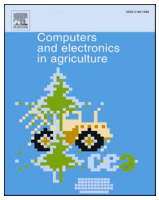
Computers and Electronics in Agriculture 183 (2021) 106066

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

journal homepage: www.elsevier.com/locate/compag



Original papers

Detection, classification, and mapping of coffee fruits during harvest with computer vision

Helizani Couto Bazame [a,*], José Paulo Molin [a], Daniel Althoff [b], Maurício Martello [a]

[a] Biosystems Engineering Department, Luiz de Queiroz College of Agriculture University of São Paulo (ESALQ/USP), Piracicaba, SP, Brazil
[b] Agricultural Engineering Department, Federal University of Viçosa (UFV), Viçosa, MG, Brazil

ARTICLE INFO

Keywords:
Precision agriculture
Convolutional neural networks
YOLO
Deep learning

ABSTRACT

In this study, an algorithm is implemented with a computer vision model to detect and classify coffee fruits and map the fruits maturation stage during harvest. The main contribution of this study is with respect to the assignment of geographic coordinates to each frame, which enables the mapping of detection summaries across coffee rows. The model used to detect and classify coffee fruits was implemented using the Darknet, an open source framework for neural networks written in C. The coffee fruits detection and classification were performed using the object detection system named YOLOv3-tiny. For this study, 90 videos were recorded at the end of the discharge conveyor of a coffee harvester during the 2020 harvest of arabica coffee (Catuaí 144) at a commercial area in the region of Patos de Minas, in the state of Minas Gerais, Brazil. The model performance peaked around the ~3300th iteration when considering an image input resolution of 800 × 800 pixels. The model presented an mAP of 84%, F1-Score of 82%, precision of 83%, and recall of 82% for the validation set. The average precision for the classes of unripe, ripe, and overripe coffee fruits was 86%, 85%, and 80%, respectively. As the algorithm enabled the detection and classification in videos collected during the harvest, it was possible to map the qualitative attributes regarding the coffee maturation stage along the crop lines. These attribute maps provide managers important spatial information for the application of precision agriculture techniques in crop management. Additionally, this study should incentive future research to customize the deep learning model for certain tasks in agriculture and precision agriculture.

1. Introduction

From an economic point of view, the mechanical harvesting of coffee is the most important agricultural operation in coffee farming. It has a large share in production costs and influences the quality of coffee produced (Matiello et al., 2015). Most of the farmers harvest the coffee fruits in a traditional method, attempting to harvest all of the coffee fruits at once. This method results in the harvesting of fruits at different stages of maturation (unripe, ripe, and overripe), which is an outcome of the coffee tree presenting uneven flowering over time (Dalvi et al., 2013).

Achieving high uniformity in coffee fruit maturation is a major challenge for the sector (Pimenta et al., 2018). Ideally, the harvest should present a majority of ripe fruits, as it is possible to obtain a final product with greater value from the cherry (ripe) coffee fruit. Harvesting larger amounts of unripe fruits or in the senescence phase (overripe) results in qualitative losses due to changes in type, drinkability, flavor, and aroma (Mesquita et al., 2008).

The knowledge of the maturation stage in coffee crops can guide agronomic treatments, management of labor resources, and planning for the future sales market. Traditionally, the maturation stage is determined by evaluating the color in samples of coffee fruits. This evaluation can be carried out visually or using colorimeters. The downside of the traditional method is the low density of sampling, which results in poor spatial representation of the coffee maturation stage. Because the imaging of fruits is an important source of information regarding their quality (Mazzia et al., 2020), the creation of a system to obtain such information in high sampling densities is essential for a trustworthy spatial representation and to assign value to the product. In this context, computer vision has been shown as a promising technique, especially with the ability to detect objects (Bresilla et al., 2019; Roy et al., 2019; Sa et al., 2016; Song et al., 2014; Tu et al., 2020; Yu et al., 2019) and provide a detailed pixel-based characterization of the color uniformity of objects (De Oliveira et al., 2016; Leme et al., 2019; Mazen and Nashat,

* Corresponding author.
E-mail addresses: helizanicouto@usp.br (H.C. Bazame), jpmolin@usp.br (J.P. Molin), daniel.althoff@ufv.br (D. Althoff), mauriciomartello@usp.br (M. Martello).

https://doi.org/10.1016/j.compag.2021.106066
Received 19 October 2020; Received in revised form 11 February 2021; Accepted 23 February 2021
Available online 4 March 2021

2019; Wu and Sun, 2013). This technique not only presents higher accuracy, but has been documented to be more versatile, faster, and can reduce labor costs (Belan et al., 2012).

Despite recent advances in computer vision techniques, only a few studies have been carried out to identify coffee fruits and to classify their maturation stage. Ramos et al. (2017) developed a machine vision system for mobile devices capable of identifying and classifying coffee fruits on branches and regardless of environmental conditions. Avendano et al. (2017) developed a system for classifying vegetative structures of coffee branches based on obtaining 2D and 3D features from videos acquired in the field. Ramos et al. (2018) determined the maturation stage of the coffee fruit on the plant by processing images acquired over the coffee harvesting period. However, the knowledge of this information in high spatial resolution in the coffee crop is still an unfilled gap. Despite the benefits of evaluating the maturation stage on the plant, this practice is not effective for a wide evaluation of the uniformity of maturation of the coffee at the field level. In this sense, some works already show advances in the spatial evaluation of crop attributes, such as the productive load in orchards and other agricultural crops (Häni et al., 2018; Koirala et al., 2019; Liu et al., 2020; Wang et al., 2019).

The advance in monitoring the spatial and temporal variability of coffee cultivation should enable the development of maps that present essential information in the diagnosis of crop variability and, consequently, in the efficient use of precision agriculture techniques (Molin et al., 2010). This information would make it possible to carry out localized interventions, with the objective not only of increasing coffee productivity, but also of increasing quality and, consequently, the amount paid for the harvested product. In addition, it would assist producers in organizing their crops, planning the number of workers needed in the grain processing stage, preparing the facilities for post-harvest service, carrying out machine maintenance and having greater control of their properties, assisting in decision making in future years (Ramos et al., 2017). Given the above, this work aims to develop and implement an algorithm based on a computer vision technique to detect, classify, and map coffee fruits during harvest.

2. Material and methods

2.1. Image acquisition

The images of the coffee fruits used in this study are frames derived from videos collected during the coffee harvest period from May 31 to June 06, 2020. The two experimental areas where harvest took place are cultivated with commercial coffee crops of the Arabica specie, variety Catuaí 144, and are situated in Patos de Minas region, in the state of Minas Gerais, Brazil (Fig. 1). The coffee crops were planted in the year 2006 and 2003 for experimental areas 1 and 2, respectively, at a density of 5000 trees hectare^{-1} with 4.0 m spacing between lines and 0.5 m between plants.

The platform used to collect videos during harvest is shown in Fig. 2. The platform consisted of a structure assembled at the side spout of a coffee harvester, just after the transverse elevator. The spout gutter was illuminated by a set of six LED lamps totaling 21 W. The videos were recorded using a camera of the mechanical shutter with a 1″ complementary metal oxide semiconductor (CMOS) sensor and 20MP. The camera was stabilized with a 3-axis Gimbal. The videos were recorded with full HD definition (1920x1080) and 100Mbps bit rate (60 fps, 720P, ISO 1600, Shutter 1/800). The frames from all videos were extracted as images and used in the study. The videos were recorded under natural conditions of daylight and including disturbances in illumination, vibration, occlusion, and overlapping of coffee fruits.

The harvest was carried out mechanically using a K3 Millennium coffee harvester (Jacto, Pompeia, Brazil), equipped with a yield monitor as described by Sartori et al. (2002). The coffee harvest operated at a working speed ranging from 0.2 to 0.7 m s^{-1} with the image capturing platform onboard.

2.2. Detection and classification of coffee fruits

The algorithm for coffee fruit detection and classification was implemented using open source neural network structures written in C and known as Darknet. The detection and classification were carried out using an object detection system called You Only Look Once (YOLO) (Redmon et al., 2016). This specific object detection system is popular for its high processing speed, as it processes images and, in a single step, predicts the objects bounding boxes and probabilities of belonging to classes.

2.2.1. YOLO background

With the goal of achieving state-of-art-performance, equivalent to other reference methods, but still maintaining its high processing speed, successive adaptations were implemented on the original version of the YOLO classifier network (YOLOv2 and YOLOv3) (Redmon and Farhadi, 2018, 2017). For detection and classification, the system presents (i) a bounding box prediction (or surrounding rectangle) of the object and (ii) a class prediction. The bounding box prediction is performed in each cell of the network prediction tensor where three bounding boxes are predicted considering three "anchor boxes" as base. Although the algorithm has the ability to adjust the dimensions of the bounding boxes, standard initial dimensions are configured for "anchor boxes" in order to facilitate network learning. The settings of the "anchor boxes" are obtained by applying the k-means clustering algorithm to the bounding boxes marked for the training dataset. For each bounding box, five predictions

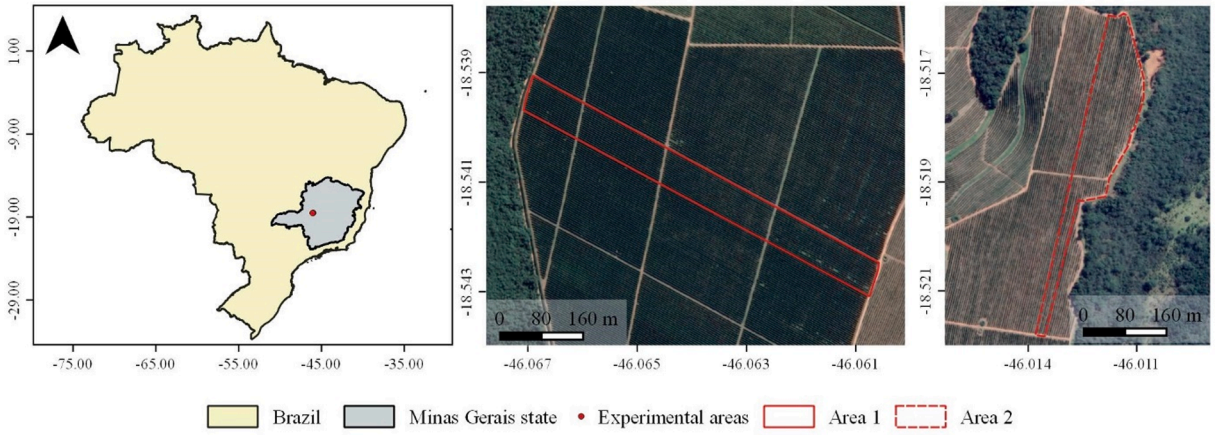


Fig. 1. Location of the experimental areas of coffee crops in the state of Minas Gerais, Brazil. Coordinate reference system: WGS84.

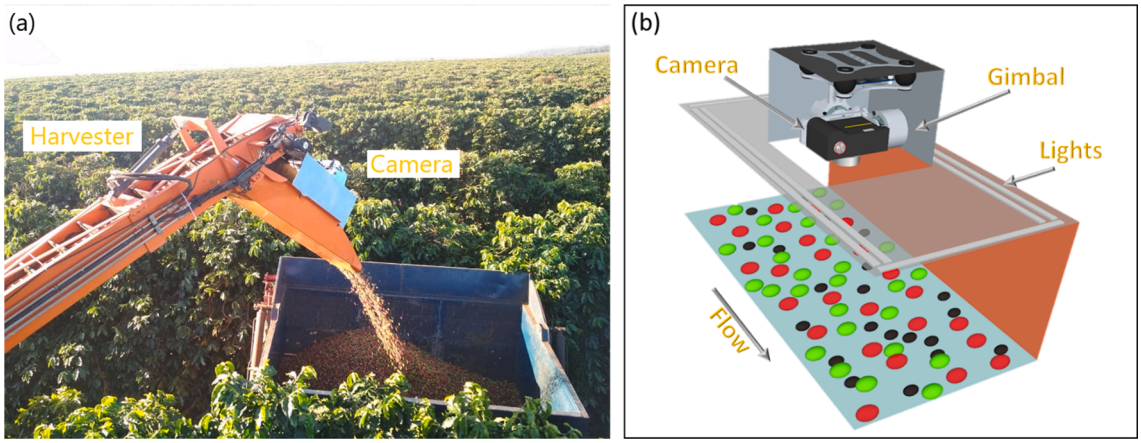


Fig. 2. Lighting and image acquisition platform mounted onboard the coffee harvester.

are made. Four of these predictions are related to the coordinates of the bounding box and one is the "objectness" prediction, i.e., probability of having an object in the bounding box predicted. During training, the sum of squared errors is used as objective function for the predictions of the bounding box coordinates, while the objectness prediction is performed using a logistic regression. The classes a box may contain are predicted for each bounding box using a multilabel classification. The multilabel classification is done with independent logistic classifiers, that is, one classifier for each class. The binary cross-entropy is used for the class predictions during training.

Along the YOLOv3 feature extractor, predictions are made at three different scales. That is, after a series of convolutional layers, the first prediction is made at a scale 32x smaller than the input image. Then, feature maps from earlier layers are concatenated with upscaled features of the current layer and more convolutional layers are added to make the second prediction. The second prediction is performed on a tensor twice the size of the first prediction. The process is repeated for the prediction on the third and last scale. This allows predictions to benefit from refined features and initial feature extractor computations (Redmon and Farhadi, 2018).

Adapted from the YOLOv3 version, YOLOv3-tiny is a network with a simplified feature extractor and with fewer convolutional layers. Similar to YOLOv3, YOLOv3-tiny also performs prediction across different scales, but only on two different scales. Despite being a less robust network, high performance is expected for simpler tasks, such as detecting circular objects. The advantage of using a simpler network is it can be trained more quickly and presents greater detection speed.

2.2.2. *Network design and training*

The types of layers, number of filters, size and stride, size of input and output tensor for each layer of the network used are shown in Table 1. This table presents an example for a network which input resolution is 608 × 608, termed here as YOLOv3-tiny-608. Considering that three anchor boxes are used per cell in the prediction tensor and that the coffee fruits classification was performed for three different classes (unripe, ripe, and overripe), the prediction layers present 24 predictions as output [(4 coordinates + 1 object prob. + 3 classes) × 3 anchors] (layers 16 and 23). The scheme for object detection and classification is shown in Fig. 3.

In addition, other dimensions of image compression (416 × 416, 608 × 608, 704 × 704, 800 × 800, and 896 × 896 pixels) for the input in the YOLOv3-tiny network were evaluated. The default anchors were recalculated using the Darknet function before training for each of the different image resolutions used. The resolution of input images for the network can be changed during inference for the detection of objects in different resolutions without the need for further training. To avoid a demand for a large number of images to train the object detection and classification model, a transfer learning technique was adopted. The model was fine-tuned on parameters (or weights) of a YOLOv3-tiny model pre-trained on the COCO data set (80 object categories, 1.5

Table 1
Example of the structure of the YOLOv3-tiny-608 model used in the study.

Layer	Type	Filters	Size/ stride	Input	Output
0	Convolutional	16	3 × 3/1	608 × 608 × 3	608 × 608 × 16
1	[1]Maxpool		2 × 2/2	608 × 608 × 16	304 × 304 × 16
2	Convolutional	32	3 × 3/1	304 × 304 × 16	304 × 304 × 32
3	Maxpool		2 × 2/2	304 × 304 × 32	152 × 152 × 32
4	Convolutional	64	3 × 3/1	152 × 152 × 32	152 × 152 × 64
5	Maxpool		2 × 2/2	152 × 152 × 64	76 × 76 × 64
6	Convolutional	128	3 × 3/1	76 × 76 × 64	76 × 76 × 128
7	Maxpool		2 × 2/2	76 × 76 × 128	38 × 38 × 128
8	Convolutional	256	3 × 3/1	38 × 38 × 128	38 × 38 × 256
9	Maxpool		2 × 2/2	38 × 38 × 256	19 × 19 × 256
10	Convolutional	512	3 × 3/1	19 × 19 × 256	19 × 19 × 512
11	Maxpool		2 × 2/1	19 × 19 × 512	19 × 19 × 512
12	Convolutional	1024	3 × 3/1	19 × 19 × 512	19 × 19 × 1024
13	Convolutional	256	1 × 1/1	19 × 19 × 1024	19 × 19 × 256
14	Convolutional	512	3 × 3/1	19 × 19 × 256	19 × 19 × 512
15	Convolutional	24	1 × 1/1	19 × 19 × 512	19 × 19 × 24
16	[2]YOLO				
17	**Route 13**				
18	Convolutional	128	1 × 1/1	19 × 19 × 256	19 × 19 × 128
19	Up-sampling		2 × 2/1	19 × 19 × 128	38 × 38 × 128
20	**Route 19 and 8**				
21	Convolutional	256	3 × 3/1	38 × 38 × 384	38 × 38 x256
22	Convolutional	24	1 × 1/1	38 × 38 × 256	38 × 38 × 24
23	YOLO				

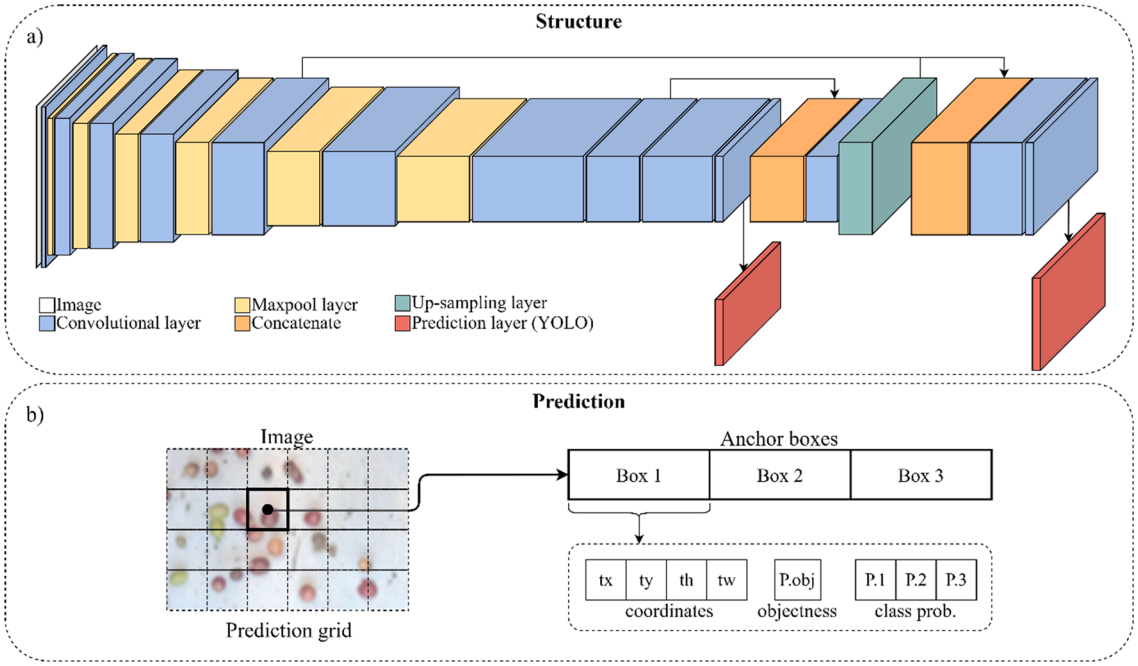


Fig. 3. (a) Structure of the YOLOv3-tiny network adapted for the study (b) the prediction scheme.

million object instances, and 330 thousand images) (Lin et al., 2014). Using weights pre-trained on a more robust database provides the model with greater capacity to extract different types of features. Although the COCO data set does not contain classes of the maturation stages of coffee fruits, the ability of the pre-trained model to extract certain features can be transferred to the new model. The pre-trained network was, therefore, additionally trained using images of the object of study (coffee fruits).

The "data augmentation" technique was also used during training. This technique randomly rotated the images and increased or decreased their exposure. Thus, the YOLOv3-tiny model is trained on a more diversified data set despite the decreased number of images in the training set (Koirala et al., 2019).

The model's responses to the images provided are the detection of the object and the class to which they belong. Thus, estimating qualitative parameters by classifying the maturation stage of coffee fruits. The fine-tuning of the network aimed to classify the objects of interest into three classes: (i) green or unripe coffee fruits; (ii) cherry or ripe coffee fruits; and (iii) raisin or overripe coffee fruits. A set of 400 images of coffee fruits, derived from frames of videos collected during the harvest, were used for the implementation of the model, of which, 280 images were used as the training set and 120 images were used as the test set (validation). The labeling of the images was performed using the Yolo mark (Bochkovskiy, 2019). The Yolo mark is a graphical user interface designed for marking the bounding boxes of objects of interest. The fruits were labeled according to the authors' visual classification based on color. The confidence threshold and the non-maximum suppression (NMS) threshold were set at 0.25 and 0.50, respectively. The model was run on a computer with the following specifications: Core™ i7-8700HQ CPU, 3.2 GHz, 64 GB RAM, and GeForce RTX 2070 graphics card with 8 GB dedicated memory. The model was implemented based on the OpenCV library and programmed in C language using the Darknet framework.

2.3. Performance criteria

The performance of the model was assessed by comparing samples of coffee fruits classified by the algorithm with the visual classification of these fruits (traditional). The criteria used to assess the performance on the training and test sets were the precision, recall, and F1-Score, which were calculated as described in Eqs. (1)–(3) (Olson and Delen, 2008) below.

$$Precision = \frac{TP}{TP + FP} \tag{1}$$

$$Recall = \frac{TP}{TP + FN} \tag{2}$$

$$F1 - Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{3}$$

where TP denotes the true positives, FP the false positives, and FN the false negatives.

In addition, the average precision of the classes (AP) (Eqs. (4) and (5)) and the mean average precision (mAP) at an intersection over the union of 50% were also calculated. After calculating the precision and recall for a class in the entire test set, we consider the average precision (AP) as the area under the Precision × Recall curve. The average precision is the precision averaged at all recall values between 0 and 1 (Mazzia et al., 2020):

$$\int_0^1 P(r)dr \tag{4}$$

This is the same as taking the area under the curve. In practice, the integral is approximated closely by a sum of the precision in each possible threshold value, multiplied by the change in the recall:

$$\sum_{k=1}^{N} P(k)\Delta r(k) \tag{5}$$

where N is the total number of images in the collection, P(k) is the precision in a cut of images k, and Δr(k) is the change in recall that occurred between cut k-1 and cut k. The mAP was obtained from the average of APs for each class.

2.4. Mapping the coffee fruits maturation stage

The geographic coordinates of each video recording, along the

harvest lines, were collected at 20 Hz with C/A code type Global Navigation Satellite System (GNSS) receiver, with Global Position System (GPS) and Globalnaya Navigatsionnaya Sputnikovaya Sistema (GLONASS) signal. Since the video frames were recorded at 60 Hz, available coordinates were matched to frames considering the time of record, and the remaining video frames were interpolated between available records. The detections from each frame were summarized in the total number of coffee fruits detected and the percentage of coffee fruits detected for each maturation stage. The summary information of the detections for each frame was then assigned to the corresponding geographic coordinates. The maps of the maturation stage of coffee fruits were developed using 90 videos collected during harvest. The interpolation of the video frame coordinates and assignment of detection information to the georeferenced points were performed using the R programming language and environment (R Core Team, 2020). The maps for each of the maturation stages of coffee fruits were developed by kriging the georeferenced points.

3. Results and discussion

3.1. Object detection performance

The performance achieved by the YOLOv3-tiny models considering different image input resolutions is presented in Fig. 4. A more significant improvement was observed by increasing the resolution of input images from 416 × 416 pixels to 608 × 608 pixels, while the performance seemed to stabilize for input resolutions above 704 × 704 pixels. The best performance was achieved considering the input resolution of 800 × 800 pixels at the ~ 3300 iteration. From this iteration, it is possible to see that the performance for the test set does not improve, and to continue the training could result in the over-fitting of the model. For this reason, the weights of the model parameters of the 3300th iteration were adopted. The YOLOv3-tiny-800 model showed mAP of 84.0%, F1-Score of 82.0%, the precision of 83.0%, and recall of 82.0% for the validation set, respectively. The metrics showed balanced results, i.e., close values of precision and recall, which is important for the model.

The performance analysis for the YOLOv3-tiny model, based on the three classes of fruit ripening stage (unripe, ripe, and overripe), are shown in Fig. 5. The model presented AP of approximately 86.0%, 85.0%, 80.0% in the classification for the validation set of unripe, ripe, and overripe, respectively. The lesser precision for the classification of overripe fruits leads us to believe that there was some confusion between ripe and overripe fruits at the time of classification, probably due to the proximity of their colors. In other words, the model sometimes failed to adequately classify the overripe fruits leading to a false positive. The model's AP for the coffee fruit detection and classification in the validation set improved when the resolution of the images used as input increased. The performance also peaked when the resolution of 800 × 800 pixels was used as input. This improvement in the result with the increase in the resolution of the images can be explained by the fact that the convolutional and pooling layers used in YOLO gradually decrease the spatial dimension with the increasing depth of the network. Thus, it may be difficult to detect some objects at lower resolutions, e.g., small or overlapped and not obvious fruits (Koirala et al., 2019).

The color of the fruits, leaves, and the amount of impurities not eliminated by the harvester pre-cleaning process can vary according to the cultivar and the growth stage. Therefore, results of detection and classification of fruits based on algorithms that consider only the color to detect objects can vary their results. On the other hand, computer vision algorithms are robust to variations in lighting, vibrations, among others. Additionally, the deep learning technique used in this study does not only consider the object color, as it automatically learns to extract many other features during training. For example, Koirala et al. (2019) evaluated the performance of six deep learning architectures to detect mango fruits in tree crown images. The authors evaluated 1515 images and obtained an F1-Score of 95.1% and a mAP of 96.7% using the YOLOv3-512 network and F1-Score of 90.0% and mAP of 93.8% using the YOLOv2-tiny-416 network.

Mazzia et al. (2020) evaluated an embedded solution in real time inspired by "Edge AI" for apple detection with the implementation of the YOLOv3-tiny algorithm on several embedded platforms. The study proved to be feasible with mAP results in the detection of 83.6%, recall of 83.0%, and precision of 69.0% using a resolution of 30 fps. The authors concluded that even for difficult scenarios such as overlapping apples, complex background, less exposure of the apple due to leaves and branches, the algorithm can detect, count, and measure the size of the apples in real time, which can help farmers and agronomists in decision-making and management of their crops.

A visual assessment of predictions for images in the test set was performed for the YOLOv3-tiny-800 trained model (Fig. 6). In these figures, although most coffee fruits have been identified and classified correctly, some fruits have not been detected nor classified. For example, the mAP was 85.6% for the first arbitrary frame (Fig. 6a) and 72.1% for

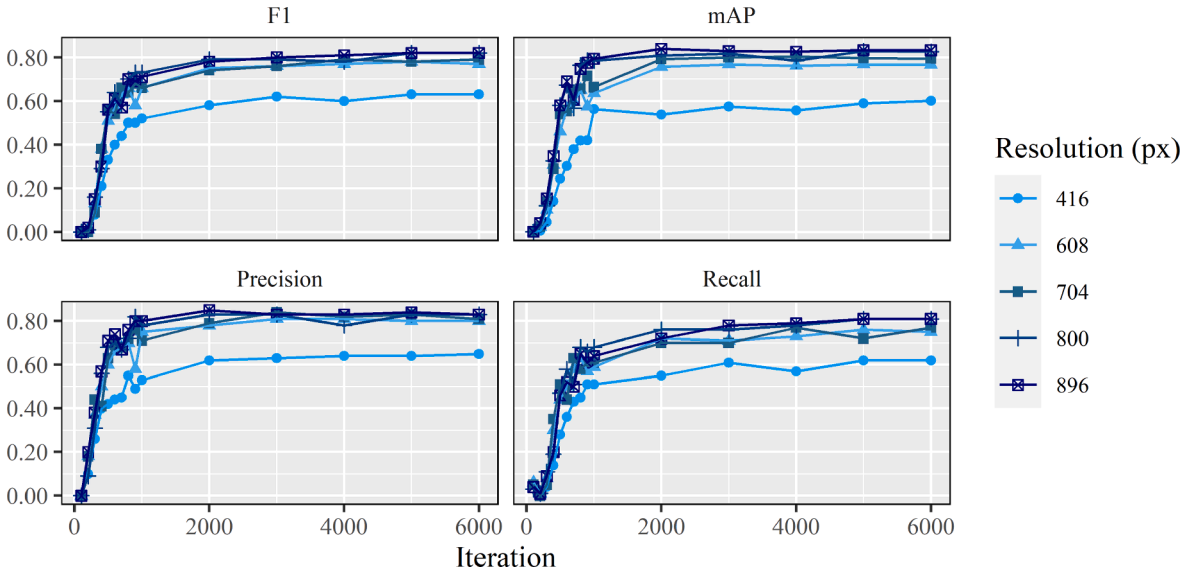


Fig. 4. Performance criteria achieved by the YOLOv3-tiny model for the validation set.

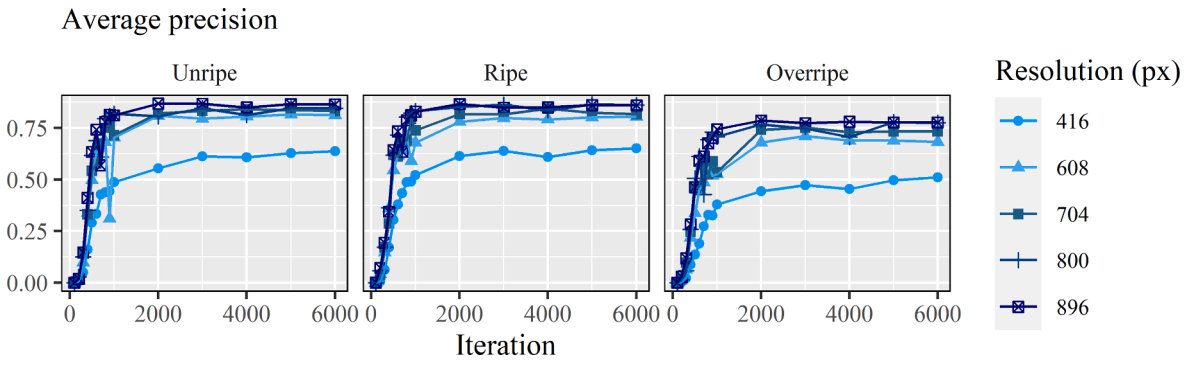


Fig. 5. Average precision achieved by the YOLOv3-tiny model by the class of coffee fruits for the validation set of.

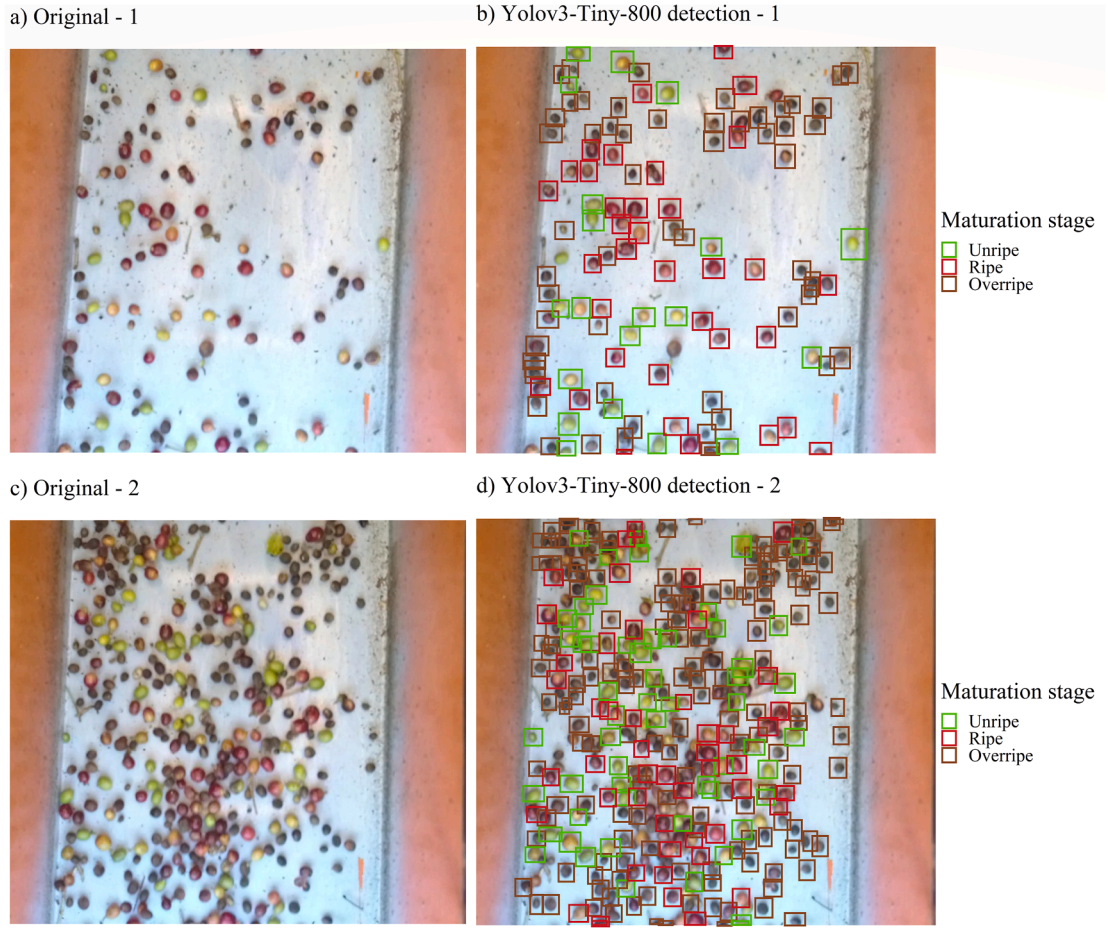


Fig. 6. Two arbitrary video frames collected during the coffee harvest: (a-c) original frames with fruits at different stages of maturation, (b-d) detection performed by the proposed model.

the second arbitrary frame (Fig. 6c). The lower mAP for the second frame can be attributed to the higher density of fruits in the frame. A higher density of fruits can lead to the overlapping of objects of interest and interfere in their detection. In addition, the downscaling of the image to 800 × 800 pixels during detection can result in clusters of fruits of similar colors blurring together.

The fact that the model performs the detection and classification of the object of interest regardless of the scenario leads us to believe that the algorithm is suitable for adverse field situations and can be a good support tool for decision making in coffee farming. However, some working conditions can diminish the confidence of the detections made by the algorithm. For example, high crop yield or high harvester forward speed can result in a high flow rate of fruits inside the spout and interfere in detections. Terrain roughness and irregularities can likewise lead to blurred images and decrease the quality of the frames collected. In contrast, the low computational demand of the YOLOv3-tiny model means that the model can be adapted and embedded, as shown by Mazzia et al. (2020), on the harvesting platform to bring real-time responses during harvest. Such information would enable the fine adjustments of the harvester for a more suitable selective harvest, i.e., the real-time correction of the vibration frequency of the harvester rods. The correct adjustment of the vibration of the rods and the machine's working speed according to the variety, size, and maturation stage of the fruits maximizes the harvest efficiency, regulates the maturation stage of

harvested coffee fruits, and reduces the operational costs (Santinato et al., 2014; Santos et al., 2015; Velloso et al., 2020).

3.2. Quality mapping: Spatial variability of coffee maturation stages

The use of images (video frames) with a small number of detections could increase the variation and distort the real proportion between the detected classes. Thus, to reduce the failures attributed to the detection and incorrect classification of coffee fruits in the images, an analysis of the distribution of the total detected fruits was performed through a time series (Fig. 7). From this analysis, a threshold of 40 fruits was chosen as the minimum number of detected fruits in a frame for the detection to be included in the coffee fruits quality mapping (Fig. 7a). Images with less than 40 fruits were, therefore, excluded from the database (Fig. 7b). This exclusion is also necessary to prevent that detection values are considered in a null detection area or that does not belong to the field.

For an arbitrary video, the total number of coffee fruits detected in each frame of the video is shown in Fig. 8a, the total number of fruits detected per class in Fig. 8b, and the proportion of classes detected per frame in Fig. 8c. Along with the frames, i.e. the harvest line, the proportion of ripe fruits was more or less stable, with a subtle peak of underripe fruits between the frames 4500 and 6500. A large proportion of overripe coffee fruits (31.4% to 75.3%) were detected in relation to ripe (21.0% to 59.6%) and underripe fruits (0–25.6%) (Fig. 8c). The overall higher proportion of overripe fruits is due to the delay to begin the harvest in the experimental areas. These areas were specifically chosen for the experiment and depended on the availability of labor during the peak of the harvest.

For the same video used in Fig. 8, Fig. 9 shows the total of detected fruits in each frame (Fig. 9a) and the detections by class (Fig. 9b) mapped along the respective harvest line. In both Figs. 8a and 9a, it is evident that the beginning of the line presented a higher number of detections. In Fig. 9b, the proportion of ripe fruits seem higher in the beginning to middle of the harvest line. The sensitivity in spatial variation is evidence of the effectiveness of the algorithm in evaluating the different stages of maturation of coffee fruits in the field. This type of information is important not only to improve the quality of the coffee drink but also to optimize production, since coffee growers could, in advance, better appreciate the contribution of the coffee maturity stage to the final quality of the drink (Baluja et al., 2013).

The coffee fruit classification at different maturation stages was spatialized, generating maps with information related to crop quality attributes (Fig. 10). Despite only one variety of coffee being cultivated in this area, the spatial variability of the qualitative attributes of the coffee crop can occur due to several factors such as microclimate, altitude, side exposure to the sun, soil type and physical-chemical properties, phenological characteristics of the plants, humidity, microfauna, among others. In Area 1 (Fig. 10a), a higher percentage of underripe fruits was harvested in relation to Area 2 (Fig. 10b), presenting an overall more balanced proportion between classes of maturation stage (Fig. 10c). Area 2 should preferably have had an earlier harvest when compared to Area 1, with a higher percentage of overripe coffee fruits. The higher percentage of overripe fruits in Area 2 is mainly located in the eastern region of the area, which may be explained by the positioning of the field. The eastern side of Area 2 is in the lowest part of the field and closest to native vegetation. The border influence from the native forest may have favored the early flowering and ripening of the coffee fruits. In commercial areas, this variability is expected. However, without technological approaches like the one presented in this study, there are no easy means to identify these aspects with high resolution and assertiveness in the field.

The identification of the different stages of maturation in the same field, consistent over the years, is valuable information for indicating preferential areas for the beginning of the harvest. The late harvest generally results in lower quality of the beverage (Läderach et al., 2011). The mapping of coffee quality also opens the opportunity to prevent the mixing of coffee fruits at different stages of maturation, favoring allotments with a higher percentage of cherry coffee fruits over unripe and overripe coffee fruits for the production of specialty coffees. This information can be of great help for coffee farmers, because such information influences the final cost of coffee production since it affects the costs of harvesting, drying handling, storage, storage infrastructure, processing, and other operations (Donizetti et al., 2011). In addition, this information becomes even more important for making it feasible to plan and include precision agriculture techniques in crop management and decision-making planning in upcoming years. However,

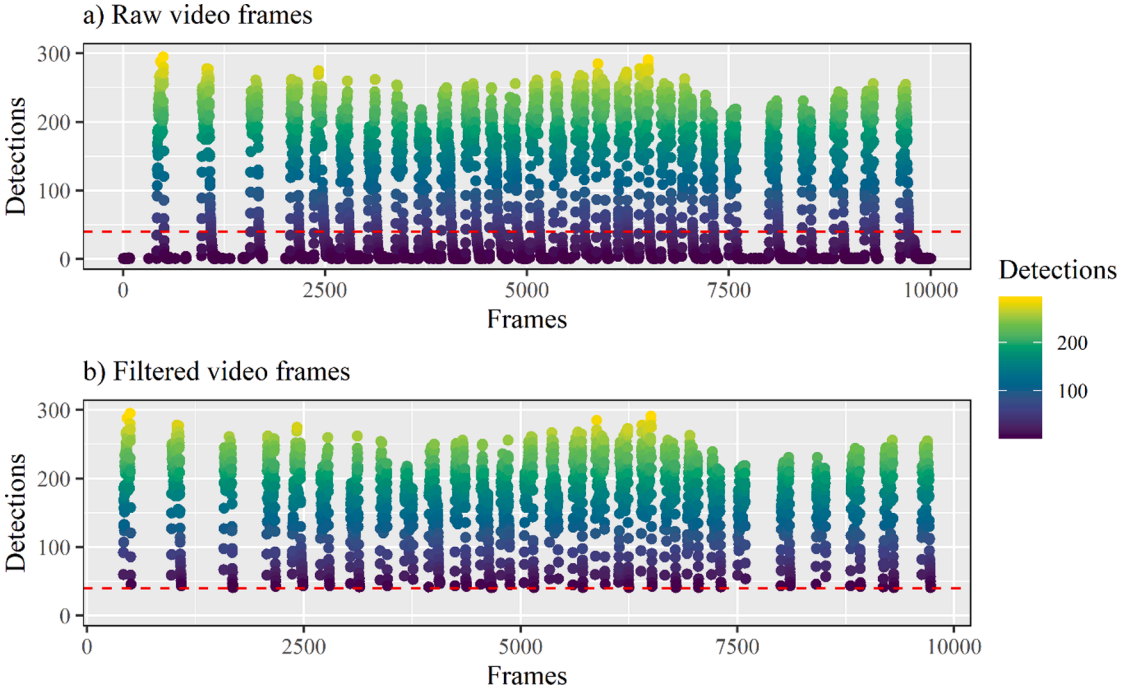


Fig. 7. Time series used to determine the threshold of minimum fruit count for quality mapping.

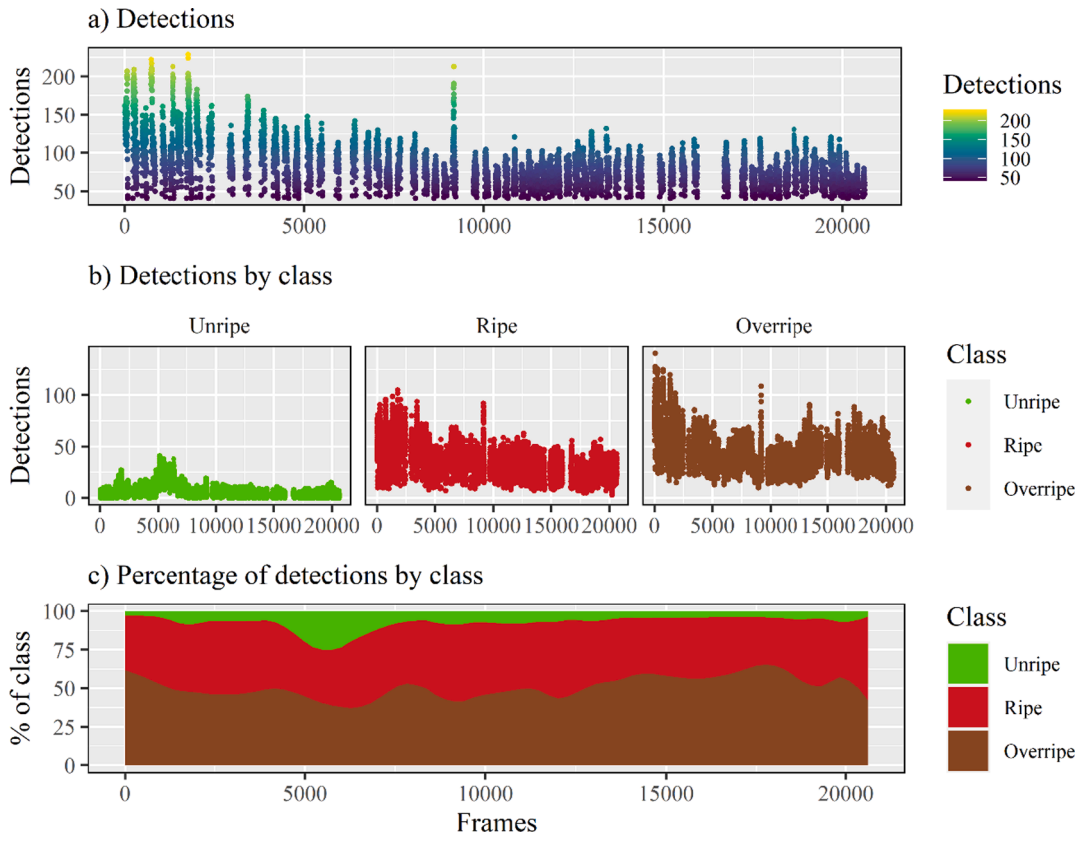


Fig. 8. For an arbitrary video, (a) the total number of detected fruits, (b) the total number of detections by ripening class, and (c) the proportion of the detected classes.

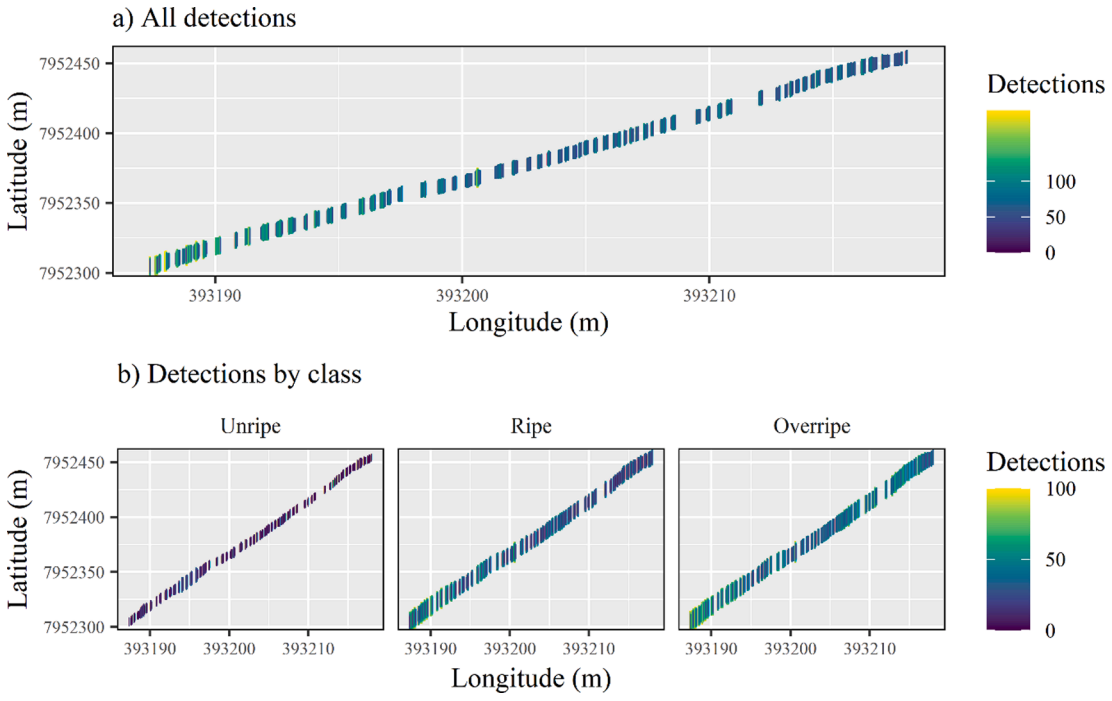


Fig. 9. For an arbitrary video, mapping of (a) total coffee fruit detections and (b) class detections along the harvest line. Coordinate reference system: WGS84 / UTM zone 23S.

recommendations are generally site-specific and difficult to generalize (Läderach et al., 2011). According to Läderach et al. (2011), practices such as shading management, soil sampling, harvesting period, fruit thinning, etc., can be better determined by on-farm experimentation by the own farmer. The practices should be evaluated by the farmer's criteria on whether they result in economic or environmental benefits.

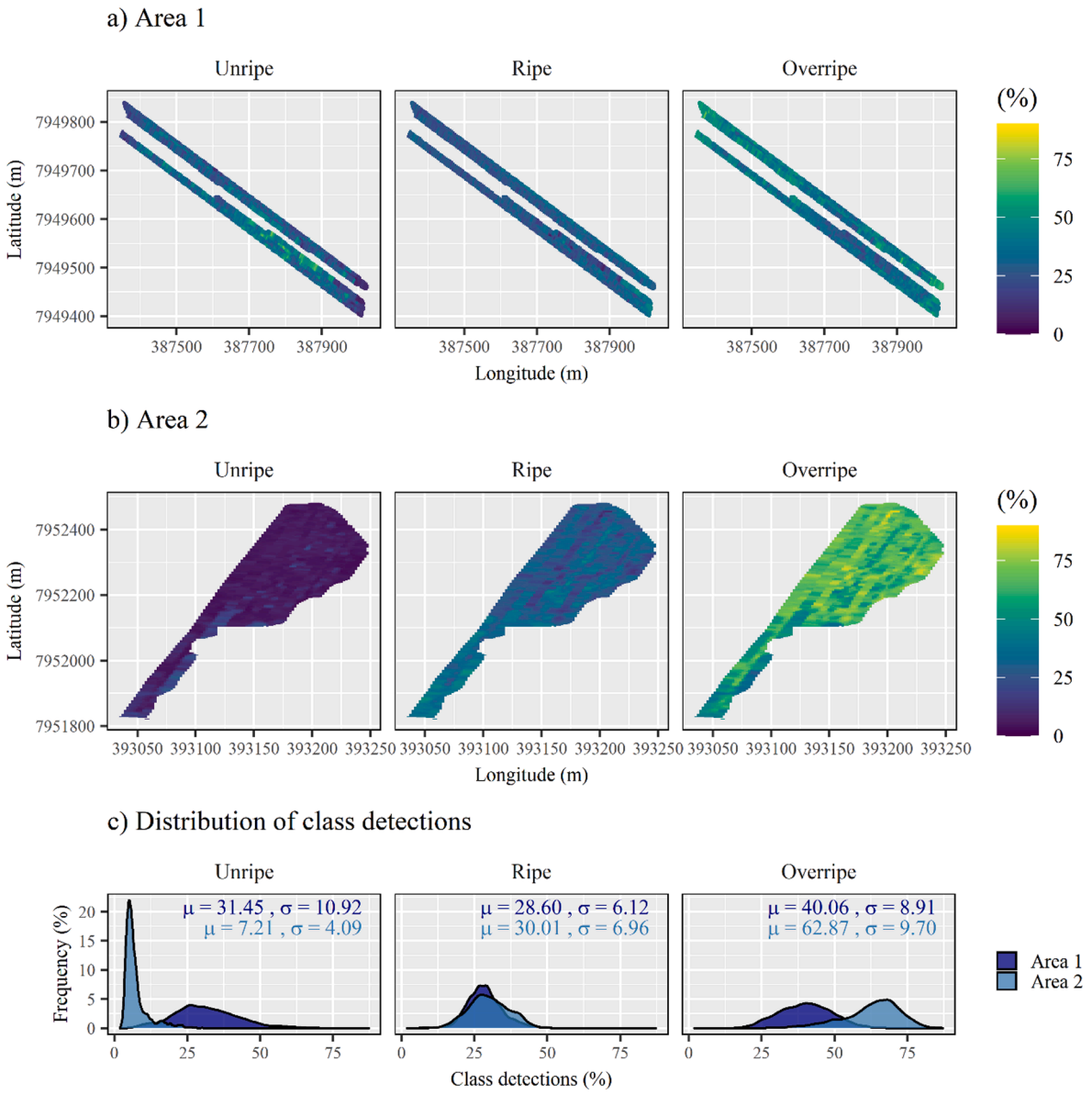


Fig. 10. Proportions of detections mapped for (a) Area 1 and (b) Area 2, and (c) the distributions of class detections (μ = mean; σ = standard deviation). Coordinate reference system: WGS84 / UTM zone 23S.

The maps that identify the spatial variability present in crops are extremely important to improve management initiatives that seek to understand this variability to manage it efficiently through precision agriculture practices (Koirala et al., 2020). The effect of factors associated with coffee maturation, reflected in the quality of the drink, justifies the assessment of the variability of attributes or characteristics of the cultivated area. That is, it is possible to conduct guided sampling to explore the possible factors that may have led to these results. This information could, consequently, support decision-making related to agricultural management practices (Ferraz et al., 2019), especially in the scope of delimiting management zones with specificities of each location. Once the management zones are defined, new actions can be taken. Therefore, the coffee crop would be ideal for proposing a precision harvesting project (Kazama et al., 2020).

4. Conclusions

The deep learning model adopted in this study supports the possibility of detecting coffee fruits and classifying their maturation stage regardless of the context they are in, i.e., contrasts between the fruit and the background, lighting and vibration conditions, harvest angle, etc. The structure of the object detection system, based on the architecture of the YOLOv3-tiny neural networks, proved to be robust and computationally efficient. The model presented as performance criteria a mAP of 84.0%, F1-Score of 82.0%, the precision of 82.0%, and recall of 83% for the validation set. The average precision for the classes of unripe, ripe, and overripe coffee fruits was 86.0%, 85.2%, and 80.0%, respectively. The low computational demand of the "tiny" version of the YOLOv3 model means that it is a great candidate, as also shown in other studies, to be adapted and embedded to bring responses in real-time during the harvest. This would enable fine adjustments in real-time for a better selective harvest.

The model was used for the detection and classification of coffee fruits in videos recorded during the coffee harvest, making it possible to map the qualitative attribute of the maturation stage of coffee over the experimental area. Mapping this attribute provides managers with information that allows the introduction of precision agriculture techniques in coffee crop management. The options for obtaining

information on the coffee maturation stage are still limited and laborious. The platform used in this study proves efficient for data collection and can be implemented for any type of coffee harvester.

This study can support future research in coffee growing to, for example, investigate the differences in maturation stage within coffee rows, analyze soil samples in the search for correlations with coffee maturation, optimize the harvester speed and the vibration of harvester rods, etc. The detection of coffee fruits in consecutive frames can also lead to the same fruits being accounted for multiple times. Thus, future research should aim for object tracking techniques that can better assess the number of coffee fruits and even generate crop yield maps.

CRediT authorship contribution statement

Helizani Couto Bazame: Conceptualization, Methodology, Software, Formal analysis, Investigation, Data curation, Writing - original draft, Visualization. **José Paulo Molin:** Conceptualization, Writing - original draft, Supervision, Resources, Project administration, Funding acquisition. **Daniel Althoff:** Methodology, Software, Formal analysis, Data curation, Writing - original draft, Visualization. **Maurício Martello:** Investigation, Formal analysis, Writing - original draft, Visualization, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

To the Guima Café Group for the experimental area, Terrena Agronegócios for supporting research, and the Coordination for the Improvement of Higher Education Personnel (in Portuguese: Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - CAPES) for granting the scholarship of authors 1 and 4 - Finance Code 001.

References

Avendano, J., Ramos, P.J., Prieto, F.A., 2017. A system for classifying vegetative structures on coffee branches based on videos recorded in the field by a mobile device. Expert Syst. Appl. 88, 178–192. https://doi.org/10.1016/j.eswa.2017.06.044.

Baluja, J., Tardaguila, J., Ayestaran, B., Diago, M.P., 2013. Spatial variability of grape composition in a Tempranillo (Vitis vinifera L.) vineyard over a 3-year survey. Precis. Agric. 14, 40–58. https://doi.org/10.1007/s11119-012-9282-5.

Belan, P.A., de Araújo, S.A., Librantz, A.F.H., 2012. Técnicas de visão computacional aplicadas no processo de calibração de instrumentos de medição com display numérico digital sem interface de comunicação de dados. Exacta 10. https://doi.org/10.5585/exacta.v10n1.3091.

Bochkovskiy, A., 2019. Yolo_mark: Windows & Linux GUI for marking bounded boxes of objects in images for training neural network. https://github.com/AlexeyAB/Yolo_mark. Accessed 06 Jun. 2020.

Bresilla, K., Perulli, G.D., Boini, A., Morandi, B., Corelli Grappadelli, L., Manfrini, L., 2019. Single-shot convolution neural networks for real-time fruit detection within the tree. Front. Plant Sci. 10 https://doi.org/10.3389/fpls.2019.00611.

Dalvi, L.P., Sakiyama, N.S., Pereira Da Silva, F.A., Cecon, P.R., 2013. Revista Agrarian 410 Qualidade de café nos estádios cereja e verde-cana via condutividade elétrica Quality of coffee cherry and sugarcane-green stage by electrical conductivity 410–414.

De Oliveira, E.M., Leme, D.S., Barbosa, B.H.G., Rodarte, M.P., Alvarenga Pereira, R.G.F., 2016. A computer vision system for coffee beans classification based on computational intelligence techniques. J. Food Eng. 171, 22–27. https://doi.org/10.1016/j.jfoodeng.2015.10.009.

Donizetti, A., Luciana, D.;, Vieira, M., Mendonça2, L., Antônio, R., Dias, A., Agda, ;, Prado, S., Rodrigo, ;, Batista, E., Sérgio, ;, Pereira, P., 2011. QUALIDADE DO CAFÉ COLHIDO EM DIFERENTES ESTÁDIOS DE MATURAÇÃO.

Ferraz, M.N., de Corrêdo, L.P., Wei, M.C.F., Molin, J.P., 2019. Spatial variability mapping of sugarcane qualitative attributes. Eng. Agrícola 39, 109–117. https://doi.org/10.1590/1809-4430-eng.agric.v39nep109-117/2019.

Häni, N., Roy, P., Isler, V., 2018. Apple counting using convolutional neural networks. In: IEEE International Conference on Intelligent Robots and Systems. Institute of Electrical and Electronics Engineers Inc, pp. 2559–2565. https://doi.org/10.1109/IROS.2018.8594304.

Kazama, E.H., da Silva, R.P., de Tavares, T.O., Correa, L.N., de Lima Estevam, F.N., de Nicolau, F.E.A., Maldonado Jr, W., 2020. Methodology for selective coffee harvesting in management zones of yield and maturation. Precis. Agric. 1–23 https://doi.org/10.1007/s11119-020-09751-1.

Koirala, A., Walsh, K.B., Wang, Z., Anderson, N., 2020. Deep learning for mango (Mangifera indica) panicle stage classification. Agronomy 10, 1–22. https://doi.org/10.3390/agronomy10010143.

Koirala, A., Walsh, K.B., Wang, Z., McCarthy, C., 2019. Deep learning for real-time fruit detection and orchard fruit load estimation: benchmarking of 'MangoYOLO'. Precis. Agric. 20, 1107–1135. https://doi.org/10.1007/s11119-019-09642-0.

Läderach, P., Oberthür, T., Cook, S., Estrada Iza, M., Pohlan, J.A., Fisher, M., Rosales Lechuga, R., 2011. Systematic agronomic farm management for improved coffee quality. F. Crop. Res. 120, 321–329. https://doi.org/10.1016/j.fcr.2010.10.006.

Leme, D.S., da Silva, S.A., Barbosa, B.H.G., Borém, F.M., Pereira, R.G.F.A., 2019. Recognition of coffee roasting degree using a computer vision system. Comput. Electron. Agric. 156, 312–317. https://doi.org/10.1016/j.compag.2018.11.029.

Lin, T.Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., Zitnick, C. L., 2014. Microsoft COCO: Common objects in context. Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics) 8693 LNCS, 740–755. https://doi.org/10.1007/978-3-319-10602-1_48.

Liu, G., Nouaze, J.C., Mbouembe, P.L.T., Kim, J.H., 2020. YOLO-tomato: A robust algorithm for tomato detection based on YOLOv3. Sensors (Switzerland) 20. https://doi.org/10.3390/s20072145.

Matiello, Santinato, Almeida, Garcia, 2015. Cultura de Café no Brasil: Manual de recomendações, 1a. ed. Fundacão procafé.

Mazen, F.M.A., Nashat, A.A., 2019. Ripeness classification of bananas using an artificial neural network. Arab. J. Sci. Eng. 44, 6901–6910. https://doi.org/10.1007/s13369-018-03695-5.

Mazzia, V., Khaliq, A., Salvetti, F., Chiaberge, M., 2020. Real-time apple detection system using embedded systems with hardware accelerators: An edge AI application. IEEE Access 8, 9102–9114. https://doi.org/10.1109/ACCESS.2020.2964608.

Mesquita, C.M. de, Rezende, J.E. de, Morais, N.C., Dias, P.T., Carvalho, R.M. de, 2008. Manual do café colheita e preparo.

Molin, J.P., de Araujo Motomiya, A.V., Frasson, F.R., di Chiacchio Faulin, G., Tosta, W., 2010. Método para avaliação de aplicaç ão de fertilizantes em taxa variávelem café. Acta Sci. - Agron. 32, 569–575. https://doi.org/10.4025/actasciagron.v32i4.5282.

Olson, D.L., Delen, D., 2008. Advanced data mining techniques, Advanced Data Mining Techniques. Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-76917-0.

Pimenta, C.J., Angélico, C.L., Chalfoun, S.M., 2018. Challengs in coffee quality: Cultural, chemical and microbiological aspects. Cienc. e Agrotecnologia. https://doi.org/10.1590/1413-70542018424000118.

Ramos, P.J., Avendaño, J., Prieto, F.A., 2018. Measurement of the ripening rate on coffee branches by using 3D images in outdoor environments. Comput. Ind. 99, 83–95. https://doi.org/10.1016/j.compind.2018.03.024.

Ramos, P.J., Prieto, F.A., Montoya, E.C., Oliveros, C.E., 2017. Automatic fruit count on coffee branches using computer vision. Comput. Electron. Agric. 137, 9–22. https://doi.org/10.1016/j.compag.2017.03.010.

Redmon, J., Divvala, S., Girshick, R., Farhadi, A., 2016. You only look once: Unified, real-time object detection. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE Computer Society, pp. 779–788. https://doi.org/10.1109/CVPR.2016.91.

Redmon, J., Farhadi, A., 2018. YOLO vol 3. Tech Rep. 1–6.

Redmon, J., Farhadi, A., 2017. YOLO9000: Better, faster, stronger. Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017 2017-Janua, 6517–6525. https://doi.org/10.1109/CVPR.2017.690.

Roy, P., Kislay, A., Plonski, P.A., Luby, J., Isler, V., 2019. Vision-based preharvest yield mapping for apple orchards. Comput. Electron. Agric. 164, 104897 https://doi.org/10.1016/j.compag.2019.104897.

Sa, I., Ge, Z., Dayoub, F., Upcroft, B., Perez, T., McCool, C., 2016. Deepfruits: A fruit detection system using deep neural networks. Sensors (Switzerland) 16. https://doi.org/10.3390/s16081222.

Santinato, F., Pereira Da Silva, R., Cassia, M.T., Santinato, R., 2014. ANÁLISE QUALI-QUANTITATIVA DA OPERAÇÃO DE COLHEITA MECANIZADA DE CAFÉ EM DUAS SAFRAS QUALITY OF OPERATION OF HARVESTING OF COFFEE AT TWO CROPS, Coffee Science.

Santos, F.L., de Queiroz, D.M., Valente, D.S.M., de Coelho, A.L.F., 2015. Simulação do comportamento dinâmico do sistema fruto-pedúnculo do café empregando o método de elementos finitos. Acta Sci. - Technol. 37, 11–17. https://doi.org/10.4025/actascitechnol.v37i1.19814.

Sartori, S., Fava, J.F.M., Domingues, E.L., Ribeiro Filho, A.C., Shiraisi, L.E., 2002. Mapping the Spatial Variability of Coffee Yield with Mechanical Harvester. In: Proceedings of the World Congress of Computers in Agriculture and Natural Resources, pp. 196–205. https://doi.org/10.13031/2013.8330.

Song, Y., Glasbey, C.A., Horgan, G.W., Polder, G., Dieleman, J.A., van der Heijden, G.W. A.M., 2014. Automatic fruit recognition and counting from multiple images. Biosyst. Eng. 118, 203–215. https://doi.org/10.1016/j.biosystemseng.2013.12.008.

Tu, S., Pang, J., Liu, H., Zhuang, N., Chen, Y., Zheng, C., Wan, H., Xue, Y., 2020. Passion fruit detection and counting based on multiple scale faster R-CNN using RGB-D images. Precis. Agric. https://doi.org/10.1007/s11119-020-09709-3.

Velloso, N.S., Rodrigues Magalhães, R., Lúcio Santos, F., Assis Rezende Santos, A., 2020. Modal properties of coffee plants via numerical simulation. Comput. Electron. Agric. 175, 105552 https://doi.org/10.1016/j.compag.2020.105552.

Wang, Z., Walsh, K., Koirala, A., 2019. Mango fruit load estimation using a video based MangoYOLO—Kalman filter—hungarian algorithm method. Sensors (Switzerland) 19. https://doi.org/10.3390/s19122742.

Wu, D., Sun, D.W., 2013. Colour measurements by computer vision for food quality control - A review. Trends Food Sci. Technol. https://doi.org/10.1016/j.tifs.2012.08.004.

Yu, Y., Zhang, K., Yang, L., Zhang, D., 2019. Fruit detection for strawberry harvesting robot in non-structural environment based on Mask-RCNN. Comput. Electron. Agric. 163, 104846 https://doi.org/10.1016/j.compag.2019.06.001.