University of São Paulo
"Luiz de Queiroz" College of Agriculture

Quantification and classification of coffee fruits with computer vision

Helizani Couto Bazame

Thesis presented to obtain the degree of Doctor in Science. Area: Agricultural Systems Engineering

Piracicaba
2021

Helizani Couto Bazame
Agricultural and Environmental Engineer

Quantification and classification of coffee fruits with computer vision

versão revisada de acordo com a resolução CoPGr 6018 de 2011

Advisor:
Prof. Dr. **JOSÉ PAULO MOLIN**

Thesis presented to obtain the degree of Doctor in Science. Area: Agricultural Systems Engineering

Piracicaba
2021

Dados Internacionais de Catalogação na Publicação
DIVISÃO DE BIBLIOTECA – DIBD/ESALQ/USP

Bazame, Helizani Couto

Quantification and classification of coffee fruits with computer vision/ Helizani Couto Bazame. - - versão revisada de acordo com a resolução CoPGr 6018 de 2011. - - Piracicaba, 2021.

77 p.

Tese (Doutorado) - - USP / Escola Superior de Agricultura “Luiz de Queiroz”.

1. Visão computacional 2. Agricultura de precisão 3. Mapas de produtividade 4. Cafeicultura I. Título

ACKNOWLEDGMENT

God, for guiding me and letting me get this far.

My family, who have always been by my side, supported, and encouraged me in all my decisions.

To my husband Daniel, for his support, love, and help in carrying out all the stages of this project.

The School of Agriculture "Luiz de Queiroz" - Unit of the University of São Paulo and the Graduate Program in Agricultural Systems Engineering (PPGESA), for the opportunity to take the course.

To Professor José Paulo Molin, for his guidance, trust, patience, and friendship.

The Coordination for the Improvement of Higher Education Personnel (CAPES), for granting the scholarship.

To my colleagues at the laboratory (LAP), in particular Lucas, Martello, Marcelo, Tatiana, Leonardo, Tiago, and Orlando for their friendship and companionship.

To my friends at ESALQ, Jessica, Larissa, Isabella, and Cleverson for their help and friendship.

To Grupo Guima Café for the experimental area and help with data collection.

To everyone who helped me in any way in the execution of this project, thank you very much.

“Todo o argumento permite sempre a discussão de duas teses contrárias, inclusive este de que a tese favorável e contrária são igualmente defensáveis.”

Protágoras de Abdera.

SUMÁRIO

RESUMO

Quantificação e classificação de frutos de café com visão computacional

O café é uma das bebidas mais consumidas e comercializadas do mundo. O conhecimento sobre a produtividade e o estágio de maturação dos frutos de café antes, e após a colheita, ainda é um desafio para o setor cafeeiro. A criação de um sistema que permita obter essa informação de forma rápida e não invasiva é fundamental para uma gestão eficiente da lavoura. O avanço do monitoramento da cultura do café deve permitir a geração de mapas que apresentem informações essenciais na diagnose da variabilidade espacial e temporal da lavoura e, consequentemente, no eficiente uso das técnicas de agricultura de precisão. Umas das alternativas utilizadas para estimar a produtividade e o estágio de maturação dos frutos de café, seria a utilização de técnicas de visão computacional baseadas na detecção e classificação de objetos. O uso de visão computacional oferece solução de baixo custo e acessível, apresentado grande potencial para a melhoria do monitoramento da lavoura de café. Este estudo foi dividido em três capítulos que apresentam o uso de modelos de visão computacional baseados na arquitetura de redes neurais YOLO para detectar frutos de café em diferentes contextos. No capítulo 1, o modelo é utilizado para detectar e classificar frutos de café na planta, uma ferramenta que pode auxiliar pequenos e grandes produtores na decisão do início da colheita de forma rápida e objetiva. No capítulo 2, o modelo é utilizado para detectar e contar frutos de café durante a colheita mecanizada, o que permite que se gere mapas de produtividade para as áreas colhidas. No capítulo 3, o modelo é utilizado para detectar e classificar os frutos de café em diferentes estágios de maturação durante a colheita mecanizada, o que permite a espacialização do estágio de maturação do café para os talhões colhidos. Os modelos de visão computacional baseados na arquitetura YOLOv4 e uma imagem de entrada com resolução de 800x800 pixels apresentaram precisão média (mAP) de 81,2%, 83,5% e 91,8% para os cenários experimentados nos capítulos 1, 2 e 3, respectivamente. O mapa de produtividade estimado a partir das detecções obtidas pelo modelo foi capaz de explicar 81% da variância do mapa de produtividade utilizado como referência. O conhecimento da variabilidade espacial e temporal de informações como produtividade e estágio de maturação são fundamentais para implantação de técnicas de agricultura de precisão na lavoura de café.

Palavras-chave: Agricultura de precisão, Visão computacional, Cafeicultura, YOLO, Colheita mecanizada

ABSTRACT

Quantification and classification of coffee fruits with computer vision

Coffee is one of the most consumed and traded beverages in the world. Knowledge about the yield and maturation stage of coffee fruits before and after the harvest is still a challenge for the coffee sector. The development of a system that allows obtaining this information quickly and non-invasively is essential for the efficient management of the crop. Advances in monitoring the coffee crop should allow for the generation of maps that present essential information for diagnosing the spatial and temporal variability of the crop and, consequently, for the efficient use of precision agriculture techniques. One of the alternatives used to estimate the yield and ripening stage of coffee fruits would be the use of computer vision techniques based on object detection and classification. The use of computer vision offers a low-cost and accessible solution, with great potential for improving the monitoring of coffee plantations. This study was divided into three chapters that present the use of computer vision models based on the YOLO neural network architecture to detect coffee fruits under different contexts. In chapter 1, the model is used to detect and classify coffee fruits on tree branches, a tool that can help small and large producers to objectively decide when to start the harvest. In chapter 2, the model is used to detect and count coffee fruits during mechanized harvesting, which allows the generation of yield maps for the harvested areas. In chapter 3, the model is used to detect and classify coffee fruits at different stages of maturation during mechanized harvesting, which allows for the spatialization of the coffee maturation stage for the harvested areas. The computer vision models based on the YOLOv4 architecture and an input image with a resolution of 800x800 pixels had mean average precision (mAP) of 81.2%, 83.5% and 91.8% for the scenarios experienced in chapters 1, 2 and 3, respectively. The yield map estimated from the detections obtained by the model was able to explain 81% of the variance of the yield map used as reference. The knowledge of the spatial and temporal variability of information such as productivity and maturation stage are essential for the implementation of precision agriculture techniques in coffee crops.

Keywords: Precision Agriculture, Computer vision, Coffee sector, YOLO, Mechanized harvesting

1. GENERAL INTRODUCTION

Brazil is coffee's largest producer and exporter in the world, and the second-largest consumer of the beverage. Coffee stands out as one of the five main agribusiness sectors in Brazil, resulting in a revenue of US$ 5.9 billion from August 2020 to July 2021 (CECAFÉ, 2021). The Brazilian 2020 coffee harvest was approximately 63.08 million 60-kg bags, cultivated mainly in the states of Minas Gerais, Espírito Santo, and São Paulo (CONAB, 2021).

The coffee crop belongs to the Rubiaceae family, genus Coffea. The two most widely cultivated species of this genus are the Coffea arabica and Coffea canephora. The arabica and canephora coffees represent about 64% and 35%, respectively, of the world's production (HAILE; KANG, 2019). Despite its economic and social relevance for Brazil, there are still several factors that limit coffee production and quality (MARIN et al., 2021).

Among the main factors affecting the coffee beverage quality is the maturity of harvested coffee fruits. The maturation of coffee fruits on plants can occur at different moments along the maturation period. Such irregularity in maturation stages can harm the beverage quality because harvesting unripe and/or overripe coffee fruits can provide the beverage astringent and bitter taste (KAZAMA et al., 2021). Thus, the pre-harvest management of the crop is fundamental for high-quality coffee production. It is also during pre-harvest when the coffee maturation stage and its final yields are estimated. Information about the maturation stage of coffee fruits helps decide the adequate moment when, and the plot where the harvest should begin (MESQUITA et al., 2008).

The coffee fruits' maturation stage is determined by assessing the color of the fruits. This assessment can be performed visually or using colorimeters. Colorimeters measure only a small part of the sample surface, which can result in an unreliable measure (OLIVEIRA et al., 2016). In contrast, the visual assessment, based on the experience of the classifier expert, can be subjective, and is prone to human error. Another disadvantage associated with these methods is that they collect data at a low density and may not be georeferenced, being unable to adequately represent the spatial variability of the coffee fruits' maturation stage inside the plot.

The coffee crop yield may be estimated from in-field samples or using a yield monitor during harvest. Sampling fruits in-field is a destructive analysis process that takes time and is expensive, thus, impractical for large areas (RAHNEMOONFAR; SHEPPARD, 2017). The yield monitor available on the market measures the volume of harvested fruits using a

conveyor belt with cells of known volume (QUEIROZ et al., 2020). The disadvantage is that this method has a high investment cost and its use is exclusive to a single brand of agricultural machinery.

A possible way to increase the number of harvested ripe (cherry) coffee fruits is via selective harvesting using either manual or mechanical harvest. Silva et al. (2015) assessed the mechanical harvest performance to selective harvest cherry coffee fruits by altering the vibration of the rods responsible for shaking the fruits off of tree branches. The authors reported an increase in both harvest efficiency and maturation index by increasing the vibration rate of harvest rods. Santinato et al. (2014) support that adequately adjusting the rods vibration and harvest speed according to coffee variety, plant size, yield, and maturation stage of fruits, can maximize the harvest efficiency and reduce operational costs. Despite the advantages of this technique, obtaining information on coffee yield and fruits' maturation stage is still a challenge faced by the global coffee industry. Thus, selective coffee harvest as it currently is can be labor- and time-demanding as it would require different information in advance or near-real-time.

The development of a system able to collect information on yield and fruits' maturation stage in a fast and non-invasive manner is needed to obtain reliable information about the field, aid decision-making, and possibly feed the harvester with parameters required to improve harvest efficiency. In this context, computer vision is a promising technique to interpret and thrive in different environments, providing accurate and specific information on crops using digital images (LU; YOUNG, 2020). According to Trucco & Verri (1998), computer vision is defined as the ensemble of computational techniques that aim to estimate physical and geometrical properties from images.

Several studies report the possibility of using computer vision techniques for detecting specific objects and providing yield estimates for crops. A study by Brogan & Edison (1974) pioneered the use of computer vision in agriculture. The authors obtained an accuracy of 98% in determining the composition of samples of maize, wheat, soybean, oat, rye, and barley from morphological attributes such as length, width, and depth. Koirala et al. (2019) used six different architectures of deep neural networks to detect mangos and estimate the orchards' yield in real-time. The authors reported an average precision of 0.983 for the test set, where images were assessed in 8 ms. Roy et al. (2019) proposed a computer vision system to map the yield of an apple orchard using digital images. The authors reported precisions ranging from 0.919 to 0.948 for different data sets.

For coffee crops, obtaining such information would help farmers make more informed decisions on crop management, disease control, and labor needs for harvest and post-harvest operations (RAHNEMOONFAR; SHEPPARD, 2017). The advance in coffee crop monitoring can help the development maps that expose the spatial variability of crop characteristics required for implementing efficient precision agriculture techniques. In this context, the main objective of this thesis was to address the monitoring of coffee yield and coffee fruits' maturation stage by benefiting from the recent advances in computer vision techniques for object detection. The thesis consists of three chapters addressing these gaps from different perspectives which are detailed below.

The (i) first chapter of this thesis, entitled "Detection of coffee fruits on tree branches using computer vision", addresses the classification of coffee fruits on tree branches pre-harvest, which can help farmers make more informed decisions from in-field pictures. Previous works developed for classifying coffee fruits on tree branches focused on extracting pre-determined features from pictures, while deep neural networks are proposed here instead. Deep neural networks can extract several features automatically and are robust to environmental conditions.

The (ii) second chapter of this thesis, entitled "Mapping coffee yield with computer vision", also proposes the use of deep neural networks to monitor coffee yield during mechanical harvest. This study shows that it is possible to track yield by counting the number of coffee fruits in georeferenced video frames. The proposed computer vision system is independent of the harvester brand.

The (iii) third and last chapter of this thesis, entitled "Detection, classification, and mapping of coffee fruits during harvest with computer vision", proposes a computer vision system based on deep neural networks to monitor the maturation stage of coffee fruits during harvest and map its spatial variability. Previous studies have only addressed this with limited success by collecting a reduced number of samples in the field and associating them to images obtained with remote sensing. This study is the first to describe the spatial variability of the coffee fruits' maturation stage by classifying hundreds of thousands of video frames obtained during the harvest.

REREFERENCES

BROGAN, W. L.; EDISON, A. R. Automatic classification of grains via pattern recognition techniques. **Pattern Recognition**, v. 6, n. 2, p. 97–103, 1974.

CECAFÉ. Conselho dos Exportadores de Café do Brasil. Relatório mensal de exportações Julho de 2021. Disponível em:< https://www.cecafe.com.br/publicacoes/relatorio-de-exportacoes/. Acesso em 23 de agosto de 2021.

CONAB. Companhia Nacional do Abastecimento. Acompanhamento da safra brasileira. v. 8, 2021. CONAB. Companhia Nacional de Abastecimento."Acompanhamento Da Safra Brasileira: Café". Monitoramento Agricola- Safra 2021. Disponível em: <https://doi.org/ISSN 2318-6852>. Acesso em: 10 de agosto de 2021.

HAILE, M.; KANG, W. H. The Harvest and Post-Harvest Management Practices' Impact on Coffee Quality. **Coffee - Production and Research**, 2019.

KAZAMA, E.H.; DA SILVA, R.P.; TAVARES, T.O.; CORREA, L. N.; ESTEVAM, F. N. L.; NICOLAU, F. E. A.; Maldonado Júnior, W. Methodology for selective coffee harvesting in management zones of yield and maturation. **Precision Agriculture**, 22, 711–733, 2021.

KOIRALA, A. et al. Deep learning for real-time fruit detection and orchard fruit load estimation: benchmarking of 'MangoYOLO'. **Precision Agriculture**, v. 20, n. 6, p. 1107–1135, 2019.

LU, Y.; YOUNG, S. A survey of public datasets for computer vision tasks in precision agriculture. **Computers and Electronics in Agriculture**, v. 178, p. 105760, 2020.

MARIN, D. B.; FERRAZ, G.A. S., SCHWERZ, F.; BARATA, R. A. P.; RAFAEL DE OLIVEIRA FARIA, R. O.; JESSICA ELLEN LIMA DIAS, J. E. L. Unmanned aerial vehicle to evaluate frost damage in coffee plants. **Precision Agriculture 2021**, p. 1–16, 11 maio 2021.

MESQUITA, C. M. DE et al. **Manual do café colheita e preparo.** Belo Horizonte Lastro Editora, , 2008.

OLIVEIRA, E. M.; LEME, D.S.; BARBOSA, B.H.G.; RODARTE, M.P.; PEREIRA, R.G.F.A. A computer vision system for coffee beans classification based on computational intelligence techniques. **Journal of Food Engineering**, v. 171, p. 22–27, 2016.

QUEIROZ, D. M.; COELHO, A. L. F.; VALENTE, D. S. M.; SCHUELLER, J, K. Sensors applied to Digital Agriculture: A review. **Revista Ciencia Agronomica**, v. 51, n. 5, p. 1–15, 2020.

RAHNEMOONFAR, M.; SHEPPARD, C. Deep Count : Fruit Counting Based on Deep Simulated Learning. **Sensors,** vol.4,.p. 1–12, 2017.

ROY, P. et al. Vision-based preharvest yield mapping for apple orchards. **Computers and Electronics in Agriculture**, v. 164, p. 104897, 1 set. 2019.

SANTINATO, F. et al. 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**, 2014.

SILVA, F. C.; SILVA, F. M.; ALVES, M. C.; FERRAZ, G. A. S.; SALES, R. S. Efficiency of coffee mechanical and selective harvesting in different vibration during harvest. **Coffee Science**, v. 10, n. 1, p. 56-64, 2015.

TRUCCO, E.; VERRI, A. Introductory Techniques for 3-D Computer Vision. n. May 2014, 1998.

2. DETECTION OF COFFEE FRUITS ON TREE BRANCHES USING COMPUTER VISION

ABSTRACT

Coffee farmers do not have efficient tools to obtain sufficient and trustworthy information about the maturation stage of coffee fruits before harvest. In this study, we propose a computer vision system to detect and classify the coffee fruits on tree branches in three classes: unripe (green), ripe (cherry), and overripe (dry). The computer vision model was trained on 387 images taken from coffee branches using a smart phone. The YOLOv3 and YOLOv4, along with their smaller versions (tiny), were assessed for the object detection. Both the YOLOv4 and YOLOv4-tiny showed better performance when compared to YOLOv3, especially when smaller network sizes are considered. The mean average precision (mAP) for a network size of 800x800 pixels was equal to 81.24%, 78.93%, 77.74%, and 77.35% for YOLOv4, YOLOv4-tiny, YOLOv3, and YOLOv3-tiny, respectively. Despite the similar performance, the YOLOv4 feature extractor was more robust when images had higher object densities, and for the detection of unripe fruits, which are generally more difficult to detect given the similar color to leaves in the background, partial occlusion by leaves and fruits, and lighting effects. This study shows the potential of computer vision systems based on deep learning to guide the decision-making of farmers in less subjective manners.

KEYWORDS: Precision agriculture, YOLO, high-quality coffee.

2.1. INTRODUCTION

Coffee is a high value-added crop grown in tropical and subtropical regions around the world (CASTRO-TANZI et al., 2014). In this scenario, Brazil is the largest producer and exporter of coffee in the world and the second-largest consumer of the drink. The coffee demand has increased along with the demand for high-quality products. The supply for such products has come especially from improvements on selective harvesting by preferably harvesting ripe fruits (RAMOS; AVENDAÑO; PRIETO, 2018). Enhancing selective harvest has made it feasible the emergence of special products. To meet the increasing demands, it has been necessary to adopt new technologies and good crop management practices that seek to improve the quality of harvested coffee without harming the environment.

Most coffee farmers do not have efficient tools to obtain sufficient and trustworthy information about the maturation stage of coffee fruits before harvest (RAMOS et al., 2017). Tracking the coffee fruits' maturation stage can aid the decision of adequate harvesting periods based on the percentage of mature fruits on tree branches (RAMOS; AVENDAÑO; PRIETO, 2018). Such information is important to adequately manage the crop and support decision-making (RODRÍGUEZ et al., 2020). Besides, the knowledge of the percentage of

fruits in different maturation stages can help coffee farmers stipulate their production losses and selling prices.

The coffee maturation stage is traditionally assessed by the colors in coffee fruits samples. The evaluation can be visual or using colorimeters. Colorimeters measure the color of the fruit surface but without spatial representativeness (OLIVEIRA et al., 2016). Another disadvantage associated with this method is that they require samples to be taken from coffee crops (destructive sampling), resulting in effective loss of yield. The visual classification can also be subjective and relies on the person's experience.

Systems based on computer vision have been largely applied in the detection and classification of fruits in recent decades (AVENDANO; RAMOS; PRIETO, 2017; BAZAME et al., 2021; HÄNI; ROY; ISLER, 2018; KOIRALA et al., 2019; LIU et al., 2020; RAMOS et al., 2017; WANG; WALSH; KOIRALA, 2019; WU et al., 2020a). Computer vision techniques have been reported as non-destructive, fast, and efficient in the detection and classification of objects. Different studies have been performed to classify the quality of coffee. Carrillo and Peñaloza (2009) pioneered research using computer vision techniques to detect and classify coffee grains. The authors obtained an accuracy of 90% for a classifier developed based on pixel values of RGB images and Mahalanobis distance to identify six classes of grains in coffee samples. OLIVEIRA et al. (2016) developed a computer vision system to classify coffee grains in groups of different market values based on their colors.

Only a few studies can be found concerning the classification of coffee fruits before the harvest, which can aid the decision-making of coffee farmers. For example, Avendano, Ramos and Prieto (2017) and Ramos et al. (2017) developed a system that builds a 3D representation of coffee branches before classifying its vegetative structures. However, the technique adopted by the authors requires that various features are extracted and then fed to the classification algorithm. Recent advances in computer vision systems based on deep learning allows that several features are learned and extracted automatically. In addition, these techniques gained popularity for their speed and accuracy.

Given the above, this study aims to implement state-of-the-art computer vision algorithms to detect and classify coffee fruits on tree branches in different maturation stages. The information concerning coffee fruits' maturation stage before harvest can help decide where and when to begin the harvest, which can increase the final product quality and, therefore, the value paid for the amount of harvested product.

2.2. MATERIAL AND METHODS

2.2.1. Data acquisition and labeling

The dataset used in this study consists of 387 RGB images of coffee fruits on tree branches (Figure 1). We used an iPhone 7 to take the pictures before the harvest from a commercial farm of arabica coffee (Catuaí 144) located in Patos de Minas, MG, Brazil. For the development of a robust computer vision model to different field conditions, the pictures were taken from different angles, sides, and for plants randomly selected across coffee lines. This resulted in a diverse scenario and under different lighting conditions. After acquiring the images, they were randomly split into a training set (~80% = 310 images) and a testing set (~20% = 77 images).

Figure 1. Image acquisition

The images were annotated considering three stages (classes) of coffee fruit maturation: unripe (green), ripe (cherry), and overripe (or dry). The annotation was carried using the graphical user interface Yolo Mark (BOCHKOVSKIY et al., 2019).

2.2.2. Computer vision algorithm

The YOLO (You Only Look Once) computer vision algorithm was chosen for the object detection in this study (REDMON et al., 2016). The YOLO is part of a family of one-stage object detectors and is popular for its speed and accuracy (WU et al., 2020a). In this study, we assessed the improvements of the YOLO latest version, YOLOv4 (BOCHKOVSKIY; WANG; LIAO, 2020), in comparison to its former version, YOLOv3

(REDMON; FARHADI, 2018). The improvements of the YOLOv4 over its former version include using the Mish activation function (MISRA, 2019), CutMix (YUN et al., 2019) and mosaic data augmentation, Cross-Stage Partial connections (CSP) (WANG et al., 2019), Cross mini-Batch Normalization (CmBN), Spatial Pyramid Pooling (SPP) (HE et al., 2015) and the Path Aggregation Network (PANet) (LIU et al., 2018) blocks, Complete Intersection over Union (CIoU) loss (ZHENG et al., 2019), among others.

Besides the YOLOv3 and YOLOv4, a smaller version of these models, termed "tiny", was also assessed. The YOLO-tiny models were developed with fewer convolutional layers and are suitable for constrained environments, as for smartphones (TANG, 2018) or microcomputers such as the Raspberry PI, NVIDIA Jetson Nano, Intel Movidius Neural Compute Stick, and Google Coral.

The object detection models were trained considering different network sizes and resampling image sizes to match the corresponding network. The network sizes adopted were 320 x 320, 416 x 416, 512 x 512, 608 x 608, 704 x 704, and 800 x 800 pixels. For training, the batch size was set to 32 in the forward pass, and the number of iterations was equal to 6000. The confidence thresholds (c) and non-maximum suppression adopted were 0.25 and 0.45, respectively. The performance criterion was tracked for each training iteration using the test set. The weights resulting in the best performance were adopted as the final weights for the model.

2.2.3. Performance evaluation

The performance of the computer vision algorithms was measured by the mean values of average precisions (mAP) obtained for all classes detected considering an intersection over union of 50%. The average precision (Eq. 1) is the average value of 11 points on the precision/recall curve for pre-determined confidence thresholds for the same class. The precision (Eq. 2) and recall (Eq. 3) are computed for 11 equally spaced confidence thresholds (c = 0.0, 0.1, …, 1.0) and precision at each recall level (Eq. 4) is interpolated by setting the maximum precision measured for a threshold which the corresponding recall r' exceeds r (Eq. 4):

$$AP = \int_0^1 p(r)dr \quad (1)$$

$$p(c) = \frac{TP}{TP+FP} \quad (2)$$

$$r(c) = \frac{TP}{TP+FN} \quad (3)$$

$$p(r) = \max_{r':r' \geq r} p(r') \quad (4)$$

AP is the average precision, TP are true positives, FP are false positives, FN are false negatives, p(r) is the precision at recall level r, p(r') is the precision at recall level r', and c is the confidence threshold.

2.3. RESULTS AND DISCUSSION

The results are presented and discussed in three subsections. The first subsection (2.3.1.) discusses the general performance obtained by the object detection algorithms and highlights the main findings of this study. The second and third subsections (2.3.2. and 2.3.3.) details more specific outcomes from the algorithms concerning performance scores for the different classes and different object densities, respectively.

2.3.1. General performance obtained by the YOLO algorithms

The performance of coffee fruit detection obtained for each YOLO algorithm and network size, as measured by their mean average precision (mAP), is displayed in Figure 2. For the YOLOv4, YOLOv4-tiny, and YOLOv3, the mAP stabilized near the network size of 608 x 608 pixels. In contrast, the performance of the YOLOv3-tiny continued to increase until the network size of 704 x 704 pixels. Despite stabilizing, the performance of the algorithms still showed slight improvements up to 800 x 800 pixels. In general, both the YOLOv4 and YOLOv4-tiny outperformed the YOLOv3. The YOLOv4-800 scored the highest mAP (81.24%), followed by YOLOv4-tiny-800 (78.93%), YOLOv3-800 (77.74%), and YOLOv3-tiny-800 (77.35%).

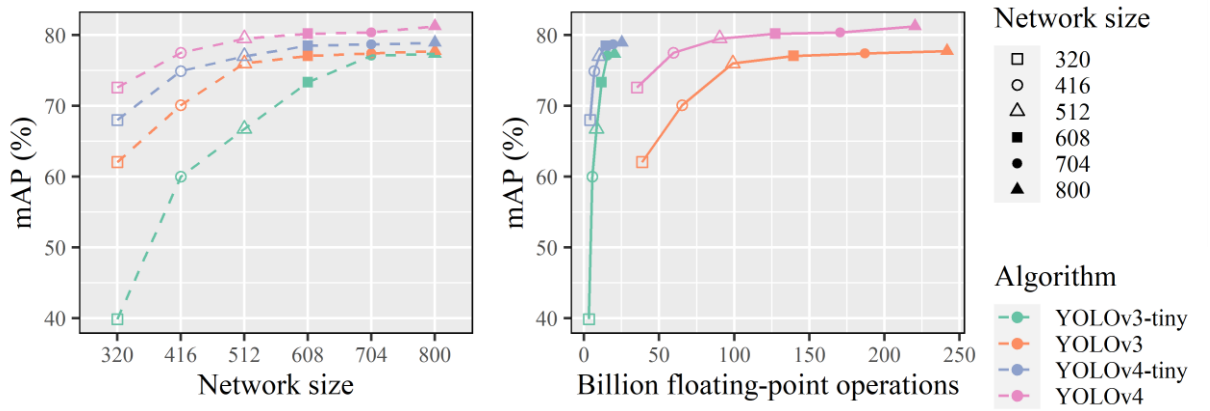

Figure 2. Performance of the different computer vision algorithms and network sizes assessed for detecting coffee fruits on branches.

The smaller the network size, largely do the YOLOv4 and YOLOv4-tiny outperforms the YOLOv3 and YOLOv3-tiny. In contrast, when larger network sizes are considered, 704 x 704 and 800 x 800 pixels, the difference in performances of the YOLOv3 and YOLOv3-tiny are negligible. Perhaps the most important outcome here is the YOLOv4-tiny outperforming the YOLOv3. This means that the updates made for the latest YOLO version were fundamental in improving its performance, even when considering a restricted number of convolutional layers. The YOLOv4-tiny requires ~89.6% less billion floating-point operations than the YOLOv3, which means its model/weights not only occupy less space in a hard drive but can be run much faster.

The detections made by the YOLO algorithms for three arbitrary images from the dataset and considering the network size of 800 x 800 pixels are shown in Figure 3. The mAP obtained for each image is also displayed in the figure, where the YOLOv4 shows to consistently outperform the other algorithms. The YOLOv4 also does a better job at detecting fruits that are overlapped (Figure 3c), or in the shade (Figure 3a). It also does better at detecting unripe (green) fruits, even when they are visually smaller in the background and in between the leaves (Figure 3b). Another adaptation that could further improve the model detection is the one suggested by LIU et al. (2020). The authors adapted the YOLO algorithm to use a circular bounding box instead of the traditional rectangular one. Because of the tomato shape, the circular bounding box allowed for a better object detection under difficult conditions of lighting, branch and leaves occlusion, and overlapping of tomatoes. In Figure 3, the YOLOv4 showed to generally better detect occluded/overlapped objects, even under difficult settings.

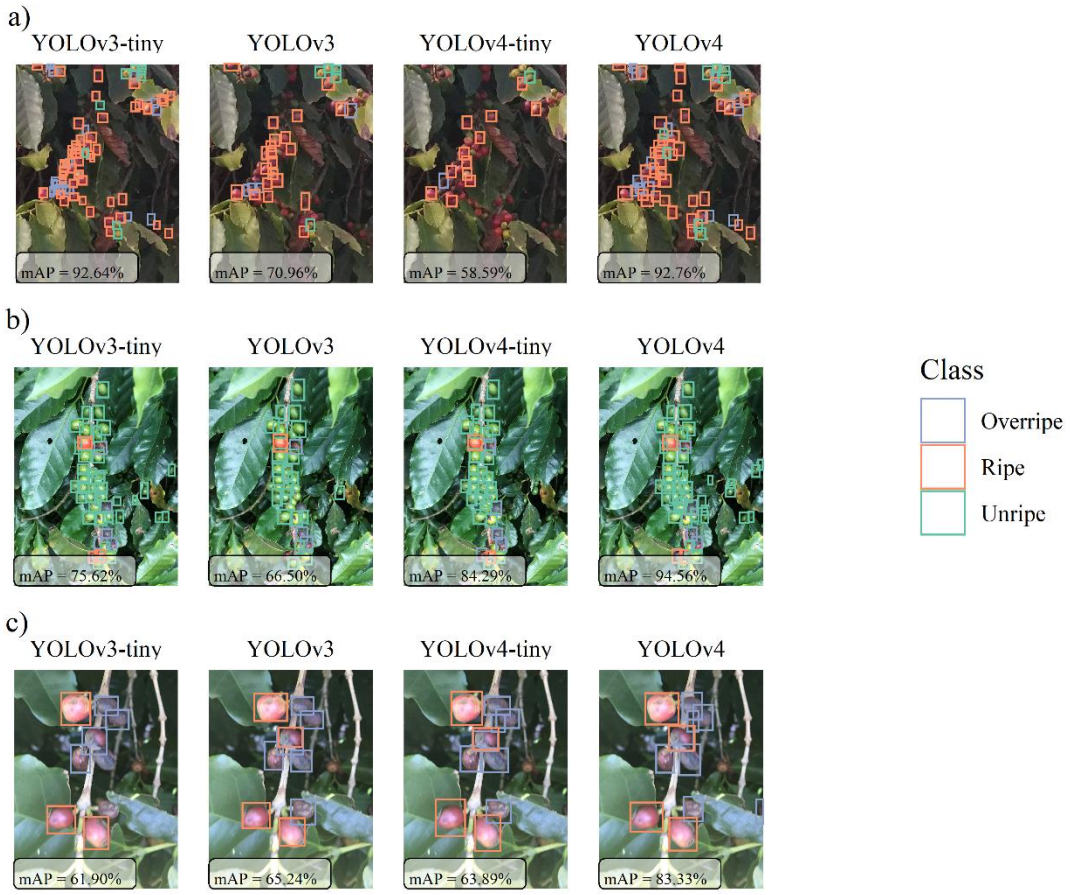

Figure 3. Coffee fruits detections made by YOLO algorithms for three arbitrary images considering a network size of 800x800 pixels.

The high performance of YOLOv3-tiny in Figure 3a deserves due attention. There seems to be a surplus of detections (bounding boxes) in the figure which, despite resulting in high recall (0.92), results in lower precision (0.70) because of the large number of false positives (see Eqs. 2 and 3). This is an outcome of the badly predicted boxes for this specific image not being adequately removed by the confidence threshold and non-maximum suppression post-processing. In contrast, the YOLOv3 model predicted coffee fruits in this figure with lower confidence, resulting in fewer boxes and higher precision (0.83) but much lower recall (0.42) and mAP. Despite a similar mAP to that obtained by the YOLOv3-tiny and YOLOv4 models for the example image in Figure 2a, the YOLOv4 resulted in far better predictions, with both high precision (0.88) and recall (0.88).

To better assess the trade-offs between precision and recall, Figure 4 shows the distribution of performance scores (mAP, precision, and recall) obtained for each image of the test set. Despite the overall higher median and mean mAP obtained from all test set images for YOLOv4, there are clear trends in the precision and recall trade-offs that can be assessed. The mAP is obtained by considering a set of different confidence thresholds, whereas the final

precision and recall are calculated assuming a pre-set confidence threshold (c = 0.25). As previously discussed, obtaining high precision at the cost of too many false positives can lead to a lower recall. For example, the YOLOv3 algorithm shows, for most network sizes, to score relatively higher precision but lower recall. In contrast, the YOLOv4 algorithm shows the opposite behavior, scoring relatively higher recall and lower precision.

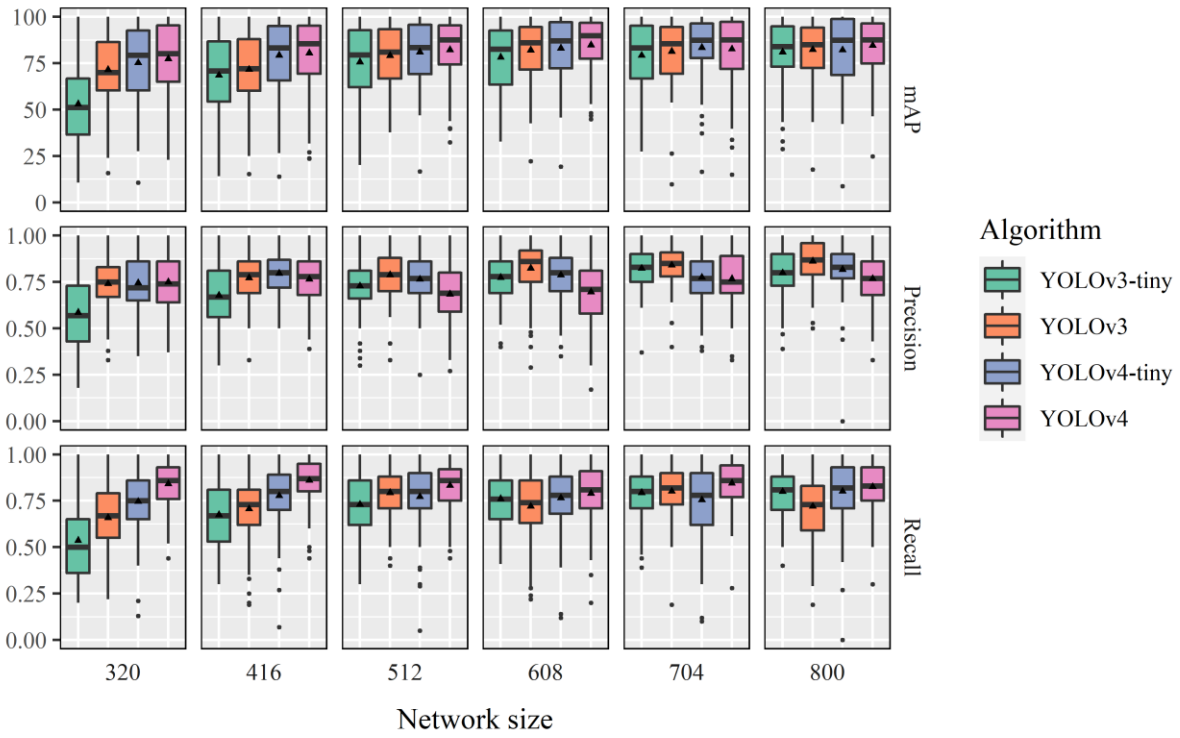

Figure 4. Distribution of performance scores obtained for each image of the test set by the different computer vision algorithms and network sizes used in this study.

Despite observing general trends for the precision-recall trade-offs for the different algorithms, the results may partially be attributed to the random process of weight adjustment during training. In this study, the models' final weights were set as the weights obtained after the training iteration that resulted in the highest mAP for the test set from all 6000 iterations. However, predictions resulting from weights scoring similar mAP can present different precision-recall trade-offs. Thus, it is up to the final user of the model to decide whether it is more important to identify all true positives regardless of obtaining a few false positives, or if predicting false positives can be detrimental/costly to the final objective. In general, similar values of precision and recall indicate a well-balanced model and a robust precision-recall trade-off.

2.3.2. Performance by detection class

The average precision (AP) obtained for each class highlights a close performance between YOLOv4 and the other models for the detection of ripe and overripe coffee fruits, especially for larger network sizes (Figure 5). For example, for ripe fruits and a network size of 800 x 800 pixels, the YOLOv4-tiny, YOLO-v3, and YOLOv3-tiny scored APs of 82.62%, 83.57%, and 80.12%, respectively, while YOLOv4 scored an AP (83.97%) higher by 1.35%, 0.40%, and 3.85%, respectively. For overripe fruits, the YOLOv4-tiny, YOLO-v3, and YOLOv3-tiny scored APs of 78.21%, 77.15%, and 75.64%, respectively, while YOLOv4 scored an AP (79.94%) higher by 1.73%, 2.79%, and 4.30%, respectively.

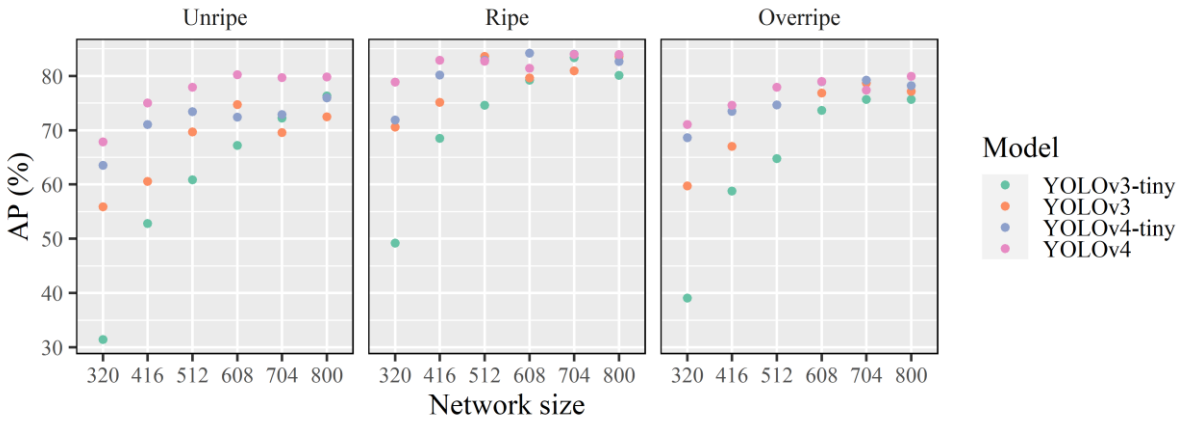

Figure 5. Performance of the different computer vision algorithms and network sizes assessed for each class of detection.

The YOLOv4 stands out, however, in the detection of unripe (green) coffee fruits, which are generally more difficult to detect because of leaves on the branches and in the background. The YOLOv4 scored an AP of 79.82% for unripe fruits and a network size of 800 x 800 pixels, which is higher by 3.86%, 7.33%, and 3.52% than those scored by the YOLOv4-tiny, YOLOv3, and YOLOv3-tiny, respectively. The difference is even higher when smaller network sizes are considered.

Computer vision technologies, whether pre-harvest or during harvest, can have the additional advantage of simultaneously evaluating the characteristics or quality of the agricultural product. For example, shape, size, or color can be determined for classification purposes. Another example is that defects and pest infestations can also be identified (SCHUELLER, 2021). RAMOS et al. (2018) developed a computer vision system that also predicts the maturation stage of coffee fruits on tree branches. The computer vision system classifies the coffee fruits after building a 3D model of on-branch coffee fruits and resulted in

classification effectiveness between 42% and 92% for the different classes of maturation stage. BAZAME et al. (2021) proposed a computer vision model to detect coffee fruits and classify their maturation stage during harvest. The authors were then able to map the maturation stage across the coffee plantation with an mAP of 86%, 85%, and 80% for unripe, ripe, and overripe fruits, respectively. The lower mAP for unripe fruits in this study, when compared to that of BAZAME et al. (2021), can be attributed to the environment where images were taken. Here, pictures were taken from on-branches coffee fruits with a diverse background, including leaves and shades, whereas BAZAME et al. (2021) collected data inside the harvester where the environment had controlled illumination and contrasting background. Besides, the authors also registered a lower mAP score for overripe fruits.

A further opportunity for the present study could be related to also predicting coffee yield from full lateral pictures of coffee plants, as has been proposed by Idol and Youkhana (2020). To obtain such information for field scales, however, it would require that images are collected along with geographic coordinates and at higher rates. Besides, the data collection at higher rates by autonomous systems has been proposed in different studies. For example, Saiz-Rubio (2021) proposed an autonomous robot to monitor vineyard water potential. Autonomous robots have even been proposed to perform actions, such as tomato harvesting (LIU et al., 2020), strawberry harvesting (XIONG et al., 2020), and weed control (WU et al., 2020b).

2.3.3 Performance for different object densities

It is harder for smaller network to detect coffee fruits in scenarios where object density is higher. This is because resizing images to lower resolution may blur the fruits' boundaries. Such behavior is evident in Figure 6, which shows lower median mAP (red dashed lines) obtained for smaller networks and steeper slopes for the ordinary least squares' regression fitted to data (blue line). For example, the YOLOv3 and YOLOv3-tiny models resulted in mAP lower than 70.04% and 57.24%, respectively, in 50% of the images in the test set for a network size of 320 x 320 pixels. The YOLOv4-tiny and YOLOv4 were more robust in extracting features and avoiding such effects for the smaller network sizes. For YOLOv4-tiny and YOLOv-4, 50% of the test set images scored mAP equal or higher than 79.26% and 80.14% for a network size of 320 x 320 pixels. WANG et al. (2021a) have proposed an adaption to the YOLOv3 model for the detection of litchi (YOLOv3-Litchi) in images with a high density of fruits. The model was adapted by the authors to have fewer convolutions than

the original YOLOv3 and predict from features maps at higher resolutions, which increased the accuracy when detecting objects in images with high densities of small fruits.

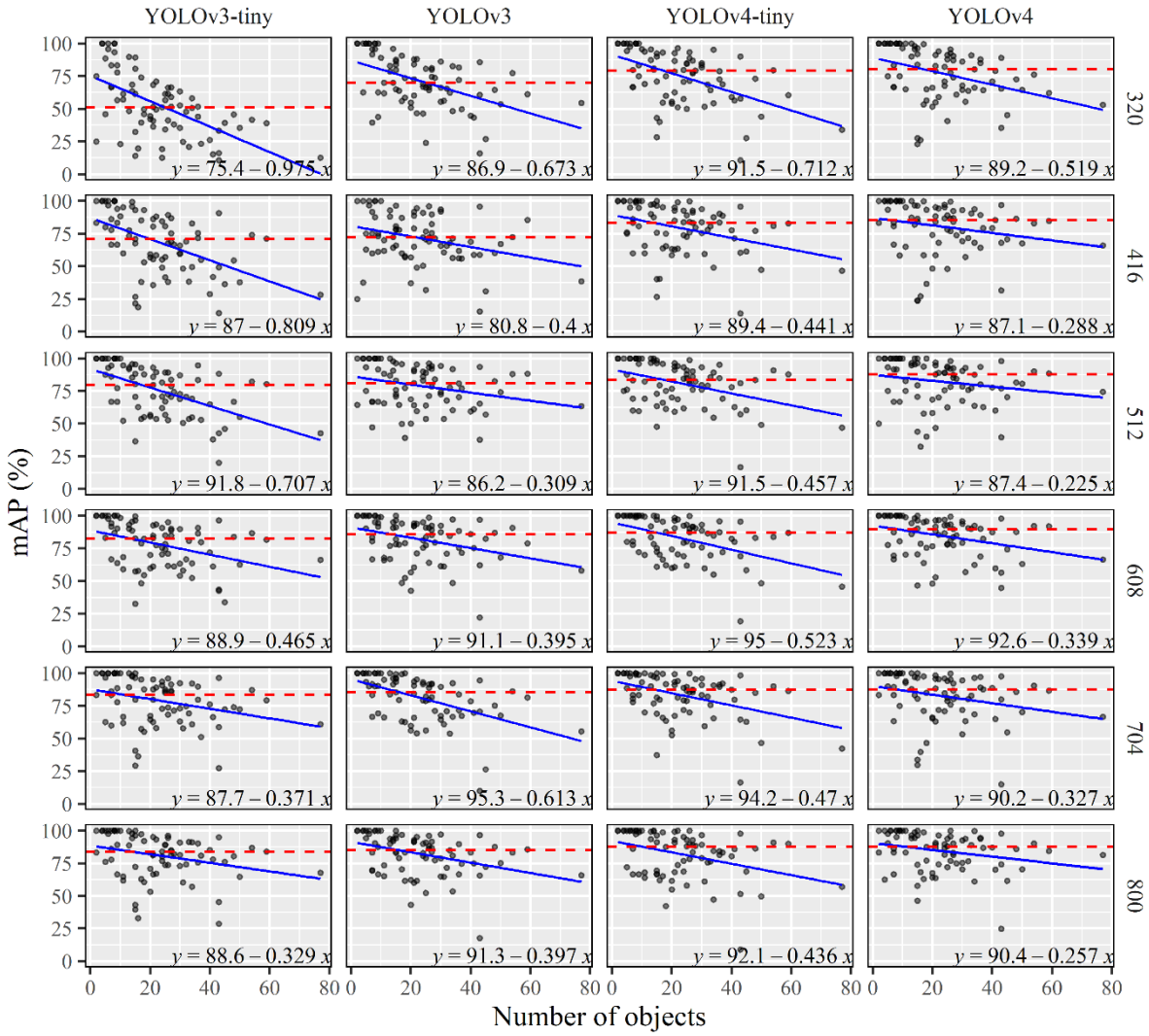

Figure 6. Performance obtained by the different computer vision algorithms and network sizes assessed for each image of the test set separately. The red dashed line represents the median mAP. The blue line represents the ordinary least squares regression fit to the data. Steeper slopes mean it is more difficult for the model to detect objects when object density is higher in the dataset.

As the network size and, therefore, resolution of resized images increased, the problem is mitigated. For example, the regression slopes for the YOLOv3-tiny models decreased from -0.975 to -0.329 for network sizes from 320 x 320 to 800 x 800 pixels. Overall, the regressions adjusted more gentle slopes (closer to 0) for scores obtained using larger network sizes. This is especially true for the YOLOv4 algorithm, which slope was only -0.257 for the network size of 800 x 800 pixels. Input images at higher resolutions mean larger networks and

usually better performance in object detection, but it may also increase the time required for predicting (WANG et al., 2021b) or constrain the model to hardware with higher computing power. The YOLOv4-tiny also performed better than YOLOv3-tiny in this regard, even at lower network sizes, which can be attributed to its more robust feature extractor.

2.4. CONCLUSIONS

In this study, the YOLOv3 and YOLOv4 object detection algorithms were implemented to detect and classify the maturation stage of coffee fruits on tree branches. The upgrades on the YOLO algorithm from v3 to v4 resulted in significant improvement in performance, even for the smaller version (YOLOv4-tiny) which is built on fewer convolutional layers. For an image input resolution of 320 x 320 pixels, the YOLOv4, YOLOv4-tiny, YOLOv3, and YOLOv3-tiny scored a mean average precision (mAP) of 72.59%, 68.00%, 62.06%, and 39.88%, respectively. For a larger networks considering images of 800 x 800 pixels, these models scored mAPs of 77.35%, 77.74%, 78.93%, and 81.24%, respectively.

The developed models generally performed better in the detection of ripe coffee fruits, which better contrast to the images' background. In contrast, the performance in detecting unripe (green) fruits was considerably lower, which can be attributed to the coffee fruits being partially occluded by leaves (similar color) and in the shade. In this regard, the YOLOv4 algorithm outperformed its former version. In addition, the YOLOv4 algorithm also resulted in detections less influenced by object density in images, all of which can be attributed to a more robust feature extractor.

Future studies could advance this research in many directions. The image acquisition could be associated with geographic coordinates or even captured by an automated system, which would allow for the spatialization of such information. The continuous collection of images from all sides of coffee plants could also be used to estimate the fruit count and, therefore, yield by the plant.

REFERENCES

AVENDANO, J.; RAMOS, P. J.; PRIETO, F. A. A system for classifying vegetative structures on coffee branches based on videos recorded in the field by a mobile device. **Expert Systems with Applications**, v. 88, p. 178–192, 2017.

BAZAME, H. C.; MOLIN, J. P.; ALTHOFF, D.; MARTELLO, M. Detection, classification, and mapping of coffee fruits during harvest with computer vision. **Computers and Electronics in Agriculture,**vol. 183, 2021.

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.

BOCHKOVSKIY, A.; WANG, C.-Y.; LIAO, H.-Y. M. YOLOv4: Optimal Speed and Accuracy of Object Detection. **arXiv**, 22 abr. 2020.

CARRILLO, E.; PEÑALOZA, A. A. Artificial vision to assure coffee-excelso beans quality. Proceedings of the 2009 Euro American Conference on Telematics and Information Systems: New Opportunities to Increase Digital Citizenship, **EATIS**, 2009.

CASTRO-TANZI, S. et al. Evaluation of a non-destructive sampling method and a statistical model for predicting fruit load on individual coffee (Coffea arabica) trees. **Scientia Horticulturae**, v. 167, p. 117–126, 2014.

HÄNI, N.; ROY, P.; ISLER, V. Apple Counting using Convolutional Neural Networks. **IEEE International Conference on Intelligent Robots and Systems**. **Anais**...Institute of Electrical and Electronics Engineers Inc., 27 dez. 2018

HE, K. et al. Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 37, n. 9, p. 1904–1916, 2015.

IDOL, T. W.; YOUKHANA, A. H. A rapid visual estimation of fruits per lateral to predict coffee yield in Hawaii. **Agroforestry Systems**, v. 94, n. 1, p. 81–93, 1 fev. 2020.

KOIRALA, A. et al. Deep learning for real-time fruit detection and orchard fruit load estimation: benchmarking of ‘MangoYOLO’. **Precision Agriculture**, v. 20, n. 6, p. 1107–1135, 2019.

LIU, G.; NOUAZE, J.C.; TOUKO MBOUEMBE, P.L.; KIM, J.H. YOLO-Tomato: A Robust Algorithm for Tomato Detection Based on YOLOv3. **Sensors**, v. 20, n. 7, 1 abr. 2020.

LIU, S. et al. Path Aggregation Network for Instance Segmentation. **Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition**, p. 8759–8768, 5 mar. 2018.

MISRA, D. Mish: A Self Regularized Non-Monotonic Activation Function. **arXiv**, 23 ago. 2019.

DE OLIVEIRA, E. M. et al. A computer vision system for coffee beans classification based on computational intelligence techniques. **Journal of Food Engineering**, v. 171, p. 22–27, 2016.

RAMOS, P. J. et al. Automatic fruit count on coffee branches using computer vision. **Computers and Electronics in Agriculture**, v. 137, p. 9–22, 1 maio 2017.

RAMOS, P. J.; AVENDAÑO, J.; PRIETO, F. A. Measurement of the ripening rate on coffee branches by using 3D images in outdoor environments. **Computers in Industry**, v. 99, p. 83–95, 1 ago. 2018.

REDMON, J. et al. **You only look once: Unified, real-time object detection**. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. **Anais**...IEEE Computer Society, 9 dez. 2016.

REDMON, J.; FARHADI, A. YOLO v.3. **Tech report**, p. 1–6, 2018.

RODRÍGUEZ, J. P. et al. A computer vision system for automatic cherry beans detection on coffee trees. **Pattern Recognition Letters**, v. 136, p. 142–153, 1 ago. 2020.

SAIZ-RUBIO, V. et al. Robotics-based vineyard water potential monitoring at high resolution. **Computers and Electronics in Agriculture**, v. 187, p. 106311, 1 ago. 2021.

SCHUELLER, J. K. Opinion: Opportunities and Limitations of Machine Vision for Yield Mapping. **Frontiers in Robotics and AI**, vol. 0, p. 18, 25 Feb. 2021.

TANG, J. GitHub - jeffxtang/yolov2_tf_ios: Object Detection with YOLOv2 and TensorFlow on iOS. Disponível em: <https://github.com/jeffxtang/yolov2_tf_ios#readme>. Acesso em: 17 jul. 2021.

WANG, C.-Y. et al. CSPNet: A New Backbone that can Enhance Learning Capability of CNN. **IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops**, v. 2020- June, p. 1571–1580, 26 nov. 2019.

WANG, H. et al. YOLOv3-Litchi Detection Method of Densely Distributed Litchi in Large Vision Scenes. **Mathematical Problems in Engineering**, v. 2021.

WANG, Z.; WALSH, K.; KOIRALA, A. Mango fruit load estimation using a video based MangoYOLO—Kalman filter—hungarian algorithm method. **Sensors (Switzerland)**, v. 19, n. 12, 2 jun. 2019.

WU, D.; LV, S.; JIANG, M.;SONG, H. Using channel pruning-based YOLO v4 deep learning algorithm for the real-time and accurate detection of apple flowers in natural environments. **Computers and Electronics in Agriculture**, v. 178, p. 105742, 1 nov. 2020a.

WU, X.; ARAVECCHIA, S.; LOTTES, P.; STACHNISS, C.; PRADALIER, C. Robotic weed control using automated weed and crop classification. **Journal of Field Robotics**, v. 37, n. 2, p. 322–340, 1 mar. 2020b.

XIONG, Y.; GE, Y.; GRIMSTAD, L.; FROM. P. J. An autonomous strawberry-harvesting robot: Design, development, integration, and field evaluation. **Journal of Field Robotics**, v. 37, n. 2, p. 202–224, 1 mar. 2020.

YUN, S. et al. CutMix: Regularization Strategy to Train Strong Classifiers with Localizable Features. **Proceedings of the IEEE International Conference on Computer Vision**, v. 2019- October, p. 6022–6031, 13 maio 2019.

ZHENG, Z. et al. Distance-IoU Loss: Faster and Better Learning for Bounding Box Regression. **arXiv**, 19 nov. 2019.

3. MAPPING COFFEE YIELD WITH COMPUTER VISION

ABSTRACT

Yield maps guide investigations into the causes of spatial and temporal variations in crop yields. The objective of this work was to implement an algorithm based on computer vision to quantify coffee fruits and to build yield maps. Data were collected in two areas of a commercial arabica coffee plantation. The images of the coffee fruits were taken from 90 videos acquired during the harvest. The YOLOv4 model was used for the detection and counting of coffee fruits. Geographic coordinates were sampled at the same time the videos were recorded and associated with video frames. The number of coffee fruit detections for each frame was converted into yield considering the average distance covered by the harvester and the distance between coffee rows. The yield maps were interpolated from the video frames' respective geographic coordinates. The YOLOv4 model had a mean average precision of 83.5%. The yield map estimated from the detections obtained by the computer vision model was able to explain 81% of the variance of the reference yield map. The main contributions of the proposed methodology are its low implementation cost and the independence of specific brands of coffee harvesters for the implementation of the image capture structure.

KEYWORDS: deep learning, precision agriculture, mechanical harvesting, YOLOv4.

3.1. INTRODUCTION

Yield monitoring and mapping of crops is a useful tool adopted for the management of its spatial and temporal variability. Yield maps are the result of every intervention realized during the crop cycle and can guide the investigation of the main causes for yield variability in the field (MALDANER et al., 2021; QUEIROZ et al., 2020). Yield monitoring is not common in coffee crops because few technologies are available and at a high cost. Yet, the development of new sensors for monitoring coffee yield has also received little attention.

The estimate of coffee yield is usually performed through the sampling of fruits in specific parts of the field before harvest begins. The disadvantage of this method is the limited number of plants and sites in the field that can be sampled, which are associated with high cost and time, and loss of sampled fruits (destructive sampling) (RODRÍGUEZ et al., 2020). Besides, this method does not provide georeferenced data at high densities and is not fit to adequately represent the spatial variability in the field (OLIVEIRA et al., 2016; IDOL and YOUKHANA, 2020).

Yield data can also be obtained using yield monitors. Sartori et al. (2002) developed the only yield monitor for coffee existent in the market. The yield monitor measures the

volume of harvest fruits using volume cells. An ultrasound sensor is used to identify when the cells are filled and the yield is estimated by accounting for the number of cells, the travel speed of the harvester, and the distance between coffee lines (MOLIN et al., 2010). The disadvantages of this method are its high investment cost and the fact that it was developed and patented for the exclusive use of a single model of coffee harvester, already discontinued in the market.

An alternative for estimating the harvested coffee fruits is the use of computer vision techniques based on object detection. These techniques allow the development of specific models that address the main characteristics of coffee crops (HAMDAN et al., 2020). This allows for the collection of data in high densities, resulting in thousands of observations by hectare. Because data can be continuously collected, they can be processed to detect spatial patterns and develop yield maps (MARIANO and MÓNICA, 2021).

Recent advances in computer vision techniques applied to the processing of images represent a large potential for their application in estimating the yield and quality of agricultural products (BAZAME et al., 2021; MOLIN et al., 2020). For example, Idol and Youkhana (2020) developed a technique to estimate coffee plant yield. The technique is based on visual estimates of the number of coffees on side views of the plant. The authors concluded that the technique can reduce the number of estimates or counts up to 75% to 85%, although it requires initial training to visually estimate the yield from side views. Rodríguez et al. (2020) also developed a mechanism based on computer vision to detect coffee fruits on plants. The authors reached their best results for arabica coffee with a precision of 0.59.

Despite the attempts described to estimate the yield of coffee fruits using computer vision, none of the studies assessed these techniques by collecting georeference data and considering all the coffee fruits over an area. Such information could result in a reliable base for coffee yield monitoring and mapping. Trustworthy yield maps allow for the adoption of precision agriculture techniques to optimize yield and machine efficiency (SANTOS et al., 2021). This was addressed in this study using a sensor embedded in a harvester. Thus, the main objective of this study was to implement an algorithm based on computer vision to detect and quantify coffee fruits during harvest and post-process the detections to produce yield maps.

3.2. MATERIAL AND METHODS

3.2.1. Study area

The experimental data were collected in two plots from a commercial farm of Arabica coffee (Catuaí 144) located in Patos de Minas, MG, Brazil (Figure 1). The coffee crops were planted in 2006 and 2013 for experimental areas 1 and 2, respectively. The coffee trees were at a density of 5.000 trees hectare^{-1} with a spacing of 4.0 m between lines and 0.5 m between plants.

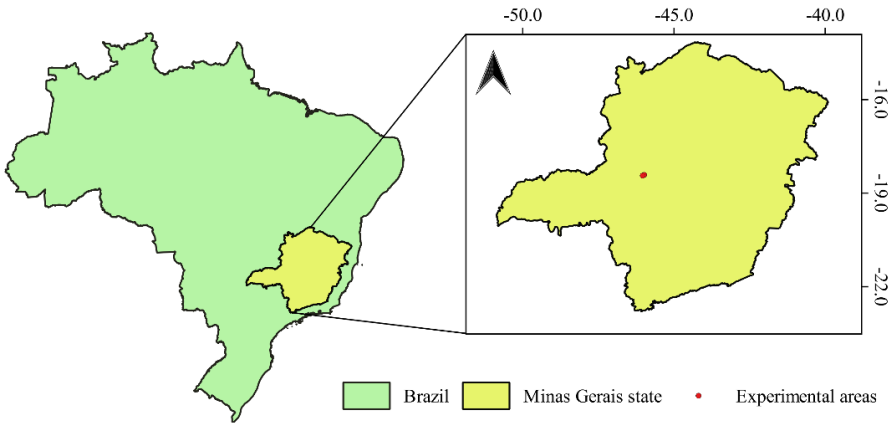

Figure 1. Experimental areas of coffee crops in Brazil.

3.2.2. Image acquisition and annotation

The images of coffee fruits were extracted from 90 videos recorded during the entire harvest and with a duration of approximately 7 to 15 minutes each. The videos were collected using an image acquisition platform mounted over the side loading spout of a coffee harvester (Figure 2). The interior of the platform was illuminated with a set of six 21W LED lamps. The camera was stabilized with a 3-axes gimbal and consisted of a 1” complementary metal oxide semiconductor (CMOS) sensor with a mechanical shutter and resolution of 20MP. The videos were recorded with full HD definition (1920x1080) and 100Mbps bit rate (60 fps, 720P, ISO 1600, Shutter 1/800).

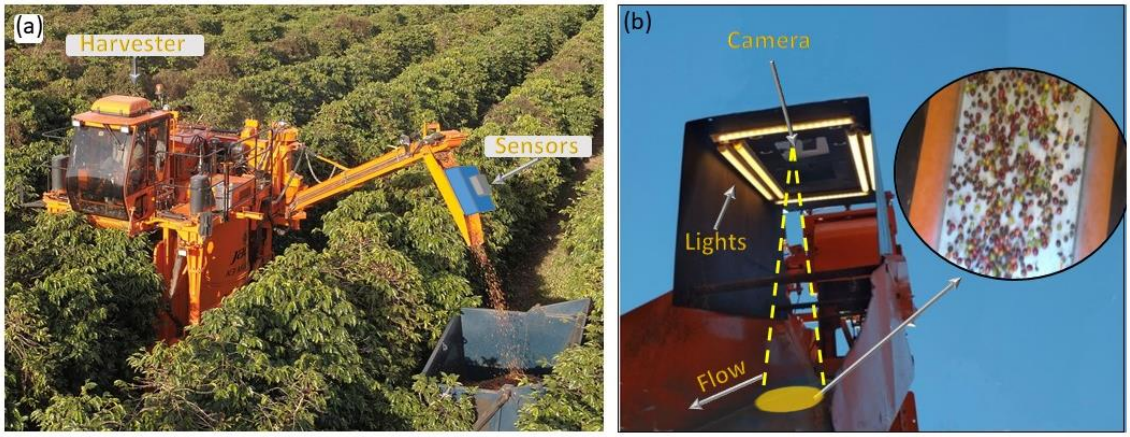

Figure 2. Image acquisition platform mounted on the side loading spout a coffee harvester.

The coffee mechanical harvest was carried out from 31 May to 06 June 2020 using a coffee harvester K3 Millennium (Máquinas Agrícolas Jacto S.A, Brazil). The harvester travel speed ranged from 0.7 to 2.5 km h^{-1} during harvest. Besides the image acquisition platform, the harvester was also embedded with the yield monitor described by Sartori et al. (2002).

For the object detection model to learn to identify coffee fruits, it is required that the model is trained on images where ground-truth annotations of the coffee fruits are available. Therefore, a data set containing 380 images acquired by the image acquisition platform was annotated by the authors using the graphical user interface Yolo Mark (Bochkovskiy, 2019). The dataset was later split into training and validation sets, each containing 260 and 120 images, respectively.

3.2.3. Computer vision algorithm: YOLOv4

The computer vision algorithm chosen for this study is the latest version of the You Only Look Once (YOLO) algorithm, the YOLOv4 (BOCHKOVSKIY et al., 2020). The YOLO algorithms are a family of one-stage object detectors that are widely known for their real-time detector speed and accuracy (WU et al., 2020). Object detectors usually consist of a "backbone" and a "head". The input image passes through the backbone, a series of convolutional layers used to extract feature maps. These feature maps are then fed to the head, where dense (one-stage, anchor-based) or sparse (two-stage, anchor-free) predictions are made (BOCHKOVSKIY et al., 2020). Like other object detectors developed in recent years, the YOLOv4 also includes a "neck" between the backbone and the head. The neck generally consists of several layers used to collect feature maps across different stages before feeding them to the predictions.

The YOLOv4 model (BOCHKOVSKIY et al., 2020) has several improvements when compared to its older versions YOLOv1 (Redmon et al., 2016), YOLOv2 (Redmon & Farhadi, 2017), and YOLOv3 (REDMON & FARHADI, 2018). The main differences between the YOLOv4 and its former version, YOLOv3, is that YOLOv4 includes Cross-stage Partial Connections (CSP) (WANG et al., 2019) into account for the Darknet53 backbone and uses the Spatial Pyramid Pooling (SPP) (HE et al., 2015) and the Path Aggregation Network (PANet) (LIU et al., 2018) in the neck instead of the Feature Pyramid Network (FPN) (LIN et al., 2016). Finally, YOLOv4 uses the YOLOv3 head for predictions. The YOLOv4 architecture and input/output dimensions are shown in Figure. 3 considering an input image of size 800 x 800 pixels and a single detection class, coffee fruits.

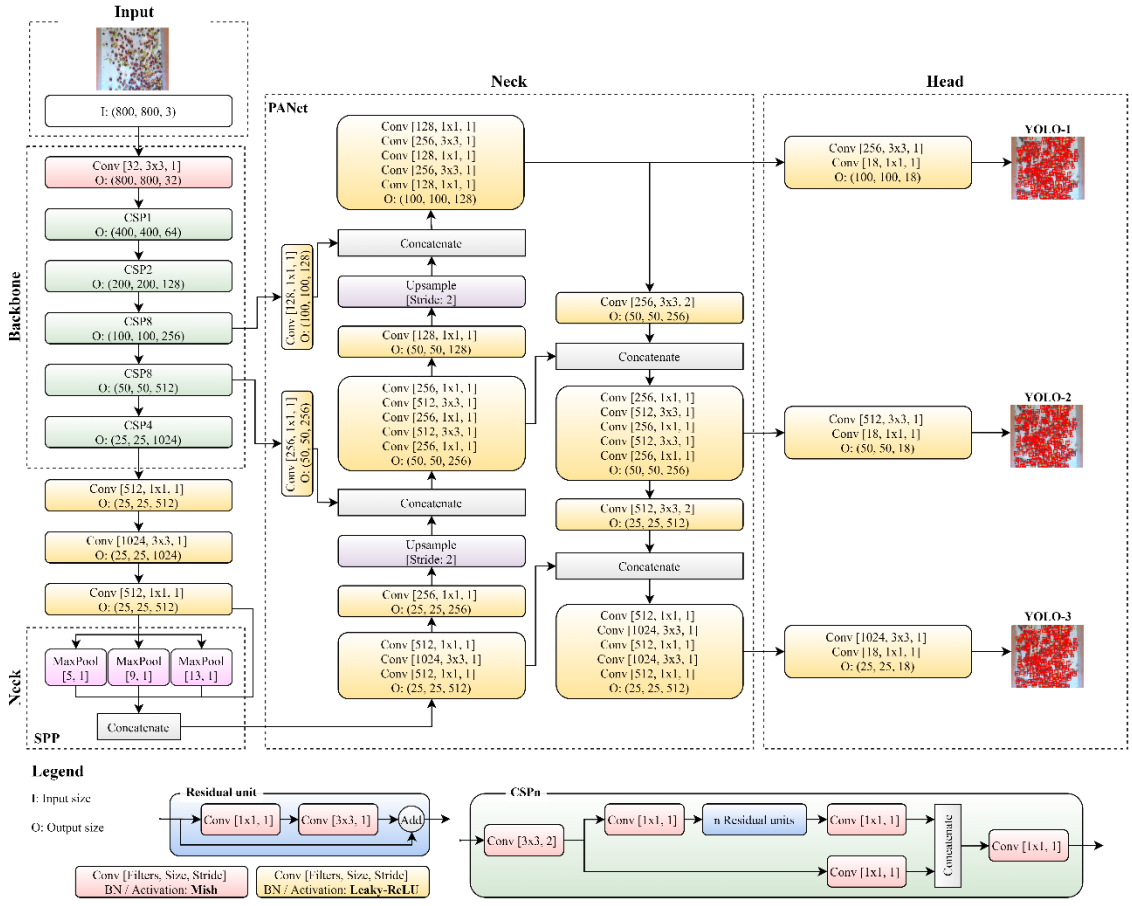

Figure 3. YOLOv4 architecture adapted to this study considering an input image of 800 x 800 pixels.

The cross-stage partial connections in CSPDarknet53 increase the learning ability of the object detector, removes computational bottlenecks, and reduces memory costs (WANG et al., 2019). The SPP block helps separate significant context features at virtually no cost of

operational speed, while the PANet is supposedly more efficient and suitable for a single GPU training than the FPN (BOCHKOVSKIY et al., 2020). Other new features in YOLOv4 also include using the Mish activation function (MISRA, 2019), CutMix (YUN et al., 2019) and mosaic data augmentation, Cross mini-Batch Normalization (CmBN) (BOCHKOVSKIY et al., 2020), and a Complete Intersection over Union (CIoU) loss (ZHENG et al., 2019) for the bounding box regression problem.

The transfer learning technique was adopted to avoid that a large number of coffee fruit images were required during training. Thus, the training of the YOLOv4 model used for the detection of coffee fruits was initialized using parameters pre-trained on the MS COCO data set (LIN et al., 2014). The model was trained for different input image resolutions. The algorithm was set to detect a single class: coffee fruits. The confidence (probability of detection) threshold and non-maximum suppression threshold were set at 0.25 and 0.50, respectively. The performance of the YOLOv4 algorithm was also benchmarked against its predecessor, YOLOv3.

The model was implemented using the Darknet, an open source neural network framework written in C and CUDA. The model training and detections were run in a CPU Core i7-8700HQ running at 3.2 GHz with 64 GB RAM and a GeForce RTX 2070 graphics card with 8 GB dedicated memory.

3.2.4. Yield monitoring and mapping

The coffee fruits detections can be converted into yield maps by distributing the detections across the harvest lines and converting fruit counts into mass. Therefore, geographical coordinates were sampled during the recordings of each video during harvest. The coordinates were sampled at a frequency of 1 Hz using a C/A code type Global Navigation Satellite System (GNSS) receiver with Global Positioning System (GPS) and Globalnaya Navigatsionnaya Sputnikovaya Sistema (GLONASS). The geographical coordinates were associated with video frames based on the time of sampling. Since the video frames were collected at a higher frequency than the geographical coordinates (60 fps), the frames were uniformly distributed along the harvest line between consecutive sampling of coordinates.

Because videos were collected at 60 fps, the same coffee fruit may be present in consecutive video frames. To avoid that the same fruit is accounted for multiple times in the detection of consecutive frames, we analyzed the average number of frames (n) it would take for the fruits to completely cross the camera screen view (Figure 4). This analysis was

performed empirically by observing video segments sampled at random in several of the recorded videos. Finally, the frames were sampled every n = 7 frames to approximate the correct count of unique coffee fruits.

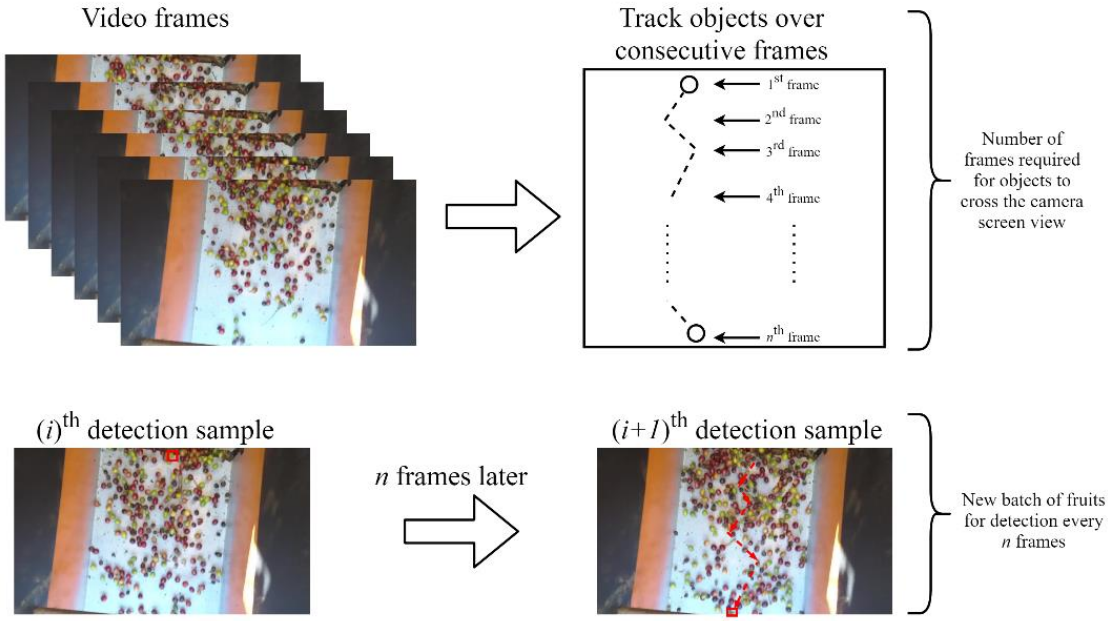

Figure 4. Scheme used for identifying the average number of frames (n) required for coffee fruits to completely cross the camera screen view. To estimate coffee yield, the frames were sampled every n frames to avoid re-counting the same fruits.

Despite the high number of frames per second, there was high variability in the number of detections in consecutive frames. This resulted in noisy coffee fruits detection series (Figure 5a). This problem is related to the scenario where data was collected, e.g., interference in illumination from the sunlight, blurring of frames due to machine vibration, impurities, and occlusion of objects. Thus, for the detection of coffee fruits sampled every 7 frames to better represent the number of harvested coffee fruits, the detection series were smoothed using the moving average of length equal to 7 (Figure 5b). The smoothed series resulted in the sampling of detection counts that avoid over or underprediction of objects.

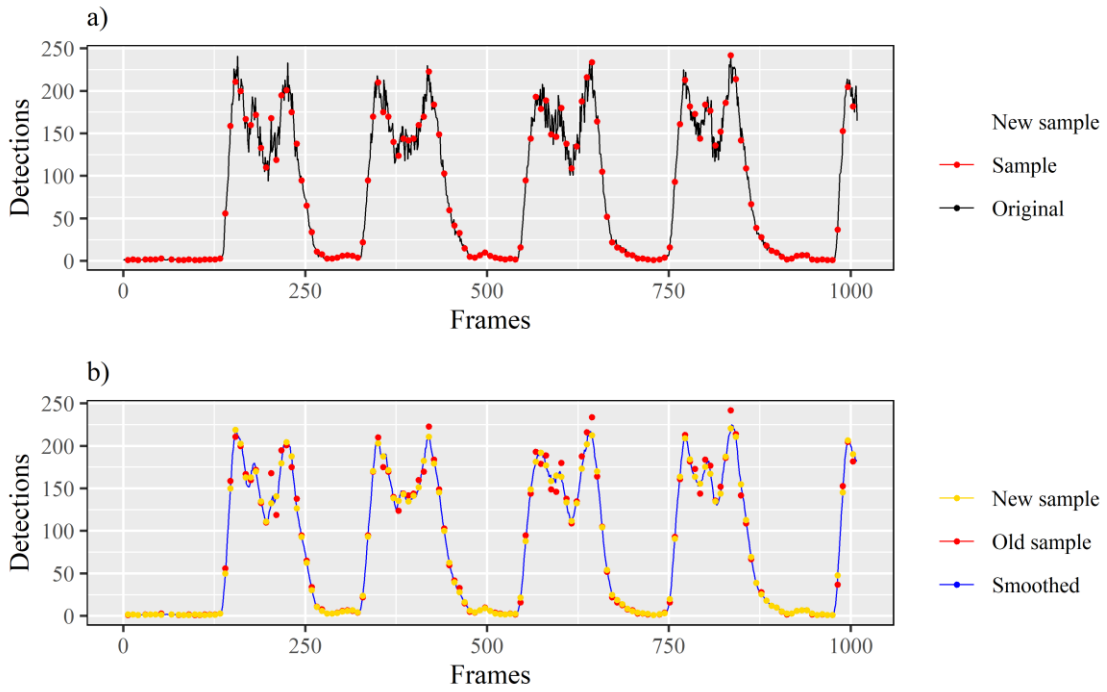

Figure 5. The number of detections made for each frame in an arbitrary video (**a**), and the detections counted smoothed by the moving average to avoid sampling detection counts when fruit detections peak upward or downward (**b**).

The detection of coffee fruits was converted into yield by considering a few conversion factors. First, the count of coffee fruits was converted into a volume of coffee fruits. For this, 70 samples of approximately 400 milliliters of coffee fruits were collected at random during harvest. First, we counted the number of coffee fruits per sample (~192 fruits per 400 milliliters). Then, the volume of coffee fruits was converted to mass. For this, the samples of one liter of coffee fruits were naturally dried on a patio, and the conversion factor was calculated (~127.7 grams of dry coffee per liter). The mass of coffee fruits concerning each frame was converted to yield (kg ha^{-1}) by considering the average distance traveled by the harvester and the distance between coffee lines of 4.0 m.

Yield maps were interpolated using the corresponding coordinates of the coffee yield estimates. The interpolation was performed using local kriging and the R programming language and environment (R Core Team, 2020). To validate the yield maps generated using both the yield monitor and the computer vision system, a controlled study was carried by storing the harvested coffee fruits in a container equipped with four weight cells. Thus, it was possible to compare the weight accumulation in the container to weight estimates provided by the yield monitor and the YOLOv4 model.

3.2.5. Performance criteria

The performance of the object detection model was assessed using the average precision (AP) of class detection for an intersection over union (IoU) of 50%. The AP summarizes the precision/recall curve defined as the mean precision at a set of eleven equally spaced confidence thresholds (c = 0.0, 0.1, …, 1.0) (Eq. 1). Precision (Eq. 2), p(c), and recall (Eq. 3), r(c), are computed for each confidence threshold and the graph is smoothed by setting the maximum precision measured for a method which the corresponding recall r' exceeds r (Eq. 4):

$$AP = \int_0^1 p(r)dr \quad (1)$$

$$p(c) = \frac{TP}{TP+FP} \quad (2)$$

$$r(c) = \frac{TP}{TP+FN} \quad (3)$$

$$p(r) = \max_{r':r' \geq r} p(r') \quad (4)$$

where AP is the average precision, p(r) is the precision at recall level r, TP are true positives, FP are false positives, FN are false negatives, p(r') is the precision at recall level r', and c is the confidence threshold. AP is usually reported as the mean average precision (mAP) of all detection classes. As we are considering only one class, mAP is equivalent to AP.

The yield map estimated using the detections from the YOLOv4 model was compared to the yield map made available by the yield monitor. Because the yield monitor outputs coffee yield in L ha^{-1}, it was also converted to mass using the conversion factor described in section 3.2.4 (127.7 grams of dry coffee per liter). The performance was assessed using the coefficient of determination (R^2) and the empirical distribution of data.

3.3. RESULTS AND DISCUSSION

The performance assessment of the models for the validation set and considering different network sizes is presented in Figure 6. The YOLOv4 model obtained higher mAP than the YOLOv3 model for all network sizes considered. Despite the mAP plateau seen for network sizes above 608 x 608 pixels, the highest performance, mAP = 83.5%, was achieved by the YOLOv4 model at a resolution of 800 x 800 pixels. The mAP is the average precision in detecting fruits and, therefore, larger values of this metric mean that a better outcome is expected in detecting coffee fruits (KUMAR et al., 2021). The results show that this model can be used as a technical reference in the implementation of coffee fruits detection systems

used to estimate the yield in coffee crops and aid the development of future technologies and research.

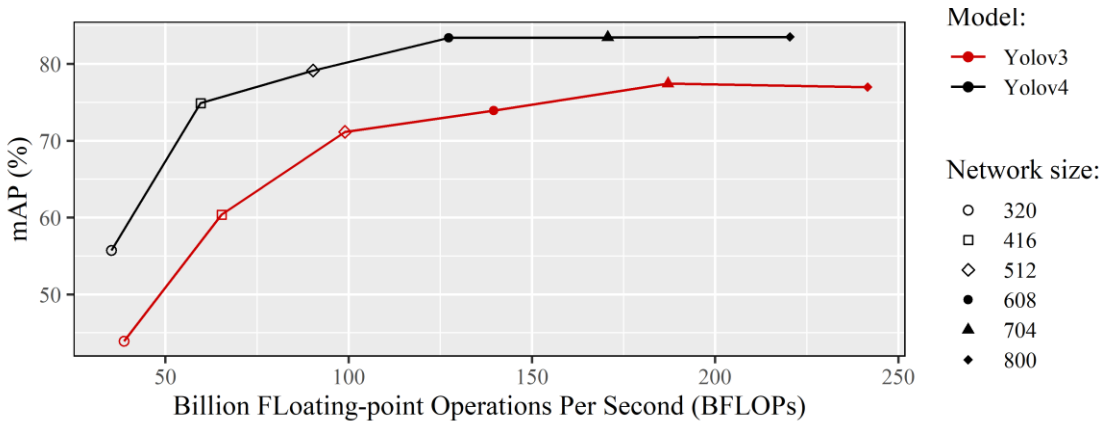

Figure 6. YOLOv3 and YOLOv4 mean average precision (mAP) and billion floating-point operations per seconds (BFLOPs) for each input resolution considered (network size).

Besides more efficient in object detection, the YOLOv4 model also showed lower computational cost (Figure 6). For example, the number of Floating Points Operations per second (FLOPs), which is a measure of computational performance, is roughly 8.8% lower for YOLOv4 when compared to YOLOv3. This means that the predictions for the newer model are performed in a shorter amount of time before post-processing the data for yield estimates.

The visual assessment of prediction results for images sampled in an arbitrary video is shown in Figure 7. Most of the coffee fruits were identified correctly, however, it was difficult for the model to detect fruits when they overlapped or when smaller (overripe/dry) fruits were close together (Figure 7g), underestimating the total number of detections (false negative).

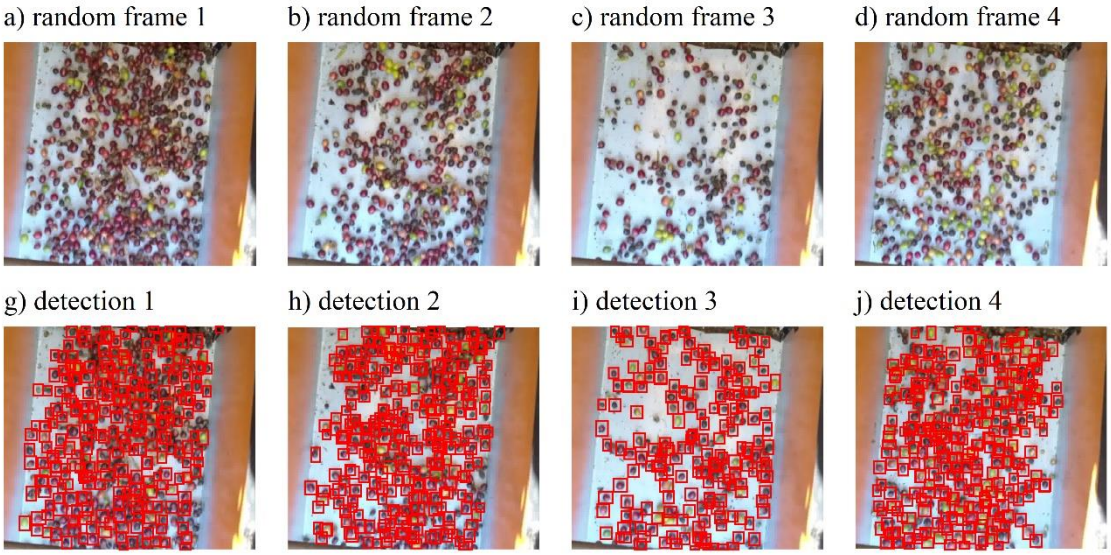

Figure 7. Detections made by the YOLOv4-800 model for random frames sampled from an arbitrary video.

The main challenges in the task for detecting coffee fruits in images were the illumination, vibration, occlusion of coffee fruits, overlapping, and impurities such as twigs and leaves. These challenges are evident when a larger load of coffee fruits is captured by the camera, as can be seen in Figure 7g when compared to figures 7h, 7i, and 7j. The average precision of the detections decreases to approximately 65% when the number of coffee fruits in the image is closer to 400 units (Figure 8).

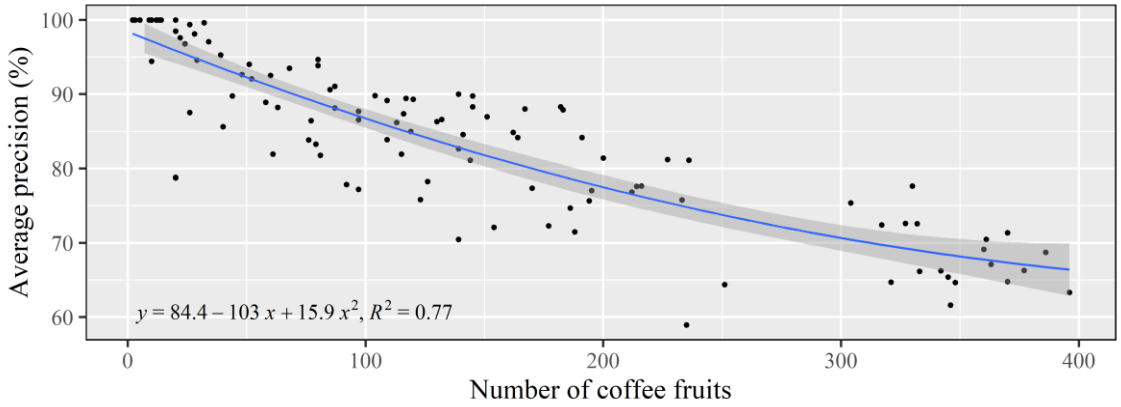

Figure 8. Average precision of coffee fruits detection using YOLOv4-800 by the number of coffee fruits in images.

After collecting the 70 samples of 1 liter of coffee fruits during harvest, we counted the number of coffee fruits per liter and the corresponding weight of dry coffee of these

samples (Figure 9). The average number of fruits per sample was 192 fruits per 400 milliliters (~480 fruits per liter), with a standard deviation of 15.4 fruits per 400 milliliters. After counting the fruits, the average weight of the samples after drying was 127.7 grams liter^{-1}, with a standard deviation of 14.2 grams liter^{-1}. These conversion factors were then used to convert the number of detections made by the YOLOv4 model to yield.

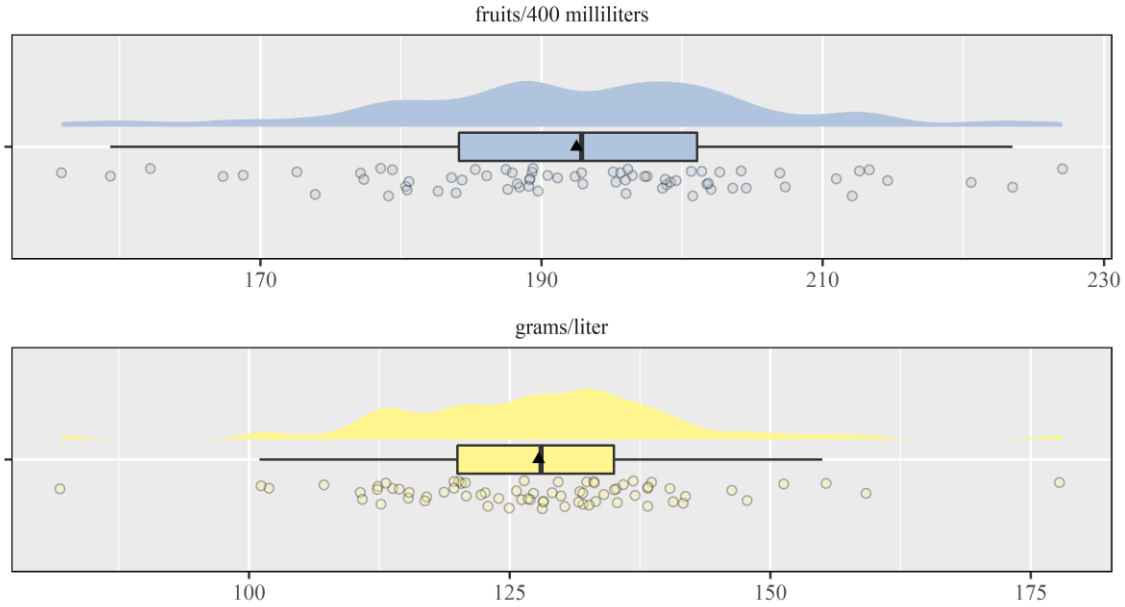

Figure 9. Number of fruits per 400 milliliters and grams of dry coffee per liter from 70 samples of coffee fruits collected during harvest.

The yield maps generated by the yield monitor of the coffee harvester and estimated using the YOLOv4-800 model are presented in Figure 10. Despite its differences, Figure 10 shows a clear resemblance of spatial patterns in yield maps from both sources. For example, both maps for area 1 show a higher yield in the northwestern part of the area and a lower yield in the southeastern part. The yield maps for area 2, on the other hand, show a lower average yield than area 1. However, both maps show a higher yield in the northern part of area 2.

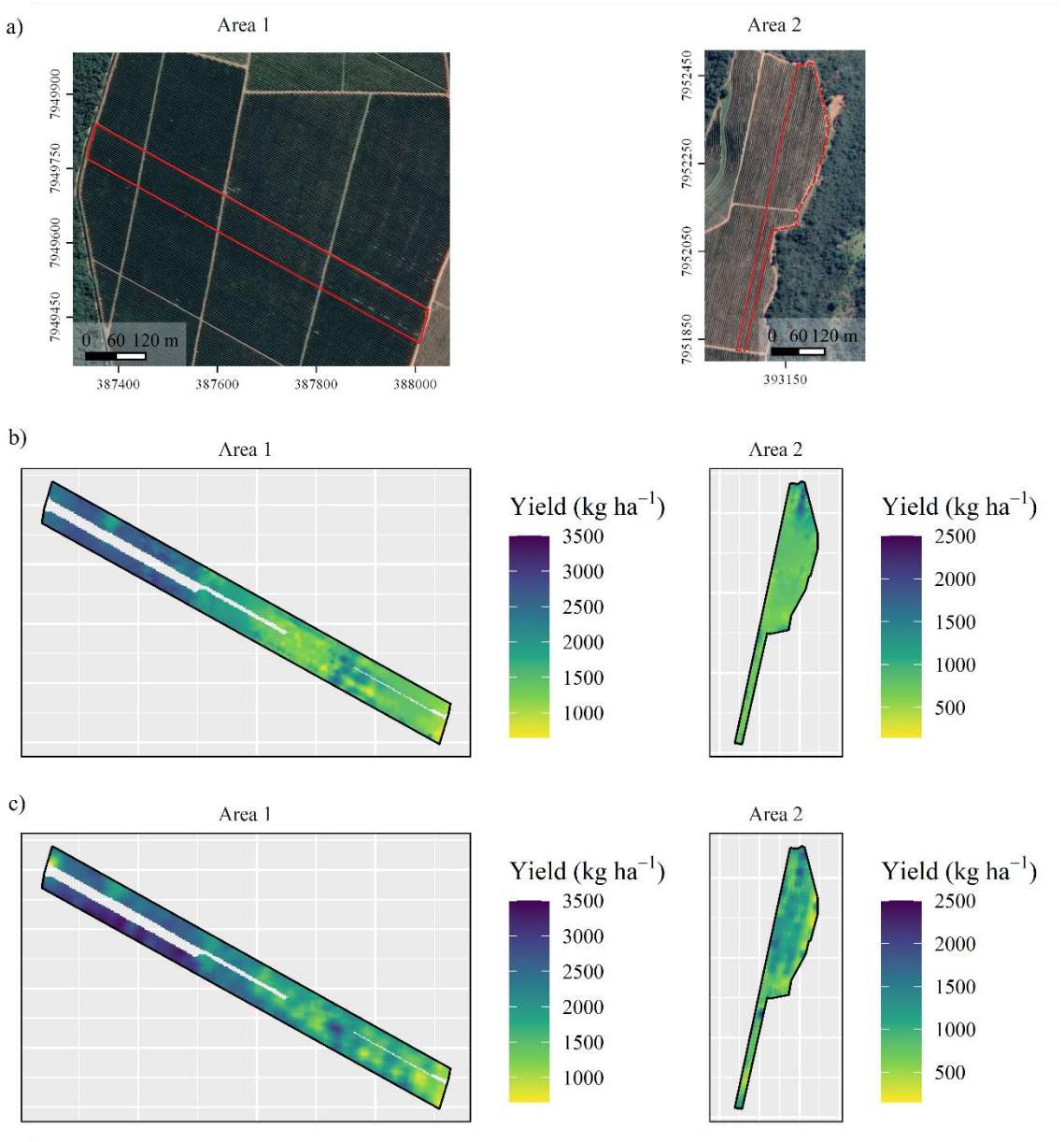

Figure 10. Study areas 1 and 2 (**a**), and yield maps estimated by the yield monitor and (**b**) by the YOLOv4-800 model (**c**).

The yield map of area 1 provided by the harvester yield monitor showed an average yield of ~1695 kg ha^{-1} and the yield map estimated based on the detections made by the YOLOv4-800 showed an average yield of ~1918 kg ha^{-1}, both showing a similar distribution of values (Figure 11a). For area 2, the average yield from the yield monitor was ~734 kg ha$^{-1,}$ and the one estimated based on the detections made by the YOLOv4-800 was ~891 kg ha^{-1} (Figure 11b). The yield maps estimated using the YOLOv4-800 detections showed an overall overestimation of yield when compared to the yield monitor. Unfortunately, the experimental areas are relatively small and with discontinuities. Nevertheless, it is worth noting that the

variability, represented by the interquartile ranges, was smaller for both yield maps in area 2 when compared to area 1, where spatial variability in yield is more notorious.

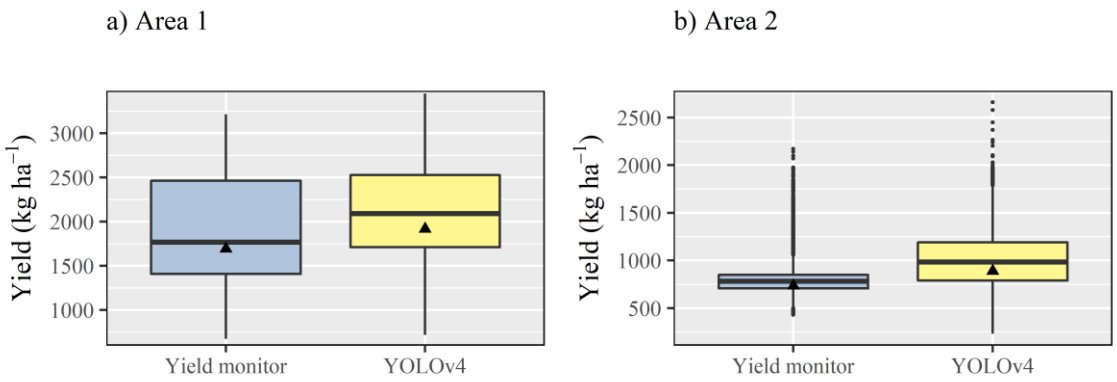

Figure 11. Distribution of yield estimated by the yield monitor and YOLOv4-800 model for (**a**) Area 1 and (**b**) Area 2.

The general overestimation of the yield maps can be attributed to different reasons. One of them is the representativeness of the conversion factors for coffee fruits to volume and mass. Despite the yield monitor using the same conversion factor from volume to mass, both factors rely on the size of fruits and perceptual of unripe, ripe, and overripe fruits in the samples considered. For area 2, this can be particularly attributed to the higher percentage of overripe fruits in the area, which are generally smaller and with a lower volume. Another reason for under or overestimating the yield is the number of frames adopted to assure that fruits are not accounted for multiple times. The numbers of frames can vary according to the harvesting travel speed, crop yield, maturation of coffee fruits, positioning of the side spout, and overall fruits load passing through the side spout. Both crop yield and fruit sizing are influenced by several factors, such as the coffee crop biennial behavior (Martins et al., 2021), soil type, climate, irrigation, altitude, and crop phenological characteristics and management (Carvalho et al., 2017; Venancio et al., 2020).

As for the detections, the mismatch between yield estimates can be explained by the area's late harvest due to the farm's machinery logistics. The harvest of this area showed a higher percentage of overripe fruits and impurities. Overripe fruits generally presented lower volume and mass, and no distinction was made between coffee fruits' maturation stage when converting to units of mass. The misclassification of impurities as coffee fruits can also lead to erroneous estimates of yield. However, as the model is trained to detect coffee fruits, the model rarely will misclassify impurities as coffee fruits. On the other hand, the yield monitor

accounts for all impurities as coffee fruits when filling its volume cells. Besides, the method proposed herein saves raw data for the entire harvest, which can be used to validate the model and look upon for further improving the methodology. In contrast, the yield monitor estimates can only be validated under specific set-ups in the field.

The yield maps provided by the yield monitor embedded on the harvester and estimated using the predictions made by the YOLOv4 model are compared pixel-wise in Figure 12. Despite a certain degree of scattering, an R^2 of 0.81 is achieved. Thus, the methodology proposed herein was able to explain approximately 81% of the variance in data measured by the yield monitor. Thus, the advantage of this methodology comes not only from its good performance but for presenting low implementation costs, since this study used only a camera, a set of LED lamps, and the image acquisition platform can be embedded in any coffee harvester. Camera performance and techniques can often be adjusted with softwares or programming languages. In addition, computer vision techniques can be suitable for large and small machines, and often, regardless of machine manufacturer and model (SCHUELLER, 2021). On the contrary, the yield monitor proposed by Sartori et al. (2002) was developed exclusively for a single company and harvester model (Queiroz et al., 2020).

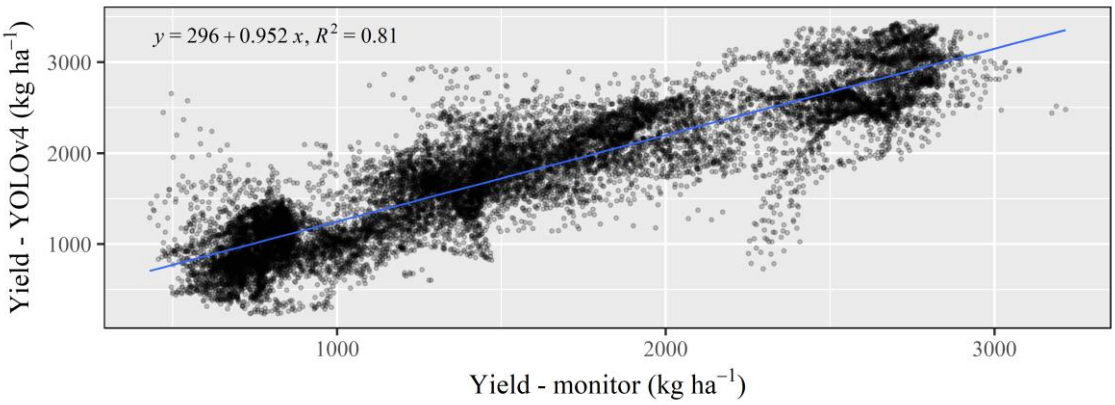

Figure 12. Scatterplot between the coffee yields estimated by the yield monitor and based on the detections made using the YOLOv4-800 model.

The controlled study where a container with weight cells is used to validate the weight of coffee fruits estimated using both the yield monitor and the YOLOv4 model is shown in Figure 13. The georeferenced data collection by each system in a coffee line of approximately 90 m is displayed in Figure 13a. For this line, the weight cells provided 1825 observations, while the yield monitor outputted only 28 data points, and the YOLOv4 provided detections for over 15 thousand video frames. Because different navigation systems were used to track

the position of each data collection, their offset was manually adjusted to match the harvest line. The volume recorded by the yield monitor and coffee fruits detected by the YOLOv4 model were converted to mass using the conversion factors described in section 3.2.4. The coffee fruits weighted in the container along the harvest line and estimated by both the yield monitor and computer vision system are displayed in Figure 13b. Because the yield monitor outputs data at a lower rate, its points were used as a reference to build double-mass curves (Figure 13c-d). For this, the nearest point from both the weight cells and YOLOv4 detections to the yield monitor points were used to compare accumulated weight along the line.

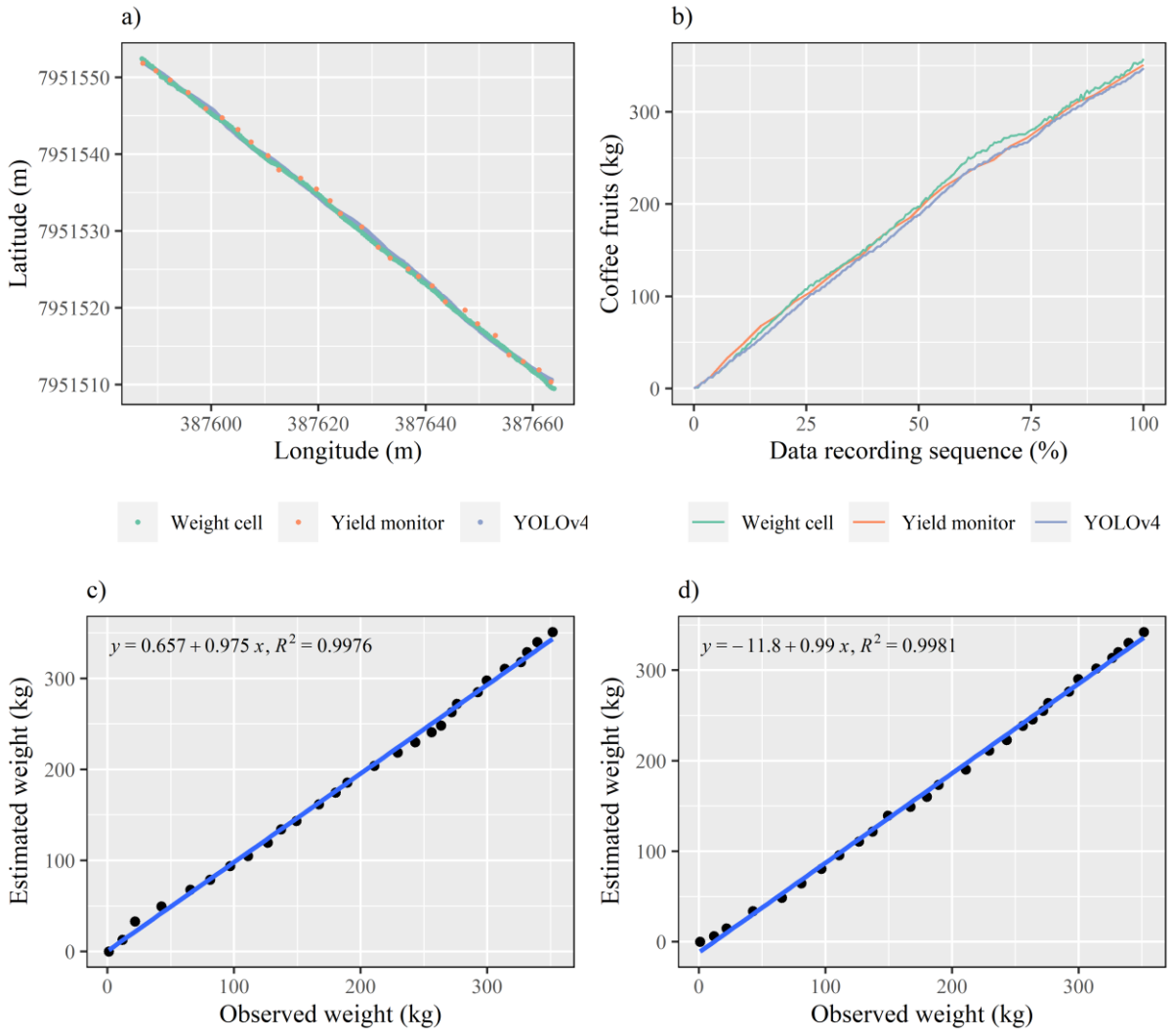

Figure 13. Georeferenced points for each data collection system **(a)**. Accumulated weight along the harvest line observed in the container equipped with weight cells and estimated by the yield monitor and YOLOv4 model. Coffee fruits weight in the container versus the accumulated weight **(b)** estimated by the yield monitor **(c)** and YOLOv4 model **(d)**.

Despite showing a high correlation between observed data and data estimated by both the yield monitor and YOLOv4, the covariance between yield estimates provided by the yield monitor and YOLOv4 model was much lower when compared for the experimental area (Figure 12, R^2 = 0.81). This is because Figure 13 compares only the data collected in a controlled study under good conditions. The entire experimental area has conditions that change a lot, as regions with much different maturation stage and levels of impurities. It was also a bigger challenge to match the coordinates of these different GNSS receivers for the entire area then for a single line in the controlled study.

The YOLOv4-800 model allowed for a fast and extensive assessment of yield at the field level. It is important to acknowledge the coffee crop specificities as a perennial crop since any factor can affect the plant yield and profitability for years (MARTINS et al., 2021). Besides, the yield maps have the potential to guide farmers to find problems that might develop for crop management strategies that consider average yield values for the entire area.

Information on spatial and temporal variability within the field has guided decision-making processes and effective management and planning of harvests for several crops (BAZAME et al., 2020; MALDANER et al., 2021). Advanced data analysis tools, such as precision agriculture practices, can help predict soil and plant attributes, which enhance how information is interpreted and, therefore, how a crop is managed (IDOL & YOUKHANA, 2020; MOLIN et al., 2020).

This study shows that yield can be estimated using computer vision models and replace the manual or less subjective ways to count coffee fruits. The results fill a gap in the literature concerning yield monitoring for coffee crops. A recommendation for future studies, given that the mAP has plateaued for higher network sizes (Figure 6), is the opportunity to make adaptions to the network structure to improve the detection precision when a large number of objects is present in the image (Figure 8). For example, Wang et al. (2021) adapted the YOLOv3 network structure to increase the output scale of feature maps before making predictions at the YOLO layers (head), which increased the model ability to detect a larger number of small objects. Obtaining more reliable yield conversion factors for coffee fruits at different maturation stages can also be adapted for a multi-classification model. Bazame et al. (2021) implemented a computer vision model to classify and map the maturation stage of coffee fruits during harvest. These studies reinforce that computer vision can increase the amount of information available for stakeholders and is an opportunity to increase coffee yield and improve beverage quality. Another recommendation would be testing other

algorithms for real-time object detection. This study shows the opportunity of using computer vision for executing simple but laborious tasks with high value for agriculture.

3.4. CONCLUSIONS

This study presents a computer vision algorithm that detects and counts coffee fruits during the mechanical harvest. The algorithm is based on the YOLOv4 neural network architecture and showed an mAP of 83.5% for images inputted at a resolution of 800 x 800 pixels. The detections performed on georeferenced videos recorded during harvest allowed for the coffee fruit counts to be translated into yield maps. The yield maps estimated from detections made by the computer vision model were able to explain 81% of the variance in reference yield maps. The main advantages of the proposed methodology are its low implementation cost and independence to specific brands of coffee harvesters. Yield mapping and monitoring provide stakeholders with information that allows for precision agriculture techniques to be adopted in crop management.

REFERENCES

BAZAME, H. C.; PINTO, F. A. C.; QUEIROZ, D. S.; QUEIROZ, D. M.; ALTHOFF, D. Spectral sensors prove beneficial in determining nitrogen fertilizer needs of *Urochloa brizantha* cv. Xaraés grass in Brazil. **Tropical Grasslands-Forrajes Tropicales**, v. 8, n. 2, p. 60–71, 30 maio 2020.

BAZAME, H. C.; MOLIN, J. P. ALTHOFF, D.; MARTELLO, M. Detection, classification, and mapping of coffee fruits during harvest with computer vision. **Computers and Electronics in Agriculture**, v. 183, p. 106066, 1 abr. 2021.

BOCHKOVSKIY, A.; WANG, C.-Y.; LIAO, H.-Y. M. YOLOv4: Optimal Speed and Accuracy of Object Detection. **arXiv**, 22 abr. 2020.

CARVALHO, L. C. C.; DA SILVA, F. M.; FERRAZ, G. A. E. S.; STRACIERI, J.; FERRAZ, P. F. P.; AMBROSANO, L. Geostatistical analysis of arabic coffee yield in two crop seasons. **Revista Brasileira de Engenharia Agricola e Ambiental**, v. 21, n. 6, p. 410–414, 2017.

HAMDAN, M. K. A.; ROVER, D. T.; DARR, M. J.; & JUST, J. Generalizable semi-supervised learning method to estimate mass from sparsely annotated images. **Computers and Electronics in Agriculture**, v. 175, p. 105533, 1 ago. 2020.

HE, K.; ZHANG, X.; REN, S.; SUN, J . Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 37, n. 9, p. 1904–1916, 2015.

IDOL, T. W.; YOUKHANA, A. H. A rapid visual estimation of fruits per lateral to predict coffee yield in Hawaii. **Agroforestry Systems**, v. 94, n. 1, p. 81–93, 1 fev. 2020.

KUMAR, A.; KALIA, A.; VERMA, K.; SHARMA, A.; KAUSHAL, M. Scaling up face masks detection with YOLO on a novel dataset. **Optik**, v. 239, p. 166744, 1 ago. 2021.

LIN, T.-Y.; DOLLÁR, P.; GIRSHICK, R.; HE, K.; HARIHARAN, B.; BELONGIE, S. Feature Pyramid Networks for Object Detection. 9 dez. 2016. Disponível em em:<http://arxiv.org/abs/1612.03144>.Acesso em: 20 de agosto de 2021.

LIN, T. Y.; MAIRE, M.; BELONGIE, S.; HAYS, J.; PERONA, P.; RAMANAN, D.; DOLLÁR, P.; ZITNICK, C. L. Microsoft COCO: Common objects in context. **Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)**, v. 8693 LNCS, n. PART 5, p. 740–755, 2014.

LIU, S.; QI, L.; QIN, H.; SHI, J.; JIA, J. Path Aggregation Network for Instance Segmentation. **Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition**, p. 8759–8768, 5 mar. 2018.

MALDANER, L. F.; CORRÊDO, L. P.; CANATA, T. F.; MOLIN, J. P. Predicting the sugarcane yield in real-time by harvester engine parameters and machine learning approaches. **Computers and Electronics in Agriculture**, v. 181, p. 105945, 1 fev. 2021.

MARIANO, C.; MÓNICA, B. A random forest-based algorithm for data-intensive spatial interpolation in crop yield mapping. **Computers and Electronics in Agriculture**, v. 184, p. 106094, 1 maio 2021.

MARTINS, R. N.; PINTO, F. A. C.; QUEIROZ, D. M.; VALENTE, D. S. M.; ROSAS, J. T. F. A novel vegetation index for coffee ripeness monitoring using aerial imagery. **Remote Sensing**, v. 13, n. 2, p. 1–16, 2 jan. 2021.

MISRA, D. Mish: A Self Regularized Non-Monotonic Activation Function. **arXiv**, 23 ago. 2019.

MOLIN, J. P.; ARAUJO MOTOMIYA, A. V.; FRASSON, F. R., CHIACCHIO FAULIN, G.; TOSTA, W. Método para avaliação de aplicaç ão de fertilizantes em taxa variávelem café. **Acta Scientiarum - Agronomy**, v. 32, n. 4, p. 569–575, 2010.

MOLIN, J. P.; BAZAME, H. C.; MALDANER, L.; CORREDO, L. P.; MARTELLO, M.; CANATA, T. F. Precision agriculture and the digital contributions for site-specific management of the fields. **Revista Ciencia Agronomica**, v. 51, n. 5, 2020.

OLIVEIRA, E. M.; LEME, D. S.; BARBOSA, B. H. G.; RODARTE, M. P.; ALVARENGA PEREIRA, R. G F. A computer vision system for coffee beans classification based on computational intelligence techniques. **Journal of Food Engineering**, v. 171, p. 22–27, 1 fev. 2016.

QUEIROZ, D. M.; COELHO, A. L. F.; VALENTE, D. S. M.; SCHUELLER, J. K. Sensors applied to Digital Agriculture: A review. **Revista Ciencia Agronomica**, v. 51, n. 5, p. 1–15, 2020.

REDMON, J.; DIVVALA, S.; GIRSHICK, R.; FARHADI, A. **You only look once: Unified, real-time object detection**. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. **Anais**...IEEE Computer Society, 9 dez. 2016

REDMON, J.; FARHADI, A. YOLO9000: Better, faster, stronger. **Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017**, v. 2017- Janua, p. 6517–6525, 2017.

REDMON, J.; FARHADI, A. YOLO v.3. **Tech report**, p. 1–6, 2018. Disponivel em:< https://pjreddie.com/media/files/papers/YOLOv3.pdf>.Acesso em: 22 de agosto de 2021

RODRÍGUEZ, J. P.; CORRALES, D. C.; AUBERTOT, J. N.; CORRALES, J. C. A Computer vision system for automatic cherry beans detection on coffee trees. **Pattern Recognition Letters**, v. 136, p. 142–153, 1 ago. 2020.

SANTOS, A. F.; CORRÊA, L. N.; LACERDA, L. N.; TEDESCO-OLIVEIRA, D.; PILON, C.; VELLIDIS, G.; DA SILVA, R. P. High-resolution satellite image to predict peanut maturity variability in commercial fields. **Precision Agriculture**, p. 1–15, 16 mar. 2021.

SARTORI, S.; FAVA, J. F. M.; DOMINGUES, E. L.; RIBEIRO FILHO, A. C.; SHIRAISI, L.E. Mapping the spatial variability of coffee yield with mechanical harvester. In: WORLD CONGRESS ON COMPUTERS IN AGRICULTURE AND NATURAL RESOURCES, 2002, Foz do Iguaçú. **Anais**… St. Joseph: ASAE, 2002. p. 196-205.

SCHUELLER, J. K. Opinion: Opportunities and Limitations of Machine Vision for Yield Mapping. **Frontiers in Robotics and AI**, vol. 0, p. 18, 25 Feb. 2021.

VENANCIO, L. P.; FILGUEIRAS, R.; MANTOVANI, E. C.; DO AMARAL, C. H.; DA CUNHA, F. F.; DOS SANTOS SILVA, F. C.; ALTHOFF, D.; DOS SANTOS, R. A.; CAVATTE, P. C. Impact of drought associated with high temperatures on coffea canephora plantations: A Case study in Espírito Santo State, Brazil. **Scientific Reports**, v. 10, n. 1, p. 1–21, 1 dez. 2020.

WANG, C.-Y. et al. CSPNet: A New Backbone that can Enhance Learning Capability of CNN. IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, 2020-June, 1571–1580. http://arxiv.org/abs/1911.11929

WANG, H.; DONG, L.; ZHOU, H.; LUO, L.; LIN, G.; WU, J.; TANG, Y. YOLOv3-Litchi Detection Method of Densely Distributed Litchi in Large Vision Scenes. **Mathematical Problems in Engineering**, v. 2021, 2021.

WU, D.; LV, S.; JIANG, M.; SONG, H. Using channel pruning-based YOLO v4 deep learning algorithm for the real-time and accurate detection of apple flowers in natural environments. **Computers and Electronics in Agriculture**, v. 178, p. 105742, 1 nov. 2020.

YUN, S.; HAN, D.; OH, S. J.; CHUN, S.; CHOE, J.; YOO, Y. CutMix: Regularization Strategy to Train Strong Classifiers with Localizable Features. **Proceedings of the IEEE International Conference on Computer Vision**, v. 2019- October, p. 6022–6031, 13 maio 2019.

ZHENG, Z.; WANG, P.; LIU, W.; LI, J.; YE, R.; REN, D. Distance-IoU Loss: Faster and Better Learning for Bounding Box Regression. **arXiv**, 19 nov. 2019.

4. DETECTION, CLASSIFICATION, AND MAPPING OF COFFEE FRUITS DURING HARVEST WITH COMPUTER VISION

Note: The article derived from this chapter has already been peer-reviewed and published online. However, this chapter presents some updates in the methodology and results.

BAZAME, H. C.; MOLIN, J. P. ALTHOFF, D.; MARTELLO, M. Detection, classification, and mapping of coffee fruits during harvest with computer vision. **Computers and Electronics in Agriculture**, v. 183, 2021. DOI: https://doi.org/10.1016/j.compag.2021.106066.

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 concerning 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 YOLOv4. 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 stabilized close to the 6000th iteration when considering an image input resolution of 800 x 800 pixels. The model presented an mAP of 92%, F1-Score of 88%, the precision of 85%, and recall of 92% for the validation set. The average precision for the classes of unripe, ripe, and overripe coffee fruits was 93%, 92%, and 91%, 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.

KEYWORDS: precision agriculture, convolutional neural networks, YOLO, deep learning.

4.1. 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; ANGÉLICO; CHALFOUN, 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 the 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; NASHAT, 2019; WU; SUN, 2013). This technique not only presents higher accuracy, but has been documented to be more versatile, faster, and can reduce labor costs (BELAN, PETERSON ADRIANO; DE ARAÚJO, SIDNEI ALVES; LIBRANTZ, 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 crops (HÄNI; ROY; ISLER, 2018; KOIRALA et al., 2019; LIU et al., 2020; WANG; WALSH; KOIRALA, 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.

4.2. MATERIAL AND METHODS

4.2.1 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 species, variety Catuaí 144, and are situated in Patos de Minas region, in the state of Minas Gerais, Brazil (Figure 1). The coffee crops were planted in the years 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.

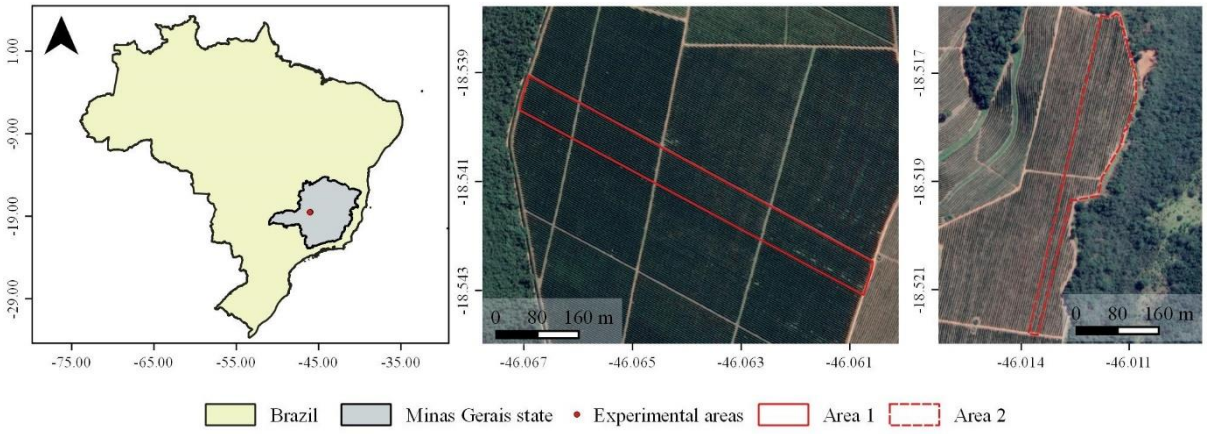

Figure 1. Location of the experimental areas of coffee crops in the state of Minas Gerais, Brazil. Coordinate reference system: WGS84.

The platform used to collect videos during harvest is shown in Figure 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 with mechanical shutter and 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 coffee harvester K3 Millennium (Máquinas Agrícolas Jacto S.A, 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.

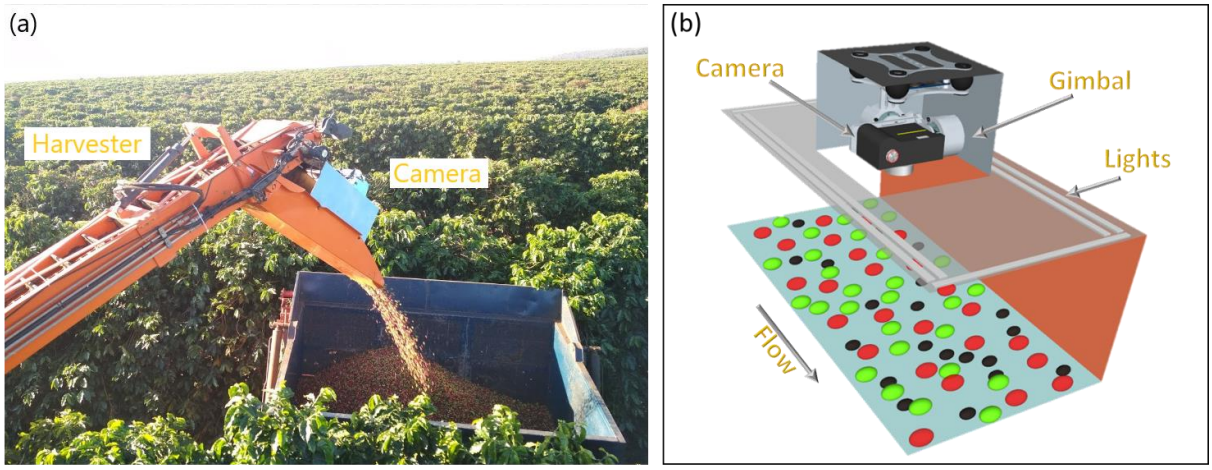

Figure 2. Lighting and image acquisition platform mounted onboard the coffee harvester.

4.2.2 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 full images and predicts the objects bounding boxes and probabilities of belonging to classes in one evaluation.

4.2.3 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, YOLOv3, and YOLOv4) (BOCHKOVSKIY; WANG; LIAO, 2020; REDMON; FARHADI, 2017, 2018). 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 the base. Although the algorithm can adjust to the dimensions of the bounding boxes during training, standard initial dimensions are configured for “anchor boxes” 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. At the prediction layer, besides the class predictions, five

predictions are made for each bounding box (Figure 3). 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. In this study, we adopted the YOLOv4 architecture, which uses a Complete Intersection over Union (CIoU) loss (ZHENG et al., 2019) for the bounding box regression problem during training. 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.

The YOLOv4 architecture uses the CSPDarknet53 backbone (WANG et al.,) as its feature extractor, which improves on Darknet53 backbone (REDMON; FARHADI, 2018) by including cross-stage partial connections (CSP). The YOLOv4 then collects feature maps from different stages using the Spatial Pyramid Pooling (SPP) (HE et al., 2015) and Path Aggregation Network (PANet) (LIU et al., 2018) as its neck. Finally, the predictions are made using output from three different stages and at three different scales using the YOLOv3 head (Figure 3). The predictions are made at scales 32, 16, and 8 times smaller than the input image. Predictions across stages can be beneficial to the model by using both refined features and initial feature extractor computations (REDMON; FARHADI, 2018).

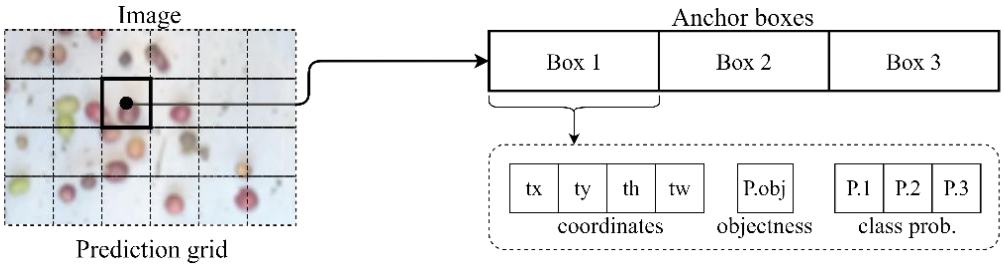

Figure 3. The prediction scheme of YOLOv3 head.

Adapted from the YOLOv4 version, YOLOv4-tiny is a network with a simplified feature extractor and with fewer convolutional layers. Similar to YOLOv4, YOLOv4-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.

4.2.4 Network design and training

The types of layers or blocks, number of filters, size and stride, size of input, and output tensor for layer/block of the network used are shown in Table 1. This table presents an example for a network which input resolution is 800 x 800 pixels, termed here as YOLOv4-800. 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) x 3 anchors] (layers 36, 40, and 44).

Table 1. Example of the structure of the YOLOv4-800 model used in the study.

	Ref.	Type	Filters	Size/Stride	Input	Output
		Resize input	-	-	-	800 x 800 x 3
Backbone	1	Conv. layer	32	3 x 3/1	800 x 800 x 3	800 x 800 x 32
	2	CSP-1	64	-	800 x 800 x 32	400 x 400 x 64
	3	CSP-2	128	-	400 x 400 x 64	200 x 200 x 128
	4	CSP-8	256	-	200 x 200 x 128	100 x 100 x 256
	5	CSP-8	512	-	100 x 100 x 256	50 x 50 x 512
	6	CSP-4	1024	-	50 x 50 x 512	25 x 25 x 1024
	7	Conv. layer	512	1 x 1/1	25 x 25 x 1024	25 x 25 x 512
	8	Conv. layer	1024	3 x 3/1	25 x 25 x 512	25 x 25 x 1024
	9	Conv. layer	512	1 x 1/1	25 x 25 x 1024	25 x 25 x 512
SPP	10	[9] MaxPool	-	5/1	25 x 25 x 512	25 x 25 x 512
	11	[9] MaxPool	-	9/1	25 x 25 x 512	25 x 25 x 512
	12	[9] MaxPool	-	13/1	25 x 25 x 512	25 x 25 x 512
	13	**[9, 10, 11, 12]**				25 x 25 x 2048
PANet	14	Conv. block (x3)	512	-	25 x 25 x 2048	25 x 25 x 512
	15	Conv. layer	256	1 x 1/1	25 x 25 x 512	25 x 25 x 256
	16	Upsample	-	-/2	25 x 25 x 256	50 x 50 x 256
	17	**Route 5**				
	18	Conv. layer	256	1 x 1/1	50 x 50 x 512	50 x 50 x 256
	19	**[16, 18]**				50 x 50 x 512
	20	Conv. block (x5)	256	-	50 x 50 x 512	50 x 50 x 256
	21	Conv. layer	128	1 x 1/1	50 x 50 x 512	50 x 50 x 128
	22	Upsample	-	-/2	50 x 50 x 128	100 x 100 x 128
	23	**Route 4**				
	24	Conv. layer	128	1 x 1/1	100 x 100 x 256	100 x 100 x 128
	25	**[22, 24]**				100 x 100 x 256
	26	Conv. block (x5)	128	-	100 x 100 x 256	100 x 100 x 128
	27	Conv. layer	256	3 x 3/2	100 x 100 x 128	50 x 50 x 256
	28	**[20, 27]**				50 x 50 x 512
	29	Conv. block (x5)	256	-	50 x 50 x 512	50 x 50 x 256
	30	Conv. layer	512	3 x 3/2	50 x 50 x 256	25 x 25 x 512
	31	**[14, 30]**				25 x 25 x 1024
	32	Conv. block (x5)	512	-	25 x 25 x 1024	25 x 25 x 512
H	33	**Route 26**				

34	Conv. layer	256	3 x 3/1	100 x 100 x 128	100 x 100 x 256
35	Conv. layer	24	1 x 1/1	100 x 100 x 128	100 x 100 x 24
36	*YOLO			100 x 100 x 24	
37	**Route 29**				
38	Conv. layer	512	3 x 3/1	50 x 50 x 256	50 x 50 x 512
39	Conv. layer	24	1 x 1/1	50 x 50 x 512	50 x 50 x 24
40	*YOLO			50 x 50 x 24	
41	**Route 32**				
42	Conv. layer	1024	3 x 3/1	25 x 25 x 512	25 x 25 x 1024
43	Conv. layer	24	1 x 1/1	25 x 25 x 1024	25 x 25 x 24
44	*YOLO			25 x 25 x 1024	

*Layer where detections take place.

In addition, other dimensions of image resampling (320 x 320, 416 x 416, 608 x 608, 704 x 704, and 800 x 800 pixels) for the input in the YOLOv4 and YOLOv4-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 model pre-trained on the COCO data set (80 object categories, 1.5 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” techniques were also used during training. These techniques include randomly rotating the images, increasing or decreasing their exposure, using CutMix (YUN et al., 2019), and mosaic data augmentation. Thus, the YOLO object detection 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.

4.2.5. 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 Equations 1 to 3 (OLSON; DELEN, 2008) below.

$$\text{Precision} = \frac{TP}{TP + FP} \quad (1)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (2)$$

$$\text{F1-Score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (3)$$

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

In addition, the average precision of the classes (AP) (Equations 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 x Recall curve. The average precision is the precision averaged at all recall values between 0 and 1 (MAZZIA et al., 2020):

$$\int_0^1 P(r)dr \quad (4)$$

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:

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

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.

4.2.6. Mapping the coffee fruits maturation stage

The geographic coordinates of each video recording were sampled using the same methodology as described in chapter 2. 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.

4.3. RESULTS AND DISCUSSION

4.3.1. Object detection performance

The performance achieved by the YOLOv4 models considering different image input resolutions is presented in Figure 4. A more significant improvement was observed by increasing the resolution of input images from 320 x 320 pixels to 608 x 608 pixels, while the performance seemed to stabilize for input resolutions equal to or above 704 x 704 pixels. The best performance was achieved considering the input resolution of 800 x 800 pixels close to the 6000th iteration. It is possible to see that the performance for the test set does not improve much from the 4000th iteration and to continue the training could result in the over-fitting of the model. For this reason, we stopped training the model after the 6000th iteration. The

YOLOv4-800 and YOLOv4-tiny-800 models showed mAPs of 91.8% and 92.9%, F1-Scores of 88.0% and 84.0%, precisions of 85.0% and 90.0%, and recalls of 92.0% and 79.0% for the validation set, respectively.

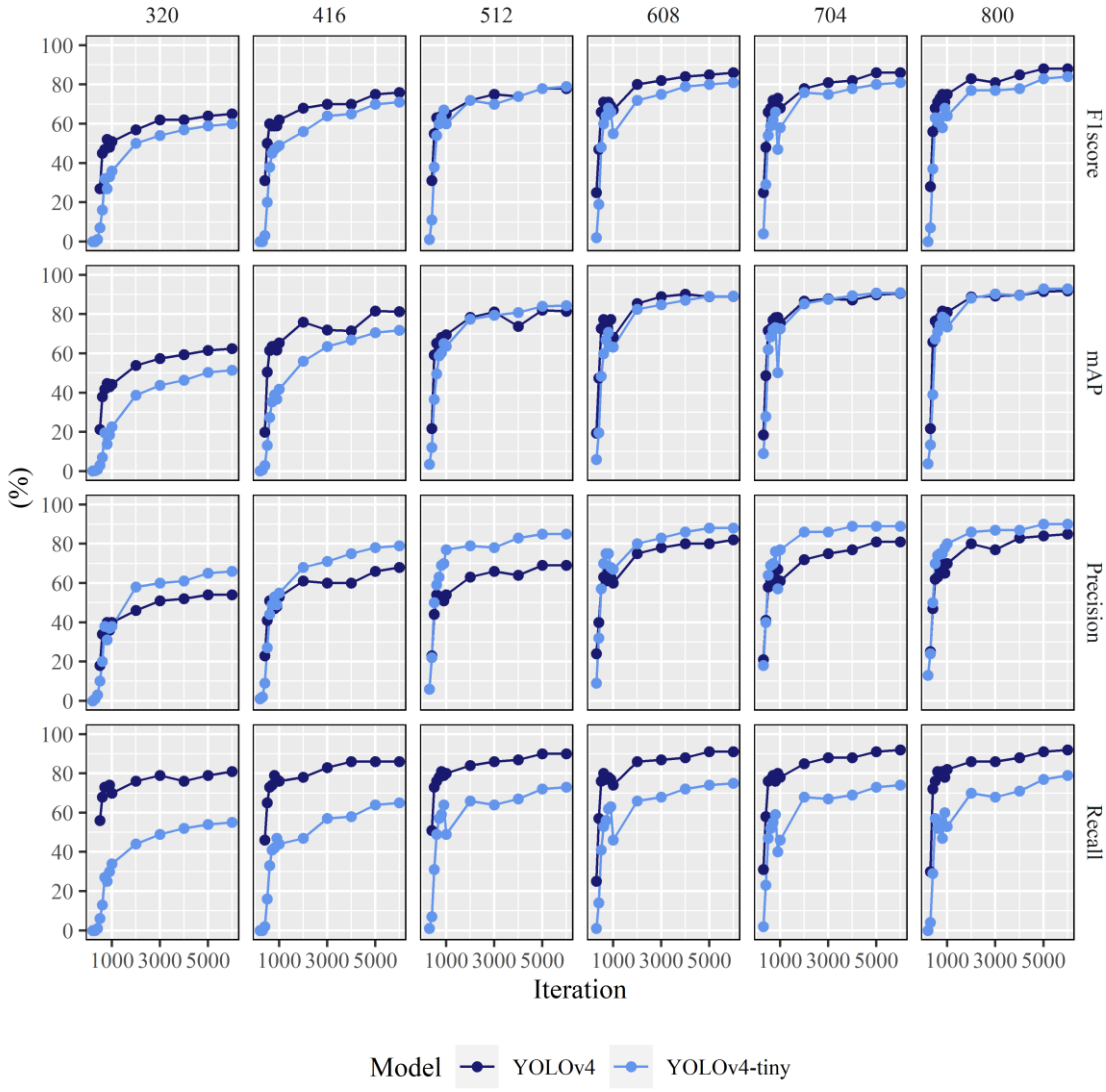

Figure 4. Performance criteria achieved by the YOLOv4 and YOLOv4-tiny models for the validation set.

The metrics showed balanced results with close values of precision and recall, which is important for the model. However, we can see that the YOLOv4-tiny models generally obtained higher precision, but lower recall when compared to YOLOv4 models. Despite presenting a higher mAP for input resolution of 800 x 800 pixels, the lower F1-Score and recall scores obtained by YOLOv4-tiny means that many objects go undetected (false

negatives). In contrast, the lower precision obtained by YOLOv4 means it sometimes mispredicts objects where are none (false positives). Overall, the higher recall obtained by YOLOv4 means it is more likely to detect coffee fruits, which can be especially useful in images with a high density of objects. Thus, the YOLOv4 model was adopted to map the coffee fruits maturation stage in the following step of this work.

The performance analysis for the YOLOv4 models, based on the three classes of fruit ripening stage (unripe, ripe, and overripe), are shown in Figure 5. For the validation set and considering an image input resolution of 800 x 800 pixels, the YOLOv4 model reached an AP of 93.0%, 91.9%, 90.7% for the classification of unripe, ripe, and overripe fruits, respectively, while the YOLOv4-tiny scored APs of 93.7%, 94.4%, and 90.6%, respectively. The lower 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 x 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, because fruits can be small or overlapped and not obvious (KOIRALA et al., 2019).

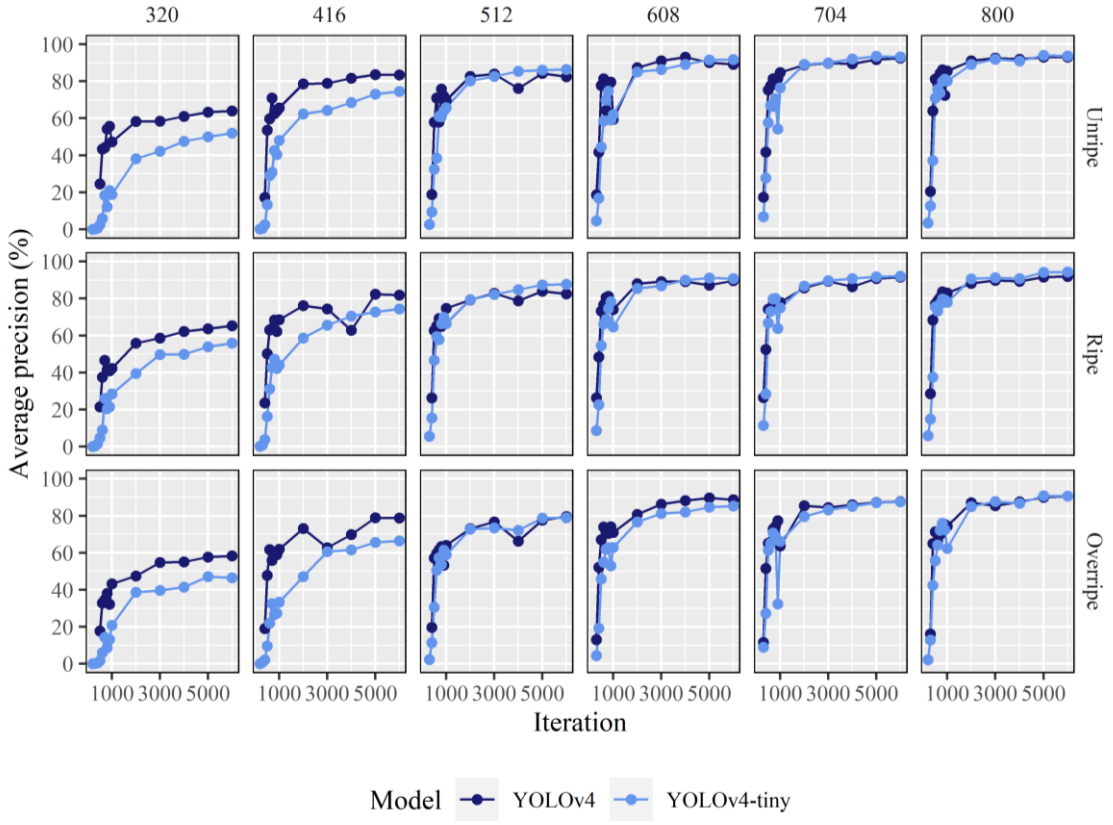

Figure 5. Average precision achieved by the YOLOv4 and YOLOv4-tiny models by the class of coffee fruits for the validation set of.

The color of the fruits, leaves and the number 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 an 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 YOLOv4-800 model (Figure 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 93.7% for the first arbitrary frame (Figure 6a-b) and 68.2% for the second arbitrary frame (Figure 6c-d). 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 x 800 pixels during detection can result in clusters of fruits of similar colors blurring together.

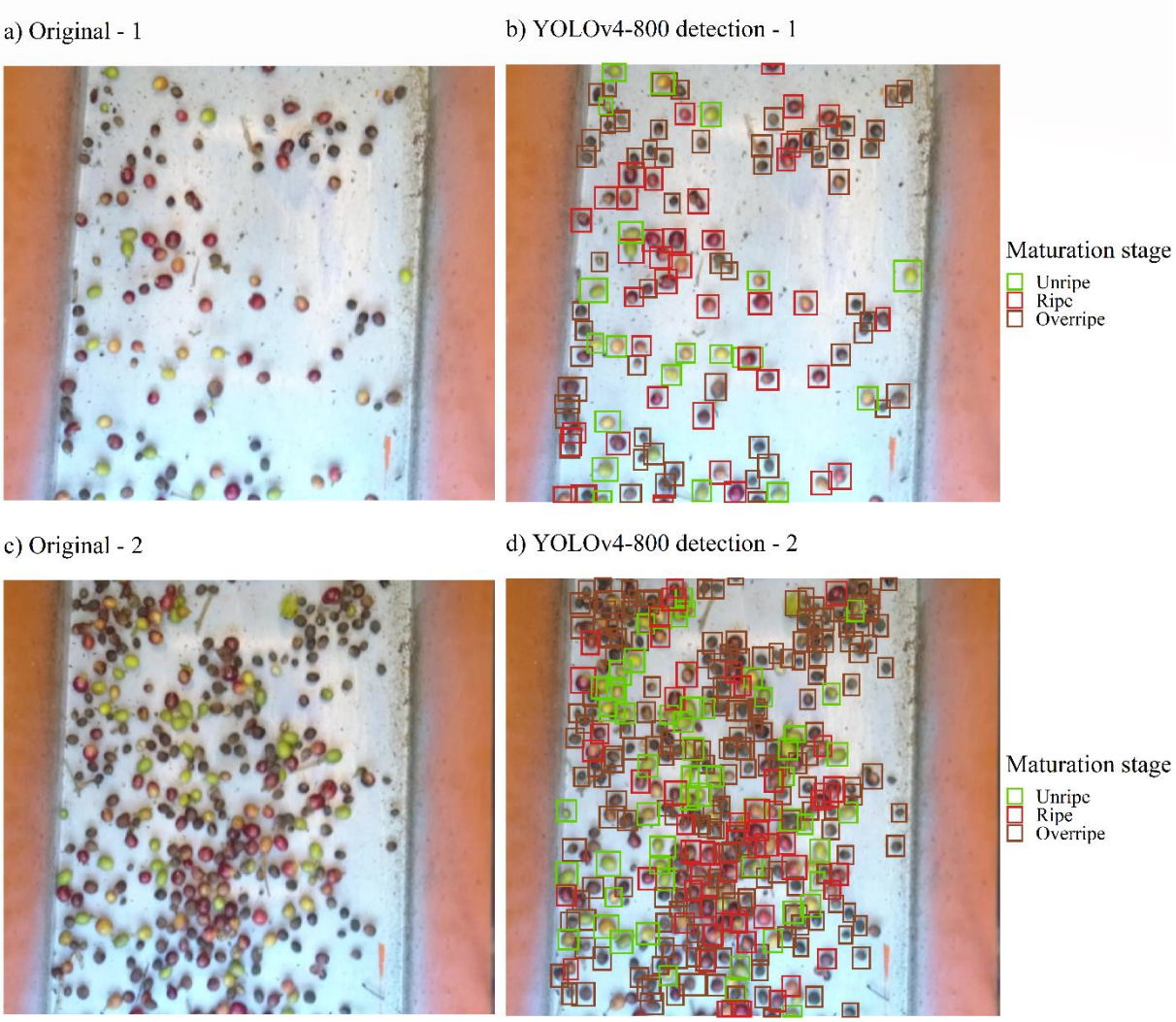

Figure 6. Two arbitrary video frames collected during the coffee harvest: (a-c) original frames with fruits at different stages of maturation, (b-d) detection performed by the proposed model.

The 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 YOLOv4-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 with 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).

4.3.2. 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 (Figure 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 (Figure 7a). Images with less than 40 fruits were, therefore, excluded from the database (Figure 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.

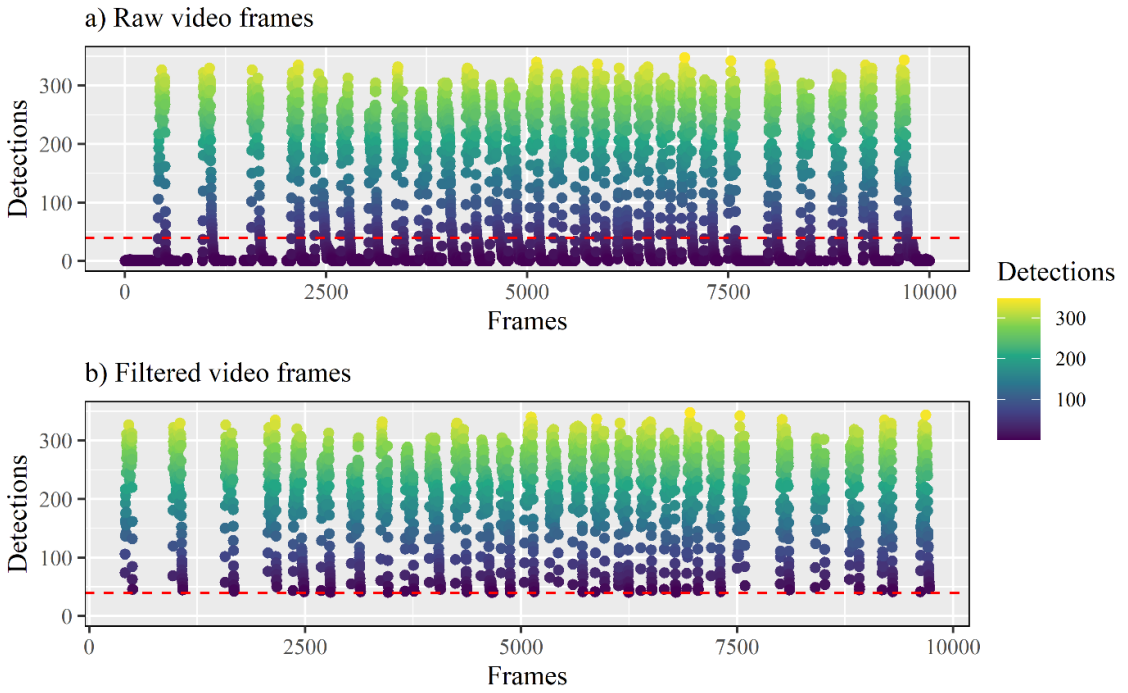

Figure 7. Time series used to determine the threshold of minimum fruit count for quality mapping.

For an arbitrary video, the total number of coffee fruits detected in each frame of the video is shown in Figure 8a, the total number of fruits detected per class in Figure 8b, and the proportion of classes detected per frame in Figure 8c. Along with the frames (harvest line), the proportion of ripe fruits was more or less stable, with a subtle increase of unripe fruits between the frames 4500 and 6500. Along the harvest line, a large proportion of overripe coffee fruits (28.9% to 71.2%) were detected in relation to ripe (23.8% to 59.3%) and unripe fruits (1.7 to 32.3%) (Figure 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.

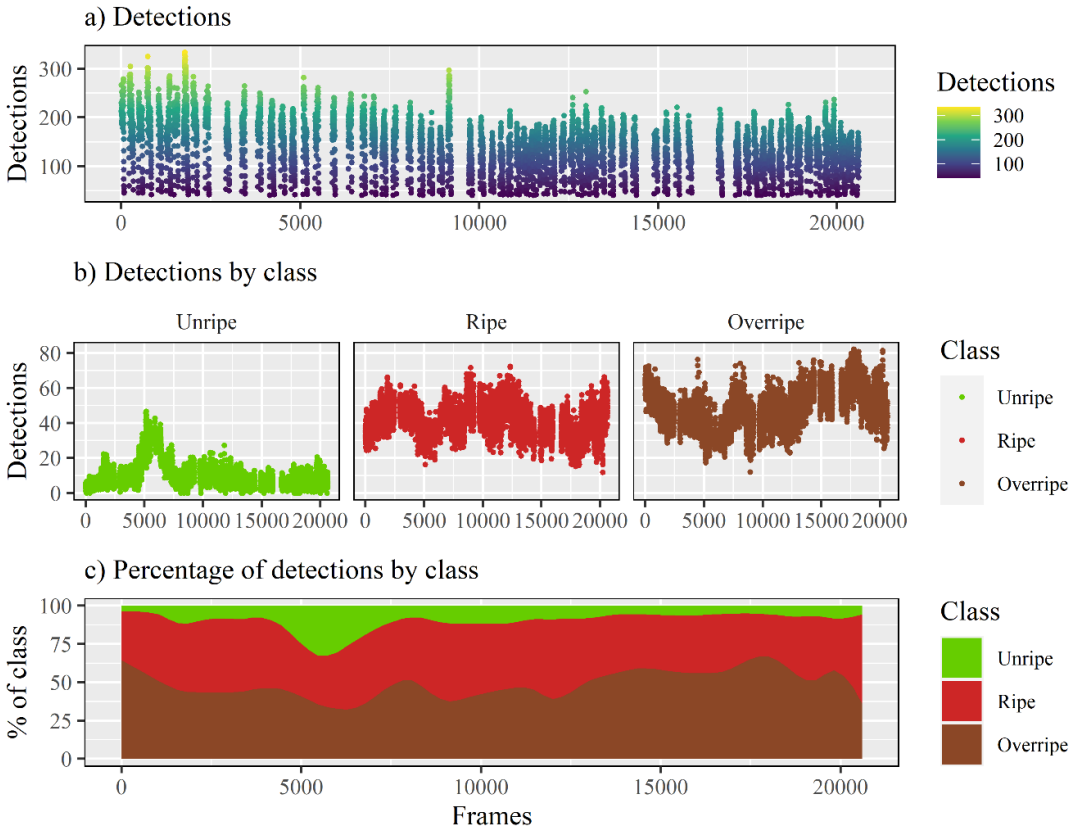

Figure 8. For an arbitrary video, (a) the total number of detected fruits, (b) the total number of detections by ripening class, and (c) the proportion of the detected classes.

For the same video used in Figure 8, Figure 9 shows the total of detected fruits in each frame (Figure 9a) and the detections by class (Figure 9b) mapped along the respective harvest line. In both Figures 8a and 9a, it is evident that the beginning of the line presented a higher number of detections. In Figure 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).

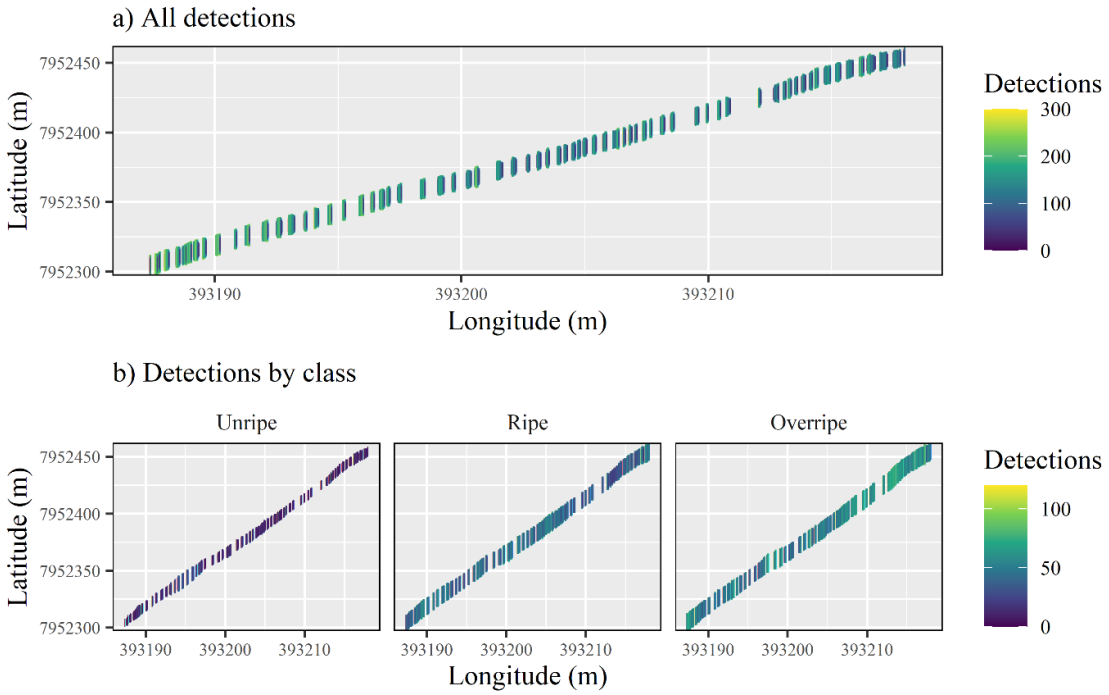

Figure 9. For an arbitrary video, mapping of (a) total coffee fruit detections and (b) class detections along the harvest line. Coordinate reference system: WGS84 / UTM zone 23S.

The coffee fruit classification at different maturation stages was spatialized, generating maps with information related to crop quality attributes (Figure 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 (Figure 10a), a higher percentage of unripe fruits was harvested in relation to Area 2 (Figure 10b), presenting an overall more balanced proportion between classes of maturation stage (Figure 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.

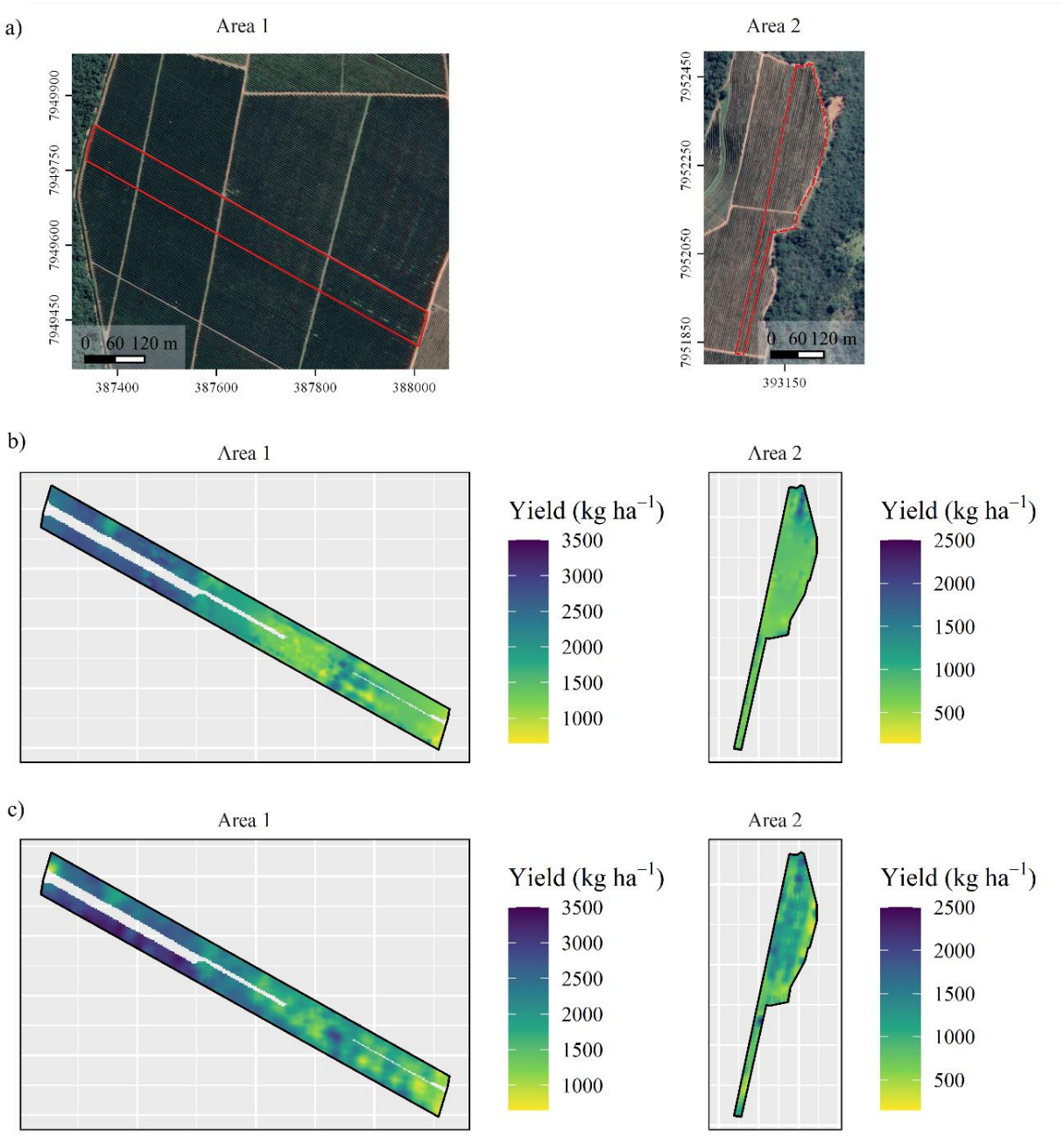

Figure 10. Proportions of detections mapped for (a) Area 1 and (b) Area 2, and (c) the distributions of class detections (μ = mean; σ = standard deviation). Coordinate reference system: WGS84 / UTM zone 23S.

The 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, 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.

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).

4.4. CONCLUSIONS

The deep learning model adopted in this study supports the possibility of detecting coffee fruits and classifying their maturation stage regardless of 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 YOLOv4 neural networks, proved to be robust and computationally efficient. The model presented as performance criteria an mAP of 91.8%, F1-Score of 88.0%, the precision of 85.0%, and recall of 92.0% for the validation set. The average precision for the classes of unripe, ripe, and overripe coffee fruits was 93.0%, 91.9%, and 90.7%, respectively. A close performance was obtained by the YOLOv4-tiny and, the low computational demand of the "tiny" version 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.

REFERENCES

AVENDANO, J.; RAMOS, P. J.; PRIETO, F. A. A system for classifying vegetative structures on coffee branches based on videos recorded in the field by a mobile device. **Expert Systems with Applications**, v. 88, p. 178–192, 1 dez. 2017.

BALUJA, J.; DIAGO, M. P.; BALDA, P.; ZORER, R.; MEGGIO, F.; MORALES, F.; TARDAGUILAJ. Spatial variability of grape composition in a Tempranillo (Vitis vinifera L.) vineyard over a 3-year survey. **Precision Agriculture**, v. 14, n. 1, p. 40–58, 9 set. 2013.

BELAN, P. A.; DE ARAÚJO, S. A.; LIBRANTZ, A. F. H. Técnicas de visão computacional aplicadas no processo de calibração de instrumentos de medição com display numérico digital sem interface de comunicação de dados. **Directory of Open Access Journals**. Disponível em: <https://doaj.org/article/990a32f8be3240979c8f69c0ddf5fb4f>. Acesso em: 31 agosto de 2020.

BOCHKOVSKIY, A.; WANG, C.-Y.; LIAO, H.-Y. M. YOLOv4: Optimal Speed and Accuracy of Object Detection. **arXiv**, 22 abr. 2020.

BRESILLA, K.; PERULLI, G. D.; BOINI, A.; MORANDI, B.; CORELLI, L.G.; MANFRINI, L.. Single-shot convolution neural networks for real-time fruit detection within the tree. **Frontiers in Plant Science**, v. 10, 16 abr. 2019.

DALVI, L. P.; NEY SUSSUMU SAKIYAMA, N. S.; SILVA, F. A. P.; CECON, P. R. Qualidade de café nos estádios cereja e verde-cana via condutividade elétrica Quality of coffee cherry and sugarcane-green stage by electrical conductivity. **Revista Agrarian**, v.6, n.22, p.410-414, 2013.

DE OLIVEIRA, E. M.; LEME, D. S.; BARBOSA, B. H. G.; RODARTE, M. P.; ROSEMARY GUALBERTO FONSECA ALVARENGA PEREIRA, G. F. A. A computer vision system for coffee beans classification based on computational intelligence techniques. **Journal of Food Engineering**, v. 171, p. 22–27, 1 fev. 2016.

DONIZETTI, A.; MENDONÇA, L. M. V. L.; DIAS, R. A. A.; PRADO, A. S.; DIAS, R. E. B. A.; PEREIRA, S. Qualidade do café colhido em diferentes estádios de maturação. **Anais** ... VII Simpósio de Pesquisa dos Cafés do Brasil, 2011.

FERRAZ, M.N.; CORRÊDO, L.P.; WEI, M.C.F.; MOLIN, J.P. Spatial variability mapping of sugarcane qualitative attributes. **Engenharia Agrícola**, v. 39, n. spe, p. 109–117, set. 2019.

HÄNI, N.; ROY, P.; ISLER, V. Apple Counting using Convolutional Neural Networks. IEEE International Conference on Intelligent Robots and Systems. **Anais** ... Institute of Electrical and Electronics Engineers Inc., 27 dez. 2018.

HE, K.; ZHANG, X.; REN, S.; SUNET, J. Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 37, n. 9, p. 1904–1916, 2015.

KAZAMA, E.H.; DA SILVA, R.P.; TAVARES, T.O.; CORREA, L. N.; ESTEVAM, F. N. L.; NICOLAU, F. E. A.; Maldonado Júnior, W. Methodology for selective coffee harvesting in management zones of yield and maturation. **Precision Agriculture,** 22, 711–733, 2021.

KOIRALA, A.; WANG, Z.; WALSH, K.; MCCARTHY, C. Deep learning for real-time fruit detection and orchard fruit load estimation: benchmarking of 'MangoYOLO'. **Precision Agriculture**, v. 20, n. 6, p. 1107–1135, 1 dez. 2019.

KOIRALA, A.; WALSH, K.B.; WANG, Z.; ANDERSON, N. Deep learning for mango (Mangifera indica) panicle stage classification. **Agronomy**, v. 10, n. 1, p. 1–22, 2020.

LÄDERACH, P.; LUNDY, M.; JARVIS, A.; RAMIREZ, J.; PEREZ PORTILLA, E.; SCHEPP, K. Systematic agronomic farm management for improved coffee quality. **Field Crops Research**, v. 120, n. 3, p. 321–329, 14 fev. 2011.

LEME, D.S.; DA SILVA, S.A.; BARBOSA, B.H.G.; BORÉM, F.M.; PEREIRA, R.G.F.A. Recognition of coffee roasting degree using a computer vision system. **Computers and Electronics in Agriculture**, v. 156, p. 312–317, 1 jan. 2019.

LIN, T. Y. et al. Microsoft COCO: Common objects in context. **Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)**, v. 8693 LNCS, n. PART 5, p. 740–755, 2014.

LIU, G.; NOUAZE, J.C.; TOUKO MBOUEMBE, P.L.; KIM, J.H. YOLO-tomato: A robust algorithm for tomato detection based on YOLOv3. **Sensors (Switzerland)**, v. 20, n. 7, 1 abr. 2020.

LIU, S. et al. Path Aggregation Network for Instance Segmentation. **Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition**, p. 8759–8768, 5 mar. 2018.

MATIELLO et al. **Cultura de Café no Brasil: Manual de recomendações**. 1ª ed. [s.l.] Fundacão procafé, 2015.

MAZEN, F. M. A.; NASHAT, A. A. Ripeness Classification of Bananas Using an Artificial Neural Network. **Arabian Journal for Science and Engineering**, v. 44, n. 8, p. 6901–6910, 1 ago. 2019.

MAZZIA, V.; KHALIQ, A.; SALVETTI, F.; CHIABERGE, M. Real-time apple detection system using embedded systems with hardware accelerators: An edge AI application. **IEEE Access**, v. 8, p. 9102–9114, 2020.

MESQUITA, C. M. DE et al. **Manual do café colheita e preparo**Belo HorizonteBelo Horizonte Lastro Editora, , 2008.

MOLIN, J. P.; ARAUJO MOTOMIYA, A. V.; FRASSON, F. R., CHIACCHIO FAULIN, G.; TOSTA, W. Método para avaliação de aplicaç ão de fertilizantes em taxa variávelem café. **Acta Scientiarum - Agronomy**, v. 32, n. 4, p. 569–575, 2010.

OLSON, D. L.; DELEN, D. **Advanced data mining techniques**. [s.l.] Springer Berlin Heidelberg, 2008.

PIMENTA, C. J.; ANGÉLICO, C. L.; CHALFOUN, S. M. Challengs in coffee quality: Cultural, chemical and microbiological aspects. **Ciência e Agrotecnologia**, vol. 42, 337–349, 2018.

RAMOS, P. J.; PRIETO, F. A.; MONTOYA, E. C.; OLIVEROS, C.E. Automatic fruit count on coffee branches using computer vision. **Computers and Electronics in Agriculture**, v. 137, p. 9–22, 1 maio 2017.

RAMOS, P. J.; AVENDAÑO, J.; PRIETO, F. A. Measurement of the ripening rate on coffee branches by using 3D images in outdoor environments. **Computers in Industry**, v. 99, p. 83–95, 1 ago. 2018.

REDMON, J. et al. **You only look once: Unified, real-time object detection**. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. **Anais**...IEEE Computer Society, 9 dez. 2016

REDMON, J.; FARHADI, A. YOLO9000: Better, faster, stronger. **Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017**, v. 2017- Janua, p. 6517–6525, 2017.

REDMON, J.; FARHADI, A. YOLO v.3. **Tech report**, p. 1–6, 2018.

ROY, P. et al. Vision-based preharvest yield mapping for apple orchards. **Computers and Electronics in Agriculture**, v. 164, p. 104897, 1 set. 2019.

SA, I. et al. Deepfruits: A fruit detection system using deep neural networks. **Sensors (Switzerland)**, v. 16, n. 8, 2016.

SANTINATO, F. et al. 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**, 2014.

SANTOS, F. L.; QUEIROZ, D. M.; VALENTE, D. S. M.; COELHO, A. L. F. Simulação do comportamento dinâmico do sistema fruto-pedúnculo do café empregando o método de elementos finitos. **Acta Scientiarum - Technology**, v. 37, n. 1, p. 11–17, 6 jan. 2015.

SARTORI, S.; FAVA, J. F. M.; DOMINGUES, E. L.; RIBEIRO FILHO, A. C.; SHIRAISI, L. E. Mapping the spatial variability of coffee yield with mechanical harvester. In: WORLD CONGRESS ON COMPUTERS IN AGRICULTURE AND NATURAL RESOURCES, 2002, Foz do Iguaçú. Anais… St. Joseph: ASAE, 2002. p. 196-205.

SONG, Y.; GLASBEY, C.A.; HORGAN, G.W.; POLDER, G.; DIELEMAN, J.A.; VAN DER HEIJDEN, G.W.A.M. Automatic fruit recognition and counting from multiple images. **Biosystems Engineering**, v. 118, n. 1, p. 203–215, 1 fev. 2014.

TU, S.; PANG, J.; LIU, H.; ZHUANG, N.; CHEN, Y.; ZHENG, C.; WAN, H.; XUEET, Y. Passion fruit detection and counting based on multiple scale faster R-CNN using RGB-D images. **Precision Agriculture**, vol.21, p.1072–1091, 2020.

VELLOSO, N. S.; MAGALHÃES, R.R.; SANTOS, F.L.; SANTOS, A.A.R. Modal properties of coffee plants via numerical simulation. **Computers and Electronics in Agriculture**, v. 175, p. 105552, 1 ago. 2020.

WANG, C.-Y. et al. CSPNet: A New Backbone that can Enhance Learning Capability of CNN. . nov.

WANG, Z.; WALSH, K.; KOIRALA, A. Mango fruit load estimation using a video based MangoYOLO—Kalman filter—hungarian algorithm method. **Sensors (Switzerland)**, v. 19, n. 12, 2 jun. 2019.

WU, D.; SUN, D. W. **C**olour measurements by computer vision for food quality control - A review. **Trends in Food Science and Technology**, vol. 29. p5-20, 2013.

YU, Y. et al. Fruit detection for strawberry harvesting robot in non-structural environment based on Mask-RCNN. **Computers and Electronics in Agriculture**, v. 163, p. 104846, 1 ago. 2019.

YUN, S. et al. CutMix: Regularization Strategy to Train Strong Classifiers with Localizable Features. **Proceedings of the IEEE International Conference on Computer Vision**, v. 2019- October, p. 6022–6031, 13 maio 2019.

ZHENG, Z.; WANG, P.; LIU, W.; LI, J.; YE, R.; REN, D. Distance-IoU Loss: Faster and Better Learning for Bounding Box Regression. arXiv, 19 nov. 2019.

5. GENERAL CONCLUSIONS

This thesis proposes a computer vision algorithm based on the YOLO neural networks architecture to detect and count coffee fruits on tree branches and during the mechanized harvest, regardless of the environmental conditions to which they may be found. In chapter 1, the detection and classification of coffee fruits on tree branches in different maturation stages using the models based on an image input resolution of 800x800 pixels and the network architectures of YOLOv3-tiny, YOLOv3, YOLOv4-tiny, and YOLOv4 scored a mean average precision (mAP) of 77.4%, 77.7%, 78.9%, and 81.2%, respectively. The YOLOv4 model showed more robustness for the defection of unripe fruits and for detecting fruits in images with a higher density of fruits.

In chapter 2, for the detection and counting of fruits during harvest, the YOLOv4 model scored a mAP of 83.5% for the image input resolution of 800x800 pixels. The yield map estimated from detections made by the computer vision model was able to explain 81% of the variance of the reference yield map. In contrast, in chapter 3, the YOLOv4 model scored an mAP of 91.8% for the detection and classification of coffee fruits in different maturation stages during harvest. The average precision for each class in the detection of unripe, ripe, and overripe coffee fruits was 93.0%, 91.9%, and 90.7%, respectively.

Monitoring the spatial variability of these attributes provides us with information that enables the introduction of precision agriculture techniques for the management of coffee crops. The study developed in this thesis can guide future research for coffee crops in the investigation of differences in fruits maturation stage along coffee lines, as in soil characteristics that can correlate to early or late maturation of coffee fruits. Monitoring such information in real-time can also help optimizing harvesting speed and the vibration of the harvester rods for selective harvest.