Inter-Row Soybean Plantation Identification in Images to Support Automatic Alignment of a Weeder Machine

Jailson Lucas Panizzon[1], André Roberto Ortoncelli[1], Alinne C. Correa Souza[1], Francisco Carlos M. Souza[1], Rafael Paes de Oliveira[1]

[1]Federal University of Technology – Paraná (UTFPR)
Dois Vizinhos, Paraná, Brazil

jailson.panizzon@gmail.com

{ortoncelli,alinnesouza,franciscosouza,raoliveira}utfpr.edu.br

Abstract. *This study explores a Computer Vision approach to identify inter-row planting in soybean areas. Related work already explores the same problem, but our work differs by focusing on inter-row identification to support the alignment of weeding machines (commonly used by small farmers who produce organic products). We created an experimental database with images collected with a camera attached to a weeder. The planting lines and inter-rows were manually labeled. To detect planting lines and inter-rows, we use two segmentation algorithms based on Convolutional Neural Networks (Mask R-CNN and YOLACT), achieving an accuracy of up to 0.656 with the interpolation of the obtained segmentation results. The segmentation results obtained made it possible to estimate the inter-rows satisfactorily. We provide a database of collected images, with the planting lines and inter-rows noted. With these results, we intend to create a solution in future work that allows automatic alignment of the weeder. We also plan to develop similar solutions for other crops (in addition to the soybeans explored in the experiments).*

1. Introduction

Different challenges are critical for sustainable agriculture, including food security, land degradation, climate change, and a growing population. One way to overcome these challenges and support sustainability is by using cutting-edge technology [Purcell et al. 2023]. Furthermore, to address population aging and the accelerating pace of life, traditional labor-intensive and risky agricultural work must be empowered by more automated control work for promising results [Cheng et al. 2023].

Autonomous and accurate navigation in agriculture is still challenging due to the complex and unstructured nature of the agricultural environment, which is a prerequisite for carrying out various tasks. With the evolution of electronics and information technology, machine vision has become a promising tool for accurate, real-time navigation. Due to the low hardware cost and the wealth of visual information, Computer Vision has been intensely studied and widely used for this purpose [Bai et al. 2023].

In this article, we call inter-rows the lines that separate the rows of crops (sequence of plants in the direction in which they were planted). Various agricultural vehicles/equipment travel/work in the inter-rows, such as, for example, weeders, inter-row seeders, sprayers, and straw rakes, among others [Kise and Zhang 2008]. Handling this equipment must be done carefully, as, in case of errors, damage to the cultivated plants may occur, causing losses and losses in productivity. Much of this kind of equipment is still controlled by human operators today.

Different studies reported related methods to the plantation's inter-row detection [Bai et al. 2023, Liang et al. 2022, Bonadies and Gadsden 2019, Basso and de Freitas 2020, Kanagasingham et al. 2020]. However, these technologies exist outside Brazil. The high cost of imports and the difficulty of technological transfer make using these systems difficult in our country, especially for small and medium-sized companies—Brazilian producers with little (or no) capital available for investment in implementations.

We started a project with a national company serving organic product producers to develop a new automatic weeding machine control technology. For this organic production, weeding is essential, as it allows weed control without pesticides. Therefore, this study focuses on detecting inter-rows in soybean plantations to enable the automatic alignment of weeder machines that are currently manually controlled by a human operator.

With the partnership, it was possible to obtain technical knowledge on the subject and acquire a base of images with a camera attached to a weeder under natural conditions of use. Our project aims to detect the inter-rows in the photos and, based on them, evaluate and correct the alignment of the equipment. In this study, we present the first results of this project, which indicate that it is possible to identify the inter-rows of planting and calculate the alignment using images acquired with a camera fixed to the weeder machine. We intend to use the results to automatically align the weeder, trying to remove the human operator. Furthermore, we also plan to replicate the technique in other crops, aiming to benefit small producers of organic products.

To segment between the lines, we explored two object segmentation techniques based on Convolutional Neural Network (CNN): Mask R-CNN [He et al. 2017] and YOLACT [Bolya et al. 2019]. Based on the segmentation results that reached an accuracy of 0.656 (with Mask R-CNN), we estimated a single line representing the direction of the planting row using a simple regression. An empirical analysis of the results makes it clear that they are satisfactory for the problem.

In addition to the experimental results presented, one of the contributions of this work is to make publicly available the experimental database used in the experiments, which includes images of soybean plantations collected with a camera attached to a weeder and with the labeling of the lines and between the lines.

The remainder of this text is organized as follows. Section 2 presents theoretical aspects that will be fundamental for understanding this work. Related work is in Section 3. The experimental environment and results are in Section 4. Section 5 presents owner results and discussions. Finally, Section 6 presents the conclusion and perspectives for future work.

2. Theoretical Aspects

This Section presents theoretical concepts related to object segmentation that are necessary to understand this study. Subsection 2.1 describes the concept of Image segmentation. Subsections 2.2 and 2.3 show the Mask R-CNN and YOLACT segmentation frameworks, respectively.

2.1. Image Segmentation

Segmentation involves the subdivision of a digital image into multiple regions or sets of pixels with similar characteristics, which may include classifying pixels with semantic labels (semantic segmentation), partitioning individual objects (instance segmentation), or both (panoptic segmentation).

Image segmentation is a fundamental task in Computer Vision with numerous important applications, such as identifying lesions for the medical field [Dias et al. 2023], plants and weeds in agriculture [Champ et al. 2020], objects in aerial images [Chakravarthy et al. 2022], among others, including inter-row planting, which is explored in this work (Section 3).

Segmentation algorithms have evolved a lot in recent years. Considering early methods such as threshold, k-means clustering, and watershed methods, improved results tend to be obtained with more advanced algorithms such as active contours, graph cuts, and conditional and random Markov fields [Haralick and Shapiro 1985]. In recent years, however, Deep Learning (DL) models based on Convolutional Neural Networks (CNN) have produced a new generation of segmentation models with notable performance improvements, often achieving the highest accuracy rates on popular benchmarks [Barbosa and Osório 2023, Minaee et al. 2021].

For this study, we selected two methods based on Deep Learning (Mask R-CNN and YOLACT) , which will presented in the following subsections.

2.2. Mask R-CNN

Mask R-CNN is a simple, flexible, and general framework for segmenting object instances, which consists of an extension of Faster R-CNN (a framework for object detection based on CNN) [Ren et al. 2016].

Mask R-CNN extends Faster R-CNN by adding a branch to predict an object mask in parallel with the existing branch for bounding box recognition; it can efficiently detect objects in an image and, at the same time, generate a segmentation mask of high-quality segmentation for each instance. Mask R-CNN is simple to train and adds only a small overhead to Faster R-CNN. Furthermore, Mask R-CNN is easy to generalize to other tasks.

When Mask R-CNN was proposed [He et al. 2017], it outperformed previous benchmarks in the COCO object instance segmentation challenge [Lin et al. 2014], efficiently detecting objects in an image and simultaneously generating a high-performance segmentation mask for each instance. Since then, Mask R-CNN has been applied to different segmentation problems [Bharati and Pramanik 2020].

2.3. YOLACT

YOLACT follows a similar principle to Mask R-CNN. While Mask R-CNN is based on the Faster R-CNN object detector, YOLACT modifies YOLO detector. One of the primary motivations for creating YOLACT is because instance segmentation methods require high computational power, making it difficult to use them in real-time application [Bolya et al. 2019].

The greater speed of YOLOACT is because it adapts a one-step object detection algorithm, unlike Mask R-CNN, which is an adaptation of a two-step algorithm. Details of the one-and two-step detection algorithms are below [Zou et al. 2023]:

- Two-stage algorithms laid the groundwork for object detection algorithms based on Deep Learning. Initially, they pinpoint potential target regions, followed by classification in a subsequent step. While known for high accuracy, these methods typically exhibit limited detection speed;
- One-stage algorithms improve detection speed, making them ideal for real-time applications on mobile devices with easy deployment. However, despite their rapid processing capabilities, they often need help to perform well when detecting dense or small objects.

YOLO is one of the leading one-step object detectors. The first YOLO version (YOLOv1) was proposed in 2006 [Redmon et al. 2016], and today it is in version 9 (YOLOv9). Notably, most versions of YOLO can achieve segmentation quality equal to or better than the two-step algorithms [Wang et al. 2024].

3. Related Works

Autonomous navigation of robots and agricultural vehicles in agricultural environments is a prerequisite for various tasks. However, accurate navigation of farming robots is still challenging due to the farm environment's complex and unstructured nature [Bai et al. 2023]. Crop row navigation is typically accomplished using vision-based cameras and global positioning system (GPS) units [Bonadies and Gadsden 2019].

Machine vision strategies were implemented to detect contours and edges of crop lines to ensure proper navigation of lines without damaging crops in recent work [Liang et al. 2022, Basso and de Freitas 2020, Kanagasingham et al. 2020]. In Liang et al. (2022), with a camera positioned on the front bumper of a tractor (at a height more significant than the height of the cotton plant at an angle of 65 degrees), images of cotton plantations were collected, which were used to identify inter-row plantings. Images collected at a rate of 30 frames per second and with a resolution of 640×480 were converted to gray scale, and cotton lines were segmented using the OTSU method [Otsu et al. 1975], which was detected at the edges using the Canny algorithm, respectively.

In Kanagasingham et al. (2022), data from Global Navigation Satellite Systems (GNSS), compass, and machine vision were integrated to create an autonomous navigation solution for a rice field weeding robot. A new crop row detection algorithm was developed to utilize images. A low-cost action camera was mounted one meter above the ground in front of the robot and oriented at a tilt angle of -35° to the horizontal. The video captured by the camera was live-streamed via the high-definition multimedia interface (HDMI) to a laptop computer. The software used for processing the images was developed based on the OpenCV library [Bradski and Kaehler 2008].

Basso and de Freitas (2020) proposed a guidance system based on digital image processing for unmanned aerial vehicles. The software consists of two algorithms. The first algorithm is Crop Row Detection, which correctly identifies crop rows. The second algorithm is the Filter Line, which generates the drive parameters sent to the flight controller. The video captured by the Raspberry Pi camera comprises frames or sequential images that provide the sensation of movement. The image is transformed to a gray level and binarized; then, crop row identification is identified using the Hough Transform [Illingworth and Kittler 1988].

Our work follows the principles of row and inter-row detection used in the literature. However, our study differs due to the position of the camera and the fact that it focuses efforts on a weeder, which is a lower-cost piece of agricultural equipment, and a solution in this context will mainly benefit small producers of organic products. We chose to use methods based on deep learning to segment the inter-rows and rows to adequately deal with different lighting conditions caused by the tractor shadow (Figures 2 and 6), and because modern versions of these algorithms (such as YOLACT) allow its execution in real-time.

4. Experimental Environment and Results

We built an experimental dataset with a GoPro HERO 7 camera attached to a weeder machine positioned facing the crop rows in parallel to the tractor at a height of 130cm from the ground and an angle of 30° as shown in Figure 1.

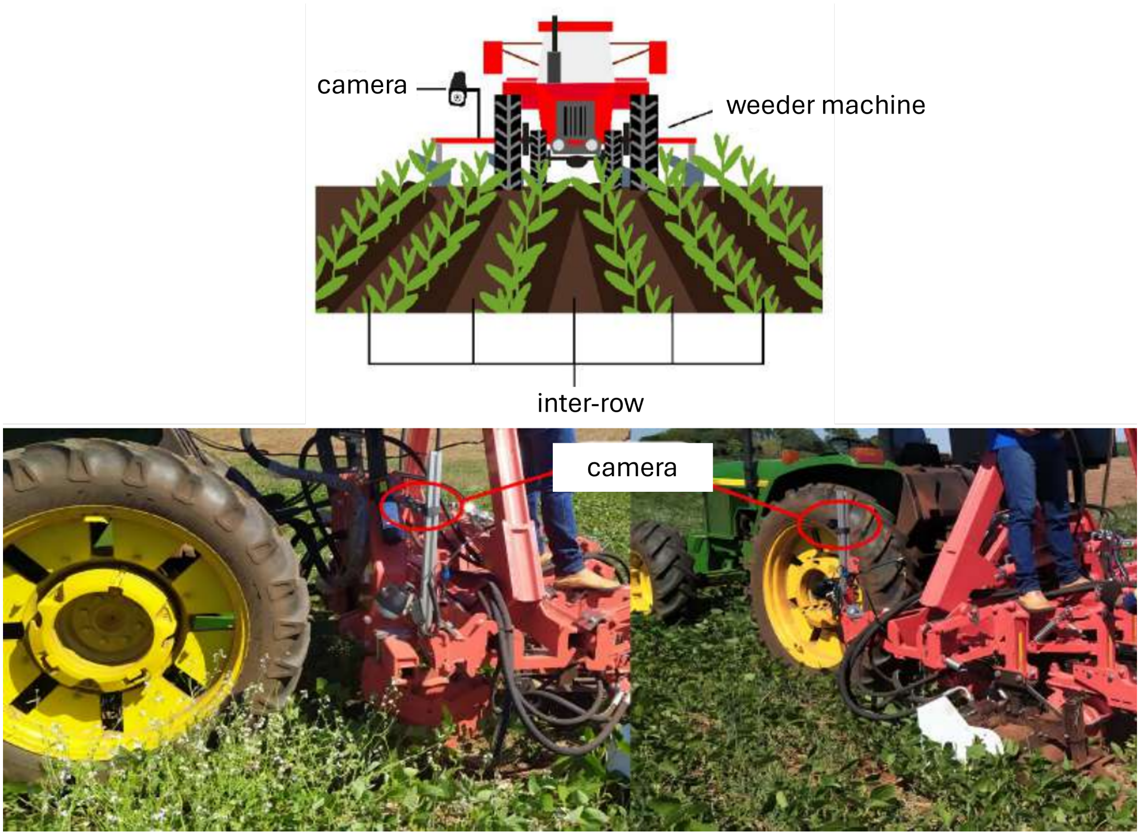

Figure 1. Camera position – experimental configuration.

The videos were recorded between 9:00 am and 5:00 pm, at a rate of 60 FPS and 720p resolution, on sunny days with a maximum cloud cover of 40% and no rain for at least 5 (five) hours antecedents, with the tractor in motion carrying the weeder machine and carried out the work of weeding (controlled by an operator who directed the equipment's handling lines)

The crop selected for the experiments was soybeans, sown with a planter from 40 cm to 50 cm between the rows, cultivated in Paraná (Brazil). The videos were recorded in a soybean plantation with heights varying between 25cm and 50cm and with infestation of invasive plants.

We resized the video frames to 480 x 270 pixels and selected one for every 120 frames to compose the experimental database. In a second step, manually, we remove identical (or very similar) images. We remove remarkably similar sequence images. At the end of this process, a set of 1117 images was selected.

We manually annotated the region of rows and inter-rows for each selected image. The VGG Image Annotator tool [Dutta and Zisserman 2019] was used for the marcation process. An example of a marked image is shown in Figure 2 - the manually added lines identify the contour of the planting lines (separating planting rows and each inter-rows).

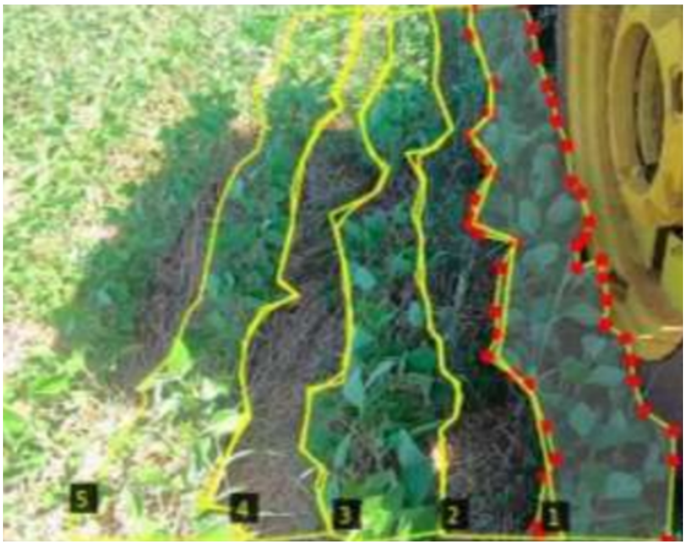

Figure 2. Dataset labeling process with VGG Image Annotator.

We divided the image set into training (70%) and testing (30%) sets. Considering the training set, we used 70% of the images for training and 30% as validation. This division is illustrated in Figure

All 1117 images of our experimental dataset, along with the annotations (in the formats used by Mask R-CNN and YOLACT) and the divisions between test, training, and validation sets that we used, are public available[1]

[1]`https://nuvem.utfpr.edu.br/index.php/s/uJcqzWy6rOV3COC`

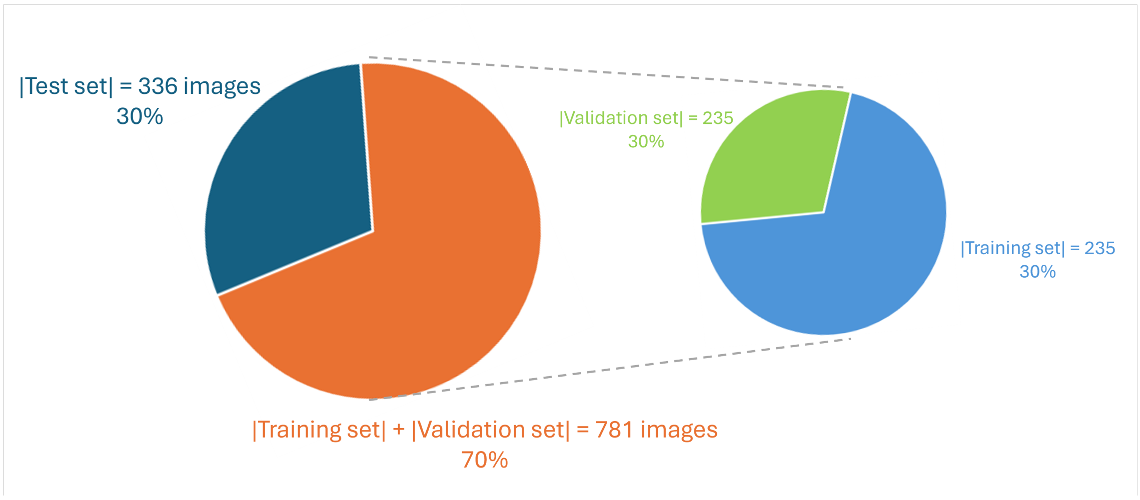

Figure 3. Composition of test, training and validation sets.

5. Experimental Results

We trained, validated, and tested Mask R-CNN and YOLACT with the experimental database. To validate the results, we used two metrics, precision, and recall, presented in Equations 1 and 2, which use the following concepts:

- True Positives (TPs): the pixels of a class A object correctly assigned to class A;
- False Positives (FPs): pixels attributed to an object of class A that do not belong to class A;
- True Negatives (TNs): pixels that do not belong to class A and were not assigned to class A; and
- False Negatives (FNs): pixels that were not assigned to class A but belong to class A.

$$Precision = \frac{TPs}{TPs + FPs} \tag{1}$$

$$Recall = \frac{TPs}{TPs + FNs} \tag{2}$$

The precision and recall results obtained in the tests carried out for the two algorithms are presented in Table 1.

Table 1. Experimental Segmentation Results.

Method	Precision	Recall
Mask R-CNN	0.656	0.437
YOLACT	0.476	0.294

Figures 4 and 5 present results of segmenting an image from the database using Mask R-CNN and YOLACT, respectively.

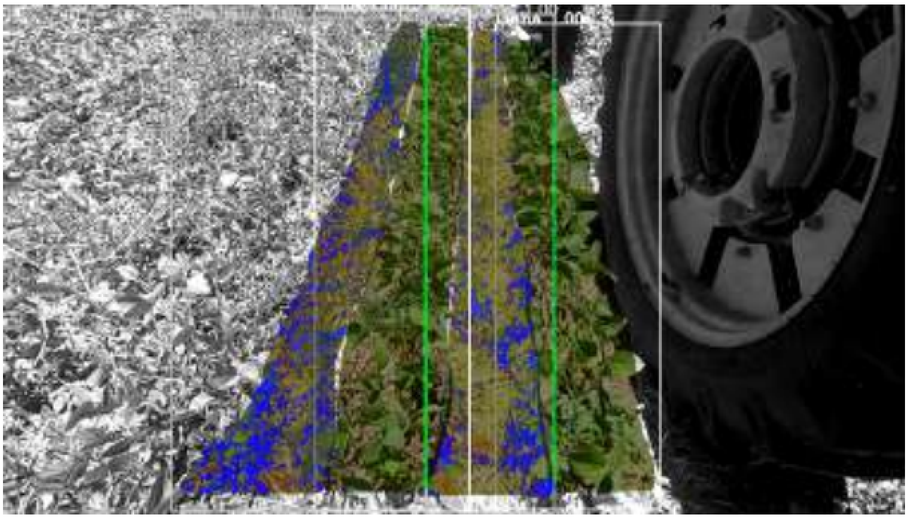

Figure 4. Mask R-CNN segmentation results.

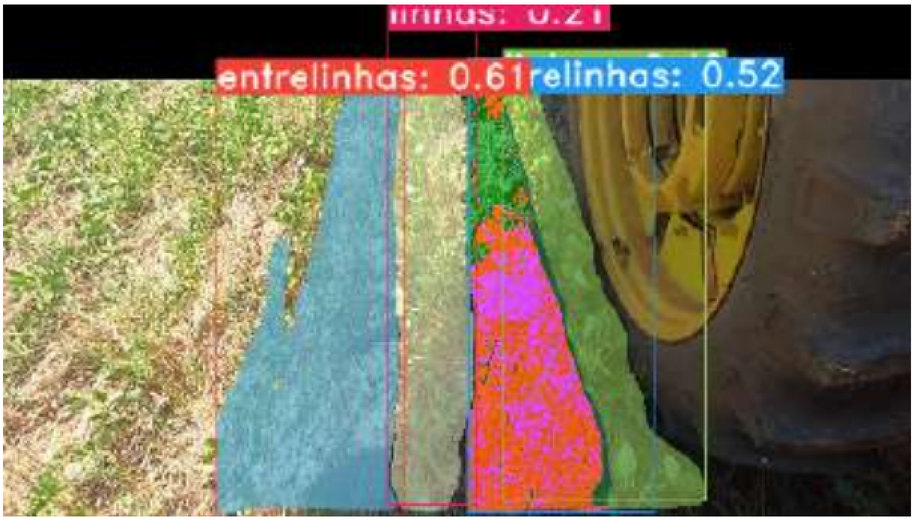

Figure 5. YOLACT segmentation results.

Even though the best segmentation accuracy result obtained is only 0.656 (with Mask R-CNN), with the segmented pixels, it was possible to satisfactorily infer (with a simple regression) lines that identify the direction of the rows and inter-rows. Figure 6 presents an example of the inference carried out with the segmentation results obtained with Mask R-CNN.

Figure 6. Rows and Inter-rows inference based on Mask R-CNN results.

6. Conclusion

This study presents an approach capable of identifying rows and inter-rows in organic soybean plantations based on images collected by a camera fixed to a weeder machine. The results can support the implementation of a solution that automatically aligns the weeder, eliminating the need for a human operator.

Suppose a solution for weeder machine alignment is developed. In that case, it should reduce costs for farmers producing organic products (one of the main target audiences for weeders) and reduce the demand for labor for this task, which has proven to be increasingly scarce.

In addition to implementing a solution that automatically aligns the weeder, we hope our results will motivate other work, such as improving the segmentation results obtained and developing techniques to identify inter-rows for other crops at different stages of plant growth (especially methods that benefit small organic product farms), and also the development of strategies that work correctly also at night.

References

Bai, Y., Zhang, B., Xu, N., Zhou, J., Shi, J., and Diao, Z. (2023). Vision-based navigation and guidance for agricultural autonomous vehicles and robots: A review. *Computers and Electronics in Agriculture*, 205:107584.

Barbosa, F. M. and Osório, F. S. (2023). Estudo de estratégia de aprendizado auto-supervisionado para aprimoramento da consistência temporal em modelo de segmentação semântica baseado em deep learning. In *Seminário Integrado de Software e Hardware*, pages 1–12.

Basso, M. and de Freitas, E. P. (2020). A uav guidance system using crop row detection and line follower algorithms. *Journal of Intelligent & Robotic Systems*, 97(3):605–621.

Bharati, P. and Pramanik, A. (2020). Deep learning techniques—r-cnn to mask r-cnn: a survey. *Computational Intelligence in Pattern Recognition*, pages 657–668.

Bolya, D., Zhou, C., Xiao, F., and Lee, Y. J. (2019). YOLACT: Real-time instance segmentation. In *IEEE/CVF International Conference on Computer Vision*, pages 9157–9166.

Bonadies, S. and Gadsden, S. A. (2019). An overview of autonomous crop row navigation strategies for unmanned ground vehicles. *Engineering in Agriculture, Environment and Food*, 12(1):24–31.

Bradski, G. and Kaehler, A. (2008). *Learning OpenCV: Computer vision with the OpenCV library*. " O'Reilly Media, Inc.".

Chakravarthy, A. S., Sinha, S., Narang, P., Mandal, M., Chamola, V., and Yu, F. R. (2022). Dronesegnet: Robust aerial semantic segmentation for uav-based iot applications. *IEEE Transactions on Vehicular Technology*, 71(4):4277–4286.

Champ, J., Mora-Fallas, A., Goëau, H., Mata-Montero, E., Bonnet, P., and Joly, A. (2020). Instance segmentation for the fine detection of crop and weed plants by precision agricultural robots. *Applications in plant sciences*, 8(7):e11373.

Cheng, C., Fu, J., Su, H., and Ren, L. (2023). Recent advancements in agriculture robots: Benefits and challenges. *Machines*, 11(1):48.

Dias, M., Santos, C., Aguiar, M., Welfer, D., Pereira, A., and Ribeiro, M. (2023). Um novo método baseado em detector de dois estágios para segmentação de instância de lesões retinianas usando o modelo mask r-cnn e a biblioteca detectron2. In *Seminário Integrado de Software e Hardware*, pages 1–12. SBC.

Dutta, A. and Zisserman, A. (2019). The via annotation software for images, audio and video. In *ACM International Conference on Multimedia*, pages 2276–2279.

Haralick, R. M. and Shapiro, L. G. (1985). Image segmentation techniques. *Computer vision, graphics, and image processing*, 29(1):100–132.

He, K., Gkioxari, G., Dollár, P., and Girshick, R. (2017). Mask R-CNN. In *IEEE International Conference on Computer Vision*, pages 2961–2969.

Illingworth, J. and Kittler, J. (1988). A survey of the hough transform. *Computer vision, graphics, and image processing*, 44(1):87–116.

Kanagasingham, S., Ekpanyapong, M., and Chaihan, R. (2020). Integrating machine vision-based row guidance with gps and compass-based routing to achieve autonomous navigation for a rice field weeding robot. *Precision Agriculture*, 21(4):831–855.

Kise, M. and Zhang, Q. (2008). Development of a stereovision sensing system for 3d crop row structure mapping and tractor guidance. *Biosystems Engineering*, 101(2):191–198.

Liang, X., Chen, B., Wei, C., and Zhang, X. (2022). Inter-row navigation line detection for cotton with broken rows. *Plant Methods*, 18(1):90.

Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., and Zitnick, C. L. (2014). Microsoft coco: Common objects in context. In *uropean Conference on Computer Vision*, pages 740–755. Springer.

Minaee, S., Boykov, Y., Porikli, F., Plaza, A., Kehtarnavaz, N., and Terzopoulos, D. (2021). Image segmentation using deep learning: A survey. *IEEE transactions on pattern analysis and machine intelligence*, 44(7):3523–3542.

Otsu, N. et al. (1975). A threshold selection method from gray-level histograms. *Automatica*, 11(285-296):23–27.

Purcell, W., Neubauer, T., and Mallinger, K. (2023). Digital twins in agriculture: Challenges and opportunities for environmental sustainability. *Current Opinion in Environmental Sustainability*, 61:101252.

Redmon, J., Divvala, S., Girshick, R., and Farhadi, A. (2016). You only look once: Unified, real-time object detection. In *IEEE conference on computer vision and pattern recognition*, pages 779–788.

Ren, S., He, K., Girshick, R., and Sun, J. (2016). Faster r-cnn: Towards real-time object detection with region proposal networks. *IEEE transactions on pattern analysis and machine intelligence*, 39(6):1137–1149.

Wang, C.-Y., Yeh, I.-H., and Liao, H.-Y. M. (2024). Yolov9: Learning what you want to learn using programmable gradient information. *arXiv preprint arXiv:2402.13616*.

Zou, Z., Chen, K., Shi, Z., Guo, Y., and Ye, J. (2023). Object detection in 20 years: A survey. *Proceedings of the IEEE*.