Article

Deep-learning *Versus* OBIA for Scattered Shrub Detection with Google Earth Imagery: *Ziziphus lotus* as Case Study

Emilio Guirado [1,†] iD**, Siham Tabik** [2,†,*] iD**, Domingo Alcaraz-Segura** [1,3,4] iD**, Javier Cabello** [1,5] iD **and Francisco Herrera** [2] iD

[1] Andalusian Center for Assessment and Monitoring of Global Change (CAESCG), University of Almería, 04120 Almería, Spain; e.guirado@ual.es (E.G.); dalcaraz@ugr.es (D.A.-S.); jcabello@ual.es (J.C.)

[2] Soft Computing and Intelligent Information System Research Group, University of Granada, 18071 Granada, Spain; herrera@decsai.ugr.es

[3] Department of Botany, Faculty of Science, University of Granada, 18071 Granada, Spain

[4] Iecolab., Interuniversitary Institute for Earth System Research in Andalusia (IISTA), University of Granada, 18006 Granada, Spain

[5] Department of Biology and Geology, University of Almería, 04120 La Cañada, Almería, Spain

* Correspondence: siham@ugr.es

† These authors contributed equally to this work.

Abstract: There is a growing demand for accurate high-resolution land cover maps in many fields, e.g., in land-use planning and biodiversity conservation. Developing such maps has been traditionally performed using Object-Based Image Analysis (OBIA) methods, which usually reach good accuracies, but require a high human supervision and the best configuration for one image often cannot be extrapolated to a different image. Recently, deep learning Convolutional Neural Networks (CNNs) have shown outstanding results in object recognition in computer vision and are offering promising results in land cover mapping. This paper analyzes the potential of CNN-based methods for detection of plant species of conservation concern using free high-resolution Google Earth™ images and provides an objective comparison with the state-of-the-art OBIA-methods. We consider as case study the detection of *Ziziphus lotus* shrubs, which are protected as a priority habitat under the European Union Habitats Directive. Compared to the best performing OBIA-method, the best CNN-detector achieved up to 12% better precision, up to 30% better recall and up to 20% better balance between precision and recall. Besides, the knowledge that CNNs acquired in the first image can be re-utilized in other regions, which makes the detection process very fast. A natural conclusion of this work is that including CNN-models as classifiers, e.g., ResNet-classifier, could further improve OBIA methods. The provided methodology can be systematically reproduced for other species detection using our codes available through (https://github.com/EGuirado/CNN-remotesensing).

Keywords: *Ziziphus lotus*; plant species detection; land cover mapping; Convolutional Neural Networks (CNNs); Object-Based Image Analysis (OBIA); remote sensing

1. Introduction

Changes in land cover and land use are pervasive, rapid, and can have significant impact on humans, the economy, and the environment. Accurate land cover mapping is of paramount importance in many applications, e.g., biodiversity conservation, urban planning, forestry, natural hazards, etc. ([1,2]). Unfortunately, land-cover mapping processes are often not accurate enough, costly, and time-consuming. In addition, frequently the classification settings for an image in one site cannot be directly applied to a different image in a different site.

In practice, land cover maps are built by analyzing remotely sensed imagery, captured by satellites, airplanes, or drones, using different classification methods. The accuracy of the results depends on the quality of the input data (e.g., spatial, spectral, and radiometric resolution of the images) and on the used classification method. The most commonly used methods can be divided into two categories: pixel-based classifiers and Object-Based Image Analysis (OBIA) ([3]). Pixel-based methods use only the spectral information available for each pixel. They are faster but ineffective in some cases, particularly for high-resolution images and heterogeneous objects detection ([4,5]). Object-based methods take into account the spectral as well as the spatial properties of image segments (i.e., set of similar neighbor pixels). They are more accurate but computationally more expensive and very time-consuming since they require high human intervention and a usually large number of iterations to obtain acceptable accuracies. Currently, the most commonly used software implementing OBIA-methods is the privative Definiens-eCognition ([6]), which provides a friendly graphical user-interface for non-programmers. There exist several free and open source OBIA-software but they are less popular ([7,8]).

To detect a specific object (e.g., a particular plant species individual) in an input image, first, the OBIA method divides the image into segments (e.g., by using a multi-resolution segmentation algorithm), and then classifies the segments based on their similarities (e.g., by using algorithms such as the k-nearest neighbor, Random forest, or Support vector machines [9–12]). This procedure has to be repeated and optimized for each single input image and the knowledge acquired (i.e., the OBIA segmentation and classification settings) from one input image cannot be directly reutilized in another.

Convolutional Neural Networks (CNNs)-based models have demonstrated impressive accuracies in object recognition and image classification in the field of computer vision ([13–16] and are starting to be used in the field of remote sensing ([17]). This success is due to the availability of larger training datasets, better algorithms, improved network architectures, faster GPUs and also improvement techniques such as data-augmentation and transfer-learning, which allow reutilization of the knowledge acquired from a set of images into other new images. Currently, the most commonly used software implementing CNNs is the open source library of Tensorflow by Google[TM] ([18]), which requires programming skills since it does not have a graphical user-interface.

This paper analyzes the potential of CNN-based methods for plant species mapping using high-resolution Google Earth[TM] images and provides an objective comparison with the state-of-the-art OBIA-based methods. As case study, it aims to map *Ziziphus lotus* shrubs, the dominant species of the European priority conservation habitat "Arborescent matorral with *Ziziphus*", which is experiencing a serious decline in Europe during recent decades ([19]) (though it is also present in North Africa and Middle East). This case is challenging since *Ziziphus lotus* individuals have diverse shapes, sizes, distribution patterns, and physiological status. In addition, distinguishing *Ziziphus lotus* shrubs from neighboring plants in remote sensing images of different regions is complex for non-experts and for automatic classification methods since the surrounding plants and the soil background strongly differ. In particular, the contributions of this work are:

- Developing an accurate and transferable CNN-based detection model for shrub mapping using free high-resolution remote sensing images, extracted from Google Earth[TM].
- Designing a new dataset that contains images of *Ziziphus lotus* individuals and bare soil with sparse vegetation for training the CNN-based model.
- Demonstrating that the use of small datasets for training the CNN-model with transfer learning from ImageNet (i.e., fine-tuning) can lead to satisfactory results that can be further enhanced by including data-augmentation, and specific pre-processing techniques.
- Comparing CNN-based models with OBIA-based methods in terms of performance, user productivity, and transferability to other regions.
- Providing a complete description of the used methodology so that it can be reproduced by other researchers for the classification and detection of this or other shrubs.

From our results, compared to OBIA, the CNN-detection model, in combination with data-augmentation, transfer-learning (fine-tuning) and a custom detection proposals technique, achieves higher precision and balance between recall and precision for detecting *Ziziphus lotus*, on two different regions, one is near and the other is far away from the training region. In addition, the detection process is faster with the CNN-detector than with OBIA, which implies a higher user productivity. Our results also suggest that OBIA-methods and software could be further improved by including CNNs-classifiers ([12]).

This paper is organized as follows. A review of related works is provided in Section 2. A description of the proposed CNN-methodology is given in Section 3. The considered study areas and how the datasets were constructed can be found in Section 4. The experimental results using CNNs and OBIA are provided in Section 5 and finally conclusions are given in Section 6.

2. Related Works

This section reviews the related works on OBIA and CNNs in land cover mapping. Then it explains how OBIA, the state-of-the-art method, is used for the detection of plant species individuals.

2.1. Land Cover Mapping

In the field of remote sensing, land cover mapping has been traditionally performed using pixel-based classifiers or object-based methods ([5]). Several papers have demonstrated that Object-Based Image Analysis (OBIA) methods are more accurate than pixel-based methods, particularly for high spatial resolution images ([3]). In the field of computer vision, object detection in an image is more challenging than scene tagging or classification because it is necessary to determine the image area that contains the searched object. In most object detection works, first a classifier is trained and then it is applied on a number of candidate windows. Recently, deep learning CNNs have started to be used for scene tagging and object detection in remotely-sensed images ([17,20–23]). However, as far as we know, there are no studies in the literature on the use of CNNs for the detection of plant species individuals in remotely-sensed images and any comparison between OBIA and deep CNNs methods.

The existing works that use deep CNNs in remotely-sensed images can be divided into two broad groups. The first group focuses on the classification of high-resolution multi-band imagery (more than three spectral bands) using CNNs-based methods ([20,21,24,25]). Most of these works reported good accuracies on well known annotated hyper-spectral scenes (e.g., the Pavia University image and the Indian pines image [26]).

The second group focuses on the classification or tagging of whole aerial RGB images (commonly called scene classification). These works also reported good accuracies on benchmark databases such as, UC-Merced dataset [27] and Brazilian Coffee Scenes dataset [28] ([22,29]). Both of these datasets contain a large number of manually labeled images. For example, the Brazilian Coffee Scenes dataset contains 50,000 of 64×64-pixel tiles, labeled as coffee (1438) non-coffee (36,577) or mixed (12,989) and UC-Merced dataset contains 2100 256×256-pixel images labeled as belonging to 21 land-use classes, with 100 images corresponding to each class. Several works have reached classification accuracies greater than 95% on these database ([22,29]).

The study that is most similar to ours is ([30]), it addresses the detection of oil-palm trees in agricultural areas using four spectral bands imagery at 0.5×0.5 m spatial resolution via CNNs. Since in plantations oil-palm trees have the same age, shape, size, and are placed at the same distance from each other, the authors could combine the LeNet-based classifier with a very simple detection technique. In addition, the authors used a large number of manually labeled training samples, 5000 palm tree samples and 4000 background samples. Our study is more challenging, because *Ziziphus lotus* is not a crop, it is a wild shrub that has very different shapes, sizes, and intensities of green color, and the surrounding plants and background soil strongly differ across regions. In addition, we will show in this paper that a much smaller training set can also lead to good results.

2.2. OBIA-Based Detection

OBIA-methods represent the state-of-the art in remote sensing for object detection [31], high-resolution land-cover mapping [32,33] and change detection [34]. However, contrarily to CNNs, OBIA-based models are not learnable models, i.e., OBIA can not directly re-utilize the learning from one image into another. The detection is performed from scratch on each new individual image. The OBIA approach is performed in two steps. First, the input image is segmented, and then each segment is assigned to a class by a classification algorithm. A simplistic flowchart of the CNNs- and OBIA-based approaches is illustrated in Figure 1b. The OBIA-detectors used in this study are implemented in eCognition 8.9 software ([6]) and work in two steps as follows:

- **Segmentation step**: First, the input image is segmented using the multi-resolution algorithm ([35]). In this step, the user has to manually optimize and initialize a set of non-dimensional parameters namely: (i) The scale parameter, to control the average image segment size by defining the maximum allowed heterogeneity (in color and shape) for the resulting image objects. The higher the value, the larger the resulting image objects. (ii) The shape versus color parameter, to prioritize homogeneity in color versus in shape or texture when creating the image objects. Values closer to one indicate shape priority, while values closer to zero indicate color priority. (iii) The compactness versus smoothness parameter, to prioritize whether producing compact objects over smooth edges during the segmentation. Values closer to one indicate compactness priority, while values closer to zero indicate smoothness priority ([36]). The results of the segmentation must be validated by analyzing the spatial correspondence between the OBIA-obtained segments and the field-digitized polygons. In this work, the geometric and arithmetic correspondance was analyzed by means of the Euclidean Distance v.2 ([37]).
- **Classification step**: Second, the resulting segments must be classified using: K-Nearest Neighbor (KNN), Random Forest (RF) or Support Vector Machine (SVM) methods. In general, several works have reported that SVM and RF obtain better accuracies ([12,38,39]). For this, the user has to introduce training sample sites for each class. Then, objects are classified based on their statistical resemblance to the training sites. The classification is validated by using an independent set of sample sites. Typically, 30% of the labeled field samples are used for training, and 70% for validation based on a confusion matrix to calculate the commission and omission errors, and the overall accuracy ([40]). Finally, to provide a fair comparison between OBIA and CNNs, we applied the same filtering method called detection proposal.

3. CNN-Based Detection for Shrub Mapping

We reformulate the problem of detecting a shrub species into a two-class problem, where the true class is "*Ziziphus lotus* shrubs" and the false class is "bare soil with sparse vegetation". To build the CNNs-based detection model, (1) we first designed a field-validated training dataset, (2) then we found the most accurate CNNs-classifier by analyzing two networks, ResNet and GoogLeNet, and considering two optimizations, fine-tuning and data-augmentation (3) during the detection process, we compared two options, the sliding-window technique and proposal detection technique to localize *Ziziphus lotus* in the test scenes. A simplistic flowchart of the CNNs- and OBIA-based approaches is illustrated in Figure 1a.
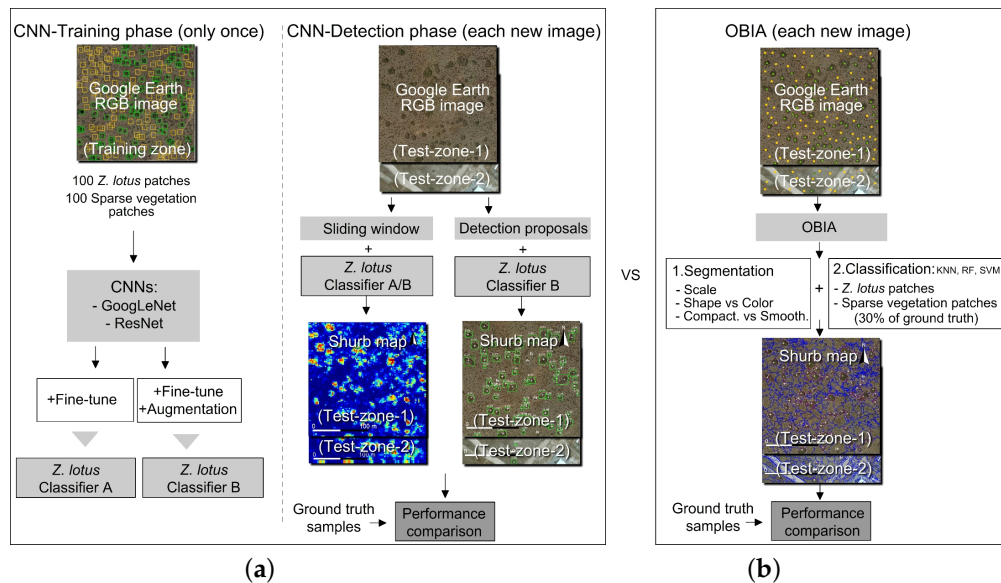
Figure 1. Flowchart of the *Ziziphus lotus* shrub mapping process using (**a**) Convolutional Neural Networks (CNNs) considering two detection approaches: sliding window and detection proposals and (**b**) Object-Based Image Analysis (OBIA). The best performance was obtained by ResNet-based classifier combined with the detection proposal technique.

3.1. Training Phase: CNN-Classifier With Fine-Tuning and Data Augmentation

In this work, we use feed-forward Convolutional Neural Networks (CNNs) for supervised classification, as they have provided very good accuracies in several applications. These methods automatically discover increasingly higher level features from data ([13,41]). The lower convolutional layers capture low-level image features (e.g., edges, color), while higher convolutional layers capture more complex features (i.e., composite of several features).

In this work, we considered the two most accurate CNNs, ResNet ([42]) and GoogLeNet ([43]). ResNet won the first place on the 2015 ILSVRC (ImageNet Large Scale Visual Recognition Competition (ILSVRC)) and is currently the most accurate and deepest CNN available. It has 152 layers and 25.5 million parameters. Its main characteristic with respect to the previous CNNs is that ResNet creates multiple paths through the network within each residual module. GoogLeNet won the first place of the 2014 ILSVRC. GoogLeNet is based on inception v3 and has 23.2 million parameters and 22 layers with learnable weights organized in four parts: (i) the initial segment, made up of three convolutional layers, (ii) nine inception v3 modules, where each module is a set of convolutional and pooling layers at different scales performed in parallel then concatenated together, (iii) two auxiliary classifiers, where each classifier is actually a smaller convolutional network put on the top of the output of an intermediate inception module, and (iv) one output classifier.

Deep CNNs, such as ResNet and GoogLeNet, are generally trained based on the prediction loss minimization. Let *x* and *y* be the input images and corresponding output class labels, the objective of the training is to iteratively minimize the average loss defined as

$$J(w) = \frac{1}{N} \sum_{i=1}^{N} L(f(w; x_i), y_i) + \lambda R(w) \tag{1}$$

This loss function measures how different is the output of the final layer from the ground truth. *N* is the number of data instances (mini-batch) in every iteration, *L* is the loss function, *f* is the predicted output of the network depending on the current weights *w*, and R is the weight decay with the Lagrange multiplier λ. It is worth mentioning that in the case of GoogLeNet, the losses of the

two auxiliary classifiers are weighted by 0.3 and added to the total loss of each training iteration. The Stochastic Gradient Descent (SGD) is commonly used to update the weights.

$$w_{t+1} = \mu w_t - \alpha \Delta J(w_t) \tag{2}$$

where μ is the momentum weight for the current weights w_t and α is the learning rate.

The network weights, w_t, can be randomly initialized if the network is trained from scratch. However, this is suitable only when a large labeled training-set is available, which is expensive in practice. Several previous studies have shown that data-augmentation ([44]) and transfer learning ([45]) help overcoming this limitation.

- **Transfer learning** (e.g., fine-tuning in CNNs). The best analogy for transfer-learning could be the way humans face a new challenge. Humans do not start the learning from scratch, they always use previous knowledge to build new one. Transfer-learning consists of re-utilizing the knowledge learnt from one problem to another related one ([46]). Applying transfer learning with deep CNNs depends on the similarities between the original and new problem and also on the size of the new training set. In deep CNNs, transfer learning can be applied via fine-tuning, by initializing the weights of the network, w_t in Equation (2), with the pre-trained weights from a different dataset.

  In general, fine-tuning the entire network (i.e., updating all the weights) is only used when the new dataset is large enough, otherwise, the model could suffer overfitting especially among the first layers of the network. Since these layers extract low-level features, e.g., edges and color, they do not change significantly and can be utilized for several visual recognition tasks. The last learnable layers of the CNN are gradually adjusted to the particularities of the problem and extract high level features.

  In this work, we have used fine-tuning on ResNet and GoogleNet. We initialized the used CNNs with the pre-trained weights of the same architectures on ImageNet dataset (around 1.28 million images over 1000 generic object classes) ([13]).

- **Data-augmentation**, also called data transformation or distortion, is used to artificially increase the number of samples in the training set by applying specific deformations on the input images, e.g., rotation, flipping, translation, cropping, or changing the brightness of the pixels. In this way, from a small number of initial samples, one can build a much larger dataset of transformed images that still are meaningful for the case study. The set of valid transformations that improves the performance of the CNN-model depends on the particularities of the problem. Several previous studies have demonstrated that increasing the size of the training dataset using different data-augmentation techniques increases performance and makes the learning of CNNs models robust to changes in scales, brightness and geometrical distortions [44,47].

3.2. Detection Phase

To obtain an accurate detection in a new image, different from the images used for training the CNN-classifier, we analyzed two approaches:

- **Sliding window** is an exhaustive technique frequently used for detection. It is based on the assumption that all the areas of the input-image are possible candidates to contain an object class. This search across the input-image can generate around 10^6 candidate windows. The detection task consists of applying the obtained CNN-classifier at all locations and scales of the input image. The sliding window approach is an exhaustive method since it considers a very large number of candidate windows of different sizes and shapes across the input image. The classifier is then run on each one of these windows. To maximize the detection accuracy, the probabilities obtained from different window sizes can be assembled into one heatmap. Finally, probability heatmaps are usually transformed into classes using a thresholding technique, i.e., areas with

probabilities higher than 50% are usually classified as the true class (e.g., *Ziziphus lotus*) and areas with probabilities lower than 50% as background (e.g., bare soil with sparse vegetation).

- **Detection proposals** are techniques that employ different selection criteria to reduce the number of candidate windows, thereby avoiding the exhaustive sliding window search ([48]). These techniques can also help to improve the detection accuracy and execution time. In general, detection proposals methods determine the set of pre-processing techniques that provides the best results. This set depends on the nature of the problem and the object of interest. From the multiple techniques that we explored, the ones that provided the best detection performance were: (i) Eliminating the background using a threshold based on its typical color or darkness (e.g., by converting the RGB image to gray scale, grays lighter than 100 digital level corresponded to bare ground). (ii) Applying an edge-detection method that filters out the objects with an area or perimeter smaller than the minimum size of the target objects (e.g., the area of the smallest *Ziziphus lotus* individual in the image, around 22 m^2).

4. Study Areas and Datasets Construction

This section describes the study areas and provides full details on how the training and test sets were built using Google Earth^{TM} images. We consider the challenging problem of detecting the *Ziziphus lotus* shrubs, since it is considered to be the key species of an ecosystem of priority conservation in the European Union (habitat 5220* habitat of 92/43/EEC Directive). During recent decades, several studies reported that *Ziziphus lotus* is declining in SE Spain, Sicily and Cyprus ([49]). In Europe, the largest population occurs in the Cabo de Gata-Níjar Natural Park (SE Spain), where an increased mortality of individual shrubs of all ages was observed in the last decade ([50,51]).

4.1. Study Areas

In this study, we considered three zones: one training-zone, for training the CNN-model, and two test zones (labeled as test-zone-1 and test-zone-2) for testing and comparing the performance of both CNN- and OBIA-based models.

- **The training-zone** used for training the CNN-based model. This zone is located in Cabo de Gata-Níjar Natural Park, 36°49′43″ N, 2°16′22″ W, in the province of Almería, Spain (Figure 2). The climate is semi-arid Mediterranean. The vegetation is scarce and patchy, mainly dominated by large *Ziziphus lotus* shrubs surrounded by a heterogeneous matrix of bare soil and small scrubs (e.g., *Thymus hyemalis*, *Launea arborescens* and *Lygeum spartum*) with low coverage ([49,52]). *Ziziphus lotus* forms large hemispherical bushes with very deep roots and 1–3 m tall that trap and accumulate sand and organic matter building geomorphological structures, called nebkhas, that constitute a shelter micro-habitat for many plant and animal species ([19,49,53]).
- **Test-zone-1** and **test-zone-2** belong to two different protected areas. Test-zone-1 is located 1.5 km west from the training-zone, 36°49′28″ N, 2°17′28″ W. Test-zone-2 is located in Rizoelia National Forest Park in Cyprus, 34°56′09″ N, 33°34′26″ E (Figure 2). These two test-zones are used for comparing the performance between CNNs and OBIA for detecting *Ziziphus lotus*.
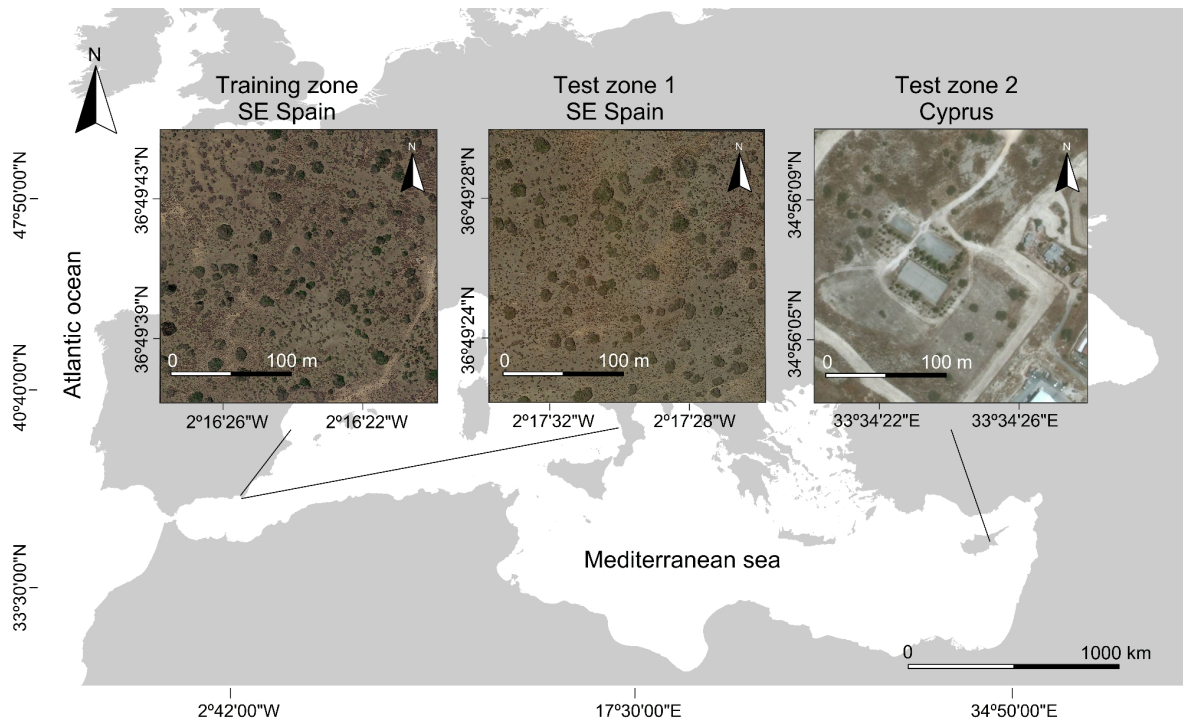
Figure 2. Localization of the three study areas used in this work: Training-zone and Test-zone-1 in Cabo de Gata-Níjar Natural Park (Spain), and Test-zone-2 in Rizoelia National Forest Park (Cyprus). The three images are 230 × 230 m with a native resolution in Google Earth^TM of 0.5 m per pixel (but downloaded as 1900 × 1900 pixel images). *Ziziphus lotus* shrubs can be seen in the three images. The used projection was geographic with the WGS84 Datum.

4.2. Datasets Construction

4.2.1. Satellite-Derived Orthoimages from Google Earth^TM

The satellite RGB orthoimages used in this work were downloaded from Google Earth^TM in European Petroleum Survey Group (EPSG) 4326 using a geographic coordinate system with the WGS84 datum. The scenes of the three areas, training-zone, test-zone-1 and test-zone-2, have an approximate size of 230 × 230 meters. The images were downloaded at two Google Earth's zoom levels: (i) the closest zoom level (i.e., 19) to the native resolutions (see below), that resulted in scenes of 456 × 456 pixels with a resolution of 0.5 m, (ii) the maximum available zoom level in their area (i.e., 21), that resulted in scenes of 1900 × 1900 pixels with an increased resolution of 0.12 m due to the smoothing applied by Google Earth. Since all results always showed better accuracies (3% better on average) with the images at 0.12 m resolution, we did not include the results at 0.5 m in the manuscript to save space. The characteristics of the satellite images used by Google to produce the orthoimages of the three areas were:

- Training-zone and test-zone-1 images (SE Spain) were captured by Worldview-2 satellite under 0% cloud cover on the 30 June 2016, with an inclination angle of 12.7°. The multispectral RGB bands have a native spatial resolution of 1.84 m, but they are pansharpened to 0.5 m by Worldview-2 using the panchromatic band. The RGB bands cover the following wavelength ranges: Red: 630–690 nm, Green: 510–580 nm, Blue: 450–510 nm.
- Test-zone-2 image (Cyprus) was captured by Pléiades-1A satellite under 0.1% cloud cover on the 8 July 2016, with an inclination angle of 29.2°. The multi spectral RGB bands have a native spatial resolution of 2 m, but they are pansharpened to 0.5 m by Pléiades-1A using the panchromatic band.

They RGB bands cover the following wavelength ranges: Red: 600–720 nm, Green: 490–610 nm, Blue: 430–550 nm.

4.2.2. Dataset for Training OBIA and for Ground Truthing

We addressed the *Ziziphus lotus* detection problem by considering two classes, (1) *Ziziphus lotus* shrubs class and (2) Bare soil with sparse vegetation class. In OBIA, the training dataset consisted of a set of georeferenced points from the same scene that we want to classify and covering the two targeted classes. Conversely, for CNNs, the training dataset consisted of two sets of images that contained each class of interest, but these images do not have to belong to same scene that we aim to classify, allowing for transferability to other regions, which is an advantage of CNNs over OBIA methods.

- In test-zone-1, 74 *Ziziphus lotus* individual were identified in the field. The perimeter of each was georeferenced in the field with a differential GPS, GS20, Leica Geosystems, Inc. From the 74 individual shrub, 30% (22 individual shrubs) were used for training and 70% (52 individuals) for validation in the OBIA method. Images containing patches from all 74 individual shrubs were used for validation in the CNN method (see below).
- In test-zone-2, 40 *Ziziphus lotus* individuals were visually identified in Google Earth by the authors using the vegetation maps and descriptions provided by local botany experts ([54]). These individuals were also validated in the field by one of the co-authors, J. Cabello. All 40 individual shrubs were used for validation in both the OBIA and CNN methods.

In both test zones, the same number of *Ziziphus lotus* individuals (74 and 40, respectively) was georeferenced for the Bare soil with sparse vegetation class (Table 1).

4.2.3. Training Dataset for the CNN-Classifier

The design of the training dataset is key to the performance of a good CNN classification model. From the 82 *Ziziphus* individuals georeferenced by botanic experts in the training-zone, we identified 100 80 × 80-pixel image patches containing *Ziziphus lotus* shrubs and 100 images for Bare soil with sparse vegetation. Examples of the labeled classes can be seen in Figure 3. We distributed the 100 images of each class into 80 images for training and 20 images for validating the obtained CNNs classifiers, as summarized in Table 1.

Table 1. Training and testing datasets for both CNN and OBIA used for mapping *Ziziphus lotus* shrubs. Bare soil: Bare soil and sparse vegetation; Img: 80 × 80-pixel image patches; Poly: digitized polygons.

	CNN Classifier		OBIA Classifier		Accuracy	
Class	**Training**	**Validation**	**Training**		**Assessment**	
	Training-Zone		**Test-Zone-1**	**Test-Zone-2**	**Test-Zone-1**	**Test-Zone-2**
Ziziphus	80 img	20 img	22 poly	0 poly	52 poly	40 poly
Bare soil	80 img	20 img	22 poly	0 poly	52 poly	40 poly

Examples of *Ziziphus lotus* shrubs　　　Examples of Bare soil and sparse vegetation
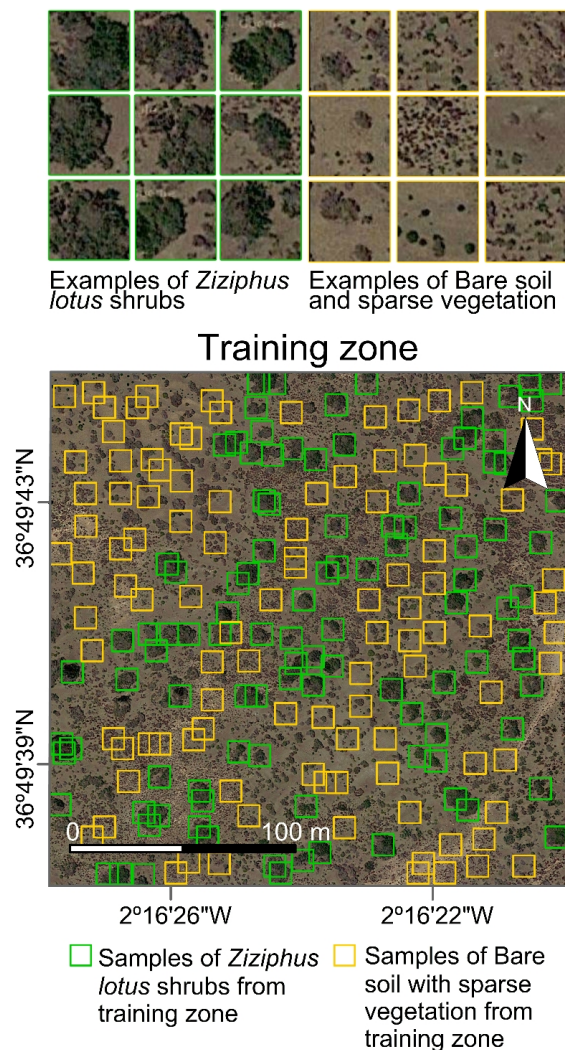
Training zone

Figure 3. The two top panels show examples of the 80×80-pixel image patches used to build the training dataset for the CNN model: (**left**) patches of *Ziziphus lotus* class, (**right**) patches of Bare soil with sparse vegetation class. The bottom panel shows the training-zone dataset with 100 *Ziziphus lotus* patches labeled with a green contour and 100 Bare soil and sparse vegetation patches labeled with yellow contour.

5. Experimental Evaluation and Discussions

This section is organized in three parts. The first part describes the steps taken to develop the best CNN-based shrub detector. For this, we used GoogLeNet and improved its baseline detection results by applying transfer-learning (fine-tuning) and data-augmentation under the sliding window approach. Then we further improved the detection by using a more powerful Network, ResNet, combined with a custom detection proposal technique.

The second part describes the steps taken to develop the best OBIA-based classification of *Ziziphus lotus* shrubs. For this, we compared three classification algorithms: KNN, RF and SVM. To ensure a fair comparison with CNNs, we re-utilized the segmentation and classification ruleset from test-zone-1 in test-zone-2, and the same threshold filtering used in the detection proposal for CNNs.

The third part provides a comparison between GoogLeNet with detection proposals, ResNet with detection proposals, OBIA-KNN, OBIA-RF, and OBIA-SVM. For the evaluation and comparison of accuracies, we used three metrics, *precision* (also called positive predictive value, i.e., how many

detected *Ziziphus lotus* are true), *recall* (also known as sensitivity, i.e., how many actual *Ziziphus lotus* were detected), and *F1 measure*, which evaluates the balance between *precision* and *recall*. Where

$$precision = \frac{True\ Positives}{True\ Positives + False\ Positives},$$

$$recall = \frac{True\ Positives}{True\ Positives + False\ Negatives},$$

and

$$F1\ measure = 2 \times \frac{precision \times recall}{precision + recall}$$

5.1. Finding the Best CNN-Based Detector

For the experiments with GoogLeNet and ResNet-based models, we have used the open source software library Tensorflow ([18]). For training CNNs, the image patches are resized from 80×80-pixels to 299×299 by GoogLeNet and to 224×224 by ResNet. Such rescaling is due to the fact that the architecture of all the layers of GoogLeNet and ResNet are adapted according to these input sizes, independently from the original resolution of the input images.

5.1.1. CNN Training With Fine-Tuning and Data-Augumentation

To improve the accuracy and reduce overfitting we (i) used fine-tuning by initializing the evaluated models with the pre-trained weights of ImageNet, and (ii) applied data augmentation techniques to increase the size of the dataset from 100 to 6000 images. In particular, for data-augmentation we applied:

- Random scale: increases the scale of the image by a factor picked randomly in [1 to 10%]
- Random crop: crops the image edges by a margin in [0 to 10%]
- Flip horizontally: randomly mirrors the image from left to right.
- Random brightness: multiplies the brightness of the image by a factor picked randomly in [0, 10].

To show the impact of data-augmentation on the performance of the detection, we analyzed the results of the GoogleNet-based classifier combined with the sliding window technique with and without data-augmentation. The results are summarized in the two first rows of Table 2. As we can observe, using only fine-tuning, the GoogLeNet-based model reached relatively good performance (77.64% precision, 89.18% recall and 83.01% F1). Adding data-augmentation further increased the performance (90.28% precision, 87.83% recall and 89.04% F1). This performance comparison was performed under the sliding-window detection approach (see next section).

5.1.2. Detection Using GoogLeNet Under the Sliding Window Approach

This section evaluates the performance of the CNN-based classifier under the sliding window approach. To assess the ability of CNNs model to detect *Ziziphus lotus* shrubs in Google Earth images, we applied the trained CNN classifiers across the entire scene of test-zone-1 by using the sliding window technique. Since the diameter of the smallest *Ziziphus lotus* individual georeferenced in the field was 4.6 m (38 0.12-m-pixels) and the largest individual in the region had a diameter of 47 m (385 0.12-m-pixels), we evaluated a range of window sizes from 38×38 to 385×385 pixels and a horizontal and vertical sliding step of about 70% the size of the sliding window, e.g., 27×27 pixels for the 38×38 sliding window, and 269×269 pixels for 385×385 sliding window.

Table 2. GoogLeNet(with and without data-augmentation) and ResNet-detection results for *Ziziphus lotus* shrubs mapping in Test-zone-1, under the sliding window approach and using a detection-proposals approach. Accuracies are expressed in terms of true positives (TP), false positives (FP), false negatives (FN), precision, recall, and F1 measure. The highest accuracies are highlighted in bold.

Detection Model	TP	FP	FN	Precision	Recall	F1
GoogLeNet (test-zone-1) +fine-tuning under **sliding window**	65	12	9	77.64%	89.18%	83.01%
GoogLeNet (test-zone-1) +fine-tuning +augmentation under **sliding window**	65	7	9	90.28%	87.83%	89.04%
GoogLeNet (test-zone-1) +fine-tuning +augmentation under **detection proposals**	69	1	5	98.57%	93.24%	95.83%
ResNet (test-zone-1) +fine-tuning +augmentation under detection proposals	**69**	**0**	**5**	**100.00%**	**93.24%**	**96.50%**

The performance of the GoogLeNet-based detector on the 1900 × 1900 pixels image corresponding to test-zone-1 is shown in Table 3 and the corresponding heatmap to each window size are illustrated in Figure 4. The best performance, highest recall and F1-measure and high precision, were obtained for a window size of 64 × 64 pixels. The time needed to perform the detection process using this window size was 291 min. This represents the execution time that would be required for *Ziziphus lotus* shrub detection on any new input image of the same dimensions, which is very time consuming to be used in larger regions or across the entire range distribution of the species along the Mediterranean region. To reduce the execution time, we next applied the detection-proposal pre-processing technique to reduce the number of candidate regions.

Table 3. CNN-detection results in Test-zone-1 at different sliding window sizes. Accuracies are expressed in terms of true positives (TP), false positives (FP), and false negatives (FN), precision, recall, F1-measure, and execution time of the detection process.

| Win. Size (Pixels)	Total # of Win.	TP	FP	FN	Precis. (%)	Recall (%)	F1-Meas. (%)	Time (min)
385×385	196	31	18	41	63.27	43.06	51.24	6.0
194×194	961	34	7	38	82.93	47.22	60.18	29.4
129×129	2209	42	6	30	87.50	58.33	70.00	67.6
97× 97	4096	59	6	13	90.77	81.94	86.13	125.4
77×77	5929	59	5	13	**92.19**	81.94	86.76	181.5
64×64	9506	65	7	7	90.28	90.28	**90.28**	291.0
55×55	13,340	65	12	7	84.42	90.28	87.25	408.4
48×48	17,292	68	16	4	80.95	94.44	87.18	529.3
42×42	22,200	70	17	2	80.46	97.22	88.05	679.6
38× 38	27,888	71	39	1	64.55	**98.61**	78.02	853.7
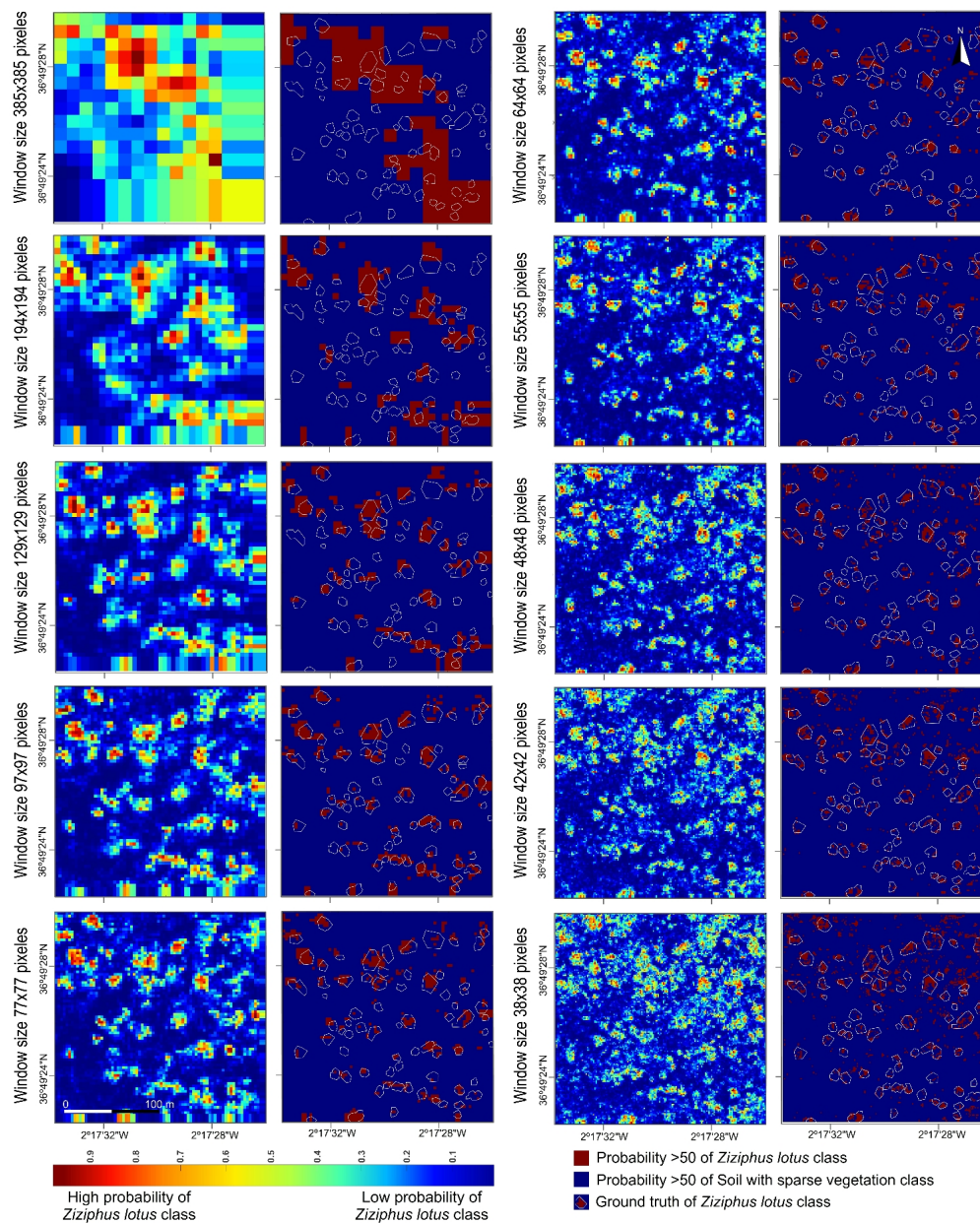
Figure 4. Maps showing the probability of *Ziziphus lotus* presence according to the CNN-classifier trained with fine-tuning and data-augmentation and applied on different sliding-window sizes from 38×38 to 385×385 pixels in Test-zone-1. The first and third columns show the heatmaps of the probability of *Ziziphus lotus* presence, and the second and fourth columns show the corresponding binary maps after applying a threshold of probability greater than 50%. The white polygons correspond to the ground-truth perimeter of each individual georeferenced in the field with a differential GPS.

5.1.3. Detection Using GoogLeNet and ResNet under a Detection Proposals Approach

This section evaluates the performance of GoogLeNet- and ResNet-based classifiers under the detection-proposal pre-processing technique. To optimize the CNN-detection accuracy and execution time, we analyzed several pre-processing techniques to generate better and faster candidate regions than with the sliding window approach. The selection of the set of pre-processing techniques that provides the best results depends on the nature of the problem and the object of interest. From the multiple techniques explored, the ones that improved the performance of the CNN detectors in this work were: (i) Eliminating the background using a threshold based on the high albedo (light color)

of the bare soil. The used detection proposals technique is illustrated in Figure 5. For this, we first converted the RGB image to gray scale and then created a binary mask-band to select only those pixels darker than 100 over 256 digital levels of gray, which was the average level of gray of the field georeferenced point of bare soil in the training-zone. (ii) Applying an edge-detection method to the previously created mask-band to select only clusters of pixels with an area greater than 180 pixels (21.6 m^2), which approximately was the size of the smallest *Ziziphus lotus* individual georeferenced in the training-zone. After applying this detection proposals technique, the number of candidate image patches to pass to the CNN detectors was 78 for test-zone-1 and 53 for test-zone-2, which significantly decreased the detection computing time.
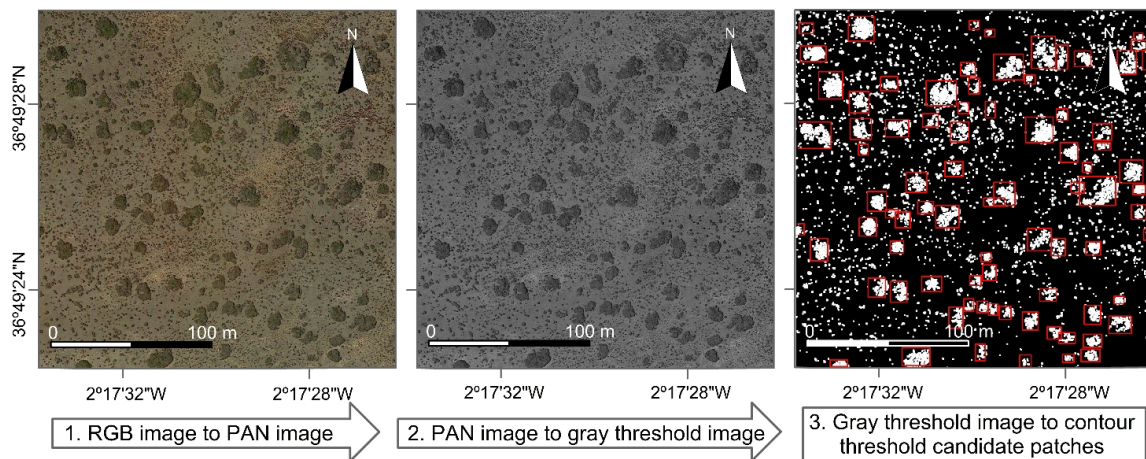

Figure 5. The used detection proposals technique consisted of, first, converting the three band image into one gray-scale band image (PAN), second, converting the gray-scale image into a binary image based on a 100 over 256 digital value threshold, and third, detecting *Ziziphus lotus* shrubs only in pixels with a digital value greater than 100. The 78 candidate patches identified in Test-zone-1 are labeled with red contour in the right panel (the 53 candidates in Test-zone-2 are not shown.)

The results of GoogLeNet- and ResNet-based detection model using the aforementioned proposal method, considering fine-tuning and data-augmentation on test-zone-1 are summarized in the last two rows of Table 2. ResNet-based classifier combined with the detection proposals technique together with fine-tuning and data-augmentation achieved the best performance. It further improved the precision and F1 of GoogLeNet-detector under the same conditions.

5.2. Finding the Best OBIA-Detector

For the experiments with the OBIA methods, we used the privative eCognition software ([6]). To determine the best segmentation, we iteratively tried all the possible combinations between the three customizable parameters: scale, ranging in [80, 160] at intervals of 5, shape and compactness, ranging in [0.1, 0.9] at intervals of 0.1. The best segmentation parameters were: scale = 110, shape = 0.3, compactness = 0.8. To obtain the best detection results using OBIA-method, we considered three classifiers, KNN, RF and SVM. The best classification configuration for KNN, RF and SVM, i.e., brightness, red band, green band, blue band and gray level co-occurrence matrix (GLCM mean) features, was determined using the Separability and Threshold tool [55]. An exhaustive search of the best configuration implied the evaluation of 1296 combinations, each segmentation test took 18 s, and each classification test took around 10 s. The whole optimization and detection process using OBIA took around 10 h for Test-zone-1. Normally, OBIA requires the user to provide training points of each input class located within the scene (test zone) we want to classify. However, to ensure a fair comparison with CNNs, we re-utilized the OBIA segmentation and classification configuration

"learned" from test-zone-1 into test-zone-2. The results of OBIA-based detection using KNN, RF and SVM are summarized in Table 4. The best results were obtained with the SVM method.

Table 4. A comparison between the best CNN-detector, ResNet-detector and OBIA, on test-zone-2, in terms of true positives (TP), false positives (FP), false negatives (FN), precision , recall and F1_measure. The highest values are highlighted in bold.

Detection Model	TP	FP	FN	Precision	Recall	F1
ResNet-based classifier (test-zone-1) +fine-tuning +augmentation under detection proposals	69	0	5	**100.00%**	93.24%	**96.50%**
OBIA-KNN (test-zone-1)	66	9	8	88.00%	89.18%	88.59%
OBIA-Random Forest (test-zone-1)	67	6	7	91.78%	90.54%	91.15 %
OBIA-SVM (test-zone-1)	72	9	2	88.88%	**97.29%**	92.90%
ResNet-based classifier (test-zone-2) +fine-tuning +augmentation under detection proposals	38	3	2	**92.68%**	**95.00%**	**93.38%**
OBIA-KNN (test-zone-2)	21	4	19	84.00%	52.50%	64.61%
OBIA-Random Forest (test-zone-2)	27	6	13	81.81%	67.5%	73.97%
OBIA-SVM (test-zone-2)	29	6	11	82.85%	72.50%	77.33%

5.3. CNN-Detector Versus OBIA-Detector

We tested the CNN-detector on two images with different radiometric and environmental characteristics, test-zone-1 (SE Spain) and test-zone-2 (Cyprus), captured by different satellites. The performance results of ResNet-model and OBIA-method in test-zone-1 (SE Spain) and test-zone-2 (Cyprus) are summarized in Table 4. As we can observe from this table, CNN-based detection model achieved significantly better detection results than OBIA on both test zones. On test-zone-1, CNN achieved higher precision, 100.00% versus 88.88%, and F1-measure, 96.50% versus 92.90%, though slightly lower recall, 93.24% versus 97.29%, than OBIA. More noticeable, on test-zone-2, CNN achieved significantly better precision 92.68% versus 82.85%, recall 95.00% versus 72.50%, and F1-measure 93.38% versus 77.33% than OBIA.

The shrub detection maps of CNN and OBIA, on test-zone-1, are shown in Figure 6(a) and (b), and on test-zone-2 are shown in Figure 6c,d. In general, both OBIA and CNN successfully detected the majority of shrubs in test-zone-1 and test-zone-2. In test-zone-1, from 74 true shrubs, OBIA detected 72 true positives while CNN detected 69 true positives. However, OBIA produced 9 false positives whereas CNN did not produce any false positive. More importantly, in test-zone-2, from 40 true shrubs, OBIA detected 29 true positives while CNN detected 38 true positives. In addition, OBIA produced 6 FP while CNN produced only 3 FP.
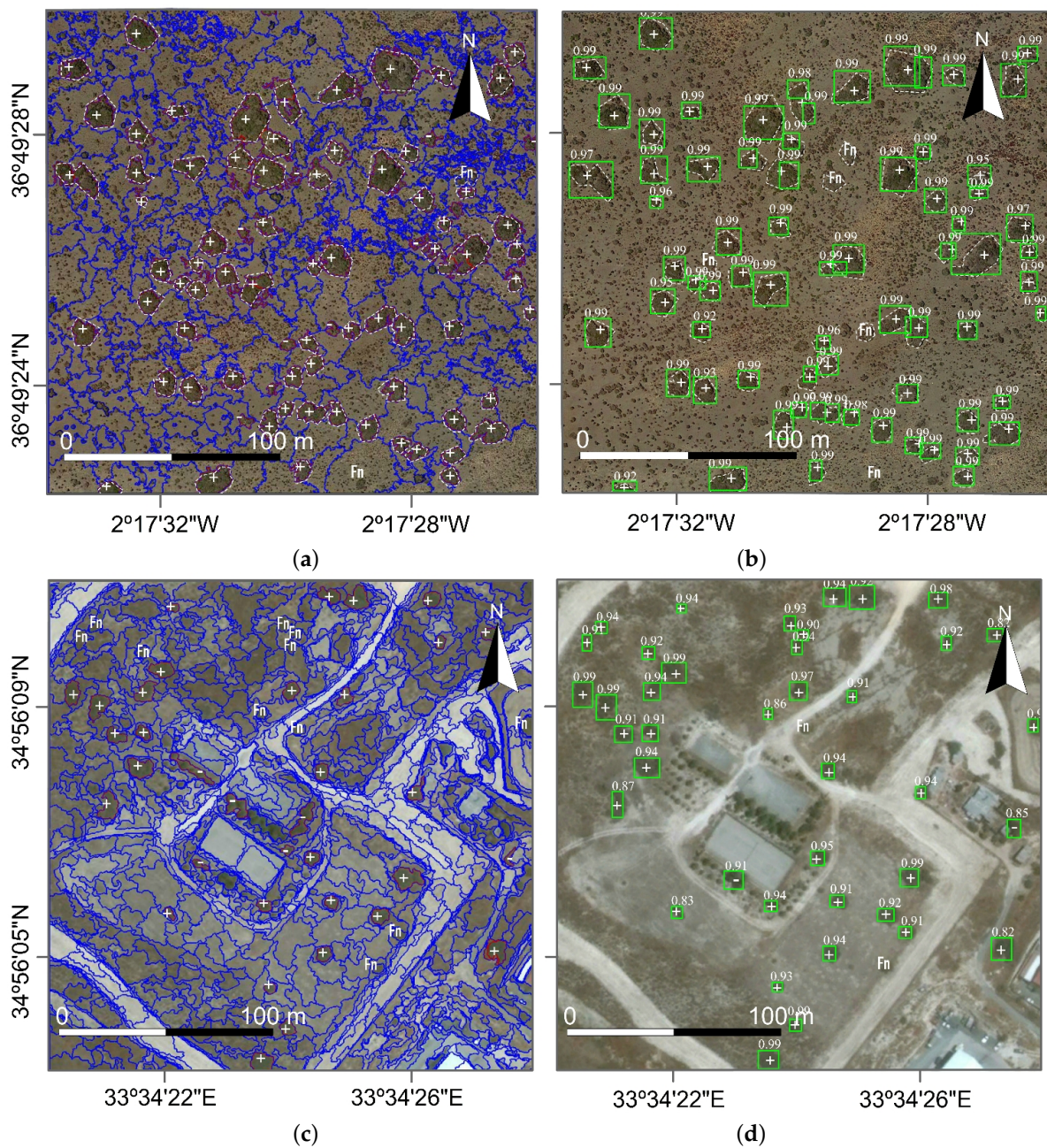
Figure 6. Shrub detection maps obtained with OBIA-based model in test-zone-1 (**a**) and test-zone-2 (**c**) and CNN-based model in test-zone-1 (**b**) and test-zone-2 (**d**). The symbols (+), (-) and (Fn) stand for true positive, false positive, and false negative, respectively. The values on the top of the bounding boxes in (2) and (4) show the probabilities, calculated by ResNet-detector, of having a *Ziziphus lotus* shrub. (**a**) OBIA detection on test-zone-1; (**b**) CNN detection on test-zone-1; (**c**) OBIA detection on test-zone-2; (**d**) CNN detection on test-zone-2.

Despite such good results, both OBIA and CNN produced under-segmentations, i.e., duplicated detections. In test-zone-1, both methods produced 3 under-segmentations. In test-zone-2, OBIA and CNN produced 1 and 0 under-segmentations, respectively. The under-segmented shrub individuals by OBIA had areas larger than 140 m^2 while the under-segmented shrub individuals by CNNs had areas between 68 m^2 and 326 m^2. This under-segmentation occurred mainly in test-zone1 and it was due to the highly heterogeneous shape and texture of some shrubs that are over-simplified in the segmentation step in OBIA and in the detection proposals step in CNNs when it converts RGB to black and white.

Such anomalous heterogeneity of test-zone-1 could be explained by the bad physiological status of the *Ziziphus lotus* shrubs due to marine intrusion and increase in defoliators [19,56]. In test-zone-2, neither OBIA nor CNN produced under-segmentation probably due to their better physiological status and more homogenous shape and texture than in test-zone-1.

A closer look at the shrubs classified as false negatives(FN) (2 FN by OBIA and 5 FN by CNNs in zone-test-1 in Figure 6) showed that those shrub individuals were in a bad health status with large extension of bare sand in their interior and with lower intensity of green color, which made them to be more easily confused with the "bare soil and sparse vegetation" class. In test-zone-2, OBIA produced 11 FN while CNN only 2 FN. One possible explanation could be the low transferability of the OBIA method, since it requires not only similar image characteristics between the learning and testing zones, but also similar characteristics of the target objects (e.g., similar shape and size). The *Z. lotus* individuals of the training zone were larger (minimum size of 46 m^2) than the individuals of test-zone-2 (minimum size of 20 m^2), which probably biased the learning on size of the OBIA-based model. Subsequently, this bias could have caused the OBIA-based model to commit more false negatives on smaller individuals in the test zone 2. On the contrary, CNNs are more robust to this problem and are highly transferable. Another possible explanation could be that OBIA is more sensitive to the radiometric difference between test-zone-1 and test-zone-2. In test-zone-2, the RGB bands were captured by Pléiades-1A satellite at a multispectral Ground Sample Distance (GSD) of 2 m. However, in test-zone-1, the RGB bands were captured by WorldView-2 at a slightly coarser multispectral GSD of 1.84 m. In both cases, RGB images were pansharpened to 0.5 m. Despite the coarser GSD of Pléiades-1A, its radiometric quality is very homogeneous, with low noise level and no saturation effects, which would compensate the small difference in GSD [57].

Overall, accuracy results were slightly better for test-zone-1 than for test-zone-2. In addition to the effect of the spatial resolution commented above, in test-zone-1 *Ziziphus lotus* does not coexist with similar shrubs in terms of size, phenology, shape and color. However, in test-zone-2 some trees do coexist, whose presence may affect the detection accuracies. In those cases, the learning of the CNN-model can be improved by using more spectral bands or temporal information, e.g., including Near Infrared, Digital Surface Model (DSM) or the seasonal NDVI dynamics [51,58].

Deep CNNs learnt and performed better on higher resolution images. This also occurs when the image spatial resolution is artificially increased by rendering, as occurs in Google Earth images. Indeed, to explore the effect of the Google Earth rendering on the CNNs performance, we analyzed a representative set of images at the native satellite spatial resolution, 0.5 m (zoom = 19), and with the downscaled resolution available in Google Earth at 0.12 m (zoom = 21). We found that CNNs performed better (accuracy of 97.73 ± 0.51) on the downscaled images with 0.12m per pixels than on the native resolution ones (accuracy of 95.23 ± 0.32) with 0.5m per pixel. This also implies that, if CNNs are trained on high-resolution images, they will progressively lose performance on low-resolution images since the shape and color maybe deteriorated.

In terms of user productivity, the training of ResNet-classifier was performed only once and took 8 min 22 s on a laptop with Intel(R) Core(TM) i5 CPU running at 2.40 GHz and 4 GB RAM. In the test phase, also called deployment phase, executing ResNet-classifier together with the detection proposals technique on the same laptop for test-zone-1 and test-zone-2 took 58.48 and 26.4 s, respectively. Whereas, finding the best configuration for OBIA on each test-zone took 10 h on the same laptop. Applying the obtained CNN-detector to any new image of similar sizes will take seconds; however, applying OBIA to a new image will take several hours. This execution can be partially reduced by using semi-automatic tools to estimate scale parameters ESP v.2 [59]. In summary, our results show that the user becomes more productive with CNNs than with OBIA and reaching higher accuracies.

6. Conclusions

In this work, we explored, analyzed and compared two detection methodologies for shrub mapping, the OBIA-based approach and the CNNs-based approach. We used a challenging case study, mapping *Ziziphus lotus*, the dominant shrub in a habitat of priority conservation interest in Europe.

Our experiments demonstrated that the ResNet-based classifier with transfer learning from ImageNet and data augmentation, together with a specific detection-proposal pre-processing technique provided better results than the state-of-the-art OBIA-based methods. In addition, an important advantage of the CNN-based detector is that it required less human supervision than OBIA, can be trained using a relatively small number of samples, and can be easily transferable to other regions or scenes with different characteristics, e.g., color, extent, light, background, or size and shape of the target objects. The lack of direct transferability was an important limitation of OBIA methods since, once calibrated for one image, the OBIA settings are not directly portable to other images (e.g., to different areas, extensions, radiometric calibrations, background color, spatial and spectral resolutions, or different sizes or shapes of the target objects).

A natural conclusion of this work is that including CNN-models as classifiers in OBIA software, e.g., a ResNet-classifier in eCognition, could make the users take advantage from the benefits of both methods, e.g., OBIA segmentation to quantify areas and CNNs for detection and classification.

Finally, the proposed CNN-based approach is based on open-source software and uses easily available Google Earth images (subject to Terms of Service), which can have huge implications for land-cover mapping and derived applications. Our CNN-based approach could be systematized and reproduced for a wide variety of detection problems. For instance, this model could be extended to a larger number of classes of shrub and tree species by including more spectral or temporal information. In any case, our CNN-based approach could support the detection and monitoring of trees and arborescent shrubs in general, which has a huge relevance for biodiversity conservation and for reducing uncertainties in carbon accounting worldwide ([60,61]). The presence of scattered trees have been recently highlighted as keystone structures capable of maintaining high levels of biodiversity and ecosystem services provision in open areas ([62]). Global initiatives could greatly benefit from CNNs, such as those recently implemented by the United Nations Food and Agricultural Organization ([60]) to estimate the overall extension of forests in drylands biomes, where they used the collaborative work of hundreds of people that visually explored hundreds of VHR images available from Google Earth to detect the presence of forests in drylands. The uncertainties in such initiatives ([61,63,64] could be decreased following our approach to build a CNN-based tree mapper). CNN-based tree and shrub detectors could serve to produce global characterizations of ecosystem structure and population abundance as part of the satellite remote sensing essential biodiversity variables initiative ([65]).

Author Contributions: E. Guirado and S. Tabik conceived and performed the experiments. E. Guirado, S. Tabik, and D. Alcaraz-Segura designed the experiments and wrote the paper. J. Cabello and F. Herrera provided improvements on the methodology and experiments, reviewed and edited the manuscript.

Conflicts of Interest: The authors declare no conflict of interest. The founding sponsors had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, and in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

CNNs   Convolutional Neural Networks
OBIA   Object-Based Image Analysis

References

1. Congalton, R.; Gu, J.; Yadav, K.; Thenkabail, P.; Ozdogan, M. Global land cover mapping: A review and uncertainty analysis. *Remote Sens.* **2014**, *6*, 12070–12093.
2. Rogan, J.; Chen, D. Remote sensing technology for mapping and monitoring land-cover and land-use change. *Prog. Plan.* **2004**, *61*, 301–325.
3. Blaschke, T. Object based image analysis for remote sensing. *ISPRS J. Photogramm. Remote Sens.* **2010**, *65*, 2–16.
4. Li, X.; Shao, G. Object-based land-cover mapping with high resolution aerial photography at a county scale in midwestern USA. *Remote Sens.* **2014**, *6*, 11372–11390.
5. Pierce, K. Accuracy optimization for high resolution object-based change detection: An example mapping regional urbanization with 1-m aerial imagery. *Remote Sens.* **2015**, *7*, 12654–12679.
6. Ecognition. Available online: http://www.ecognition.com (accessed on 5 May 2017)
7. Knoth, C.; Nüst, D. Reproducibility and Practical Adoption of GEOBIA with Open-Source Software in Docker Containers. *Remote Sens.* **2017**, *9*, 290.
8. Teodoro, A.; Araújo, R. Exploration of the OBIA methods available in SPRING noncommercial software to UAV data processing. In Proceedings of the Earth Resources and Environmental Remote Sensing/GIS Applications V, Amsterdam, The Netherlands, 10 October 2014; Volume 9245.
9. Pal, M. Random forest classifier for remote sensing classification. *Int. J. Remote Sens.* **2005**, *26*, 217–222.
10. Heumann, B.W. An object-based classification of mangroves using a hybrid decision tree—Support vector machine approach. *Remote Sens.* **2011**, *3*, 2440–2460.
11. Feng, Q.; Liu, J.; Gong, J. UAV remote sensing for urban vegetation mapping using random forest and texture analysis. *Remote Sens.* **2015**, *7*, 1074–1094.
12. Ma, L.; Li, M.; Ma, X.; Cheng, L.; Du, P.; Liu, Y. A review of supervised object-based land-cover image classification. *ISPRS J. Photogramm. Remote Sens.* **2017**, *130*, 277–293.
13. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet Classification with Deep Convolutional Neural Networks. In Proceedings of the 25th International Conference on Neural Information Processing Systems, Lake Tahoe, NV, USA, 3–6 December 2012; Curran Associates Inc.: Red Hook, NY, USA, 2012; Volume 1, pp. 1097–1105.
14. Le, Q.V. Building high-level features using large scale unsupervised learning. In Proceedings of the 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, Vancouver, BC, Canada, 26–31 May 2013; pp. 8595–8598.
15. Hinton, G.; Deng, L.; Yu, D.; Dahl, G.; Mohamed, A.; Jaitly, N.; Senior, A.; Vanhoucke, V.; Nguyen, P.; Sainath, T.; et al. Deep Neural Networks for Acoustic Modeling in Speech Recognition: The Shared Views of Four Research Groups. *IEEE Signal Process. Mag.* **2012**, *29*, 82–97.
16. Sainath, T.N.; Mohamed, A.; Kingsbury, B.; Ramabhadran, B. Deep convolutional neural networks for LVCSR. In Proceedings of the 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, Vancouver, BC, Canada, 26–31 May 2013; pp. 8614–8618.
17. Zhu, X.X.; Tuia, D.; Mou, L.; Xia, G.S.; Zhang, L.; Xu, F.; Fraundorfer, F. Deep learning in remote sensing: A review. *arXiv Prepr.* **2017**, arXiv:1710.03959.
18. Abadi, M.; Agarwal, A.; Barham, P.; Brevdo, E.; Chen, Z.; Citro, C.; Corrado, G.; Davis, A.; Dean, J.; Devin, M.; et al. Tensorflow: Large-scale machine learning on heterogeneous distributed systems. *arXiv Prepr.* **2016**, arXiv:1603.04467.
19. Tirado, R. 5220 Matorrales arborescentes con *Ziziphus* (*). In *VV. AA., Bases ecológicas Preliminares Para la Conservación de Los Tipos de Hábitat de Interés Comunitario en Espana*; Ministerio de Medio Ambiente, Medio Rural y Marino: Madrid, Spain, 2009.
20. Zhao, W.; Du, S.; Emery, W. Object-Based Convolutional Neural Network for High-Resolution Imagery Classification. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* **2017**, *10*, 3386–3396.
21. Längkvist, M.; Kiselev, A.; Alirezaie, M.; Loutfi, A. Classification and segmentation of satellite orthoimagery using convolutional neural networks. *Remote Sens.* **2016**, *8*, 329.

22. Hu, F.; Xia, G.; Hu, J.; Zhang, L. Transferring deep convolutional neural networks for the scene classification of high-resolution remote sensing imagery. *Remote Sens.* **2015**, *7*, 14680–14707.

23. Santara, A.; Mani, K.; Hatwar, P.; Singh, A.; Garg, A.; Padia, K.; Mitra, P. BASS Net: Band-Adaptive Spectral-Spatial Feature Learning Neural Network for Hyperspectral Image Classification. *arXiv Prepr.* **2016**, arXiv:1612.00144.

24. Ding, C.; Li, Y.; Xia, Y.; Wei, W.; Zhang, L.; Zhang, Y. Convolutional Neural Networks Based Hyperspectral Image Classification Method with Adaptive Kernels. *Remote Sens.* **2017**, *9*, 618.

25. Liang, H.; Li, Q. Hyperspectral imagery classification using sparse representations of convolutional neural network features. *Remote Sens.* **2016**, *8*, 99.

26. Hyperspectral Remote Sensing Scenes. Available online: http://www.ehu.eus/ccwintco/index.php?title=Hyperspectral_Remote_Sensing_Scenes (accessed on 25 Nov 2017).

27. UC Merced Land Use Dataset. Available online: http://weegee.vision.ucmerced.edu/datasets/landuse.html (accessed on 25 Nov 2017).

28. Brazilian Coffee Scenes Dataset. Available online: http://www.patreo.dcc.ufmg.br/2017/11/12/brazilian-coffee-scenes-dataset/ (accessed on 25 Nov 2017).

29. Castelluccio, M.; Poggi, G.; Sansone, C.; Verdoliva, L. Land use classification in remote sensing images by convolutional neural networks. *arXiv Prepr.* **2015**, arXiv:1508.00092

30. Li, W.; Fu, H.; Yu, L.; Cracknell, A. Deep Learning Based Oil Palm Tree Detection and Counting for High-Resolution Remote Sensing Images. *Remote Sens.* **2016**, *9*, 22.

31. Tiede, D.; Krafft, P.; Füreder, P.; Lang, S. Stratified Template Matching to Support Refugee Camp Analysis in OBIA Workflows. *Remote Sens.* **2017**, *9*, 326.

32. Laliberte, A.S.; Rango, A.; Havstad, K.M.; Paris, J.F.; Beck, R.F.; McNeely, R.; Gonzalez, A.L. Object-oriented image analysis for mapping shrub encroachment from 1937 to 2003 in southern New Mexico. *Remote Sens. Environ.* **2004**, *93*, 198–210.

33. Hellesen, T.; Matikainen, L. An object-based approach for mapping shrub and tree cover on grassland habitats by use of LiDAR and CIR orthoimages. *Remote Sens.* **2013**, *5*, 558–583.

34. Stow, D.; Hamada, Y.; Coulter, L.; Anguelova, Z. Monitoring shrubland habitat changes through object-based change identification with airborne multispectral imagery. *Remote Sens. Environ.* **2008**, *112*, 1051–1061.

35. Baatz, M.; Schäpe, A. Multiresolution segmentation: An optimization approach for high quality multi-scale image segmentation. *Angew. Geogr. Inf. XII* **2000**, *58*, 12–23.

36. Tian, J.; Chen, D.M. Optimization in multi-scale segmentation of high-resolution satellite images for artificial feature recognition. *Int. J. Remote Sens.* **2007**, *28*, 4625–4644.

37. Liu, Y.; Bian, L.; Meng, Y.; Wang, H.; Zhang, S.; Yang, Y.; Shao, X.; Wang, B. Discrepancy measures for selecting optimal combination of parameter values in object-based image analysis. *ISPRS J. Photogramm. Remote Sens.* **2012**, *68*, 144–156.

38. Belgiu, M.; Drăguţ, L. Random forest in remote sensing: A review of applications and future directions. *ISPRS J. Photogramm. Remote Sens.* **2016**, *114*, 24–31.

39. Ghosh, A.; Joshi, P. A comparison of selected classification algorithms for mapping bamboo patches in lower Gangetic plains using very high resolution WorldView 2 imagery. *Int. J. Appl. Earth Observ. Geoinf.* **2014**, *26*, 298–311.

40. Congalton, R.; Green, K. *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices*; CRC Press: Boca Raton, FL, USA, 2008.

41. Guo, Y.; Liu, Y.; Oerlemans, A.; Lao, S.; Wu, S.; Lew, M. Deep learning for visual understanding: A review. *Neurocomputing* **2016**, *187*, 27–48.

42. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.

43. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going deeper with convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 1–9.

44. Tabik, S.; Peralta, D.; Herrera-Poyatos, A.; Herrera, F. A snapshot of image pre-processing for convolutional neural networks: Case study of MNIST. *Int. J. Comput. Intell. Syst.* **2017**, *10*, 555–568.

45. Shin, H.; Roth, H.R.; Gao, M.; Lu, L.; Xu, Z.; Nogues, I.; Yao, J.; Mollura, D.; Summers, R.M. Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. *IEEE Trans. Med. Imaging* **2016**, *35*, 1285–1298.

46. Pan, S.J.; Yang, Q. A survey on transfer learning. *IEEE Trans. Knowl. Data Eng.* **2010**, *22*, 1345–1359.

47. Peralta, D.; Triguero, I.; Garcia, S.; Saeys, Y.; Benitez, J.; Herrera, F. On the use of convolutional neural networks for robust classification of multiple fingerprint captures. *Int. J. Intell. Syst.* **2018**, *33*, 313–230.

48. Hosang, J.; Benenson, R.; Dollar, P.; Schiele, B. What makes for effective detection proposals. *IEEE Trans. Pattern Anal. Mach. Intell.* **2016**, *38*, 814–830.

49. Tirado, R.; Pugnaire, F. Shrub spatial aggregation and consequences for reproductive success. *Oecologia* **2003**, *136*, 296–301.

50. Guirado, E. Factores que Afectan a la Distribucion Especial de Vegetacion Freatofita (Zipiphus lotus) en el Acuifero Costero de Torre Garcia (Sureste de España). Master Thesis, University of Almería, Almería, Spain, 2015.

51. Guirado, E.; Alcaraz-Segura, D.; Rigol-Sanchez, J.; Gisbert, J.; Martinez-Moreno, F.; Galindo-Zaldivar, J.; Gonzalez-Castillo, L.; Cabello, J. Remote sensing-derived fractures and shrub patterns to identify groundwater dependence. *Ecohydrology* **2017**, accepted .

52. Rivas Goday, S.; Bellot, F. Las formaciones de *Zizyphus lotus* (L.) Lamk., en las dunas del Cabo de Gata. *Anales del Instituto Español de Edafología, Ecología y Fisiología Vegetal* **1944**, *3*, 109–126.

53. Lagarde, F.; Louzizi, T.; Slimani, T.; El Mouden, H.; Kaddour, K.; Moulherat, S.; Bonnet, X. Bushes protect tortoises from lethal overheating in arid areas of Morocco. *Environ. Conserv.* **2012**, *39*, 172–182.

54. Manolaki, P.; Andreou, M.; Christodoulou, C. Improving the conservation status of priority habitat types 1520 and 5220 at Rizoelia National Forest Park. EC LIFE Project. Available online: http://www.life-rizoelia. eu (accessed on 25 October 2017).

55. Nussbaum, S.; Niemeyer, I.; Canty, M. *SEATH—A New Tool for Automated Feature Extraction in the Context of Object-Based Image Analysis*; 1st International Conference; Object-Based Image Analysis (OBIA): Salzburg, Austria, 2006.

56. García García, J.; Sánchez Caparrós, A.; Castillo, E.; Marín, I.; Padilla, A.; Rosso, J. Hidrogeoquímica de las aguas subterráneas en la zona de Cabo de Gata. In *López-Geta JA, Gómez JD, De la Orden JA, Ramos G. and Rodríguez L. Tecnología de la Intrusión de Agua de mar en Acuíferos Costeros: Países Mediterráneos*; IGME: Madrid, Spain, 2003; pp. 413–422.

57. Poli, D.; Remondino, F.; Angiuli, E.; Agugiaro, G. Radiometric and geometric evaluation of GeoEye-1, WorldView-2 and Pléiades-1A stereo images for 3D information extraction. *ISPRS J. Photogramm. Remote Sens.* **2015**, *100*, 35–47.

58. Basaeed, E.; Bhaskar, H.; Al-Mualla, M. CNN-based multi-band fused boundary detection for remotely sensed images. In Proceedings of the 6th International Conference on Imaging for Crime Prevention and Detection (ICDP-15), IET, London, UK, 15–17 July 2015; p. 11.

59. Drăguţ, L.; Csillik, O.; Eisank, C.; Tiede, D. Automated parameterisation for multi-scale image segmentation on multiple layers. *ISPRS J. Photogramm. Remote Sens.* **2014**, *88*, 119–127.

60. Bastin, J.; Berrahmouni, N.; Grainger, A.; Maniatis, D.; Mollicone, D.; Moore, R.; Patriarca, C.; Picard, N.; Sparrow, B.; Abraham, E.; et al. The extent of forest in dryland biomes. *Science* **2017**, *356*, 635–638.

61. Schepaschenko, D.; Fritz, S.; See, L.; Bayas, J.C.L.; Lesiv, M.; Kraxner, F.; Obersteiner, M. Comment on "The extent of forest in dryland biomes". *Science* **2017**, *358*, eaao0166.

62. Prevedello, J.A.; Almeida-Gomes, M.; Lindenmayer, D. The importance of scattered trees for biodiversity conservation: A global meta-analysis. *J. Appl. Ecol.* **2017**, in press, doi:10.1111/1365-2664.12943.

63. Griffith, D.M.; Lehmann, C.E.; Strömberg, C.A.; Parr, C.L.; Pennington, R.T.; Sankaran, M.; Ratnam, J.; Still, C.J.; Powell, R.L.; Hanan, N.P.; et al. Comment on "The extent of forest in dryland biomes". *Science* **2017**, *358*, eaao1309.

64. de la Cruz, M.; Quintana-Ascencio, P.F.; Cayuela, L.; Espinosa, C.I.; Escudero, A. Comment on "The extent of forest in dryland biomes". *Science* **2017**, *358*, eaao0369.

65. Pettorelli, N.; Wegmann, M.; Skidmore, A.; Mücher, S.; Dawson, T.P.; Fernandez, M.; Lucas, R.; Schaepman, M.E.; Wang, T.; O'Connor, B.; et al. Framing the concept of satellite remote sensing essential biodiversity variables: challenges and future directions. *Remote Sens. Ecol. Conserv.* **2016**, *2*, 122–131.