

Article

Cork Oak Production Estimation Using a Mask R-CNN

André Guimarães ^{1,*}, Maria Valério ², Beatriz Fidalgo ², Raúl Salas-Gonzalez ², Carlos Pereira ^{1,3}
and Mateus Mendes ^{1,4,*}

¹ Instituto Superior de Engenharia de Coimbra, Polytechnic of Coimbra—ISEC, 3030-199 Coimbra, Portugal

² School of Agriculture of Coimbra, Polytechnic of Coimbra—ESAC, 3045-093 Coimbra, Portugal

³ Departamento de Eng. Informática, CISUC—Centre for Informatics and Systems of the University of Coimbra, Pólo II, Rua Sílvio Lima, 3030-290 Coimbra, Portugal

⁴ Departamento de Eng. Eletrotécnica e Computadores, ISR—Institute of Systems and Robotics of the University of Coimbra, Pólo II, Rua Sílvio Lima, University of Coimbra, 3030-194 Coimbra, Portugal

* Correspondence: a21240003@isec.pt (A.G.); mmendes@isec.pt (M.M.)

Abstract: Cork is a versatile natural material. It can be used as an insulator in construction, among many other applications. For good forest management of cork oaks, forest owners need to calculate the volume of cork periodically. This will allow them to choose the right time to harvest the cork. The traditional method is laborious and time consuming. The present work aims to automate the process of calculating the trunk area of a cork oak from which cork is extracted. Through this calculation, it will be possible to estimate the volume of cork produced before the stripping process. A deep neural network, Mask R-CNN, and a machine learning algorithm are used. A dataset of images of cork oaks was created, where targets of known dimensions were fixed on the trunks. The Mask R-CNN was trained to recognize targets cork regions, and so the area of cork was estimated based on the target dimensions. Preliminary results show that the model presents a good performance in the recognition of targets and trunks, registering a mAP@0.7 of 0.96. After obtaining the mask results, three machine learning models were trained to estimate the cork volume based on the area and biometric parameters of the tree. The results showed that a support vector machine produced an average error of 8.75%, which is within the error margins obtained using traditional methods.

Keywords: forest management; *Quercus suber*; cork volume; machine learning; mask R-CNN



Citation: Guimarães, A.; Valério, M.; Fidalgo, B.; Salas-Gonzalez, R.; Pereira, C.; Mendes, M. Cork Oak Production Estimation Using a Mask R-CNN. *Energies* **2022**, *15*, 9593. <https://doi.org/10.3390/en15249593>

Academic Editor: Paul Stewart

Received: 31 October 2022

Accepted: 14 December 2022

Published: 17 December 2022

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



Copyright: © 2022 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 (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The cork oak forest is an asset of strategic importance for Portugal. It is formed by multiple-use forest ecosystems in which the main product is cork. Cork has a high economic value due to its technological properties, namely elasticity, impermeability, and good thermal insulation. Due to its low thermal conductivity and impermeability, it has been used in construction since pre-historical times. Nowadays, it is transformed and used in a variety of construction materials such as blocks, cork rubber or insulation sheets. It can also be transformed into grains and incorporated into filling materials. It can be used indoors and outdoors, in floors, walls or under tiles as a roof insulator. It can also be transformed and incorporated into paint, thus improving the paint's thermal properties. Cork materials in general have service life of over 50 years.

According to the last Portuguese national forest inventory (NFI), in 2013 the cork oak occupied approximately 23% of the forest territory [1]. Portugal is currently the largest cork exporter in the world, controlling a 62.4% share of the world trade in cork [2]. Cork oak is characterized by the development of cork, that is, a prominent bark in a continuous layer which involves the entire trunk and branches of the tree. Cork extraction (stripping) occurs periodically, usually on a nine- or ten-year cycle, and does not damage the tree. Afterwards, a new bark begins to form on the exposed stem surface [3]. The maximum height of cork stripping is established by law and is determined by the cork harvesting coefficient (HC),

which must not exceed the value of 3.0 and is defined as the ratio between stem harvesting height (H) and stem perimeter over cork, at breast height [4].

To estimate cork production, the IFN in Portugal uses circular field plots of 2000 m², where measurements of different dendrometric characteristics of the trees are carried out. However, given the sampling intensity adopted, the IFN results only produce results at the national and regional level and are not sufficiently detailed to guide management decisions at the forest or stand level [5]. Consequently, at the level of forest exploitation, new stand field inventories must be made to estimate cork oak production.

Traditional inventory methods are based on field measurements of diameter at breast height (DBH) and height of the cork oak. DBH is measured using a caliper or diameter tape. The hypsometer is used for measuring total tree height, crown height, the stem height free from branches and the height where stem begins to fork. The volume and cork production are then estimated using equations based on the DBH and tree height [6]. Alternatively, another non-destructive method, the Bitterlich relascope, can be used to estimate trunk diameters at various heights. Subsequently, the Smalian or Newton formulae can be applied to each section of the trunk separately to estimate more precisely the volume of each section [7,8]. However, this method is very laborious and expensive, issues which prevent its wider application. Therefore, it is necessary to develop new methodologies for calculating the volume of standing trees that are non-destructive, accurate and reduce forest management costs [9].

With the advancement of technology, remote sensing allows us to obtain several attributes of trees and forest stands [5]. These methods require less time and make measurement work easier. Nevertheless, the use of this technology remains limited by the costs involved, the difficulties of data processing, the lack of equipment and the requirement of specialized personnel [9].

Faced with this problem, the present work aims to propose a method to automatically estimate the volume of cork in a cork oak tree, using a Mask R-CNN deep neural network to find the area of interest and another machine learning algorithm to predict the volume of oak from the area and other data. The Mask R-CNN is an extension of the previously proposed Faster Region-Based Convolutional Network (Faster R-CNN) [10], which solves instance segmentation tasks.

A dataset of images collected in the field was used in the training process, containing a total of 62 images.

The article is organized as follows. Section 2 presents a literature review, describing some works carried out in the scope of volume calculation using image processing and deep learning methods. Section 3 presents the evolution of R-CNN neural networks. Section 4 describes the Mask R-CNN architecture in more detail. Section 5 describes the characteristics of the dataset and the difficulties encountered and addresses the methodology followed in the implementation of the deep learning model. Section 6 presents the obtained results of the Mask R-CNN model. Section 7 describes the methodology for calculating the trunk area after obtaining the mask from the output of Mask R-CN. Section 8 shows the results obtained with three different machine learning algorithms to estimate the final volume of cork. The last section presents the conclusions drawn from this study, as well as the analysis through the results obtained in the experiment.

2. Literature Review

The following subsections describe previous work based on computer vision methods and on deep learning to estimate the biometric parameters of the tree. The techniques used are capable of obtaining accurate results.

2.1. Traditional Methods

Depending on the degree of detail required in measuring the volume of a tree, it may be necessary to calculate only the volume of the trunk or to include the tree branches. In

this project, the focus will be only on the trunk volume since it is in this section of the tree that the cork is extracted.

Currently, there is a significant variety of techniques that are used to measure the volume of a tree. Considering that the trunk does not have the same diameter along its height, regardless of technique or method used, it is always necessary to segment the trunk into different sections. The volume is calculated for each section, and the partial volumes of each section will be considered to determine the total volume of the trunk.

Trunk volume can be obtained through direct measurement by a destructive method, where it is necessary to cut the tree and measure the diameter along the trunk using a measuring tape [11], or by remote measurement methods, performed with the support of professional measuring equipment that allows us to carry out a ground-level measurement. With the advancement of technology, the use of remote measurement methods allows us to spend less time measuring and make the measurement work easier, although the acquisition of this type of equipment can be very expensive.

Regarding common measuring instruments, hypsometers are the most used when the objective is to measure the height of a tree. Hypsometers can measure the height of the tree with high precision considering trigonometry and atmospheric pressure. It is possible to estimate the volume of a tree using binary equations based on DBH (diameter at breast height) and the height calculated by the hypsometer. These binary equations are called hypsometric relations. The hypsometric relations represent the height-diameter relationship of the tree, using data obtained from a set of trees. Through these data, the relationship between the DBH and the height of the tree is established, and a function is calculated to represent this relationship.

2.2. Computer Vision Methods

Zhang and Huang [12] present a method of measuring the height of a tree based on image processing. In this method, images of trees were collected, and three red dots were placed on each tree to serve as markers. One of the points was placed at the base of the tree, another one a meter from the base and the last one at the maximum possible height of the tree. The photographs were taken perpendicular to the ground, forming a 90-degree angle. During image processing, the coordinates of the three marking points were extracted. In order to extract the coordinates of the upper marking point, a model called HSI color model was used. The authors propose a method in which the image segmentation is performed on the three components of the HSI model, managing to segment the tree of all other objects present in the background of the image. After segmentation, the image is converted into a binary format and the coordinates are extracted through progressive scanning. Finally, the height of the tree is calculated using the theory of similarity between triangles. Experimental results indicate that the relative measurement error corresponding to the tree height prediction is about 4%.

Han [13] proposed a method to estimate the volume of a tree also using image processing, such as the method proposed by Zhang and Huang [12]. Two red marking points were placed on the tree trunk before the photograph was taken. After extraction, both the trunk and the marking points, the edge and the central axis of the trunk, were adjusted through the construction of a curve, so that it represented a better fit to the collected data. The method proposed by this study proved to be viable since it presented relative measurement errors in the order of 5.4%.

Another approach to calculating the volume was proposed by Putra et al. [14]. Here, the use of optical sensors, namely a smartphone camera, was evaluated. They also used image processing methods to estimate the circumference of trees in homogeneous and productive forests, especially seringueira and albizia plantations, with a real-time measurement approach. The images were captured at approximately one meter from the tree and with the camera pointed, parallel to the ground, at the trunk area, which allows us to calculating the diameter at breast height. Measurements performed using the camera showed acceptable accuracy, with a coefficient of determination of 95% and an RMSE of

approximately 7.9 cm. These correspond to a relative measurement error of about 9.4%. Despite the accuracy, this method is only applicable to trees with relatively circular shapes, and there are also several aspects that affect measurement errors such as the presence of inclined trees, irregular geometric shapes and the computer vision segmentation methods themselves, which sometimes are not the most suitable.

In the study by Coelho et al. [15], the volume of Corsican pine trees was estimated using computer vision techniques and classical formulae for volume determination. The area of the pine tree in the image was determined using a Grab Cut algorithm. Among the methods developed to estimate the height and volume of a tree, the ones that presented the best results were the use of hypsometric relations to calculate the height, and the Newton method to calculate the volume of the tree. These methods showed average errors of 12.18% and 10.90%, respectively.

This last study presents a viable volume calculation method for use as it shows similar errors when compared with traditional methods, allowing the reduction of time spent in the field and the high associated costs. The error of this method serves as a reference for this study.

Several studies have been carried out using LiDAR data to estimate measurements of trees, such as the DBH, height and volume. The methods used vary, but the more common ones use tree reconstruction or automatic methods to detect the tree and estimate its parameters.

Gonzalez et al. [16] used this method when scanning 29 trees (the specimen was not named, only that 9 were from Peru, 10 from Indonesia and 10 from Guyana). The scan was performed using a RIEGL VZ-400 3D terrestrial laser scanner. This scanner has a horizontal field of view of 360° and a vertical one of 100°, with a scan resolution of 0.06°. According to the authors, this method showed an RMSE of 3.29 m³ when estimating the volume of the referred trees.

Heurich M. [17] tested the use of airborne LiDAR for delineating individual trees. Airborne LiDAR are LiDAR systems mounted on vehicles such as airplanes, helicopters, and drones. These systems have been more commonly used for large-area retrieval of forest structural parameters due to the benefits of cost-efficiency. In this study, a cohort of 2584 trees, comprised of *Picea abies* and *Fagus sylvatica*, were scanned, and 76.9% of the trees could be recognized. Then, regression equations were used to determinate tree height, DBH and volume of single trees, resulting in a coefficient of determination (R^2) of 0.97 for tree height and of 0.90 for volume.

Although LiDAR techniques show relatively good results, they are expensive and time consuming.

2.3. Deep Learning Methods

In the study by Juyal et al. [18] a method was proposed. The aim was to estimate the diameter and height of a tree using Mask R-CNN neural networks, and to calculate biomass, a key indicator of ecological and vegetation management processes. The Mask R-CNN network was used to detect the tree and a reference (white rectangular square held parallel to the tree in the image). As the reference detected by the neural network has known dimensions, it was possible to estimate the circumference and height of the tree and, consequently, its volume. To train and test the neural network, images obtained by the Forest Research Institute in Dehradun were used. There is no indication of which species of trees were used. The use of Convolutional Neural Networks of the Mask R-CNN type allowed us not only to detect the object in the image, but also to produce a high-quality result and, therefore, to obtain a better understanding at the pixel level and thus to delimit the object with high accuracy.

The results show an average error of 8% for diameter detection and 12% for height detection.

This method is in part similar to the present approach. However, the present approach is applied to cork oak trees and aims to estimate the area of the trunk and the volume of cork.

3. Evolution of R-CNN

A Convolutional Neural Network (CNN) is a deep learning model with the ability to receive an input image, assign an importance (learning weights and biases) to various features/objects of the image and then able to differentiate them from each other. The pre-processing required in the deep neural network ConvNet is much less than that in other classification algorithms. While filters in primitive methods are handmade, requiring a lot of prior knowledge, CNNs have the ability to autonomously learn these characteristics [19].

A standard CNN cannot be used in this type of problem because the network output is variable. A possible approach to solve this type of problem would be to divide the image into different regions of interest and use a CNN to detect the presence of the object within each region. The problem with this type of approach is that the objects that are intended to be identified may be in different locations and have different aspects. That is, to correctly identify the objects in the image, a large number of regions of interest would have to be selected, something which would require a high and not viable processing power.

In order to overcome the problem of having to identify many regions of interest, with the aim of identifying objects quicker and reducing the necessary processing capacity, several neural network architectures have been proposed to aim at object detection. One of them is called R-CNN and was proposed by Girshick et al. [20]. This neural network uses an algorithm called selective search [21]. This can, instead of trying to classify a huge number of regions of interest, start by classifying just 2000 regions.

Even if R-CNN has partially solved some of the issues mentioned above, there are still some problems that this neural network was not able to fully mitigate. In fact, although it has been beneficial to reduce the number of regions of interest to be classified, it still takes a lot of training time to be able to classify the 2000 regions. Another negative factor is the amount of time needed to test an image, which makes the method unfeasible to implement for a real-time response. Another disadvantage also identified in this type of network is that the selective search algorithm, which selects candidate regions, does not have the ability to learn, which results in an incorrect identification of some candidate regions.

These disadvantages led to the proposal of new neural networks, based on the previous R-CNN and following a similar approach, having the advantage of being faster in object detection. These new networks are called Fast R-CNN [22] and Faster R-CNN [10].

4. Mask R-CNN

In the study published by Kaiming He et al. [23], an approach is proposed which is capable of efficiently detecting objects in an image while also generating a high-quality segmentation mask for each instance. The proposed method is called Mask R-CNN and it is an extension of the previously proposed method called Faster R-CNN. In addition to the bounding box recognition layer present in the Faster R-CNN model, another convolutional layer, capable of simultaneously predicting the object mask, was added. This approach proves to be effective, simple, and solid, facilitating the implementation of solutions for recognition problems at the pixel level of an image, something that the Faster R-CNN model is unable to solve.

The Faster R-CNN model has two outputs, the class to which the object belongs and a bounding box of the same object. Conversely, in the proposed model a third output, was added where the binary mask of each region of interest is obtained (Region of Interest—ROI). Thus, the Mask R-CNN model is an architecture created with the aim of integrating object detection with semantic segmentation. In practice, the Faster R-CNN architecture was combined with networks called Fully Convolutional Networks (FCNs), creating the Mask R-CNN.

Since an image pixel specification is required, it was necessary to make some adjustments to the Faster R-CNN architecture. It was verified that the regions of the feature map, resulting from the ROI Pooling layer, presented slight deviations from the regions of the original image. Due to this problem, the authors introduced the concept of ROI Align.

In the present work, the Mask R-CNN model was implemented in Python 3.7.11, using Keras 2.3.1 and Tensorflow 1.15.5.

5. Methodology for Mask R-CNN Model

This section describes the construction of the dataset, including data augmentation methodologies and the metrics used to evaluate the neural model.

The objective of this work is to create a model that follows a three-step approach in order to be able to predict the cork volume of a cork oak. The first stage, described in this section, is to create a deep learning model that receives as input an image of a cork oak and generates as output the mask of the part of the trunk from which the cork is extracted.

5.1. Dataset

The dataset, which has been made available by the authors to the public at [<https://www.kaggle.com/datasets/andreguim/cork-oak-segmentation>, accessed on 25 October 2022], consists of images of cork oaks obtained before the trees undergo the stripping process. Two or three targets were affixed on each tree trunk, depending on the height of the tree. The targets are detected by the neural network, and then used to calibrate the calculation of the segmented trunk area. A tripod was used to fix the camera so that the photographs were taken with the optical axis parallel to the ground. A preliminary dataset had 55 images, and it was later extended to 62 images after more pictures were taken in the field. Figure 1 shows an image of a cork oak with three targets.



Figure 1. Sample image of the dataset.

This dataset includes images of several types of trunks characteristic of trees of this species. Namely, there are straighter trunks, more curved trunks, as well as trunks with or without bifurcations. Despite having few instances, the dataset continues to be satisfactory for training models based on the Mask R-CNN network, since this network frequently allows us to obtain satisfactory results even in small datasets, as in this circumstance.

To train the model in a supervised way, the training and test images were annotated. The contours of the trunks and the respective targets were marked using the MakeSense.AI web tool. In addition to marking the images, this tool allows the extraction of files in the COCO JSON format, used for training Mask R-CNN networks. After annotation, the resulting dataset was randomly divided into segments of 80% for training, 11% for validation and 9% for testing. That resulted in 50 images belonging to the training set, 7 to the validation set and 5 to the test set. To minimize the problem of working with a small dataset, data augmentation techniques were applied, which are described below.

In order to reduce the error in obtaining the mask from the photograph, the photograph should be taken by a camera with good resolution (the better the resolution, the better the results will be), with favorable climatic conditions for a good perception of the tree and at a time of day when there is enough light to be able to clearly identify the part of the trunk from which the cork will be extracted.

It is also advisable that the images are taken close to the tree and that the camera is pointed as parallel to the ground as possible.

Therefore, each tree was photographed twice in different positions with respect to the tree at a distance of 5 m, and two more times at a distance of 10 m. The camera used was a Canon EOS 200D with an EFS 18–55 mm lens. All the fieldwork took place during the spring–summer season with days of good sunlight, i.e., under open sky conditions. Additionally, we tried to take the photographs without shade, i.e., with direct exposure to the sun (light), and in a position in which it avoids having another tree behind it that could confuse the surface of the working tree with another tree.

5.2. Data Augmentation

As mentioned before, object detection algorithms based on deep learning require a large amount of data to perform properly. However, the dataset used in this project has a small size. To minimize this problem, data augmentation techniques were used to increase the amount of training instances.

The data augmentation process generates new instances from the original data, using transformation methods such as rotation, translation and resizing [24]. The application of those techniques minimizes the overfitting problem.

In particular, the techniques used were: image rotation; image mirror; translation; and brightness adjustment. These transformations were applied during the training process, using the implementation that is available in the Mask R-CNN network.

5.3. Model Evaluation

In object segmentation, three distinct tasks are performed: one to determine if an object exists in the image, another to find the location of the object, and another to draw a binary mask over the object. Additionally, a typical dataset will have more than one class, whose distribution is not uniform, which is also the case in the data used in the present work.

In the present work, the Mean Average Precision (mAP) [25] was used as a criterion to evaluate the models. The global value is determined by calculating the mAP across all classes and at all intersection over union (IoU) boundaries [26]. The IoU metric, also known as the Jaccard index, allows us to quantify the percentage of overlap between the target mask and the mask predicted by the neural network. This metric is closely related to the Dice coefficient, which is often used as a loss function during training.

In a simplified way, the IoU metric measures the number of common pixels between the target and prediction masks, divided by the total number of pixels present in both masks, and can be calculated by the following formula:

$$IoU = \frac{TargetMask \cap PredictedMask}{TargetMask \cup PredictedMask}$$

The intersection ($A \cap B$) is composed of the pixels found in both the predicted mask and the real mask of the object, while the union ($A \cup B$) is made up of all the pixels found in either the predicted or target mask.

In the analysis of the results, three mAP limits are considered: mAP@0.5, mAP@0.7 and mAP@0.9. The mAP@0.5 measures the performance of the model with respect to object segmentation when the predicted mask overlaps by at least half of the real box. The mAP@0.7 shows the performance of the model in relation to object segmentation when the predicted mask overlaps by at least 70% over the real box. The mAP@0.9 follows the same logic as the previous ones, being that it represents the performance of the model in relation

to the segmentation of the object when the predicted mask overlaps by at least 90% of the real box.

Several experiments were carried out which allowed the evaluation of the different models. The experiments aimed to improve the performance of the model, through different technical approaches, namely: a modification of values of certain hyperparameters, an increase in the amount of dataset images and the use of some data augmentation techniques. The best model is presented in this paper.

6. Results of the Mask R-CNN Model

In this section, the configurations of the best models are described, and the results are presented and analyzed.

For the model construction, as referred above, a dataset containing 62 images was used. This was divided into 50 images for training and 7 for validation, leaving the other 5 for testing.

The settings defined for this experiment are shown in Table 1:

Table 1. Settings of the experiment.

| Parameter | Value |
|-----------------------------|-------------------------|
| GPU_COUNT | 1 |
| IMAGES_PER_GPU | 3 |
| STEPS_PER_EPOCH | 500 |
| VALIDATION_STEPS | 50 |
| BACKBONE | Resnet101 |
| TRAIN_ROIS_PER_IMAGE | 70 |
| NUM_CLASSES | 3 |
| RPN_ANCHOR_SCALES | (32, 64, 128, 256, 512) |
| RPN_NMS_THRESHOLD | 0.7 |
| RPN_TRAIN_ANCHORS_PER_IMAGE | 256 |
| IMAGE_MIN_DIM | 1024 |
| IMAGE_MAX_DIM | 1024 |

According to Mask R-CNN documentation, the IMAGES_PER_GPU parameter, used to define how many images are trained at once per GPU, consumes a large amount of memory. Since the GPU used to train the neural network has only 6 GB of dedicated memory, the value used in this experiment was only one image per GPU. It was also decided that the default values of learning rate and momentum should be kept.

The STEPS_PER_EPOCH parameter defines the number of steps in each training epoch. Although the dataset contains few images, this parameter was set to 500 to avoid spending a lot of training time on updates, performed by the TensorBoard, for providing the measurements and visualizations needed during the workflow [27]. The VALIDATION_STEPS parameter defines the number of validations performed at the end of each training period, and a value of 50 was chosen.

The maximum number of regions of interest to be considered in the final layers of the neural network is defined by the TRAIN_ROIS_PER_IMAGE parameter, which was parameterized with a value of 70. This value is relatively low compared to the one defined in the mask article RCNN (512). However, due also to the high memory consumption required, it was decided to carry out this experiment using the mentioned value.

The defined batch size is obtained by multiplying the parameters GPU_COUNT and IMAGES_PER_GPU, which in the case of this experiment means that the batch size value is 3. In this project, low batch values are used because the graphics card memory quickly exceeds its limit as this value is increased.

As already mentioned, due to the very small size of the dataset, four data augmentation techniques were used to increase the amount of training images:

- Image Rotation between -10° and $+10^\circ$;
- Vertical Flip (image mirroring);
- Translation of 10%;
- Brightness Adjustment (Add -30 to 30 to the brightness-related channels of the image).

It was decided to define the application of the techniques as following a probability of 83%, that is, each image in the dataset has an 83% probability of undergoing some modification and being reinserted again into the training set. In addition, the techniques began to be applied randomly and not sequentially, with each one being applied separately to each image, using the *imgaug* library. *Imgaug* [28] is a library for image augmentation in machine learning experiments.

Figure 2 illustrates the result of segmentation and classification, using the model generated in this experiment and shows a forked trunk with three legs, a very common situation in cork oak trees in Portugal. The analysis shows how it was possible to clearly identify the area of the trunk that had already been stripped (left image) and that will be stripped again (right image). The image shows a minor problem in the upper zone on the right leg, where the line separating the stripped and unstripped zone is not clearly visible.

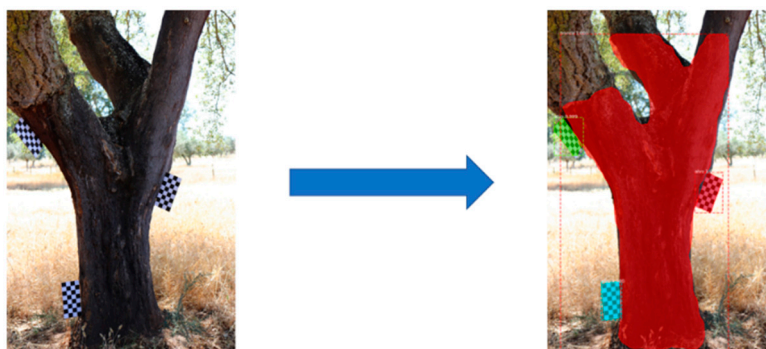


Figure 2. Example of segmentation and classification of objects.

Based on the image described, it is possible to verify that the performance of the model seems quite good.

After the training phase, the model was evaluated according to the aforementioned metrics. The results obtained are presented in Table 2.

Table 2. Performance results of the experiment.

| mAP@0.5 | mAP@0.7 | mAP@0.9 |
|---------|---------|---------|
| 0.964 | 0.964 | 0.577 |

The results allow us to conclude that the model is performing well, since its mAP@0.7 is above 96%.

The mAP@0.5 and mAP@0.7 values make this model viable for use in the volume calculation. Even so, there may be some space for progression, as mAP@0.9 may still be improved.

The results also show the need to carefully collect the images to minimize interference from other vegetation and the effect of light and against light.

7. Calculation of the Area of Cork to Be Stripped

After training and validating the detection model, the second stage of this work begins by creating a method for calculating the area of the mask resulting from the output of the

neural network. This was developed using Python 3.7.13, leveraging the availability of packages for computer vision. For this, firstly, since the images had different dimensions, it was necessary to preprocess them, in which a resize to 1024×1024 was applied so that the detection was performed correctly. Figure 3 shows examples of masks obtained by applying the previously generated model.

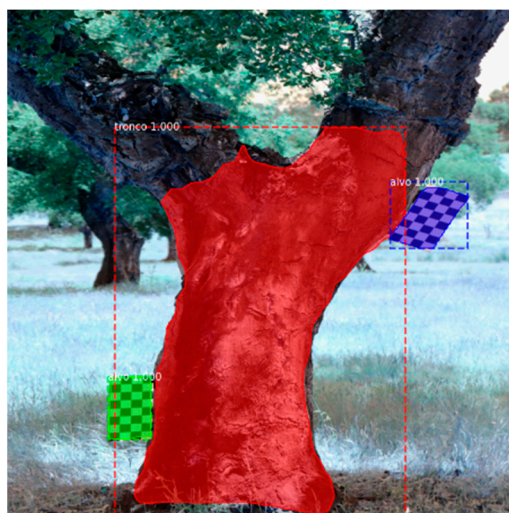


Figure 3. Resulting masks (trunk and targets).

After the detection of the masks (trunk and targets), the area represented by each pixel, in cm^2 , was estimated. The estimate was obtained through the ratio between the area in square pixels (px^2) of the targets, calculated from the masks using the `minAreaRect` method of the OpenCV library, and the actual known area in square centimeters (cm^2). The ratio was calculated for each of the targets present in the image, thus resulting in a list containing the different ratios and the designation of the corresponding targets. The formula for calculating the ratio is shown below:

$$Ratio = \frac{MaskArea(px^2)}{RealArea(cm^2)}$$

The variable ratio corresponds to the ratio between areas (px^2 to cm^2), the `MaskArea` is the calculated area, in px^2 , through the obtained mask and the `RealArea` is the known area of the target, in cm^2 . The value resulting from the application of the formula presented above was rounded to two decimal places.

With the previously calculated target ratios, it became possible to obtain the trunk mask area. For this, the trunk area in px^2 was obtained, which was calculated through the contour of the trunk mask using the contour area method again. However, to calculate the area in cm^2 , it was necessary to use the ratios previously calculated using the masks of the targets. Due to the natural characteristics of the trees and the positioning of the targets on their trunks, it was necessary to find a solution for identifying the most correct ratio to be used in the calculation of the trunk area. It was decided to use the average of the ratios of the different targets and to obtain the ratio to be used in the calculation of the trunk area, hereinafter called `FRatio`.

The trunk area, in cm^2 , is derived by dividing the calculated area, in px^2 , by the `FRatio`. The formula for calculating the trunk area is shown below:

$$Area(cm^2) = \frac{CalculatedTrunkArea(px^2)}{FRatio}$$

The value resulting from the application of the formula presented above was rounded to two decimal places and later converted into m^2 .

8. Volume Estimation

After calculating the area, three machine learning algorithms were tested to estimate the final volume of cork, which is the third stage of this project. Linear and non-linear regression algorithms were used, namely LinearRegression, Support Vector Regression (SVR) and MLPRegressor, all taken from the Python sk-learn library. All those algorithms were trained and tested using the same dataset. All stages of the model training process are presented below.

8.1. Biometric Parameters Dataset

A forest inventory was conducted in the spring–summer season. The data were collected in the field in circular plots of $1000 m^2$. In this inventory, the data recorded were the following: identification of the farm to which the trees belonged (property), tree number, tree diameter at the base. These were taken at 1.30 m and at 2 m height. At these heights of the trunk, the thickness of the cork (Eco030, Eco130; Eco200), the total height of the tree, the height of the first bough, the height of the first fork and the height of the stem without boughs or forks, as well as the diameter of the crown in the north-south and east-west directions were measured. Finally, the type of cork (segundeira or ama-dia) was recorded. Using these data, it was possible to estimate the trunk area and its volume. In addition to these features, another feature was added that corresponds to the area calculated using the developed methodology described above (area (m^2)). It should be noted that most of these features were not used for training, but only for dataset analysis, as will be explained in the next section.

Biometric records of some trees from the dataset of images were used in this study. It was not possible to collect records of all trees and, therefore, the dataset for volume estimation contains only data from 18 trees.

Since the dataset has a small size, it was necessary to resort to some data augmentation techniques. For that we used Roboflow, a computer vision developer framework for better data preprocessing and model training techniques, available from [<https://roboflow.com>, accessed on 20 July 2022]. The data augmentation techniques were applied to images whose data were present in the biometric parameter's dataset. The following transformations were applied:

- Crop: 0% Minimum Zoom, 10% Maximum Zoom;
- Flip: Vertical;
- Brightness: Between -5% and $+5\%$.

After applying the above techniques, the dataset now contains 100 instances since the area calculation algorithm was applied to all images, both transformed and non-transformed. Figure 4 illustrates part of the final dataset.

Once the process of increasing the number of records was concluded, the correlation between the dataset features and the extracted cork volume was analyzed. It was found that the base diameter and the diameter at breast height (1.30 m) were the ones that showed the best correlation, 0.98 and 0.99, respectively. Despite the high correlation, this fact can be explained by the fact that there are few records in the dataset.

The correlation between the areas (real and calculated) and the cork volume was also analyzed. It was found that the calculated area and the real area presented a correlation of 0.87 and 0.85, respectively. This reveals that the trunk area is closely linked with the cork volume. Figure 5 illustrates the correlation between the different features and the cork volume.

| | Property | Parcel | No_tree | Type | D030 | D130 | D200 | Eco030 | Eco130 | Eco200 | v (m3) | area (m2) | calculated_area (m2) |
|-----|----------|--------|---------|-----------|-------|-------|------|--------|--------|--------|--------|-----------|----------------------|
| 0 | mouchao | 2 | 1 | Sb amadia | 40.36 | 40.23 | 0.0 | 3.10 | 2.80 | 0.00 | 0.166 | 0.52 | 0.84 |
| 1 | mouchao | 2 | 1 | Sb amadia | 40.36 | 40.23 | 0.0 | 3.10 | 2.80 | 0.00 | 0.166 | 0.52 | 0.81 |
| 2 | mouchao | 2 | 1 | Sb amadia | 40.36 | 40.23 | 0.0 | 3.10 | 2.80 | 0.00 | 0.166 | 0.52 | 0.77 |
| 3 | mouchao | 2 | 1 | Sb amadia | 40.36 | 40.23 | 0.0 | 3.10 | 2.80 | 0.00 | 0.166 | 0.52 | 0.76 |
| 4 | mouchao | 2 | 1 | Sb amadia | 40.36 | 40.23 | 0.0 | 3.10 | 2.80 | 0.00 | 0.166 | 0.52 | 0.82 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 95 | mouchao | 11 | 12 | Sb amadia | 58.30 | 58.10 | 0.0 | 2.77 | 3.07 | 2.27 | 0.346 | 0.78 | 1.58 |
| 96 | mouchao | 11 | 12 | Sb amadia | 58.30 | 58.10 | 0.0 | 2.77 | 3.07 | 2.27 | 0.346 | 0.78 | 1.70 |
| 97 | mouchao | 11 | 12 | Sb amadia | 58.30 | 58.10 | 0.0 | 2.77 | 3.07 | 2.27 | 0.346 | 0.78 | 1.74 |
| 98 | mouchao | 11 | 12 | Sb amadia | 58.30 | 58.10 | 0.0 | 2.77 | 3.07 | 2.27 | 0.346 | 0.78 | 1.71 |
| 99 | mouchao | 11 | 12 | Sb amadia | 58.30 | 58.10 | 0.0 | 2.77 | 3.07 | 2.27 | 0.346 | 0.78 | 1.76 |

100 rows × 13 columns

Figure 4. Part of the Biometric Parameters Dataset.

```

v (m3)                1.000000
D130                  0.992616
D030                  0.985677
calculated_area (m2) 0.881204
area (m2)             0.868051
D200                  0.532158
Parcel                0.309632
Eco200                0.255604
No_tree               -0.202410
Eco130                -0.252046
Eco030                -0.321028
Name: v (m3), dtype: float64

```

Figure 5. Correlation between features and volume.

The following section describes all the experiments carried out to find the best algorithm to estimate the cork volume through the area previously calculated.

8.2. Results of Machine Learning Models

In this section, we consider various regressors and compare their performance in terms of the data collected in the field. We used the mean absolute error (MAE), the mean squared error (MSE) and root mean squared error (RMSE) as performance indicators. The MAE refers to the mean error between the predictions made and the actual values of these observations, taking the average over all observations. The MSE measures the average squared difference between the estimated values and the actual value. Therefore, values closer to zero are better. Due to the square, large errors are emphasized and have a relatively greater effect on the value of the performance metric. Finally, the RMSE is a metric that tells the square root of the average squared difference between the predicted values and the actual values in a dataset. The lower the RMSE, the better a model fits a dataset. All these metrics are useful to analyze the performance of regression models.

Before starting the training of the models, some features were removed. This is because, in a real situation, using only the camera, it is not possible to have access to certain data, such as the diameter and real area of the trunk, the cork thickness and the volume (target of our model). Therefore, the mentioned features were removed, as well as the tree number and the plot. The final model includes three features, the property, the type of cork of the tree and the area calculated using the developed method. The resulting dataset was randomly divided into 80% for training and 20% for testing.

Three machine learning algorithms were used, a LinearRegressor, Support Vector Regression (SVR) and MLPRegressor.

Table 3 shows the MAE, MSE and RMSE of each algorithm with the three features mentioned above.

Table 3. Performance results of the machine learning models.

| | MSE | RMSE | %MAE |
|-----------------|---------|---------|--------|
| LinearRegressor | 0.00096 | 0.03095 | 10.22% |
| SVR | 0.00078 | 0.02788 | 8.75% |
| MLPRegressor | 0.00084 | 0.02903 | 10.79% |

It is possible to verify, through the data presented in Table 3, that the SVR algorithm presented the best results, with the average error being 8.75% for the volume estimation. In any case, the other algorithms produced positive results, with the LinearRegressor algorithm showing an average error of 10.22% and the MLPRegressor algorithm showing an error of 10.79%.

9. Conclusions and Future Work

This study proposes a model based on the Mask R-CNN neural network. The model aims to detect and segment the trunk of a cork oak, making it possible to extract the mask from it. The extracted mask will be used to calculate the volume of cork that the cork oak is expected to produce.

The preliminary experimental results demonstrate that the model presents a good performance in the segmentation of the instances in the image. The mAP@0.5 and mAP@0.7 of the best model have values greater than 0.96, which makes this model viable to be used. The mAP@0.9 registered a value of 0.58, which shows that there is still room for future evolution.

The fact that the dataset has images of trunks of different shapes, something that often happens in cork oak trunks, can be beneficial to increase the generalization of the model. However, the size of the dataset is still small, so there is scope to improve the results obtained, and in future works we intend to annotate and include a greater number of instances.

The cork volume was calculated, using machine learning models that received data from the mask of the trunk that was extracted, the characteristics and biometric data of the trees. The SVR algorithm proved to be the best algorithm for this problem, presenting an average error of 8.75%, which makes this model perfectly viable to be used. This error is also very good when compared to other methods for volume calculation, described in the state of the art.

Future work includes obtaining more images and also enrich the biometric parameters dataset with other significant features.

Author Contributions: Methodology, A.G, B.F, R.S.-G., C.P. and M.M.; Software, A.G.; Data curation, M.V. and R.S.-G.; Writing—original draft, A.G.; Writing—review & editing, M.V., B.F, R.S.-G., C.P. and M.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The dataset and scripts have been made available to the public at <https://www.kaggle.com/datasets/andreguim/cork-oak-segmentation>.

Acknowledgments: The authors would like to thank Herdade do Mouchão for allowing the collection of photographs and measurements of cork oaks from the Herdade, data that were used in the scope of this work.

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

References

1. ICNF. *IFN6—Áreas dos usos do solo e das Espécies Florestais de Portugal Continental*; Resultados Preliminares; Instituto da Conservação da Natureza e das Florestas: Lisboa, Portugal, 2013; 33p, (Available only in Portuguese).
2. APCOR. Cork Yearbook 2020. Available online: <https://www.apcor.pt/en/portfolio-posts/apcor-year-book-2020/> (accessed on 4 April 2022).
3. Pereira, H. *Cork: Biology, Production and Uses*; Elsevier: Amsterdam, The Netherlands, 2007.
4. Ministério da Agricultura, do Desenvolvimento Rural e das Pescas (MADRP)., Diário da República de Portugal—I Série—A., Decreto-Lei n.º 169/2001 de 25 de Maio, (Available only in Portuguese). Available online: <https://data.dre.pt/eli/dec-lei/169/2001/05/25/p/dre/pt/html> (accessed on 4 April 2022).
5. Kangas, A.; Astrup, R.; Breidenbach, J.; Fridman, J.; Gobakken, T.; Korhonen, K.T.; Maltamo, M.; Nilsson, M.; Nord-Larsen, T.; Næsset, E.; et al. Remote sensing and forest inventories in Nordic countries—Road map for the future. *Scand. J. For. Res.* **2018**, *33*, 397–412. [[CrossRef](#)]
6. ICNF. *IFN6—Relatório Final do Inventário Florestal Nacional Portugal Continental*; Instituto da Conservação da Natureza e das Florestas: Lisboa, Portugal, 2015; (Available only in Portuguese).
7. Van Laar, A.; Alparslan, A. *Forest Mensuration*; Springer: Dordrecht, The Netherlands, 2007; Volume 13.
8. West, P.W. *Tree and Forest Measurement*, 2nd ed.; Springer: Berlin/Heidelberg, Germany, 2009. [[CrossRef](#)]
9. Marzulli, M.I.; Raunonen, P.; Greco, R.; Persia, M.; Tartarino, P. Estimating tree stem diameters and volume from smartphone photogrammetric point clouds. *Forestry* **2020**, *93*, 411–429. [[CrossRef](#)]
10. Gavrilescu, R.; Zet, C.; Foşalău, C.; Skoczylas, M.; Cotovanu, D. Faster r-cnn:an approach to real-time object detection. In Proceedings of the 2018 International Conference and Exposition on Electrical and Power Engineering (EPE), Iasi, Romania, 18–19 October 2018; pp. 0165–0168.
11. Koirala, A.; Montes, C.R.; Bullock, B.P.; Wagle, B.H. Developing taper equations for planted teak (*Tectona grandis* L.f.) trees of central lowland Nepal. *Trees For. People* **2021**, *5*, 2666–7193. [[CrossRef](#)]
12. Zhang, J.; Huang, X.Y. Measuring method of tree height based on digital image processing technology. In Proceedings of the 2009 1st International Conference on Information Science and Engineering, ICISE, Nanjing, China, 26–28 December 2009; pp. 1327–1331.
13. Han, D. Standing tree volume measurement technology based on digital image processing. In Proceedings of the International Conference on Automatic Control and Artificial Intelligence (ACAI 2012), Xiamen, China, 3–5 March 2012; pp. 1922–1925.
14. Putra BT, W.; Ramadhani, N.J.; Soedibyoy, D.W.; Marhaenanto, B.; Indarto, I.; Yualianto, Y. The use of computer vision to estimate tree diameter and circumference in homogeneous and production forests using a non-contact method. *For. Sci. Technol.* **2021**, *17*, 32–38. [[CrossRef](#)]
15. Coelho, J.; Fidalgo, B.; Crisóstomo, M.M.; Salas-González, R.; Coimbra, A.P.; Mendes, M. Non-Destructive Fast Estimation of Tree Stem Height and Volume Using Image Processing. *Symmetry* **2021**, *13*, 374. [[CrossRef](#)]
16. Gonzalez de Tanago, J.; Lau, A.; Bartholomeus, H.; Herold, M.; Avitabile, V.; Raunonen, P.; Martius Ch Goodman, R.C.; Disney, M.; Manuri, S.; Burt, A.; et al. Estimation of above-ground biomass of large tropical trees with terrestrial LiDAR. *Methods Ecol. Evol.* **2018**, *9*, 223–234. [[CrossRef](#)]
17. Heurich, M. Automatic recognition and measurement of single trees based on data from airborne laser scanning over the richly structured natural forests of the Bavarian Forest National Park. *For. Ecol. Manag.* **2008**, *255*, 2416–2433. [[CrossRef](#)]
18. Juyal, P.; Sharma, S. Estimation of Tree Volume Using Mask R-CNN based Deep Learning. In Proceedings of the 2020 11th International Conference on Computing, Communication and Networking Technologies, ICCCNT 2020, Kharagpur, India, 1–3 July 2020.
19. Sharma, N.; Jain, V.; Mishra, A. An Analysis of Convolutional Neural Networks for Image Classification. *Procedia Comput. Sci.* **2018**, *132*, 377–384. [[CrossRef](#)]
20. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Columbus, OH, USA, 23–28 June 2014.
21. Uijlings, J.R.; Van De Sande, K.E.; Gevers, T.; Smeulders, A.W. Selective search for object recognition. *Int. J. Comput. Vis.* **2013**, *104*, 154–171. [[CrossRef](#)]
22. Girshick, R. Fast r-cnn. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 7–13 December 2015.
23. He, K.; Gkioxari, G.; Dollar, P.; Girshick, R. Mask r-cnn. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017.
24. Roth, H.R.; Lu, L.; Liu, J.; Yao, J.; Seff, A.; Cherry, K.; Kim, L.; Summers, R.M. Improving Computer-Aided Detection Using Convolutional Neural Networks and Random View Aggregation. *IEEE Trans. Med. Imaging* **2016**, *35*, 1170–1181. [[CrossRef](#)] [[PubMed](#)]
25. Everingham, M.; Van Gool, L.; Williams, C.K.I.; Winn, J.; Zisserman, A. The PASCAL Visual Object Classes Challenge 2012 (VOC2012) Results. 2012. Available online: <http://www.pascal-network.org/challenges/VOC/voc2012/workshop/index.html> (accessed on 16 May 2022).

26. Rezatofighi, H.; Tsoi, N.; Gwak, J.; Sadeghian, A.; Reid, I.; Savarese, S. Generalized intersection over union: A metric and a loss for bounding box regression. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 16–17 June 2019; pp. 658–666.
27. Get Started with TensorBoard. 2022. Tensorflow. Available online: https://www.tensorflow.org/tensorboard/get_started (accessed on 3 May 2022).
28. Imgaug 0.4.0 Documentation. 2020. Imgaug. Available online: <https://imgaug.readthedocs.io/en/latest> (accessed on 16 May 2022).