision	Recall
Unchanged	86%	93%
Artificialised	89%	81%
Naturalised	71%	68%

The naturalised class presents the worst results out of all classes. A possible reason is the ambiguity of the definition of a naturalised surface, detailed in section 4.1.1. A recently demolished area may easily still be classified as a road, as opposed to bare soil, due to its imperviousness, but be identified as a naturalisation in the ground-truth image.

Figure 4 presents results obtained for different parcels.

(a) 2018. (b) 2020. (c) Change Map.

(d) 2018. (e) 2020. (f) Change Map.

Figure 4: Close-up on images (in RGB) and artificialisation maps (green and red in change maps correspond to naturalised and artificialised pixels) of parcels 34172000KM0031 (top) and 34172000EZ0159 (bottom).

It is possible to observe that the tool correctly identifies artificialisation.

5. Conclusions

In the present work, a tool capable of generating an indicator of urban artificialisation from multitemporal very high-resolution stereo satellite imagery, was presented. The tool generates artificial maps which correctly identify artificialised and naturalised regions, with a higher spatial resolution (0.5m) than the current solution used by the french government: CLC+.

A comparative analysis between the proposed segmentation approach and alternative approaches validated the approach. In particular, it was possible to verify the superior performance of a multi-label deep learning approach compared to a radiometric based approach or even a mono-label based approach.

Verification and validation of the porposed solution was performed at parcel scale using a validation dataset developed at CNES.

There are several leads to follow in future work.

Regarding the image segmentation module, the labelling method currently generates incomplete and noisy data, which influences the training of the classifier network. A future step could involve either the search for a more accurate label for training or the implementation of a network resistant to incomplete and noisy labels.

Regarding the validation process, because the validation dataset is at parcel scale, it is not possible to quantitatively validate the system at a finer level of resolution. An interesting future step could be to explore the possibility of developing of a validation dataset with a higher resolution.

Finally, all tests performed on the system used a single pair of images. An important development in the project is the robustification of the system relative to different input conditions, such as lighting, acquisition angles, period of acquisition and region analysed. Currently, the tool is being tested in the Eolab for projects in Toulouse and the Senegal.

The tool developed is currently being used in CNES missions in Toulouse, Paris and Senegal.

Acknowledgements

The author would like to thank professor Alexandre Bernardino, for all the support provided during the development of this work; and the faculty, Instituto Superior Técnico (IST), part of the University of Lisbon, for providing the necessary background for the thesis.

The author would also like to thank CNES for providing the resources and the project that inspired this work, as well as, technical insight and support from the various actors at EOLab, the laboratory inside CNES where the project was born. A special thanks goes to Solange Lemai-chenevier and Olivier Queyrut, who provided guidance through-

out most of the thesis and heavily contributed to its accomplishment.

References

[1] H. A. Afify. Evaluation of change detection techniques for monitoring land-cover changes: A case study in new burg el-arab area. *Alexandria Engineering Journal*, 50(2):187–195, 2011.

[2] Airbus. Pléiades imagery user guide. `https://www.intelligence-airbusds.com/imagery/constellation/pleiades/`. Accessed in 2022-08-22.

[3] F. Bovolo. A multilevel parcel-based approach to change detection in very high resolution multitemporal images. *IEEE Geoscience and Remote Sensing Letters*, 6(1):33–37, 2008.

[4] CNES. Optical calibration. `https://www.orfeo-toolbox.org/packages/doc/test-cookbook-workshops/Applications/app_OpticalCalibration.html`. Accessed in 2022-10-24.

[5] CNES. From raw image to calibrated product. `https://www.orfeo-toolbox.org/CookBook/recipes/optpreproc.html`, 2022. Accessed in 2022-10-24.

[6] CNES. Orfeo toolbox. `https://www.orfeo-toolbox.org/`, 2022. Accessed in 2022-10-20.

[7] S. Coast. Openstreetmap. `https://www.openstreetmap.org`. Accessed in 2022-09-01.

[8] R. Cresson, M. Grizonnet, and J. Michel. Orfeo toolbox applications. *QGIS and generic tools*, 1:151–242, 2018.

[9] M. de la transition écologique et solidaire. Sitadel. `https://www.data.gouv.fr/en/datasets/base-des-permis-de-construire-et-autres-autorisations-durbanisme-sitadel/`. Accessed in 2022-09-01.

[10] M. de la transition écologique et solidaire. Évaluation du taux d'artificialisation en france : comparaison des sources teruti-lucas et fichiers fonciers. `https://www.statistiques.developpement-durable.gouv.fr/sites/default/files/2019-08/datalab-56-evaluation-du-taux-d-artificialisation-en-france-aout2019.pdf`. Accessed in 2022-08-19.

[11] M. de la transition écologique et solidaire. Objective "zero net artificialization": which levers should be used to protect soils? `https://www.strategie.gouv.fr/sites/strategie.gouv.fr/files/atoms/files/dp_-_artificialisation_-_gb.pdf`, 2019. Accessed in 2022-08-19.

[12] S. Diek, F. Fornallaz, M. E. Schaepman, and R. De Jong. Barest pixel composite for agricultural areas using landsat time series. *Remote Sensing*, 9(12):1245, 2017.

[13] ESA. Pleiades instruments. `https://earth.esa.int/eogateway/missions/pleiades`. Accessed in 2022-08-22.

[14] ESA. Urban atlas. `https://land.copernicus.eu/local/urban-atlas`, 2022. Accessed in 2022-09-01.

[15] D. Lallement, P. Lassalle, Y. Ott, R. Demortier, and J. Delvit. Bulldozer: An automatic self-driven large scale dtm extraction method from digital surface model. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 43:409–415, 2022.

[16] P.-S. Liao, T.-S. Chen, P.-C. Chung, et al. A fast algorithm for multilevel thresholding. *J. Inf. Sci. Eng.*, 17(5):713–727, 2001.

[17] P. Lier, C. Valorge, and X. Briottet. Imagerie spatiale: Des principes d'acquisition au traitement des images optiques pour l'observation de la terre. 2008.

[18] D. Maune. Digital elevation model technologies and applications: The dem users manual. 01 2007.

[19] M. Milczarek, A. Robak, and A. Gadawska. Sentinel water mask (swm) - new index for water detection on sentinel-2 images. 7th Advanced Training Course on Land Remote Sensing, 09 2017.

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

[21] C. Rother, V. Kolmogorov, and A. Blake. Grabcut -interactive foreground extraction using iterated graph cuts. *ACM Transactions on Graphics (SIGGRAPH)*, August 2004.

[22] J. W. Rouse Jr, R. H. Haas, J. Schell, and D. Deering. Monitoring the vernal advancement and retrogradation (green wave effect) of natural vegetation. Technical report, 1973.

[23] D. Rüfenacht, C. Fredembach, and S. Süsstrunk. Automatic and accurate shadow detection using near-infrared information. *IEEE transactions on pattern analysis and machine intelligence*, 36(8):1672–1678, 2013.

[24] S. Saha, F. Bovolo, and L. Bruzzone. Unsupervised deep change vector analysis for multiple-change detection in vhr images. *IEEE Transactions on Geoscience and Remote Sensing*, 57(6):3677–3693, 2019.

[25] S.A.L.S.A. The 6s code is a basic rt code used for calculation of lookup tables in the modis atmospheric correction algorithm. `http://6s.ltdri.org/`. Accessed in 2022-10-24.

[26] A. Shafique, G. Cao, Z. Khan, M. Asad, and M. Aslam. Deep learning-based change detection in remote sensing images: a review. *Remote Sensing*, 14(4):871, 2022.

[27] Y. T. Solano-Correa, F. Bovolo, and L. Bruzzone. An approach for unsupervised change detection in multitemporal vhr images acquired by different multispectral sensors. *Remote Sensing*, 10(4):533, 2018.

[28] A. Song, Y. Kim, and Y. Han. Uncertainty analysis for object-based change detection in very high-resolution satellite images using deep learning network. *Remote Sensing*, 12(15):2345, 2020.

[29] E. F. Vermote, D. Tanré, J. L. Deuze, M. Herman, and J.-J. Morcette. Second simulation of the satellite signal in the solar spectrum, 6s: An overview. *IEEE transactions on geoscience and remote sensing*, 35(3):675–686, 1997.

[30] M. Volpi and D. Tuia. Dense semantic labeling of subdecimeter resolution images with convolutional neural networks. *IEEE Transactions on Geoscience and Remote Sensing*, 55(2):881–893, 2016.

[31] L. Xu, W. Jing, H. Song, and G. Chen. High-resolution remote sensing image change detection combined with pixel-level and object-level. *IEEE Access*, 7:78909–78918, 2019.

[32] D. Youssefi, J. Michel, E. Sarrazin, F. Buffe, M. Cournet, J.-M. Delvit, C. L'Helguen, O. Melet, A. Emilien, and J. Bosman. Cars: A photogrammetry pipeline using dask graphs to construct a global 3d model. In *IGARSS 2020-2020 IEEE International Geoscience and Remote Sensing Symposium*, pages 453–456. IEEE, 2020.

[33] W. Zhang, J. Qi, P. Wan, H. Wang, D. Xie, X. Wang, and G. Yan. An easy-to-use airborne lidar data filtering method based on cloth simulation. *Remote sensing*, 8(6):501, 2016.