Revista Engenharia na Agricultura

CLASSIFICATION OF SOYBEAN PODS USING THE DEEP LEARNING TECHNIQUES

Poliana Maria da Costa Bandeira[1]* (iD), Flora Maria de Melo Villar[1] (iD), Iara Aparecida Dias[1] (iD) & Priscila Pascali da Costa Bandeira[1] (iD)

1 - Federal University of Vicosa, Department of Agricultural Engineering, Vicosa, Minas Gerais, Brazil

ABSTRACT

Crop productivity estimate aims at the economic definitions about crop, agricultural management, and land use, among others. However, it is common to observe the use of visual methods to estimate the productivity of the soybean crop through the classification of pods, resulting in a slow, costly method besides being susceptible to human errors. Thus, the objective of this work was to carry out the training of two deep learning methods to classify soybean pods according to the number of grains based on images obtained using a smartphone. Data collection was carried out at the Federal University of Viçosa (UFV). Data consisted of capturing images from a smartphone and training two deep learning models: Mask R-CNN and YOLOv4. To capture the images, the soybean pods were pulled from the plants and placed in a white-bottom container. This procedure occurred for each plant collected. Both models tended towards a better classification for the two- and three-grain pods, reaching a value of 90% for the F1 score metric. This fact may have occurred because of the greater amount of these two types of pods present in the chosen cultivars. Finally, the potential of using deep learning to classify soybean pods based on the number of grains was observed.

CLASSIFICAÇÃO DE VAGENS DA SOJA POR MEIO DE TÉCNICAS DE APRENDIZADO PROFUNDO

RESUMO

A estimativa de produtividade de culturas agrícolas visa as definições econômicas acerca da cultura, gestão agrícola, uso de terras, entre outros. Porém, é comum observar o uso de métodos visuais para realizar a estimativa de produtividade da cultura da soja por meio da classificação de vagens, tornando o processo lento, oneroso e susceptível a erros humanos. Dessa forma, com essa pesquisa, objetivou-se realizar o treinamento de dois modelos do aprendizado profundo para classificar as vagens de soja quanto ao número de grãos a partir de imagens obtidas por smartphone. A coleta dos dados ocorreu na Universidade Federal de Viçosa (UFV), os dados consistiram na captura de imagens a partir de um smartphone e no treinamento de dois modelos de aprendizado profundo: Mask R-CNN e YOLOv4. Para a captura das imagens, as vagens da soja foram arrancadas das plantas e depositadas em um recipiente de fundo branco. Este procedimento ocorreu para cada planta coletada. Ambos os modelos apresentaram uma tendencia de melhor classificação para as vagens de dois e três grãos, alcançando o valor de 90% para a métrica pontuação F1. Possivelmente, este fato ocorreu devido a maior quantidade presente desses dois tipos de vagens nas cultivares escolhidas. Ao final, observou-se o potencial do uso do aprendizado profundo para classificação de vagens da soja com base no número de grãos.

INTRODUCTION

The soybean crop is highly valued in international production due to its high nutritional value and its productive potential (MAUAD et al. 2011). Thus, estimating soybean productivity becomes a strategy for rural producers, agro-industrial companies, and even the government itself because of the impact that this crop generates on the country's economy (RAMOS et al., 2017). However, it is observed that the most used method to estimate soybean productivity depends on human sight, which is a costly and error-prone method.

The use of unconventional techniques to obtain information to estimate the productivity of crops is being applied in the field (ALVES et al., 2018; CHAN et al., 2020; MILLER et al., 2018). These technologies aim to reduce human interference, making the process less costly and less susceptible to human errors. Among current technologies, deep learning techniques stand out, seeking to solve issues from visual data.

Deep learning is an artificial intelligence technique that uses artificial neural networks to learn to recognize patterns in complex data. This technique may be a contribution to solving problems that require visual estimates of data, such as visualization in images. Thus, it is important to choose a tool that can be implemented in image capture equipment. Smartphones are an example of an affordable and popular tool for capturing images. Its applicability in the field can be seen to estimate crop yields (TEDESCO-OLIVEIRA et al. 2020), and disease classification (NGUGI et al., 2020), among others. The union of two technologies, smartphones, and deep learning, can generate facilitating and agile information to estimate the productivity of a crop of extreme domestic economic importance.

Specifically for the soybean crop, deep learning has been used to identify and diagnose diseases and pests that affect crop productivity and to select soybean varieties resistant to atypical conditions (ETIENNE et al., 2021; ZHU et al., 2019). Thus, the objective of this work was to carry out the training of deep learning models in the classification of soybean pods based on the number of grains they have, which is a factor that can contribute to the calculation of soybean productivity estimates.

MATERIAL AND METHODS

The material and methods were described following a chronology, in which data were initially collected, then processed, and finally evaluated. the process in which the work was developed is briefly described in Figure 1.
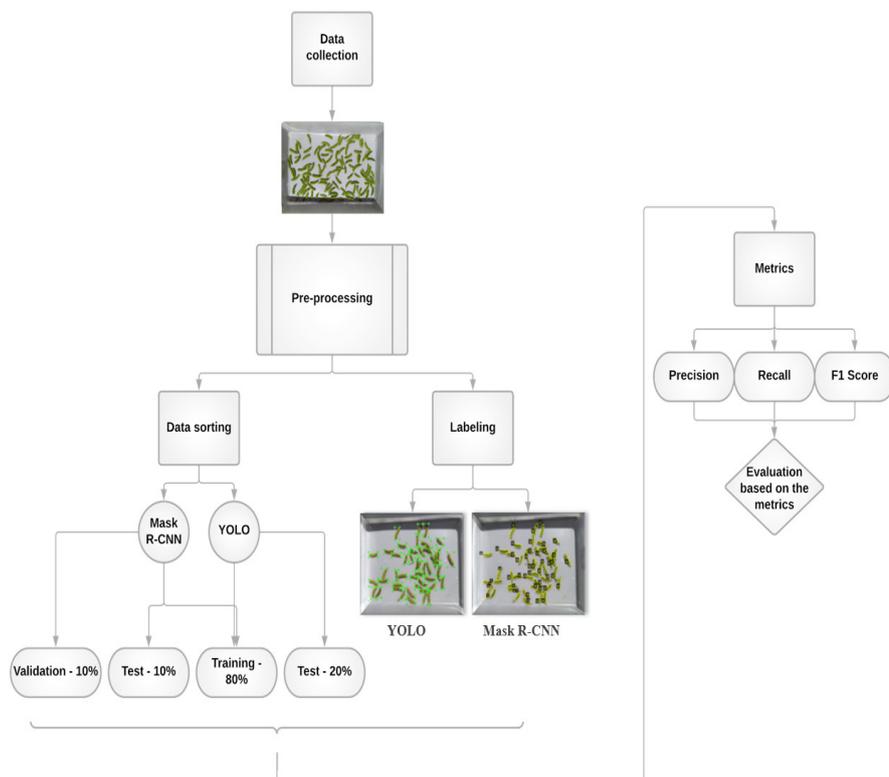
Figure 1. Flowchart showing the material and methods used in the execution of this work

Data collection

In this work, two soybean cultivars were used, TMG 7063 IPRO (cultivar 1) and TMG 7363 RR (cultivar 2). The cultivars were sown in the experimental area of the Federal University of Viçosa (UFV) in Viçosa, state of Minas Gerais. The area was sown on December, 22020, with an average of 26 to 30 plants per 1 m².

The acquisition of digital images, for the construction of the database, was from February 18 and March 4, 2021, with soybean plants in the phenological stages between R6 and R8. The images were captured in the field, and the images were captured in the period from 8:00 a.m. to 12 p.m. This time variation in capturing images is important to obtain data with different levels of luminosity, making the models capable of working in different lighting conditions. As a result, the models can handle images at different times of the day or in environments with shadows and irregular lighting.

The plants used in the experiment were selected at random, including plants of the two cultivars that varied in size, from small to large. To capture the images, the plants were collected from the experimental area. All pods were removed from the plants and placed in a white-bottom container (Figure 2). This procedure was performed for all collected plants and there was no standardization of the distance between the camera and the container filled with pods. Distance non-standardization is chosen to allow deep learning models to be able to classify objects in different situations, making them practical and useful for those who use them. Furthermore, it is not necessary to use any additional equipment to assist in fixing the distance between the pods and the camera lens, resulting in a simple and accessible process.

The images were obtained using a smartphone that has a dual camera of 48 megapixels and five megapixels, with a 1/2" sensor and a 1.8-focal aperture. Throughout the collection, the smartphone settings were kept on automatic. According to the characteristics of the mobile device used, the generated images had a size of 3000 x 4000 pixels. In all, 495 images of both cultivars were captured, as seen in Figure 2, each image containing more than one pod. Therefore, the total database had 23193 pods. The Mask R-CNN model allows the use of three subsets: training (80% of the database), validation (10% of the database), and test (10% of the database), while the YOLO model allows up to the moment, only the use of the training subsets (80% of the database) and test (20% of the database).

Model pre-processing and selection

For the classification of pods, two deep-learning models were chosen: Mask R-CNN (He et al.,

Figure 2. Image captured from pods of a single plant using the image acquisition method

2018) and YOLOv4. The Mask R-CNN stands out for its efficiency in instance segmentation, being an extension of the Faster R-CNN model with the addition of mask prediction that can surround the object of interest, showing the exact location of the object in the image. This model has been widely used in agriculture, being a solution to problems in the agricultural sector (DE CARVALHO et al., 2021; Lee et al., 2020; MEKHALFI et al., 2021; VALICHARLA, 2021). In turn, YOLO (You Only Look Once) (REDMON et al., 2016) has several versions, with improvements and minor modifications among them. In general, the YOLO model is known for its high detection speed, a result of the simultaneous processing that takes place inside it, which allows the determination of the bounding box coordinates and classification of the objects of interest in a single moment (REDMON et al., 2016).

As supervised models require labels for their training, it was necessary to label each pod to generate a true bounding box. With this box, the model can identify the object of interest and extract the inherent characteristics of that object, allowing model learning. In this work, the objects of interest were the pods, which were labeled based on the number of grains they had: "one", "two", "three" and "four". The work consisted of using different models, so it was necessary to label the images in different software, as each model requires a different extension for the labels. In Figure 3a, the labeling for the Mask R-CNN model is presented, in which the online software VGG Image Annotator was used (DUTTA; ZISSERMAN, 2019). Figure 2b, on the other hand, shows the labeling for YOLOv4 with the LabelImg program (TZUTALIN, 2015).

The input data for training the Mask R-CNN model was composed of the validation and training set, while the YOLO model used a single training set. In both pieces of training, the transfer of learning technique was used, which consists of reusing basic information about images (color, edges, among others) of a model that had already been trained in a new problem. The intention is to reduce the training time as the model does not start from zero.
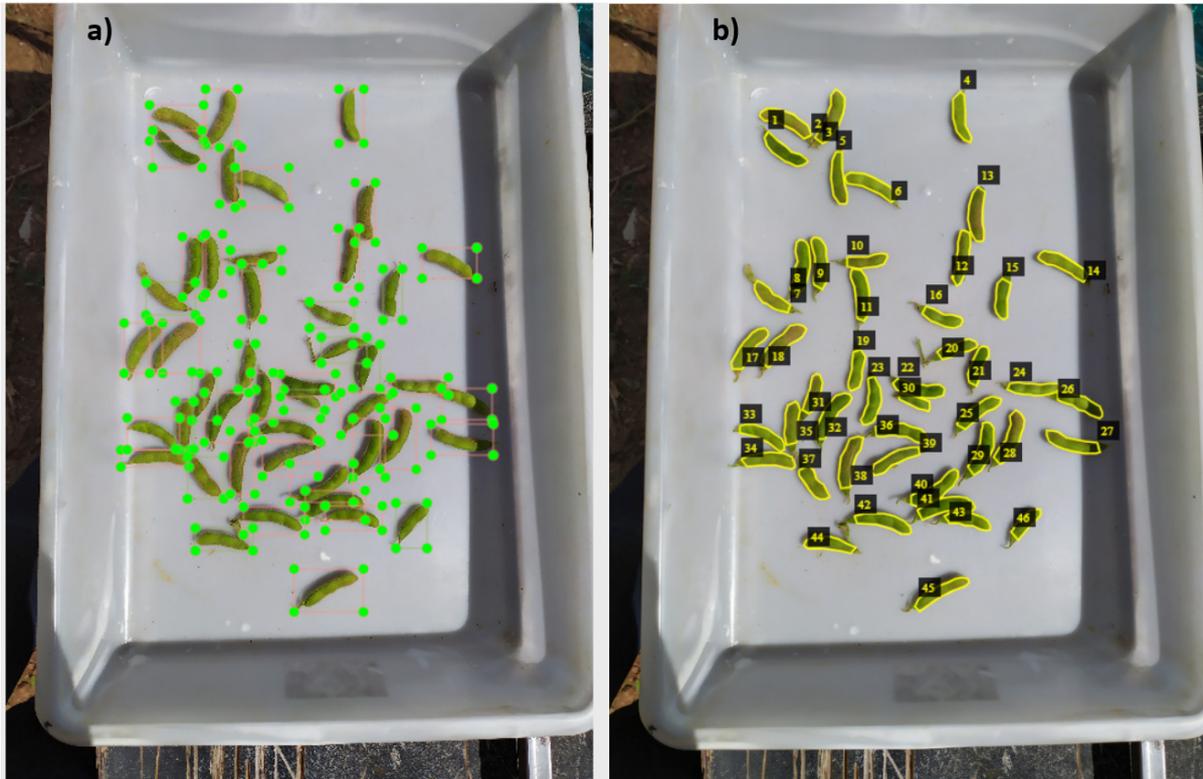
Figure 3. Labeling performed for the LabelImg for the YOLO model (a) and labeling performed in VGG Image Annotator for the Mask R-CNN model (b)

Evaluation of the models

The metrics of Precision (Equation 1), recall (Equation 3), and F1 score (Equation 3) were used to evaluate both models. Precision quantifies the proportion of predicted values that are true values. Recall measures the proportion of actual values that the model was able to classify correctly. Finally, the F1 score is the harmonic mean between the two values, precision and recall, meaning the average performance of a single class.

$$P = \frac{TP}{TP + FP} * 100 \tag{1}$$

$$R = \frac{TP}{TP + FN} * 100 \tag{2}$$

$$F1 = 2 * \frac{P * R}{P + R} * 100 \tag{3}$$

Where,
P = precision, %;
R = recall, %;
F1 = F1 score, %;
TP = True Positive, dimensionless;
FP = False Positive, dimensionless;
FN = False Negative, dimensionless.

RESULTS AND DISCUSSION

As shown in Figure 4, the result of the evaluative metrics using the Mask R-CNN model and Yolov4. In both models, it was observed that classes "three" and "two" achieved the best performance compared to classes "one" and "four". This result can be related to the number of pods of each class per plant. The cultivars used had a higher incidence of two- and three-grain pods. Therefore, the superiority in the classification may be the result of this higher incidence, providing a greater number of samples for training both models and causing the presented disparity.
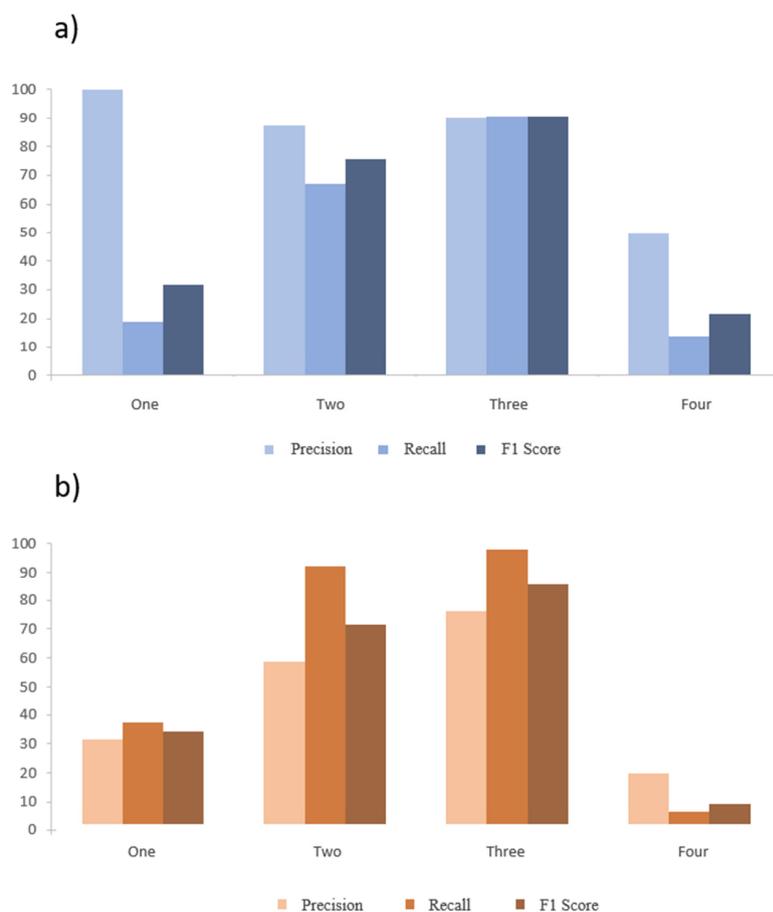
Figure 4. Performance of the Mask R-CNN (a) and YOLOv4 (b) models regarding the classification of soybean pods according to the number of grains they contain, using the Metrics of Precision, Recall, and F1 score

According to the results of Figure 4, it was observed that the Mask R-CNN model presented a value of 100% in class "one" accuracy. This metric informs the percentage of predictions that match the actual value. Thus, even without the correct prediction of all objects of this class, the model able to classify as class "one" was correct. On the other hand, when analyzing the recall, the opposite occurs, with a value of 18.75% (Figure 4a). This metric reports the percentage of actual values that were predicted correctly. At this point, the model failed because it classified incorrectly or did not detect all pods in class "one".

In the work carried out by Yu et al., (2019), the Mask R-CNN model was used to detect strawberry fruits, in which it reached 95.78% and 95.41% for accuracy and recall, respectively. In the experiment carried out by Ganesh et al., (2019), from different color channels, maximum values of 97.54%, 86.73%, and 88.67% were reached for accuracy, recall, and F1 score in the classification of oranges. Using YOLOv3-tiny, Mazzia et al. (2020) found the performance for detecting apple fruits, achieving results of 83% and 69% for accuracy and recall, respectively. In the aforementioned works, the object of interest had a visually different color from the vegetation. In the current work, however, the models were required to overcome the difficulty of classifying similar objects, whose differences were based on subtle characteristics

such as size and shape, which are less expressive traits in comparison to the color characteristic.

Among the models, a better performance was observed for the Mask R-CNN model. This fact may be due to the training time as they are models with different constitutions, it is possible that for YOLO a longer training time was required since the time of observation and extraction of information about each image is reduced in comparison to Mask R-CNN. However, it was observed that in some images, both models presented similar difficulties. As seen in Figure 5, it is observed that both models were not able to detect the same pod. The visible absence of grain boundaries in this pod may have been the factor that led to the models' inability to detect it correctly. This is because it can be an important characteristic for the models to classify them correctly.

Despite the differences found in the results, both models present satisfactory values (Figure 4) and are close to those in the literature (AFONSO et al., 2020; DAVIS et al., 2020; HUANG et al., 2020; UZAL et al., 2018; XU et al., 2020). Such a fact is important as the classification of soybean pods is a crucial factor for estimating productivity, as well as for the genetic improvement of the crop. By removing dependence on human sight, errors tend to be minimized, as the machine does not reduce functioning and does not get tired like human beings. In addition to being a method that
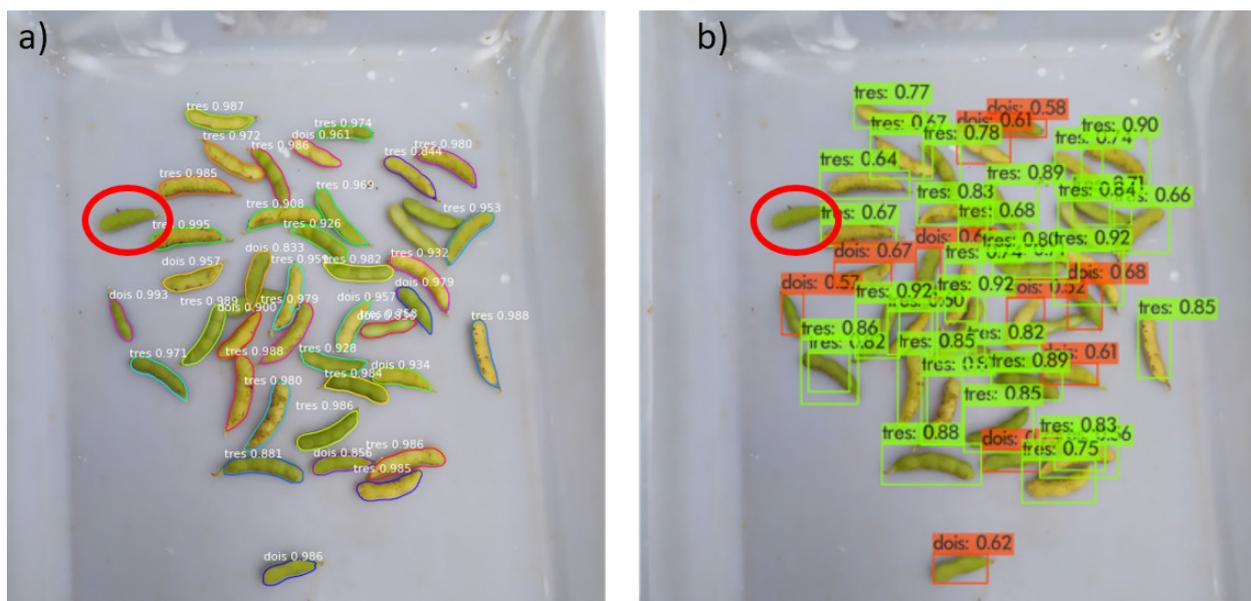
Figure 5. Classification with the use of Mask R-CNN (a) and YOLOv4 (b) on the test subset, where both failed to classify the same pod

will contribute to the reduction of errors, the work presented in this article uses a tool that is easy to be manipulated and access, which is the smartphone. This ensures simplicity in using the presented method. Thus, the future contributions of this research will be to improve the models to obtain the count of the number of pods and grains as a return, thus contributing to possible methods to estimate productivity.

Among the analyzed models, as mentioned, both achieved similar values. However, the primary difference between the models is the computational need. The Mask R-CNN model requires more computational demand compared to the YOLO model. For that reason, using YOLO for this occasion would be more affordable.

CONCLUSION

- Two deep learning models were analyzed and both show results greater than 70% (considering the F1 score). Furthermore, for classifying three-grain pods, both models showed accuracy above 80%. However, due to computational needs, the YOLO model stands out, making it more accessible. As a result, the potential for using deep learning associated with popular tools such as smartphones to obtain data is observed.

ACKNOWLEDGMENTS

AUTHORSHIP CONTRIBUTION STATEMENT

BANDEIRA, P. M. C.: Conceptualization, Data curation, Methodology, Software, Writing – original draft; **VILLAR, F. M. M.:** Funding acquisition, Project administration, Resources, Supervision, Writing – review & editing; **DIAS, I. A.:** Data curation, Formal Analysis, Investigation, Methodology; **BANDEIRA, P. P. C.:** Validation, Visualization, Writing – review & editing.

DECLARATION OF INTERESTS

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.

REFERENCES

AFONSO, M. et al. Tomato Fruit Detection and Counting in Greenhouses Using Deep Learning. **Frontiers in Plant Science**, v. 11, p. 1759, 19 nov. 2020.

ALVES, G. R. et al. Estimativa da produtividade de soja com redes neurais artificiais. **Acta Scientiarum - Agronomy**, v. 40, n. 1, p. 1-9, 2018.

CHAN, M.; WEI, F.; MOLIN, P. Soybean Yield Estimation and Its Components : A Linear Regression Approach. n. Ml, 2020.

DAVIS, C. C. et al. A New Method for Counting Reproductive Structures in Digitized Herbarium Specimens Using Mask R-CNN. **Frontiers in Plant Science**, v. 11, n. July, p. 1-13, 2020.

DE CARVALHO, O. L. F. et al. Instance segmentation for large, multi-channel remote sensing imagery using mask-RCNN and a mosaicking approach. **Remote Sensing**, v. 13, n. 1, p. 1-24, 2021.

DUTTA, A.; ZISSERMAN, A. The VIA annotation software for images, audio, and video. **MM 2019 - Proceedings of the 27th ACM International Conference on Multimedia**, p. 2276-2279, 2019.

ETIENNE, A. et al. Deep learning-based object detection system for identifying weeds using uas imagery. **Remote Sensing**, v. 13, n. 24, p. 1-22, 2021.

GANESH, P. et al. Deep Orange: Mask R-CNN based Orange Detection and Segmentation. **IFAC-PapersOnLine**, v. 52, n. 30, p. 70-75, 2019.

HE, K. et al. Mask R-CNN. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 42, n. 2, p. 386-397, 2020.

HUANG, Z. et al. DC-SPP-YOLO: Dense connection and spatial pyramid pooling based YOLO for object detection. **Information Sciences**, v. 522, p. 241-258, 2020.

LEE, J. et al. Artificial intelligence approach for tomato detection and mass estimation in precision agriculture. **Sustainability (Switzerland)**, v. 12, n. 21, p. 1-15, 2020.

MAUAD, M.; BERTOLOZE SILVA, T. L.; ALMEIDA NETO, A. I.; ABREU, V. G. Influência da densidade de semeadura sobre características agronômicas na cultura da soja. **Agrarian**, [S. l.], v. 3, n. 9, p. 175-181, 2011.

MAZZIA, V., Khaliq, A., Salvetti, F., & Chiaberge, M. (2020). Real-time apple detection system using embedded systems with hardware accelerators: An edge AI application. **IEEE Access**, 8, 9102-9114.

MEKHALFI, M. L. et al. Detecting crop circles in google earth images with mask R-CNN and YOLOv3. **Applied Sciences (Switzerland)**, v. 11, n. 5, p. 1-12, 2021.

MILLER, J. J. et al. Characterizing soybean vigor and productivity using multiple crop canopy sensor readings. **Field Crops Research**, v. 216, n. November 2017, p. 22-31, 2018.

NGUGI, L. C.; ABELWAHAB, M.; ABO-ZAHHAD, M. Tomato leaf segmentation algorithms for mobile phone applications using deep learning. **Computers and Electronics in Agriculture**, v. 178, n. September, p. 105788, 2020.

RAMOS, P. J. et al. Automatic fruit count on coffee branches using computer vision. **Computers and Electronics in Agriculture**, v. 137, p. 9-22, 2017.

REDMON, J. et al. You only look once: Unified, real-time object detection. **Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition**, v. 2016- Decem, p. 779-788, 2016.

TEDESCO-OLIVEIRA, D. et al. Convolutional neural networks in predicting cotton yield from images of commercial fields. **Computers and Electronics in Agriculture**, v. 171, n. October 2019, p. 105307, 2020.

TZUTALIN. LabelImg. Git code (2015). https:// github.com/tzutalin/labelImg

UZAL, L. C. et al. Seed-per-pod estimation for plant breeding using deep learning. **Computers and Electronics in Agriculture**, v. 150, n. April, p. 196-204, 2018.

VALICHARLA, S. K. Weed Recognition in Agriculture : A Mask R-CNN Approach Weed Recognition in Agriculture : A Mask R-CNN Approach. 2021.

XU, B. et al. Automated cattle counting using Mask R-CNN in quadcopter vision system. **Computers and Electronics in Agriculture**, v. 171, n. October 2019, p. 105300, 2020.

YU, Y. et al. Fruit detection for strawberry harvesting robot in non-structural environment based on Mask-RCNN. **Computers and Electronics in Agriculture**, v. 163, n. June, p. 104846, 2019.

ZHU, S. et al. Identification of Soybean Varieties Using Hyperspectral Imaging Coupled with Convolutional Neural Network. **Sensors 2019,** Vol. 19, Page 4065, v. 19, n. 19, p. 4065, 20 set. 2019.