








## Article

# Peaches Detection Using a Deep Learning Technique—A Contribution to Yield Estimation, Resources Management, and Circular Economy

Eduardo T. Assunção <sup>1,2</sup> , Pedro D. Gaspar <sup>1,2,\*</sup> , Ricardo J. M. Mesquita <sup>1</sup> , Maria P. Simões <sup>3</sup> ,  
António Ramos <sup>3</sup> , Hugo Proença <sup>4</sup>  and Pedro R. M. Inacio <sup>4</sup> 

- <sup>1</sup> C-MAST Center for Mechanical and Aerospace Science and Technologies, University of Beira Interior, 6201-001 Covilha, Portugal; eduardo.assuncao@ubi.pt (E.T.A.); ricardo.mesquita@ubi.pt (R.J.M.M.)
- <sup>2</sup> Department of Electromechanical Engineering, University of Beira Interior, Rua Marquês d'Ávila e Bolama, 6201-001 Covilha, Portugal
- <sup>3</sup> School of Agriculture, Polytechnic Institute of Castelo Branco, 6000-084 Castelo Branco, Portugal; mpaulasimoes@ipcb.pt (M.P.S.); aramos@ipcb.pt (A.R.)
- <sup>4</sup> Instituto de Telecomunicações, Department of Computer Science, University of Beira Interior, 6201-001 Covilha, Portugal; hugomcp@di.ubi.pt (H.P.); prmi@ubi.pt (P.R.M.I.)
- \* Correspondence: dinis@ubi.pt

**Abstract:** Fruit detection is crucial for yield estimation and fruit picking system performance. Many state-of-the-art methods for fruit detection use convolutional neural networks (CNNs). This paper presents the results for peach detection by applying a faster R-CNN framework in images captured from an outdoor orchard. Although this method has been used in other studies to detect fruits, there is no research on peaches. Since the fruit colors, sizes, shapes, tree branches, fruit bunches, and distributions in trees are particular, the development of a fruit detection procedure is specific. The results show great potential in using this method to detect this type of fruit. A detection accuracy of 0.90 using the metric average precision (AP) was achieved for fruit detection. Precision agriculture applications, such as deep neural networks (DNNs), as proposed in this paper, can help to mitigate climate change, due to horticultural activities by accurate product prediction, leading to improved resource management (e.g., irrigation water, nutrients, herbicides, pesticides), and helping to reduce food loss and waste via improved agricultural activity scheduling.

**Keywords:** convolutional neural network; deep learning; fruit detection; precision agriculture; sustainability



**Citation:** Assunção, E.T.; Gaspar, P.D.; Mesquita, R.J.M.; Simões, M.P.; Ramos, A.; Proença, H.; Inacio, P.R.M. Peaches Detection Using a Deep Learning Technique—A Contribution to Yield Estimation, Resources Management, and Circular Economy. *Climate* **2022**, *10*, 11. <https://doi.org/10.3390/cli10020011>

Academic Editors: Jong Ahn Chun, Hen-I Lin and Daeha Kim

Received: 21 December 2021

Accepted: 10 January 2022

Published: 18 January 2022

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



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

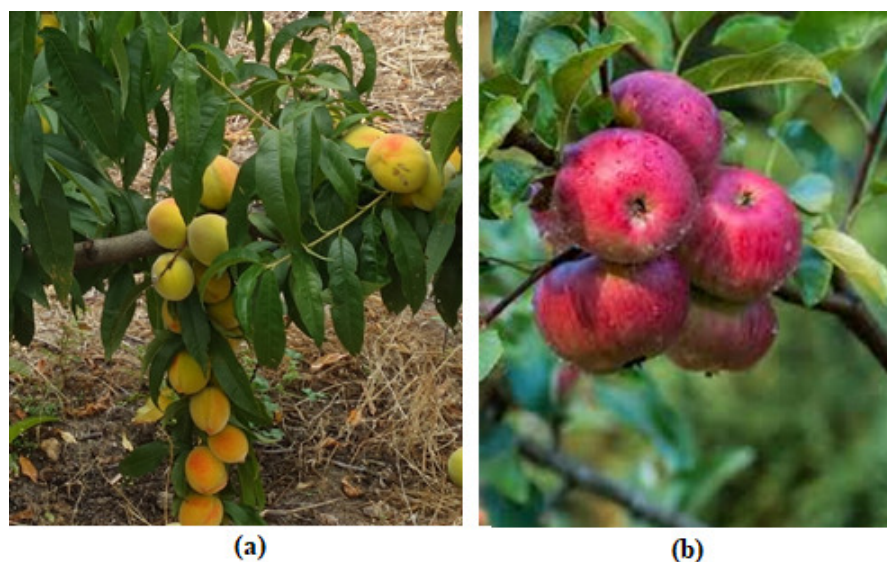
Sustainable agricultural practices can improve product availability in a supply chain. Some traditional practices can dangerously compromise soil structure and an orchard's abilities to adapt to extreme weather events, climate change, and the stresses of intensive harvest production. All methods, whether new or traditional, must be examined when performing sustainable agriculture practices [1].

Agriculture directly impacts climate change as its activities contribute to a portion of global anthropogenic greenhouse gas (GHG) emissions. The main GHGs produced by agriculture are nitrous oxide (N<sub>2</sub>O) and carbon dioxide (CO<sub>2</sub>). N<sub>2</sub>O is produced by the microbial transformation of nitrogen in soils during applying synthetic fertilizers to the ground, and CO<sub>2</sub> is produced by changes in above- and below-ground carbon stocks caused by land-use changes [2]. Continuous monitoring can provide an up-to-date status of a field in real-time, based on the selected parameters of the plants and fields. This approach is made possible by a computer vision framework on the fields. Fruit detection via computer vision allows monitoring and control of yield, contributing to higher production. Higher yields indicate the possibility of rising production, with less water and without extending

the area under agriculture, which means less deforestation and a reduction of natural resource consumption, and GHG emissions.

The application of precision agriculture techniques can help to mitigate climate change. Deep neural networks emerge as powerful tools in precision agriculture as they can provide accurate product prediction, leading to improved resource management, such as irrigation water, nutrients, herbicides, and pesticides, and help reduce food loss and waste by improved agricultural activity scheduling. Alibabaei et al. [3] investigated the ability of an encoder–decoder long short-term memory model, to model the daily evapotranspiration for threes. Assunção et al. [4] used computer vision, based on convolutional neural networks, to classify peach fruit diseases.

Yield estimation is a branch of precision agriculture. It enables planning for cropping, operations, inventory management, and other ancillary services (e.g., fruit pickup by a robot). There are several published works in the field of fruit yield estimation. For example, Häni et al. [5] presented a methodology for apple production estimation. Dorj et al. [6] developed a system for citrus fruit detection and counting. Bargoti and Underwood [7] presented a work on the detection of mangoes, almonds, and apples in orchards. A fruit yield estimation pipeline begins with detection, followed by tracking and counting. The fruit detection phase is critical to the performance of the yield estimation system. Most state-of-the-art fruit detection systems are based on a convolutional neural network (CNN) [8], such as object detection, segmentation, among others. These methods automatically extract features from the appearance of fruit images (i.e., colors and shapes). In this context, the development of fruit detection is specific (because the fruit colors, sizes, shapes, clusters, and distributions in the trees are particular). Figure 1 illustrates this problem. Considering the lack of work with peaches, this paper presents the results for peach detection using the object detection framework faster region-based convolutional neural network (Faster R-CNN). Another contribution is that the images are from a non-controlled environment (outdoors). That is, they come from a natural orchard.



**Figure 1.** Particularity of fruit type: (a) peach. (b) apple.

Bargoti and Underwood [9] presented a methodology to obtain an estimate of red and green apples. For the detection phase, they used the traditional computer vision approach based on hue, saturation, and value (HSV) color space for segmentation. With the development of convolutional neural networks (CNNs), many authors use them for fruit detection based on segmentation and object detection approaches. Puttemans et al. [10] proposed a watershed and trinocular stereo triangulation-based segmentation with a cascade of weak classifiers to perform a strawberry and apple detector model. Bargoti

and Underwood [11] trained a CNN, where its output was the probability that each pixel belonged to a fruit. Then, the output result was used to generate a binary mask for segmentation. Häni et al. [12] proposed the CNN segmentation model, named U-NET, for apple segmentation and circular Hough transform for fruit detection. The original purpose for this network is medical image segmentation. The author reports an F1 score of 0.858 for detection.

CNNs have also made massive contributions in improving image classification and object detection. In this regard, the object detection framework R-CNN [13] and its variants fast R-CNN [14], faster R-CNN [15], and mask R-CNN [16] are widely used in the literature. Sa et al. [17] proposed faster R-CNN to detect peppers, apples, avocados, mangoes, strawberries, and oranges. However, only the images of peppers are from the orchard. The rest of the fruit images are from the internet (Google Images). CNNs rely on large amounts of data for training to avoid overfitting, and they have good generalization. To achieve good performance with a relatively small set of training data, a transfer learning technique is often used.

Recently, you only look once (YOLO), another branch of object detection based on CNN, became a framework for fruit detection. Koirala et al. [18] applied this model to detect mangoes in orchard images. They used two metrics for evaluation, F1 and average precision (AP), and reported results of 0.96 and 0.98, respectively. Liu et al. [19] applied the mask R-CNN method to detect cucumber fruits. They chose ResNet-101 as the backbone and proposed a modification in the scales and aspect ratios of the anchor boxes. For evaluation, the authors used an F1 score and obtained a score of 0.894 for the test images.

Other studies were developed to detect fruits. However, fruit detection procedures are specific due to the particular distribution of fruit colors, sizes, shapes, branches, and bundles. This study provides the first application (and, consequently, the annotated dataset) of peach detection that is constrained by (1) peaches with different colors, (2) peaches overlapped, and (3) peaches occluded by leaves. The faster R-CNN computer vision framework is used for fruit detection in peach orchards. Fruit detection can be used to count the number of fruits during the culture, and knowing that value, irrigation and all other cultural practices can be conveniently scheduled, and the correct amount of required inputs for fertilization and plant protection can be acquired. This approach improves the productivity and competitiveness of farmers while ensuring environmental concerns. Thus, it can be considered a precision agriculture application based on a deep neural network to help mitigate climate change, as continuous monitoring provides an up-to-date status of the field culture and allows monitoring and control of yields, which may be used as decision-making support systems that lead to increased production with less consumption of natural resources and GHG emissions.

## 2. Materials and Methods

### 2.1. Dataset

The images used to compose the dataset were taken with the Eken H9R camera in a peach orchard in Beira Interior, Portugal. All images that composed the training dataset were from the same orchard and were taken on the same day. A particular feature for this peach dataset is its yellowish color. This feature can be a good way to check the generalization ability of the model when testing images from other orchards with a different peach color. Figure 2 shows a training sample image. The original images have a resolution of  $5472 \times 3648$  (pixels), RGB channels with 24-bit color depths. The relationship between the object (fruit) size and image size affects the detection performance. This means that a small object is difficult to detect in a large image. This is due to the peculiarity of the CNN polling operation and anchor size, which is part of the object detection algorithm. To mitigate this problem, we crop each original image in for quadrants. Each of them with a resolution of  $2736 \times 1824$  (pixels). Figure 3 illustrates this process.





**Figure 2.** Training sample image.



**Figure 3.** Process of cropping the original image into four parts.

For training, 200 images were used, with 1934 annotated fruits. For testing, 40 images (from the same orchard) with 410 annotated peaches were used.

In addition, a small dataset was created from another orchard (Quinta Nova, also in Beira Interior, Portugal), which has a different color, to check the generalization of the model. Figure 4 shows an image that composes the test dataset. It can be seen that the peaches have a reddish color and the illumination varies a lot compared to the training image. This subset of test images has the same resolution as the training images ( $2736 \times 1824$ ). With seven samples, 91 peaches were annotated.



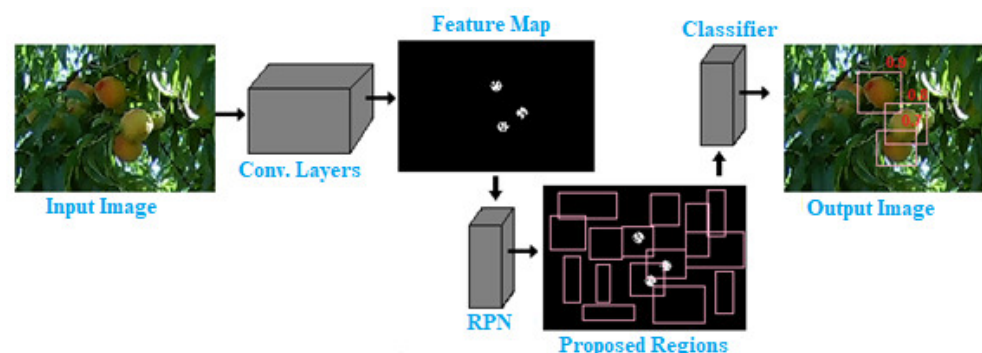
**Figure 4.** Testing sample image.

## 2.2. Object Detection Framework

Object detection is a branch of computer vision that aims to find and classify objects in an image. Histograms of oriented gradients for human detection (HOG) [20] is one of the most popular and traditional (i.e., hand-crafted) methods used for object detection. Nowadays, traditional object detections are replaced by modern models based on CNNs (e.g., faster R-CNN [15], single shot multibox detector (SSD) [21], you only look once (YOLO) [22]).

In this paper, the faster R-CNN model was used as the basis. The faster R-CNN is a two-stage model in which a CNN backbone is used for feature extraction in the first stage (e.g., VGG16 [23], ResNet [24], Inception [25], and others). Figure 5 shows a simplified workflow for the faster R-CNN in the context of fruit detection. Here, the outputs of the convolution layers are called feature maps. The last layer is used as an input to the region proposal network (RPN), which generates the regions containing possible objects. These regions of interest (ROIs) are used for object classification and bounding box prediction.

Figure 5 shows the faster R-CNN object detection model in the context of fruit detection.



**Figure 5.** Faster R-CNN in the context of fruit detection.

The TensorFlow model application programming interface (API) was used to perform the experiments [26]. This API implements the object detection frameworks faster R-CNN and SSD. In this work, the faster R-CNN with the Inception v2 backbone was used because it has better detection accuracy compared to the other variants implemented in this API. In addition, transfer learning was performed on the COCO dataset [27]. Training and testing



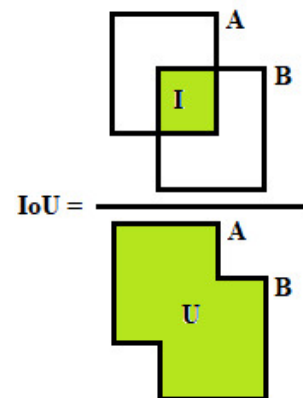
were performed on a desktop with Intel(R) Core(TM) i7-4790 CPU @ 3.60 GHz, 8 GB RAM, and an NVIDIA RTX 2080 graphics card with 8 G memory.

### 2.3. Evaluation Metric

Intersection over union (IOU), defined in (1), is a metric used to evaluate object detection tasks [28]. It provides a value for bounding box prediction. The value obtained from IOU can be used to indicate whether the prediction is true positive (TP), false positive (FP), or false negative (FN), depending on the threshold used. For example, if the IoU threshold is 0.5 and the IoU value for a (positive class) prediction is 0.7, then the prediction is classified as true positive (TP). On the other hand, if the IoU value is 0.3, then the prediction is classified as false negative (FN).

$$IOU = \frac{A \cap B}{A \cup B}, \quad (1)$$

where  $A$  is the ground truth area and  $B$  is the detected area. Figure 6 illustrates the IOU metric.



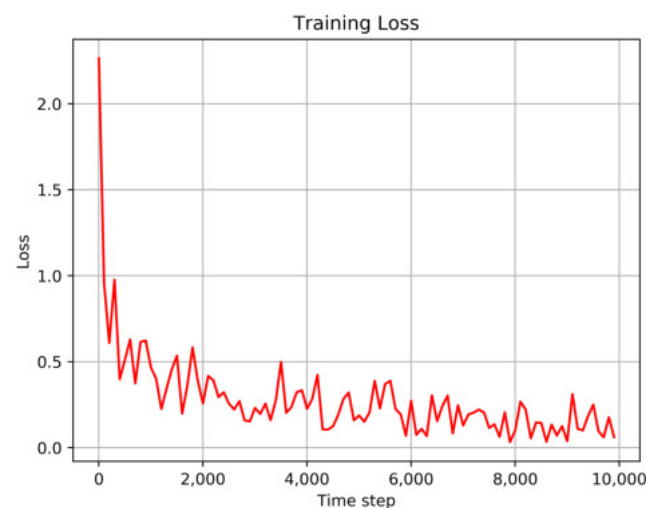
**Figure 6.** Intersection over union illustration.

Evaluations on the test images were performed using the PASCAL VOC metric average precision (AP) for an IOU threshold of 0.5.

## 3. Results and Discussions

### 3.1. Training Result

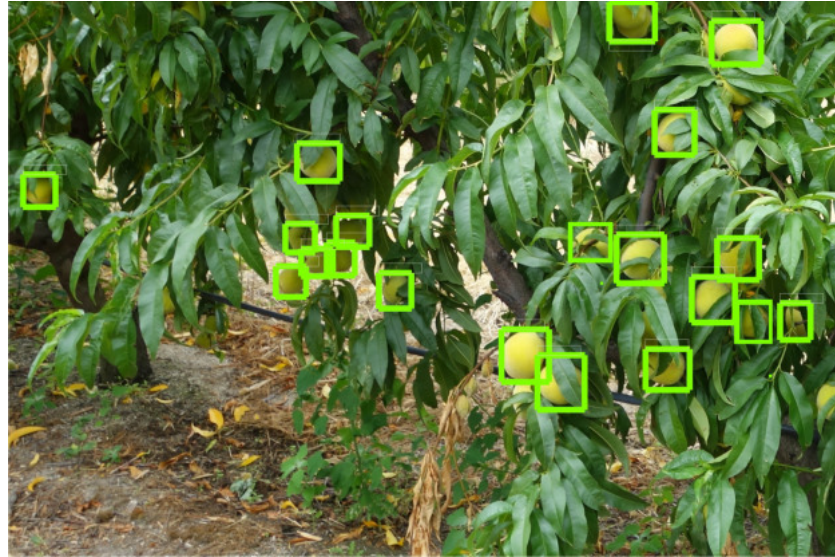
The training loss curve of faster R-CNN with Inception v2 is shown in Figure 7. The final loss was about 0.1 and the loss curve started to stabilize after 8000 iterations.



**Figure 7.** Training loss curve.

### 3.2. Test Results in the Same Orchard Where the Training Was Performed

Detection performance was assessed by the metric AP with IOU = 0.5 and achieved a value of 0.90. Figures 8 and 9 show the visual results.



**Figure 8.** Example (1) of peach detection by faster R-CNN with Inception v2 backbone.



**Figure 9.** Example (2) of peach detection by faster R-CNN with Inception v2 backbone.

### 3.3. Test Results in a Different Orchard

The following results and discussion are intended to show the evaluation of the model in a different orchard, with many variations, compared to the orchard in which the model was trained, to evaluate more rigorously the generalization ability of the model. The test was performed on seven images with 91 peach fruits and resulted in an AP of 0.77.

Figure 10 shows the visual result for the first test image. The peculiarity of this image is the illumination, in which the light is too reflected in the leaves. The color of the fruit is also reddish. These features are completely different from the images found in the training dataset. As can be seen in Figure 10a,b, the model predicted non-occluded and clustered fruits well, but failed to detect some occluded fruits. In this image, the model only fails in the most difficult case of detection (very occluded). To improve the accuracy of the model, samples from this orchard need to be included in the training dataset.



Figure 11 shows the result for another test image, where the lighting is not good, as the exposure of the camera did not help to distinguish fruits from leaves. Moreover, there were several small fruits. In Figure 11a,b, it can be seen that the model did not detect many small fruits. This issue is a common problem with object detection models. One way to solve this problem is to take pictures close to the trees.



(a)



(b)

**Figure 10.** Test result in a different orchard (sample-1): (a) shows a test image and its manually annotated fruits; (b) shows the predicted fruits detection by the model. The green rectangular boxes are corrected detections, and the red circles are illustrative missing detections.





(a)



(b)

**Figure 11.** Test result in a different orchard (sample-2): (a) shows a test image and its manually annotated fruits; (b) shows the predicted fruit detection by the model. The green rectangular boxes are corrected detections, and the red circles are illustrative missing detections.

The distinctive feature of the Figure 12 is the many yellow leaves. This feature might confuse the model, since the color of the fruits in the training images is mostly yellow. Figure 12a,b shows the excellent detection results for this particular image. The model copes well with yellow leaves, which could affect the prediction of the model. There is only one false detection for yellow leaves, one for occluded and two for small fruits.





(a)



(b)

**Figure 12.** Test result in a different orchard (sample-3): (a) shows a test image and its manually annotated fruits; (b) shows the predicted fruits detection by the model. The green rectangular boxes are corrected detections, and the red circles are illustrative missing detections.

Figure 13 has two special features: many yellow leaves and reddish fruits. Again, the model shows good generalization. For this image, there was only one miss detection for yellow leaves and one miss detection for occluded fruit. The model did well for clustered and red fruits.





(a)



(b)

**Figure 13.** Test results in a different orchard (sample-4): (a) shows a test image and its manually annotated fruits; (b) shows the predicted fruits detection by the model. The green rectangular boxes are corrected detections, and the red circles are illustrative missing detections.

### 3.4. Final Discussions

The results show great potential in using the faster R-CNN model for peach detection in an uncontrolled environment. The model can deal well with occluded fruits, bunches of fruits, and light changes, which are features of a non-controlled environment. For a more accurate evaluation of the model, a test was conducted in another orchard. The results showed that the model is robust, as it handles image detections that are completely different from the training data (e.g., the color of the fruit, the color of the leaves, and the lighting). One possible way to improve the detection accuracy is to extend this work by

adding images to the training data from a sequence of frames (from a video). This approach may help the model to better handle the problem of occlusion. Moreover, one may add images from different orchards and take the images closer to the tree.

#### 4. Conclusions

This article contributes toward climate change mitigation through the application of precision agriculture using deep neural networks by providing the field's current status, resulting in higher yields and, consequently, less deforestation, and a reduction in natural resource consumption and greenhouse gas emissions. In this work, we applied the deep learning approach faster R-CNN object detection model to evaluate the detection of peaches in images from an orchard. Peaches have a specific color, size, shape, fruit clustering, and distribution in a tree. The results show that the model handles all of these peculiarities well, in terms of peach fruit detection, and achieves an AP of 0.90 for the test split images belonging to the same orchard of the training, and an AP of 0.77 for the test images belonging to a different orchard (that has a feature discussed earlier). This shows great performance when compared to hand-crafted fruit detection models. For future work, we propose performing the detection in a sequence of frames, to detect hidden fruits more easily. Moreover, one may add images from different orchards and take the images closer to the tree.

**Author Contributions:** Conceptualization: P.D.G.; data curation: E.T.A. and R.J.M.M.; formal analysis: E.T.A., P.D.G., M.P.S., H.P. and P.R.M.I.; funding acquisition: P.D.G.; investigation: E.T.A. and R.J.M.M.; methodology: E.T.A. and P.D.G.; project administration: P.D.G.; resources: R.J.M.M., M.P.S. and A.R.; software: E.T.A.; supervision: P.D.G.; validation: E.T.A.; visualization: E.T.A.; writing—original draft: E.T.A.; writing—review and editing: P.D.G. and H.P. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research work is funded by the PrunusBot project—autonomous controlled spraying aerial robotic system and fruit production forecast, operation no. PDR2020-101-031358 (leader), consortium no. 340, initiative no. 140, promoted by PDR2020, and co-financed by the EAFRD and the European Union under the Portugal 2020 program.

**Acknowledgments:** The authors are thankful to Fundação para a Ciência e Tecnologia (FCT) and R&D Unit “Center for Mechanical and Aerospace Science and Technologies” (C-MAST), under project UIDB/00151/2020, for the opportunity and the financial support to carry on this project. The contributions of Hugo Proença and Pedro Inácio in this work were supported by FCT/MEC through FEDER—PT2020 Partnership Agreement under Project UIDB//50008/2021.

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

#### References

1. Ontario Ministry of Agriculture. Introduction to Sustainable Agriculture. 2016. Available online: <http://www.omafra.gov.on.ca/english/busdev/facts/15-023.htm> (accessed on 11 October 2021).
2. Balafoutis, A.; Beck, B.; Fountas, S.; Vangeyte, J.; van der Wal, T.; Soto, I.; Gómez-Barbero, M.; Barnes, A.P.; Eory, V. Precision Agriculture Technologies positively contributing to GHG emissions mitigation, farm productivity and economics. *Sustainability* **2017**, *9*, 1339. [CrossRef]
3. Alibabaei, K.; Gaspar, P.; Lima, T.M. Modeling evapotranspiration using Encoder-Decoder Model. In Proceedings of the 2020 International Conference on Decision Aid Sciences and Application (DASA), Sakheer, Bahrain, 8–9 November 2020; pp. 132–136.
4. Assunção, E.; Diniz, C.; Gaspar, P.; Proença, H. Decision-making support system for fruit diseases classification using Deep Learning. In Proceedings of the 2020 International Conference on Decision Aid Sciences and Application (DASA), Sakheer, Bahrain, 8–9 November 2020; pp. 652–656.
5. Hãni, N.; Roy, P.; Isler, V. Apple Counting using Convolutional Neural Networks. In Proceedings of the 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Madrid, Spain, 1–5 October 2018; pp. 2559–2565. [CrossRef]
6. Dorj, U.O.; Lee, M.; Yun, S.-s. An yield estimation in citrus orchards via fruit detection and counting using image processing. *Comput. Electron. Agric.* **2017**, *140*, 103–112. [CrossRef]
7. Bargoti, S.; Underwood, J. Deep fruit detection in orchards. In Proceedings of the 2017 IEEE International Conference on Robotics and Automation (ICRA), Singapore, 29 May–3 June 2017; pp. 3626–3633. [CrossRef]



8. Koirala, A.; Walsh, K.; Wang, Z.; McCarthy, C. Deep learning—Method overview and review of use for fruit detection and yield estimation. *Comput. Electron. Agric.* **2019**, *162*, 219–234. [[CrossRef](#)]
9. Wang, Q.; Nuske, S.; Bergerman, M.; Singh, S. Automated Crop Yield Estimation for Apple Orchards. In *Experimental Robotics, Proceedings of the 13th International Symposium on Experimental Robotics, Québec City, QC, Canada, 18–21 June 2012*; Desai, J.P., Dudek, G., Khatib, O., Kumar, V., Eds.; Springer International Publishing: Heidelberg, Germany, 2013; pp. 745–758. [[CrossRef](#)]
10. Puttemans, S.; Vanbrabant, Y.; Tits, L.; Goedemé, T. Automated visual fruit detection for harvest estimation and robotic harvesting. In *Proceedings of the 2016 Sixth International Conference on Image Processing Theory, Tools and Applications (IPTA), Oulu, Finland, 12–15 December 2016*; pp. 1–6. [[CrossRef](#)]
11. Bargoti, S.; Underwood, J.P. Image Segmentation for Fruit Detection and Yield Estimation in Apple Orchards. *J. Field Robot.* **2017**, *34*, 1039–1060. [[CrossRef](#)]
12. Häni, N.; Roy, P.; Isler, V. A Comparative Study of Fruit Detection and Counting Methods for Yield Mapping in Apple Orchards. *arXiv* **2020**, arXiv:1810.09499.
13. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. In *Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014*; pp. 580–587. [[CrossRef](#)]
14. Girshick, R. Fast R-CNN. In *Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 7–13 December 2015*; pp. 1440–1448. [[CrossRef](#)]
15. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 1137–1149. [[CrossRef](#)] [[PubMed](#)]
16. He, K.; Gkioxari, G.; Dollár, P.; Girshick, R. Mask R-CNN. In *Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017*; pp. 2980–2988.
17. Sa, I.; Ge, Z.; Dayoub, F.; Upcroft, B.; Perez, T.; Mccool, C. DeepFruits: A Fruit Detection System Using Deep Neural Networks. *Sensors* **2016**, *16*, 1222. [[CrossRef](#)] [[PubMed](#)]
18. Koirala, A.; Walsh, K.B.; Wang, Z.X.; McCarthy, C. Deep learning for real-time fruit detection and orchard fruit load estimation: Benchmarking of ‘MangoYOLO’. *Precis. Agric.* **2019**, *20*, 1107–1135. [[CrossRef](#)]
19. Liu, X.; Zhao, D.; Jia, W.; Ji, W.; Ruan, C.; Sun, Y. Cucumber Fruits Detection in Greenhouses Based on Instance Segmentation. *IEEE Access* **2019**, *7*, 139635–139642. [[CrossRef](#)]
20. Dalal, N.; Triggs, B. Histograms of oriented gradients for human detection. In *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’05), San Diego, CA, USA, 20–26 June 2005*; Volume 1, pp. 886–893.
21. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.E.; Fu, C.Y.; Berg, A. SSD: Single Shot MultiBox Detector. In *Proceedings of the ECCV 2016, Amsterdam, The Netherlands, 11–14 October 2016*.
22. Redmon, J.; Divvala, S.; Girshick, R.B.; Farhadi, A. You Only Look Once: Unified, Real-Time Object Detection. In *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016*; pp. 779–788.
23. Liu, S.; Deng, W. Very deep convolutional neural network based image classification using small training sample size. In *Proceedings of the 2015 3rd IAPR Asian Conference on Pattern Recognition (ACPR), Kuala Lumpur, Malaysia, 3–6 November 2015*; pp. 730–734. [[CrossRef](#)]
24. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016*; pp. 770–778. [[CrossRef](#)]
25. Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; Wojna, Z. Rethinking the Inception Architecture for Computer Vision. In *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016*; pp. 2818–2826. [[CrossRef](#)]
26. Yu, H.; Chen, C.; Du, X.; Li, Y.; Rashwan, A.; Hou, L.; Jin, P.; Yang, F.; Liu, F.; Kim, J.; et al. TensorFlow Model Garden. 2020. Available online: <https://github.com/tensorflow/models> (accessed on 12 October 2021).
27. Lin, T.Y.; Maire, M.; Belongie, S.; Hays, J.; Perona, P.; Ramanan, D.; Dollár, P.; Zitnick, C.L. Microsoft COCO: Common Objects in Context. In *Proceedings of the ECCV 2014, Zurich, Switzerland, 6–12 September 2014*; Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T., Eds.; Springer International Publishing: Cham, Switzerland, 2014; pp. 740–755.
28. Rezatofighi, H.; Tsoi, N.; Gwak, J.; Sadeghian, A.; Reid, I.; Savarese, S. Generalized Intersection Over Union: A Metric and a Loss for Bounding Box Regression. In *Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019*; pp. 658–666. [[CrossRef](#)]