





Article

Optimizing the Recognition and Feature Extraction of Wind Turbines through Hybrid Semantic Segmentation Architectures

Miguel-Ángel Manso-Callejo * , Calimanut-Ionut Cira , Ramón Alcarria  and José-Juan Arranz-Justel 

Departamento de Ingeniería Topográfica y Cartografía, E.T.S.I. en Topografía, Geodesia y Cartografía, Universidad Politécnica de Madrid, 28031 Madrid, Spain; ionut.cira@upm.es (C.-I.C.); ramon.alcarria@upm.es (R.A.); josejuan.arranz@upm.es (J.-J.A.-J.)

* Correspondence: m.manso@upm.es

Received: 22 October 2020; Accepted: 11 November 2020; Published: 13 November 2020



Abstract: Updating the mapping of wind turbines farms—found in constant expansion—is important to predict energy production or to minimize the risk of these infrastructures during storms. This geoinformation is not usually provided by public mapping agencies, and the alternative sources are usually consortiums or individuals interested in mapping and study. However, they do not offer metadata or genealogy, and their quality is unknown. This article presents a methodology oriented to optimize the recognition and extraction of features (wind turbines) using hybrid architectures of semantic segmentation. The aim is to characterize the quality of these datasets and help to improve and update them automatically at a large-scale. To this end, we intend to evaluate the capacity of hybrid semantic segmentation networks trained to extract features representing wind turbines from high-resolution images and to characterize the positional accuracy and completeness of a dataset whose genealogy and quality are unknown. We built a training dataset composed of 5140 tiles of aerial images and their cartography to train six different neural network architectures. The networks were evaluated on five test areas (covering 520 km² of the Spanish territory) to identify the best segmentation architecture (in our case, LinkNet as base architecture and EfficientNet-b3 as the backbone). This hybrid segmentation model allowed us to characterize the completeness—both by commission and by omission—of the available georeferenced wind turbine dataset, as well as its geometric quality.

Keywords: semantic segmentation; quality control; wind turbines; feature extraction

1. Introduction

The representation of geographic elements in vector models is generally produced using various capturing and digitization technologies in which the human factor intervenes, that is, they are not free from errors. The human factor also means that it is not possible to faithfully represent a geographical reality because it is in constant change due to extreme natural actions (earthquakes, tsunamis, fires) or continuous (tectonic movements, climate change, etc.), or due to anthropic activities. This lack of accuracy in cartography and geographic information sources is manifested by omission and commission errors, and by geometric location errors.

Thanks to the advances in parallel processing technologies (Graphics Processing Units or GPUs) and artificial intelligence (AI), methodologies are able to be developed for the efficient large-scale processing of data with artificial neural networks (ANN), and obtain high success rates in predictions in different areas. One such example is Deep Learning (DL) applications in remote sensing, where it is

possible to process aerial images through semantic segmentation techniques and assign land cover labels to every pixel in an image [1].

There are several works in the literature related to the use of DL techniques for monitoring the condition of wind turbines [2], or for their identification in images captured from autonomous vehicles (drones) for supporting maintenance tasks [3]. However, we did not find papers describing experiences related to the use of semantic segmentation for the extraction of wind turbines. These types of features can have different dimensions (blade diameters from 30 m to 140 m, hub heights from 40 m to 100 m), depending on their capacity to produce energy. The complexity of these structures derives from difficulties in identifying them in high-resolution aerial images due to the large areas of shadow generated, the orientations of the shadows (depending on the time the aerial image was captured), and the errors generated in the orthorectification process.

Interest in applying computer vision techniques to remotely sensed imagery has increased following the introduction of convolutional neural networks like AlexNet [4], VGGNet [5], or Inception [6]. Supported by these advancements, modern segmentation techniques emerged. Current semantic segmentation implementations use convolutional layers to downsample (encode) the images and transposed convolutions in the upsampling part (decoder) to resize the image to the original dimensions.

The two most popular semantic segmentation architectures are U-Net [7] and LinkNet [8]. They both follow the encoder–decoder structure (presented in Figure 1), which allows the use of backbone networks at each encoding step and introduce skip connections between two corresponding paths as an intent to provide local context and transfer information. State-of-the-art segmentation models focus on improving the predictions and exploring the information over larger receptive fields, on developing new links between parts of the segmentation model to provide more local context or on introducing lateral connections to allow the detection of objects at multiple scales.

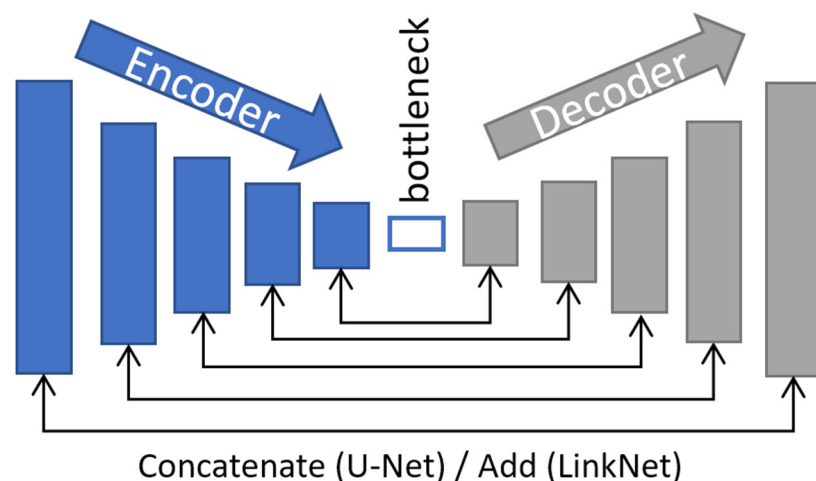


Figure 1. Simplified representation of the Encoder–Decoder Structure present in U-Net and LinkNet. Note: The encoder downsamples the input image using convolutional layers, while the decoder part upsamples the representation from the bottleneck to the original input size using transposed convolutions and enriches it using the skip connections.

The backbone networks also evolved to improve the efficiency of segmentation operation. Moving away from standard CNN architectures like VGGNet [5], new backbone networks specialized in segmentation were introduced. Some of the most popular backbone networks specialized in semantic segmentation are EfficientNet [9] and SEResNeXt [10] (based on ResNetXt [11] but using “Squeeze-and-Excitation” (SE) blocks [10]).

In remote sensing applications, semantic segmentation aims to assign a land cover class to every pixel of an image. This task becomes the operation of extracting geospatial objects from remote sensing

images and is challenging due to the high variety of geospatial objects and their complexity (many displaying similar hyperspectral signatures and structural properties) and due to the strong trade-off between downsampling (which enables the abstraction and obtaining of richer information) and objects extraction (requiring a strong local context). Most of the existing methods use supervised learning to extract the geometries using their radiometric, spatial, and photometric features.

Deep learning applied to remote sensing has proven to be useful for aerospatial objects extraction [12] and aircraft recognition [13,14], for automobile detection [15], for building detection and mapping [16,17], for roof segmentation [18,19], and for vegetation and buildings segmentation [20]. Other works in the relevant literature focused on building frameworks for extracting multi-class geospatial objects (including bridges, ships, tennis courts) [21], on extracting pylons of electric transport lines [22,23], on detecting vehicles [24,25], on extracting photovoltaic panels using random forest techniques [26] or through semantic segmentation [27], or road segmentation using adversarial spatial pyramid networks [28]. One of the main drawbacks is related to the small study areas taken into account, featuring favorable scenarios (as also pointed out in [29]).

For this reason, in this article, we will explore the feasibility of training various segmentation models to extract and map wind turbines in optical high-resolution remote sensing orthorectified images. We will use an available dataset (point-type geometry) that georeferences 90% of the wind turbines in Spain, but whose genesis and sources are unknown, and therefore their quality is also unknown. We set ourselves the tasks of exploring the ability to use DL models in extracting wind turbines from aerial imagery and to characterize the quality (positional accuracy and completeness) of the said dataset in order to answer the following research questions:

- i. Which semantic segmentation network generates the best results in the task of wind turbines segmentation from high-resolution aerial imagery?
- ii. Can a DL model trained to extract wind turbines help in characterizing the quality of an available dataset in terms of positional accuracy and completeness?

To address these challenges, we propose a methodology for creating optimal Deep Learning-based solutions to extract the ground geometry of wind turbines by means of semantic segmentation using high-resolution aerial orthoimagery—for having them freely available, but we could have used high-resolution optical images obtained by remote sensing—and a methodology to evaluate the positional accuracy and the completeness of a dataset containing the geospatial objects.

The main contributions of this research are:

- We propose a methodology for evaluating the goodness of the wind turbine predictions extracted through semantic segmentation operations.
- We propose a methodology for evaluating the positional accuracy of the geospatial object extracted using segmentation techniques as a means to assess the quality of a dataset containing the wind turbines.
- We train five different semantic segmentation networks on a novel dataset composed of 5,140 tiles of high-resolution aerial orthoimagery and their corresponding cartographic mask containing the wind turbine feature.
- We study how changes in a network's structure (depthwise) and learning techniques affect a segmentation model's performance in the wind turbine extraction task.

The use of the best performing network, through the proposed methodology, allows to extract the features that represent wind turbines with few false-negatives and generate a set of data to use to compare and characterize the quality of another dataset of which the quality is unknown.

2. Data and Methodology

No official source with georeferenced wind turbines has been found. Some businesses in the field offer lists of their windmill farms registries, in some cases, even mentioning the coordinates of

the centroids of the wind farms (e.g., “The Wind Power”, by the Spanish Eolic Business Association, or the Spanish electric transport operator, e.sios). Unfortunately, they are not very useful for detecting the individual wind turbines that form the windmill farms. Nonetheless, we identified the websites of the Association of Renewable Energy Maintenance Companies (“Asociación de Empresas de Mantenimiento de Energías Renovables”, AEMER) and the database generated by Antonio Marín [30], and obtained a dataset containing 18,000 georeferenced wind turbines distributed among the entire Spanish mainland and the Balearic Islands (as seen in Figure 2).

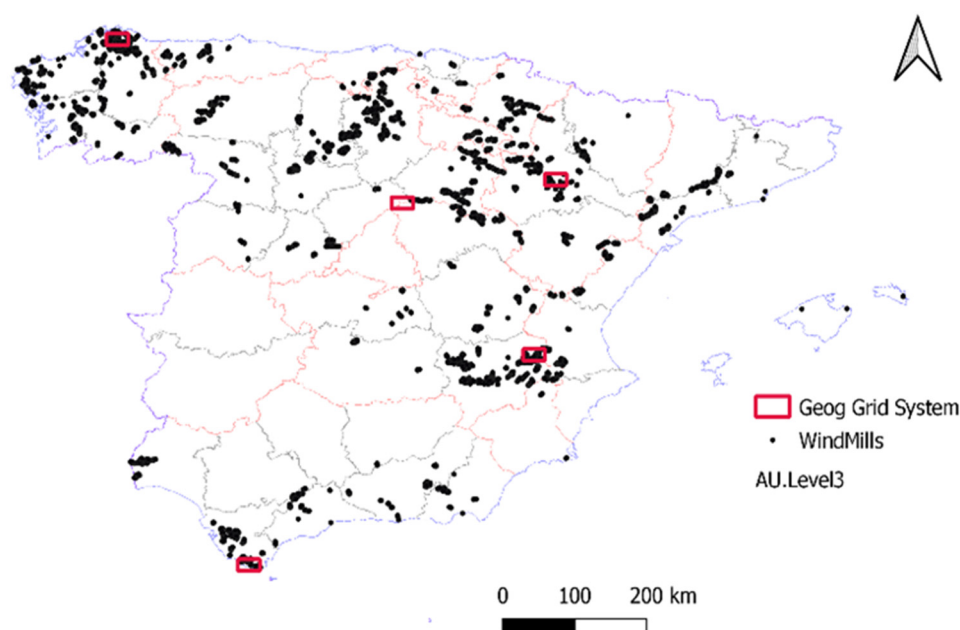


Figure 2. Geographical distribution of wind power facilities in the Spanish mainland and the Balearic Islands. Note: The location of the five areas studied in this work is marked with a red rectangle.

However, we promptly identified various quality problems related to this data repository: location errors consisting of systematic misplacements of the objects, wind turbines that do not exist in the aerial images (they may have been planned and even built, but are inexistent in the latest orthoimage freely obtained from the National Geographic Information Center (“Centro Nacional de Información Geográfica”, CNIG), and wind turbines not registered in the database. In Figure 3a–d, we can find examples of georeferencing problems and the challenges generated by the differences in types, sizes, and shadows of wind turbines.

To generate the training data bank, we conducted a visual inspection of the available wind turbine dataset (vector format) superimposed on the orthoimages. We applied a general criterion of representation for the footprint of the wind turbine’s base (a circle with a radius of 3 m), regardless of their size and height (since no such information was available at the individual level). The criterion of 3 m of radius for the bases of the wind turbines was adopted, after we analyzed the number of turbines aggregated by power (from 100 KW to 20 MW) and verified that 80% of the installed turbines had powers between 500 KW and 2 MW, typically using the last one, with a base diameter of 5 m. A radius of 3 m was selected for its representation so that the positioning errors of the centers, on an orthophoto of 0.5 m resolution, do not affect and the base is contained in that circle. This way, we generated segmentation masks (raster format) that contain information about the classes of each pixel “Wind Turbine Exists” or “No wind turbine”, the pixels defining the base of the wind turbines being easy to identify.

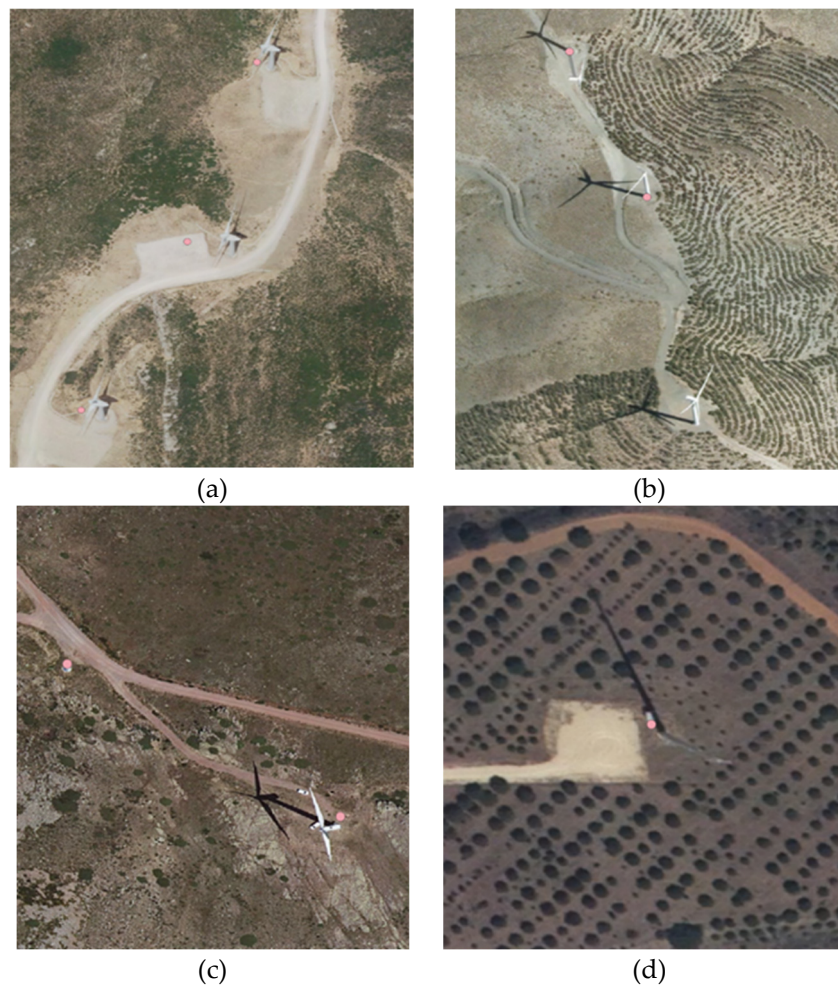


Figure 3. (a) Misplaced coordinates resulting in positional errors, (b) the lower wind turbine is not georeferenced, (c) the turbine on the left has not yet been built and (d) an element that is not a turbine, but a pole for measuring atmospheric parameters and/or a lightning rod. Notes: We can also observe that the shadows are different in each image, due to the sun's inclination at the moment the image was captured. Since all the imagery has the same scale, we can also observe the differences in the sizes of the wind turbines.

For the tagging operation, we divided the latest orthoimage tiles of the PNOA [31] (Spanish “Plan Nacional de Ortofotografía Aérea de España”) into tiles of 256×256 pixels at a scale of scale 1:2132 (~0.6 m pixel resolution) using a web application built for the tagging purpose. This web tool allows the simultaneous visualization of the aerial orthoimagery and its cartographic representation, and provides tools that allow the operators to work in parallel during the labelling operations. This same work logic has been successfully employed in other works related to the binary recognition of continuous geospatial elements (road network) in orthophotographs using deep neural networks ([32,33]). This way, we obtained a dataset of tiles containing 5140 aerial images of the wind turbines and their cartographic representations. This training dataset represents 28% of the wind turbines whose positional error and completeness accurate that we want to analyze. In Figure 4, we can find some illustrative examples of differences in size and textures around the turbines.

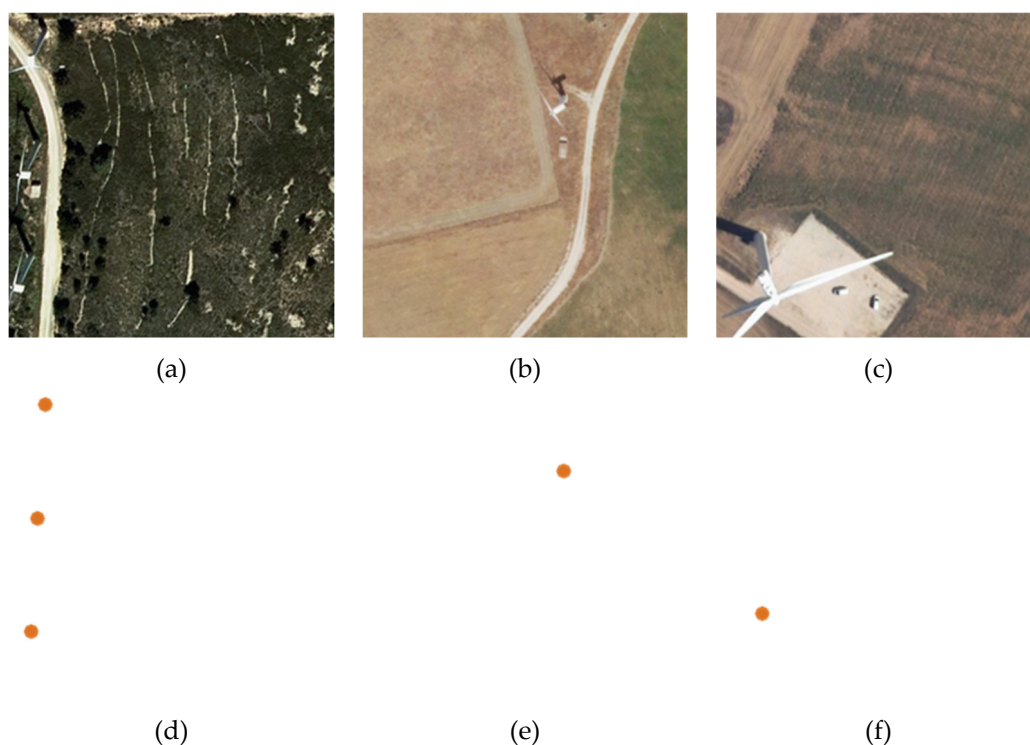


Figure 4. Examples of sets of aerial images divided into tiles of 256×256 pixels (a–c) and their correspondent cartographic representation (d–f) used for training the models. The differences in size of the wind turbines can be observed in the left and central images, as opposed to the one on the right.

The methodology proposed in this paper allows the design of Deep Learning solutions to segment wind turbines in high-resolution aerial orthoimagery through feature extraction.

As specified in the introduction, segmentation models are usually built using a segmentation architecture coupled with a base network (or backbone). In [34], we studied the appropriateness of using hybrid segmentation networks for extracting complex geospatial elements and contrasted their performance in comparison to state-of-the-art segmentation architectures, obtaining improvements in performance metrics of 2.7–3.5% when compared to the original architectures trained from scratch. The best performing hybrid models were featuring U-Net [7] as segmentation architecture with SEResNeXt50 [10] as a backbone network and LinkNet [8] coupled with EfficientNet [9] as backbone networks and will be used in this study. We also considered the b0, b1, b2 and b3 variations of EfficientNet as backbones to analyze how the depth of the backbone will impact the segmentation results.

2.1. Methodology for Evaluating the Goodness of the Semantic Segmentation Predictions

To evaluate the goodness of the predictions returned by the semantic segmentation models, we selected five areas of Spanish geography with extensions of 28×19 km each (National Geographical Grid System at 1:50,000 scale) from the H0008, H0383, H0432, H0767 and H1077 cartographic sheets (represented by red rectangles in Figure 2). The methodology applied for their evaluation consisted of eight steps, which are graphically summarized in Figure 5 and described below.

As a first step, traverse from North to South and West to East the latest orthoimage downloaded from the CNIG in ECW format, and extract the tiles. Given that the latest orthoimages have a spatial resolution of $0.25 \text{ m} \times 0.25 \text{ m}$ and the training of the network was performed on tiles with a pixel size of 0.597 m , the resolution of the latest orthophoto was halved (to 0.5 m/pixel). For this purpose, tiles of 512×512 pixels were extracted, these images being collapsed afterwards by choosing one out of every 4 pixels.

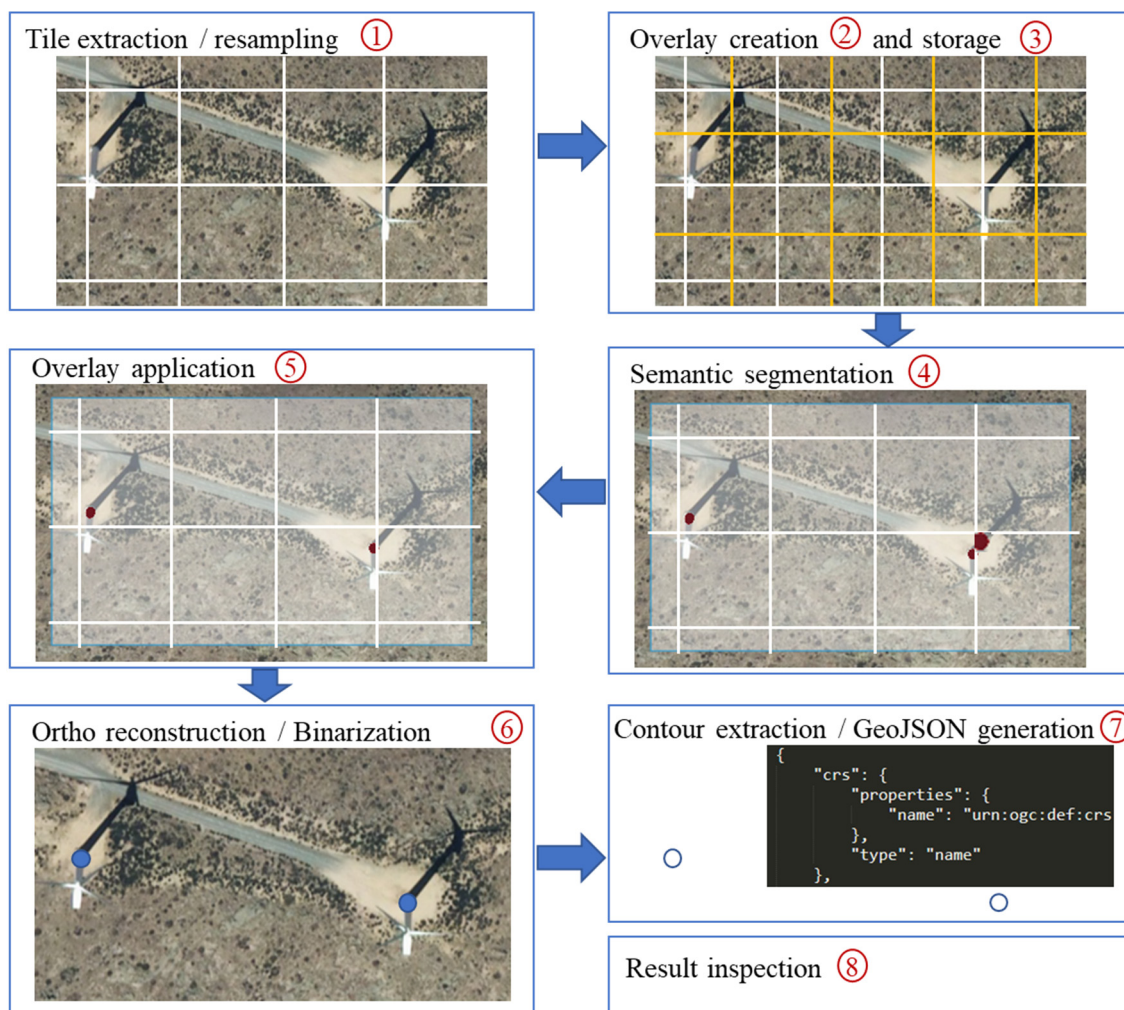


Figure 5. Graphical representation of methodology for evaluating the goodness of the semantic segmentation predictions.

Next, to avoid the common semantic segmentation problem where objects near the boundaries of an image are inaccurately extracted due to the lack of context, we apply a modified version of the method proposed by [29], where the predictions are averaged with overlays. Our proposed method consists of processing for each original 256×256 -pixel tile, four secondary tiles of the same size, their center being the centers of the four quarters of the original tile (as shown in Figure 6). In this way, we are able to keep the central 64×64 pixels and discard the pixels farthest from the center. This strategy also penalizes the computation as proposed by [29], but does not average or mask results.

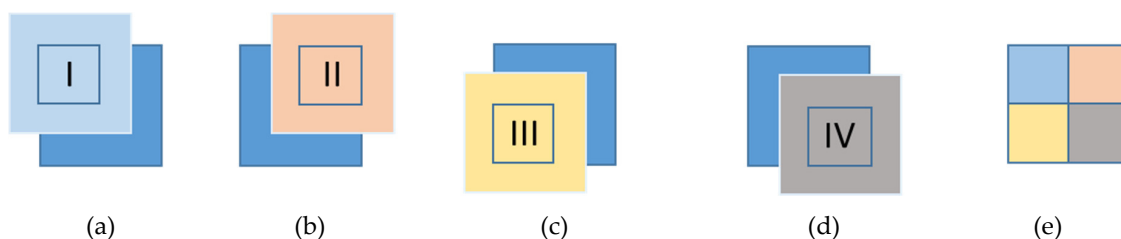


Figure 6. Abstractization of applying the methodology proposed in Section 2.1 to a random tile. We can see how four subimages are centered in each quadrant of the final tile (I, II, III, IV) of size 256×256 (a–d), and how the peripheral part of all of them is discarded to form an image of the same size in which the predictions close to the corners are not considered (e).

As a third step, a georeferenced 256×256 tiff image is constructed from the four matrices of 64×64 cells (featuring numerical values in the range (0,1) and stored on the hard disk (Figure 6).

Next, the algorithm runs through the entire orthoimage in a Z-shape (generating approximately 32,000 prediction images per analyzed region).

In the fourth step, the created overlay is used to reconstruct a prediction image with dimensions similar to the original one, through a merge operation (by generating a mosaic). The prediction image and the original will not have the same dimensions, since a 256×256 -pixels image cannot be centered at the edges of the original image and a 64-pixel strip is wasted at each edge. On a practical level, this is not a problem since orthoimages provided by CNIG are openly available and extend beyond the boundaries of the sheets, in such a way that two adjacent orthoimages always overlap in a strip.

Next, once the orthoimage is reconstructed, a binarization operation is carried out over the mosaic of predictions, by saturating the probabilities that exceed a parameterizable threshold to 1, and the rest to 0.

As a penultimate step, the binarized image is vectorized to generate the polygons of the bases of the identified wind turbines, in an operation similar to the extraction of the contour lines from a digital elevation model. The centroids of these polygons are calculated and stored as turbine coordinates in a GeoJSON file, to facilitate their visualization and subsequent analysis.

At last, each turbine identified in the five selected regions is analyzed and contrasted with the orthophotography itself to determine errors in the identification, whether they are false-positives or false-negatives, thus evaluating the goodness of the trained hybrid segmentation model.

2.2. Methodology for Evaluating the Positional Accuracy of the Dataset

The data generated by the semantic segmentation operation in areas where wind turbines are presumed to exist was considered as reference data to determine the positional accuracy of the initial dataset. In this operation, the methodology applied consisted of:

We start by iterating to evaluate the latest orthoimage available (centered on the initial known positions of the wind turbines) with the trained semantic segmentation model to obtain a matrix of 256 rows \times 256 columns containing the probability values in the range (0,1).

Next, binarize the probability matrix by saturating the high probability values to 1 and the low values to 0 in a way similar to the “nearblack or nearwhite” function.

Then, build a single band geotiff image that shares the geometric characteristics of the input orthoimage (pixel size and origin).

Afterwards, vectorize the binary image to extract the elliptical polygons (ideally circular) representing the wind turbine bases.

Finally, calculate and store the centroids of these polygons as the true coordinates of the wind turbines, adding as feature attributes the original coordinates in a GeoJSON file, to facilitate their subsequent analysis and visualization.

Once these five steps are completed, analyze the pairs of coordinates obtained to eliminate pairs of mismatched coordinates (where several turbines are identified in the same tile and the calculated coordinates correspond to only one of them) and statistically analyze the errors. The most precise data are considered those obtained from the segmentation operation, as they are obtained from orthoimages generated by an official governmental institution and their error is limited since the source of the data, its genealogy and its geometric quality are known.

It can be observed that steps 2–5 of the methodology proposed for evaluating the positional accuracy are the same as those of the methodology proposed in Section 2.1. However, this time, we work with smaller volumes of information (tiles of 256×256 pixels), instead of images of large dimensions ($54,000$ columns \times $28,000$ rows).

3. Implementation of the Proposed Methodology

The implementation of the methodology started with the creation of the training and test sets. Training neural networks require a large number of training examples to obtain a model capable of delivering good predictions. Creating these training banks is costly, not only due to the volume of the images required, but also because we need to ensure a correct correspondence between the segmentation masks (cartographic representation) of the wind turbines and the aerial images containing them. In Figure 4, we can find some examples of the aerial images and the representation of the bases of the wind turbines used for the training process. The dataset resulting from the tagging operation (5140 sets of tiles and their correspondent geographic representation) was randomly split into the training set (containing 4112 sets) and the test set (containing 1028 sets), following the 80–20% division criteria.

We built the networks using the deep learning “Segmentation Models” library [35] (based on Keras [36], with TensorFlow 1.14 [37] as backend) and trained them on an NVIDIA 2060 GPU. We modified the standard configurations of U-Net and LinkNet by replacing the default backbones both in the encoder and the decoder parts with networks that proved to be better suited for image segmentation tasks. The following hybrid segmentation models were considered for training: LinkNet-EfficientNet-b0, LinkNet-EfficientNet-b1, LinkNet-EfficientNet-b2, LinkNet-EfficientNet-b3 and U-Net-SEResNeXt50.

These models were trained with sets of input tiles and their segmentation mask via stochastic gradient descent to optimize the segmentation function over the training set. This method performs an update of the model’s weights and combined with backpropagation is the standard training algorithm for ANN. As hyperparameters, we considered Adam [38] to optimize the binary cross-entropy Jaccard loss function. This loss function measures the similarity between the prediction (pr) and the ground-truth (gt) when working with a network ending in a sigmoid function, $L(gt, pr) = -gt * \log(pr) - (1 - gt) * \log(1 - pr)$.

The segmentation model takes an RGB image of size $256 \times 256 \times 3$ and outputs a map of size $256 \times 256 \times 1$, by keeping the depth-wise argmax of each pixel. The pre-processing of the tiles begins with normalizing the intensity values of the pixels from the range of (0,255) to (0 and 1), this way avoiding the computation of very large numbers.

To overcome the overfitting behavior, we applied online data augmentation, exposing the model to more aspects of the data. Some of the transformations applied include random crops, random rotation and brightness, horizontal and vertical flips, and contrast and gamma shifts. We also applied feature mapping (transfer learning) from the ILSVRC dataset [39] to start from pre-trained weights instead of randomly initialized weights. This way, we ensure a better convergence of the model [40].

We started with a learning rate of 0.01; this value is reduced with a factor of 0.1 when the metrics plateaued for more than 10 epochs (up to a minimum value of $1 \times e^{-5}$). The learning process stopped when the performance stalled for more than 10 epochs or when the network started to display overfitting behavior (a pronounced higher performance on the train set and a lower performance on the test set). The models trained were stored in the open h5 format. In Figure 7, we can observe the behavior displayed by the models during the training process.

To use the models stored in the h5 exchange format, a virtual Python with Conda environment was configured on a Windows OS computer, while installing and configuring the GDAL v3.0.2 [41] and NumPy [42] libraries and dependencies.

The developed scripts rely on NumPy to handle multidimensional arrays of numbers (the initial 3-dimensional arrays of the RGB orthoimages and the final monoband array containing the prediction), to perform the decimation of the orthophotographs from higher spatial resolution (0.25 m) to a lower spatial resolution (0.5 m), to merge the centers of the prediction images (as described in Figure 6), and to perform the thresholding or binarization of the prediction images.

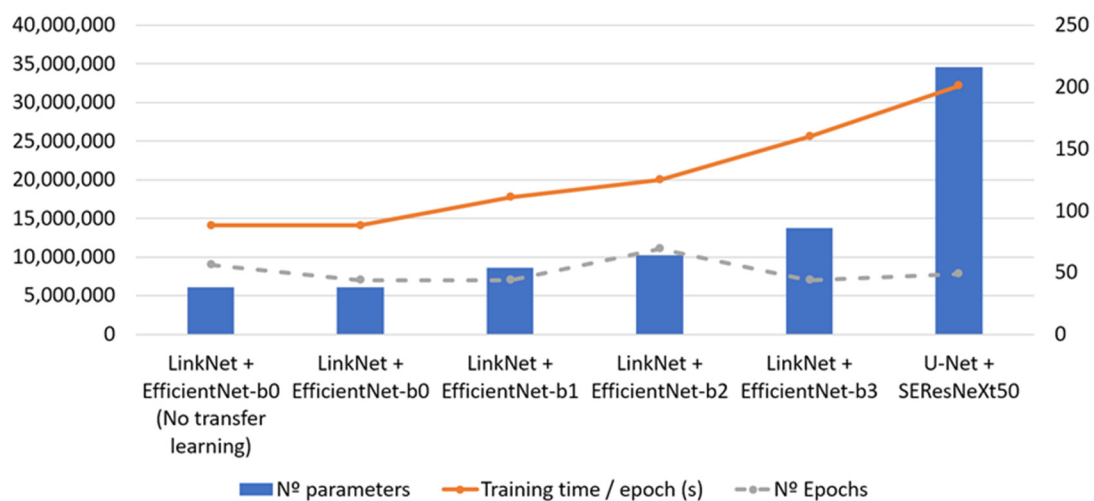


Figure 7. Description of the relation between the training time, the depth of the network and the number of epochs until convergence displayed by the semantic segmentation models during the training process.

The GDAL/ORG libraries were used to open the images stored in ECW format, to create new images both in-memory and in GeoTIFF format, to create vector datasets in memory or files in GeoJSON format, and to generate the polygons that enclose the pixels classified as belonging to the “Wind Turbine Exists” category (by applying the “*ContourGenerate*” function). To reduce the number of false-positives, we used GDAL to calculate the centroid of the feature corresponding to the georeferenced wind turbine, to eliminate the small errors related to the misclassification of other objects as a wind turbine, and to filter objects whose area is smaller than a selected threshold.

4. Results and Discussion

Table 1 shows the results of a systematical evaluation operation of the five test areas (presented in Figure 2) with the semantic segmentation networks trained to extract the positions of the wind turbines and features the number of wind turbines elements identified by the models.

Table 1. Results obtained from applying the methodology proposed in Section 2.1 (related to the semantic segmentation evaluation) to the five test areas selected.

| Sheet/Model | Nº of Wind Turbines | LinkNet + EfficientNet-b0 (No Transfer Learning) | LinkNet + EfficientNet-b0 | LinkNet + EfficientNet-b1 | LinkNet + EfficientNet-b2 | LinkNet + EfficientNet-b3 | U-Net + SEResNeXt50 |
|-------------|---------------------|--|---------------------------|---------------------------|---------------------------|---------------------------|---------------------|
| H1077 | 427 | 815 | 684 | 917 | 836 | 558 | 875 |
| H0383 | 481 | 1749 | 1889 | 2270 | 1702 | 1803 | 932 |
| H0767 | 278 | 477 | 344 | 980 | 714 | 459 | 391 |
| H0342 | 136 | 395 | 245 | 622 | 398 | 246 | 293 |
| H0008 | 598 | 995 | 787 | 982 | 2042 | 626 | 978 |

The lowest numbers of wind turbines detected in each area highlighted in green. This criterion is selected to minimize the number of false-positives detected by the network. When results are very close (almost the same), several networks are marked as candidates for that area. We can see that the model that shows the most figures highlighted in green is the hybrid network following the LinkNet architecture, with EfficientNet-b3 as the backbone (LinkNet + EfficientNet-b3).

In Table 2, we can find an evaluation of the completeness accuracy of the best performing model (identified in Table 1) in the five test areas.

Table 2. Results obtained from applying the second methodology proposed in Section 2.2 (related to the evaluation of completeness accuracy) to the five test areas with the hybrid network LinkNet + EfficientNet-b3.

| Wind Turbines | Original | Detected with Deep Learning | | | Completeness Original Data | |
|---------------|----------|-----------------------------|---------|---------|----------------------------|------------|
| Map sheet | # | # | False + | False − | Omission | Commission |
| H0008 (W) | 598 | 626 | 28 | 1 | 9 | 8 |
| H0383 (N-E) | 481 | 1303 | 813 | 3 ** | 12 | 0 |
| H0432 (C) | 136 | 246 | 106 | 0 | 4 | 0 |
| H0767 (E) | 278 | 459 | 177 | 0 | 6 | 1 |
| H1077 (S) | 427 | 558 | 215 | 85 * | 6 | 18 |

* Different Wind Turbines support structures (similar to high voltage towers). ** Three small wind turbines built next to an urban environment on a different support structure.

Heterogeneity can be observed in the results, mainly with regard to the number of detected false-positives which is generally large. This fact motivates the need for a cartographic reviewing process to eliminate the false-positives from the generated predictions, despite successfully performing joining and vectorizing operations. In the dataset generated for sheet H0383, this number seems excessive since the number of false-positives is greater than the true ones. False-negative values (when the model omits an element of the wind turbine) are acceptable; in two of the analyzed scenes there are none, while in a third the is only one, and. In the remaining two sheets, false-negatives are due to the existence of very small wind turbines (H0383) or with another type of support structure that is not cylindrical (H1077), but rather on the structure similar to a high-tension tower (as seen in Figure 8a,b) or are mounted on three cylinders that form the edges of a pyramid trunk to support the turbine (as seen in Figure 8c).

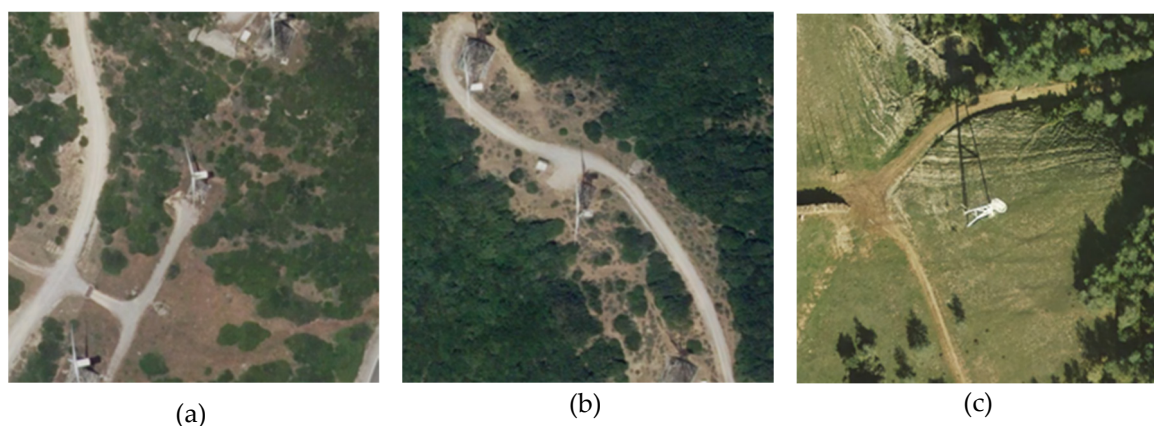


Figure 8. Unconventional types of wind turbines identified. In the left (a) and center (b), the turbine support structures look like high voltage transport pylons, while the example on the right (c) shows a structure consisting of three pillars in the shape of a stool.

Regarding the quality of the available dataset, the evaluation of the completeness in the exhaustively revised sheets shows that there are percentages of errors by omission ranging from 0.7% to 2.9%. However, there are no omission errors in three cases and, in the others, it represents 0.3% and 1.3%. If these results are extrapolated to the entire geography of the dataset, it confirms the trend in the proportions of errors by omission derived from the constructions of new wind turbines. At the end of 2019, the sector stated [43] the existence of 20,940 Wind Turbines in Spain. The 16.3% rate of the omission error suggested by this value is not too misaligned with the rates obtained by the model, knowing that the date of several of the orthophotographs used is 2017. We must also consider that the growth of the global windmill farm is not caused by the growth of existing ones, but rather by the creation of new parks (not registered in the dataset).

Regarding the quality of the positional accuracy of the initial data (18,000 features), we found 259 cases where the best performing segmentation model (LinkNet + EfficientNet-b3) has not identified any wind turbines within the 256×256 pixel tiles (centered on the coordinates of the objects). In 181 of these cases, a visual inspection confirmed the absence of the wind turbines (a commission error rate of 1.4% of the set, this value being higher than the systematic analysis of five well-distributed areas). On the contrary, in 69 cases, the network generated false-negatives, which represent a rate of 0.38%, within the range of values obtained in the areas studied. However, of those 69 cases, only 22 can be considered false-negative, 45 cases being caused by a special type of wind turbine base (not cylindrical, as can be seen in Figure 8a, as was the case in sheet H1077). For the remaining two cases, the orthophotographs shows different types of turbines (as seen in the Figure 8b). In addition, another nine wind turbines have a low power production capacity; therefore, they are smaller and do not physically resemble normal wind turbines.

By applying the proposed methodology systematically to the 18,000 wind turbines, a set of 18,602 positional relationships was obtained, corresponding to 17,741 different features (18,000 minus the previous 259). This difference in observations is because, in some cases, several turbines were identified in the same tile, all of them being stored. In the subsequent analysis, these cases are filtered (step 6 from Section 2.2), reaching the result of 17,741 turbines correctly identified by the network. The filtering was carried out by applying a distance criterion between the initial coordinates and those obtained from the deep learning evaluation; 861 observations being eliminated this way.

After filtering the data, we obtained a dataset that is 98.56% of the original and we proceeded to characterize the positional error in the planimetric coordinates (X, Y). The results obtained are presented in Table 3. This allows us to conclude that the dataset has positional accuracy characteristics that are a little worse for the Y component of the planimetry (0.5 m), a value that matches the resolution of the images used. In any case, the RMSE (root-mean-square error) is approximately 7 m according to this first analysis.

Table 3. Mean errors, standard deviation, and root-mean-square error (RMSE) in the planimetric coordinates (X, Y).

| | Min | Max | Mean | Sigma | RMSE |
|---|----------|---------|---------|--------|--------|
| X | −71.58 m | 74.27 m | −0.17 m | 6.94 m | 6.95 m |
| Y | −73.89 m | 74.48 m | −0.32 m | 7.42 m | 7.43 m |

To determine the main ranges of errors in both components, the histogram of the distances was calculated by grouping the values into 19 bins. We can see that, for approximately 75% of the georeferenced wind turbines, the error in X is between −1 and 5 m and between −3 and 4 m for the Y coordinate. Both histograms are shown in Figure 9.

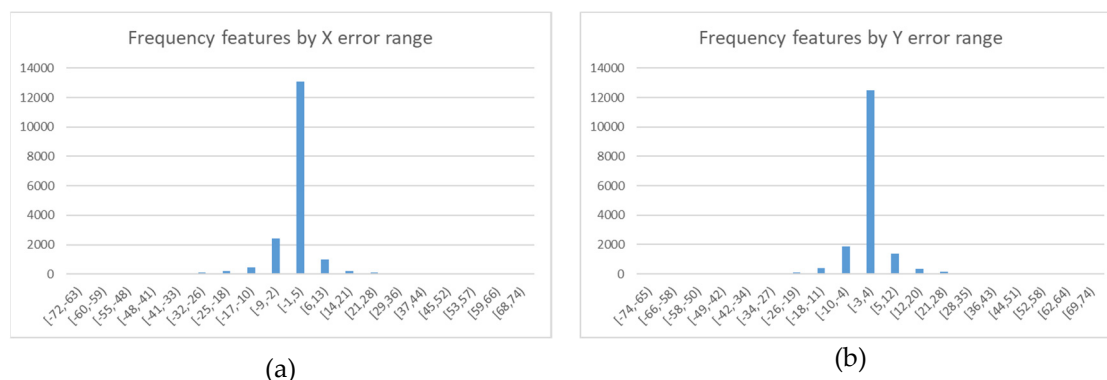


Figure 9. Histograms of positional errors in the X-coordinate (a) and Y-coordinate (b) grouped in 19 bins.

5. Conclusions

In this paper, we explored the capacity of different hybrid semantic segmentation models to extract wind turbines in high-resolution aerial orthoimagery. These models were trained on a dataset generated especially for this purpose and their generalization capacity (goodness of the predictions) was evaluated on five different areas of the Spanish territory (regions of 28 km × 19 km, represented with red rectangles in Figure 2).

We also proposed a new strategy to automatically deal with the problem of inaccurate extractions near the boundaries of the images. This strategy is described in Figure 6 and consists of automatically (1) dividing a given tile into four quadrants, (2) extract four secondary images with the same dimensions correspondent to each quadrant from the orthoimage with higher resolution, (3) segment each secondary tile, and (4) only consider their central part in the final segmentation map of the original tile. By applying this strategy, we achieved considerable improvements when extracting wind turbines close to the edges.

Moreover, we proposed a methodology as a workflow that allows the generation of a continuous image with the semantic segmentation predictions and its subsequent vectorization to extract the centers from the bases of the wind turbines. This methodology enabled the comparison of the performance of several hybrid segmentation models in different regions of the Spanish geography and the selection of the one delivering the lowest number of false-positives (in our case, LinkNet + EfficientNet-b3).

We also proposed a methodology as a workflow that has allowed the analysis of the quality in terms of positional accuracy and completeness of an unofficial dataset containing georeferenced wind turbines. In Section 5, we presented the statistical discrepancies of the georeferenced data (Table 3) and the discrepancies in completeness by terms of omission and commission (Table 2). As for the quality of completeness, the values presented are objective, but must be handled with caution, as the date of the latest available orthoimage does not correspond with the latest update of the analyzed dataset. This situation was observed in areas where wind turbines were registered, although they appear under construction in the imagery.

Regarding the geometric quality of the georeferencing of the evaluated dataset, the errors are more dispersive in the X coordinate than the Y coordinate and the values are significantly higher than the resolution of the aerial orthoimagery used 0.5 m (pixel size).

It has been proven that based on the proposed methodology, it is possible to evaluate the quality parameters of datasets whose genealogy and metadata are known using deep learning and semantic segmentation techniques. However, this evaluation has an important cost in the creation of the training datasets of the networks and the tests to select the architecture or hybrid network that generates better results for the application domain (in our case, wind turbines).

The results of the experiments carried out with the six hybrid semantic segmentation networks show that applying Feature Mapping from networks pre-trained with millions of images (e.g., ImageNet Large Scale Visual Recognition Challenge) helps to obtain better results when the training dataset is not very large, as in this case (5140 tiles), as seen in Figure 7 and Table 1. In our case, the best performing segmentation architecture uses LinkNet as base architecture and EfficientNet-b3 as the backbone.

As for the second research question, we found that it is possible to use semantic segmentation models to extract features with greater positional accuracy, so that they can be used to characterize the quality of a dataset whose genealogy is unknown. We were also able to evaluate the performance of the networks in detecting the wind turbines by systematically reviewing and comparing the predictions with the ground-truth in five different regions from the Spanish territory, this way obtaining the goodness of the results in terms of false-positives and false-negatives. This is a great achievement from a productive point of view, since it noticeably reduces the human effort required to update the cartography. This way, a systematic and very laborious revision is avoided; only places where the network has extracted features need to be reviewed.

The methodology proposed to characterize the quality in the studied terms (geometric and completeness) can be applied to any other type of feature that can be modeled as point representation in the cartography (e.g., electric towers, telecommunications antennas). Additionally, it can improve

the quality of collaboratively generated datasets (e.g., crowdsourcing) in which quality criteria are not well defined and the quality of each contributor's contributions cannot be quantified.

Author Contributions: Conceptualization, M.-Á.M.-C., C.-I.C. and R.A.; Data curation, M.-Á.M.-C. and J.-J.A.-J.; Formal analysis, M.-Á.M.-C., C.-I.C. and J.-J.A.-J.; Investigation, M.-Á.M.-C., C.-I.C. and R.A.; Methodology, M.-Á.M.-C. and C.-I.C.; Software, M.-Á.M.-C., C.-I.C. and R.A.; Supervision, M.-Á.M.-C. and R.A.; Writing—original draft, M.-Á.M.-C. and C.-I.C.; Writing—review and editing, M.-Á.M.-C., C.-I.C., R.A., and J.-J.A.-J. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

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

References

1. 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. [\[CrossRef\]](#)
2. Stetco, A.; Dinmohammadi, F.; Zhao, X.; Robu, V.; Flynn, D.; Barnes, M.; Keane, J.; Nenadic, G. Machine learning methods for wind turbine condition monitoring: A review. *Renew. Energy* **2019**, *133*, 620–635. [\[CrossRef\]](#)
3. Abedini, F.; Bahaghighat, M.; S'hoyan, M. Wind turbine tower detection using feature descriptors and deep learning. *Facta Univ. Ser. Electron. Energ.* **2020**, *33*, 133–153. [\[CrossRef\]](#)
4. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet classification with deep convolutional neural networks. *Commun. ACM* **2017**, *60*, 84–90. [\[CrossRef\]](#)
5. Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv* **2014**, arXiv:1409.1556.
6. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going deeper with convolutions. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 1–9.
7. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In Proceedings of the Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015, Munich, Germany, 5–9 October 2015; Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F., Eds.; Springer International Publishing: Cham, Switzerland, 2015; pp. 234–241.
8. Chaurasia, A.; Culurciello, E. LinkNet: Exploiting Encoder Representations for Efficient Semantic Segmentation. In Proceedings of the 2017 IEEE Visual Communications and Image Processing (VCIP), St. Petersburg, FL, USA, 10–13 December 2017; pp. 1–4. [\[CrossRef\]](#)
9. Tan, M.; Le, Q.V. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. *arXiv* **2019**, arXiv:1905.11946.
10. Hu, J.; Shen, L.; Sun, G. Squeeze-and-Excitation Networks. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 7132–7141.
11. Xie, S.; Girshick, R.B.; Dollár, 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 2017), Honolulu, HI, USA, 21–26 July 2017; pp. 5987–5995.
12. Cai, B.; Jiang, Z.; Zhang, H.; Zhao, D.; Yao, Y. Airport Detection Using End-to-End Convolutional Neural Network with Hard Example Mining. *Remote Sens.* **2017**, *9*, 1198. [\[CrossRef\]](#)
13. Zuo, J.; Xu, G.; Fu, K.; Sun, X.; Sun, H. Aircraft Type Recognition Based on Segmentation with Deep Convolutional Neural Networks. *IEEE Geosci. Remote Sens. Lett.* **2018**, *15*, 282–286. [\[CrossRef\]](#)
14. Li, Y.; Fu, K.; Sun, H.; Sun, X. An Aircraft Detection Framework Based on Reinforcement Learning and Convolutional Neural Networks in Remote Sensing Images. *Remote Sens.* **2018**, *10*, 243. [\[CrossRef\]](#)
15. Ding, P.; Zhang, Y.; Deng, W.-J.; Jia, P.; Kuijper, A. A light and faster regional convolutional neural network for object detection in optical remote sensing images. *ISPRS J. Photogramm. Remote Sens.* **2018**, *141*, 208–218. [\[CrossRef\]](#)
16. Alidoost, F.; Arefi, H. A CNN-Based Approach for Automatic Building Detection and Recognition of Roof Types Using a Single Aerial Image. *PFG J. Photogramm. Remote Sens. Geoinf. Sci.* **2018**, *86*, 235–248. [\[CrossRef\]](#)

17. Ma, J.; Wu, L.; Tang, X.; Liu, F.; Zhang, X.; Jiao, L. Building Extraction of Aerial Images by a Global and Multi-Scale Encoder-Decoder Network. *Remote Sens.* **2020**, *12*, 2350. [\[CrossRef\]](#)
18. Chen, Q.; Wang, L.; Wu, Y.; Wu, G.; Guo, Z.; Waslander, S.L. Aerial Imagery for Roof Segmentation: A Large-Scale Dataset towards Automatic Mapping of Buildings. *arXiv* **2018**, arXiv:1807.09532. [\[CrossRef\]](#)
19. Yang, H.L.; Yuan, J.; Lunga, D.; Laverdiere, M.; Rose, A.; Bhaduri, B. Building Extraction at Scale Using Convolutional Neural Network: Mapping of the United States. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2018**, *11*, 2600–2614. [\[CrossRef\]](#)
20. Shorter, N.; Kasparis, T. Automatic Vegetation Identification and Building Detection from a Single Nadir Aerial Image. *Remote Sens.* **2009**, *1*, 731–757. [\[CrossRef\]](#)
21. Li, Y.; Zhang, Y.; Huang, X.; Yuille, A.L. Deep networks under scene-level supervision for multi-class geospatial object detection from remote sensing images. *ISPRS J. Photogramm. Remote Sens.* **2018**, *146*, 182–196. [\[CrossRef\]](#)
22. Dutta, T.; Sharma, H.; Vellaiappan, A.; Balamuralidhar, P. Image Analysis-Based Automatic Detection of Transmission Towers using Aerial Imagery. In *Pattern Recognition and Image Analysis*; Paredes, R., Cardoso, J.S., Pardo, X.M., Eds.; Lecture Notes in Computer Science; Springer International Publishing: Cham, Switzerland, 2015; Volume 9117, pp. 641–651. ISBN 978-3-319-19389-2.
23. Tragulnuch, P.; Chanvimaluang, T.; Kasetkasem, T.; Ingprasert, S.; Isshiki, T. High Voltage Transmission Tower Detection and Tracking in Aerial Video Sequence using Object-Based Image Classification. In Proceedings of the 2018 International Conference on Embedded Systems and Intelligent Technology & International Conference on Information and Communication Technology for Embedded Systems (ICESIT-ICICTES), Khon Kaen, Thailand, 7–9 May 2018; pp. 1–4.
24. Lu, J.; Ma, C.; Li, L.; Xing, X.; Zhang, Y.; Wang, Z.; Xu, J. A Vehicle Detection Method for Aerial Image Based on YOLO. *J. Comput. Commun.* **2018**, *6*, 98–107. [\[CrossRef\]](#)
25. Cao, Y.; Wang, G.; Yan, D.; Zhao, Z. Two Algorithms for the Detection and Tracking of Moving Vehicle Targets in Aerial Infrared Image Sequences. *Remote Sens.* **2016**, *8*, 28. [\[CrossRef\]](#)
26. Malof, J.M.; Bradbury, K.; Collins, L.M.; Newell, R.G.; Serrano, A.; Wu, H.; Keene, S. Image features for pixel-wise detection of solar photovoltaic arrays in aerial imagery using a random forest classifier. In Proceedings of the 2016 IEEE International Conference on Renewable Energy Research and Applications (ICRERA), Birmingham, UK, 20–23 November 2016; pp. 799–803.
27. Yu, J.; Wang, Z.; Majumdar, A.; Rajagopal, R. DeepSolar: A Machine Learning Framework to Efficiently Construct a Solar Deployment Database in the United States. *Joule* **2018**, *2*, 2605–2617. [\[CrossRef\]](#)
28. Shamsolmoali, P.; Zareapoor, M.; Zhou, H.; Wang, R.; Yang, J. Road Segmentation for Remote Sensing Images using Adversarial Spatial Pyramid Networks. *arXiv* **2020**, arXiv:2008.04021. [\[CrossRef\]](#)
29. Dong, R.; Li, W.; Fu, H.; Gan, L.; Yu, L.; Zheng, J.; Xia, M. Oil palm plantation mapping from high-resolution remote sensing images using deep learning. *Int. J. Remote Sens.* **2020**, *41*, 2022–2046. [\[CrossRef\]](#)
30. Marín, A. Wtg_Spain_27052017. Available online: <https://amezet.carto.com/builder/fd430f8c-41cb-11e7-876d-0ecd1babdde5/embed> (accessed on 5 November 2020).
31. Instituto Geográfico Nacional Plan Nacional de Ortofotografía Aérea. Available online: <https://pnoa.ign.es/caracteristicas-tecnicas> (accessed on 25 November 2019).
32. De la Fuente Castillo, V.; Díaz-Álvarez, A.; Manso-Callejo, M.-Á.; Serradilla García, F. Grammar Guided Genetic Programming for Network Architecture Search and Road Detection on Aerial Orthophotography. *Appl. Sci.* **2020**, *10*, 3953. [\[CrossRef\]](#)
33. Cira, C.-I.; Alcarria, R.; Manso-Callejo, M.-Á.; Serradilla, F. A Framework Based on Nesting of Convolutional Neural Networks to Classify Secondary Roads in High Resolution Aerial Orthoimages. *Remote Sens.* **2020**, *12*, 765. [\[CrossRef\]](#)
34. Cira, C.-I.; Alcarria, R.; Manso-Callejo, M.-Á.; Serradilla, F. A Deep Learning-Based Solution for Large-Scale Extraction of the Secondary Road Network from High-Resolution Aerial Orthoimagery. *Appl. Sci.* **2020**, *10*, 7272. [\[CrossRef\]](#)
35. Yakubovskiy, P. *Segmentation Models*; GitHub: San Francisco, CA, USA, 2019; Available online: https://github.com/qubvel/segmentation_models (accessed on 16 October 2020).
36. Chollet, F. *Keras*. 2015. Available online: <https://github.com/fchollet/keras> (accessed on 16 October 2020).

37. Abadi, M.; Barham, P.; Chen, J.; Chen, Z.; Davis, A.; Dean, J.; Devin, M.; Ghemawat, S.; Irving, G.; Isard, M.; et al. TensorFlow: A system for large-scale machine learning. In Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI '16), Savannah, GA, USA, 2–4 November 2016; p. 21.
38. Kingma, D.P.; Ba, J. Adam: A Method for Stochastic Optimization. *arXiv* **2014**, arXiv:1412.6980.
39. Russakovsky, O.; Deng, J.; Su, H.; Krause, J.; Satheesh, S.; Ma, S.; Huang, Z.; Karpathy, A.; Khosla, A.; Bernstein, M.; et al. ImageNet Large Scale Visual Recognition Challenge. *Int. J. Comput. Vis.* **2015**, *115*, 211–252. [[CrossRef](#)]
40. Gupta, S.; Girshick, R.B.; Arbeláez, P.A.; Malik, J. Learning Rich Features from RGB-D Images for Object Detection and Segmentation. In *Proceedings of the 13th European Conference on Computer Vision (ECCV 2014), Zurich, Switzerland, 6–12 September 2014*; Proceedings, Part VII; Fleet, D.J., Pajdla, T., Schiele, B., Tuytelaars, T., Eds.; Springer: Berlin/Heidelberg, Germany, 2014; Volume 8695, pp. 345–360.
41. GDAL/OGR Contributors. *GDAL/OGR Geospatial Data Abstraction Software Library*; Open Source Geospatial Foundation: Chicago, IL, USA, 2020; Available online: <https://gdal.org/index.html> (accessed on 30 March 2020).
42. Van Der Walt, S.; Colbert, S.C.; Varoquaux, G. The NumPy array: A structure for efficient numerical computation. *Comput. Sci. Eng.* **2011**, *13*, 22–30. [[CrossRef](#)]
43. Asociación Empresarial Eólica-Spanish Wind Energy Association-Energía Eólica La eólica en España. Available online: <https://www.aeeolica.org/sobre-la-eolica/la-eolica-espana> (accessed on 7 November 2020).

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).