International Journal of Geo-Information

MDPI

Article

Instance Segmentation for Governmental Inspection of Small Touristic Infrastructure in Beach Zones Using Multispectral High-Resolution WorldView-3 Imagery

Osmar Luiz Ferreira de Carvalho [1], Rebeca dos Santos de Moura [2], Anesmar Olino de Albuquerque [2], Pablo Pozzobon de Bem [2], Rubens de Castro Pereira [3,4], Li Weigang [1], Dibio Leandro Borges [1], Renato Fontes Guimarães [2], Roberto Arnaldo Trancoso Gomes [2] and Osmar Abílio de Carvalho Júnior [2,*]

1 Department of Computer Science, University of Brasília, Brasília 70910-900, Brazil; osmarcarvalho@ieee.org (O.L.F.d.C.); weigang@unb.br (L.W.); dibio@unb.br (D.L.B.)
2 Department of Geography, University of Brasília, Brasília 70910-900, Brazil; moura.santos@aluno.unb.br (R.d.S.d.M.); anesmar@ieee.org (A.O.d.A.); pablo.bem@aluno.unb.br (P.P.d.B.); renatofg@unb.br (R.F.G.); robertogomes@unb.br (R.A.T.G.)
3 Institute of Informatics, Federal University of Goiás, Goiânia 74690-900, Brazil; rubens.castro@ufg.br
4 Information Technology Center, EMBRAPA Rice and Beans, Santo Antônio de Goiás 86085-981, Brazil
* Correspondence: osmarjr@unb.br

check for updates

Citation: de Carvalho, O.L.F.; de Moura, R.d.S.; de Albuquerque, A.O.; de Bem, P.P.; de Castro Pereira, R.; Weigang, L.; Borges, D.L.; Guimarães, R.F.; Gomes, R.A.T.; de Carvalho Júnior, O.A. Instance Segmentation for Governmental Inspection of Small Touristic Infrastructure in Beach Zones Using Multispectral High-Resolution WorldView-3 Imagery. *ISPRS Int. J. Geo-Inf.* **2021**, *10*, 813. https://doi.org/10.3390/ijgi10120813

Academic Editors: Wolfgang Kainz, Cristina Ponte Lira and Rita González-Villanueva

Received: 5 October 2021
Accepted: 26 November 2021
Published: 30 November 2021

Abstract: Misappropriation of public lands is an ongoing government concern. In Brazil, the beach zone is public property, but many private establishments use it for economic purposes, requiring constant inspection. Among the undue targets, the individual mapping of straw beach umbrellas (SBUs) attached to the sand is a great challenge due to their small size, high presence, and agglutinated appearance. This study aims to automatically detect and count SBUs on public beaches using high-resolution images and instance segmentation, obtaining pixel-wise semantic information and individual object detection. This study is the first instance segmentation application on coastal areas and the first using WorldView-3 (WV-3) images. We used the Mask-RCNN with some modifications: (a) multispectral input for the WorldView3 imagery (eight channels), (b) improved the sliding window algorithm for large image classification, and (c) comparison of different image resizing ratios to improve small object detection since the SBUs are small objects (<32^2 pixels) even using high-resolution images (31 cm). The accuracy analysis used standard COCO metrics considering the original image and three scale ratios (2×, 4×, and 8× resolution increase). The average precision (AP) results increased proportionally to the image resolution: 30.49% (original image), 48.24% (2×), 53.45% (4×), and 58.11% (8×). The 8× model presented 94% AP50, classifying nearly all SBUs correctly. Moreover, the improved sliding window approach enables the classification of large areas providing automatic counting and estimating the size of the objects, proving to be effective for inspecting large coastal areas and providing insightful information for public managers. This remote sensing application impacts the inspection cost, tribute, and environmental conditions.

Keywords: Mask-RCNN; multispectral; deep learning; object detection

1. Introduction

Public land management is essential for the effective use of natural resources with implications for economic, social, and environmental issues [1]. Government policies establish public areas in ecological, social, or safety-relevant regions (i.e., natural fields and historic spaces), offering services ranging from natural protection to recreation [1,2]. However, managing public interests to promote social welfare over private goals is a significant challenge. Especially in developing countries, recurrent misuse of public land [3], and illegal invasions (i.e., the use of public lands for private interests) [4] are among the most common problems.

Coastal zone areas concentrate a large part of the world population, despite being environmentally sensitive with intense natural processes (erosion, accretion, and natural disasters) [5] and constant anthropic threats (marine litter, pollution, and inappropriate use) [6,7]. The coastal zone is a priority for developing programs for continuous monitoring and misuse detection. In Brazil, coastal areas belong to the Federal Government, considering the distance of 33 meters from the high-medium water line in 1831 (known as "navy land"). Beaches and water bodies have guaranteed public access according to the Brazilian Forest Code. Therefore, Brazilian legislation establishes measures for public use, economic exploitation, environmental preservation, and recovery considering coastal areas' socio-environmental function. The inspection of beach areas in Brazil is a challenge, as the Union's Heritage Secretariat does not have complete and accurate information about this illegal occupation throughout the country. The undue economic exploitation of the urban beach strip leads to an increase in the number of illegal constructions, a reduction in government revenue due to non-registration, environmental problems, visual pollution, beach litter, among others. Many illegal installations in urban beaches are masonry constructions for private or commercial use. In addition, tourist infrastructure for food and leisure extends several straw beach umbrellas (SBUs) (fixed in the sand by local traders) to the sand strip without permission. Given the potential impact on the environment and the local economy, the monitoring and enforcement to curb private business development in public spaces must be constant and efficient [5], mainly to avoid uncontrolled tourism development [8,9]. The inspection must ensure the legal requirements, avoid frequent changes that lead to lawful gaps, and minimize differences arising from conflicts of interest.

Conventionally, the inspection process imposes a heavy burden on state and federal agencies, containing few inspectors with low frequency on site. In this regard, geospatial technologies and remote sensing techniques are valuable for public managers since they enable monitoring changes in the landscapes and understanding different patterns and behaviors. Thus, an excellent potential for remote sensing application by government control agencies is detecting unauthorized constructions in urban areas [10,11]. Several review articles address the use of remote sensing and geospatial technology in coastal studies [12–16]. Currently, geospatial technology is a key factor for the development and implementation of an integrated coastal management, allowing a spatial analysis for studies of environmental vulnerability, landform change (erosion and accretion), disaster management, protected areas, ecosystem, economic, and risk assessment [17–20].

However, few remote sensing studies focus on the detection of tourist infrastructure objects on the beach for inspection. Beach inspection requires high-resolution images and digital image processing algorithms that identify, count, and segment small objects of interest, such as the SBUs. Among the remote sensing data, high-resolution orbital images have the advantage of periodic availability and coverage of large areas at a moderate cost, unlike aerial photographs and unmanned aircraft systems (UASs) of limited accessibility. Typically, high-resolution satellite images acquire a panchromatic band (from 1 meter to sub-metric resolutions) and multispectral bands (spectral bands of blue, green, red, and near-infrared with spatial resolutions ranging from 1 to 4 m), such as IKONOS (Panchromatic: 1 m; Multispectral: 4 m), OrbView-3 (Panchromatic: 1 m; Multispectral: 4 m), QuickBird (Panchromatic: 0.6 m; Multispectral: 2.4 m), GeoEye-1 (Panchromatic: 0.41 m; Multispectral: 1.65 m), and Pleiades (Panchromatic: 0.5 m; Multispectral: 2 m). Unlike the satellites mentioned above, the WorldView-2 (WV2) and WorldView-3 (WV3) images present a differential for combining the panchromatic band (0.3 m resolution) with eight multispectral bands (Resolution 1, 24 m): coastal (400–450 nm), blue (450–510 nm), green (510–580 nm), yellow (585–625 nm), red (655–690 nm), red edge (705–745 nm), near-infrared 1 (NIR1) (770–895 nm), and near-infrared 2 (NIR2) (860–1040 nm). Therefore, WorldView-2 and WorldView-3 have additional spectral bands compared to other sensors (coastal, yellow, red edge, and NIR2), valuable for urban mapping [21]. Therefore, the conjunction of the spectral and spatial properties of the WorldView-2 and WorldView-3 images is an advantage in the detailed classification process in complex urban environments. Few studies

assess infrastructure detection on the beach. Llausàs et al. [22] conducted a study on private swimming pools on the Catalan coast to estimate water use from WorldView-2 images and Geographic Object-Based Image Analysis (GEOBIA). Papakonstantinou et al. [23] used UAS images and GEOBIA to detect tourist structures in the coastal region of the Santorini and Lesvos islands. Despite the wide use of the GEOBIA, deep learning (DL) segmentation techniques demonstrate greater efficiency than GEOBIA in the following factors: (a) greater precision and efficiency; (b) high ability to transfer knowledge to other environments and different attributes of objects (light, color, size, shape, and background); (c) requires less human supervision; and (d) less noise interference [24–27].

The DL methods promote a revolution in several fields of science, including visual recognition [28], natural language processing [29,30], speech recognition [31,32], object detection [33,34], medical image analysis [35–37], person identification [38–40], and drug discovery [41–43] and genomics [44,45]. Like other fields of knowledge, DL achieves state-of-the-art performance in remote sensing [46–48] with a significant increase in articles after 2014 [49]. In a short period, several review articles have reported about DL and sensing, considering different applications [50,51], digital image processing methods [46,49,52–55], types of images [56–60], and environmental studies [61]. DL algorithms use neural networks [62], a structure composed of weighted connections between neurons that iteratively learn high and low-level features such as textures and shapes through gradient descent [63]. Moreover, convolutional neural networks (CNN) have great usability in image processing because of their ability to process data in multi-dimensional arrays [64]. There are many applications with CNN models, e.g., classification, object detection, semantic segmentation, instance segmentation, among others [46]. The best method often depends on the problem specification.

Nonetheless, instance segmentation and object detection networks enable a distinct identification for elements belonging to the same class, suitable for multi-object identification and counting. A drawback when comparing instance segmentation and object detection networks is real-time processing, in which instance segmentation usually presents an inference speed lower than object detection. Nevertheless, instance segmentation models bring more pixel-wise information, crucial to determining the exact object dimensions.

However, instance segmentation brings difficulties in its implementation. The first is the annotation format, where most instance segmentation models use a specific annotation format that is not straightforward from traditional annotations. The second is that most algorithm uses conventional red, green, and blue (RGB) images, whereas remote sensing images often present more spectral channels and varied dimensions. The third problem is adjusting the training images to a specific size to train the models. To classify a large area requires post-processing procedures. Object detection algorithms require only the bounding box coordinates, which are much more straightforward than instance segmentation that requires each object's bounding boxes and polygons.

Another recurrent problem is the poor performance of DL algorithms on small objects since they present low resolutions and a noisy representation [65]. Common objects in context (COCO) [66] characterizes objects sizes within three categories: (a) small objects (area $< 32^2$ pixels); (b) medium objects ($32^2 <$ area $< 96^2$); and (c) large objects (area $> 96^2$ pixels). The average precision (AP) score (main metric) has nearly half of the performance on small objects within the COCO challenge than on medium and large objects. According to a review article by Tong et al. [67], few studies focus on small object detection, and despite the subject's relevance, the current state is far from acceptable in most scenarios and still underrepresented in the remote sensing field. In this regard, an effective method is to increase the image dimensions. In this way, the small objects will have more pixels, differentiating them from noise.

The present research aims to effectively identify, count, and estimate SBU areas using multispectral WordView-3 (WV-3) imagery and instance segmentation to inspect and control tourist infrastructure properly. Very few works use instance segmentation on the remote sensing field, and none of those use WV-3 images or in beach areas. Thus,

our contributions are threefold: (1) a novel application of instance segmentation using multispectral WV-3 images on beach areas, (2) leverage the existing method for classifying large areas using instance segmentation, and (3) analyze and compare the effect of the DL image tiles and their metrics.

2. Materials and Methods

The methodology is subdivided into the following steps: (A) dataset; (B) instance segmentation approach; (C) image mosaicking using sliding window; and (D) performance metrics (Figure 1).

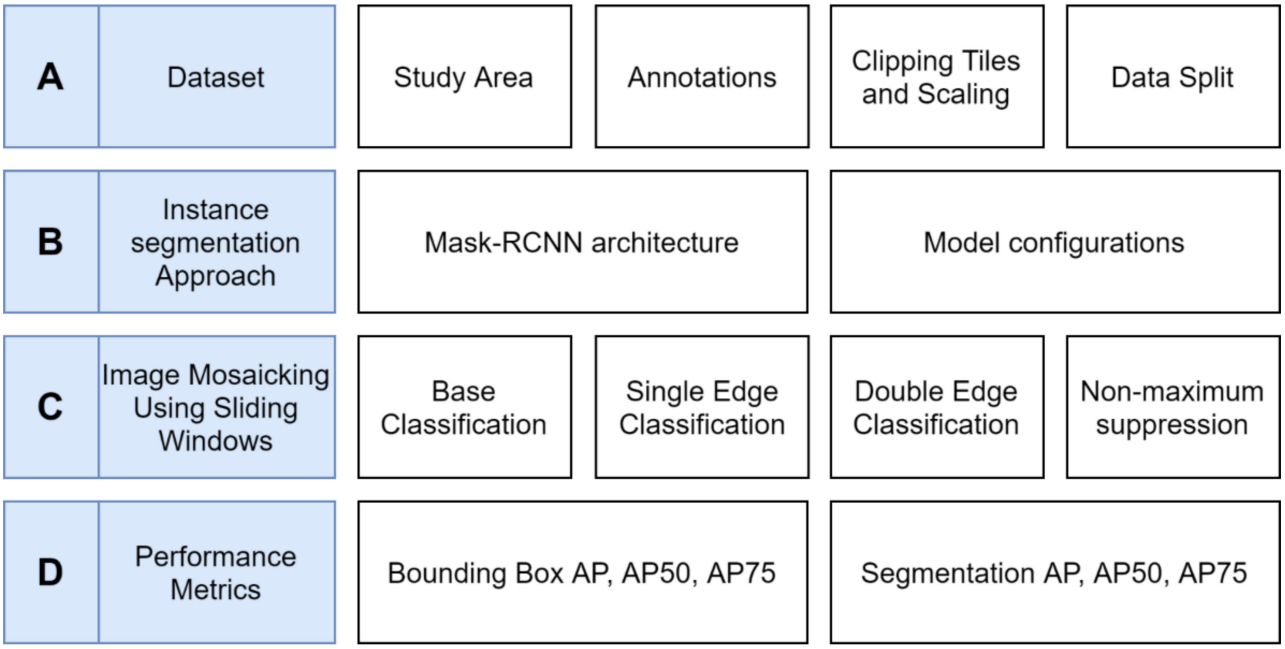

Figure 1. Methodological flowchart.

2.1. Dataset

2.1.1. Study Area

The study area was Praia do Futuro in Fortaleza, Ceará, Brazil, with intense tourist activities (Figure 2). The research used WorldView-3 images from 17 September 2017 and 18 September 2018, provided by the European Space Agency (ESA) with a total length of 400 km^2. The WorldView-3 images combine the acquisition of panchromatic (with 0.31 m resolution) and multispectral (with 1.2 m resolution and eight spectral bands) bands. Thus, we use the Gram–Schmidt pan-sharpening method [68] with nearest neighbor resampling to maximize image resolution and preserve spectral values [69]. The pan-sharpening technique aims to combine the multispectral images (with low spatial resolution and narrow spectral band) with the panchromatic image (with high spatial resolution and wide spectral band), extracting the best characteristics of both data and merging in a product that favors the data interpretation [70]. The Gram–Schmidt technique presents high fidelity in rendering spatial features, being a fast and straightforward method.

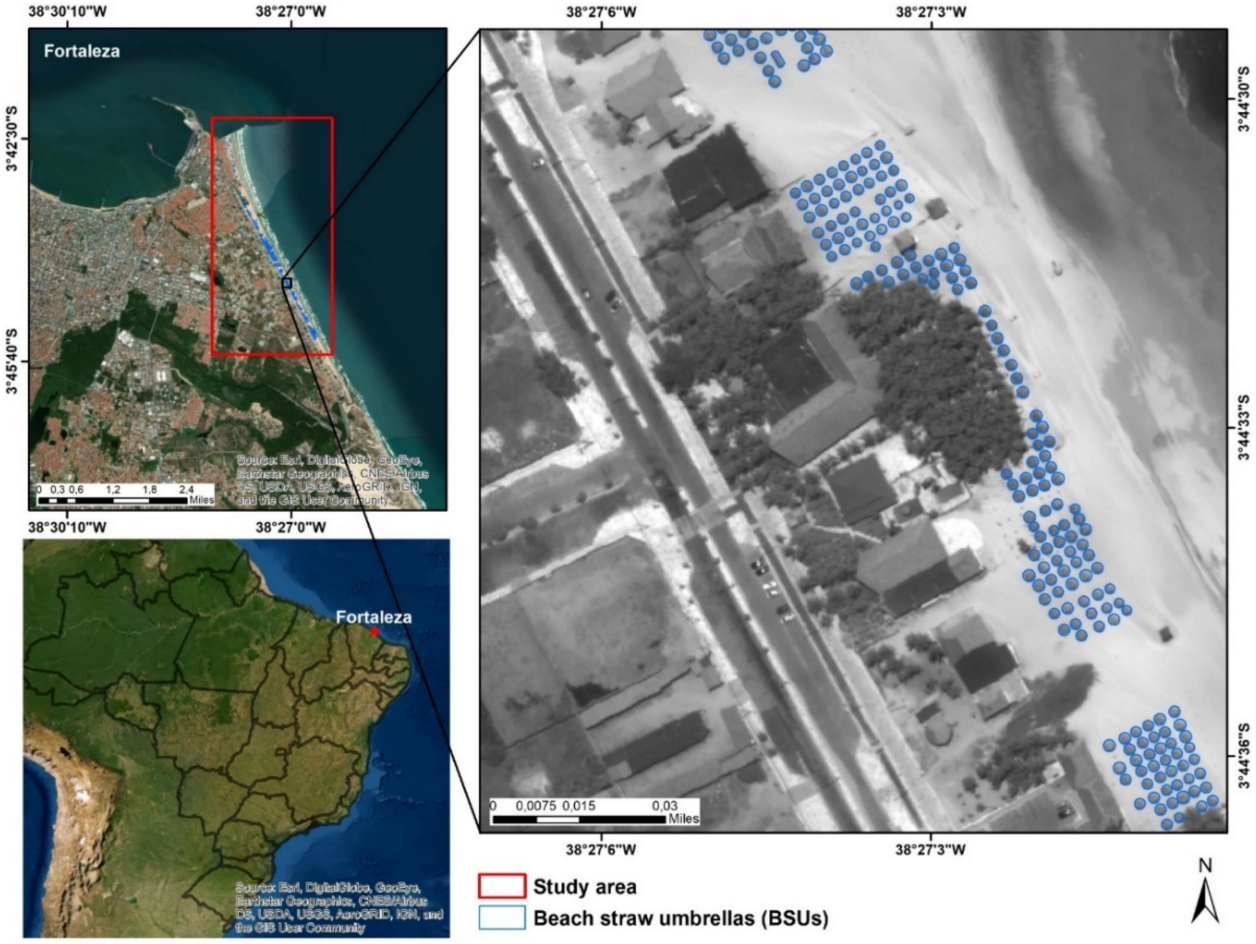

Figure 2. Study area with a zoomed area from the WorldView-3 image.

2.1.2. Annotations

Annotations assign specific labels to the objects of interest, consisting of the ground truth in model training. Instance segmentation programs use the COCO annotation format, such as Detectron2 software [71] with the Mask-RCNN model [72]. Consequently, several annotation tools have been proposed for traditional photographic images considering the COCO format, such as LabelMe [73,74], Computer Vision Annotation Tool (CVAT) [75], RectLabel (https://rectlabel.com, accessed on 5 October 2021), Labelbox (https://labelbox.com, accessed on 5 October 2021), and Visual Object Tagging Tool (VoTT) (https://github.com/microsoft/VoTT, accessed on 5 October 2021). In remote sensing studies, an extensive collection of annotation tools is present in Geographic Information Systems (GIS) with several procedures to capture, store, edit and display georeferenced data. Therefore, an alternative to taking advantage of all the technology developed for spatial data is to convert the output data from the GIS program to COCO annotation format. In the present research, we converted GIS data to the COCO annotation format [66] from the program developed in the C++ language proposed by Carvalho et al. [76]. Thus, the SBUs' ground truth digitization used ArcGIS software. Since instance segmentation requires a unique identifier (ID) for each object, each SBU had a different value (from 1 to N, with N being the total number of SBUs).

2.1.3. Clipping Tiles and Scaling

Our research targets are very small ($<16^2$ pixels) and very crowded in most cases. A powerful yet straightforward operation to improve small objects' detection is to scale the input image [67]. We evaluated the ratios of $2\times$, $4\times$, and $8\times$ the original image. The cropped

tiles considered 64 × 64 pixels in the original image, which increased proportionally with the different scaling ratios (128 × 128, 256 × 256, and 512 × 512, respectively).

2.1.4. Data Split

For supervised DL tasks, the usage of three sets is beneficial to evaluate the proposed model. The training set usually presents most of the samples, which is where the algorithm will understand the patterns. However, the training set alone is insufficient since the final model may be overfitting or underfitting. In this regard, the validation set plays a crucial role in keeping track of the model progress. A common approach is to save the model with the best performance on the validation set. Nevertheless, this procedure also brings a bias. With that said, the model is often preferable to be done using an independent test set. Thus, we distributed the cropped tiles into training, validation, and test sets as listed in Table 1. The number of instances shows a high object concentration, with an average of nearly ten objects per 64 × 64 pixel image.

Table 1. Data split in the training validation and testing sets with their respective number of images and instances.

Set	Number of Images	Number of Instances
Train	185	1780
Validation	40	631
Test	45	780

2.2. Instance Segmentation Approach

2.2.1. Mask-RCNN Architecture

Facebook Artificial Intelligence Research (FAIR) introduced the Mask-region-based convolutional neural network (Mask-RCNN) as an extension of previous studies for object detection architectures: RCNN [77], Fast-RCNN [78], and Faster-RCNN [79]. The Mask-RCNN uses the Faster-RCNN as a basis with the addition of a segmentation branch that performs a binary segmentation on each detected bounding box using a fully convolutional network (FCN) [80] (Figure 3).

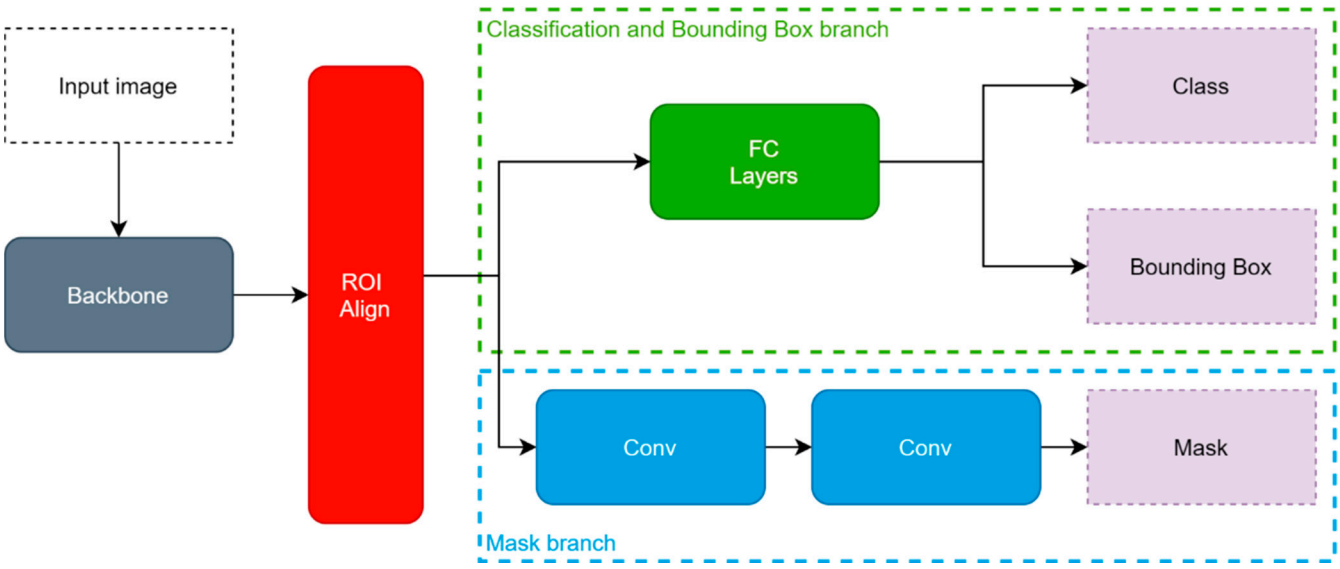

Figure 3. Mask-RCNN architecture.

The region-based algorithms present a backbone structure (e.g., ResNets [81], ResNeXts [82], or other CNNs) followed by a region proposal network (RPN). However, the Mask-RCNN has a region of interest (RoI) align mechanism, in contrast to the RoIPool. The benefit of

this method is a better alignment of each RoI with the inputs that removes any quantization problems on the RoI's boundaries. Succinctly, the model aims to identify the bounding boxes, classify the bounding box classes, and apply a pixel-wise mask on the bounding box objects. The loss function considers the three elements, being the sum of the bounding box loss ($Loss_{bbox}$), mask loss ($Loss_{mask}$), and classification loss ($Loss_{class}$), in which $Loss_{mask}$ and $Loss_{class}$ are log loss functions, and $Loss_{bbox}$ is the L1 loss.

Moreover, we use the Detectron2 [71] software, which uses the Pytorch framework. Since this architecture is usually applied to traditional images (3 channels), it requires some adjustments to be compatible with the WV-3 imagery (TIFF format and have more than three channels) [76].

2.2.2. Model Configurations

To train the Mask-RCNN model, we made the necessary source code changes for compatibility and applied z-score normalization based on the training set images. We only used the ResNeXt-101-FPN (X-101-FPN) backbone since the objective is to analyze scaling.

Regarding hyperparameters, we applied: (a) Stochastic gradient descent (SGD) optimizer with a learning rate of 0.001 (divided by ten after 500 iterations); (b) 256 ROIs per image; (c) five thousand iterations; (d) different anchor box scales for the original image (4, 8, 12, 16, 32), 2× scale image (8, 16, 24, 32, 64), 4× scale image (16, 32, 48, 64, 128), and 8× scale image (32, 64, 48, 128, 256). To avoid overfitting, we applied the following augmentations on the training images: (a) random horizontal flip, (b) random vertical flip, (c) random rotation. Finally, we used Nvidia GeForce RTX 2080 TI GPU with 11GB memory to process and train the model.

2.3. Image Mosaicking Using Sliding Windows

In remote sensing, the images often present interest areas much larger than the images used in training, validation, and testing. This problem requires some post-processing procedures. This process is not straightforward since the edges of the frames usually present errors. In this context, the sliding window technique has been used to various semantic segmentation problems [83–86], in which the authors establish a step value (usually less than the frame size) and take the average from the overlapping pixels to attenuate the border errors. The problem persists in object detection and instance segmentation since predictions from adjacent frames would output distinct partial predictions for the same object. Recently, de Carvalho et al. [76] proposed a mosaicking strategy for object detection using a base classifier (Figure 4B), vertical edge classifier (Figure 4C), and horizontal edge classifier (Figure 4E). Our research adapted the method by adding a double edge classifier since some errors may persist (https://github.com/osmarluiz/Straw-Beach-Umbrella-Detection, accessed on 5 October 2021).

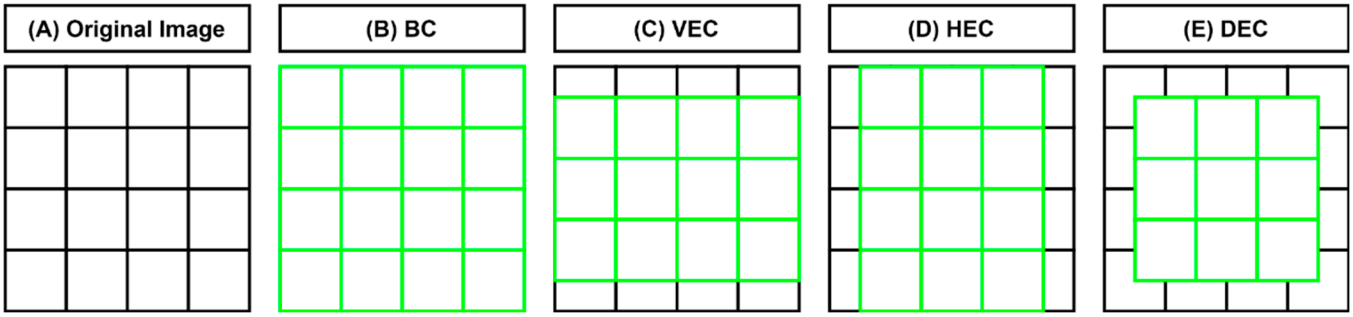

Figure 4. Scheme of the mosaicking procedure, with the (**A**) original image separation into tiles, (**B**) base classification (BC), (**C**) vertical edge classification (VEC), (**D**) horizontal edge classification (HEC), and (**E**) double edge classification (DEC).

2.3.1. Base Classification

The first step is to apply a base classifier (BC) (considering all elements) using a sliding window starting at x = 0 and y = 0, and stride values of 512 (Figure 5B). This procedure produces partial classification on the frame's edges between consecutive frames, resulting in more than one imperfect classification for the same object, which is a misleading result.

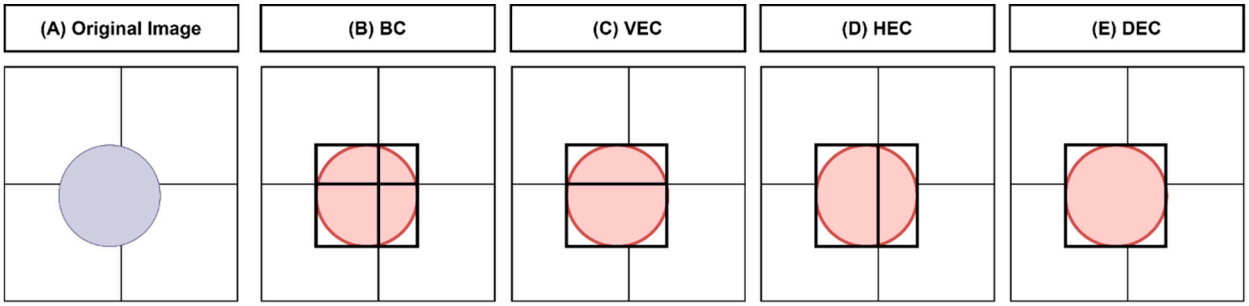

Figure 5. Example of classifications from each mosaicking procedure considering the (**A**) original image, (**B**) base classifier (BC), (**C**) vertical edge classifier (VEC), (**D**) horizontal edge classifier (HEC), and (**E**) double edge classifier.

2.3.2. Single Edge Classification

The second step is to classify objects located in the borders (partially classified objects by the BC). We applied the vertical edge classifier (VEC) to classify elements in consecutive frames vertical-wise, composed of a sliding window that starts at x = 256 and y = 0 (Figure 5C). Similarly, to horizontal-wise consecutive frames, we applied the horizontal edge classifier (HEC), with a sliding window that starts at x = 0 and y = 256 (Figure 5D). Both strategies use 512-pixel strides. In addition, to avoid the high computational cost, the VEC and HED only classify objects that start before the center of the image (x < 256 for the VEC and y < 256 for the HEC) and end after the image's center (x > 256 for the VEC and y > 256 for the HEC).

2.3.3. Double Edge Classifier

An additional problem for crowded object areas such as SBUs are objects located at the BC borders horizontal-wise and vertical-wise, presenting a double edge error (DEC). Thus, we enhanced the mosaicking by applying a new sliding window, starting at x = 256 and y = 256 with 512-pixel strides (Figure 5E).

2.3.4. Non-Maximum Suppression Sorted by Area

Furthermore, each object located at the images' borders may present more than one classification for the same object, partial classifications from consecutive BC frames (incorrect classifications), and a unique, complete classification (correct classification) from the HEC, VEC, or DEC (Figure 5). The elimination of excessive boxes used the non-maximum suppression ordered by area, guaranteeing only the classification of the most significant element (complete object). Figure 5 shows an example of an element located at double edges, where the DEC classification is the largest and the correct one.

2.4. Performance Metrics

The model evaluation considered the following COCO metrics [66]: average precision (AP), AP50, and AP75. The AP is a ranking metric that calculates the area under the precision-recall curve. However, in object detection, it is crucial to determine a minimum overlap between the predicted bounding box and the ground truth bounding box to evaluate a correct classification. Thus, another element is the intersection over union (IoU) (Figure 6).

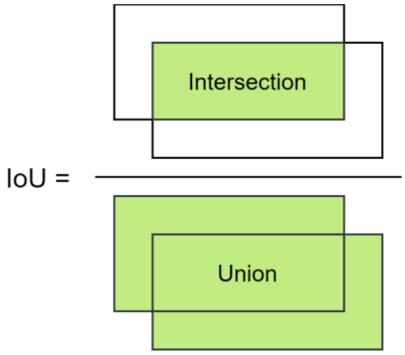

Figure 6. Representation of the Intersection over Union metric.

In this regard, the COCO AP considers the average among ten intersection over union (IoU) thresholds (from 0.5 to 0.95 with 0.05 steps), while AP50 and AP75 scores consider a fixed threshold of 0.5 and 0.75.

3. Results

3.1. Performance Metrics

Table 2 lists the detection (Box) and segmentation (Mask) results with different image scaling ratios and the X-101-FPN backbone. Results on the original image presented similar results compared to the COCO dataset scores. Moreover, scaling presented significant improvement, in which 2× scaling increased nearly 20% in the AP score, and 8× scaling increased nearly 30% AP.

Table 2. COCO metrics (AP, AP50, and AP75) for segmentation (mask) and detection (box) on the different ratio images.

Ratio (Size)	Type	AP	AP50	AP75
8× (512 × 512)	Box	58.12	94.56	66.06
	Mask	56.76	93.73	63.86
4× (256 × 256)	Box	53.45	93.01	60.76
	Mask	52.89	92.21	58.87
2× (128 × 128)	Box	48.24	89.66	46.54
	Mask	49.09	90.24	49.84
1× (64 × 64)	Box	30.49	74.68	15.68
	Mask	36.69	77.42	27.50

Small objects negatively affect the strictest metrics (highest IoU, e.g., AP75). Slight errors in the bounding box position on small objects (with fewer pixels) significantly reduce the IoU (implying low AP scores). In turn, the mistakes are much less impactful when increasing the image dimensions. However, a limitation to the indefinite increase in the image dimensions is the high computational cost.

3.2. Scene Classification

We used the X-101-FPN model with the best scaling ratio (8×) scores, applying it in a 3072 × 2048 pixel image (also using 8× scaling) to validate the mosaicking technique. Figure 7A demonstrates a satisfactory classification even in crowded areas. This process excluded 66 partial classifications in total (Figure 7B), and the trained model has proven to distinguish SBUs from other elements such as tourist beach umbrellas.

Figure 8 shows three zoomed areas (1, 2, and 3) where the top images present the complete (correct) classification results, whereas the bottom images show the partial (incorrect) classifications deleted by the non-max suppression sorted by area algorithm. Figure 8A–C shows the DEC, VEC, and HEC, respectively. Another interesting point is

that example 3.2 shows that one of the partial predictions has greater confidence than the correct prediction (97% against 96%), demonstrating that the non-maximum suppression ordered by area brings improved results.

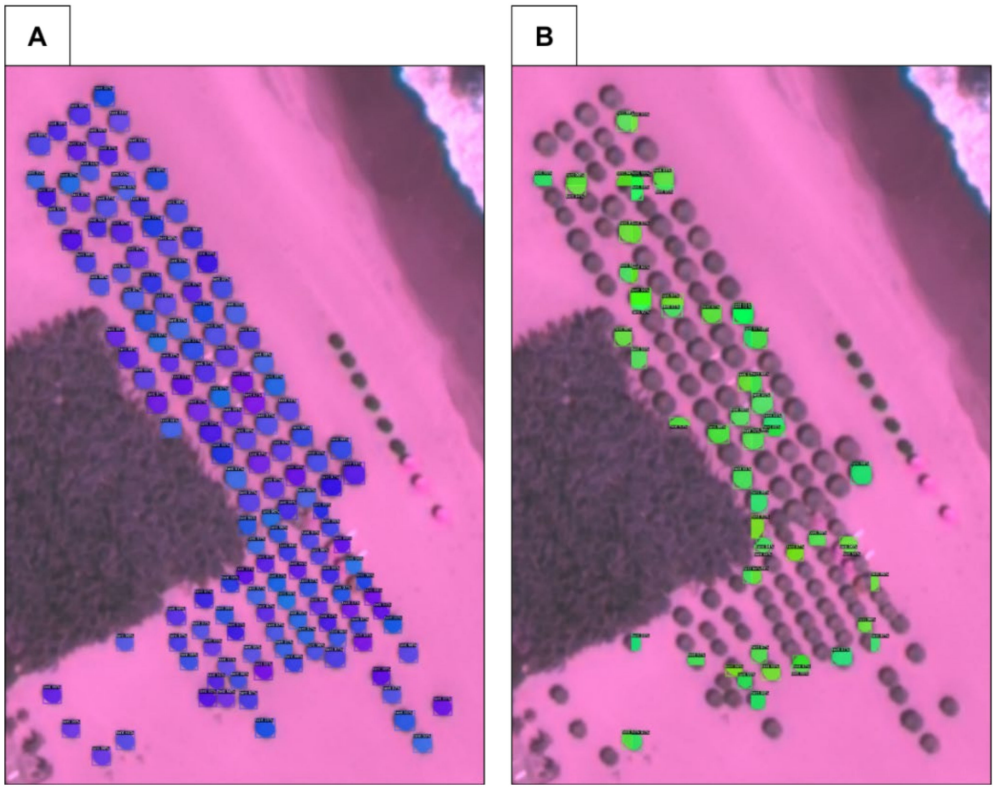

Figure 7. Classifications considering the correct classifications (**A**) and the deleted partial classifications from each object (**B**).

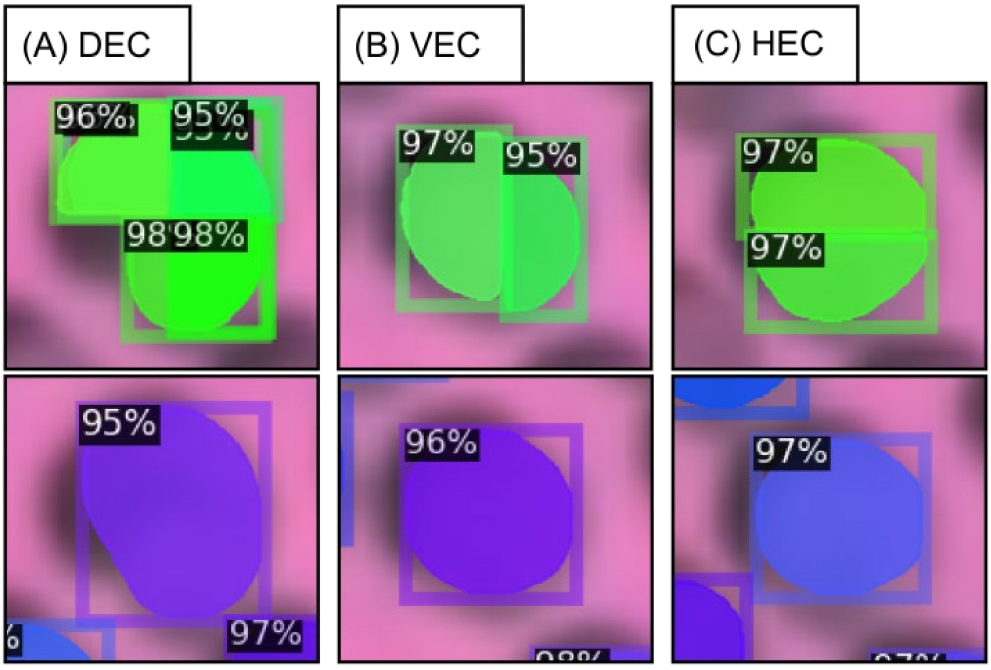

Figure 8. Representation of the three distinct classification scenarios, considering the (**A**) double edge classification (DEC), (**B**) vertical edge classification (VEC), and (**C**) horizontal edge classification (HEC).

Table 3 lists quantitative values that may be very helpful in decision making. This methodology enables automatic counting and detection within large areas using Mask-RCNN. The sizes of the SBUs are very similar, with the average and median sizes very close and a standard deviation of 0.2 m^2. In addition, the algorithm was able to differentiate very close objects, showing a good usage of instance segmentation models for crowded regions.

Table 3. Analysis of the detected objects regarding their counting, average size, median size, minimum size, maximum size, and standard deviation, considering the 8× scaled image.

Description	Result
Count	148 SBUs
Average SBU size	4172 pixels (5.8 m^2)
Median SBU size	4027 pixels (5.6 m^2)
SBU Standard Deviation	161.60 pixels (0.2 m^2)
Minimum SBU Size	2693 pixels (3.8 m^2)
Maximum SBU Size	7278 pixels (10.2 m^2)
Average SBU size	4172 pixels (5.8 m^2)
Median SBU size	4027 pixels (5.6 m^2)

4. Discussion

Instance segmentation is a state-of-the-art computer vision segmentation method that enables many practical approaches for identifying objects at the pixel level. Most instance segmentation studies use large datasets (e.g., COCO [66], Cityscapes [87], or Mapillary Vistas [88]) in a ready-to-use format. Developing datasets for instance segmentation is highly complex and labor intensive, requiring annotation experts and a suitable storage format for DL models. Difficulties worsen for orbital remote sensing images by the need to choose the places of each image tile and the existence of very little annotation software available that considers geospatial data's particularities. With that said, in in a Web of Science search up to November 11, considering the keywords "instance segmentation", "remote sensing", and "deep learning", we found only 22 peer-reviewed journal articles. Despite the gains in efficiency and quality of results, the limited number of papers using instance segmentation demonstrates the difficulties reported. The present research demonstrates that instance segmentation allows a significant gain in inspection efficiency in coastal areas that have not yet been explored. Within these 22 articles, Soloy et al. [89] also explored the beach areas, but with a different approach, as the authors used photos taken by the iPhone to quantify grain size on pebble beaches.

4.1. Multichannel Instance Segmentation Studies

Few studies addressed instance segmentation using multi-channel imagery. Most studies use RGB images [90–92] or even three-channel images from the combination of digital orthophoto map and near-infrared band from the Landsat-8 [93]. The usage of multi-channels in remote sensing is widespread, allowing for more efficient detection than traditional RGB images (e.g., camera photos). Basically, there are four scenarios in remote sensing for using multi-channel inputs: (1) sensors with many spectral bands, (2) time series, (3) change detection, and (4) a combination between these characteristics (e.g., a time series of multispectral images). Using multispectral imagery, de Carvalho et al. [76] made a study on center pivot irrigation systems using Landsat-8 images. The authors compared the usage of seven channels with the traditional RGB, showing a difference of 3% in the AP metric when using more channels. Hao et al. [94] used a multiband input for the Mask-RCNN for identifying tree-crowns and estimating their height. Concerning time series applications, de Albuquerque et al. [95] used Sentinel-1 time series (up to eleven channels) for mapping center pivots. The authors reported an increased performance when including more time frames. In a different approach, de Albuquerque et al. [27] used Sentinel-2 time series (up to 24 channels), considering four spectral bands per temporal frame for effectively mapping regions with a cloud presence.

4.2. Methods for Large Area Classification

A significant problem is that a DL adaptation for remote sensing applications uses large-size images. In this regard, the present research used mosaicking with sliding windows for object detection/instance segmentation. This procedure is more common in semantic segmentation approaches using overlapping pixels [84–86,96]. The method uses a sliding window with a step size smaller than the frame dimensions, causing overlapping. Averaging the overlapping areas mitigates errors, providing better accuracy metrics and visual results. However, for instance segmentation models, the procedure must consider the bounding boxes. In this sense, we modified the method proposed by de Carvalho et al. [76], introducing the double edge classifier (DEC) that is more efficient in extremely crowded areas, such as the SBUs. The methodology effectively eliminates frame discontinuity problems by considering the prediction under the best circumstance, providing a viable solution for mapping large areas.

The capability to apply an instance segmentation algorithm over a large area enables a thorough scene understanding, which is vital for public inspection. For example, our study allows automatic counting of all SBUs and a series of other statistics, such as average size, median size, and standard deviation of the sizes, among others. These quantitative results increase the amount of information for public managers to act. In addition, it is possible to extract the exact location of each element just by getting the coordinates of each bounding box.

4.3. Small Object Problem

Small objects often underperform in many datasets. For example, in the COCO dataset, the AP_{small} metric is much lower than the AP_{medium} and AP_{large} metrics. This effect is related to increasing noise with decreasing object size. In the review of Tong et al. [67], image scaling is a straightforward approach to improve small object detection. Nevertheless, no study compares the effect of different scaling and improved object detection. In this regard, this research compares three scaling ratios for mapping SBUs, which are very small objects. This comparison can guide other studies further studies of small object detection in other scenarios. Our results show that image scaling (even as an image augmentation built-in method) may be a plausible and effective solution. The AP metrics increased more than 20%, considering eight times the original size. Even so, doubling the dimensions already provided a significant increase. This analysis is relevant since increasing the image dimensions might present computational problems (e.g., memory, and processing time).

Some other alternatives have been studied for detecting small objects. Zhang et al. [97] proposed a scale adaptive proposal network by modifying the Faster-RCNN architecture. This innovative approach has broad applications where there are datasets of many different sizes. Nonetheless, considering different scales might not be enough for very small objects, especially for AP scores, where few mistakes in the bounding box drastically reduce this accuracy metric. Generative adversarial networks (GAN) algorithms also present advances in studies with small objects [65]. In remote sensing, Ren et al. [98] proposed an advanced end-to-end GAN to increase image resolution and apply the Faster-RCNN network in object detection. Therefore, a viable alternative for future studies would be the development of algorithms using GAN for surveillance in coastal areas. In the traditional RGB images from the COCO data set, Kisantal et al. [99] made an augmentation system based on copying and pasting small objects into different images to increase the representativeness of a small object in a larger number of images. This augmentation is a promising strategy for datasets with different scaled images. However, it can be computationally expensive in multichannel imaging and in detecting many small objects.

4.4. Accuracy Metric Analysis for Small Objects

Even though there is broad applicability of the COCO metrics for instance segmentation datasets (including the COCO dataset), the AP50 is the most appropriate metric for analyzing small objects (especially in datasets in which all objects are small) since very

few mistakes drop the performance metrics significantly. Figure 9 shows two theoretical examples A and B, in which the prediction and the ground truth bounding boxes have the exact spatial dimensions.

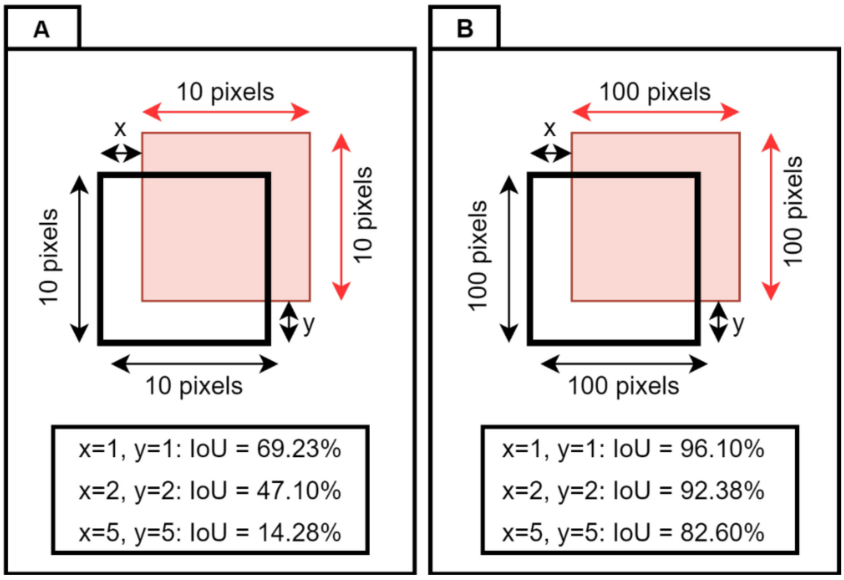

Figure 9. Theoretical examples of predictions on (**A**) small and (**B**) large objects, and their corresponding IoU in different overlapping scenarios.

When considering small objects, a slight mistake of one pixel horizontally and vertically has an IoU of 69.25%, impacting the AP and AP75 metric. A one-pixel error in a 100×100-pixel bounding box generates a 96.10% IoU, showing the attenuation of slight errors in larger objects. This research shows that the simple increase in object dimensions allows the algorithm to have a better accuracy score. Therefore, generating ground truth data, especially for small objects, must be done rigorously to avoid misleading metrics.

4.5. Policy Implications

The Brazilian Government is responsible for the administration and inspection of federal properties. According to Normative Instruction No. 23, of 18 March 2020, the inspection action may have a preventive or coercive nature, requiring a field inspector to investigate possible irregularities committed against federal properties. The inspection action is predominantly coercive through denunciation, when the improper action is consolidated, leaving only the repair of the damage. The lack of preventive action causes an increase in unlawful acts and the filing of numerous lawsuits, with deprivation of use of areas and legal uncertainty.

In Brazil, beach areas are public properties protected by environmental legislation (CONAMA resolution No. 303 of 20 March 2002) as permanent preservation areas and consist of Navy land, where private occupation (private, commercial, or industrial) requires payment of a fee for the use of the public area. Beach areas are constant targets of economic exploitation and improper tourism and need constant surveillance. In this context, the development of remote and semi-automated methods of surveillance of property misuse becomes fundamental.

Therefore, the instance segmentation of multispectral remote sensing images demonstrates a high potential to establish an effective action with a solid preventive impact due to the rapid infraction detection. However, the procedure should be improved, including other activities without prior authorization in coastal areas such as landfills, deforestation, construction, fences, or other improvements, which could be developed in future lines of research.

5. Conclusions

The automatic remote sensing detection of tourist infrastructure in beach areas is essential for government surveillance, requiring quick and periodic information for decision making. The coastal regions of Brazil are government property, being areas with specific taxation for use and environmental protection. This study proposed a methodology based on instance segmentation to identify straw beach umbrellas (SBUs), the most common tourist structure on Brazilian beaches. The developed method integrates different solutions for the use of instance segmentation in remote sensing data: (1) multi-channel models, (2) small object detection, and (3) classification of large areas. Therefore, we modified Detectron2's Mask-RCNN model to account for multi-channel image inputs in TIFF format, compared different scaling ratios on the original image, and improved the existing method for classifying large images using the sliding window technique. Our results show that increasing image dimensions significantly improve the AP metric from 30% to 58%. In addition, the less strict metric (AP50) showed results from 74% to 94%. Image scaling is a computationally expensive solution, so we initially considered the original image dimensions of 64 × 64 pixels. In addition, even though we evaluated up to eight times the original dimensions (resulting in a 512 × 512 image), a two-times resizing already provides a significant increase. Thus, the research needs to define a trade-off in computational cost and in the quality of predictions.

Another problem is the accumulation of errors on the frame edges, which intensify with overcrowded objects. Our innovative proposal to use double edge classification (DEG) solved the problem simply and efficiently. The architecture of all exposed methods is a suitable solution for accurately detecting small objects in large areas using multispectral data, providing insightful information for public managers. For example, statistical analysis of the SBUs on a 3072 × 2048 test image identified 148 objects with an average size of 5.8 m^2. The bounding box centroid established the exact geographic location. Future studies in this area will consider more beach elements, exploring objects and background elements, and other segmentation tasks such as panoptic segmentation.

Author Contributions: Conceptualization, Osmar Luiz Ferreira de Carvalho, Osmar Abílio de Carvalho Júnior, and Rebeca dos Santos de Moura; methodology, Osmar Luiz Ferreira de Carvalho, Osmar Abílio de Carvalho Júnior, Pablo Pozzobon de Bem, and Rebeca dos Santos de Moura; software, Osmar Luiz Ferreira de Carvalho, Rebeca dos Santos de Moura, and Pablo Pozzobon de Bem; validation, Osmar Luiz Ferreira de Carvalho, Anesmar Olino de Albuquerque, and Rebeca dos Santos de Moura; formal analysis, Osmar Luiz Ferreira de Carvalho, Rubens de Castro Pereira, Osmar Abílio de Carvalho Júnior, and Renato Fontes Guimarães; investigation, Osmar Luiz Ferreira de Carvalho, Rebeca dos Santos de Moura, and Renato Fontes Guimarães; resources, Osmar Abílio de Carvalho Júnior, Roberto Arnaldo Trancoso Gomes, and Renato Fontes Guimarães; data curation, Osmar Luiz Ferreira de Carvalho, Anesmar Olino de Albuquerque, and Pablo Pozzobon de Bem; writing—original draft preparation, Osmar Luiz Ferreira de Carvalho and Osmar Abílio de Carvalho Júnior; writing—review and editing, Osmar Luiz Ferreira de Carvalho, Osmar Abílio de Carvalho Júnior, Rebeca dos Santos de Moura, Rubens de Castro Pereira, Dibio Leandro Borges, and Li Weigang; visualization, Osmar Luiz Ferreira de Carvalho, Rubens de Castro Pereira, Osmar Abílio de Carvalho Júnior, and Rebeca dos Santos de Moura; supervision, Osmar Abílio de Carvalho Júnior, Roberto Arnaldo Trancoso Gomes, Renato Fontes Guimarães, Dibio Leandro Borges, and Li Weigang; project administration, Osmar Luiz Ferreira de Carvalho, Osmar Abílio de Carvalho Júnior, Roberto Arnaldo Trancoso Gomes, and Renato Fontes Guimarães; funding acquisition, Osmar Abílio de Carvalho Júnior, Roberto Arnaldo Trancoso Gomes, and Renato Fontes Guimarães. All authors have read and agreed to the published version of the manuscript.

Funding: The following institutions funded this research: National Council for Scientific and Technological Development (434838/2018-7), Coordination for the Improvement of Higher Education Personnel, and the Union Heritage Secretariat of the Ministry of Economy.

Data Availability Statement: Data available on request from the authors.

Acknowledgments: The authors thank the following institutions: National Council for Scientific and Technological Development (CNPq) for the fellowship granted to professors Osmar Abílio de

Carvalho Júnior, Roberto Arnaldo Trancoso Gomes, and Renato Fontes Guimarães; Coordination for the Improvement of Higher Education Personnel (CAPES) for postgraduate assistance; Union Heritage Secretariat of the Ministry of Economy for financial support; and the European Space Agency (ESA) for image supply within the project "Surveillance of union properties areas using deep learning technique in satellite images". Special thanks are given to the research group of the Laboratory of Spatial Information System of the University of Brasilia for technical support.

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

References

1. Brown, G.; Weber, D.; de Bie, K. Assessing the value of public lands using public participation GIS (PPGIS) and social landscape metrics. *Appl. Geogr.* **2014**, *53*, 77–89. [CrossRef]
2. DeFries, R.; Hansen, A.; Turner, B.L.; Reid, R.; Liu, J. Land use change around protected areas: Management to balance human needs and ecological function. *Ecol. Appl.* **2007**, *17*, 1031–1038. [CrossRef] [PubMed]
3. Belal, A.A.; Moghanm, F.S. Detecting urban growth using remote sensing and GIS techniques in Al Gharbiya governorate, Egypt. *Egypt. J. Remote Sens. Space Sci.* **2011**, *14*, 73–79. [CrossRef]
4. Dacey, S.; Song, L.; Pang, S. An intelligent agent based land encroachment detection approach. In Proceedings of the International Conference on Neural Information Processing, Daegu, Korea, 3–7 November 2013; pp. 585–592.
5. Brown, G.; de Bie, K.; Weber, D. Identifying public land stakeholder perspectives for implementing place-based land management. *Landsc. Urban Plan.* **2015**, *139*, 1–15. [CrossRef]
6. Martin, C.; Parkes, S.; Zhang, Q.; Zhang, X.; McCabe, M.F.; Duarte, C.M. Use of unmanned aerial vehicles for efficient beach litter monitoring. *Mar. Pollut. Bull.* **2018**, *131*, 662–673. [CrossRef] [PubMed]
7. Serra-Gonçalves, C.; Lavers, J.L.; Bond, A.L. Global review of beach debris monitoring and future recommendations. *Environ. Sci. Technol.* **2019**, *53*, 12158–12167. [CrossRef]
8. Gladstone, W.; Curley, B.; Shokri, M.R. Environmental impacts of tourism in the Gulf and the Red Sea. *Mar. Pollut. Bull.* **2013**, *72*, 375–388. [CrossRef]
9. Burak, S.; Dog˘an, E.; Gaziog˘lu, C. Impact of urbanization and tourism on coastal environment. *Ocean Coast. Manag.* **2004**, *47*, 515–527. [CrossRef]
10. He, Y.; Ma, W.; Ma, Z.; Fu, W.; Chen, C.; Yang, C.-F.; Liu, Z. Using Unmanned Aerial Vehicle Remote Sensing and a Monitoring Information System to Enhance the Management of Unauthorized Structures. *Appl. Sci.* **2019**, *9*, 4954. [CrossRef]
11. Varol, B.; Yılmaz, E.Ö.; Maktav, D.; Bayburt, S.; Gürdal, S. Detection of illegal constructions in urban cities: Comparing LIDAR data and stereo KOMPSAT-3 images with development plans. *Eur. J. Remote Sens.* **2019**, *52*, 335–344. [CrossRef]
12. Lira, C.; Taborda, R. Advances in Applied Remote Sensing to Coastal Environments Using Free Satellite Imagery. In *Remote Sensing and Modeling*; Finkl, C., Makowski, C., Eds.; Springer: Cham, Switzerland, 2014; pp. 77–102.
13. Parthasarathy, K.S.S.; Deka, P.C. Remote sensing and GIS application in assessment of coastal vulnerability and shoreline changes: A review. *ISH J. Hydraul. Eng.* **2019**, 1–13. [CrossRef]
14. McCarthy, M.J.; Colna, K.E.; El-Mezayen, M.M.; Laureano-Rosario, A.E.; Méndez-Lázaro, P.; Otis, D.B.; Toro-Farmer, G.; Vega-Rodriguez, M.; Muller-Karger, F.E. Satellite Remote Sensing for Coastal Management: A Review of Successful Applications. *Environ. Manag.* **2017**, *60*, 323–339. [CrossRef]
15. El Mahrad, B.; Newton, A.; Icely, J.; Kacimi, I.; Abalansa, S.; Snoussi, M. Contribution of Remote Sensing Technologies to a Holistic Coastal and Marine Environmental Management Framework: A Review. *Remote Sens.* **2020**, *12*, 2313. [CrossRef]
16. Ouellette, W.; Getinet, W. Remote sensing for Marine Spatial Planning and Integrated Coastal Areas Management: Achievements, challenges, opportunities and future prospects. *Remote Sens. Appl. Soc. Environ.* **2016**, *4*, 138–157. [CrossRef]
17. Ibarra-Marinas, D.; Belmonte-Serrato, F.; Ballesteros-Pelegrín, G.; García-Marín, R. Evolution of the Beaches in the Regional Park of Salinas and Arenales of San Pedro del Pinatar (Southeast of Spain) (1899–2019). *ISPRS Int. J. Geo-Inf.* **2021**, *10*, 200. [CrossRef]
18. Rifat, S.; Liu, W. Measuring Community Disaster Resilience in the Conterminous Coastal United States. *ISPRS Int. J. Geo-Inf.* **2020**, *9*, 469. [CrossRef]
19. Sahana, M.; Hong, H.; Ahmed, R.; Patel, P.P.; Bhakat, P.; Sajjad, H. Assessing coastal island vulnerability in the Sundarban Biosphere Reserve, India, using geospatial technology. *Environ. Earth Sci.* **2019**, *78*, 304. [CrossRef]
20. Poompavai, V.; Ramalingam, M. Geospatial Analysis for Coastal Risk Assessment to Cyclones. *J. Indian Soc. Remote Sens.* **2013**, *41*, 157–176. [CrossRef]
21. Momeni, R.; Aplin, P.; Boyd, D. Mapping Complex Urban Land Cover from Spaceborne Imagery: The Influence of Spatial Resolution, Spectral Band Set and Classification Approach. *Remote Sens.* **2016**, *8*, 88. [CrossRef]
22. Llausàs, A.; Hof, A.; Wolf, N.; Saurí, D.; Siegmund, A. Applicability of cadastral data to support the estimation of water use in private swimming pools. *Environ. Plan. B Urban Anal. City Sci.* **2019**, *46*, 1165–1181. [CrossRef]
23. Papakonstantinou, A.; Doukari, M.; Stamatis, P.; Topouzelis, K. Coastal Management Using UAS and High-Resolution Satellite Images for Touristic Areas. *Int. J. Appl. Geospat. Res.* **2019**, *10*, 54–72. [CrossRef]

24. 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]
25. Huang, H.; Lan, Y.; Yang, A.; Zhang, Y.; Wen, S.; Deng, J. Deep learning versus Object-based Image Analysis (OBIA) in weed mapping of UAV imagery. *Int. J. Remote Sens.* **2020**, *41*, 3446–3479. [CrossRef]
26. Liu, T.; Abd-Elrahman, A.; Morton, J.; Wilhelm, V.L. Comparing fully convolutional networks, random forest, support vector machine, and patch-based deep convolutional neural networks for object-based wetland mapping using images from small unmanned aircraft system. *GIScience Remote Sens.* **2018**, *55*, 243–264. [CrossRef]
27. De Albuquerque, A.O.; Ferreira de Carvalho, O.L.; e Silva, C.; Saiaka Luiz, A.; De Bem, P.P.; Gomes, R.A.T.; Guimaraes, R.F.; de Carvalho Júnior, O.A.A. Dealing with Clouds and Seasonal Changes for Center Pivot Irrigation Systems Detection Using Instance Segmentation in Sentinel-2 Time Series. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 8447–8457. [CrossRef]
28. Voulodimos, A.; Doulamis, N.; Doulamis, A.; Protopapadakis, E. Deep Learning for Computer Vision: A Brief Review. *Comput. Intell. Neurosci.* **2018**, *2018*, 1–13. [CrossRef] [PubMed]
29. Sun, S.; Luo, C.; Chen, J. A review of natural language processing techniques for opinion mining systems. *Inf. Fusion* **2017**, *36*, 10–25. [CrossRef]
30. Young, T.; Hazarika, D.; Poria, S.; Cambria, E. Recent Trends in Deep Learning Based Natural Language Processing [Review Article]. *IEEE Comput. Intell. Mag.* **2018**, *13*, 55–75. [CrossRef]
31. Nassif, A.B.; Shahin, I.; Attili, I.; Azzeh, M.; Shaalan, K. Speech Recognition Using Deep Neural Networks: A Systematic Review. *IEEE Access* **2019**, *7*, 19143–19165. [CrossRef]
32. Zhang, Z.; Geiger, J.; Pohjalainen, J.; Mousa, A.E.-D.; Jin, W.; Schuller, B. Deep Learning for Environmentally Robust Speech Recognition. *ACM Trans. Intell. Syst. Technol.* **2018**, *9*, 1–28. [CrossRef]
33. Zhao, Z.-Q.Q.; Zheng, P.; Xu, S.-T.T.; Wu, X. Object Detection With Deep Learning: A Review. *IEEE Trans. Neural Netw. Learn. Syst.* **2019**, *30*, 3212–3232. [CrossRef]
34. Sharma, V.; Mir, R.N. A comprehensive and systematic look up into deep learning based object detection techniques: A review. *Comput. Sci. Rev.* **2020**, *38*, 100301. [CrossRef]
35. Zhou, T.; Ruan, S.; Canu, S. A review: Deep learning for medical image segmentation using multi-modality fusion. *Array* **2019**, *3–4*, 100004. [CrossRef]
36. Liu, S.; Wang, Y.; Yang, X.; Lei, B.; Liu, L.; Li, S.X.; Ni, D.; Wang, T. Deep Learning in Medical Ultrasound Analysis: A Review. *Engineering* **2019**, *5*, 261–275. [CrossRef]
37. Serte, S.; Serener, A.; Al-Turjman, F. Deep learning in medical imaging: A brief review. *Trans. Emerg. Telecommun. Technol.* **2020**. [CrossRef]
38. Wu, D.; Zheng, S.-J.; Zhang, X.-P.; Yuan, C.-A.; Cheng, F.; Zhao, Y.; Lin, Y.-J.; Zhao, Z.-Q.; Jiang, Y.-L.; Huang, D.-S. Deep learning-based methods for person re-identification: A comprehensive review. *Neurocomputing* **2019**, *337*, 354–371. [CrossRef]
39. Bharathi, B.; Shamily, P.B. A review on iris recognition system for person identification. *Int. J. Comput. Biol. Drug Des.* **2020**, *13*, 316. [CrossRef]
40. Kaur, P.; Krishan, K.; Sharma, S.K.; Kanchan, T. Facial-recognition algorithms: A literature review. *Med. Sci. Law* **2020**, *60*, 131–139. [CrossRef] [PubMed]
41. Dana, D.; Gadhiya, S.; St Surin, L.; Li, D.; Naaz, F.; Ali, Q.; Paka, L.; Yamin, M.; Narayan, M.; Goldberg, I.; et al. Deep Learning in Drug Discovery and Medicine; Scratching the Surface. *Molecules* **2018**, *23*, 2384. [CrossRef]
42. Lavecchia, A. Deep learning in drug discovery: Opportunities, challenges and future prospects. *Drug Discov. Today* **2019**, *24*, 2017–2032. [CrossRef]
43. Chen, H.; Engkvist, O.; Wang, Y.; Olivecrona, M.; Blaschke, T. The rise of deep learning in drug discovery. *Drug Discov. Today* **2018**, *23*, 1241–1250. [CrossRef]
44. Koumakis, L. Deep learning models in genomics; are we there yet? *Comput. Struct. Biotechnol. J.* **2020**, *18*, 1466–1473. [CrossRef]
45. Talukder, A.; Barham, C.; Li, X.; Hu, H. Interpretation of deep learning in genomics and epigenomics. *Brief. Bioinform.* **2021**, *22*, bbaa177. [CrossRef] [PubMed]
46. Ma, L.; Liu, Y.; Zhang, X.; Ye, Y.; Yin, G.; Johnson, B.A. Deep learning in remote sensing applications: A meta-analysis and review. *ISPRS J. Photogramm. Remote Sens.* **2019**, *152*, 166–177. [CrossRef]
47. Liu, L.; Ouyang, W.; Wang, X.; Fieguth, P.; Chen, J.; Liu, X.; Pietikäinen, M. Deep Learning for Generic Object Detection: A Survey. *Int. J. Comput. Vis.* **2020**, *128*, 261–318. [CrossRef]
48. Li, K.; Wan, G.; Cheng, G.; Meng, L.; Han, J. Object detection in optical remote sensing images: A survey and a new benchmark. *ISPRS J. Photogramm. Remote Sens.* **2020**, *159*, 296–307. [CrossRef]
49. Cheng, G.; Xie, X.; Han, J.; Guo, L.; Xia, G.-S. Remote Sensing Image Scene Classification Meets Deep Learning: Challenges, Methods, Benchmarks, and Opportunities. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2020**, *13*, 3735–3756. [CrossRef]
50. Ball, J.E.; Anderson, D.T.; Chan, C.S. Comprehensive survey of deep learning in remote sensing: Theories, tools, and challenges for the community. *J. Appl. Remote Sens.* **2017**, *11*, 042609. [CrossRef]
51. Li, J.; Huang, X.; Gong, J. Deep neural network for remote-sensing image interpretation: Status and perspectives. *Natl. Sci. Rev.* **2019**, *6*, 1082–1086. [CrossRef] [PubMed]
52. Hoeser, T.; Bachofer, F.; Kuenzer, C. Object detection and image segmentation with deep learning on earth observation data: A review-part II: Applications. *Remote Sens.* **2020**, *12*, 3053. [CrossRef]

53. Khelifi, L.; Mignotte, M. Deep Learning for Change Detection in Remote Sensing Images: Comprehensive Review and Meta-Analysis. *IEEE Access* **2020**, *8*, 126385–126400. [CrossRef]
54. Li, Y.; Zhang, H.; Xue, X.; Jiang, Y.; Shen, Q. Deep learning for remote sensing image classification: A survey. *WIREs Data Min. Knowl. Discov.* **2018**, *8*, e1264. [CrossRef]
55. Zhang, L.; Zhang, L.; Du, B. Deep Learning for Remote Sensing Data: A Technical Tutorial on the State of the Art. *IEEE Geosci. Remote Sens. Mag.* **2016**, *4*, 22–40. [CrossRef]
56. Paoletti, M.E.; Haut, J.M.; Plaza, J.; Plaza, A. Deep learning classifiers for hyperspectral imaging: A review. *ISPRS J. Photogramm. Remote Sens.* **2019**, *158*, 279–317. [CrossRef]
57. Parikh, H.; Patel, S.; Patel, V. Classification of SAR and PolSAR images using deep learning: A review. *Int. J. Image Data Fusion* **2020**, *11*, 1–32. [CrossRef]
58. Signoroni, A.; Savardi, M.; Baronio, A.; Benini, S. Deep Learning Meets Hyperspectral Image Analysis: A Multidisciplinary Review. *J. Imaging* **2019**, *5*, 52. [CrossRef] [PubMed]
59. Zhu, X.X.; Tuia, D.; Mou, L.; Xia, G.; Zhang, L.; Xu, F.; Fraundorfer, F. Deep Learning in Remote Sensing: A Comprehensive Review and List of Resources. *IEEE Geosci. Remote Sens. Mag.* **2017**, *5*, 8–36. [CrossRef]
60. Vali, A.; Comai, S.; Matteucci, M. Deep Learning for Land Use and Land Cover Classification Based on Hyperspectral and Multispectral Earth Observation Data: A Review. *Remote Sens.* **2020**, *12*, 2495. [CrossRef]
61. Yuan, Q.; Shen, H.; Li, T.; Li, Z.; Li, S.; Jiang, Y.; Xu, H.; Tan, W.; Yang, Q.; Wang, J.; et al. Deep learning in environmental remote sensing: Achievements and challenges. *Remote Sens. Environ.* **2020**, *241*, 111716. [CrossRef]
62. Schmidhuber, J. Deep learning in neural networks: An overview. *Neural Netw.* **2015**, *61*, 85–117. [CrossRef]
63. Nogueira, K.; Penatti, O.A.B.; dos Santos, J.A. Towards better exploiting convolutional neural networks for remote sensing scene classification. *Pattern Recognit.* **2017**, *61*, 539–556. [CrossRef]
64. Lecun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [CrossRef]
65. Li, J.; Liang, X.; Wei, Y.; Xu, T.; Feng, J.; Yan, S. Perceptual generative adversarial networks for small object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1222–1230.
66. Lin, T.-Y.; Maire, M.; Belongie, S.; Hays, J.; Perona, P.; Ramanan, D.; Dollár, P.; Zitnick, C.L. Microsoft COCO: Common Objects in Context. In *Computer Vision—ECCV 2014. Lecture Notes in Computer Science, vol 8693*; Fleet, D., Tomas, P., Schiele, B., Tuytelaars, T., Eds.; Springer: Cham, Switzerland, 2014; pp. 740–755; ISBN 978-3-319-10601-4.
67. Tong, K.; Wu, Y.; Zhou, F. Recent advances in small object detection based on deep learning: A review. *Image Vis. Comput.* **2020**, *97*, 103910. [CrossRef]
68. Laben, C.A.; Brower, B.V. Process for Enhancing the Spatial Resolution of Multispectral Imagery Using Pan-Sharpening. USA Patent 6,011,875, 4 January 2000.
69. Johansen, K.; Duan, Q.; Tu, Y.-H.; Searle, C.; Wu, D.; Phinn, S.; Robson, A.; McCabe, M.F. Mapping the condition of macadamia tree crops using multi-spectral UAV and WorldView-3 imagery. *ISPRS J. Photogramm. Remote Sens.* **2020**, *165*, 28–40. [CrossRef]
70. Ghassemian, H. A review of remote sensing image fusion methods. *Inf. Fusion* **2016**, *32*, 75–89. [CrossRef]
71. Wu, Y.; Kirillov, A.; Massa, F.; Lo, W.-Y.; Girshick, R. Detectron2. Available online: https://github.com/facebookresearch/detectron2 (accessed on 3 March 2021).
72. 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; IEEE: Venice, Italy, 2017; pp. 2980–2988.
73. Torralba, A.; Russell, B.C.; Yuen, J. LabelMe: Online Image Annotation and Applications. *Proc. IEEE* **2010**, *98*, 1467–1484. [CrossRef]
74. Russell, B.C.; Torralba, A.; Murphy, K.P.; Freeman, W.T. LabelMe: A Database and Web-Based Tool for Image Annotation. *Int. J. Comput. Vis.* **2008**, *77*, 157–173. [CrossRef]
75. Sekachev, B.; Nikita, M.; Andrey, Z. Computer Vision Annotation Tool: A Universal Approach to Data Annotation. Available online: https://software.intel.com/en-us/articles/computer-vision-annotation-tool-a-universal-approach-to-data-annotation (accessed on 30 October 2021).
76. De Carvalho, O.L.F.; de Carvalho Júnior, O.A.A.; de Albuquerque, A.O.; de Bem, P.P.; Silva, C.R.; Ferreira, P.H.G.; Moura, R.D.S.D.; Gomes, R.A.T.; Guimarães, R.F.; Borges, D.L.D.L. Instance Segmentation for Large, Multi-Channel Remote Sensing Imagery Using Mask-RCNN and a Mosaicking Approach. *Remote Sens.* **2021**, *13*, 39. [CrossRef]
77. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. In Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; IEEE: Columbus, OH, USA, 2014; Volume 1, pp. 580–587.
78. Girshick, R. Fast R-CNN. In Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 7–13 December 2015; IEEE: Santiago, Chile, 2015; pp. 1440–1448.
79. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 1137–1149. [CrossRef]
80. Shelhamer, E.; Long, J.; Darrell, T. Fully Convolutional Networks for Semantic Segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 640–651. [CrossRef]

81. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June2016; IEEE: Las Vegas, NV, USA, 2016; Volume 45, pp. 770–778.
82. Xie, S.; Girshick, R.; Dollar, P.; Tu, Z.; He, K. Aggregated Residual Transformations for Deep Neural Networks. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; IEEE: Honolulu, HI, USA, 2017; pp. 5987–5995.
83. Audebert, N.; Boulch, A.; Randrianarivo, H.; Le, B.; Ferecatu, M.; Lefèvre, S.; Marlet, R.; Audebert, N.; Boulch, A.; Randrianarivo, H.; et al. Deep learning for urban remote sensing. In Proceedings of the 2017 Joint Urban Remote Sensing Event (JURSE), Dubai, United Arab Emirates, 6–8 March 2017.
84. Da Costa, L.B.; de Carvalho, O.L.F.; de Albuquerque, A.O.; Gomes, R.A.T.; Guimarães, R.F.; de Carvalho Júnior, O.A. Deep semantic segmentation for detecting eucalyptus planted forests in the Brazilian territory using sentinel-2 imagery. *Geocarto Int.* **2021**, 1–13. [CrossRef]
85. Da Costa, M.V.C.V.; de Carvalho, O.L.F.; Orlandi, A.G.; Hirata, I.; De Albuquerque, A.O.; e Silva, F.V.; Guimarães, R.F.; Gomes, R.A.T.; de Carvalho Júnior, O.A. Remote Sensing for Monitoring Photovoltaic Solar Plants in Brazil Using Deep Semantic Segmentation. *Energies* **2021**, *14*, 2960. [CrossRef]
86. De Albuquerque, A.O.; de Carvalho Júnior, O.A.; de Carvalho, O.L.F.; de Bem, P.P.; Ferreira, P.H.G.; de Moura, R.D.S.; Silva, C.R.; Trancoso Gomes, R.A.; Fontes Guimarães, R. Deep Semantic Segmentation of Center Pivot Irrigation Systems from Remotely Sensed Data. *Remote Sens.* **2020**, *12*, 2159. [CrossRef]
87. Cordts, M.; Omran, M.; Ramos, S.; Rehfeld, T.; Enzweiler, M.; Benenson, R.; Franke, U.; Roth, S.; Schiele, B. The Cityscapes Dataset for Semantic Urban Scene Understanding. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; IEEE: Las Vegas, NV, USA, 2016; Volume 29, pp. 3213–3223.
88. Neuhold, G.; Ollmann, T.; Bulo, S.R.; Kontschieder, P. The Mapillary Vistas Dataset for Semantic Understanding of Street Scenes. In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; IEEE: Venice, Italy, 2017; pp. 5000–5009.
89. Soloy, A.; Turki, I.; Fournier, M.; Costa, S.; Peuziat, B.; Lecoq, N. A deep learning-based method for quantifying and mapping the grain size on pebble beaches. *Remote Sens.* **2020**, *12*, 3659. [CrossRef]
90. Zhao, W.; Persello, C.; Stein, A. Building outline delineation: From aerial images to polygons with an improved end-to-end learning framework. *ISPRS J. Photogramm. Remote Sens.* **2021**, *175*, 119–131. [CrossRef]
91. Li, Y.; Xu, W.; Chen, H.; Jiang, J.; Li, X. A Novel Framework Based on Mask R-CNN and Histogram Thresholding for Scalable Segmentation of New and Old Rural Buildings. *Remote Sens.* **2021**, *13*, 1070. [CrossRef]
92. Wu, Q.; Feng, D.; Cao, C.; Zeng, X.; Feng, Z.; Wu, J.; Huang, Z. Improved Mask R-CNN for Aircraft Detection in Remote Sensing Images. *Sensors* **2021**, *21*, 2618. [CrossRef]
93. Lv, Y.; Zhang, C.; Yun, W.; Gao, L.; Wang, H.; Ma, J.; Li, H.; Zhu, D. The Delineation and Grading of Actual Crop Production Units in Modern Smallholder Areas Using RS Data and Mask R-CNN. *Remote Sens.* **2020**, *12*, 1074. [CrossRef]
94. Hao, Z.; Lin, L.; Post, C.J.; Mikhailova, E.A.; Li, M.; Chen, Y.; Yu, K.; Liu, J. Automated tree-crown and height detection in a young forest plantation using mask region-based convolutional neural network (Mask R-CNN). *ISPRS J. Photogramm. Remote Sens.* **2021**, *178*, 112–123. [CrossRef]
95. De Albuquerque, A.O.; de Carvalho, O.L.F.; e Silva, C.R.; de Bem, P.P.; Gomes, R.A.T.; Borges, D.L.; Guimarães, R.F.; Pimentel, C.M.M.; de Carvalho Júnior, O.A. Instance segmentation of center pivot irrigation systems using multi-temporal SENTINEL-1 SAR images. *Remote Sens. Appl. Soc. Environ.* **2021**, *23*, 100537. [CrossRef]
96. Audebert, N.; Le Saux, B.; Lefèvre, S. Segment-before-Detect: Vehicle Detection and Classification through Semantic Segmentation of Aerial Images. *Remote Sens.* **2017**, *9*, 368. [CrossRef]
97. Zhang, S.; He, G.; Chen, H.-B.; Jing, N.; Wang, Q. Scale Adaptive Proposal Network for Object Detection in Remote Sensing Images. *IEEE Geosci. Remote Sens. Lett.* **2019**, *16*, 864–868. [CrossRef]
98. Ren, Y.; Zhu, C.; Xiao, S. Small Object Detection in Optical Remote Sensing Images via Modified Faster R-CNN. *Appl. Sci.* **2018**, *8*, 813. [CrossRef]
99. Kisantal, M.; Wojna, Z.; Murawski, J.; Naruniec, J.; Cho, K. Augmentation for small object detection. *arXiv Prepr.* **2019**, arXiv:1902.07296.