Article

Mask R-CNN and OBIA Fusion Improves the Segmentation of Scattered Vegetation in Very High-Resolution Optical Sensors

Emilio Guirado [1,2,*], Javier Blanco-Sacristán [3], Emilio Rodríguez-Caballero [4,5], Siham Tabik [6], Domingo Alcaraz-Segura [7,8], Jaime Martínez-Valderrama [1] and Javier Cabello [2,9]

1 Multidisciplinary Institute for Environment Studies "Ramon Margalef" University of Alicante, Edificio Nuevos Institutos, Carretera de San Vicente del Raspeig s/n San Vicente del Raspeig, 03690 Alicante, Spain; jaime.mv@ua.es
2 Andalusian Center for Assessment and monitoring of global change (CAESCG), University of Almeria, 04120 Almeria, Spain; jcabello@ual.es
3 College of Engineering, Mathematics and Physical Sciences, University of Exeter, Penryn Campus, Cornwall TR10 9EZ, UK; jb1230@exeter.ac.uk
4 Agronomy Department, University of Almeria, 04120 Almeria, Spain; e.rodriguez-caballer@mpic.de
5 Centro de Investigación de Colecciones Científicas de la Universidad de Almería (CECOUAL), 04120 Almeria, Spain
6 Department of Computer Science and Artificial Intelligence, University of Granada, 18071 Granada, Spain; siham@ugr.es
7 Department of Botany, Faculty of Science, University of Granada, 18071 Granada, Spain; dalcaraz@ugr.es
8 iEcolab, Inter-University Institute for Earth System Research, University of Granada, 18006 Granada, Spain
9 Department of Biology and Geology, University of Almeria, 04120 Almeria, Spain
* Correspondence: e.guirado@ual.es

Abstract: Vegetation generally appears scattered in drylands. Its structure, composition and spatial patterns are key controls of biotic interactions, water, and nutrient cycles. Applying segmentation methods to very high-resolution images for monitoring changes in vegetation cover can provide relevant information for dryland conservation ecology. For this reason, improving segmentation methods and understanding the effect of spatial resolution on segmentation results is key to improve dryland vegetation monitoring. We explored and analyzed the accuracy of Object-Based Image Analysis (OBIA) and Mask Region-based Convolutional Neural Networks (Mask R-CNN) and the fusion of both methods in the segmentation of scattered vegetation in a dryland ecosystem. As a case study, we mapped *Ziziphus lotus*, the dominant shrub of a habitat of conservation priority in one of the driest areas of Europe. Our results show for the first time that the fusion of the results from OBIA and Mask R-CNN increases the accuracy of the segmentation of scattered shrubs up to 25% compared to both methods separately. Hence, by fusing OBIA and Mask R-CNNs on very high-resolution images, the improved segmentation accuracy of vegetation mapping would lead to more precise and sensitive monitoring of changes in biodiversity and ecosystem services in drylands.

Keywords: deep-learning; fusion; mask R-CNN; object-based; optical sensors; scattered vegetation; very high-resolution

1. Introduction

Dryland biomes cover ~47% of the Earth's surface [1]. In these environments, vegetation appears scattered [2] and its structure, composition and spatial patterns are key indicators of biotic interactions [3], regulation of water, and nutrient cycles at landscape level [4]. Changes in the cover and spatial patterns of dryland vegetation occur in response to land degradation processes [5]. Hence, methods to identify and characterize vegetation patches and their structural characteristics can improve our ability to understand dryland functioning and to assess desertification risk [5–8]. Progress has been made using remote sensing tools in this regard (e.g., quantification of dryland vegetation structure at landscape scale [9], monitoring vegetation trends [10], spatial patterns identifying ecosystem

multifunctionality [11], characterizing flood dynamics [12], among many others). However, the improvement in the accuracy of vegetation cover measurement is still being studied to obtain maximum performance from data and technology. Estimating and monitoring changes in vegetation cover through remote sensing is key for dryland ecology and conservation [6]. Both historical temporal and spatial data are the base for remote sensing studies to identify the functioning and structure of vegetation [13,14].

The analysis of very high-resolution images to detect and measure vegetation cover and its spatial arrangement across the landscape starts typically by segmenting the objects to be identified in the images [7]. Object-Based Image Analysis (OBIA) [15] and Mask Region-based Convolutional Neural Networks (Mask R-CNN) [16] are among the most used and state-of-the-art segmentation methods. Though they provide a similar product, both methods rely on very different approaches. OBIA combines spectral information from each pixel with its spatial context [17,18]. Similar pixels are then grouped in homogenous objects that are used as the basis for further classification. Mask R-CNN, on the other hand, a type of artificial intelligence whose functioning is inspired by the human brain provides transferable models between zones and semantic segmentation with unprecedented accuracy [19,20]. Besides, fusion has recently been used to improve spectral, spatial, and temporal resolution from remote sensing images [21–23]. However, the fusion of methods for vegetation mapping has not been evaluated.

Remote sensing studies based on very high-resolution images have increased in the last years (e.g., [24–27]), partly because of the availability of Google Earth images worldwide [28–30] and the popularization of unmanned aerial vehicles (UAV). Although these images have shown a high potential for vegetation mapping and monitoring [31–33], two main problems arise when they are used. First, higher spatial resolution increases the spectral heterogeneity among and within vegetation types, resulting in a salt and pepper effect in their identification that does not correctly characterize the actual surface [34]. Second, the processing time of very high-resolution images and the computational power required is larger than in the case of low-resolution images [35]. Under these conditions, traditional pixel-based analysis has proved to be less accurate than OBIA or Mask R-CNN for scattered vegetation mapping [15,36]. There are many applications for OBIA [37–39] and deep learning segmentation methods [40,41]. For example, mapping greenhouses [42], monitoring disturbances affecting vegetation cover [5], or counting scattered trees in Sahel and Sahara [43]. These methods have been compared with excellent results in both segmenting and detecting tree cover and scattered vegetation [7,44,45]. However, greater precision is always advisable in problems of very high sensitivity [46]. Despite methodological advances, selecting the appropriate image source is key to produce accurate segmentations of objects, like in vegetation maps [47,48], and there is no answer to the question of which image or method to choose for segmenting objects. Understanding how the spatial resolution of the imagery used affects these segmentation methods or the fusing of both is key for their correct application to obtain better accuracy in object segmentation in vegetation mapping in drylands.

To evaluate which is the most accurate method between OBIA and Mask R-CNN to segment scattered vegetation in drylands and to understand the effect of the spatial resolution of the images used in this process, we assessed the accuracy of these two methods in the segmentation of scattered dryland shrubs and compared how final accuracy varies as does spatial resolution. We also check the accuracy of the fusion of both methods.

This work is organized as follows. Section 2 describes the study area, the dataset used, and the methodologies tested. Section 3 describes the experiments addressed to assess the accuracies of the methods used. The experimental results and discussion are presented in Section 4, and conclusions are given in Section 5.

2. Materials and Methods

2.1. Study Area

We focused on the community of *Ziziphus lotus* shrubs, an ecosystem of priority conservation interest at European level (habitat 5220* of Directive 92/43/EEC), located in Cabo de Gata-Níjar Natural Park (36°49'43'' N, 2°17'30'' W, SE Spain), one of the driest areas of continental Europe. This type of vegetation is scarce and patchy, which appears surrounded by a matrix of bare soil and small shrubs (e.g., *Launea arborescens*, *Lygeum spartum* and *Thymus hyemalis*). *Z. lotus* is a facultative phreatophyte [49] and forms large hemispherical canopies (1–3 m tall) that constitute fertility islands where many other species of plants and animals live [50]. These shrubs are long-lived species contributing to the formation of geomorphological structures, called nebkhas [51], that protect from the intense wind erosion activity that characterizes the area, thereby retaining soil, nutrients, and moisture.

2.2. Dataset

The data set consisted of two plots (Plot 1 and Plot 2) with 3 images of different spatial resolution in each one. The plots had an area of 250 × 250 m with scattered *Z. lotus* shrubs. The images were obtained from optical remote sensors in the visible spectral range, Red, Green and Blue bands (RGB) and spatial resolutions of < 1 m/pixel:

- A 0.5 × 0.5 m spatial resolution RGB image obtained from Google Earth [52].
- A 0.1 × 0.1 m spatial resolution image acquired using an RGB camera sensor of 50 megapixels (Hasselblad H4D) equipped with a 50 mm lens and charge-coupled device (CCD) sensor of 8176 pixels × 6132 pixels mounted on a helicopter with a flight height of 550 m.
- A 0.03 × 0.03 m spatial resolution image acquired using a 4K pixels resolution RGB camera sensor on a professional UAV Phantom 4 UAV (DJI, Shenzhen, China) and with a flight height of 40 m.

2.3. OBIA

OBIA-based segmentation is a method of image analysis that divides the image into homogeneous objects of interest (i.e., groups of pixels also called segments) based on similarities of shape, spectral information, and contextual information [17]. It identifies homogeneous and discrete image objects by setting an optimal combination of values for three parameters (i.e., Scale, Shape, and Compactness) related to their spectral and spatial variability. There are no unique values for any of these parameters, and their final combination always depends on the object of interest, so finding this optimal combination represents a challenge due to the vast number of possible combinations. First, it is necessary to establish an appropriate Scale level depending on the size of the object studied in the image [43]; for example, low Scale values for small shrubs and high Scale values for large shrubs [44,45]. Recent advances have been oriented in developing techniques (e.g., [53–59]) and algorithms (e.g., [60–63]) to automatically find the optimal value of the Scale parameter [64], which is the most important for determining the size of the segmented objects [65,66]. The Shape and the Compactness parameters must be configured too. While high values of the Shape parameter prioritize the shape over the colour, high values of the Compactness parameter prioritize compactness of the objects over the smoothness of their edges [67].

2.4. Mask R-CNN

In this problem of locating and delimiting the edges of dispersed shrubs, we used a computer vision technique named instance segmentation [68]. Such technique infers a label for each pixel considering other nearby objects, thus including the boundaries of the object. We used Mask R-CNN segmentation model [16], which extends Faster R-CNN detection model [16] and provides three outputs for each object: (i) a class label, (ii) a bounding box that delimits the object and (iii) a mask which delimits the pixels that constitute each object.

In the binary problem addressed in this work, Mask R-CNN generates for each predicted object instance a binary mask (values of 0 and 1), where values of 1 indicate a *Z. lotus* pixel and 0 indicates a bare soil pixel.

Mask R-CNN relies on a classification model for the task of feature extraction. In this work, we used ResNet 101 [69] to extract increasingly higher-level characteristics from the lowest to the deepest layer levels.

The learning process of Mask R-CNN is influenced by the number of epochs, which is the number of times the network goes through the training phase, and by other optimizations such as transfer-learning or data-augmentation (see Section 3.2). Finally, the $1024 \times 1024 \times 3$ band image input is converted to $32 \times 32 \times 2048$ to represent objects at different scales via the characteristic network pyramid.

2.5. Segmentation Accuracy Assessment

The accuracy of the segmentation task in this work was assessed with respect to ground truth by using the Euclidean Distance v.2 (ED2; [70]), which evaluates the geometric and arithmetic discrepancy between reference polygons and the segments obtained during the segmentation process. Both types of discrepancy need to be assessed. As reference polygons, we used the perimeter of 60 *Z. lotus* shrubs measured with photo-interpretation in all images by a technical expert. We estimated the geometric discrepancy by the "Potential Segmentation Error" (PSE; Equation (1)), defined as the ratio of the total area of each segment obtained in the segmentation that falls outside the reference segment and the total area of reference polygons as:

$$PSE = \frac{\Sigma |s_i - r_k|}{\Sigma |r_k|} \tag{1}$$

where PSE is the "Potential Segmentation Error", r_k is the area of the reference polygon and si is the overestimated area of the segment obtained during the segmentation. A value of 0 indicates that segments obtained from the segmentation fit well into the reference polygons. Conversely, larger values indicate a discrepancy between reference polygons and the segments.

Although the geometric relation is necessary, it is not enough to describe the discrepancies between the segments obtained during the segmentation process and the corresponding reference polygons. To solve such problem the ED2 index includes an additional factor, the "Number-of-Segmentation Ratio" (NSR), that evaluates the arithmetic discrepancy between the reference polygons and the generated segments (Equation (2)):

$$NSR = \frac{abs(m - v)}{m} \tag{2}$$

where NSR is the arithmetic discrepancy between the polygons of the resulting segmentation and the reference polygons and abs is the absolute value of the difference of the number of reference polygons, m, and the number of segments obtained, v.

Thus, the ED2 can be defined as the joint effect of geometric and arithmetic differences (Equation (3)), estimated from PSE and NSR, respectively, as:

$$ED2 = \sqrt{(PSE)^2 + (NSR)^2} \tag{3}$$

where ED2 is Euclidean Distance v.2, PSE is Potential Segmentation Error, and NSR is Number-of-Segmentation Ratio. According to Liu et al. [70], values of ED2 close to 0 indicate good arithmetic and geometric coincidence, while high values indicate a mismatch between them.

3. Experiments

We set several experiments to assess the accuracy of the two different OBIA and Mask R-CNN segmenting scattered vegetation in drylands. We used the images of Plot 1 to test the OBIA and Mask R-CNN segmentation methods. The images of Plot 2 were used

for the training phase in Mask R-CNN experiments exclusively (Figure 1). In Section 3.1, we describe OBIA experiments, focused on detecting the best parameters (i.e., Scale, Shape and Compactness) of a popularly used "multi-resolution" segmentation algorithm [71]. In Section 3.2. we described the Mask R-CNN experiments, in which we first evaluated the precision in the detection of shrubs (capture or notice the presence of shrubs) and second how accurate is the segmentation of those shrubs. Finally, in Section 3.3. we described the fusion of both methods and compared all the accuracies between them in Section 4.3.
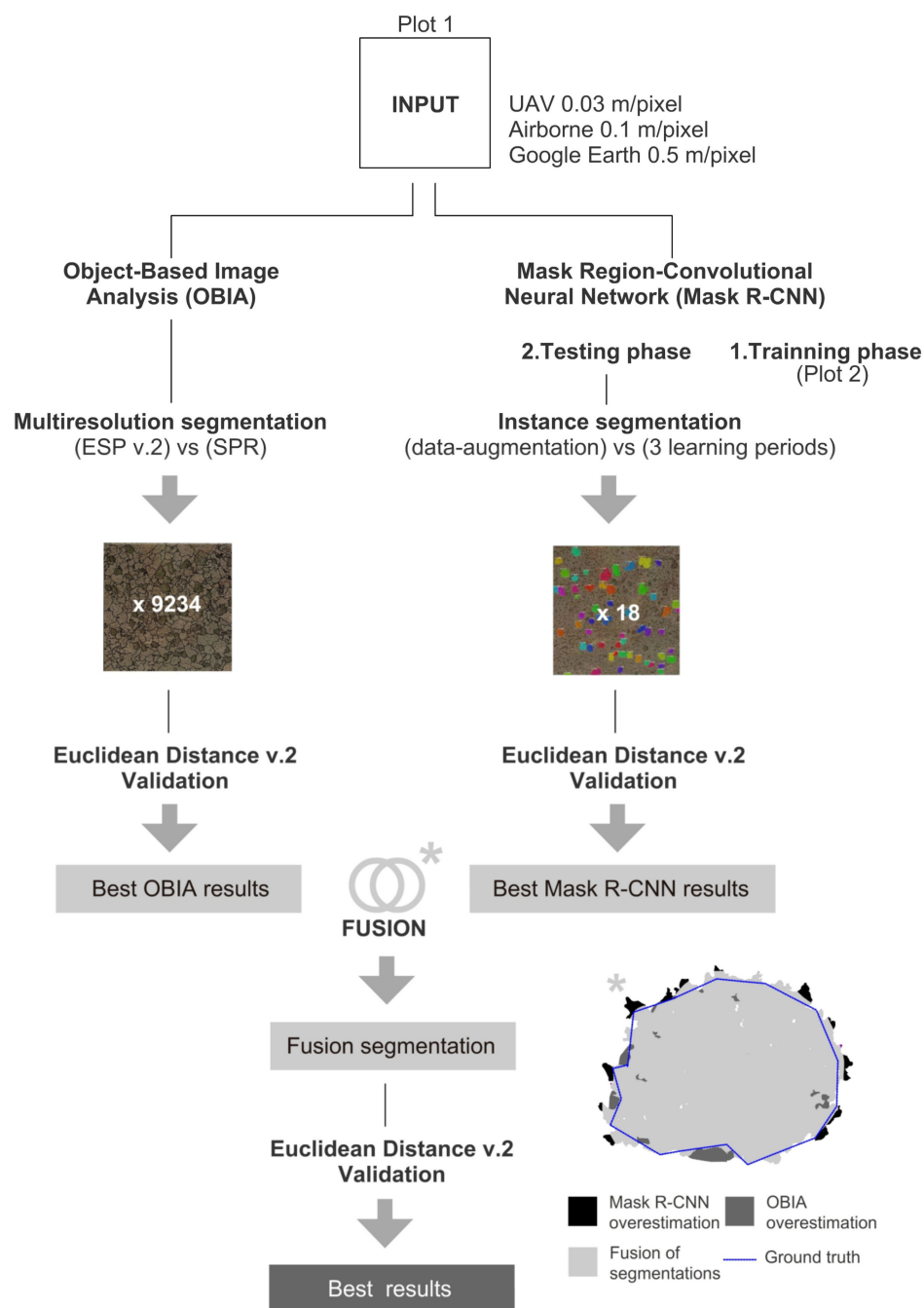

Figure 1. Workflow with the main processes carried out in this work. Asterisk shows an example of the result of the fusion of the segmentation results from OBIA and Mask R-CNN. OBIA: Object-Based Image Analysis; Mask R-CNN: Mask Region-based Convolutional Neural Networks; ESP v.2: Estimation of Scale Parameter v.2; SPR: Segmentation Parameters Range.

3.1. OBIA Experiments

To obtain the optimal value of each parameter of the OBIA segmentation, we use two approaches:

(i)   A ruleset called Segmentation Parameters Range (SPR) in eCognition v8.9 (Definiens, Munich, Germany) with the "multi-resolution" algorithm that segmented the images of Plot 1 by systematically increasing the Scale parameter in steps of 5 and the Shape and Compactness parameters in steps of 0.1. The Scale parameter ranged from 80 to 430, and the Shape and the Compactness from 0.1 to 0.9. We generated a total of 9234 results with possible segmentations of *Z. lotus* shrubs. The Scale parameter ranges were evaluated considering the minimum cover size (12 m^2) and maximum cover size (311 m^2) of the shrubs measured in the plot and the pixel size.

(ii)   We also performed the semi-automatic method Estimation of Scale Parameter v.2 (ESP2; [70]) to select the best scale parameter. This tool performs semi-automatic segmentation of multiband images within a range of increasing Scale values (Levels), while the user previously defines the values of the Compactness and Shape parameters. Three options available in the ESP2 tool were tested: a) the hierarchical analysis Top-down (HT), starting from the highest level and segmenting these objects for lower levels; b) the hierarchical analysis Bottom-up (HB), which starts from the lower level and combines objects to get larger levels; and c) analysis without hierarchy (NH), where each scale parameter is generated independently, based only on the level of the pixel [64].

3.2. Mask R-CNN Experiments

Mask R-CNN segmentation is divided in two phases: i) Training and ii) Testing phases. In the training phase, we selected 100 training polygons representing 100 shrub individuals with different sizes. The sampling was done using VGG Image Annotator [72] to generate a JSON file, which includes the coordinates of all the vertices of each segment, equivalent to the perimeter of each shrub. To increase the number of samples and reduce overfitting of the model, we applied data-augmentation and transfer-learning:

- Data augmentation aims to artificially increase the size of the dataset by slightly modifying the original images. We applied the filters of vertical and horizontal flip; Scale decrease and increase in the horizontal and vertical axis between 0.8 to 1.2; Rotation of 0 to 365 degrees; Shearing factor between −8 to 8; Contrast normalization with values of 0.75 and 1.5 per channel; Emboss with alpha 0, 0.1; Strength with 0 to 2.0; Multiply 0.5 and 1.5, per channel to change the brightness of the image (50–150% of the original value).

- Transfer-learning consists in using knowledge learnt from one problem to another related one [73], and we used it to improve the neural network. Since the first layers of a neural network extract low-level characteristics, such as colour and edges, they do not change significantly and can be used for other visual recognition works. As our new dataset was small, we applied fine adjustment to the last part of the network by updating the penultimate weights, so that the model was not overfitting, as mainly occurs between the first layers of the network. We specifically used transfer-learning on ResNet 101 [69] and used Region-based CNN with the pre-trained weights of the same architectures on COCO dataset (around 1.28 million images over 1000 generic object classes) [74].

We tested three different learning periods (100 steps per epoch) per model:

(A)   40 epochs with transfer-learning in heads,
(B)   80 epochs with 4 fist layers transfer-learning,
(C)   160 epochs with all layers transfer-learning.

We trained the algorithm based on the ResNet architecture with a depth of 101 layers with each of the three proposed spatial resolutions. We then evaluated the trained models in all possible combinations between the resolutions. We evaluated the use of data-

augmentation and transfer-learning from more superficial layers to the whole architecture with different stages in the training process. Particularly:

(1.1) Trained with UAV images.
(1.2) Trained with UAV images and data-augmentation.
(2.1) Trained with airborne images.
(2.2) Trained with airborne images and with data-augmentation.
(3.1) Trained with Google Earth images.
(3.2) Trained with Google Earth images and data-augmentation.

We did the test phase using Plot 1. To identify the most accurate experiments, we evaluated the detection of the CNN-based models, and determined their Precision, Recall, and F1-measure [75] as:

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}},$$ (4)

$$\text{Recall} = \frac{\text{True Positives}}{\text{True positives} + \text{False Negatives}},$$ (5)

$$\text{F1} - \text{measure} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$ (6)

3.3. Fusion of OBIA and Mask R-CNN

We combined the most accurate segmentations obtained using OBIA and Mask R-CNN, according to ED2 values (Figure 1). We let o_i denote the i-th OBIA polygon within the OBIA segmentation, O, and m_j denote the j-th Mask R-CNN polygon within the Mask R-CNN segmentation, C. Then we have O = {o_i: i = 1, 2, ..., m} and C = {c_j: j = 1, 2, ..., n}. Here, the subscripts i and j are sequential numbers for the polygons of the OBIA and Mask R-CNN segmentations, respectively. m and n indicate the total numbers of the objects segmented with OBIA and Mask R-CNN, respectively. m and n must be equal. Finally, the corresponding segment data sets extracted (Equation (7)) by the fusion are considered a consensus among the initially segmented objects as:

$$OC_{ij} = \text{area}O_i \cap \text{area}C_j$$ (7)

where OC_{ij} is the intersected area between the segments of the OBIA segmentation (O_i) and the area of the segments of the Mask R-CNN segmentation (C_j).

Finally, we estimate ED2 values of the final segmentation using validation shrubs from Plot 1, and we compared it with segmentation accuracy obtained by the different methods.

4. Results and Discussion

4.1. OBIA Segmentation

In total, 9234 segmentations were performed by SPR, 3078 for each image type (e.g., Google Earth, airborne and UAV). OBIA segmentation accuracy using the SPR presented large variability (Table 1), with values of ED2 ranging between 0.05 and 0.28. Segmentation accuracy increased with image spatial resolution. Thus, the higher the spatial resolution, the higher the Scale values and more accurate the segmentation was. This result was represented by a decrease in ED2 values of 0.14, 0.10 and 0.05 for Google Earth, airborne and UAV images, respectively. The best combinations of segmentation parameters along the different images were (Figure 2): (i) for the Google Earth image, Scale values ranging from 105 to 110, low Shape values of 0.3 and high Compactness values from 0.8 to 0.9; (ii) for the orthoimage from the airborne sensor, Scale values between 125 and 155, Shape of 0.6 and Compactness of 0.9; and (iii) for the UAV image, the optimal segmentation showed the highest Scale values, ranging from 360 to 420, whereas Shape and Compactness values were similar to the values of the Google Earth image.

Table 1. Segmentation accuracies of Object-Based Image Analysis (OBIA) among the three spatial resolutions evaluated. For each segmentation type, only the most accurate combination of Scale, Shape, and Compactness is shown. ESP2/HB: Estimate Scale Parameter v.2 (ESP2) with Bottom-up Hierarchy; ESP2/HT: ESP2 with Top-down Hierarchy; ESP2/NH: ESP2 Non-Hierarchical; SPR: Segmentation with Parameters Range. Closer values to 0 indicate accurate segmentations. In bold the most accurate results.

Image Source	Resolution (m/Pixel)	Segmentation Method	Segmentation Parameters			Segmentation Quality	
			Scale	Shape	Compactness	ED2	Average Time (s)
Google Earth	0.5	ESP2/HB	100	0.6	0.9	0.25	365
		ESP2/HT	105	0.7	0.5	0.26	414
		ESP2/NH	105	0.5	0.1	0.28	2057
		SPR	**90**	**0.3**	**0.8**	**0.2**	**18**
Airborne	0.1	ESP2/HB	170	0.5	0.9	0.14	416
		ESP2/HT	160	0.5	0.9	0.15	650
		ESP2/NH	160	0.5	0.5	0.14	3125
		SPR	**155**	**0.6**	**0.9**	**0.1**	**24**
UAV	0.03	ESP2/HB	355	0.3	0.7	0.12	5537
		ESP2/HT	370	0.5	0.7	0.11	8365
		ESP2/NH	350	0.5	0.7	0.1	40,735
		SPR	**420**	**0.1**	**0.8**	**0.05**	**298**

When we applied the semi-automatic method ESP2 to estimate the optimum value of the Scale parameter, we observed a similar pattern to that described for the SPR, with an increase in accuracy when increasing spatial resolution. The highest value of ED2 was for the Google Earth image segmentation results (ED2 = 0.25), decreasing for the orthoimage from the airborne sensor (ED2 = 0.15) and reaching the minimum value (best) in the UAV image (ED2 = 0.12). However, the results obtained by ESP2 were worse than the results obtained by the SPR method in all the images analysed (Table 1) with the largest differences in the image with the lowest spatial resolution (Google Earth). In the Google Earth images, the best method of analysis of the three options presented by the ESP2 tool was the hierarchical bottom level, with acceptable ED2 values, lower than 0.14 (Table 1). For the airborne images, the results were equal to Google Earth images (hierarchical bottom level). Conversely, the segmentation of the UAV image produced the best ED2 values when applying the ESP2 without hierarchical level. The computational time for the segmentation of the images was higher in ESP2 than SPR approach. In addition, the computation time of the analysis was also influenced by the number of pixels to analyse, it increased in higher spatial resolution images in computer with a Core i7-4790K, 4 GHz and 32G of RAM memory (Intel, Santa Clara, CA, USA) (Table 1).
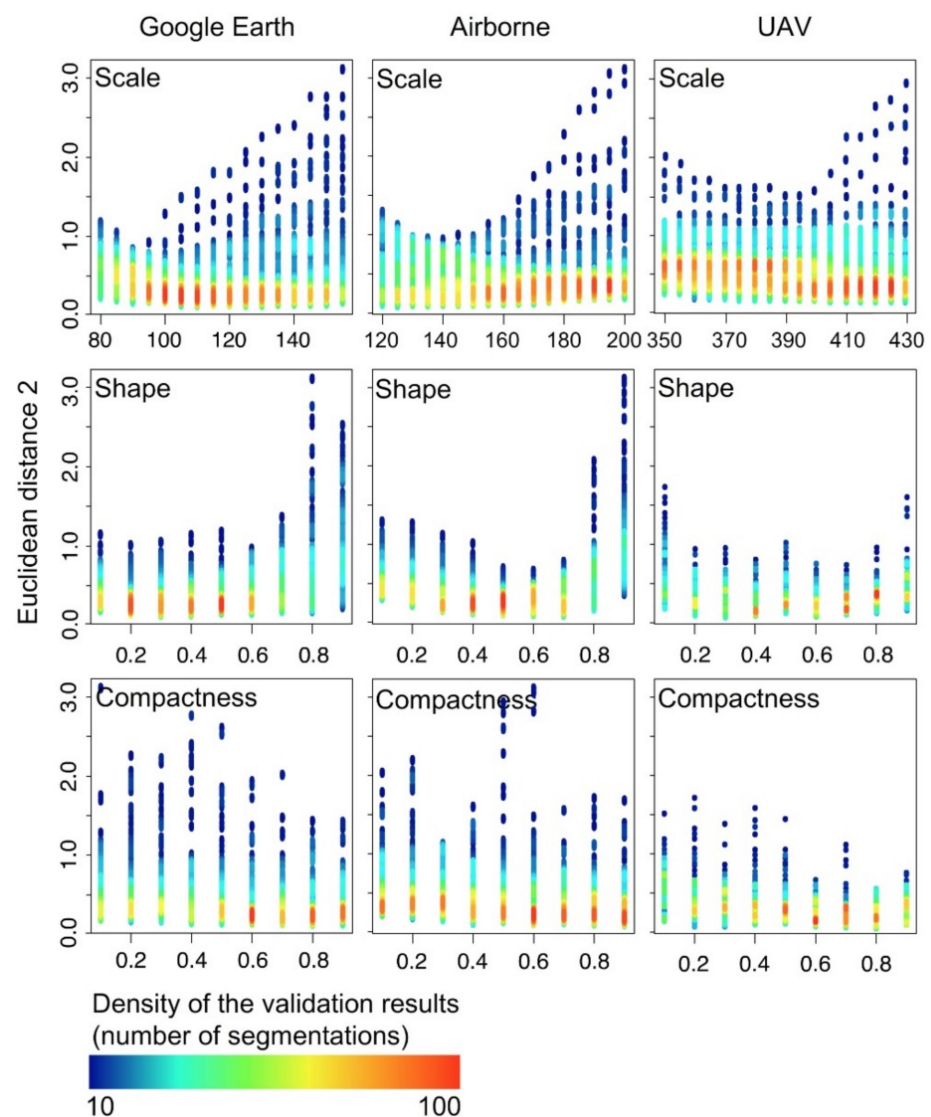
Figure 2. Relationship between Scale, Shape and Compactness parameters (X axis) evaluated using Euclidean distance v.2 (ED2; Y axis) in 9234 Object-based image analysis (OBIA) segmentations from Google Earth, Airborne and unmanned aerial vehicle (UAV) images. The rainbow palette shows the density of validation results. In red high density and in blue low density.

4.2. Mask R-CNN Segmentation

4.2.1. Detection of Scattered Shrubs

We obtained the best detection results for the models trained and evaluated with UAV images (F1-measure = 0.91) and the models trained with the highest number of epochs and data-augmentation activated (Table 2). The best transfer from a UAV trained model to a test with another resolution was to the image from the airborne sensor. Nevertheless, the Google Earth test image produced a similar result of F1-measure = 0.90. We consider that a model trained with data-augmentation and very high spatial resolution images (0.03 m/pixel) can generalize well to less accurate images such as those from Google Earth (0.5 m/pixel). Furthermore, when we trained the models with Google Earth images, we observed that it also generalised well to more precise resolutions (F1-measure = 0.90). For this reason, the detection of *Z. lotus* shrubs might be generalizable from any resolution less than 1 m/pixel.

Table 2. Test results of Mask Region-based Convolutional Neural Networks (Mask R-CNN) experiments in three different spatial resolutions images. TP: True Positive; FP: False Negative; FN: False Negative. Precision, Recall, and F1-measure were used for detection results. In bold the most accurate results.

Experiments/Image		TP	FP	FN	Precision	Recall	F1
	UAV	55	5	10	0.92	0.85	0.88
1.1.A	Airborne	56	4	9	0.93	0.86	0.90
	GE	50	1	15	0.98	0.77	0.86
	UAV	59	6	6	0.91	0.91	0.91
1.1.B	Airborne	60	7	5	0.90	0.92	0.91
	GE	55	2	10	0.96	0.85	0.90
	UAV	**55**	**1**	**10**	**0.98**	**0.85**	**0.91**
1.1.C	Airborne	52	3	13	0.94	0.80	0.87
	GE	53	0	12	1	0.81	0.89
	UAV	53	1	12	0.98	0.82	0.89
1.2.A	Airborne	54	1	11	0.98	0.83	0.90
	GE	42	3	23	0.93	0.65	0.76
	UAV	55	1	10	0.98	0.85	0.91
1.2.B	Airborne	50	2	15	0.96	0.77	0.85
	GE	50	2	15	0.96	0.77	0.85
	UAV	**56**	**3**	**8**	**0.95**	**0.87**	**0.91**
1.2.C	Airborne	52	3	13	0.94	0.80	0.87
	GE	54	1	12	0.98	0.81	0.89
	UAV	41	0	24	1	0.63	0.77
2.1.A	Airborne	38	0	27	1	0.58	0.74
	GE	34	1	31	0.97	0.52	0.68
	UAV	47	0	18	1	0.72	0.84
2.1.B	**Airborne**	**55**	**3**	**10**	**0.95**	**0.85**	**0.89**
	GE	50	1	16	0.98	0.76	0.85
	UAV	52	1	13	0.98	0.80	0.88
2.1.C	Airborne	58	3	7	0.95	0.88	0.91
	GE	54	1	12	0.98	0.82	0.89
	UAV	31	0	34	1	0.48	0.65
2.2.A	Airborne	48	1	17	0.98	0.74	0.84
	GE	38	1	27	0.97	0.58	0.73
	UAV	38	1	27	0.97	0.58	0.73
2.2.B	Airborne	46	1	19	0.98	0.71	0.82
	GE	47	3	18	0.94	0.72	0.82
	UAV	46	1	19	0.98	0.70	0.82
2.2.C	**Airborne**	**51**	**2**	**14**	**0.96**	**0.78**	**0.86**
	GE	50	2	15	0.96	0.77	0.85
	UAV	37	0	28	1	0.57	0.73
3.1.A	Airborne	43	0	22	1	0.66	0.80
	GE	41	1	24	0.98	0.63	0.77
	UAV	48	1	17	0.98	0.74	0.84
3.1.B	Airborne	51	1	14	0.98	0.78	0.87
	GE	**54**	**1**	**11**	**0.98**	**0.83**	**0.90**
	UAV	52	1	13	0.98	0.80	0.88
3.1.C	Airborne	52	1	13	0.98	0.80	0.88
	GE	54	2	11	0.96	0.83	0.89
	UAV	54	1	11	0.98	0.83	0.90
3.2.A	Airborne	56	4	9	0.93	0.86	0.90
	GE	53	2	12	0.96	0.82	0.88
	UAV	**56**	**3**	**9**	**0.95**	**0.86**	**0.90**
3.2.B	Airborne	54	5	11	0.92	0.83	0.87
	GE	53	3	12	0.95	0.82	0.88
	UAV	54	3	11	0.95	0.83	0.89
3.2.C	Airborne	52	3	13	0.95	0.80	0.87
	GE	52	3	13	0.95	0.80	0.87

4.2.2. Segmentation Accuracy for Detected Shrubs

The best segmentation accuracy was obtained with the models trained and tested with the same source of images, reaching values of ED2 = 0.07 in Google Earth ones. However, when the model trained with Google Earth images was tested in a UAV image, the ED2 resulted in 0.08. Moreover, the effect of data-augmentation was counterproductive in models trained with airborne images and only lowered ED2 (best results) in models trained with the UAV image. In general, data-augmentation helped to generalise between images but did not obtain a considerable increase in precision in models trained and tested with the same image resolution (Table 3 and Figure 3).

Table 3. Segmentation accuracies of Mask Region-based Convolutional Neural Networks (Mask R-CNN). PSE: Potential Segmentation Error; NSR: Number Segmentation Ratio; ED2: Euclidean Distance v.2. In bold the most accurate results.

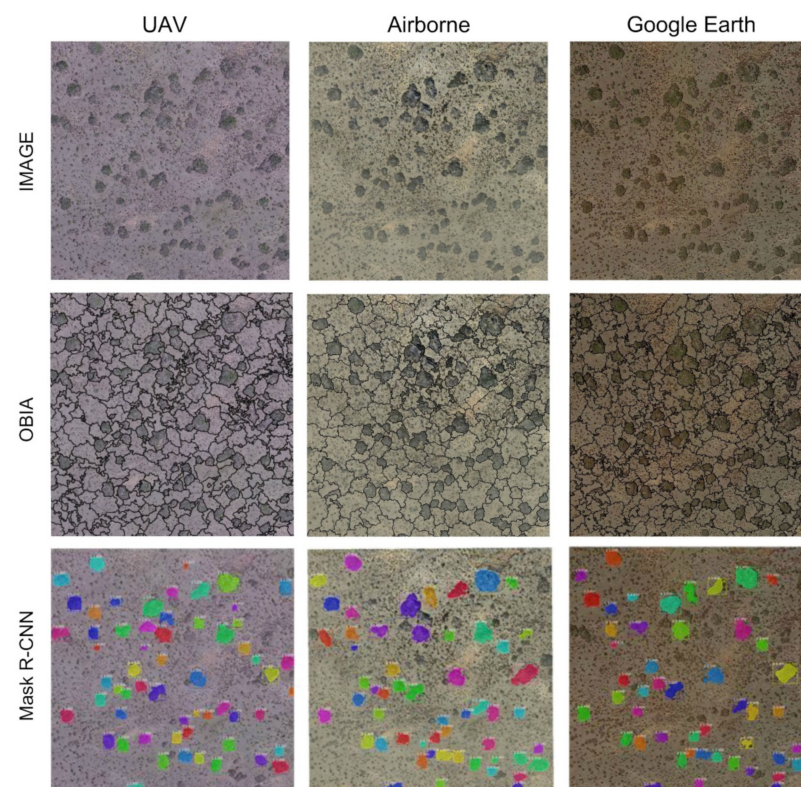| Best Experiment	Image Train	Image Test	PSE	NSR	ED2
1.1.C	UAV	UAV	0.0532	0.1290	0.1396
1.2.C	UAV	UAV	0.0512	0.0967	0.1095
2.1.C	**Airborne**	**Airborne**	**0.0408**	**0.0645**	**0.0763**
2.2.C	Airborne	Airborne	0.0589	0.0645	0.0873
3.1.B	**GE**	**GE**	**0.0414**	**0.0645**	**0.0767**
3.2.B	**GE**	**UAV**	**0.0501**	**0.0645**	**0.0816**

Figure 3. Examples of segmentation of images from Plot 1 using Object-based Image Analysis (OBIA; **Top**) and Mask Region-based Convolutional Neural Networks (Mask R-CNN; **Down**) on Google Earth, Airborne and Unmanned Aerial Vehicle (UAV) images. The different colours in the Mask R-CNN approach are to differentiate the shrubs individually.

4.3. Fusion of OBIA and Mask R-CNN

Our results showed that the fusion between OBIA and Mask R-CNN methods in very high-resolution RGB images is a powerful tool for mapping scattered shrubs in drylands.

We found that the individual segmentations by using OBIA and Mask R-CNN independently were worse than the fusion of both. The accuracy of the fusion of OBIA and Mask R-CNN was higher than the accuracies of the separate segmentations (Table 4), being the most accurate segmentation of all the experiments tested in this work, with an ED2 = 0.038. However, the fusion between results on Google Earth images only improved the ED2 by 0.02. Therefore, the fusion of both segmentation methods provided the best segmentation over the previous methods (OBIA (ED2 = 0.05) and Mask R-CNN (ED2 = 0.07)), in very high-resolution images to segment scattered vegetation in drylands. Moreover, by merging the results of both methodologies (OBIA ∩ Mask R-CNN), the accuracy increases with an ED2 = 0.03.

Table 4. Segmentation accuracies of the fusion of Object-Based Image Analysis (OBIA) and Mask Region-based Convolutional Neural Networks (Mask R-CNN). PSE: Potential Segmentation Error; NSR: Number Segmentation Ratio; ED2: Euclidean Distance v.2. In bold the most accurate results.

Best Experiment	Best OBIA (ED2)	Best Mask R-CNN (ED2)	PSE	NSR	ED2
1.1.C	**0.05**	**0.13**	**0.02**	**0.03**	**0.0386**
1.2.C	0.05	0.10	0.02	0.03	0.0417
2.1.C	0.10	0.07	0.02	0.03	0.0388
2.2.C	0.10	0.08	0.05	0.06	0.0395
3.1.B	0.20	0.07	0.00	0.06	0.0645
3.2.B	0.20	0.08	0.00	0.06	0.0645

To our knowledge, the effect of mixing these two methodologies has not been studied until the date, and it might be vital to improving future segmentation methods. As can be seen in the conceptual framework (Figure 1), it is reasonable to think that the higher the resolution and, therefore, the higher the detail at the edges of vegetation represented in the images, the fusion will improve the final precision of the segmentation. Nevertheless, in images with lower resolution, the fusion improved but to a minor degree.

The spatial resolution of the images affected the accuracy of the segmentation, providing outstanding results in all segmentation methods and spatial resolutions. However, according to [57], we observed that the spatial resolution and Scale parameter played a key role during the segmentation process and controlled the accuracy of the final segmentations. In non-fusion segmentation methods (OBIA or Mask R-CNN) the segmentation accuracy was higher in the spatial resolution image from UAV and OBIA up to ED2 = 0.05. However, when the object to be segmented is larger than the pixel size of the image, the spatial resolution of the image is of secondary importance [37,57,76,77]. For this reason, as the scattered vegetation in this area presents a mean size of 100 m^2 [5], corresponding to 400 pixels of Google Earth image, only slight increases in segmentation accuracy were observed as the spatial resolution increased. Moreover, the overestimation of the area of each shrub was not significant as the images spatial resolution increased. Therefore, Google Earth images could be used to map scattered vegetation in drylands, if the plants to be mapped are larger than the pixel size. This result opens a wide range of new opportunities for vegetation mapping in remote areas where UAV or airborne image acquisition is difficult or acquiring commercial imagery of very high-resolution is very expensive. These results are promising and highlight the usefulness of free available Google Earth images for big shrubs mapping with only a negligible decrease in segmentation accuracy when compared with commercial UAV or airborne images. However, the segmentation of vegetation could be better if we use the near infrared NIR band since vegetation highlights in this range of the spectrum (e.g., 750 to 2500 nm) or used in vegetation indices such as the normalized difference vegetation index (NDVI) or Enhanced vegetation index (EVI). Finally, very high spatial resolution UAV images need much more computational time and are expensive and not always possible to obtain at larger scales in remote areas, hampering their use.

5. Conclusions

Our results showed that both OBIA and Mask R-CNN methods are powerful tools for mapping scattered vegetation in drylands. However, both methods were affected by the spatial resolution of the orthoimages utilized. We have shown for the first time that the fusion of the results from these methods increases, even more, the precision of the segmentation. This methodology should be tested on other types of vegetation or objects in order to prove to be fully effective. We propose an approach that offers a new way of fusing these methodologies to increase accuracy in the segmentation of scattered shrubs and should be evaluated on other objects in very high-resolution and hyperspectral images.

Using images with very high spatial resolution could provide the required precision to further develop methodologies to evaluate the spatial distribution of shrubs and dynamics of plant populations in global drylands, especially when utilizing free-to-use images, like the ones obtained from Google Earth. Such evaluation is of particular importance in drylands of developing countries, which are particularly sensitive to anthropogenic and climatic disturbances and may not have enough resources to acquire airborne or UAV imagery. For these reasons, future methodologies as the one presented in this work should focus on using freely available datasets.

In this context, the fusion of OBIA and Mask R-CNN could be extended to a larger number of classes of shrub and tree species or improved with the inclusion of more spectral and temporal information. Furthermore, this approach could improve the segmentation and monitoring of the crown of trees and arborescent shrubs in general, which are of particular importance for biodiversity conservation and for reducing uncertainties in carbon storages worldwide [78]. Recently, scattered trees have been identified as key structures for maintaining ecosystem services provision and high levels of biodiversity [43]. Global initiatives could benefit largely from CNNs, including those recently developed by FAO [79] to provide the forest extent in drylands. The uncertainties in this initiative [80,81] might be reduced implementing our approach CNN-based to segment trees. Tree and shrub segmentation methods could provide a global characterization of forest ecosystem structures and population abundances as part of the critical biodiversity variables initiative [82,83]. In long-lived shrubs where the precision of the segmentation is key for monitoring the detection of disturbances (e.g., pests, soil loss or seawater intrusion) [5]. Finally, the monitoring of persistent vegetation with minimal cover changes over decades could benefit from fusion approaches in the segmentation methods proposed.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

Abbreviation	Description
CCD	Charge-Coupled Device
ED2	Euclidean Distance v.2
ESP2	Estimation Scale Parameter v.2
ETRS	European Terrestrial Reference System
HB	Bottom-up Hierarchy
HT	Top-down Hierarchy
JSON	JavaScript Object Notation
NH	Non-Hierarchical
NSR	Number-of-Segmentation Ratio
OBIA	Object-based Image Analysis
R-CNN	Region—Convolutional Neural Networks
RGB	Red Green Blue
SPR	Segmentation Parameters Range
UAV	Unmanned aerial vehicle
UTM	Universal Transverse Mercator
VGG	Visual Geometry Group

References

1. Koutroulis, A.G. Dryland changes under different levels of global warming. *Sci. Total Environ.* **2019**, *655*, 482–511. [CrossRef] [PubMed]
2. Puigdefábregas, J. The role of vegetation patterns in structuring runoff and sediment fluxes in drylands. *Earth Surf. Process. Landf.* **2005**, *30*, 133–147. [CrossRef]
3. Ravi, S.; Breshears, D.D.; Huxman, T.E.; D'Odorico, P. Land degradation in drylands: Interactions among hydrologic–aeolian erosion and vegetation dynamics. *Geomorphology* **2010**, *116*, 236–245. [CrossRef]
4. Gao, Z.; Sun, B.; Li, Z.; Del Barrio, G.; Li, X. Desertification monitoring and assessment: A new remote sensing method. In Proceedings of the 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Beijing, China, 10–15 June 2016.
5. Guirado, E.; Blanco-Sacristán, J.; Rigol-Sánchez, J.; Alcaraz-Segura, D.; Cabello, J. A Multi-Temporal Object-Based Image Analysis to Detect Long-Lived Shrub Cover Changes in Drylands. *Remote Sens.* **2019**, *11*, 2649. [CrossRef]
6. Guirado, E.; Alcaraz-Segura, D.; Rigol-Sánchez, J.P.; Gisbert, J.; Martínez-Moreno, F.J.; Galindo-Zaldívar, J.; González-Castillo, L.; Cabello, J. Remote-sensing-derived fractures and shrub patterns to identify groundwater dependence. *Ecohydrology* **2018**, *11*, e1933. [CrossRef]
7. Guirado, E.; Tabik, S.; Alcaraz-Segura, D.; Cabello, J.; Herrera, F. Deep-learning Versus OBIA for Scattered Shrub Detection with Google Earth Imagery: *Ziziphus lotus* as Case Study. *Remote Sens.* **2017**, *9*, 1220. [CrossRef]
8. Kéfi, S.; Guttal, V.; Brock, W.A.; Carpenter, S.R.; Ellison, A.M.; Livina, V.N.; Seekell, D.A.; Scheffer, M.; van Nes, E.H.; Dakos, V. Early warning signals of ecological transitions: Methods for spatial patterns. *PLoS ONE* **2014**, *9*, e92097. [CrossRef]
9. Cunliffe, A.M.; Brazier, R.E.; Anderson, K. Ultra-fine grain landscape-scale quantification of dryland vegetation structure with drone-acquired structure-from-motion photogrammetry. *Remote Sens. Environ.* **2016**, *183*, 129–143. [CrossRef]
10. Brandt, M.; Hiernaux, P.; Rasmussen, K.; Mbow, C.; Kergoat, L.; Tagesson, T.; Ibrahim, Y.Z.; Wélé, A.; Tucker, C.J.; Fensholt, R. Assessing woody vegetation trends in Sahelian drylands using MODIS based seasonal metrics. *Remote Sens. Environ.* **2016**, *183*, 215–225. [CrossRef]
11. Berdugo, M.; Kéfi, S.; Soliveres, S.; Maestre, F.T. Author Correction: Plant spatial patterns identify alternative ecosystem multifunctionality states in global drylands. *Nat. Ecol. Evol.* **2018**, *2*, 574–576. [CrossRef]
12. Mohammadi, A.; Costelloe, J.F.; Ryu, D. Application of time series of remotely sensed normalized difference water, vegetation and moisture indices in characterizing flood dynamics of large-scale arid zone floodplains. *Remote Sens. Environ.* **2017**, *190*, 70–82. [CrossRef]
13. Tian, F.; Brandt, M.; Liu, Y.Y.; Verger, A.; Tagesson, T.; Diouf, A.A.; Rasmussen, K.; Mbow, C.; Wang, Y.; Fensholt, R. Remote sensing of vegetation dynamics in drylands: Evaluating vegetation optical depth (VOD) using AVHRR NDVI and in situ green biomass data over West African Sahel. *Remote Sens. Environ.* **2016**, *177*, 265–276. [CrossRef]
14. Taddeo, S.; Dronova, I.; Depsky, N. Spectral vegetation indices of wetland greenness: Responses to vegetation structure, composition, and spatial distribution. *Remote Sens. Environ.* **2019**, *234*, 111467. [CrossRef]
15. Blaschke, T. Object based image analysis for remote sensing. *ISPRS J. Photogramm. Remote Sens.* **2010**, *65*, 2–16. [CrossRef]
16. He, K.; Gkioxari, G.; Dollar, P.; Girshick, R. Mask R-CNN. In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017.

17. Zhang, J.; Jia, L. A comparison of pixel-based and object-based land cover classification methods in an arid/semi-arid environment of Northwestern China. In Proceedings of the 2014 Third International Workshop on Earth Observation and Remote Sensing Applications (EORSA), Changsha, China, 11–14 June 2014.

18. Amitrano, D.; Guida, R.; Iervolino, P. High Level Semantic Land Cover Classification of Multitemporal Sar Images Using Synergic Pixel-Based and Object-Based Methods. In Proceedings of the IGARSS 2019—2019 IEEE International Geoscience and Remote Sensing Symposium, Yokohama, Japan, 28 July–2 August 2019.

19. Li, S.; Yan, M.; Xu, J. Garbage object recognition and classification based on Mask Scoring RCNN. In Proceedings of the 2020 International Conference on Culture-oriented Science & Technology (ICCST), Beijing, China, 28–31 October 2020.

20. Zhang, Q.; Chang, X.; Bian, S.B. Vehicle-Damage-Detection Segmentation Algorithm Based on Improved Mask RCNN. *IEEE Access* **2020**, *8*, 6997–7004. [CrossRef]

21. Ghassemian, H. A review of remote sensing image fusion methods. *Inf. Fusion* **2016**, *32*, 75–89. [CrossRef]

22. Belgiu, M.; Stein, A. Spatiotemporal Image Fusion in Remote Sensing. *Remote Sens.* **2019**, *11*, 818. [CrossRef]

23. Moreno-Martínez, Á.; Izquierdo-Verdiguier, E.; Maneta, M.P.; Camps-Valls, G.; Robinson, N.; Muñoz-Marí, J.; Sedano, F.; Clinton, N.; Running, S.W. Multispectral high resolution sensor fusion for smoothing and gap-filling in the cloud. *Remote Sens. Environ.* **2020**, *247*, 111901. [CrossRef]

24. Alphan, H.; Çelik, N. Monitoring changes in landscape pattern: Use of Ikonos and Quickbird images. *Environ. Monit. Assess.* **2016**, *188*, 81. [CrossRef]

25. Mahdianpari, M.; Granger, J.E.; Mohammadimanesh, F.; Warren, S.; Puestow, T.; Salehi, B.; Brisco, B. Smart solutions for smart cities: Urban wetland mapping using very-high resolution satellite imagery and airborne LiDAR data in the City of St. John's, NL, Canada. *J. Environ. Manag.* **2020**, 111676, In press. [CrossRef]

26. Mahdavi Saeidi, A.; Babaie Kafaky, S.; Mataji, A. Detecting the development stages of natural forests in northern Iran with different algorithms and high-resolution data from GeoEye-1. *Environ. Monit. Assess.* **2020**, *192*, 653. [CrossRef] [PubMed]

27. Fawcett, D.; Bennie, J.; Anderson, K. Monitoring spring phenology of individual tree crowns using drone—Acquired NDVI data. *Remote Sens. Ecol. Conserv.* **2020**. [CrossRef]

28. Hu, Q.; Wu, W.; Xia, T.; Yu, Q.; Yang, P.; Li, Z.; Song, Q. Exploring the Use of Google Earth Imagery and Object-Based Methods in Land Use/Cover Mapping. *Remote Sens.* **2013**, *5*, 6026–6042. [CrossRef]

29. Venkatappa, M.; Sasaki, N.; Shrestha, R.P.; Tripathi, N.K.; Ma, H.O. Determination of Vegetation Thresholds for Assessing Land Use and Land Use Changes in Cambodia using the Google Earth Engine Cloud-Computing Platform. *Remote Sens.* **2019**, *11*, 1514. [CrossRef]

30. Sowmya, D.R.; Deepa Shenoy, P.; Venugopal, K.R. Feature-based Land Use/Land Cover Classification of Google Earth Imagery. In Proceedings of the 2019 IEEE 5th International Conference for Convergence in Technology (I2CT), Bombay, India, 29–31 March 2019.

31. Li, W.; Buitenwerf, R.; Munk, M.; Bøcher, P.K.; Svenning, J.-C. Deep-learning based high-resolution mapping shows woody vegetation densification in greater Maasai Mara ecosystem. *Remote Sens. Environ.* **2020**, *247*, 111953. [CrossRef]

32. Uyeda, K.A.; Stow, D.A.; Richart, C.H. Assessment of volunteered geographic information for vegetation mapping. *Environ. Monit. Assess.* **2020**, *192*, 1–14. [CrossRef]

33. Ancin–Murguzur, F.J.; Munoz, L.; Monz, C.; Hausner, V.H. Drones as a tool to monitor human impacts and vegetation changes in parks and protected areas. *Remote Sens. Ecol. Conserv.* **2020**, *6*, 105–113. [CrossRef]

34. Yu, Q.; Gong, P.; Clinton, N.; Biging, G.; Kelly, M.; Schirokauer, D. Object-based Detailed Vegetation Classification with Airborne High Spatial Resolution Remote Sensing Imagery. *Photogramm. Eng. Remote Sens.* **2006**, *72*, 799–811. [CrossRef]

35. Laliberte, A.S.; Herrick, J.E.; Rango, A.; Winters, C. Acquisition, Orthorectification, and Object-based Classification of Unmanned Aerial Vehicle (UAV) Imagery for Rangeland Monitoring. *Photogramm. Eng. Remote Sens.* **2010**, *76*, 661–672. [CrossRef]

36. Whiteside, T.G.; Boggs, G.S.; Maier, S.W. Comparing object-based and pixel-based classifications for mapping savannas. *Int. J. Appl. Earth Obs. Geoinf.* **2011**, *13*, 884–893. [CrossRef]

37. Hossain, M.D.; Chen, D. Segmentation for Object-Based Image Analysis (OBIA): A review of algorithms and challenges from remote sensing perspective. *ISPRS J. Photogramm. Remote Sens.* **2019**, *150*, 115–134. [CrossRef]

38. Arvor, D.; Durieux, L.; Andrés, S.; Laporte, M.-A. Advances in Geographic Object-Based Image Analysis with ontologies: A review of main contributions and limitations from a remote sensing perspective. *ISPRS J. Photogramm. Remote Sens.* **2013**, *82*, 125–137. [CrossRef]

39. Johnson, B.A.; Ma, L. Image Segmentation and Object-Based Image Analysis for Environmental Monitoring: Recent Areas of Interest, Researchers' Views on the Future Priorities. *Remote Sens.* **2020**, *12*, 1772. [CrossRef]

40. Guo, Y.; Liu, Y.; Georgiou, T.; Lew, M.S. A review of semantic segmentation using deep neural networks. *Int. J. Multimed. Inf. Retr.* **2018**, *7*, 87–93. [CrossRef]

41. Singh, R.; Rani, R. Semantic Segmentation using Deep Convolutional Neural Network: A Review. *SSRN Electron. J.* **2020**. [CrossRef]

42. Aguilar, M.A.; Aguilar, F.J.; García Lorca, A.; Guirado, E.; Betlej, M.; Cichon, P.; Nemmaoui, A.; Vallario, A.; Parente, C. Assessment of multiresolution segmentation for extracting greenhouses from worldview-2 imagery. *ISPRS-Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2016**, *XLI-B7*, 145–152. [CrossRef]

43. Brandt, M.; Tucker, C.J.; Kariryaa, A.; Rasmussen, K.; Abel, C.; Small, J.; Chave, J.; Rasmussen, L.V.; Hiernaux, P.; Diouf, A.A.; et al. An unexpectedly large count of trees in the West African Sahara and Sahel. *Nature* **2020**, *587*, 78–82. [CrossRef]

44. Guirado, E.; Tabik, S.; Rivas, M.L.; Alcaraz-Segura, D.; Herrera, F. Whale counting in satellite and aerial images with deep learning. *Sci. Rep.* **2019**, *9*, 14259. [CrossRef]

45. Guirado, E.; Alcaraz-Segura, D.; Cabello, J.; Puertas-Ruíz, S.; Herrera, F.; Tabik, S. Tree Cover Estimation in Global Drylands from Space Using Deep Learning. *Remote Sens.* **2020**, *12*, 343. [CrossRef]

46. Zhou, Y.; Zhang, R.; Wang, S.; Wang, F. Feature Selection Method Based on High-Resolution Remote Sensing Images and the Effect of Sensitive Features on Classification Accuracy. *Sensors* **2018**, *18*, 2013. [CrossRef]

47. Foody, G.; Pal, M.; Rocchini, D.; Garzon-Lopez, C.; Bastin, L. The Sensitivity of Mapping Methods to Reference Data Quality: Training Supervised Image Classifications with Imperfect Reference Data. *ISPRS Int. J. Geo-Inf.* **2016**, *5*, 199. [CrossRef]

48. Lu, D.; Weng, Q. A survey of image classification methods and techniques for improving classification performance. *Int. J. Remote Sens.* **2007**, *28*, 823–870. [CrossRef]

49. Torres-García, M.; Salinas-Bonillo, M.J.; Gázquez-Sánchez, F.; Fernández-Cortés, A.; Querejeta, J.L.; Cabello, J. Squandering Water in Drylands: The Water Use Strategy of the Phreatophyte *Ziziphus lotus* (L.) Lam in a Groundwater Dependent Ecosystem. *Am. J. Bot.* **2021**, *108*, 2, in press.

50. Tirado, R.; Pugnaire, F.I. Shrub spatial aggregation and consequences for reproductive success. *Oecologia* **2003**, *136*, 296–301. [CrossRef] [PubMed]

51. Tengberg, A.; Chen, D. A comparative analysis of nebkhas in central Tunisia and northern Burkina Faso. *Geomorphology* **1998**, *22*, 181–192. [CrossRef]

52. Fisher, G.B.; Burch Fisher, G.; Amos, C.B.; Bookhagen, B.; Burbank, D.W.; Godard, V. Channel widths, landslides, faults, and beyond: The new world order of high-spatial resolution Google Earth imagery in the study of earth surface processes. *Google Earth Virtual Vis. Geosci. Educ. Res.* **2012**, *492*, 1–22. [CrossRef]

53. Li, M.; Ma, L.; Blaschke, T.; Cheng, L.; Tiede, D. A systematic comparison of different object-based classification techniques using high spatial resolution imagery in agricultural environments. *Int. J. Appl. Earth Obs. Geoinf.* **2016**, *49*, 87–98. [CrossRef]

54. Yu, W.; Zhou, W.; Qian, Y.; Yan, J. A new approach for land cover classification and change analysis: Integrating backdating and an object-based method. *Remote Sens. Environ.* **2016**, *177*, 37–47. [CrossRef]

55. Yan, J.; Lin, L.; Zhou, W.; Ma, K.; Pickett, S.T.A. A novel approach for quantifying particulate matter distribution on leaf surface by combining SEM and object-based image analysis. *Remote Sens. Environ.* **2016**, *173*, 156–161. [CrossRef]

56. Colkesen, I.; Kavzoglu, T. Selection of Optimal Object Features in Object-Based Image Analysis Using Filter-Based Algorithms. *J. Indian Soc. Remote Sens.* **2018**, *46*, 1233–1242. [CrossRef]

57. Lefèvre, S.; Sheeren, D.; Tasar, O. A Generic Framework for Combining Multiple Segmentations in Geographic Object-Based Image Analysis. *ISPRS Int. J. Geo-Inf.* **2019**, *8*, 70. [CrossRef]

58. Hurskainen, P.; Adhikari, H.; Siljander, M.; Pellikka, P.K.E.; Hemp, A. Auxiliary datasets improve accuracy of object-based land use/land cover classification in heterogeneous savanna landscapes. *Remote Sens. Environ.* **2019**, *233*, 111354. [CrossRef]

59. Gonçalves, J.; Pôças, I.; Marcos, B.; Mücher, C.A.; Honrado, J.P. SegOptim—A new R package for optimizing object-based image analyses of high-spatial resolution remotely-sensed data. *Int. J. Appl. Earth Obs. Geoinf.* **2019**, *76*, 218–230. [CrossRef]

60. Ma, L.; Cheng, L.; Li, M.; Liu, Y.; Ma, X. Training set size, scale, and features in Geographic Object-Based Image Analysis of very high resolution unmanned aerial vehicle imagery. *ISPRS J. Photogramm. Remote Sens.* **2015**, *102*, 14–27. [CrossRef]

61. Zhang, X.; Du, S.; Ming, D. Segmentation Scale Selection in Geographic Object-Based Image Analysis. *High Spat. Resolut. Remote Sens.* **2018**, 201–228.

62. Yang, L.; Mansaray, L.; Huang, J.; Wang, L. Optimal Segmentation Scale Parameter, Feature Subset and Classification Algorithm for Geographic Object-Based Crop Recognition Using Multisource Satellite Imagery. *Remote Sens.* **2019**, *11*, 514. [CrossRef]

63. Mao, C.; Meng, W.; Shi, C.; Wu, C.; Zhang, J. A Crop Disease Image Recognition Algorithm Based on Feature Extraction and Image Segmentation. *Traitement Signal* **2020**, *37*, 341–346. [CrossRef]

64. 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. [CrossRef]

65. Torres-Sánchez, J.; López-Granados, F.; Peña, J.M. An automatic object-based method for optimal thresholding in UAV images: Application for vegetation detection in herbaceous crops. *Comput. Electron. Agric.* **2015**, *114*, 43–52. [CrossRef]

66. Josselin, D.; Louvet, R. Impact of the Scale on Several Metrics Used in Geographical Object-Based Image Analysis: Does GEOBIA Mitigate the Modifiable Areal Unit Problem (MAUP)? *ISPRS Int. J. Geo-Inf.* **2019**, *8*, 156. [CrossRef]

67. Blaschke, T.; Lang, S.; Hay, G. *Object-Based Image Analysis: Spatial Concepts for Knowledge-Driven Remote Sensing Applications*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2008; ISBN 9783540770589.

68. Watanabe, T.; Wolf, D.F. Instance Segmentation as Image Segmentation Annotation. In Proceedings of the 2019 IEEE Intelligent Vehicles Symposium (IV), Paris, France, 9–12 June 2019.

69. Demir, A.; Yilmaz, F.; Kose, O. Early detection of skin cancer using deep learning architectures: Resnet-101 and inception-v3. In Proceedings of the 2019 Medical Technologies Congress (TIPTEKNO), Izmir, Turkey, 3–5 October 2019.

70. 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. [CrossRef]

71. Nussbaum, S.; Menz, G. eCognition Image Analysis Software. In *Object-Based Image Analysis and Treaty Verification*; Springer: Berlin/Heidelberg, Germany, 2008; pp. 29–39.

72. Dutta, A.; Gupta, A.; Zissermann, A. VGG Image Annotator (VIA). Available online: http://www.robots.ox.ac.uk/~{}vgg/software/via (accessed on 11 December 2020).

73. Shin, H.-C.; 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. [CrossRef] [PubMed]

74. Caesar, H.; Uijlings, J.; Ferrari, V. COCO-Stuff: Thing and Stuff Classes in Context. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018.

75. Wang, B.; Li, C.; Pavlu, V.; Aslam, J. A Pipeline for Optimizing F1-Measure in Multi-label Text Classification. In Proceedings of the 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA), Orlando, FL, USA, 17–20 December 2018.

76. Zhan, Q.; Molenaar, M.; Tempfli, K.; Shi, W. Quality assessment for geo–spatial objects derived from remotely sensed data. *Int. J. Remote Sens.* **2005**, *26*, 2953–2974. [CrossRef]

77. Chen, G.; Weng, Q.; Hay, G.J.; He, Y. Geographic object-based image analysis (GEOBIA): Emerging trends and future opportunities. *Gisci. Remote Sens.* **2018**, *55*, 159–182. [CrossRef]

78. Cook-Patton, S.C.; Leavitt, S.M.; Gibbs, D.; Harris, N.L.; Lister, K.; Anderson-Teixeira, K.J.; Briggs, R.D.; Chazdon, R.L.; Crowther, T.W.; Ellis, P.W.; et al. Mapping carbon accumulation potential from global natural forest regrowth. *Nature* **2020**, *585*, 545–550. [CrossRef] [PubMed]

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

80. 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. [CrossRef]

81. 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. [CrossRef]

82. Fernández, N.; Ferrier, S.; Navarro, L.M.; Pereira, H.M. Essential Biodiversity Variables: Integrating In-Situ Observations and Remote Sensing Through Modeling. *Remote Sens. Plant Biodivers.* **2020**, *18*, 485–501.

83. Vihervaara, P.; Auvinen, A.-P.; Mononen, L.; Törmä, M.; Ahlroth, P.; Anttila, S.; Böttcher, K.; Forsius, M.; Heino, J.; Heliölä, J.; et al. How Essential Biodiversity Variables and remote sensing can help national biodiversity monitoring. *Glob. Ecol. Conserv.* **2017**, *10*, 43–59. [CrossRef]