Estimating Tomato Fruit Masses through Image Processing and Artificial Intelligence

Elognissè Erasme Guérin AGOSSADOU[1,†], Mahugnon Géraud AZEHOUN PAZOU[1,*,†], Régis Donald HONTINFINDE[1,†] and Ahmed Dooguy Kora[2,†]

[1]Université nationale des sciences, Technologie, Ingénierie et Mathématiques (UNSTIM), POBox 486, SOGBO ALIHO, Abomey, Benin
[2]EDMI, Cheikh Anta Diop University. Dakar, Senegal.

Abstract

The integration of intelligent and connected production systems has positioned artificial intelligence (AI) as a pivotal component in society's digital transformation, becoming indispensable. Leveraging the vast amounts of data generated, AI can now make critical decisions to mitigate potential disasters. This study focuses on developing a method that combines computer vision and machine learning algorithms to estimate tomato weights. A dataset of tomato images was compiled, and a modified Mask R-CNN algorithm was employed to detect, segment, and extract individual fruit masks. Various regression models were evaluated to predict tomato weight based on visual features. The results on the test dataset indicate that this approach can estimate the number and total weight of tomatoes with approximately 93% accuracy. This research highlights the potential for automated monitoring of market garden crop yields through AI.

Keywords

tomato fruit mass estimation, image processing, prediction models, Neural network, deep learning, pix2pix, rcnn

1. Introduction

Agriculture faces major challenges in sustainably feeding a growing global population, making accurate crop yield estimation essential for informed decision-making by farmers. While traditional methods such as field surveys can be helpful, they are often limited by issues of accuracy, cost, and time efficiency.

Tomato (Solanum lycopersicum) is a crucial vegetable crop globally, boasting 183 million tonnes in 2018 [1]. Native to Central and South America, the tomato was introduced to Europe in the 16th century, quickly gaining popularity for its delicious, nutrient-rich fruits loaded with vitamins, minerals, and antioxidants [2]. Major producers include China, India, the United States, and Turkey, with significant cultivation also occurring in African nations such as Nigeria, Egypt, Morocco, and Algeria, primarily for local consumption [3]. Tomatoes are generally classified into two main varieties: determinate, which have limited growth, and indeterminate, which continue growing throughout their lifecycle. Whether cultivated in open fields or under protective covers like greenhouses, tomato farming requires careful irrigation due to the plant's deep taproot system. Furthermore, challenges such as pest infestations—like downy mildew and Botrytis necessitate the use of appropriate cultivation practices and phytosanitary measures to ensure optimal yields.

Several approaches have been investigated in the literature to address the challenge of fruit weight estimation. For instance, Yamamoto et al. [4] developed a method to accurately count individual tomato fruits from images of plants grown in a laboratory setting. This method employed decision trees to analyze pixel color characteristics, achieving precise pixel-level segmentation. Post-processing was then applied to group pixels corresponding to fruits, enabling the extraction and counting of fruit centroids. The study reported a detection precision of 0.88 and recall of 0.80, demonstrating the method's efficacy in controlled environments for tomato detection and counting.

In Indonesia, the increasing demand for tomatoes necessitates efficient post-harvest handling. A study by Sari et al. [5] proposed a sorting system that categorizes tomatoes based on color, size, and weight using image processing with the OpenCV [6] library. The system sorts tomatoes into red, yellow, and green categories and measures dimensions by identifying the outermost points of the detected fruits. It utilizes a weight sensor for mass measurement. The prototype, which incorporates a webcam, Arduino, and conveyor system, achieved 100% accuracy in color detection and 95% in weight measurement, although dimensional measurement accuracy was only 5%.

Van Daalen et al. [7] examined the application of augmented reality (AR) in agriculture, focusing on detecting tomato ripeness using the 3D scanning capabilities of the HoloLens [8]. Their experimental setup, which included various tomato varieties, highlighted both the opportunities and challenges of using AR for hands-free tasks like training and harvesting in greenhouse environments.

Similarly, Lee et al. [9] proposed an artificial intelligence-based system for tomato detection and mass estimation, utilizing multi-class detection and instance-wise segmentation. By analyzing a tomato image dataset with a calibrated vision system, the study demonstrated a high correlation between fruit dimensions and mass. Their method achieved a mean absolute percentage error of 7.09%, showcasing the effectiveness of computer vision and machine learning for automating tasks such as yield monitoring and fruit sizing.

In another study, Nyalala et al. [10] developed seven regression models, including Support Vector Regression (SVR) [11] and artificial neural networks (ANNs) [12] with different training algorithms. These models effectively estimated fruit weight and volume, offering significant potential for improvements in fruit sorting and grading processes.

Basak et al. [13] introduced a non-destructive method for estimating strawberry fruit weight using machine learning models. By analyzing 900 samples from three different strawberry cultivars, they used image processing to calcu-

Cotonou'24: Conférence Internationale des Technologies de l'Information et de la Communication de l'ANSALB, June 27–28, 2024, Cotonou, BENIN

*Corresponding author.
†These authors contributed equally.
agossadourin@gmail.com (E. E. G. AGOSSADOU); geraud.pazou@gmail.com (M. G. A. PAZOU); hontinfinde7@gmail.com (R. D. HONTINFINDE); ahmed.kora@esmt.sn (A. D. Kora)

late pixel numbers. Linear regression (LR) and non-linear SVR models were applied, resulting in training and testing accuracies of 96.3% and 89.6%, respectively.

This study focuses on applying recent advancements in computer vision, particularly object detection, and machine learning algorithms to estimate tomato weight from real-world images. The subsequent sections describe the equipment used, the structure and composition of the dataset, and the methodology employed to generate accurate quantitative measures such as projected surface area and total weight for detected fruits. Our findings demonstrate the effectiveness of this approach. Additionally, we discuss the challenges faced and propose recommendations for future research.

2. Material and Methods

2.1. Dataset

The data used in this study consists of tomato fruit images collected both online and in the field under real-world conditions. The dataset includes a total of 180 images obtained online and 100 images taken in the field, containing a total of 1143 tomato fruit instances. Table 1 illustrates the composition of our dataset.

Images captured in the field helped to collect additional information such as actual fruit area and actual fruit weight, which enriches the dataset by providing accurate and relevant measurements for tomato fruit weight estimation. Table 2 presents additional insights concerning field-captured images. Upon analysis of the table, the average fruit weight is 35.30 g , with a standard deviation of 14.56 g . The average true area is 2673.48 mm^2, with a standard deviation of 873.68 mm^2. Quartile values provide insights into the distribution of the data. Thus, 25% of the fruits have a weight of less than 25.21 g, 50% have a weight of less than 37.00 g , and 75% have a weight of less than 43.49 g. For the actual surface area, 25% of fruits have an area less than $2,024.93$ mm^2, 50% have an area less than $2,779.53$ mm^2, and 75% have an area less than $3,219.12$ mm^2.

2.2. Methods

To estimate tomato fruit weights, we developed a four steps approach (see figure 1)

2.2.1. Detection, segmentation and extraction of tomato fruit masks

To train our segmentation model, we prepared a dataset of tomato images, labeled in the COCO format. The dataset consisted of 180 images containing 1043 instances of tomatoes, sourced from both the internet and field photography, and annotated using the ROboflow platform. We employed the Mask R-CNN instance segmentation model through the Detectron2 framework, selecting the mask_rcnn_R_50_FPN_3x configuration developed by Facebook AI Research. This model, pre-trained on the COCO dataset, combines the Mask R-CNN architecture with a ResNet-50 backbone and Feature Pyramid Network (FPN) for high-performance, multi-scale object detection.

Table 1
Dataset Overview

Source	Number of images	Number of fruit instances
Online	180	1043
Field-collected	100	100
Total	**280**	1143

Table 2
Additional information on images taken in the field

	weight	real_surface (mm^2)
count	100.000000	100.000000
mean	33.341900	2565.479377
std	13.884898	912.439551
min	9.930000	856.037079
25%	19.932500	1723.114236
50%	35.955000	2609.542487
75%	42.877500	3186.808853
max	63.760000	4931.281258

2.2.2. Projected Surface Area Estimation of Each Tomato

To evaluate the projected area of each tomato from images, a dataset was constructed, including individual images of tomatoes, their actual weight in grams, the total number of pixels in the image, the number of pixels corresponding tomato (obtained by semantic segmentation), and the total area of the image in square meters, obtained by camera calibration.

The estimation of the projected area took place in two steps: first, the segmentation mask allows us to calculate the area in pixels occupied by the tomato in the image. Then, a camera calibration converted this pixel area into an actual metric area, using a coin as a reference object. By photographing the tomatoes under the same conditions as the reference piece, the resulting conversion factor was used to convert the pixel area of each fruit into a measure of its actual projected area in metric units. This method uses a rule of three, where the actual surface area of the tomato (A_{tomato}) is estimated based on the number of pixels corresponding to the tomato in the image (P_{tomato}), using the conversion factor established during calibration: $\frac{A_{ref}}{P_{ref}}$.

$$A_{tomate} = P_{tomate} \times \frac{A_{ref}}{P_{ref}} \tag{1}$$

With this method, we were able to estimate the real surface area of each tomato in physical space from segmentation in image space, thanks to precise calibration using a reference object.

2.2.3. Tomato Mass Estimation

To estimate the weight of the tomatoes based on their projected surface area, we tested several regression models, including Simple Linear Regression (SLR), Multiple Linear Regression (MLR), and Partial Least Squares Regression (PLSR). These models aimed to establish a mathematical relationship between the surface area (independent variable) and the weight (dependent variable) of the tomatoes.

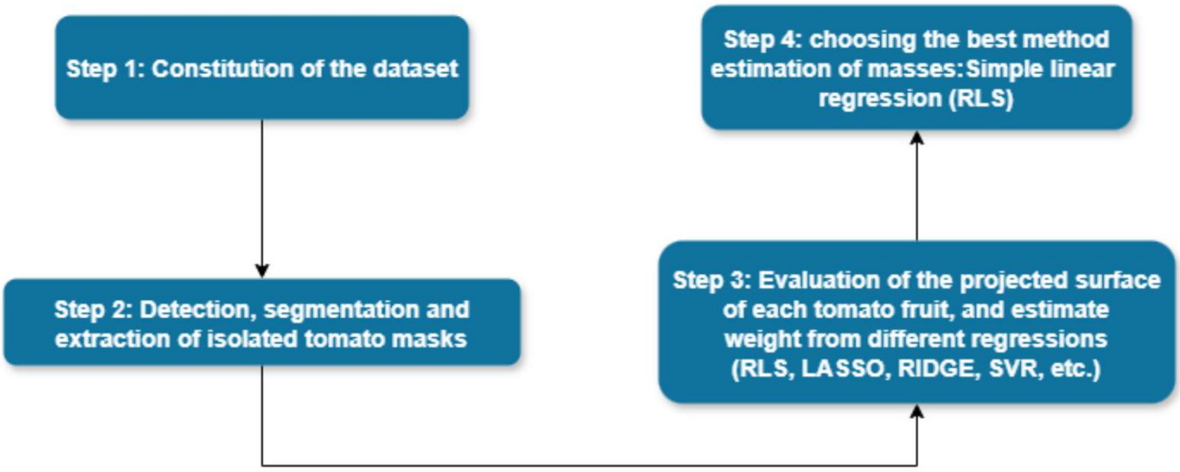

Figure 1: Summary illustration of the methodology

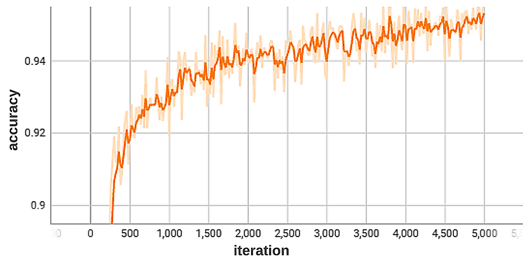

Figure 2: Model accuracy

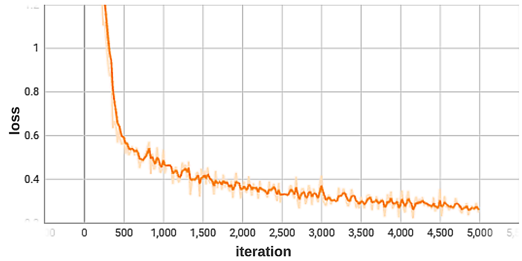

Figure 3: Evolution of the cost function

The performance of each model was evaluated on a validation set consisting of 20% of the total dataset, collected under real-world conditions. Standard metrics, such as Root Mean Square Error (RMSE) and the Coefficient of Determination (R^2), were employed to assess model accuracy. We also applied 10 -fold cross-validation to each model to reduce the likelihood of overfitting.

Figure 1 depicts the summary of the methodology adopted in this study.

3. Results and Discussion

3.1. Results

Figure 2 illustrates the model's accuracy, while Figure 3 depicts the evolution of the cost function

The performance of the model was evaluated on the test set consisting of 19 images containing a total of 149 tomato annotations. The Average Precision (AP) metric was used to quantify the model's ability to correctly detect and segment tomatoes under various conditions.

Table 3 presents the results obtained for the detection and semantic segmentation tasks. We observe an average AP of 55.9% for detection and 54.6% for segmentation on different IoU thresholds between 0.5 and 0.95. The model achieves better performance on large fruits (AP of 66.1% in detection) than on small tomatoes (AP of 30.3%).

These results confirm the model's effectiveness in detecting and segmenting tomatoes in real-world conditions. Further data annotation and model optimization are expected

Table 3
Model results in terms of Average Precision

Metric	**AP**	**AP50**	**AP75**	**APm**	**APl**
Detection	55.901	74.083	62.361	30.294	66.144
Segmentation	54.591	73.763	61.112	24.978	64.943

to enhance performance.

The projected surface area of each fruit was derived from the segmented mask by calculating the pixel area, then converting it to real-world units using camera calibration information as defined in Equation 1. This method achieved a precision of approximately 95.

For tomato weight estimation, a subset of the dataset containing real-world images was used, which included precise data on both the actual weight of each tomato and their projected surface area. A mathematical relationship between the weight and projected area was established through the evaluation of several regression methods. The algorithms tested included Least Squares Regression (LSR), Multiple Linear Regression (MLR), and Support Vector Machines (SVM), and their performance was compared using cross-validation and Mean Square Error (MSE) as the evaluation metric.

Table 4 highlights the performance metrics of the tested models.

Among the evaluated models, Lasso Regression achieved the best performance, with a MAE of 5,776 and an MSE of 62.99.

The corresponding model equation is:

Table 4
Performance metrics of different models

	MSE	MAE	RSE	$\mathbf{R^2}$
Linear Regression	67.465310	5.959565	8.110772	0.614756
Lasso Regression	**62.990660**	**5.775707**	**7.900871**	**0.659433**
Ridge Regression	64.222324	5.820851	7.789839	0.662985
ElasticNet Regression	65.214001	5.919661	8.063410	0.534604
SVR	81.623252	6.884133	8.980888	0.564414
Random Forest	67.078331	6.002012	8.102465	0.622985
AdaBoost Regression	76.441269	6.757964	8.621712	0.578526
KNeighbors Regression	68.750815	6.179380	8.225651	0.634068
Decision Tree	126.243306	8.132200	11.068062	0.322372

Table 5
Prediction results on the test set

Projected area	actual weight	Estimated weight	Absolute error	Relative error (%)
3219.122984	48.370	42.042340	6.327660	13.081785
2566.503710	30.760	33.377463	2.617463	8.509306
3279.246427	38.690	42.840604	4.150604	10.727847
2635.552676	30.600	34.294231	3.694231	12.072651
1273.816970	105.590	16.214358	89.375642	84.644040
2733.490428	30.360	35.594558	5.234558	17.241629
2521.044293	28.530	32.773894	4.243894	14.875199
3122.755376	37.570	40.762860	3.192860	8.498430
3501.459234	50.850	45.790941	5.059059	9.948985
2535.848511	35.070	32.970451	2.099549	5.986738
3098.277947	41.740	40.437871	1.302129	3.119618
...	...	...	...	...
2782.959320	26.520	36.251361	9.731361	36.694423
2436.892034	33.990	31.656598	2.333402	6.864966
2810.053656	37.080	36.611095	0.468905	1.264578
3040.780757	44.260	39.674477	4.585523	10.360424
3229.282192	46.620	42.177225	4.442775	9.529762
Total	1361.29	1257.726200	96.563799	7.09

$$M = 0.01327708 \times PA - 0.69821033 \quad (2)$$

Table 3.1 presents the prediction results on the test dataset, where our model achieved a relative error of 7.09% in estimating the total weight. When applied in an autonomous field system, this method shows great potential to enhance yield estimation efficiency, helping farmers save time and reduce labor costs.

3.2. Discussion

The study employed a multi-step methodology to estimate tomato fruit weights from images. First, a Mask R-CNN model, using the mask_rcnn_R_50_FPN_3x configuration, was trained on a dataset of 180 images containing 1043 tomato instances. After detection and segmentation, the projected surface area of each tomato was estimated using a calibrated conversion from pixel area to metric units, achieving approximately 95% accuracy. For weight estimation, several regression models were evaluated on a subset of real-world images with known weights and projected areas. Among the regression models evaluated, the Lasso Regression algorithm demonstrated superior performance in estimating tomato weights. This model achieved a Mean Absolute Error (MAE) of 5.776 grams and a Mean Squared Error (MSE) of 62.99 grams2. Our model outperformed the approach described by Lee et al. [9], which reported an MAE of 7.09 grams for a similar tomato weight estimation task.

When applied to the test dataset, this model achieved a relative error of 7.09% in estimating the total weight of tomatoes. These results demonstrate the potential of this combined approach for automated tomato yield estimation, although the ideal conditions of the study (fully visible fruits) suggest that further research is needed to address real-world challenges such as occlusion.

While this study yielded promising results, it's important to acknowledge its primary limitation: the experiments were conducted under idealized conditions that do not fully represent real-world agricultural environments. All tomatoes in the study were fully visible and unobstructed, which rarely occurs in actual fields where fruits are often partially hidden by leaves, branches, or other fruits. This idealization may lead to overly optimistic performance estimates.

To bridge this gap and enhance the model's practical applicability, future research will focus on developing robust occlusion handling techniques, such as implementing advanced image processing algorithms for reconstructing partially obscured fruits or using ellipse fitting methods to estimate the full shape of partially visible tomatoes.

Additionally, creating more representative datasets that reflect the challenging conditions found in real agricultural settings, including various levels of occlusion and diverse growth stages, will be crucial. By addressing these limitations and training on more diverse and challenging datasets, future iterations of this system could significantly improve in accuracy and robustness, making it a more reliable tool for automated agricultural yield estimation in real-world scenarios.

4. Conclusion

This study successfully introduced an innovative approach for accurately assessing tomato crop yields through the use of advanced image processing, computer vision, and artificial intelligence techniques. The results align closely with the objectives of estimating both the quantity and total weight of fruits, highlighting the practical benefits of this methodology for farmers.

Looking ahead, future enhancements will focus on refining the approach by integrating multispectral imaging to improve data acquisition. Additionally, algorithmic advancements, including image generation and ellipse fitting techniques, will be employed to tackle challenges related to occlusion. These developments will enhance the model's scalability and robustness, facilitating large-scale deployment in real-world agricultural settings. The anticipated implementation of this approach in automated systems that utilize drones and ground-based robots presents exciting opportunities for digital agriculture, paving the way for precise, efficient, and automated yield estimation.

References

[1] Food and Agriculture Organization, Agricultural production statistics, n.d. Retrieved from https://www.fao.org/3/cc3751en/cc3751en.pdf.

[2] M. Dorais, D. Ehret, A. Papadopoulos, Tomato (solanum lycopersicum) health components: From the seed to the consumer, Phytochemistry Reviews 7 (2008) 231–250. doi:10.1007/s11101-007-9085-x.

[3] WordAtlas, The world's leading tomato producing countries, n.d. Retrieved from https://www.worldatlas.com/articles/which-are-the-world-s-leading-tomato-producing-countries.html.

[4] K. Yamamoto, W. Guo, Y. Yoshioka, S. Ninomiya, On plant detection of intact tomato fruits using image analysis and machine learning methods, Sensors 14 (2014) 12191–12206. URL: https://www.mdpi.com/1424-8220/14/7/12191. doi:10.3390/s140712191.

[5] M. I. Sari, R. Fajar, T. Gunawan, R. Handayani, The use of image processing and sensor in tomato sorting machine by color, size, and weight, JOIV : International Journal on Informatics Visualization (2022). URL: https://api.semanticscholar.org/CorpusID:250542375.

[6] Opencv: Open source computer vision library, https://opencv.org/, n.d. Accessed: 2024-10-02.

[7] T. van Daalen, J. Peller, J. Balendonck, Determining fresh tomato weight using depth images from an ar headset, IFAC-PapersOnLine 55 (2022) 119–123. URL: https://www.sciencedirect.com/science/article/pii/S2405896322027586. doi:10.1016/j.ifacol.2022.11.125.

[8] Microsoft hololens, https://www.microsoft.com/fr-fr/hololens?msockid=1255574f41cb6082275f4248408c611d, n.d. Accessed: 2024-10-02.

[9] J.-S. Lee, H. Nazki, J. Baek, Y. Hong, M. hun Lee, Artificial intelligence approach for tomato detection and mass estimation in precision agriculture, Sustainability (2020). URL: https://api.semanticscholar.org/CorpusID:228852288.

[10] I. Nyalala, C. Okinda, Q. Chao, P. Mecha, T. Korohou, Z. Yi, S. Nyalala, Z. Jiayu, L. Chao, C. Kunjie, Weight and volume estimation of single and occluded tomatoes using machine vision, International Journal of Food Properties 24 (2021) 818–832. URL: https://doi.org/10.1080/10942912.2021.1933024. doi:10.1080/10942912.2021.1933024. arXiv:https://doi.org/10.1080/10942912.2021.1933024.

[11] F. Zhang, L. J. O'Donnell, Chapter 7 - support vector regression, in: A. Mechelli, S. Vieira (Eds.), Machine Learning, Academic Press, 2020, pp. 123–140. URL: https://www.sciencedirect.com/science/article/pii/B9780128157398000079. doi:10.1016/B978-0-12-815739-8.00007-9.

[12] K. O'Shea, R. Nash, An introduction to convolutional neural networks, 2015. URL: https://arxiv.org/abs/1511.08458. arXiv:1511.08458.

[13] J. K. Basak, B. Paudel, N. E. Kim, N. C. Deb, B. G. Kaushalya Madhavi, H. T. Kim, Non-destructive estimation of fruit weight of strawberry using machine learning models, Agronomy 12 (2022). URL: https://www.mdpi.com/2073-4395/12/10/2487. doi:10.3390/agronomy12102487.