sustainability

MDPI

Article

Leaf Disease Segmentation and Detection in Apple Orchards for Precise Smart Spraying in Sustainable Agriculture

Gary Storey *, Qinggang Meng and Baihua Li

Department of Computer Science, Loughborough University, Loughborough LE11 3TU, UK; q.meng@lboro.ac.uk (Q.M.); b.li@lboro.ac.uk (B.L.)
* Correspondence: g.storey@lboro.ac.uk

Abstract: Reduction in chemical usage for crop management due to the environmental and health issues is a key area in achieving sustainable agricultural practices. One area in which this can be achieved is through the development of intelligent spraying systems which can identify the target for example crop disease or weeds allowing for precise spraying reducing chemical usage. Artificial intelligence and computer vision has the potential to be applied for the precise detection and classification of crops. In this paper, a study is presented that uses instance segmentation for the task of leaf and rust disease detection in apple orchards using Mask R-CNN. Three different Mask R-CNN network backbones (ResNet-50, MobileNetV3-Large and MobileNetV3-Large-Mobile) are trained and evaluated for the tasks of object detection, segmentation and disease detection. Segmentation masks on a subset of the Plant Pathology Challenge 2020 database are annotated by the authors, and these are used for the training and evaluation of the proposed Mask R-CNN based models. The study highlights that a Mask R-CNN model with a ResNet-50 backbone provides good accuracy for the task, particularly in the detection of very small rust disease objects on the leaves.

Keywords: computer vision; instance segmentation; sustainable agriculture

Citation: Storey, G.; Meng, Q.; Li, B. Leaf Disease Segmentation and Detection in Apple Orchards for Precise Smart Spraying in Sustainable Agriculture. *Sustainability* **2022**, *14*, 1458. https://doi.org/10.3390/su14031458

Academic Editor: Michael S. Carolan

Received: 30 December 2021
Accepted: 24 January 2022
Published: 27 January 2022

1. Introduction

The application of chemicals such as fungicides, pesticides and herbicides via spraying in agriculture, with the aim to control plant disease, pests and weeds and ensure good crop yields, is common. Current agricultural spraying practices have been identified as unsustainable due to the environmental and health issues [1,2]. These environmental issues include land contamination which can be caused by chemical run-off from the field and drainage issues due to the volume of the chemicals applied [3]. On the other hand, chemical drift is where sprayed chemicals are carried by the wind onto the neighbouring crops/fields or residential areas can result in crop contamination and grass contamination in children's playgrounds [1]. Trace chemicals from spraying can also be found within the crop itself with studies showing high levels presence in fruits and vegetables which can be toxic when consumed [4]. Furthermore, there are health risks associated with exposure to the chemicals to groups including production workers, formulators, sprayers, mixers, loaders and agricultural farm workers [2].

Legislative bodies such as the European Commission and French government have formally identified the unsustainable usage of chemicals in agriculture as a serious environmental and health issues and have introduced incentive policies which target a reduction in chemical use as a key option to reduce the contamination of the environment [5]. Strategies such as Integrated Pest Management (IPM) are being championed by the European Commission, which has the aim to provide a less aggressive approach to managing agricultural land. One of the principle aims of IPM is that chemicals applied will be as specific as possible for the target and shall have the least side effects on human health, non-target organisms and the environment [6]. This approach has the potential to play an important role in reductions through sustainable agricultural methods.

One area in which efforts have been made to introduce sustainable practices is through improving spraying technology, which can also provide an economic benefit to the farmer through use of less chemicals. New spraying technologies have shown the potential to significantly improve through a reduction in chemicals used [7]. Traditionally the application of fungicides, herbicides and pesticides through spraying has been performed aggressively and uniformly on crops with an overuse of chemical to ensure the eradication of diseases, weeds and pests, respectively. Both hydraulic and hydro-pneumatic sprayers are used for this process in a number of forms including manually operated backpack sprayers, self-propelled units and tractor operated sprayers. While mechanical and electrical advances play a part in improving spraying systems, another important area is that of smart spraying systems, which aims to apply artificial intelligence (AI) and computer vision to provide more intelligent and precise spraying [8].

AI and more specifically computer vision can be used to provide information to a spraying system, this could be target location, size and disease state. Computer vision techniques have been previously researched in the detection of crop diseases [9–11] and weeds [7] in different areas of agriculture. Abdulridha et al. [9] detected laurel wilt disease in avocado fruit using aerial multi-spectral imaging and using these images to classifying healthy from stressed avocado plants. Machine learning algorithms namely the multi-layer perceptron and K-nearest neighbour where applied to find the best multi-spectral band combinations and perform the classification. Vine leaf disease was the topic of Pantazi et al. [10], where a pipeline of computer vision method using segmentation was applied to detect disease state. The pipeline consists of using the GrabCut algorithm for the segmentation of foreground leaves, the foreground leaf pixels have features extracted using Local Binary Patterns and their histograms were utilised for training a one class Support Vector Machine. In Partel et al. [7], a YOLO object detection based model was applied to identify between the portulaca weed, sedge and pepper plant types from aerial photography. The Plant Pathology Challenge 2020 dataset was introduced in Thapa et al. [11], which also provided some benchmarks results on image level disease classification using a ResNet-50 network pre-trained on ImageNet. Although the system performed well on image with single disease present, it struggled with images with multiple disease symptoms.

While promising results have been shown in the classification of crop disease from leaf images [12], there are some limitation regarding the datasets in which each image consists of single leaf with a contrasting background, this is not comparable to real-world conditions where leaves are bunched together, with other foliage as the background. This issue was also a motivation for the introduction of the Plant Pathology Challenge 2020 dataset [11]. Object detection methods such as YOLO [7] have been applied with good accuracy to agriculture-based detection but the nature of bounding box based detection is not always precise, especially when the shape of the detected objects is not rectangular. Therefore, object detection is not accurate enough to effectively direct a smart spraying system. Segmentation methods can theoretically provide more precise detection due to the pixel level accuracy, but many methods do not have the capability to distinguish between instances of the same object class in an image, a strength of object detection. This downside can be seen in Pantazi et al. [10] where a disease can only be detected at an image level rather than leaf level if many leaves are present in a single image. Precisely classifying each pixel of an image for each object detected has the potential to provide precision location information to a smart spraying system.

In this paper, an initial study is presented with the aim is to address the challenge of providing pixel level classification on images that depict real-life scenarios (i.e., multiple leaves per image with different disease states), which has the potential to provide accurate instructions to a smart spraying system capable of reducing chemical usage. Specifically, we propose an instance segmentation based system for leaf and rust disease identification in apple orchards. In apple and pear orchards, rust disease can be a common issue that can be effectively managed through the application of fungicides that are also used for apple scab disease. It is especially problematic in areas with an abundance of cedar trees which

creates high inoculum levels, the cause of cedar rust disease. Rust disease presents as a distinctive visual pattern on the leaves and fruit as a shown Figure 1.

Figure 1. Rust disease on the leaves of an apple tree.

The proposed system uses the state-of-the-art Mask R-CNN [13] method for instance segmentation from standard RGB images. Both ResNet [14] and MobileNetV3 [15] convolutional neural networks are trained and evaluated as the feature extraction backbone of Mask R-CNN. The publicly available Plant Pathology Challenge 2020 dataset [11] provides images for training and evaluating the proposed systems. As this dataset does not include segmentation maps, we manually annotated a subset of the images to allow. To our knowledge, no previous research has used instance segmentation methods in for this application. The main contributions of this paper are as follows:

1. A Mask R-CNN based instance segmentation system for precise leaf and rust identification on apple trees from RGB images;
2. Manually annotated segmentation maps for a subset of the Plant Pathology Challenge 2020 data set;
3. Benchmark evaluations of the proposed system for object detection, segmentation and disease detection accuracy, on the newly annotated images.

This paper is organised as follows: Section 2 reviews the related works. Section 3 describes the proposed instance segmentation system for leaf and rust disease identification. Section 4 describes the performance of the proposed system. Finally, the discussion and scope for future works are in Section 5.

2. Related Works

In this section, computer vision methods applicable in the domain of disease and pest detection are discussed. Firstly an overview of key object detection and segmentation methods are discussed, followed by an overview of some key architectures which can be applied for feature learning. Finally, there is a brief overview of previous research detection within the agricultural domain that can enable smart spraying systems.

2.1. Object Detection and Segmentation

Detection and segmentation are two fundamental computer vision techniques which have been applied to many image classification tasks. Object detection is the task of classifying objects in an image, and the prediction is returned as bounding box with a class label. Segmentation refers to the process of labelling each pixel within an image as an object class, providing fine grain detail of objects classes within an image. The difference between these approaches is that semantic segmentation models have no understanding

of the number of objects in an image, for example, if there are many people in an image, whereas instances also can identify the individual people.

Three popular architectures have emerged to provide object detection through the use of internal region proposal (RPN), these being Faster R-CNN [16], YOLO [17] and Single Shot Detection (SSD) [18]. Both SSD and YOLO are defined as single shot detection methods and differ from Faster R-CNN which can be thought of as a two shot method, as it uses an RPN and object classifier. Faster R-CNN incorporates an RPN into the Fast R-CNN architecture, and the RPN consists of three convolutional filters and take the features generated from an initial shared set of convolutional layers. The RPN learns to determine if a proposed region is a background or a foreground object. The idea of anchors for generating proposal boxes is applied where each pixel of the scaled image has an associated proposal window at different scales and ratios with the anchor at the centre. Following foreground/background classification, foreground objects of interest are then classified as a specific object type with an associated bounding box for the object. Their inception numerous modifications have been introduced to improve the speed and accuracy of the methods, including feature pyramid networks [19] and YOLOv4 [20]. The main drawback with bounding box based detection is that it is unable to provide the precision of segmentation which can be an issue in some applications requiring precise location information. This is specifically true for objects such as trees where the bounding box can contain many background pixels as object pixels.

Semantic segmentation has seen a number of techniques that have increased performance over recent years. The encoder–decoder architecture is one such contribution which consists of two parts. Initially, the encoder is applied on the input reducing the spatial dimension of feature maps, with the purpose of capturing longer range information. The decoder recovers the object detail and spatial dimension commonly by up-sampling the learning deconvolution or nearest neighbour interpolation which is computationally less intensive as it does not require learning. Major methods that use encoder–decoder include SegNet [21], which reuses the pooling indices from the encoder with in the decoder, while U-Net [22] adds skip connections from the encoder features to the corresponding decoder levels' activations. Atrous convolution layers were introduced in DeepLab [23], which provided a wider field of view at the same computational cost as regular convolutional layers and have provided increased accuracy in segmentation tasks.

Instance segmentation has applied techniques from both object detection and semantic segmentation. Specifically, the idea of region proposals has proven popular. Initial approaches were segmentation-first based and included DeepMask [24], which initially learns segment candidates and these proposals are then classified by Fast R-CNN [16]. The main issue with the segmentation-first based methods was that they were slow. Other instance segmentation methods such as InstanceCut [25] and Pixelwise Instance Segmentation [26] build upon the success of semantic segmentation. These methods aim to cut out the different objects based upon the class values and grouping of the semantic segmentation output. Mask R-CNN [13] was based upon an instance-first strategy, which applies a parallel prediction of segmentation masks and class labels. These parallel channels simultaneously address object classes, boxes and masks, making the system fast. Mask R-CNN introduces region-of-interest alignment to overcome systematic errors on overlapping instances and spurious edges exhibited in previous instance-first methods [27].

2.2. Convolutional Architectures

There has been two differing directions when developing convolutional architectures used for learning feature maps in deep learning. One direction has been the development of complex, deep and high parameters networks which aim to have high task accuracy at the expense of computational efficiency. The second direction is the creation of computationally efficient models that can be ran on mobile and embedded systems with an adequate level of accuracy for a given task. This section will mostly consider the second as they are more relevant to running embedded systems that can be deployed within agricultural settings.

ResNet [14] architectures have been highly successful for image classification and segmentation tasks. The capacity to develop very deep ResNet architectures is due to the use of shortcut connections. Shortcut connections allow the data signal to bypass one layer and move to the next layer in the sequence, permitting the gradients to flow from later layers to the early layers during back propagation in the training phase. However, ResNet architectures can be very large and have more compact variations using 18 and 50 layers, while sacrificing some accuracy provides higher levels at a smaller computational cost due the reduced number of parameters in the model.

Mobile-specific models that aim to be computationally efficient have been built on increasingly more efficient building blocks. Depthwise separable convolutions were initially introduced as an efficient replacement for traditional convolution layers in MobileNetV1 [28]. Depthwise separable convolutions separate the spatial filtering from the feature generation mechanism. Spatial filtering applies a light weight convolution while heavier 1×1 pointwise convolutions are used for feature generation. MobileNetV2 [29] introduced using linear bottleneck and inverted residual structure in order to make even more efficient layer structures by leveraging the low rank nature of the problem. Specifically using a 1×1 expansion convolution followed by depthwise convolutions and a 1×1 projection layer. The input and output are connected with a residual connection if and only if they have the same number of channels. This structure maintains a compact representation at the input and the output while expanding to a higher-dimensional feature space internally to increase the expressiveness of nonlinear per-channel transformations. Most recently, MobileNetV3 [15] mobiles have shown excellent performance increase over the previous versions in both object detection and segmentation tasks through the introduction of the NetAdapt algorithm [30] into the architecture. NetAdapt automatically adapts a pre-trained deep neural network to a mobile platform when given a resource budget. MobileNetV3 defines two specific architectures, MobileNetV3-Large and MobileNetv3-Small, with small reduced layers to provide a lower computational cost.

2.3. Alternative Architectures

Action recognition, object detection and tracking that resides in the spatio-temporal domain with features extracted from videos rather than single images have also been prominent areas for research as it address temporal changes from frame to frame. Recurrent Neural Networks (RNN) and Long Short-Term Memory (LSTM) networks have been shown to perform well in these areas. In Shu et al. [31], a novel Skeleton-joint Co-attention RNN was proposed to capture the spatial coherence among joints in the human skeleton with the aim of predicting human motion, this outperform-related method in this area. A Coherence Constrained Graph LSTM was proposed by Tang et al. [32] to address the group activity recognition problem by exploring human motion. This method applies a temporal confidence gate and a spatial confidence gate to control the memory state while an attention mechanism was employed to quantify the contribution of a certain motion. These spatio-temporal based methods could potentially be applied to other groups recognition such as tree and leaves in the agricultural domain. In object tracking, swarm intelligence and evolutionary algorithms have been shown to perform well. In Perez-Cham et al. [33], a hybrid algorithm combining swarm intelligence and evolutionary algorithm principles named Honeybee Search Algorithm was proposed which was inspired by the search for food of honeybees. A key element of this work was the capability to run the algorithm in parallel on the GPU accelerating the processing time.

A further issue in deep learning is the need for large amounts of training data; to address these issues, in [34], a weakly shared Deep Transfer Networks (DTNs) was presented. Its capability included transferring labelling information across heterogeneous domains, especially from text domain to image domain. This method can adequately mitigate the problem of insufficient image training data by bringing in rich labels from the text domain. Another challenge in computer vision is that of adversarial attacks where an input image is minimally manipulated causing the system to miss-classify objects. A Deep Genetic

Programming method called Brain Programming was introduced by Olague et al. [35] that is robust to adversarial attacks in comparison to the convolutional AlexNet model when compared on two artworks databases.

2.4. Agricultural Applications

Disease and weed detection has been an important research topic for sustainable agriculture, with numerous methods being applied to the problem. A hybrid method for lesion detection and classification on the citrus fruits and leaves was proposed by Sharif et al. [36]. This method constructed a codebook using texture, colour and geometric features and implemented feature selection based on PCA, skewness and entropy. A method for automated crop disease identification in leaves using Local Binary Patterns for feature extraction and One Class Classification for classification of the disease state was proposed [10]. A system for weed detection for smart spraying was proposed by Partel et al. [7]: This method used a YOLOv3 model to perform object detection to identify weeds from plants. The Plant Pathology Challenge 2020 [11] provided a dataset of apple disease images and a benchmark method for classification using a ResNet-50 network pre-trained on ImageNet dataset. Argüeso et al. [37] applied a Convolutional Neural Network using a few shot learning algorithm for plant leaf classification using deep learning with small datasets. A segmentation method for disease detection at the leaf scale using a colour features and region-growing was proposed in Jothiaruna et al. [38]. Disease of avocado plants was researched in Abdulridha et al. [9], using a multi-step approach including image acquisition from two cameras, image segmentation using region of interest and polygon region of interest, a multilayer perceptron for feature extraction and KNN classification. A multidimensional feature compensation residual neural network (MDFC–ResNet) model for fine-grained disease identification was proposed by Hu et al. [12]. The MDFC–ResNet identifies three dimensions, namely, species, coarse-grained disease and fine-grained disease and fuses multidimensional data to provide recognition results on the AI Challenger Crop Disease Identification dataset. For a more comprehensive overview of this area, readers are referred to the following survey papers [39,40].

3. Proposed System

The proposed system uses the Mask R-CNN method [13] to perform leaf and disease object detection and segmentation. Both ResNet-50 [14] and MobileNetV3 [15] networks are tested as the the feature extraction CNN backbone within Mask R-CNN to understand their specific suitability in for the task. In this section, an overview of the system architecture is provided, and a discussion regarding the annotation of segmentation masks for the dataset and system training parameters used is presented.

3.1. Mask R-CNN

Mask R-CNN [13] is a state-of-the-art instance segmentation method which extends the Faster R-CNN object detection method [16] by adding an object mask branch to the model. Mask R-CNN, therefore, has three distinctive outputs for a given object with these being the class label, object bounding box and object mask. An overview of Mask R-CNN is shown within Figure 2.

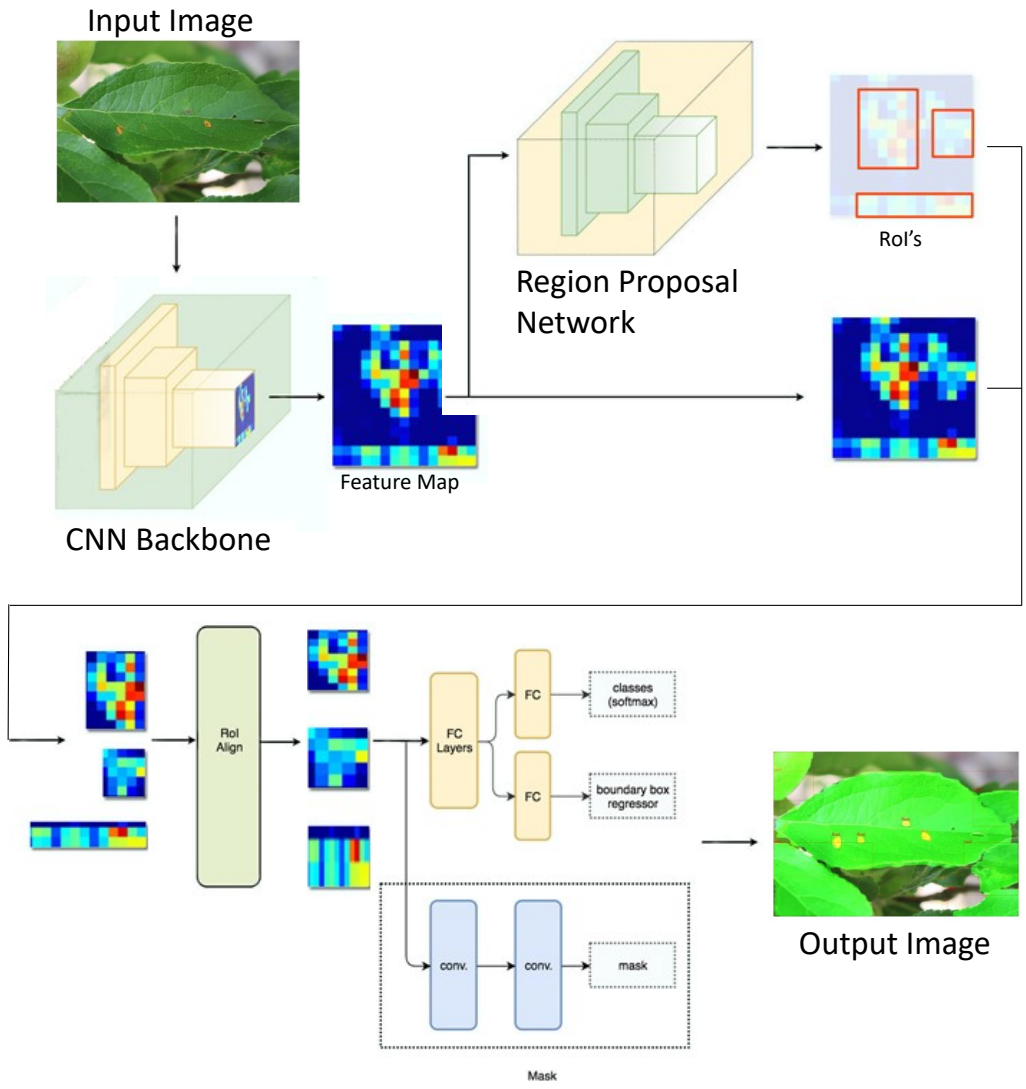

Figure 2. Proposed system architecture.

The functionality inherited from the Faster R-CNN network architecture is as follows: Firstly, a convolutional backbone is used to learn the features associated with the object detection task. Secondly, the region proposal network (RPN) layer learns n regions of interests of probable object locations within images. Finally, a region of interest pooling layer (RoIPool) and a set of fully connected layers are used to extracts features using from each candidate region of interest and performs classification and bounding-box regression while collapsing RoIPool features into short output vectors of class label and bounding box offset coordinates. RoIPool was introduced in Faster R-CNN to create uniform size feature maps from RoI of different dimensions, which is introduced as objects that are not all a uniform size. RoIPool introduces some misalignments between the RoI and the extracted features, which is problematic for accuracy when predicting object masks due to hard quantisations. The RoIAlign layer is, therefore, introduced in Mask R-CNN to address this issue, which instead uses bilinear interpolation to compute the exact values of the input features at four regularly sampled locations in each RoI bin and aggregates the result by using max or average values. This method provided a large improvement for mask prediction without any negative effects relative to object detection.

Mask R-CNN also added a new branch for segmentation mask predictions, rather than fully connected layers as used for the other tasks. In order to extract pixel-to-pixel correspondence for RoI segmentation, a fully convolution network (FCN) [41] is applied. This requires fewer parameters and is more accurate than fully connected layers for this

task. Binary masks for each RoI are predicted. For a single RoI, the mask predicts the pixels that constitute the object and those that do not. This approach to mask prediction and class prediction means that the generation of masks can be learnt for every class without competition among classes. Traditionally, semantic segmentation uses FCN's in which the mask and class are intrinsically linked. Multi-task loss is used to train the Mask R-CNN for each of the three tasks, and the total loss of the network is described as follows:

$$loss_{total} = loss_{cls} + loss_{box} + loss_{mask} \tag{1}$$

where $loss_{total}$ is the total network loss, comprising the sum of the classification task loss as $loss_{cls}$, the bounding box regression task loss as $loss_{box}$ and the mask prediction task loss as $loss_{mask}$. Classification loss $loss_{cls}$ is a softmax loss function given by the following.

$$loss_{cls} = \frac{1}{N_{cls}} \sum_{i=n} -(1 - p_i^*) \cdot log(1 - p_i) - p_i^* \cdot log(p_i) \tag{2}$$

It is used for learning an object (leaf, rust, etc.) ($p_i = 1$) and a non-object ($p_i = 0$), where p_i^* is the ground truth class label and p_i the predicted class for the ith anchor, respectively. This loss function is normalised by N_{cls}, which is the training mini-batch size. Bounding box regression loss $loss_{box}$ is defined as follows:

$$loss_{box} = \frac{1}{N_{box}} \sum_{i=n} p_i^* smooth_{L1}(t_i - t_i^*) \tag{3}$$

where for the ith anchor, the $L1$ loss between the ground-truth box t_i^* and the predicted bounding box t_i is calculated. Both t_i^* and t_i are vectors representing the four parameterised coordinates of the predicted bounding box. Only positive anchors affect the loss as described by the term $p_i^* smooth_{L1}$. N_{box} is the total number of anchors that normalises the loss function. Mask prediction loss $loss_{mask}$ is defined as the average binary cross-entropy loss as follows:

$$loss_{mask} = -\frac{1}{m \times n} \sum_{i=n} y_{ij} \log \hat{y}_{ij}^k + (1 - y_{ij}) \log(1 - \hat{y}_{ij}^k) \tag{4}$$

where y_{ij} is the label of pixel (i, j) in the ground truth mask for an $m \times n$ region. $\hat{y}_{ij}^k$ is the prediction for the same pixel location for the ground-truth class k. Region size is applied to normalise this.

3.2. Feature Extraction Backbone and Feature Pyramid Network

The Mask R-CNN architecture provides a flexible approach to the CNN based backbone used for feature extraction. In this paper, three backbones are evaluated for the tasks, these being ResNet-50 [14] and MobileNetV3-Large and MobileNetV3-Large-Mobile [15]. These backbones have been selected due to their different properties, with ResNet-50 being the largest in terms of parameters and the slowest for inference, and MobileNetV3-Large-Mobile is correspondingly the smallest and fastest. None of these backbones are very large; thus, they can also potentially be deployed on many types of devices including embedded systems.

Each backbone network in Mask R-CNN also applies a feature pyramid network (FPN) [19]. Detecting objects in different scales is challenging, specifically when objects are small. FPN is a set of layers that replaces the original feature extractor RPN of Faster R-CNN in Mask R-CNN. The FPN generates multiple feature map layers, which have been shown to improve feature quality compared with a regular feature pyramids for object detection.

3.3. Plant Data Set and Model Training

The publicly available Plant Pathology Challenge 2020 data set [11] provides images and disease labelling used in the training and evaluation of the proposed method. The dataset consists of 3462 images of apple tree leaves and fruit, split evenly into 1821 testing and training images with an associated class of diseased or not plants. As the dataset currently had no ground truth segmentation masks, manual annotation was undertaken by the authors to evaluate the use of instance segmentation for this application. A subset of challenging images from the dataset was manually annotated for the task of segmentation: In total, 142 images where annotated, 101 for usage in training and 41 images for evaluation. Figure 3 shows examples of labelling added to the images by the authors. Due to the depth of field in the images, only leaves and apples with significant details were considered for labelling, and blurry objects were not labelled. In total, four classes were used in the segmentation annotation; these were leaf, apple, rust and background.

Figure 3. Segmentation map annotation examples created for this paper.

Training of the of the proposed system was conducted on a desktop PC with an Nvidia RTX Titan GPU and AMD Ryzen 7-3700X CPU with 32GB of RAM. The PyTorch framework was used, specifically using the Mask R-CNN model and ResNet-50 and MobileNetV3 models pre-trained on the Coco data set [42] available via the TorchVision 0.10 library. Training for each model was ran for 20 epochs using the SGD optimiser with a learning rate of 0.005, momentum of 0.9 and weight decay 0.0005. The only data augmentation method applied was the subset of images from the Plant Pathology Challenge 2020, which was random horizontal flipping.

4. Results

To evaluate the capability of the proposed system for the tasks of leaf and rust disease identification, experiments were conducted for object detection and segmentation precision and disease detection. This also establishes some benchmark results on the new set of segmentation annotations for the Plant Pathology Challenge 2020 dataset. Three different feature extraction backbones to Mask R-CNN were trained as described in Section and evaluated, and these were ResNet-50, Large MobileNetV3-Large and MobileNetV3-Large-320 (a mobile optimised MobileNetV3-Large model provided in PyTorch), respectively. The results of these evaluations are discussed in this section.

4.1. Object Detection and Instance Segmentation Evaluation

A key aspect of the proposed system is that it is both capable of detecting objects to a good level and can generate an accurate segmentation mask for the found object. The metrics for object detection and segmentation were generated using widely used Coco tools evaluation tool [42]. Average Precision (AP) is a commonly applied for object detection task and is adapted here at two different Intersection over Union (IoU), where an IoU of 0.5 is a standard indicator of model accuracy, while 0.75 highlights more precise detection and segmentation capabilities of a model.

Tables 1 and 2 show the AP results for object detection and segmentation respectively. The Mask R-CNN model with the ResNet-50 backbone performs best in all tasks, specifically in segmentation evaluation. To visually inspect the strengths and weaknesses of each backbone, Figure 4 provides examples of the output. It is clear that the ResNet backbone is the only model that handles small object detection well, which is one of the more difficult challenges. ResNet-50 shows great potential for detecting the rust on leaves, as the largest network in terms of depth it handles spatial features at lower resolutions, whereas MobileNetV3-Large struggles with very small areas of rust and MobileNetV3-Large-Mobile loses the fine grain detail needed to effectively learn and detect these small size features through a reduction in image resolution to improve its speed. All models show the capability to detect leaves. While segmentation results are good, the difficulty in determining leaf edges could be improved in all models, although background leaves and blurred edges in the images make this very challenging. Table 3 shows the inference speed of the Mask R-CNN system with each backbone, ResNet-50 is the slowest of the models only 0.04 seconds slower than MobileNetV3-Large. MobileNetV3-Large-Mobile is three times faster but has significant issues with detection accuracy of very small objects as a result of speed optimisation.

Table 1. Object detection results—average precision (AP); intersection over Union (IoU).

Backbone	AP @ IoU = 0.50	AP @ IoU = 0.75
ResNet-50	0.486	0.282
MobileNetV3-Large	0.369	0.273
MobileNetV3-Large-Mobile	0.277	0.158

Table 2. Segmentation results—average precision (AP); intersection over Union (IoU).

Backbone	AP @ IoU = 0.50	AP @ IoU = 0.75
ResNet-50	0.489	0.368
MobileNetV3-Large	0.351	0.173
MobileNetV3-Large-Mobile	0.264	0.127

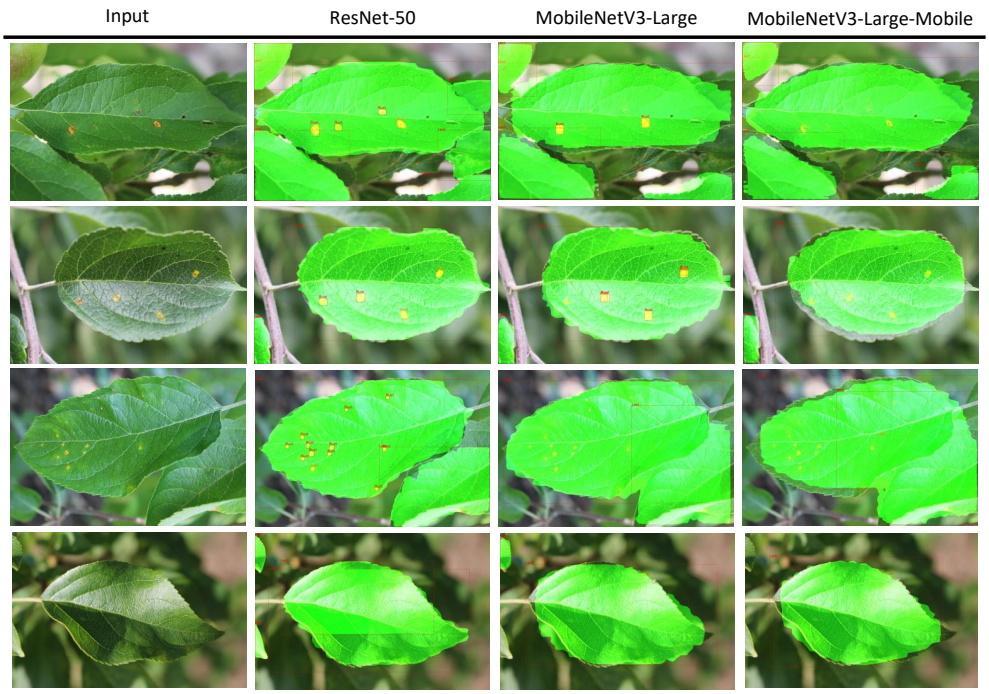

Figure 4. Examples of segmentation results from the proposed system.

Table 3. Inference time.

Backbone	Time (s.)
ResNet-50	0.26
MobileNetV3-Large	0.22
MobileNetV3-Large-Mobile	0.07

4.2. Disease Detection

This evaluation of the model looks at how well the proposed system can determine leaves with rust against healthy leaves. The Plant Pathology Challenge 2020 dataset [11] provides a set of disease (rust) and healthy labels for each image in the training sets. These ground truth labels are used for them in the subset of data used to train and evaluate this system. Any object detection of the class rust denotes a diseased leaf, while the presence of a detected leaf but no rust is determined to be healthy.

Table 4 shows the results of this evaluation in terms of detection accuracy. Only Mask R-CNN with a ResNet-50 backbone has a significantly high accuracy of 80.5%. This is due to the model's capability to focus on small areas which other backbone models do not handle well, as discussed in the previous section. Figure 5 shows the confusion matrix, showing that the main area in which errors lie is the misclassification of healthy leaves as having rust. Analysis of the output shows a number of reasons for these errors. Firstly, as shown in Figure 6, some leaves look to potentially have some rust; although ground truth labelling does not detail rust, this could be down to labelling errors, or these marks may not be rust at all but it is hard even for a human to tell exactly what this small mark is in the images. One other limitation is that some leaves have markings similar to rust, and these are not labelled and are not common in the training set on healthy leaves. These errors have more to do with current limitations in the dataset and the segmentation annotation specifically, which could be rectified in future research with a larger and more varied training set of data and labels.

Table 4. Disease detection accuracy.

Backbone	Accuracy
ResNet-50	80.5%
MobileNetV3-Large	68.3%
MobileNetV3-Large-Mobile	53.7%

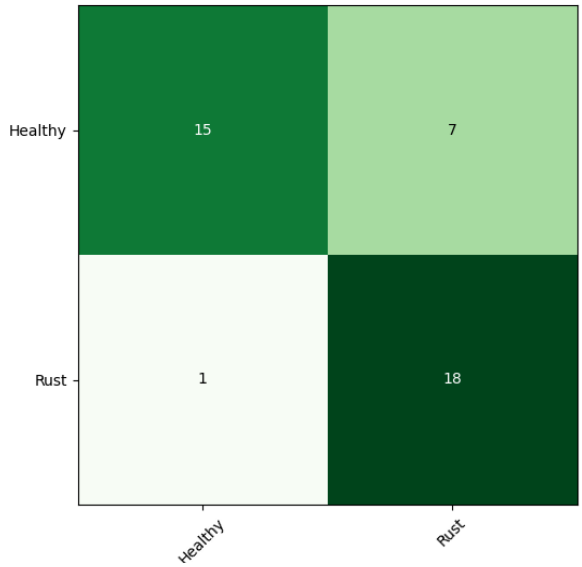

Figure 5. Disease detection confusion matrix—ResNet-50.

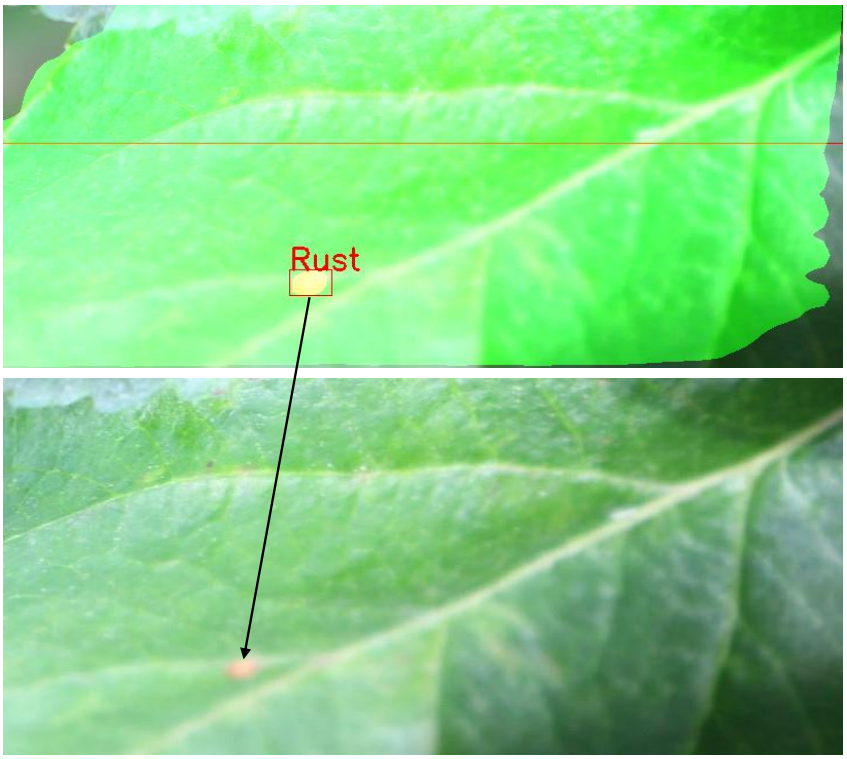

Figure 6. Disease detection—questionable errors.

5. Discussion

In this paper, a study presenting a method for the precise detection of leaves and rust disease was proposed and evaluated. The proposed system used a Mask R-CNN model while testing three different backbones for their performance: These were ResNet-50 and MobileNetV3-Large and MobileNetV3-Large-Mobile. The study results presented show promise especially when using the ResNet-50 backbone, which has the capability to find and segment very small objects such as rust on leaves accurately. On the other hand, the other backbones work well for leaf detection and segmentation with MobileNetV3-Large-Mobile well suited to low-powered CPU-based devices such as smart-phones, making it a cost-effective method to potentially deploy. The labelling with segmentation information also provide a foundation to further explore these models in future work, with the opportunity to continue labelling to create a larger dataset.

The study highlights the potential for instance segmentation in precision agricultural practice and smart spraying systems. One key improvement over object detection methods alone is that rather than a binary distinction of healthy leaf vs. rust-infected leaf, the capacity for precision is present. For example, this includes the ability to identify different leaves and also the rust that is present on that specific leaf. This can be obtained with some simple post-processing of the segmentation masks and the overlapping location of rust and leaf. These data could identify the leaf size and then the area covered by the rust, influencing the amount of fungicide applied for more sustainable agricultural practices.

This research provides a foundation for further studies. In particular, it provides information on the annotated segmentation maps for a subset of the Plant Pathology Challenge 2020 dataset [11], which can be made available to the research community along with the benchmark results produced.

Author Contributions: Conceptualization, G.S.; methodology, G.S.; software, G.S.; validation, G.S.; formal analysis, G.S.; investigation, G.S.; data curation, G.S.; writing—original draft preparation, G.S.; writing—review and editing, Q.M. and B.L.; visualization, G.S.; funding acquisition, Q.M. and B.L. All authors have read and agreed to the published version of the manuscript.

Funding: Supported by Innovate UK (grant number 104016): AgriRobot—Autonomous Agricultural Robot System for Precision Spraying.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Publicly available datasets were analyzed in this study. This data can be found here: https://www.kaggle.com/c/plant-pathology-2020-fgvc7.

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

References

1. Linhart, C.; Niedrist, G.H.; Nagler, M.; Nagrani, R.; Temml, V.; Bardelli, T.; Wilhalm, T.; Riedl, A.; Zaller, J.G.; Clausing, P.; et al. Pesticide contamination and associated risk factors at public playgrounds near intensively managed apple and wine orchards. *Environ. Sci. Eur.* **2019**, *31*, 28. [CrossRef]
2. Simon, S.; Brun, L.; Guinaudeau, J.; Sauphanor, B. Pesticide use in current and innovative apple orchard systems. *Agron. Sustain. Dev.* **2011**, *31*, 541–555. [CrossRef]
3. Creech, C.F.; Henry, R.S.; Werle, R.; Sandell, L.D.; Hewitt, A.J.; Kruger, G.R. Performance of Postemergence Herbicides Applied at Different Carrier Volume Rates. *Weed Technol.* **2015**, *29*, 611–624. [CrossRef]
4. Aktar, W.; Sengupta, D.; Chowdhury, A. Impact of pesticides use in agriculture: Their benefits and hazards. *Interdiscip. Toxicol.* **2009**, *2*, 1–12. [CrossRef] [PubMed]
5. Lefebvre, M.; Langrell, S.R.; Gomez-y-Paloma, S. Incentives and policies for integrated pest management in Europe: A review. *Agron. Sustain. Dev.* **2015**, *35*, 27–45. [CrossRef]
6. Dara, S.K. The New Integrated Pest Management Paradigm for the Modern Age. *J. Integr. Pest Manag.* **2019**, *10*, 1–9. [CrossRef]
7. Partel, V.; Kakarla, S.C.; Ampatzidis, Y. Development and evaluation of a low-cost and smart technology for precision weed management utilizing artificial intelligence. *Comput. Electron. Agric.* **2019**, *157*, 339–350. [CrossRef]
8. Ampatzidis, Y. Applications of Artificial Intelligence for Precision Agriculture. *EDIS* **2018**, *2018*, 1–5. [CrossRef]
9. Abdulridha, J.; Ehsani, R.; Abd-Elrahman, A.; Ampatzidis, Y. A remote sensing technique for detecting laurel wilt disease in avocado in presence of other biotic and abiotic stresses. *Comput. Electron. Agric.* **2019**, *156*, 549–557. [CrossRef]
10. Pantazi, X.E.; Moshou, D.; Tamouridou, A.A. Automated leaf disease detection in different crop species through image features analysis and One Class Classifiers. *Comput. Electron. Agric.* **2019**, *156*, 96–104. [CrossRef]
11. Thapa, R.; Zhang, K.; Snavely, N.; Belongie, S.; Khan, A. The Plant Pathology Challenge 2020 data set to classify foliar disease of apples. *Appl. Plant Sci.* **2020**, *8*, e11390. [CrossRef] [PubMed]
12. Hu, W.J.; Fan, J.; Du, Y.X.; Li, B.S.; Xiong, N.; Bekkering, E. MDFC-ResNet: An Agricultural IoT System to Accurately Recognize Crop Diseases. *IEEE Access* **2020**, *8*, 115287–115298. [CrossRef]
13. He, K.; Gkioxari, G.; Dollar, P.; Girshick, R. *Mask R-CNN*. In Proceedings of the IEEE International Conference on Computer Vision (ICCV 2017), Venice, Italy, 22–29 October 2017; pp. 2980–2988.
14. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2016), Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
15. Howard, A.; Sandler, M.; Chen, B.; Wang, W.; Chen, L.C.; Tan, M.; Chu, G.; Vasudevan, V.; Zhu, Y.; Pang, R.; et al. Searching for mobileNetV3. In Proceedings of the IEEE International Conference on Computer Vision (ICCV 2019), Seoul, Korea, 27 October–2 November 2019; pp. 1314–1324.
16. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2015**, *39*, 1137–1149. [CrossRef] [PubMed]
17. Redmon, J.; Farhadi, A. YOLO9000: Better, Faster, Stronger 2016. Available online: https://arxiv.org/abs/1612.08242 (accessed on 24 November 2021).
18. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.Y.; Berg, A.C. *SSD: Single Shot MultiBox Detector*; Springer: Cham, Switzerland, 2016; pp. 21–37. [CrossRef]
19. Lin, T.Y.; Dollár, P.; Girshick, R.; He, K.; Hariharan, B.; Belongie, S. *Feature Pyramid Networks for Object Detection*. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2017), Honolulu, HI, USA, 21–26 July 2017; pp. 936–944.
20. Bochkovskiy, A.; Wang, C.Y.; Liao, H.Y.M. YOLOv4: Optimal Speed and Accuracy of Object Detection. 2020. Available online: https://arxiv.org/abs/2004.10934 (accessed on 21 November 2021).
21. Badrinarayanan, V.; Kendall, A.; Cipolla, R. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 2481–2495. [CrossRef] [PubMed]
22. Ronneberger, O.; Fischer, P.; Brox, T. *U-net: Convolutional Networks for Biomedical Image Segmentation*; Springer: Berlin/Heidelberg, Germany, 2015; Volume 9351, pp. 234–241. [CrossRef]
23. Chen, L.C.; Papandreou, G.; Kokkinos, I.; Murphy, K.; Yuille, A.L. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. *IEEE Trans. Pattern Anal. Mach. Intell.* **2018**, *40*, 834–848. [CrossRef]

24. Pinheiro, P.O.; Lin, T.Y.; Collobert, R.; Dollár, P. *Learning to Refine Object Segments*; Springer: Berlin/Heidelberg, Germany, 2016; Volume 9905, pp. 75–91. [CrossRef]
25. Kirillov, A.; Levinkov, E.; Andres, B.; Savchynskyy, B.; Rother, C. InstanceCut: From Edges to Instances with MultiCut. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2017), Honolulu, HI, USA, 21–26 July 2017; pp. 7322–7331.
26. Arnab, A.; Torr, P.H. Pixelwise Instance Segmentation with a Dynamically Instantiated Network. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2017), Honolulu, HI, USA, 21–26 July 2017; pp. 879–888.
27. Li, Y.; Qi, H.; Dai, J.; Ji, X.; Wei, Y. Fully Convolutional Instance-Aware Semantic Segmentation. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2017), Honolulu, HI, USA, 21–26 July 2017; pp. 4438–4446.
28. Howard, A.G.; Zhu, M.; Chen, B.; Kalenichenko, D.; Wang, W.; Weyand, T.; Andreetto, M.; Adam, H. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. 2017. Available online: https://arxiv.org/abs/1704.04861 (accessed on 24 November 2021).
29. Sandler, M.; Howard, A.; Zhu, M.; Zhmoginov, A.; Chen, L.C. MobileNetV2: Inverted Residuals and Linear Bottlenecks. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 4510–4520.
30. Yang, T.J.; Howard, A.; Chen, B.; Zhang, X.; Go, A.; Sandler, M.; Sze, V.; Adam, H. NetAdapt: Platform-Aware Neural Network Adaptation for Mobile Applications. In *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*; Springer: Berlin/Heidelberg, Germany, 2018; Volume 11214, pp. 289–304.
31. Shu, X.; Zhang, L.; Qi, G.J.; Liu, W.; Tang, J. Spatiotemporal Co-attention Recurrent Neural Networks for Human-Skeleton Motion Prediction. *IEEE Trans. Pattern Anal. Mach. Intell.* **2021**. [CrossRef]
32. Tang, J.; Shu, X.; Yan, R.; Zhang, L. Coherence Constrained Graph LSTM for Group Activity Recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2019**, 44, 636–647. [CrossRef]
33. Perez-Cham, O.E.; Puente, C.; Soubervielle-Montalvo, C.; Olague, G.; Aguirre-Salado, C.A.; Nuñez-Varela, A.S. Parallelization of the Honeybee Search Algorithm for Object Tracking. *Appl. Sci.* **2020**, *10*, 2122. [CrossRef]
34. Shu, X.; Qi, G.J.; Tang, J.; Wang, J. Weakly-Shared deep transfer networks for heterogeneous-domain knowledge propagation. In Proceedings of the MM 2015—Proceedings of the 2015 ACM Multimedia Conference, Brisbane, Australia, 26–30 October 2015; pp. 35–44. [CrossRef]
35. Olague, G.; Ibarra-Vázquez, G.; Chan-Ley, M.; Puente, C.; Soubervielle-Montalvo, C.; Martinez, A. A Deep Genetic Programming Based Methodology for Art Media Classification Robust to Adversarial Perturbations. In *International Symposium on Visual Computing*; Springer: Berlin/Heidelberg, Germany, 2020; Volume 12509, pp. 68–79. [CrossRef]
36. Sharif, M.; Khan, M.A.; Iqbal, Z.; Azam, M.F.; Lali, M.I.U.; Javed, M.Y. Detection and classification of citrus diseases in agriculture based on optimized weighted segmentation and feature selection. *Comput. Electron. Agric.* **2018**, *150*, 220–234. [CrossRef]
37. Argüeso, D.; Picon, A.; Irusta, U.; Medela, A.; San-Emeterio, M.G.; Bereciartua, A.; Alvarez-Gila, A. Few-Shot Learning approach for plant disease classification using images taken in the field. *Comput. Electron. Agric.* **2020**, *175*, 105542. [CrossRef]
38. Jothiaruna, N.; Sundar, K.J.A.; Karthikeyan, B. A segmentation method for disease spot images incorporating chrominance in Comprehensive Color Feature and Region Growing. *Comput. Electron. Agric.* **2019**, *165*, 104934. [CrossRef]
39. Loey, M.; ElSawy, A.; Afify, M. Deep learning in plant diseases detection for agricultural crops: A survey. *Int. J. Serv. Sci. Manag. Eng. Technol.* **2020**, *11*, 41–58. [CrossRef]
40. Sandhu, G.K.; Kaur, R. *Plant Disease Detection Techniques: A Review*. In Proceedings of the IEEE International Conference on Automation, Computational and Technology Management (ICACTM 2019), London, UK, 24–26 April 2019; pp. 34–38.
41. Dai, J.; Li, Y.; He, K.; Sun, J. *R-FCN: Object Detection via Region-Based Fully Convolutional Networks*; Curran Associates Inc.: Red Hook, NY, USA, 2016; pp. 379–387.
42. Lin, T.Y.; Maire, M.; Belongie, S.; Hays, J.; Perona, P.; Ramanan, D.; Dollár, P.; Zitnick, C.L. *Microsoft COCO: Common Objects in Context*; Springer: Berlin/Heidelberg, Germany, 2014; Volume 8693, pp. 740–755. [CrossRef]