# Corn Plant Counting Using Deep Learning and UAV Images

Bruno T. Kitano, Caio C. T. Mendes, André R. Geus, Henrique C. Oliveira, and Jefferson R. Souza

*Abstract*—**The adoption of new technologies, such as unmanned aerial vehicles (UAVs), image processing, and machine learning, is disrupting traditional concepts in agriculture, with a new range of possibilities opening in its fields of research. Plant density is one of the most important corn *(Zea mays L.)* yield factors, yet its precise measurement after the emergence of plants is impractical in large-scale production fields due to the amount of labor required. This letter aims to develop techniques that enable corn plant counting and the automation of this process through deep learning and computational vision, using images of several corn crops obtained using a low-cost unmanned aerial vehicle (UAV) platform assembled with an RGB sensor.**

*Index Terms*—**Deep learning (DL), plant counting, precision agriculture.**

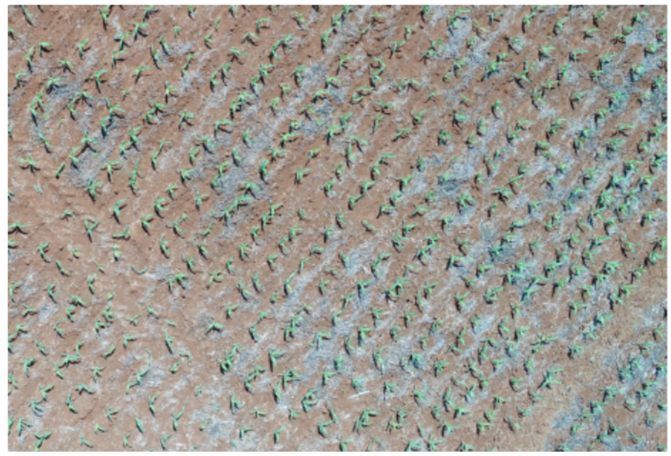

Fig. 1. Example of a corn crop image acquired by a UAV.

## I. Introduction

The demand for food in the world is growing and will be a significant challenge in the coming decades. With the accelerated pace of global population growth, especially in developing countries, it is estimated that by 2050, the population will reach 10 billion people, increasing the demand for food by 60% [1].

Several measures will be required to improve food security, such as enhancing food distribution globally and increasing the yield of current crops. It will be essential to have more efficiency in rural properties, increasing productivity per hectare without increasing the amount of arable area. Therefore, modernizing production systems and making them smarter and more efficient will play a vital role in this context. Such tools will be the basis for an unprecedented increase in productivity and will bring the farmer information and the control over variables previously unattainable.

Among the crops with significant global economic importance is corn, with a planted area of about 188.63 million hectares in 2018 [2]. It is estimated that by 2023, corn production in Brazil will grow 47% [3], being driven by constant technological advances and led by the digital transformation of the sector. The corn crop in Brazil has main destinations, human and animal consumption, being also used as a raw material to produce ethanol in countries such as the USA.

Plant density, stand, or number of individuals per hectare (Fig. 1) is one of the most important yield components in corn crops [4], as well as the number of ears per plant, grains per ear, and ear weight. Achieving the planned stand is a combination of various factors, such as the use of quality seeds, appropriate machinery, correct planting speed, and favorable weather conditions during sowing.

The quantification of plant density has essential applications in agricultural experimentation and for the farm management. For the former, this information is crucial in breeding programs, selection, and technical positioning of new hybrids, whereas for the latter, it is possible to analyze the planting quality, predict productivity, and make important decisions during the management of the crop and for the next seasons.

However, the possibilities to quantify this factor are limited. The most commonly used method is the manual measurement, but cost and labor are restrictive factors for greater accuracy and larger areas. Aiming this problem, digital approaches are being considered as solutions in this work, making use of tools such as unmanned aerial vehicles (UAVs), artificial intelligence, and digital image processing (DIP). The use of UAVs in agricultural properties has been growing exponentially [1], with a greater focus on image acquisition for plant counting and analysis of the installed crops. In this application, the UAV flies over a predetermined area of the farm, and then, the visual computing system brings a vital part of the desired information [2], [3].

Manuscript received February 4, 2019; revised June 4, 2019 and July 11, 2019; accepted July 14, 2019. This work was supported in part by UNICAMP, Federal University of Uberlândia (UFU) and in part by CNPq under Grant 400699/2016-8. *(Corresponding author: Jefferson R. Souza.)*

B. T. Kitano, C. C. T. Mendes, A. R. Geus, and J. R. Souza are with the Faculty of Computing, Federal University of Uberlândia, Uberlândia 38408-100, Brazil (e-mail: brunotkitano@ufu.br; caioctmendes@gmail.com; geus.andre@ufu.br; jrsouza@ufu.br).

H. C. Oliveira is with the School of Civil Engineering, Architecture and Urban Planning, University of Campinas, Campinas 13083-970, Brazil (e-mail: oliveira@fec.unicamp.br).

Color versions of one or more of the figures in this letter are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/LGRS.2019.2930549

After the images are captured and corrected considering camera tilt and relief displacement, a product called orthomosaic is generated from the studied area, aiming to have the general image of the field in a single scale. The orthomosaic, in turn, can be used to visually evaluate the field, which will bring context about its state and quality.

Through the orthomosaic, it is possible to evaluate the parameters of low complexity, such as planting quality, failures, uniformity of the planted crop, and occurrence of points of attention, such as flooding caused by excessive rainfall. However, the visual analysis carried out by a human operator in this context is limited and cannot quantify important information of greater complexity, such as the number of plants in a certain region, total area of failures, and altimetric variations in the field. Due quantification of these indexes is only possible with the application of computational techniques on these images, in the case of extensive planting areas.

The purpose of this research is to apply machine learning (ML) techniques, more specifically, deep learning (DL) to measure the final population of plants in corn crops using RGB aerial images collected by a low-cost UAV. The main contributions of this work are as follows.

1) Develop a tool for plant counting of corn crops on commercial farming using aerial images.
2) Propose an approach for counting plants using DL.
3) Define the technical aspects of corn plants that may be the limiting factors in the adoption of AI methodologies for plant counting, such as the vegetative stage of the plants, flight heights, and plant densities.

The remainder of this letter is organized as follows: Section II will cover related works regarding this topic, Section III will detail the proposed methodology, and results and conclusion will be discussed in Sections IV and V, respectively.

## II. Related Work

This session describes the related works in the area that has similarities with the proposed method, counting of plants using ML techniques and computational vision.

Ribera *et al.* [5] detail the procedure of counting sorghum plants using convolutional neural networks (CNN), adopting the architectures AlexNet [6], Inception-v2 [7], Inception-v3 [8], and Inception-v4 [9] with some modifications, to estimate the number of plants in the image using regression instead of classification. From the last layer of the network, which consists of a single neuron indicating the number of plants in the image, the plant count is regressed. The best result was obtained using the Inception-v3 architecture, with a mean absolute percentage error (MAPE) of 6.7%.

In a similar work carried out in [10], eucalyptus plants are detected and counted using the Inception-v2, ResNet 101, and Inception ResNet V2 architectures. The ResNet 101 model obtained the best accuracy, but the author denotes the need to use it after post-DIP, due to the computational demand generated by the model. In this context, the Inception-v2 network demonstrated the highest level of accuracy per processing time among the three models, presenting results only 1.3% less accurate than ResNet 101, but with results generated seven times faster. Owing to the low computational cost, it has mentioned the possibility of performing the processing and detection of the plants in the aircraft itself.

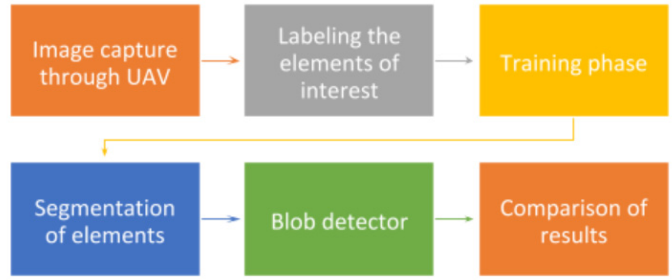


Fig. 2. Diagram of the methodology. Data flow is represented by arrows.

In turn, Shrestha and Steward [11] use computational vision to automate the counting of plants in corn, developing a system of sensing plant population with images obtained by a terrestrial vehicle. Gndinger and Schmidhalter [10] explore the digital count of maize plants using DIP methodologies, first isolating the corn plants of collected aerial images through color spectral differences and then correlating the number of green pixels present in the image with manual counting of the plants.

Fan *et al.* [12] proposed a new algorithm based on deep neural networks to detect tobacco plants in images captured by UAVs. Three stages were considered in this approach: first, the extraction of several candidate tobacco plants using morphological operations and the watershed segmentation. Then, the classification of candidate tobacco and nontobacco plants was made using a deep convolutional network, followed by a step of postprocessing to improve the exclusion of nontobacco plant regions.

The proposed approach uses images taken from RGB sensors, embedded in a low-cost UAV, as opposed to some works mentioned in this section, which uses images collected with multispectral sensors. In this letter, we also present a novel method of plant counting using DL in different flight heights, plant densities, and growth stages, through the application of the U-Net architecture, enabling training and classification with a relatively small data set.

## III. Proposed Methodology

Fig. 2 shows the techniques applied to perform the counting of corn plants using aerial images acquired from an RGB camera in a UAV platform.

The images were captured from the fields located in the Triângulo Mineiro region of Minas Gerais in Brazil and were composed of corn crops, as shown in Fig. 1. The UAV used in this work was the DJI Phantom 4 Pro. Its RGB camera has a 1-in CMOS sensor and 20 megapixels.

The elements of interest (corn and ground) were labeled using the software LabelMe [13] to create the training and validation data sets. For these two data sets, 634 labels for corn plants and 271 labels for ground were annotated, between 29 images.

The CNN architecture used to perform the segmentation of the images was the U-Net [14]. It is characterized by having two stages or sections, in which the first performs a series of convolutions and max-pooling layers to incrementally decrease the spatial resolution of the input while increasing the number of feature channels. This initial section is very similar to the architecture proposed in [15]. Section II uses convolutions and transposed convolutions, essentially reverting the effects of the first part, reducing feature channels, and increasing the resolution. One relevant characteristic of this architecture is the use of "skip" connections, which concatenates channelwise the results of convolutions performed before each max-pooling to the results of each transposed convolution. These skip connections are matched so that the result before the first max-pool layer is concatenated to the result of the last transposed convolution. The benefit of such connections is that they allow for the network to learn an initial result in the first few training steps without the need for modifying most network parameters. In others words, they allow the training process to be incremental, where the parameters present in the middle of the architecture can be the last to be learned and, as a result, they help to avoid local minima.

The chosen architecture uses a $3 \times 3$ filter for convolutions except for the last layer and $2 \times 2$ windows for max-pooling and transposed convolutions layers. Every convolution is followed by a rectified linear unit. After the last convolution layer, we have a final $1 \times 1$ convolution, where the number of filters is 2, since we have two classes (corn and ground). Hence, the result of the network is two matrices, with the same spatial size as the input where each one represents the score for each class. During inference, the softmax function is applied channelwise transforming the score into a probability distribution. During training, we use the cross-entropy loss.

After obtaining the corn/ground segmentation, we find the plants close to each other, and then, we used an opening morphological operator [16], [17] to separate the objects. In a morphological operation, a given "structuring element" is applied, transforming a nonsegmented image (input/original image) into a segmented image (output). The opening operation is an erosion followed by dilation and can be used to separate objects in the image, removing noises and protrusions.

Finally, blob detection methodologies (based on computational vision) were applied over the segmented images for the quantification of the number of corn plants. The blob detector [18], [19] aims to identify each connected group of segmented pixels, in our case representing the corn plant. Through the parameters of each blob, we identified the number of individuals present in the image. The manual benchmarking of the areas will be held as ground truth which is defined by agriculture experts and also for the efficiency comparison of the proposed method of work.

## IV. Experimental Results

The system used to acquire the images (Fig. 3) is equipped with an RGB camera (low spectral resolution—different from most of the approaches that use at least the infrared band) and an on-board global positioning system (GPS) receiver, powered by a lithium polymer battery, assembled with a system for navigation control, flight execution, and sensor activation.

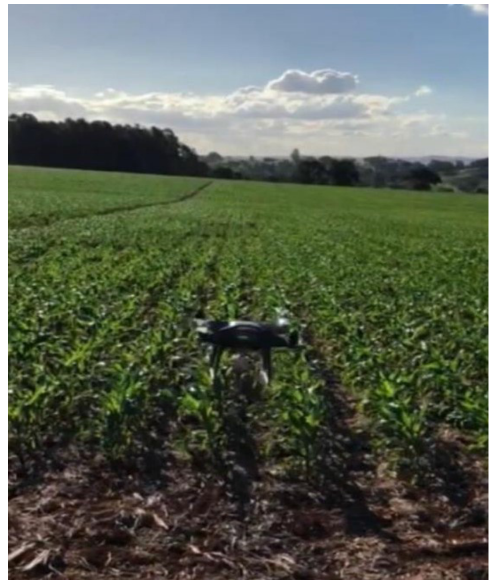

Fig. 3. DJI Phantom 4 Pro robotic platform used in the experiments.

Aerial images were captured with a 1" 20MP CMOS sensor, outputting images of $5472 \times 3648$ pixels, with a flying height varying between 1 and 20 m to obtain a range of different situations for the system to be tested.

### A. Training, Model Selection, and Testing

A predefined route was determined for the UAV to capture the images. The data set consists of 50 images, of which 25 were used for training, four for validation, and 21 for testing purposes. The training data set is composed of 513 labels created for corn plants and 215 for ground, while for validation, a total of 121 labels for corn plants and 56 labels for ground were created.

In the test data set, three different factors we analyzed to test our approach: plant density, flight altitude, and growth stage. The proposed methodology was tested with three different densities: 45 000, 70 000, and 90 000 plants per hectare; flight heights varying in three levels of 10, 15, and 20 m; and three different growth stages, V4, V6, and V8 (respectively, 4, 6, and 8 leaves). Three repetitions for each factor were considered to attest the consistency of the results, and some images of one factor could be used on another, resulting in a total of 21 images used in the test data set.

The training and validation data sets contained images of corn plants in growth stages between V2 and V8, with altitudes varying from 1 to 20 m. The labels were created using the software LabelMe (Fig. 4), in which the region samples provided for the optimization process follow a pixel-perfect pattern, that is, all the pixels of the given label were made sure to be related only to a specific object. In this proposed methodology, the nonannotated regions are not considered during the loss calculation in the optimization process of the weights of the neural network, thus implying in the training results and subsequently in an optimal network validation.

The training used the U-Net architecture, in which a given model was created using both training and validation data sets. The maximum resolution used for the training phase was $1696 \times 1344$ pixels due to the high processing demand and

TABLE I
RESULT OF THE CORN PLANT COUNTING ON TEST DATA USING THE PROPOSED METHODOLOGY

| Image | Factor | Altitude (meters) | Growth stage | Density of plants/hectare | Reference | Methodology | Residual | Residual (abs(%)) |
|---|---|---|---|---|---|---|---|---|
| 1 | Density | 15 | 4 | 45.731 | 75 | 71 | 4 | 5.3 |
| 2 | Density | 15 | 4 | 48.780 | 80 | 74 | 6 | 7.5 |
| 3 | Density | 15 | 4 | 47.560 | 78 | 71 | 7 | 9.0 |
| 4 | Density | 15 | 4 | 70.731 | 116 | 96 | 20 | 17.2 |
| 5 | Density | 15 | 4 | 73.170 | 121 | 103 | 18 | 14.9 |
| 6 | Density | 15 | 4 | 69.512 | 114 | 99 | 15 | 13.2 |
| 7 | Density | 15 | 4 | 91.463 | 150 | 70 | 80 | 53.3 |
| 8 | Density | 15 | 4 | 94.512 | 155 | 90 | 65 | 41.9 |
| 9 | Density | 15 | 4 | 93.902 | 154 | 100 | 54 | 35.1 |
| 10 | Flight height | 10 | 4 | 71.951 | 118 | 101 | 17 | 14.4 |
| 11 | Flight height | 10 | 4 | 72.560 | 119 | 105 | 14 | 11.8 |
| 12 | Flight height | 10 | 4 | 68.292 | 112 | 99 | 13 | 11.6 |
| 4 | Flight height | 15 | 4 | 70.731 | 116 | 96 | 20 | 17.2 |
| 5 | Flight height | 15 | 4 | 73.170 | 120 | 103 | 17 | 14.2 |
| 6 | Flight height | 15 | 4 | 69.512 | 114 | 99 | 15 | 13.2 |
| 13 | Flight height | 20 | 4 | 70.121 | 115 | 112 | 3 | 2.6 |
| 14 | Flight height | 20 | 4 | 70.121 | 115 | 106 | 9 | 7.8 |
| 15 | Flight height | 20 | 4 | 70.731 | 116 | 106 | 10 | 8.6 |
| 13 | Growth stage | 20 | 4 | 70.121 | 115 | 112 | 3 | 2.6 |
| 14 | Growth stage | 20 | 4 | 70.121 | 115 | 106 | 9 | 7.8 |
| 15 | Growth stage | 20 | 4 | 70.731 | 116 | 106 | 10 | 8.6 |
| 16 | Growth stage | 20 | 6 | 71.951 | 118 | 95 | 23 | 19.5 |
| 17 | Growth stage | 20 | 6 | 70.121 | 115 | 100 | 15 | 13.0 |
| 18 | Growth stage | 20 | 6 | 73.780 | 121 | 97 | 24 | 19.8 |
| 19 | Growth stage | 20 | 8 | 73.170 | 120 | 90 | 30 | 25.0 |
| 20 | Growth stage | 20 | 8 | 70.731 | 116 | 82 | 34 | 29.3 |
| 21 | Growth stage | 20 | 8 | 71.341 | 117 | 87 | 30 | 25.6 |

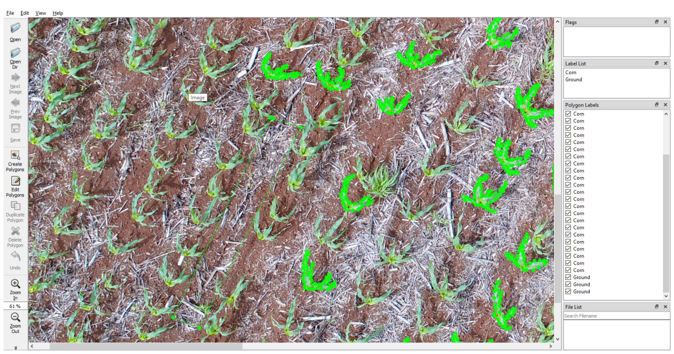

Fig. 4. Procedure for label creation using the software LabelMe.

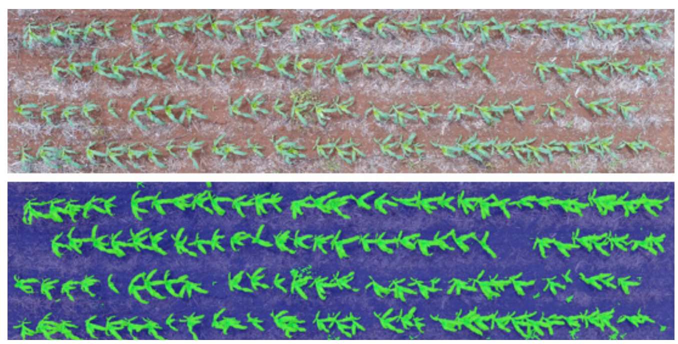

Fig. 5. Segmentation inference applied to the testing data.

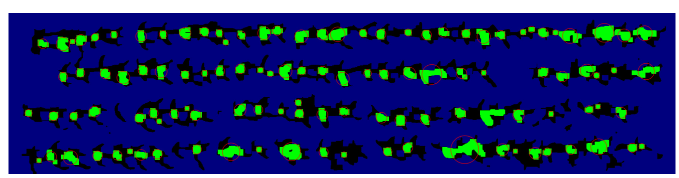

Fig. 6. Resultant image after the opening operator and blob detection.

limitation of the available hardware. As for the testing data sets, all images are in 4k resolution, but due to the flying height, the number of pixels per inch has a slight variation.

### *B. Segmentation Using U-Net Deep Learning*

Using the model generated, a segmentation inference is applied to the test data. The segmentation was done with the obtained training data to segregate corn plants from the ground and assigned a determined color to each one them—green for corn and blue for ground (Fig. 5). The training was finalized with a training loss of 0.03 and an accuracy of 0.994. The best validation accuracy obtained was 0.999 after 14 epochs.

The loss values are used to indicate the parameters that generated the best model. Smaller amounts are desirable and are calculated on the training and validation sets; then, it is expected to decrease with consecutive iterations or epochs. The validation accuracy indicates how accurate the model was when classifying pixels as corn plants and ground using the ground-truth data as a comparison.

### *C. Applying Morphological Operators and Blob Detection*

As noticed in Fig. 5, in some cases, the plants are very close to each other. To solve this, an opening morphological operator was used.

The total number of plants is obtained by applying a blob detector, which consists of the identification of a pixel group (area) that corresponds to the size of the plant. As the average area of corn plants is similar, this technique leads to a suitable approach for the counting phase. Fig. 6 shows the resultant image after the opening and blob detector algorithms.

### D. Discussion of Results

The manual plant counting (ground truth) and the results for the automatic counting are shown in Table I, along with the respective residual value (difference between the manual and the automatic counting, indicated as Reference and Methodology, respectively) obtained by the proposed method for each factor and tested images. The residual in percentage was calculated dividing the residual number by the ground truth.

For the plant density factor, the average residual was around 7.3% for the range of 45 000 plants, 15.1% for 70 000 plants, and 43.4% for 90 000 plants. Positive values occurred as a result of an incomplete detection, which can happen due to the presence of plants that were close to each other in the original image and, in this case, were segmented and identified as only one individual. The presence of less-developed corn plants that are close to each other represents a challenge for the opening operator because the area can be similar to a single more developed plant, not being possible to properly segregate these two objects. For this factor, we observe that our approach presented better results for lower plant densities.

Considering altitude, for flight heights of 10, 15, and 20 m, the average residual was 12.6%, 14.9%, and 6.4%, respectively, corresponding the best result for the highest flight height. Although higher spatial resolution and a, subsequently, higher level of detail are obtained in lower flight heights, to separate plants from the ground, it is not desirable to have a greater level of detail from one to the other, which happens in lower flight heights, because the ground has a greater homogeneity level from plants regarding to color and shade. Lower residual values were observed on a similar level of detail between these two features, ground and plants, which occurs on higher flight heights (lower scales) as confirmed by the obtained results.

For the growth stage, among the three different stages tested, the average residual was 6.4% considering 4 leaves, 17.5% for 6 leaves, and 26.7% for 8 leaves. For our approach, the best result was obtained with earlier growth stages, where we correlated the loss in performance due to the amount of overlap between leaves in more advanced growth stages.

Our methodology, in some cases, also identified incorrectly some plants that are not corn but have the same shape and color, such as weeds and straw. This denotes the importance to label other objects present in the image and train the U-Net to segment these objects for better results.

The best overall result was obtained on Image 13 (Table I), with a residual percentage of 2.6%, while Image 7 presented the worst result of the experiment, with a residual percentage of 53.3%.

## V. Conclusion

Based on the obtained results of the proposed methodology, it is possible to affirm that the proposed methodology is shown as a feasible alternative for corn plant counting. However, some challenges, already demonstrated in this letter, were found and should be treated so that the accuracy improves and the results are as close as possible to ground truth.

Considering the experiments and the proposed methodology, the best results achieved for corn plant counting were obtained from the images of fields with plant densities of around 70 000 plants/hectare, at the earliest growth stage (V4) and a flight height of 20 m.

As future work, tests will be performed to identify other objects present in the images that can influence the automatic counting, such as weeds and straw, and improve the algorithm so that the segmentation is more precise, especially in the cases of more advanced growth stages and higher plant densities. In addition, the next steps consider the possibility of applying the same approach for other crops, such as soybean and cotton.

## Acknowledgment

The authors would like to thank NVIDIA Corporation for donating Titan Xp used for this research.

## References

[1] FAO, IFAD, UNICEF, WFP, and WHO. (2018). *The State of Food Security and Nutrition in the World 2018. Building Climate Resilience for Food Security and Nutrition*. Accessed: Jan. 29, 2018. [Online]. Available: http://www.fao.org/3/I9553EN/i9553en.pdf

[2] USDA. (2018). *World Agricultural Supply and Demand Estimates*. Accessed: Jan. 29, 2018. [Online]. Available: https://www.usda.gov/oce/commodity/wasde/

[3] MarkEsalq. (2018). *O Aumento da Populacão Mundial e a Demanda por Alimentos [The Increase in World Population and Food Demand]*. Accessed: Dec. 27, 2018. [Online]. Available: http://markesalq.com.br/o-aumento-da-populacao-mundial-edemanda-por-alimentos

[4] R. L. Nielson, "Advanced farming systems and new technologies for the maize industry," in *Proc. FAR Maize Conf.*, 2012, pp. 1–15.

[5] J. Ribera, Y. Chen, C. Boomsma, and E. J. Delp, "Counting plants using deep learning," in *Proc. IEEE Global Conf. Signal Inf. Process. (GlobalSIP)*, Nov. 2017, pp. 1344–1348.

[6] A. Krizhevsky, "One weird trick for parallelizing convolutional neural networks," 2014, *arXiv:1404.5997*. [Online]. Available: https://arxiv.org/abs/1404.5997

[7] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. Int. Conf. Mach. Learn.*, 2015, pp. 448–456.

[8] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Dec. 2016, pp. 2818–2826.

[9] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, "Inception-v4, inception-ResNet and the impact of residual connections on learning," in *Proc. 31st AAAI Conf. Artif. Intell.*, 2017, pp. 4278–4284.

[10] F. Gnädinger and U. Schmidhalter, "Digital counts of maize plants by unmanned aerial vehicles (UAVs)," *Remote Sens.*, vol. 9, no. 6, p. 544, 2017.

[11] D. S. Shrestha and B. L. Steward, "Automatic corn plant population measurement using machine vision," in *Trans. ASAE*, vol. 46, no. 2, p. 559, 2003.

[12] Z. Fan, J. Lu, M. Gong, H. Xie, and E. Goodman, "Automatic tobacco plant detection in UAV images via deep neural networks," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 3, pp. 876–887, Mar. 2018.

[13] B. C. Russell, A. Torralba, K. P. Murphy, and W. T. Freeman, "Labelme: A database and Web-based tool for image annotation," *Int. J. Comput. Vis.*, vol. 77, nos. 1–3, pp. 157–173, 2008.

[14] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, 2015, pp. 234–241.

[15] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," Sep. 2014, *arxiv:1409.1556*. [Online]. Available: https://arxiv.org/abs/1409.1556

[16] J. Serra, *Image Analysis and Mathematical Morphology*. San Francisco, CA, USA: Academic, 1982.

[17] R. M. Haralick and L. G. Shapiro, *Computer and Robot Vision*. Boston, MA, USA: Addison-Wesley, 1992.

[18] R. Acevedo-Avila, M. Gonzalez-Mendoza, and A. Garcia-Garcia, "A linked list-based algorithm for blob detection on embedded vision-based sensors," *Sensors*, vol. 16, no. 6, p. 782, 2016.

[19] M. Paralic, "Fast connected component labeling in binary images," in *Proc. 35th Int. Conf. Telecommun. Signal Process.*, Jul. 2012, pp. 706–709.