



WildFruiP: Estimating Fruit Physicochemical Parameters from Images Captured in the Wild

Diogo J. Paulo^{1,2}(✉) , Cláudia M. B. Neves^{3,4} ,
Dulcineia Ferreira Wessel^{3,4} , and João C. Neves^{1,2} 

¹ University of Beira Interior, Covilhã, Portugal

² NOVA LINCS - NOVA Laboratory for Computer Science and Informatics,
Lisbon, Portugal

diogo.paulo@ubi.pt

³ Polytechnic Institute of Viseu, Viseu, Portugal

⁴ LAQV-REQUIMTE, Department of Chemistry,
University of Aveiro, Aveiro, Portugal

Abstract. The progress in computer vision has allowed the development of a diversity of precision agriculture systems, improving the efficiency and yield of several processes of farming. Among the different processes, crop monitoring has been extensively studied to decrease the resources consumed and increase the yield, where a myriad of computer vision strategies has been proposed for fruit analysis (e.g., fruit counting) or plant health estimation. Nevertheless, the problem of fruit ripeness estimation has received little attention, particularly when the fruits are still on the tree. As such, this paper introduces a strategy to estimate the maturation stage of fruits based on images acquired from handheld devices while the fruit is still on the tree. Our approach relies on an image segmentation strategy to crop and align fruit images, which a CNN subsequently processes to extract a compact visual descriptor of the fruit. A non-linear regression model is then used for learning a mapping between descriptors to a set of physicochemical parameters, acting as a proxy of the fruit maturation stage. The proposed method is robust to the variations in position, lighting, and complex backgrounds, being ideal for working in the wild with minimal image acquisition constraints. Source code is available at <https://github.com/Diogo365/WildFruiP>.

Keywords: Fruit Physicochemical Parameters · Computer Vision · Maturation Stage

1 Introduction

The agricultural industry is a vital sector of the global economy and plays a crucial role in the human food supply. Food production is constantly evolving, and technology has been a significant ally in this process. Vision-based systems

and artificial intelligence have enabled significant improvements in quality and productivity. In this context, this work aims to develop a computer vision strategy for automatically estimating fruit maturation stages from a single photo acquired by handheld devices. This will allow for a more accurate and efficient evaluation of the fruit production process, reducing resource waste and increasing agricultural sector productivity.

While computer vision systems have provided valuable information for crop monitoring, the assessment of fruit ripeness has received limited attention. Existing systems primarily rely on drone-based monitoring methods, which do not allow for the determination of the specific maturation stage of each fruit. Alternatively, some researchers have developed systems to classify the maturation stage of fruit after harvesting, which may not be particularly useful for farmers.

Considering the importance of determining fruit ripeness for optimal harvesting decisions, we introduce a method for determining the maturation stage of fruits using a single photo acquired from handheld devices while the fruit is still on the tree. The proposed method relies on an image segmentation model to crop and align the fruit, which a CNN subsequently analyses for extracting a visual descriptor of the fruit (illustrated in Fig. 1). The maturation stage is defined by a set of physicochemical parameters that are inferred from the visual descriptor using a regression model. To allow the learning of the image segmentation and regression model, we collected a dataset of 400 images of figs and prickly pears and their corresponding physicochemical parameters. To the best of our knowledge, this is the first dataset comprising both visual and physicochemical data, and we expect it to be of particular interest to the research community for carrying out studies of the relationship between the chemical properties of fruits and their visual appearance. The dataset used in this work is publicly available on <https://github.com/Diogo365/WildFruiP>.

Our main contributions in this work are as follows:

- We introduce a strategy for fruit ripeness estimation capable of operating in images acquired in the wild while the fruit is still in the tree.
- We assessed the performance of the proposed method in determining a set of physicochemical parameters of a fruit using a single image obtained in the visible light spectrum.
- To foster the research on the problem of fruit ripeness estimation from visual data, we introduce a dataset comprising 400 images from two fruit species and their respective physicochemical parameters, which serve as a proxy to the fruit maturation stage.

2 Related Work

2.1 Detection Methods

Object detection in images is a crucial task in computer vision, which had a tremendous progress in the last years due to the emergence of deep learning.

Several works have taken advantage of this progress for fruit detection. In [12], Parvathi *et al.* proposed an enhanced model of Faster R-CNN [10] for detecting coconuts in images with complex backgrounds to determine their ripeness. The performance of the model was evaluated on a dataset containing real-time images and images from the Google search engine. The results showed that the improved Faster R-CNN model achieved better detection performance compared to other object detectors such as SSD [7], YOLO [9], and R-FCN [2].

2.2 Segmentation Methods

Image segmentation is crucial for fruit image analysis, as it allows separating fruits from other parts of the image, such as leaves or background. Mask R-CNN [4] is an instance segmentation method that has proven effective in object segmentation tasks and has been extensively used for fruit analysis applications.

Siricharoen *et al.* [13] proposed a three-phase deep learning approach [13] to classify pineapple flavor based on visual appearance. First, a Mask R-CNN segmentation model was used for extracting pineapple features from the YCbCr color space. Then, a residual neural network pre-trained on COCO and ImageNet datasets was utilized for flavor classification. The authors concluded that their model successfully captured the correlation between pineapple visual appearance and flavor.

Ni *et al.* [8] developed an automated strategy for blueberry analysis. They employed a deep learning-based image segmentation method using the Mask R-CNN model to count blueberries and determine their ripeness. The results indicated variations among the cultivars, with ‘Star’ having the lowest blueberry count per cluster, ‘Farthing’ exhibiting less ripe fruits but compact clusters, and ‘Meadowlark’ showing looser clusters. The authors highlighted the need for objective methods to address fruit ripeness inconsistency caused by annotation inconsistencies in the trained model.

2.3 Methods for Estimating Fruit Ripeness in Images

Several strategies have been introduced to enable pre-harvest in-field assessment of fruit ripeness using handheld devices [6]. However, most approaches rely on the non-visible light spectrum, requiring thus dedicated hardware [11].

Regarding the approaches devised for visible light spectrum, most of them use CNNs for the estimation of fruit ripeness. Appe *et al.* [1] proposed a model for tomato ripeness estimation using transfer learning. They relied on the VGG16 architecture, where the top layer was replaced with a multilayer perceptron (MLP). The proposed model with fine-tuning exhibited improved effectiveness in tomato ripeness detection and classification. In another work, Sabzi *et al.* [12] developed an innovative strategy for estimating the pH value of oranges from three different varieties. A neural network was combined with the particle swarm optimization [5] to select the most discriminative features from a total of 452 features obtained directly from segmented orange images. This approach was able to rely on a subset of six features to obtain an accurate estimation of the pH values across different orange varieties.

In short, few approaches were devised for addressing the problem of fruit ripeness estimation from visual data, ranging from traditional extraction of hand-crafted features to deep-learning-based methods.

3 Proposed Method

The proposed approach can be broadly divided into three principal phases: the detection and segmentation of fruits in an image, the alignment and cropping of the fruit, and the determination of the physicochemical parameters of the fruit. The pipeline of this method is presented in Fig. 1.

3.1 Fruit Detection and Segmentation

This phase aims at removing the spurious information from the image keeping only the fruit region. Accordingly, the fruit is segmented automatically using the Mask R-CNN [4] allowing the prediction of a binary mask containing the pixels where exists a specific type of fruit. Considering the specificity of this task, the Mask R-CNN was fine-tuned on the proposed dataset, allowing thus it to generalize to the fruits targeted in this problem. To address the problem of multiple fruits in the image, we establish that the fruit to be analysed should be in the center of the image, and thus the remaining masks are discarded.

Figure 2 depicts the results obtained by applying Mask R-CNN to both figs and prickly pears.

3.2 Image Alignment

Considering that fruit orientation varies significantly in the images, it is particularly important to enforce a standard alignment to ease the learning of the fruit analysis model.

Considering the general shape of fruits, we propose to approximate their silhouette using an ellipse. Also, we concluded that the silhouette of the fruit can be modeled using the segmentation mask obtained from the previous phase.

Let M be the segmentation mask, and consider the general equation of the ellipse:

$$\frac{((x - x_0) \cos \theta + (y - y_0) \sin \theta)^2}{a^2} + \frac{(-(x - x_0) \sin \theta + (y - y_0) \cos \theta)^2}{b^2} = 1, \quad (1)$$

where (x_0, y_0) are the coordinates of the ellipse's center, a and b are the horizontal and vertical semi-axes, respectively and θ is the ellipse orientation, with $\theta \in [-\frac{\pi}{2}, \frac{\pi}{2}]$. The boundary of M is determined using the convex hull of the (x, y) points of M , and least square fitting [3] is used to determine x_0 , y_0 , a , b , and θ . The rotation angle θ is then used to rotate the original image and crop the fruit region based on the minimum bounding box containing the ellipse obtained. The results of the fruit alignment can be observed in Fig. 3.

3.3 Determination of Physicochemical Parameters

In the third phase, a CNN model is used to learn a visual descriptor which can encode the discriminative information regarding the physicochemical parameters of the fruit. A multi-layer perceptron is used as a regression model to infer the nine physicochemical parameters from the visual descriptors. The CNN and the regression model were trained in an end-to-end manner using the mean-squared error loss, and a k-fold cross-validation technique was adopted due to the reduced amount of training data.

4 Dataset

Considering the unavailability of public datasets comprising fruit images and their corresponding maturation stage or physicochemical parameters, we acquired 4 photos each from 60 figs and 40 prickly pears from local farmers. The fruits were subsequently harvested and analysed in the lab to extract 9 characteristics that are typically correlated with the maturation state of the fruit. The physical and chemical parameters obtained are listed in Table 1.

Table 1. Range of values for the physicochemical parameters used in this project.

Attribute	Range of Values
<i>TSS ($^{\circ}$ Brix)</i>	[13.5;19.0]
<i>Hardness (N)</i>	[1.1;61.4]
<i>pH</i>	[4.2;6.2]
<i>mass (g)</i>	[24.3;209.6]
<i>L</i>	[21.1;56.7]
<i>a</i>	[-16.4;23.2]
<i>b</i>	[2.6;26.8]
<i>length (cm)</i>	[59.6;110.0]
<i>diameter (cm)</i>	[39.0;62.9]

To allow the development of a custom image segmentation model, we annotated the complete set of 400 images using the CVAT tool. An exemplar from each of the fruit species and its corresponding annotations can be observed in Fig. 4. To foster the research on the problem of estimating fruit ripeness from visual data, we make our dataset publicly available¹.

¹ <https://github.com/Diogo365/WildFruiP>.

5 Experiments

This section reports the performance of the proposed method for the problem of physicochemical parameter estimation from images of figs and prickly pears acquired using handheld devices. Tests are conducted using the aligned, and misaligned/cropped dataset using different neural networks. Also, we compare the proposed approach with a state-of-the-art method devised for inferring fruit physicochemical parameters from visual data.

5.1 Implementation Details

Detection and Segmentation. The backbone of the Mask R-CNN was a Residual Neural Network (ResNet), specifically the ResNet50 variant integrated into the PyTorch framework. Prior annotations were necessary for each fruit, including bounding boxes, labels, and masks to train the model. The data augmentation transformations were resizing, horizontal flipping, brightness and contrast adjustment. After defining the necessary transformations for data processing, the annotated initial dataset was split into 80% for training and 20% for testing. Finally, with the separated datasets and processed data, the model was trained for 50 epochs using the stochastic gradient descent optimizer with a learning rate of 0.001.

Determination of Physicochemical Parameters. The training data consisted of a set of images and their corresponding physicochemical parameters. Each parameter was normalized using a linear transformation estimated from the training data. A lightweight CNN architecture (ResNet18) was used for extracting 2048 dimensional visual descriptors from the aligned fruit images and a multi-layer perceptron was exploited for the estimation of nine parameters from the visual descriptors. The configurations used are presented in Table 2. All models were trained for a maximum of 100 epochs using the Early Stopping regularization technique and all of our experiments were conducted on PyTorch with NVIDIA GeForce RTX 3060 GPU and with Intel(R) Core(TM) i7-10700 CPU @ 2.90GHz. The inference times reported in Table 3 were obtained by executing the model on this hardware configuration.

Table 2. Configuration used for training the CNN.

Batch Size	16
Epochs	100
Learning Rate	0.001
Optimizer	Adam
Image Size	320 × 320

Table 3. Inference time and total size of the different models.

Models	Inference Time	Storage Size
MobileNetV2	10.1 ms	8.7 MB
ResNet-18	10.2 ms	89.9 MB

5.2 Metrics

To assess the performance of the proposed model, four metrics were employed, mean squared error (MSE), mean absolute error (MAE), mean absolute percentage error (MAPE), and the coefficient of determination (R^2). They are defined as follows:

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2, \tag{2}$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|, \tag{3}$$

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100, \tag{4}$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \tag{5}$$

In these equations, n represents the total number of data points or observations in the evaluation set, y_i denotes the true value of the dependent variable for the i^{th} observation, \hat{y}_i represents the predicted value of the dependent variable for the i^{th} observation, \bar{y} denotes the mean value of the dependent variable across all observations.

Table 4. Performance of the proposed approach. The R^2 value (mean \pm std) determined for both species denotes a strong predictive power for some physicochemical parameters. Also, the comparison with the approach of Sabzi *et al.* [12], evidences a clear improvement in all parameters.

Attributes	Sabzi <i>et al.</i> [12]		Proposed Method		
	Prickly Pears	Figs	Prickly Pears	Figs	
TSS ($^{\circ}$ Brix)	-0.81 \pm 0.04	-	0.18 \pm 0.07	-	
Hardness (N)	-2.66 \pm 0.70	-1.13 \pm 0.66	0.51 \pm 0.08	0.68 \pm 0.03	
pH	-1.26 \pm 0.27	-	0.13 \pm 0.11	-	
mass (g)	-0.83 \pm 0.14	-	0.22 \pm 0.11	-	
Color	L	-3.39 \pm 0.84	-3.26 \pm 0.01	0.25 \pm 0.09	0.75 \pm 0.04
	a	-6.71 \pm 0.03	-5.96 \pm 1.73	0.83 \pm 0.03	0.87 \pm 0.01
	b	-4.63 \pm 0.71	-3.23 \pm 0.69	0.42 \pm 0.08	0.79 \pm 0.03
length	-0.17 \pm 0.14	-0.04 \pm 0.08	-0.03 \pm 0.08	0.40 \pm 0.03	
diameter	-0.65 \pm 0.28	-0.49 \pm 0.34	0.37 \pm 0.07	0.28 \pm 0.04	

5.3 Performance of the Proposed Approach

The proposed method was assessed in the evaluation split of both prickly pear and fig images using k-fold validation and repeating the training and evaluation process 10 times. The results are reported in Table 4.

The analysis of the results with respect to prickly pears shows a moderate correlation (refer to R^2) between some physicochemical parameters and the predictions of the network obtained from the fruit image. All parameters showed a positive correlation except for the length. Insufficient relevant information in the image might explain the lack of correlation with the prickly pear’s length. Visual features such as shape, color, or texture are not informative about the length of a fruit. A strong predictive power was obtained for the ‘a’ parameter, hardness, and ‘b’ parameters, with correlations of 0.83, 0.51, and 0.42, respectively. The ‘a’ parameter represented the fruit chromaticity from green to red, which is strongly correlated with the fruit ripeness. Hardness also corresponds to ripeness, as riper figs are typically less firm. However, the sugar content, measured by the TSS ($^{\circ}$ Brix) had a weak correlation possibly due to the dataset small size.

Table 5. Results obtained by the proposed model with the misaligned, and aligned datasets using prickly pears.

Attributes	R2		MSE		MAE		MAPE		
	Misaligned Image	Aligned Image	Misaligned Image	Aligned Image	Misaligned Image	Aligned Image	Misaligned Image	Aligned Image	
TSS ($^{\circ}$ Brix)	-0.29 ± 0.42	0.18 ± 0.07	0.94 ± 0.25	0.63 ± 0.04	0.78 ± 0.10	0.65 ± 0.03	5.13 ± 0.64	4.27 ± 0.16	
Hardness (N)	0.41 ± 0.13	0.51 ± 0.08	6.97 ± 1.46	5.89 ± 0.95	2.07 ± 0.21	1.92 ± 0.19	12.12 ± 1.15	11.25 ± 1.43	
pH	-0.18 ± 0.25	0.13 ± 0.11	0.01 ± 0.00	0.01 ± 0.00	0.09 ± 0.01	0.08 ± 0.00	1.50 ± 0.14	1.31 ± 0.06	
mass (g)	0.05 ± 0.15	0.22 ± 0.11	1022.83 ± 162.76	854.54 ± 131.64	25.10 ± 1.94	23.06 ± 1.56	21.95 ± 1.75	20.50 ± 1.33	
Color	L	-0.18 ± 0.31	0.25 ± 0.09	17.61 ± 5.07	10.55 ± 1.03	3.20 ± 0.40	2.45 ± 0.15	6.71 ± 0.82	5.14 ± 0.30
	a	0.65 ± 0.11	0.83 ± 0.03	27.50 ± 8.16	13.50 ± 2.39	4.03 ± 0.46	2.93 ± 0.24	193.38 ± 54.40	130.26 ± 21.16
	b	0.15 ± 0.15	0.42 ± 0.08	5.02 ± 0.86	3.43 ± 0.52	1.79 ± 0.17	1.44 ± 0.10	8.39 ± 0.85	6.80 ± 0.47
length	-0.37 ± 0.21	-0.03 ± 0.08	127.65 ± 21.27	97.24 ± 8.54	8.92 ± 0.83	7.80 ± 0.33	11.02 ± 1.17	9.72 ± 0.41	
diameter	0.08 ± 0.21	0.37 ± 0.07	28.70 ± 6.02	19.96 ± 2.41	4.22 ± 0.42	3.57 ± 0.23	8.34 ± 0.87	7.03 ± 0.51	

Regarding the performance attained on figs, only six out of nine parameters were evaluated due to insufficient data for TSS ($^{\circ}$ Brix), pH, and mass. Nevertheless, our approach demonstrated a better aptitude for estimating physicochemical parameters in this fruit species (figs), likely due to the disparity in dataset sizes.

Regarding the comparison with the state-of-the-art, the method of Sabzi *et al.* [12] significantly underperformed when compared with our approach. The main justification for this difference is the fact that the method of Sabzi *et al.* [12] was originally intended to analyse fruit images in controlled scenarios (the method was devised for pH estimation of oranges in a uniform background). However, the images obtained when the fruits are still on the tree are inherently more challenging due to the varying pose, lighting, and complexity of the background.

5.4 Impact of Alignment Phase

In this experiment, the model was trained using misaligned/cropped, and aligned images for the nine physicochemical parameters.

Prickly Pears. Upon analyzing Table 5, it was observed that ablating the alignment of the images, as expected, led to worst results, making the difference between negative and positive values of R^2 as is the case of the TSS ($^{\circ}$ Brix), pH and L parameters.

Table 6. Results obtained by the proposed model with the misaligned, and aligned datasets using figs.

Attributes	R2		MSE		MAE		MAPE		
	Misaligned Image	Aligned Image	Misaligned Image	Aligned Image	Misaligned Image	Aligned Image	Misaligned Image	Aligned Image	
TSS ($^{\circ}$ Brix)	-	-	-	-	-	-	-	-	
Hardness (N)	0.72 \pm 0.05	0.68 \pm 0.03	77.39 \pm 12.30	88.20 \pm 9.20	6.40 \pm 0.78	6.76 \pm 0.46	137.34 \pm 29.17	138.02 \pm 21.14	
pH	-	-	-	-	-	-	-	-	
mass (g)	-	-	-	-	-	-	-	-	
Color	L	0.73 \pm 0.06	0.75 \pm 0.04	17.95 \pm 4.04	16.55 \pm 2.25	3.31 \pm 0.37	3.17 \pm 0.23	9.61 \pm 0.99	9.18 \pm 0.66
	a	0.87 \pm 0.01	0.87 \pm 0.01	11.20 \pm 1.16	11.01 \pm 1.20	2.63 \pm 0.18	2.61 \pm 0.16	84.93 \pm 12.27	97.79 \pm 8.78
	b	0.77 \pm 0.03	0.79 \pm 0.03	10.30 \pm 1.12	9.39 \pm 1.24	2.50 \pm 0.15	2.40 \pm 0.18	25.21 \pm 1.34	23.89 \pm 3.05
length	0.47 \pm 0.04	0.40 \pm 0.03	46.36 \pm 3.48	53.68 \pm 2.58	5.34 \pm 0.19	5.87 \pm 0.15	7.03 \pm 0.32	7.75 \pm 0.24	
diameter	0.38 \pm 0.08	0.28 \pm 0.04	8.42 \pm 0.96	9.47 \pm 0.53	2.27 \pm 0.12	2.38 \pm 0.07	4.85 \pm 0.26	5.12 \pm 0.16	

Figs. Regarding figs, upon analyzing Table 6 it was observed that the aligned dataset yielded slightly better results compared to the misaligned dataset. The aligned dataset led to improved training results for the color parameters, while the misaligned dataset performed better for shape features (diameter and length).

The diameter parameter proved challenging but outperformed the worst parameter in the prickly pear experiment.

5.5 Impact of Model Architecture

Considering that the proposed approach is planned to work in handheld devices with low computational resources, the proposed method is based on a lightweight architecture. To determine the best architecture for the problem, we compared the impact of the architecture on the performance of the proposed approach, as well as, on the inference time.

Therefore, for this experiment, we assessed the performance of our approach using two lightweight architectures: MobileNetV2 and ResNet18. The comparison of the model size and inference time of the different architectures is provided in Table 3, while Table 7 reports the performance of our approach along the different architectures.

It is interesting to observe that ResNet18 was able to consistently attain the best results over all parameters and simultaneously for both fruit species.

Despite its larger size, we claim that the superior predictive power obtained justifies its use in this problem. Also, it is important to note that the inference time is equivalent for both models.

5.6 Hard Samples

To further explain the obtained results, an additional test was conducted to identify figs where the proposed method significantly deviated from the correct physicochemical parameters using the MAE metric (less sensitive to outliers than MSE). Figure 5 shows the images of the two fruit species where the proposed approach had the largest MAE.

Table 7. Results obtained for the R^2 metric on prickly pears and figs utilizing the ResNet18 and MobileNetv2.

Attributes	Prickly Pears		Figs	
	MobileNetV2	ResNet18	MobileNetV2	ResNet18
TSS ($^{\circ}$ Brix)	0.15 \pm 0.08	0.18 \pm 0.07	-	-
Hardness (N)	0.49 \pm 0.06	0.51 \pm 0.08	0.62 \pm 0.03	0.68 \pm 0.03
pH	0.06 \pm 0.10	0.13 \pm 0.11	-	-
mass (g)	0.07 \pm 0.07	0.22 \pm 0.11	-	-
Color	L	0.08 \pm 0.09	0.64 \pm 0.03	0.75 \pm 0.04
	a	0.76 \pm 0.04	0.80 \pm 0.02	0.87 \pm 0.01
	b	0.28 \pm 0.09	0.42 \pm 0.08	0.68 \pm 0.03
length	-0.10 \pm 0.10	-0.03 \pm 0.08	0.24 \pm 0.03	0.40 \pm 0.03
diameter	0.26 \pm 0.05	0.37 \pm 0.07	0.16 \pm 0.06	0.28 \pm 0.04

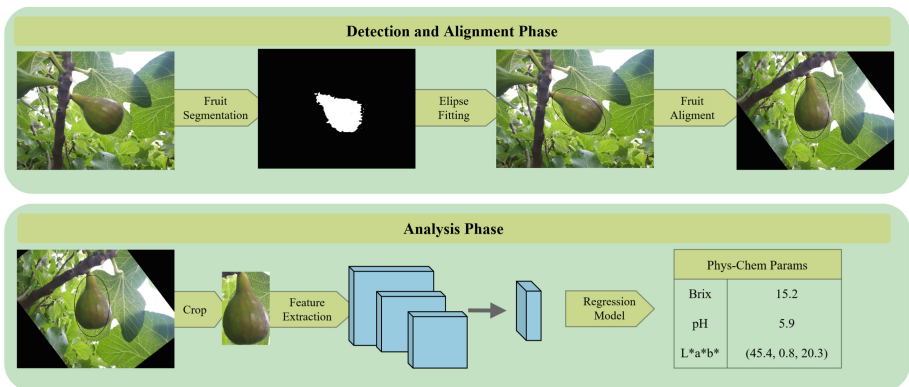


Fig. 1. Pipeline of the proposed approach. The fruit image is given to an image segmentation approach, which determines the fruit mask. Using the mask, a fitting process is performed to enclose the fruit in an ellipse, and the rotation angle of the ellipse is used to align the fruit in the image. Afterwards, the fruit is cropped using the bounding boxes also extracted from the segmentation mask. The cropped fruit is fed into a CNN for extracting a visual descriptor which is subsequently mapped to a set of physicochemical parameters through a regression model.

Several factors affect the performance of the model, including luminosity differences, blur, variations in fig shapes (length, diameter, and mass), and limitations in training due to a lack of examples of unripe figs.



Fig. 2. Fruit detection and segmentation. The Mask R-CNN was fine-tuned to provide a rough segmentation of the fruit allowing to discard irrelevant regions of the image in the analysis of the data. Even though the masks are not so accurate in the border, it is important to note that the accuracy of the segmentation mask is not crucial for the overall approach.

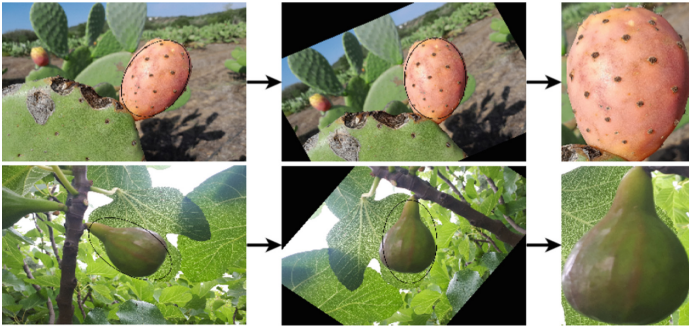


Fig. 3. Proposed alignment process. The fruit is approximated using an ellipse, which allows to obtain the rotation angle for image alignment and cropping the aligned fruit. The alignment process is depicted for the two fruit species considered in this study.



Fig. 4. Samples from the proposed dataset. Our dataset comprises 400 images of two fruit species and their corresponding physicochemical parameters. Also, we provide the location of fruit in the image using manually annotated bounding boxes and segmentation masks.



Fig. 5. Hard to predict samples. The five samples from the two fruit species that have the highest absolute error over the nine physicochemical parameters.

6 Conclusion and Future Work Prospects

In this work, we introduced an approach for estimating the maturation stage of fruit images acquired in the wild using handheld devices. The proposed approach relied on an innovative alignment strategy that increased the robustness to pose variations. Also, we introduced a novel dataset containing images with significant variations in lighting, and diversity of the background. The experimental validation of the proposed approach showed a strong correlation with some physicochemical parameters, which can serve as a proxy to determine the maturation stage of the fruits considered in this study. On the other hand, our approach was capable of remarkably surpassing a state-of-the-art approach specifically designed for fruit maturation estimation. To further validate the proposed method, we carried out several experiments, which showed that the alignment phase increased the performance of the method. Also, the analysis of the most challenging image samples evidenced that blur and brightness variation were the major causes of failure. In the future, we expect that our approach can be incorporated into a mobile application, providing farmers with an easy-to-use fruit ripeness estimation tool for efficient control and informed decision-making in agriculture.

Acknowledgements. This work was funded by the InovFarmer.MED project, which is part of the PRIMA Programme. Also, it was supported by NOVA LINC3 (UIDB/04516/2020) with the financial support of FCT/IP.

References

1. Appe, S.R.N., Arulselvi, G., Balaji, G.N.: Tomato ripeness detection and classification using VGG based CNN models. *Int. J. Intell. Syst. Appl. Eng.* **11**(1), 296–302 (2023)

2. Dai, J., Li, Y., He, K., Sun, J.: R-FCN: object detection via region-based fully convolutional networks. In: *Proceedings of the Advances in Neural Information Processing Systems (NIPS)*, vol. 29. Curran Associates, Inc. (2016)
3. Fitzgibbon, A., Pilu, M., Fisher, R.: Direct least square fitting of ellipses. *IEEE Trans. Pattern Anal. Mach. Intell. (TPAMI)* **21**(5), 476–480 (1999)
4. He, K., Gkioxari, G., Dollár, P., Girshick, R.: Mask R-CNN. In: *Proceedings of the International Conference on Computer Vision (ICCV)*, pp. 2980–2988 (2017)
5. Kennedy, J., Eberhart, R.: Particle swarm optimization. In: *Proceedings of the International Conference on Neural Networks (IJCNN)*, vol. 4, pp. 1942–1948 (1995)
6. Li, B., Lecourt, J., Bishop, G.: Advances in non-destructive early assessment of fruit ripeness towards defining optimal time of harvest and yield prediction - a review. *Plants* **7**(1), 3 (2018)
7. Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., Berg, A.C.: SSD: single shot multibox detector. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) *ECCV 2016. LNCS*, vol. 9905, pp. 21–37. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46448-0_2
8. Ni, X., Li, C., Jiang, H., Takeda, F.: Deep learning image segmentation and extraction of blueberry fruit traits associated with harvestability and yield. *Horticult. Res.* **7**, 110 (2020)
9. Redmon, J., Divvala, S., Girshick, R., Farhadi, A.: You only look once: unified, real-time object detection. In: *Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 779–788 (2016)
10. Ren, S., He, K., Girshick, R., Sun, J.: Faster R-CNN: towards real-time object detection with region proposal networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **39**(6), 1137–1149 (2017)
11. Rizzo, M., Marcuzzo, M., Zangari, A., Gasparetto, A., Albarelli, A.: Fruit ripeness classification: a survey. *Artif. Intell. Agric.* **7**, 44–57 (2023)
12. Sabzi, S., Javadikia, H., Arribas, J.I.: A three-variety automatic and non-intrusive computer vision system for the estimation of orange fruit pH value. *Measurement* **152**, 107–298 (2020)
13. Siricharoen, P., Yomsatieankul, W., Bunsri, T.: Recognizing the sweet and sour taste of pineapple fruits using residual networks and green-relative color transformation attached with mask R-CNN. *Postharvest Biol. Technol.* **196**, 112–174 (2023)