Computers and Electronics in Agriculture 188 (2021) 106263

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

journal homepage: www.elsevier.com/locate/compag

Original papers

Towards selective and automatic harvesting of broccoli for agri-food industry

Antonio García-Manso *, Ramón Gallardo-Caballero, Carlos J. García-Orellana, Horacio M. González-Velasco, Miguel Macías-Macías

Instituto de Computación Científica Avanzada (ICCAEx), Universidad de Extremadura, E-06006, Spain

ARTICLE INFO

Keywords:
Deep learning
Object detection
Broccoli

ABSTRACT

Broccoli is a vegetable grown worldwide due to its good nutritional properties. The harvest of this product is done selectively by hand depending on their size and state of maturation for both fresh market and agri-food industry. The final aim of our work is the development of a machine that is able to automatically harvest only those broccoli heads that have the size and ripeness suitable for the agri-food industry, besides discarding those overripe or with diseases. One critical element in such a machine is a vision system that locates and classifies the broccoli heads present in photographic images, to trigger later a cutting module. In this paper, we present an approach to that vision system, based on deep learning techniques. The proposed algorithm, running in a relatively cheap hardware, is able to work in real time, locating broccoli heads in 640 × 480 px digital images, and classifying then into *harvestable*, *immature* and *wasted* classes. Tested with images taken in real conditions, with many heads partially hidden by leaves, the system was able to correctly locate and classify up to 97% of the cases presented in the test set.

1. Introduction

Broccoli (*Brassica oleraceae va. Cymosa*), is a plant of the family Brasicaceae or Cruciferae. This plant has abundant green fleshy flower heads, arranged in the shape of a tree, with branches that are born from a thick edible stem. The great mass of heads is surrounded by leaves. It belongs to the group of cabbages together with cauliflower, cabbage and red cabbage. In mid-latitudes, its harvesting season and best time to eat it is from September to June. It has good nutritional properties such as the contribution of vitamins A, C and niacin; minerals such as potassium, calcium, sodium and magnesium and fibre. For all these reasons, over 26 million tons are produced worldwide (data from FAOSTAT 2017 FAO (2019)). The largest producers are China, India, the United States, Spain, Mexico and Italy.

Traditionally, all production of broccoli, both for the fresh market, where products reach consumers directly, and for agri-food industry, where products reach consumers once processed, is harvested by hand Mullaney and Weinroth (2019). The harvest must be selective due to the different stages of maturity that can be found within the same plantation. In our region the agricultural workers specialized in this task of harvesting only cut the heads of broccoli that have an optimal maturity stage (i.e. neither overripe nor underripe), no defects or diseases, and a size that is the maximum possible with the two previous conditions, and always larger than a size requested by the industry they are destined to.

Generally, the optimal size for harvesting a head of broccoli is around 16 cm in diameter (Fig. 1(a)). However, the objective of farmers will always be to maximize their crop yield. Therefore, the broccoli should be allowed to grow as much as possible (Fig. 1(b)) before cutting, whenever possible due to its maturity stage. This means that each head of broccoli has to be considered for harvesting not only depending on its size, but also on its state of maturity, before it reaches a stage where it is no longer harvestable (Fig. 1(c)).

In Fig. 1(a) it can be seen that, despite the broccoli head size, the grains are very small, and the head completely compact and hermetic. Therefore, its harvest at that moment is not advisable since it can grow more and reach larger size and, hence, higher weight. In Fig. 1(b) we show the broccoli in its optimum state for harvesting. The head should be semi-spherical, compact, smooth, with compact, homogeneous, green to violet florets. The flowers must be small, completely closed, without protruding individually in the florets. It can be seen that the size of the

* Corresponding author.
E-mail addresses: agmanso@unex.es (A. García-Manso), rgallardo@unex.es (R. Gallardo-Caballero), cjgarcia@unex.es (C.J. García-Orellana), hmgvelas@unex.es (H.M. González-Velasco), mmacias@unex.es (M. Macías-Macías).

https://doi.org/10.1016/j.compag.2021.106263
Received 26 November 2020; Received in revised form 9 April 2021; Accepted 8 June 2021
Available online 23 July 2021

grain that will give rise to the flowers is larger than in Fig. 1(a). Finally, in Fig. 1(c) a broccoli head is observed with a diameter bigger than the two previous ones. Now the top part is no longer compact and airtight, an the texture of the grains has also changed. All this indicates that its state of maturity has plate-ned and in a short time the flowers will break out. In other words, it has been left too long to be harvested. In Fig. 1(c) can also be observed the characteristic spots due to the fungus botrytis, which makes the head not suitable for consumption, and therefore it should be discarded.

In order to maximize the crop yield, plantations are harvested at least once a week, subject to weather conditions, and two or three successive rounds are performed in each session. Therefore, a very high workload and personnel cost are involved in a plantation of this type, being approximately 12% of production costs Anejo no 20 (2020).

Due to its large production and worldwide spread, there are currently one commercially available broccoli harvester by RoboVeg RoboVeg (2021), and one patented prototype by MYCOM MYCOM (2021). RoboVeg company has developed a machine that must be attached to a tractor, and uses a vision system to detect the broccoli heads and a robotic arm to cut them. On the other hand, the prototype by MYCOM has autonomous movement, and the broccoli heads detected by the machine are cut by a fixed blade. Though both machines are reported to perform very well, with 2-3 s to locate and harvest the broccoli heads, given their characteristics they are mainly intended for large broccoli growers, and very few medium or small growers can afford one of them. In this context, our global aim of this research line is the development of a simple and low-cost prototype of an automatic and selective harvester for broccoli, which could be affordable for small plantations.

In these prototypes, a critical part is always the system that analyses images in order to locate and classify (i.e., detect) the broccoli heads, thus replacing the human visual perception. Several approaches to automatic localization of broccoli within images can be found in literature. The first study Ramirez (2006) used texture-based analysis in RGB images, but it was very simple (only considered 13 images) and its results are not very significant. More interesting is Blok et al. (2016), where a method based on texture filters is proposed, analysing 228 broccoli heads. Though it presented a high precision (99,5%), the recall was not so high due to generalization problems of the algorithm. Later, in Blok et al. (2021), the same authors introduced deep learning techniques in their location and segmentation algorithm, achieving a better image generalization than in previous studies. They also report the development of a cutting robot, but few details are given.

Most of the studies include machine learning in the techniques used to locate broccoli heads. In Kusumam et al. (2016), Kusumam et al. (2017) they also address the localization of broccoli in the field, this time using 3D vision (with Kinect-2), a Support Vector Machine classifier and a temporal filter, to provide the 3D localization of the heads with up to 95% of precision. In Le Louedec et al. (2020), another system for localization of broccoli heads is presented, which is based on 3D information obtained from RGBD sensors. In their technique they trained a CNN auto-encoder for the task of semantic segmentation using the 3D information, outperforming the results described by Kusumam et al. (2016), and reporting very high inference speeds, making the technique suitable for real time applications. In Bender et al. (2020) they presented a system to locate plants (not heads) of broccoli and

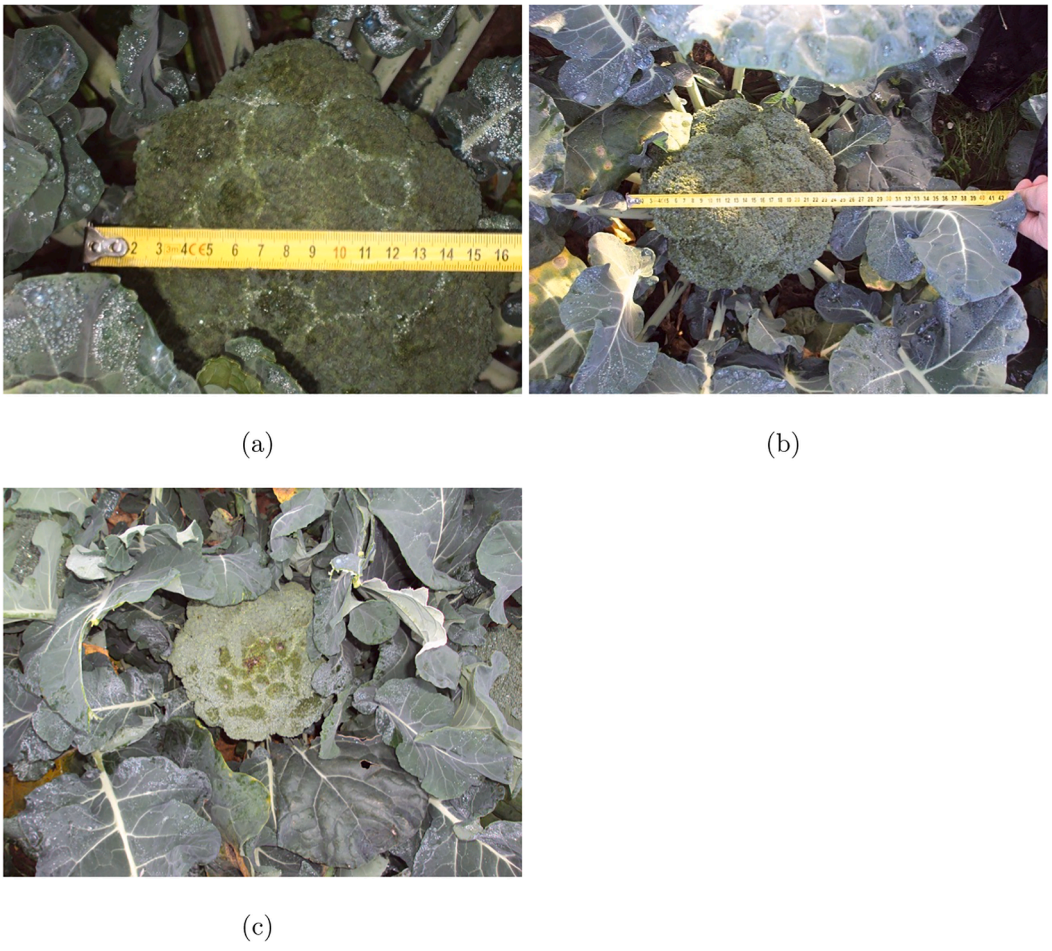

Fig. 1. 1(a) Optimal broccoli for harvesting considering only the size. 1(b) Optimal broccoli for harvesting by size and state of maturity. 1(c) Broccoli not optimal for harvest due to disease (botrytis fungus).

cauliflower with Faster Region-Based convolutional neural network, and reported 95% mean average precision (mAP). However, they did not locate the head, essential for harvesting. Finally, in Zhou et al. (2020), another system for the segmentation and localization of broccoli heads is presented, based in other type of deep convolutional neural network. In this case, the segmented images are used later to estimate the quality of the broccoli head, considering the yellowness observed.

Unfortunately, all the studies described above only approach the localization of broccoli heads, and only in two of them they evaluate their condition of *harvestable*, exclusively based on the size. However, it is very important for a selective harvester to determine if the localized head is *harvestable* (and we have to cut it), *immature* (and we let it grow, so it will be considered in the following rounds of harvesting) or *wasted* (because it is overripe or have defects or diseases, and we can cut and discard it). This way, another classification step is needed.

However, given the high performance obtained by deep neural networks in the localization of broccoli heads, and the ability of this networks to classify the proposals into predefined classes Koirala et al. (2019), we hypothesised that the classification can be made along with the localization, in one unique step, with one of those networks. This way, and taking into account the global aim of our research described above, the objective of this work is to test this hypothesis through the development of a robust vision system that, considering 2D images taken in the field without controlled illumination conditions:

- Locates and classifies (i.e. detects) the broccoli heads, determining if they are harvestable or not.
- Can be implemented in a low-cost and low-power hardware, and is able to work in real time. This way, it can be incorporated our prototype with a minimum cost.
- Performs well in the less restrictive case, i.e., in the harvesting of broccoli for agri-food industry.

Therefore, what this work apports is a new technique that locates and classifies the broccoli heads in a single step, and which can be used in real time with a relatively cheap hardware, providing methodology to implement these network models, once trained, on NVIDIA Jetson Nano NVIDIA (2019), as will be discussed in the following section, where the materials and methods are described.

2. Materials and methods

2.1. Image dataset

Taking into account our objectives described above, in this study we considered images of the broccoli varieties used for commercial industry (Parthenon, Monaco, Marathon, …). These types of broccoli go to factories where the florets that make up the head are separated and then deep-frozen. For broccoli going to the fresh market, other parameters for optimal maturity are considered, and therefore this study is not suitable, though it could be adapted.

The images used to train and test our algorithm were captured with non-professional cameras in plantations around Badajoz, Spain. These images were taken in the field, with non-controlled natural illumination. Due to this, the heads of broccoli in the images are often partially hidden by leaves, or covered with water droplets of dew, as can be seen in Fig. 2 (a) and (b). In total, 6139 images were captured with broccoli heads, distributed among images of broccoli considered harvestable (optimal state of maturity, decided by an expert harvester), immature (they can grow more) and wasted (not suitable for consumption due defects, diseases such as botrytis fungus, or excesive ripeness).

Trying to obtain a more robust visual system, two kinds of images were captured in the process. Firstly, some of the images where taken by hand using two different cameras: Sony-DSC-S950, 1/2.3 inch CCD sensor, with a resolution of 2592×1944 px (618 images) and Nikon-E4600, 1/2.5 inch CCD sensor, with a resolution of 2288×1712 px (57 images), both with autofocus. This images were taken at a fixed distance from the ground, in several sessions. On the other hand, more images were taken using a Sony IMX179 image sensor camera (1/3.2 inch) placed on a tractor, automatically capturing over 10 images per second while the tractor was moving at a speed of approximately 1 km/h (5464 images), simulating a possible real scenario, in which we need to process the images at this rate.

Therefore, in this study we considered a total of 6165 images of broccoli heads, taken from the 6139 photographs, of which 2778 were considered as harvestable (Figs. 1(b) and 2(b)), 2805 as immature (Figs. 1(a) and 2(a)) and 582 as wasted. The images were randomly separated into learning and test sets as shown in Table 1.

Once taken, the images were marked (correct location of broccoli heads) and labelled (correct classification: harvestable, immature, wasted) by an expert farmer of our region, using an application developed for the task, which stores data in a database.

Table 1
Composition of the training sets.

	Harvestable	*Immature*	*Wasted*	
Learning (L)	2233	2247	451	4931
Test (T)	545	558	131	1234
Total	2778	2805	582	6165

(a) (b)

Fig. 2. 2(a) Image of a *immature* head of broccoli originally captured at a size of 2592×1944 px with a Sony-DSC-S950 camera, 1/2.3 inch CCD sensor. 2(b) Image of a *harvestable* head of broccoli originally captured at a size of 2288×1712 px with a Nikon-E4600 camera, 1/2.5 inch CCD sensor.

2.2. Modules and development kits

The NVIDIA Jetson Nano was used to implement the image analysis. It is a small and powerful computer that allows running multiple neural networks in parallel for applications such as image classification, object detection, segmentation or voice processing. In addition, it has a fully configurable 40-pin General Purpose Input/Output (GPIO) connector through which other devices can be easily controlled, such as the tool needed to cut a broccoli head. This device has a power consumption of 5 watts, a particularly important parameter for embedded systems that depend on a battery. In fact, it could be easily connected via a dc-dc converter (12 to 5 V) to the battery of the tractor system to which we intend to attach our harvesting prototype, once developed. Though other manufacturers and models of embedded microprocessors available in the market, as can be seen in the studies carried out in Embedded microprocessor benchmark consortium (2020), the NVIDIA Jetson Nano was chosen because, as we will show later, it met our performance requirements, and its price was very low.

2.3. Choice of learning algorithm

To estimate automatically the maturity state and of the broccoli heads, two different task are required: the location of the head within the image, and the analysis of it, using a classifier. Currently there are several network models that allow the location and classification of objects in a unified way (object detection), reducing the total processing time. Faster R-CNN Ren et al. (2017), SSD Liu et al. (2016), Retina Net Lin et al. (2017) or all three versions of YOLO Redmon et al. (2016), Redmon and Farhadi (2017), Redmon and Farhadi (2018) are examples of models that perform detection of objects with low inference times. Among the previous models, Faster R-CNN has become the reference model Koirala et al. (2019), due to to the accuracy and robustness of the predictions it generates, which is why we have chosen it as the engine of our system.

The aforementioned network models are based on a deep convolutional neural network trained on a large set of images such as ImageNet Deng et al. (2009). Networks of type ResNet He et al. (2016) or VGG16 Simonyan and Zisserman (2015), are commonly used as baseline for deep network models. In fact, ResNet-50 architecture presents a good trade-off between precision and inference-time compared to other architectures, as shown in Canziani et al. (2016).

For our system, we decided to use a pre-trained model in order to take advantage of what is known as transfer learning Tan et al. (2018). A model trained in ImageNet will have essentially "learned" both the low-level and high-level features in the images. This is of paramount importance for the development of a real-world application like the one we are dealing with, where the number of training data will always be low, mainly due to the cost in time and resources to obtain the images, and the need of an expert in the maturity state of broccoli to label those images. Fully training this type of network requires adjusting over 23 million parameters (network weights for a ResNet −50 model), which is very difficult to achieve without ImageNet or other large dataset, such as COCO. However, the net weights pre-trained in ImageNet can be used as a starting point, adjusting then in a last training stage to locate and classify the broccoli heads into its different states of maturity.

Though there are several popular environments that can be used for that last training, as TensorFlow Abadi et al. (2015), PyTorch Paszke et al. (2017), Caffe Jia et al. (2014) or Detectron Girshick et al. (2018) (now integrated into PyTorch), we selected Detectron because of its ease of use, and our previous experience with this environment. In any case, the choice of the environment should not lead to very different results in training.

In order to be used in the NVIDIA Jetson Nano for inference, the model was transformed using NVIDIA TensorRT NVIDIA (2020a,b), obtaining a very good performance when processing an image size of 640 × 480 px, as we shall see in the results section.

2.4. Network training

As we have previously mentioned, the Faster R-CNN model is built on a pre-trained ResNet-50 backbone. The Faster R-CNN region generation block is implemented with a Feature Pyramid Network (FPN) (Lin et al., 2017) which, unlike the standard model, uses only square base ROIs. Since broccoli can typically be contained in a 1:1 aspect ratio bounding box, using training regions with other aspect ratios to adjust the model can result in elongated proposals that cover more than one broccoli head, reducing the efficiency of the detection system.

As mentioned above, the training was carried out with Detectron, on a GTX 1070 Ti GPU card with 8 GB RAM. It was configured to use minibatch stochastic gradient descent (SGD) with 2 images per GPU and 256 ROIs per image. Hence, the number of ROIs per training minibatch was 512. We used a weight decay of 0.0001 and momentum of 0.9. The model was trained for 60000 iterations, that are equivalent to 24 epochs. The initial learning rate was set to 0.0025, and it is reduced by a factor of 10 each 20000 iterations. Besides, we used horizontal image flipping for data augmentation.

2.4.1. Metrics used for evaluation

To evaluate system performance, four different metrics were considered: precision, recall, F1-score and Average Precision (AP).

To determine whether a proposal generated by the system has located a head of broccoli, we must use an overlap metric with respect to the reference marks. In object detection, Intersection over Union (IoU) is used to measure the accuracy of a detection proposal. This parameter reaches a value of 1.0 when the proposal is perfectly overlapped with the ground-truth bounding box, and a minimum value of 0.5 is usually considered a good object detection Everingham et al. (2015). The formal definition of IoU is given by Eq. 1, where P is the area of the proposal and G is the area of the ground-truth mark.

$$IoU(P,G) = \frac{|P \cap G|}{|P \cup G|} \tag{1}$$

As the network can generate slightly different proposals for the same object, we must use a non-maximum proposal suppression algorithm (NMS) at its output. A NMS algorithm permits to obtain non-overlapping proposals from the complete list of proposals produced by the network, ordered desdendingly with respect to their score and with the highest level of confidence. To measure the overlap between two proposals, IoU given by Eq. 1 is commonly used. In this work, we have set an overlap threshold of 0.5 for the NMS algorithm, so any candidate with at least this level of IoU with a lower score proposal will be eliminated from the final list.

In order to determine performance, all the elements in the final list of proposals are compared with the ground-truth marks, using IoU as the overlap metric. This way, a proposal is considered to be a true positive (TP) when its IoU with a reference mark exceeds a previously set IoU threshold. On the other hand, if a proposal does not reach that IoU threshold with any ground-truth mark, it will be considered a false positive (FP). And finally, each ground-truth mark for which no overlapping proposal is obtained from the network, i.e. each non-localized broccoli head, is considered as a false negative (FN). Using these markers, the usual metrics of precision, recall and F1-score are defined (Koirala et al., 2019). For unbalanced, multiclass test sets, these parameters are calculated for each category, and the final value is the weighed average using the number of items in each category as weights. Precision for class C_i is calculated as the number of *correctly* predicted objects of that class divided by *all* predicted objects of class C_i. On the other hand, recall for class C_i is the number of correctly predicted objects of class C_i divided by the number of the objects of class C_i present in the test set. And finally,

$$F1score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{2}$$

With this definitions, precision informs about the ability of a classifier to identify relevant objects, while recall measures the ability of the classifier to find all relevant cases. Finally, F1-score informs whether a model has been adjusted to favour precision over recall or vice versa, and reaches a maximum value of 1 when both are maximum, decreasing when precision or recall decreases.

The area under the precision-recall curve (AP) can also be used as a single metric to summarize the performance of the object detection model. For a given IoU threshold to determine TPs, a model with high precision at all recall levels will have a high AP score. In a multi class object detection task the mean Average Precision is used (mAP), where individual AP is averaged over all classes. Multiple variants of this metric have been defined, and it is very common to calculate AP as an average of the results obtained with different IoU thresholds. Therefore, AP can be calculated over a range of thresholds at 10 different IoU ranging from 0.50 to 0.95 at 0.05 step-size, usually denoted, AP@[.50:.05:.95].

2.5. Broccoli head size estimation

According to geometrical optics laws, the broccoli head size can be estimated from the images using Eq. 3:

$$R = \frac{D \times O \times P}{FL} \tag{3}$$

where R is the real object size (mm); D, the distance to the object (mm); O, the object size within the image (px); P, the pixel size (μm) and FL, the focal length (mm). From this parameters, the most difficult to obtain is the distance to the object, because it has to be recorded when the image is taken. Though it could be measured using a RGBD camera or a LiDAR module, this complicates our vision system, and, in our view, it is not necessary. Given the new methods of working and monitoring the plantations, it can be assumed that the plants of the same harvest grow in an homogeneous way, so that they are all at approximately the same distance from the ground. This way, if we take the images at a fixed distance from the ground, as we did with all our cameras, this parameter only has to be measured once in each session. Obviously, this condition is very easy to maintain when the camera is placed on the tractor, as we did with our Sony IMX179.

Considering that the relationship between R and D is linear, an error of E_D mm in the measure of D, will give rise to an error of

$$E_R = \frac{E_D \times O \times P}{FL} \tag{4}$$

Therefore, for an image taken with the Sony-DSC-S950 ($FL = 5.8$ mm, $P = 2.37$ μm), in which we have a broccoli head of 500 px size (at least 1/4 of the horizontal dimension of the image), an error as large as $\pm$10 cm in D, will produce an error in our estimation of $\pm$1 cm.

Anyway, as we discussed in the introduction, the concrete size is not critical for agri-food market, as workers decide the moment of harvesting based mainly on the state of maturity. In this work, we used this estimation to study the relation between our classes and the sizes of the broccoli heads.

3. Results and discussion

Once the Faster R-CNN network was trained with our images, we tested its performance on the whole test set. The mean inference time with the GPU system used for training was 59 ms per image. When performing this operation using NVIDIA Jetson Nano we found that the system required 134.6 s to process the 1234 images in the test set. Thus, the mean inference time is 109 ms per image with the low cost computer. Therefore, as we had raised in the objectives, it allows us to operate the detection subsystem in real time.

3.1. Evaluation of multi-class object detection

First of all, we have carried out a study of the system performance with the minimum score required for detection proposals. The objective of this study was to maximize the F1 score, obtaining as a result that the maximum efficiency of the detector is achieved by discarding the proposals that do not reach a minimum score of 0.7.

Table 2 shows the performance parameters of our system, for a configuration which discards proposals with a score lower than 0.7, as previously stated, and counts a true positive if the IoU overlap criteria given in Section 2.4.1 is met. The table shows, for the three classes of interest and in a simplified way, the confusion matrix over the test set, together with the *Precision*, *Recall* and *F1* rates.

As a first result, it should be noted that more than 97% of the broccoli heads are found on the main diagonal of the confusion matrix, and therefore correctly located and well classified. This result informs us of a high overall performance when separating the proposed classes.

When analysing the deviations from the main diagonal, we first observe that the modelling achieved for the *harvestable* and *immature* classes generates a maximum erroneous transfer between both classes of less than 1.8%. This result is interesting, considering that one of the indicators of the state of maturity is the size of the broccoli head, which has not been used as a direct or indirect input to the classifier.

The behaviour of the system with the heads to be discarded (*Wasted*) indicates that it is capable of correctly separating 95% of the instances, allowing the wrong harvesting of only 3% of the cases to be discarded.

With the level of certainty required for a valid detector proposal, we can see that the configured system is highly robust. This translates into six false negatives in the immature class, which can be harvested in a later pass, so they would not affect negatively the harvest. And there is only a false positive, also in this class, which would not trigger the cutting mechanism as immature, and therefore would not affect the yield of the harvest.

To study the response of our detector with the size of the heads classified as *harvestable*, we have analysed the mean diameter of the proposals generated for each class. The results obtained are shown in Table 3, together with its associated standard deviations.

On average, we can see that the heads classified as *harvestable* have a larger diameter than those classified as *immature*, as we expected. And the moderate standard deviations associated with each metric indicate that the diameter for harvestable and immature classes are different. The high value of the standard deviation for *immature* class points out greater variability in size associated with this class. However, considering the high yield achieved when detecting the *harvestable* and *immature* heads, it seems clear that the size information is not decisive for the detector to determine the real maturity state of the broccoli head.

Fig. 3 shows three cases well classified as *harvestable*, *immature* and

Table 2
Performance parameters over the test set for the selected system configuration. High recall and precision rates are obtained for the three classes considered. The results correspond to the optimum detector configuration with a score > 0.7 and an IoU > 0.5.

	Predicted label				
True label	*Harvestable*	*Immature*	*Wasted*	FN	Recall
Harvestable	535	8	2	0	0.98
Immature	10	541	1	6	0.97
Wasted	4	2	125	0	0.95
FP	0	1	0		
Precision	0.97	0.98	0.97		
F1-score	0.97	0.98	0.95		

Table 3
Average sizes of the proposals generated by the network together with their associated standard deviations. The average size of the proposals classified as *harvestable* exceeds that of the immature ones.

Class	Diameter (cm)	(σ)
Harvestable	16.42	2.26
Inmmature	10.51	2.71
Wasted	16.67	2.09

wasted (botrytis). As can be seen in figures (b) and (c) the system is able of performing correctly despite the reflections produced by the sun in the photograph.

Fig. 4 shows misclassified cases. As can be seen, many times it is difficult, even for an expert, to correctly classify cases by looking at just one photograph. And it is very likely that the expert was wrong when classifying the case shown in the photograph 4(f) as wasted, perhaps confused by the reflections produced by the sun, and so the system could have classified it correctly.

3.2. Global results

Table 4 shows the results obtained on the test set indicated in Table 1, in terms of AP, AP_{50} and AP_{75} (average precision rates assessed at 0.50 and 0.75 IoU threshold), in addition to AP on medium (AP_M) and large (AP_L) objects as indicators of the evaluation of the detection process.

AP is averaged for all categories (classes). In this context, it is traditionally known as *mean average precision* (mAP). It shows us the percentage of hits on all classes. In this case, we obtained a value over 83%, which can be considered as acceptable. We can also appreciate that the values for AP_{50} and AP_{75} are very close, which means that the proposals of the network are adjusted to the marks made by the expert, thus detecting quite well the dimensions of the broccoli heads.

4. Conclusions and future work

At the sight of our results, with 97% of the broccoli heads in the test set correctly located and classified, we can conclude that the localization of broccoli heads in images and its classification can be accomplished in one unique step with a Faster R-CNN deep neural network. Besides, as far as we know, this is the first work that identifies the harvestable broccoli heads by its state of maturity, and not only by the size. Also, wasted heads are detected, allowing to improve the overall quality of the harvest.

On the other hand, the inference step was implemented in a low-cost hardware, cheaper than an industrial computer, and performed at a rate of ten images per second, so it is suitable for our real time application. This is very important because, integrated with an image capture system, it can decide in real time if a broccoli head should be harvested or not, triggering a cutting module in the right moment.

As future work, we consider two different lines. In one hand, we intend to continue the development of a complete harvester prototype, integrating first the image capture module and our broccoli heads detector. The integrated system should be thoroughly tested in real conditions before adding the mechanical cutting module, which will be

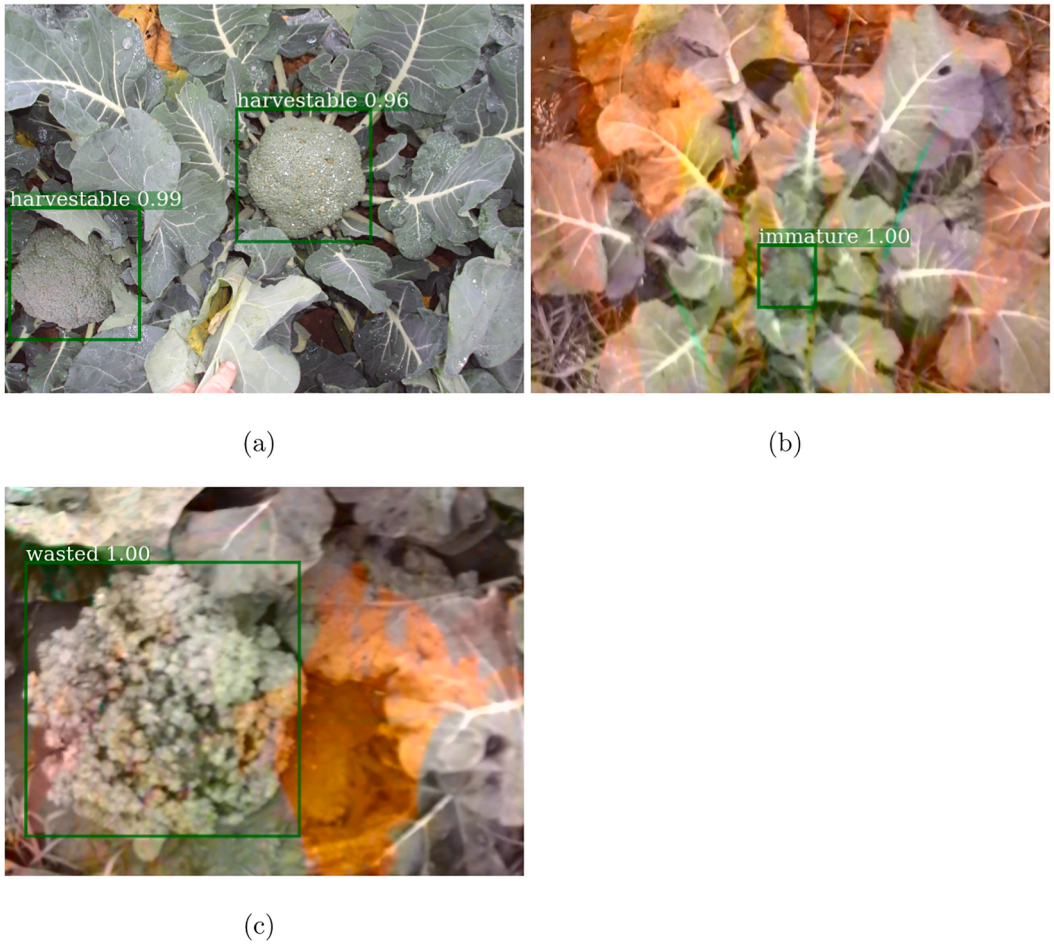

Fig. 3. Correct class. 3(a) Two correctly detected harvestable broccoli heads on the same image. 3(b) Well classified immature broccoli head. 3(c) Well classified wasted broccoli head.

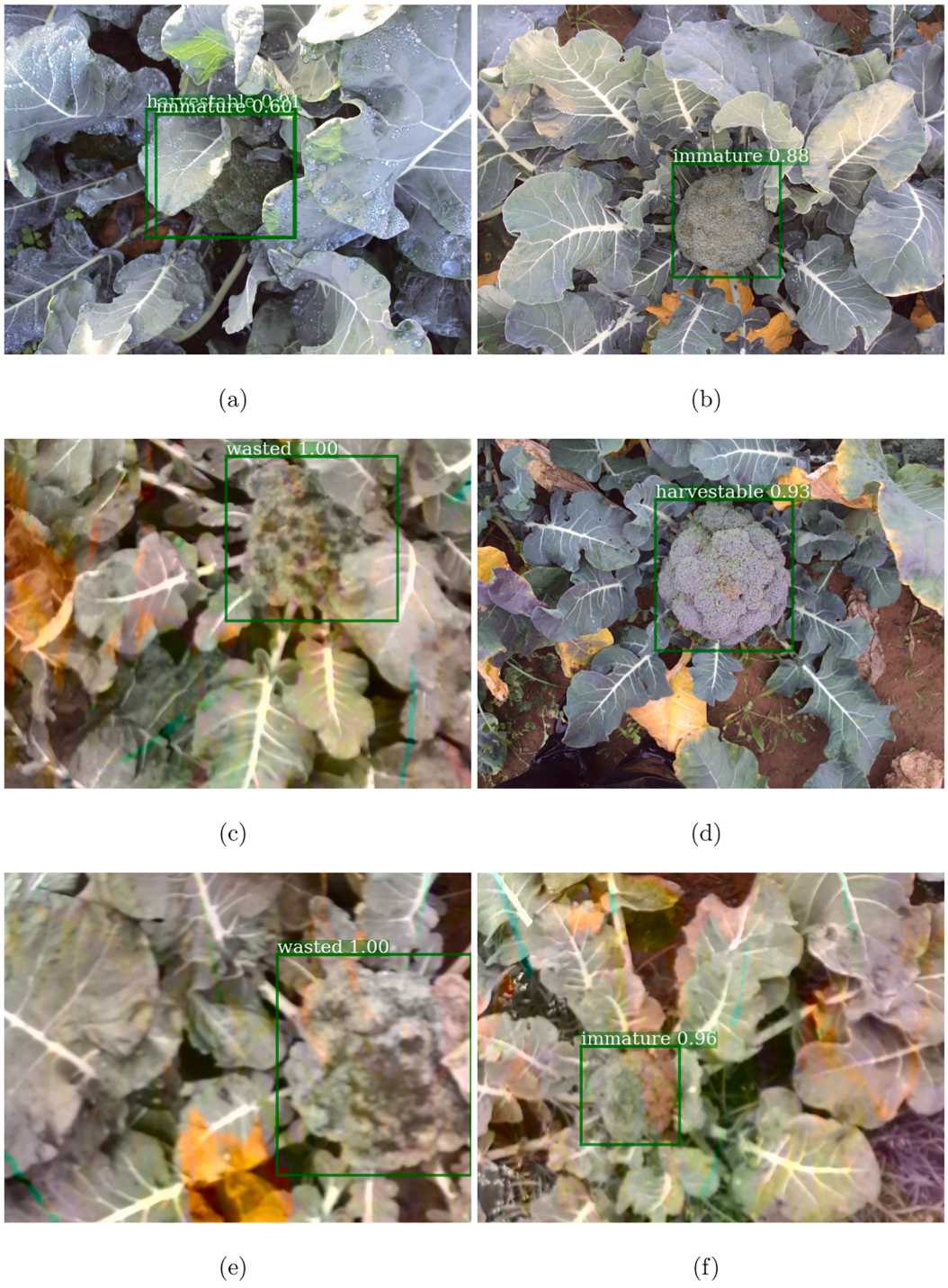

Fig. 4. Wrong class. 4(a) Immature broccoli head classiffed as harvestable. 4(b) harvestable broccoli head classified as immature. 4(c) harvestable broccoli head classified as wasted. 4(d) Wasted broccoli head classified as harvestable. 4(e) Immature broccoli head classified as wasted. 4(f) Wasted broccoli head classified as immature.

Table 4
Detection results on the test set.

Method	AP	AP_{50}	AP_{75}	AP_M	AP_L
Result	0.8347	0.9851	0.9816	0.6915	0.8381

precisely synchronised in order to harvest every localized broccoli head that is classified as harvestable.

Regarding our vision system, we intend to improve it in two ways. First, in order to increase image generalization, our image database should be increased, adding some other varieties of broccoli grown in our region, along with other lighting conditions (acquiring images at different times of the day). This will make our classifier even more robust. Besides, we are considering to develop an optional image capture module with artificial, controlled illumination, with the aim of making the prototype usable at night.

CRediT authorship contribution statement

Antonio García-Manso: Conceptualization, Software, Investigation, Data curation, Writing – original draft, Writing – review & editing. **Ramón Gallardo-Caballero:** Software, Writing – review & editing, Data curation. **Carlos J. García-Orellana:** Validation, Supervision. **Horacio M. González-Velasco:** Conceptualization, Writing – review & editing, Supervision. **Miguel Macías-Macías:** Supervision.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We thank Fernando Gómez Viseas for helping us in all matters related to the mechanical part of the image capture system as well as for the development of a first version of the mechanical prototype for cutting broccoli heads. We also want to thank Francisco García Manso, who gave us access to his broccoli plantations and advised us on all aspects related to broccoli cultivation.

References

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G.S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., Levenberg, J.,Mané, D., Monga, R., Moore, S., Murray, D., Olah, C., Schuster, M., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Viégas, F., Vinyals, O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y., Zheng, X., 2015. TensorFlow: Large-scale machine learning on heterogeneous systems, Accessed: 25-05-2020. URL http://tensorflow.org/.

Anejo no 20, estudio de costes y análisis de sensibilidad del proyecto, www.juntaex.es/filescms/con03/uploaded_files/SectoresTematicos/DesarrolloRural/Regadios/RegadioArroyoDelCampo/MemoriaYAnejos/AN.20-EstudioDeCostesYAnalisisDeSensibilidad.pdf, Accessed: 08-06-2020.

Bender, A., Whelan, B., Sukkarieh, S., 2020. A high-resolution, multimodal data set for agricultural robotics: A ladybird's-eye view of brassica. J. Field Robot. 37 (1), 73–96. https://doi.org/10.1002/rob.21877. https://onlinelibrary.wiley.com/doi/abs/10.1002/rob.21877.

Blok, P.M., Barth, R., Berg, W.V.D., 2016. Machine vision for a selective broccoli harvesting robot. IFAC-PapersOnLine 49 (16), 61–71. https://doi.org/10.1016/j.ifacol.2016.10.013.

Blok, P.M., van Evert, F.K., Tielen, A.P.M., van Henten, E.J., Kootstra, G., 2021. The effect of data augmentation and network simplification on the image-based detection of broccoli heads with mask r-cnn. J. Field Robot. (38), 85–104. https://doi.org/10.1002/rob.21975.

Canziani, A., Paszke, A., Culurciello, E., 2016. An analysis of deep neural network models for practical applications. arXiv:1605.07678.

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., Fei-Fei, L., 2009. ImageNet: A Large-Scale Hierarchical Image Database. In: CVPR09.

Embedded microprocessor benchmark consortium, 2020. MLMark - An EEMBC Benchmark, https://www.eembc.org/mlmark/scores.php, Accessed: 25-05-2020.

Everingham, M., Eslami, S.M.A., Van Gool, L., Williams, C.K.I., Winn, J., Zisserman, A., 2015. The pascal visual object classes challenge: A retrospective. Int. J. Comput. Vision 111 (1), 98–136. https://doi.org/10.1007/s11263-014-0733-5.

FAO, 2019. Food and agriculture organization of the united nations. food and agriculture data, http://www.fao.org/faostat/en/, Accessed: 25-05-2020.

Girshick, R., Radosavovic, I., Gkioxari, G., Dollár, P., He, K., 2018. Detectron, https://github.com/facebookresearch/detectron, Accessed: 25-05-2020.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770–778.

Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., Guadarrama, S., Darrell, T., 2014. Caffe: Convolutional Architecture for Fast Feature Embedding. In: Proceedings of the 22Nd ACM International Conference on Multimedia, MM '14. ACM, New York, NY, USA, pp. 675–678.

Koirala, A., Walsh, K.B., Wang, Z., McCarthy, C., 2019. Deep learning – method overview and review of use for fruit detection and yield estimation. Comput. Electron. Agric. 162, 219–234. https://doi.org/10.1016/j.compag.2019.04.017.

Kusumam, K., Krajník, T., Pearson, S., Cielniak, G., Duckett, T., 2016. Can you pick a broccoli? 3D-vision based detection and localisation of broccoli heads in the field. In: 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 646–651.

Kusumam, K., Krajník, T., Pearson, S., Duckett, T., Cielniak, G., 2017. 3d-vision based detection, localization, and sizing of broccoli heads in the field. J. Field Robot. 34 (8), 1505–1518. https://doi.org/10.1002/rob.21726.

Le Louedec R., Montes, HA., Duckett, T., Cielniak, G., 2020. Segmentation and detection from organised 3D point clouds: A case study in broccoli head detection. In: The IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops.

Lin, T., Goyal, P., Girshick, R., He, K., Dollár, P., 2017. Focal loss for dense object detection. In: 2017 IEEE International Conference on Computer Vision (ICCV), pp. 2999–3007. https://doi.org/10.1109/ICCV.2017.324.

Lin, T., Dollar, P., Girshick, R., He, K., Hariharan, B., Belongie, S., 2017. Feature pyramid networks for object detection. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE Computer Society, Los Alamitos, CA, USA, pp. 936–944. https://doi.org/10.1109/CVPR.2017.106.

Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., Berg, A.C., 2016. Ssd: Single shot multibox detector. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (Eds.), Computer Vision – ECCV 2016, Springer International Publishing, pp. 21–37. doi: 10.1007/978-3-319-46448-0_2.

Mullaney, S., Weinroth, M., 2019. Food source information. broccoli, https://fsi.colostate.edu/broccoli1/, Accessed: 25-05-2020.

MYCOM, 2021. Mycom, http://www.mycom-japan.co.jp/n_topics/bah2-1800.html, Accessed: 22-03-2021.

NVIDIA, 2019. Autonomous machines, https://www.nvidia.com/es-es/autonomous-machines/embedded-systems/jetson-nano/, Accessed: 25-05-2020.

NVIDIA, 2020a. NVIDIA TensorRT, https://developer.nvidia.com/tensorrt, Accessed: 25-05-2020.

NVIDIA, 2020b. Jetson Nano: Deep Learning Inference Benchmarks, https://developer.nvidia.com/embedded/jetson-nano-dl-inference-benchmarks, Accessed: 25-05-2020.

Ramirez, R., 2006. Computer vision based analysis of broccoli for application in a selective autonomous harvester.

Redmon, J., Farhadi, A., 2017. Yolo9000: Better, faster, stronger. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition, pp. 6517–6525. https://doi.org/10.1109/CVPR.2017.690.

J. Redmon, A. Farhadi, Yolov3: An incremental improvement, arXiv preprint arXiv: 1804.02767 (2018).

Redmon, J., Divvala, S., Girshick, R., Farhadi, A., 2016. You only look once: Unified, real-time object detection. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition, pp. 779–788. https://doi.org/10.1109/CVPR.2016.91.

Ren, S., He, K., Girshick, R., Sun, J., 2017. Faster R-CNN: Towards real-time object detection with region proposal networks. IEEE Trans. Pattern Anal. Mach. Intell. 39 (6), 1137–1149.

Paszke, A., Gross, S., Chintala, S., Chanan, G., Yang, E., DeVito, Z., Lin, Z., Desmaison, A., Antiga, L., and Lerer, A. (2017). Automatic Differentiation in PyTorch. NIPS 2017 Workshop on Autodiff.

RoboVeg, 2021. Roboveg harvesters, https://www.roboveg.com/harvesters, Accessed: 22-03-2021.

Simonyan, K., Zisserman, A., 2015. Very deep convolutional networks for large-scale image recognition. In: ICLR 2015 Conference Proceedings, San Diego, USA.

Tan, C., Sun, F., Kong, T., Zhang, W., Yang, C., Liu, C., 2018. A Survey on Deep Transfer Learning. In: Kurková, V., Manolopoulos, Y., Hammer, B., Iliadis, L., Maglogiannis, I. (Eds.), Artificial Neural Networks and Machine Learning – ICANN 2018. Springer International Publishing, Cham, pp. 270–279. https://doi.org/10.1007/978-3-030-01424-7_27.

Zhou, C., Hu, J., Xu, Z., Yue, J., Ye, H., Yang, G., 2020. A monitoring system for the segmentation and grading of broccoli head based on deep learning and neural networks. Front. Plant Sci. 11, 402. https://doi.org/10.3389/fpls.2020.00402.