2023

Detection of Grape Clusters in Images using Convolutional Neural Network

Mohammad Osama Shahzad
National University of Sciences and Technology, Islamabad, Pakistan

Anas Bin Aqeel
National University of Sciences and Technology, Islamabad, Pakistan

Waqar Shahid Qureshi
Technological University Dublin, waqar.shahid@gmail.com

Funder: This research received no external funding.

Detection of Grape Clusters in Images using Convolutional Neural Network

Muhammad Osama Shahzad[1*], Anas Bin Aqeel[1], and Waqar Shahid Qureshi[1,2]
[1] Department of Mechatronics Engineering, National University of Sciences and Technology, H-12, Islamabad, Pakistan
[2] School of Computer Science, TU Dublin, Dublin 7, Ireland
*email: usamashahzad19@gmail.com

2023 International Conference on Robotics and Automation in Industry (ICRAI) | 978-1-6654-6472-7/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICRAI57502.2023.10089582

***Abstract*— Convolutional Neural Networks and Deep Learning have revolutionized every field since their inception. Agriculture has also been reaping the fruits of developments in mentioned fields. Technology is being revolutionized to increase yield, save water wastage, take care of diseased weeds, and also increase the profit of farmers. Grapes are among the highest profit-yielding and important fruit related to the juice industry. Pakistan being an agricultural country, can widely benefit by cultivating and improving grapes per hectare yield. The biggest challenge in harvesting grapes to date is to detect their cluster successfully; many approaches tend to answer this problem by harvest and sort technique where the foreign objects are separated later from grapes after harvesting them using an automatic harvester. Currently available systems are trained on data that is from developed or grape-producing countries, thus showing data biases when used at any new location thus it gives rise to a need of creating a dataset from scratch to verify the results of research. Grape is available in different sizes, colors, seed sizes, and shapes which makes its detection, through simple Computer vision, even more challenging. This research addresses this issue by bringing the solution to this problem by using CNN and Neural Networks using the newly created dataset from local farms as the other research and the methods used don't address issues faced locally by the farmers. YOLO has been selected to be trained on the locally collected dataset of grapes.**

Keywords—Object Recognition, Grapes, YOLO, Convolutional Neural Network, Deep Learning, Object Detection

I. INTRODUCTION

Grape is one of the most important fruit in the world with huge profitability and usage. In Botanical terms, a grape is considered a berry. It is one of the very few fruits that grow in a cluster of 15 to 300. Grapes' different colours include red, green, and purple; grapes can be seedless, with big seeds, and have a wide variety of flavors ranging from different degrees of sweet and sour. Grape juice is used in cooking to enhance umami. Grapes are used to extract fruit juice, and wine and are also consumed as toppings, jams, vinegar, grapeseed oil, and raisins. Grapes are 81% water and 16% carbohydrates, have negligible fats and a percentage of protein, and also dietary fiber, which is an important part of everyday diet. Grapes are a good source of vitamin C and K. Grapes are cultivated globally at 7 mil ha which makes them one of the leading fruit. Its total production in 2016 was 77.4 mil tons (valued at $68.3 billion) [1]. Red grapes are a major source of resveratrol. Resveratrol has chemo-preventive and therapeutic properties. It is useful in controlling diabetes and has been linked to reduced colon cancer [2].

Pakistan being an agriculture-based economy and a region suitable for grapes cultivation has a huge potential for grape production and not just earned by exports but can also use it to set up and develop its sister industries. By doing so, we will be bringing cash to the farmers which will result in further progression of the agriculture sector. Globally Pakistan is ranked 56th in terms of production and 96 in terms of exports of grapes [1]. There is a huge potential for it to increase its grape productivity by focusing on increasing its yield per hectare. Right now Pakistan's yield is just 37% of the average global yield per hectare [1] and the rate of increase of production in Pakistan is also much lower than the global average.

A lot of research is being carried out related to grapes, mainly to increase their productivity and enhance their taste. It's an ancient fruit and archaeological remains suggest that mankind started growing grapes as early as 6500 B.C. [2]. In agriculture, AI and Deep Learning is the forerunner and help scientists to tackle challenges of food storage, food production, and disease management. The population is expected to surge to 250 billion by 2050 and to meet their demands, 70% of the increase in food production is needed [3]. As grape is the third most valuable crop globally after potatoes and tomatoes [4], their demand and need are going to increase exponentially with time. Due to this factor, extensive research on this fruit is needed at the time and will prove to be fruitful.

YOLO object detection algorithm will be used to detect the individual clusters of grapes using bounding boxes and after successful object detection, more complex tasks will be carried out. The tasks may include improving the process of the fruit harvest, disease detection, yield estimation, anomaly detection, and targeted spraying; for such innovative tasks, a reliable algorithm is required. This study is to develop such state of an art algorithm that is reliable and can match the required standards.

The purpose of this study is to collect a dataset from scratch in the agricultural environment and then make the data useful by annotating it and making it go through a tough process of acceptability, so data doesn't have any bias in its core. In the process of annotation, the objects of interest were carefully identified and labeled. Then lastly, train a reliable algorithm to test it on test data and verify its performance. The algorithm acceptability rate should be satisfactory moreover the aim is also for it to be scalable and implementable in real-time which can only be achieved if the training is done with no biasness and has high precision and recall.

II. RELATED WORK

Machine Learning and Deep Learning are the developing fields of the present that are greatly impacting the surroundings directly and indirectly. It has changed the way we see things today and has affected every single part of the current daily lives of the masses.

978-1-6654-6472-7/23/$31.00 © 2023 IEEE

In the field of agriculture, ML (machine learning) and DL (deep learning) has brought a revolution. These are core technologies driving AI-based (artificial intelligent based) robots, autonomous spray drones, mapping, and autonomous harvesting. Deep learning-based models are predicting the crop maturity index, health, and yield, also a huge amount of agricultural data is being gathered to further improve the field of agriculture. AI has dramatically increased the crop yield and profit margin of the farmers. Berenstein found out that usage of sprays can be reduced by 30% if we detect and spray 90% of the clusters of grapes [5]. This will not only save a huge amount of resources but increase farmers' profit margins and lessen the pollution that is being caused owing to the excessive usage of pesticides.

Zabawa & Kicherer [6] worked on the detection of single grapevine berries. They annotated 32 images into three classes naming berry, edges, and background. Every berry was surrounded by an edge while the remaining image was termed as background. They were able to achieve accuracy ranging from 84% to 95%; however, 28% of detected berries were False positive (Type II error); the problem was tackled by incorporating different methods including image filtering.

Aquino & Millan's [7] work in grape yield prediction is also noteworthy and mentionable. They integrated a camera with a vehicle specifically designed to be used on a grape farm. The camera was triggered by the movement of the vehicle. The data was collected during nighttime using LED as the artificial light source. They were able to limit the average square error to 0.16 kg per vine and RMSE was 0.48kg for an image segment consisting of three vines. Another such work was by Ralph Linker & Kelman [8] who also worked on yield estimation, but of apples, in the nighttime. They used the specular light (light reflected from the apple surface) during nighttime to detect the fruit, favorable results for which were achieved by aligning the camera with the light source. Nellithimaru et al. [16] presented a FAST R-CNN-based model of grape counting and 3D reconstruction of a vineyard algorithm that used camera equipment with an air blower to accurately model plants by hindering leaves movement from the object of interest. Many others devised models for grape detection and yield estimation exist, but few notable works include Font et al. [17] work on yield estimation using artificial illumination at night time, Huerta et al. [18] work on creating a 3D model from images and using it to estimate yield and lastly Nuske et al. [19] work on the berry detection during night time. Most of these works cover the detection problem during the night or training the dataset during the daytime is less. Nevertheless, these methods provide a non-invasive and automated way of grape detection and are being used in different applications from spraying to yield estimation, etc.

Object detection in orchards is not a new thing and is spread across many fruits and crops such as Bargoti et al. [9] worked on the detection and yield estimation of apples, Huang et al. [10] worked on citrus detection system using a mobile platform, Lim et al. [11] worked on detecting the kiwi fruit flowers in orchard environment, Borianne et al. [12] worked on the detection of mangoes detection and detection of immature peaches by Kurtulmus et al. [13] are only a few to be named. Object detection is generally performed using CNN or DNN techniques such as Fast R-CNN [14], RNN, and YOLO [15]. Though a lot of work is being carried out in the fields of object detection and yield estimation no reliable method has yet been discovered with enough accuracy. Generally, the yield estimation algorithms work fine against the average grape cluster but show bias when it encounters a weak cluster. Moreover, the scarcity of goods and enough data is another reason for less efficient or overfitted models.

III. Materials & Methods

This section contains all the necessary information regarding how the model was created and what steps were followed to train it; the factors that influenced training and how we overcame them. The data collection was the most important step that required physical presence in fields, followed by pre-processing and data annotation after which the object detection algorithm was trained on the cleaned and labeled data. The results of the trained object detection algorithm were tested on the test dataset and a wide array of performance parameters were calculated using the test dataset. Different batches of data were trained and tested to verify the results.

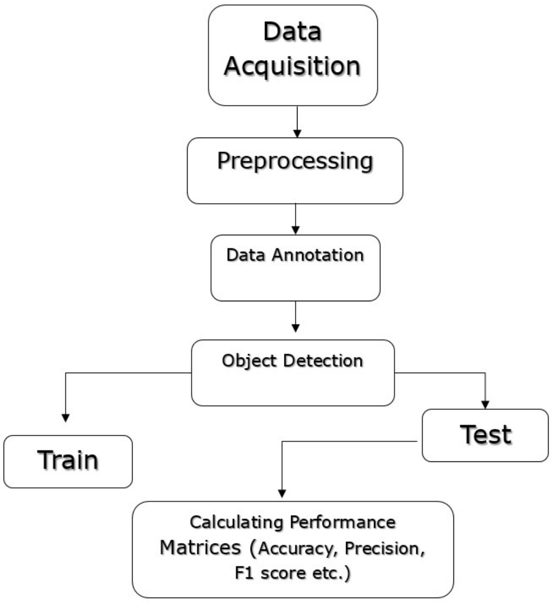
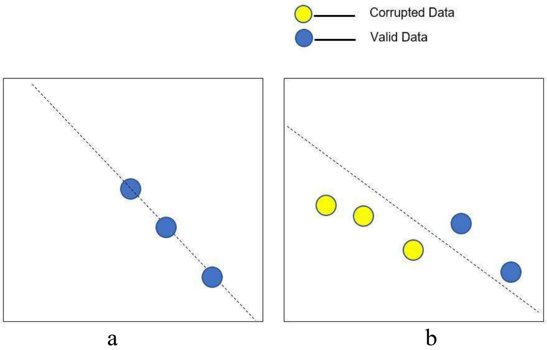

Data Acquisition
Preprocessing
Data Annotation
Object Detection
Train
Test
Calculating Performance Matrices (Accuracy, Precision, F1 score etc.)

Fig. 1. Block diagram of the steps in the methodology

A. Collecting Data

In any of the Deep Learning studies and experiments, the role of data is as important as the algorithm itself because the data is what trains the algorithm. If the data is not taken care of and contains any kind of business, it will also be transferred to the model and the model will reflect it while testing or after being deployed. It might compromise the data's effectiveness to perform in unfamiliar conditions. Thus, it is essential to collect data with set parameters and strictly follow those parameters. It has been found useful to go through the collected data to gauge its usability as the training dataset.



(a) Model converges cleanly with clean data (b) Inclusion of bad data affects the performance of the model
Fig. 2. Impact of Bad Data on Deep Learning Model

Data not fulfilling those parameters must be filtered out of the training data. If there is corrupted data involved in training, it will influence and affect the model performance and will lead to more False Positive detections, which is also known as Type II error, and will also reduce the model True Positive accuracy by increasing the value of mean error and the model will not converge accurately (as shown in Fig. 1).

In this study, a comprehensive and dedicated dataset was created from scratch for which, two different sensors were used. Data were captured at multiple angles and different times of the day so that images with diverse lightning conditions are part of the dataset. A camera with auto-focus and manual focus modes was used with multiple ISO levels and color values. Images were taken at different distances and conditions like multiple cluster images, single cluster images etc. This will make sure that our dataset has contrasting images of different types and cover multiple details of the farm. It will make the dataset diverse and bias free which will make the model converge faster.

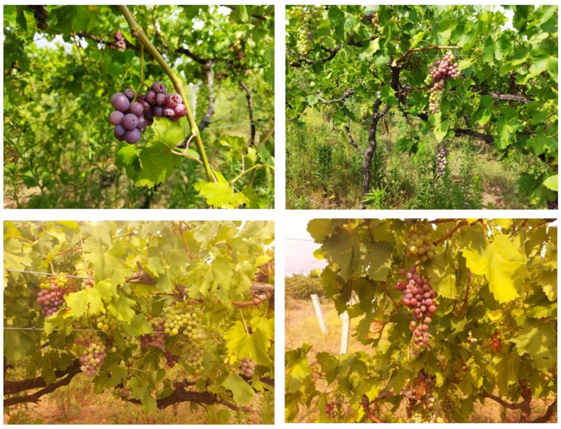


Fig. 3. Images from the newly created grape dataset

B. *Pre-processing*

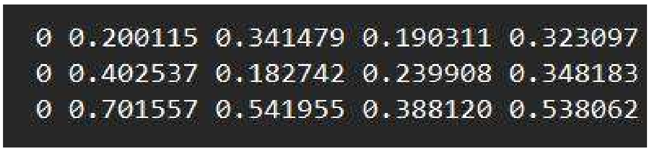
The images are needed to be passed through certain steps to reduce abnormalities and ensure the availability of only high-quality data to the training algorithm. For this, we reduced the size of the image, increased the contrast of images, and also discarded the unsuitable images. Image size was reduced using batch normalization which gives us the ability to repeat a certain process on multiple images.

This process needed to individually view images keenly and single out bad images [20]. This process is important as this will filter out the biases in data and bad data from compromising the efficiency of the training model thus making the model robust and even more intelligent and the results will be showcased as improvement in its performance parameters.

C. *Data Annotation*

Data were annotated using a toolbox known as LabelImg. The fruits were labeled and as this is a single-class classification problem, the label used was “Grape”. The format for data annotation was YOLO format which is suitable to label data that’ll be fed to the YOLO algorithm. Data Annotation is a very important step as this process tells the model about the object of interest and annotating faulty data may result in poor performance of the algorithm [19].

Annotation being saved in YOLO format has five parameters and is saved as a *.txt* file. The first numeric value represents the class of the object. The second and third represents the center of the bounding box in term of x and y coordinates respectively while the fourth and fifth value represents the length of x-axis and y-axis respectively. These values are normalized.



```
0 0.200115 0.341479 0.190311 0.323097
0 0.402537 0.182742 0.239908 0.348183
0 0.701557 0.541955 0.388120 0.538062
```

Fig. 4. A File showing Annotations in YOLO format

D. *Training*

The models were trained using Google Colab, remotely. The dataset, along with annotations, was uploaded on Google Drive which not only made the training process faster but also made it independent of a particular hardware or workstation.

Two different YOLO v3-based models were trained with varying numbers of images and sizes of datasets to verify the results and models’ performance. The first dataset contained 838 images while the second dataset carried 1172 images. Datasets were divided into training and test datasets with a ratio of 80/20 [20] where 80% forming the training dataset and 20% being the test. Colab GPU was used as virtual training hardware, which is a great substitute for normal mid-range GPUs.

The first model used the approach of fine-tuning where we used pre-trained weights to start training. The second model used the approach of transfer learning where we took advantage of results learned from the first model and used the same weights as initial weights. The first model was trained in one go and it took 17 hours and 48 minutes of training for it to converge; it converged at 4000 epochs (as shown in Fig.. 5). The second model was trained in batches and multiple intervals by pausing the training and restarting the training using the weights of the previous session; it took 28 hours for it to be trained and the model converged at 4100 epochs.

978-1-6654-6472-7/23/$31.00 © 2023 IEEE

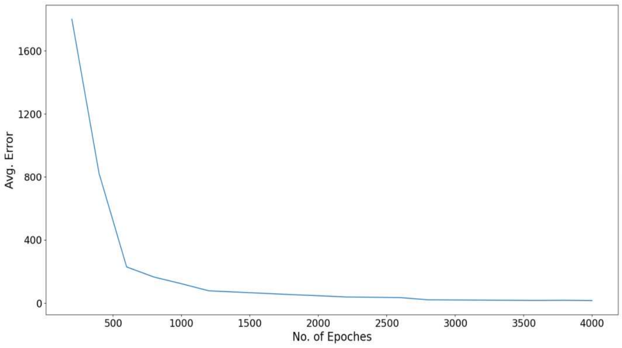


Fig. 5. Graph Showcasing convergence of Model 1

E. Testing

Testing is the essence of the whole procedure where the conclusion is drawn and results are calculated. The test dataset is fed to the weights of trained models and their different types of set performance parameters are calculated. If the model outperforms the performance parameters, then the model is considered successful else it is re-designed and optimized. Models are tested using multiple world-like different situations and the model is considered good only if it can perform better in the unseen environment and new data.

TABLE 1: CONFUSION MATRIX OF MODEL 2

True Positives	False Positives
214	19
05	N/A
False Negative	True Negative

We downloaded the trained weights after the training process has been completed and testing was done completely offline (result shown in Fig. 6). In testing, the test dataset was fed to the model that consisted of images carefully selected to be able to show contrast properties and unseen data selected randomly. Bounding boxes were drawn in this method over the object of interest. Interaction over union was calculated on the ground truth of the bounding box over the box detected while the accuracy, precision, and recall are all calculated from the confusion matrix (shown in Tables 1 and 2). The F1 score was calculated from the value of precision and recall.

A separate script was written for testing where the inputs are the weight of the trained model and the images or the frame of the video. The algorithm outputs the numerical data in YOLO format (shown in Fig. 4) which is then used to draw on the input image in form of a bounding box (a few of the instances are shown in Fig. 7).

IV. Results and Discussion

The process of result declaration is most important as the conclusion of the whole research is based on it. Two different datasets were trained using YOLO v3 which is an improvement over the original YOLO [21]. The results were tested on the trained model and the truth table was calculated for all the models respectively. The truth table consist of the detected, wrongly detected, and not detected instances.

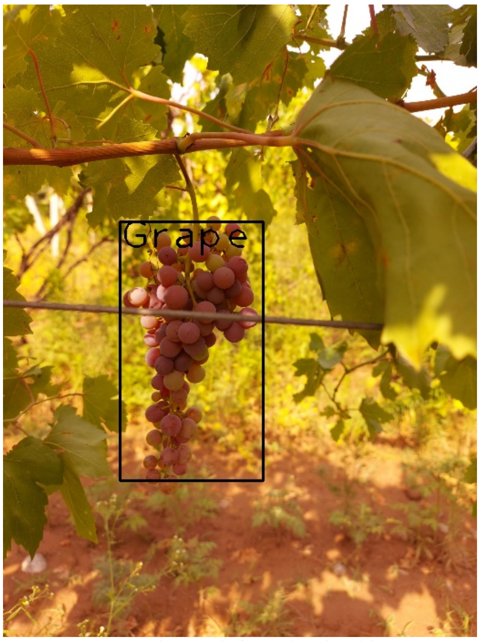


Fig. 6. The result from the algorithm

The confusion matrix (shown in Table 1) shows the values calculated from the test dataset of Model 1. It constituted a total of 150 instances. Among those instances, 136 were True Positive while 11 were wrongly detected as positive (thus False Positive) and lastly, 3 instances should have been detected as positive but not detected so (False Negative). False Positives and False Negatives represents Type-I and Type-II errors respectively.

TABLE 2: CONFUSION MATRIX OF MODEL 1

True Positives	False Positives
136	11
03	N/A
False Negative	True Negative

The result of Model 1 is quite satisfying but to reinforce our findings and be sure of the data performing efficiently without showing any bias or overfitting, we ought to test it more with models trained on even larger datasets. We tested another model with a larger and better dataset. Its test dataset consisted of 238 instances and the training dataset was also considerably larger thus this dataset was able to learn more features due to more and better-labeled data. Among all the instances, 214 of the instances of Model 2 were rightly detected as labeled and thus are placed in the "True Positive" block (as shown in Table 2). 19 instances were False Positive and only 5 instances were detected as False Negatives. Both Models 1 & 2 don't have any value in True Negative as that block applies only to multiclass classification and not to object detection problems. Model 2 results complement the result of Model 1 and verify the efficiency and performance of models as well as data.

978-1-6654-6472-7/23/$31.00 © 2023 IEEE

TABLE 3: PERFORMANCE PARAMETERS OF MODEL 1 & MODEL 2

Model	Accuracy	Recall	Precision	F1 Value
Model 1	90.66	97.84	92.51	0.95
Model 2	88.74	98.02	90.3	0.94

Table 3 concludes the result as it shows all the performance parameters calculated by the confusion matrixes of both models. We compared the precision, recall, accuracy, and F1 score of both models. The accuracy of Model 2 is slightly lower (88.74%) than that of Model 1 (90.66%). The precision and recall are almost the same and this is also affirmed by the F1 score which is 0.95 for Model 1 and 0.94 for Model 2. By these results, we conclude that both the models complement the result of each other, and training results are satisfactory for the model to be used in a real-time grape farm environment for grape bunch detection.

V. CONCLUSION

This research was an extensive amalgam of on-groundwork, data handling & labeling, and training DNN models. A few models were created after making a viable and excellent dataset that can comprehend contrasted information. Datasets were trained using pre-trained weights for faster convergence. This research proposed a model that can work under multiple lightning conditions at many different angles. The proposed models have been trained on a large dataset that consisted of data taken at a grape farm using multiple camera sensors at different angles and distances which is not a case in much previous research [22]. Multiple models were trained using different features of the dataset and results were calculated using different parameters rather than only relying on accuracy. It was made sure that the model doesn't overfit or underfit; also a balance between precision and recall was achieved which leads to a high F1 score of 0.95. The IOU of these models also falls under 'good' criteria, its value is ranging between 0.854 and 0.865. The models were trained using google collab and showed promising and intended results. The models give high accuracy (shown in Table 3) and both models complement each other despite having intersections in their datasets. Models were fast and trained on an algorithm that can process data in real-time which makes the models able to be implemented for real-time calculation in a real-world environment. In the future of this study, we aim to further increase the detection rate and performance parameters of the models and train a few other state-of-the-art models, like SSD or RNN [23] to compare the results and find the best model for such a problem.

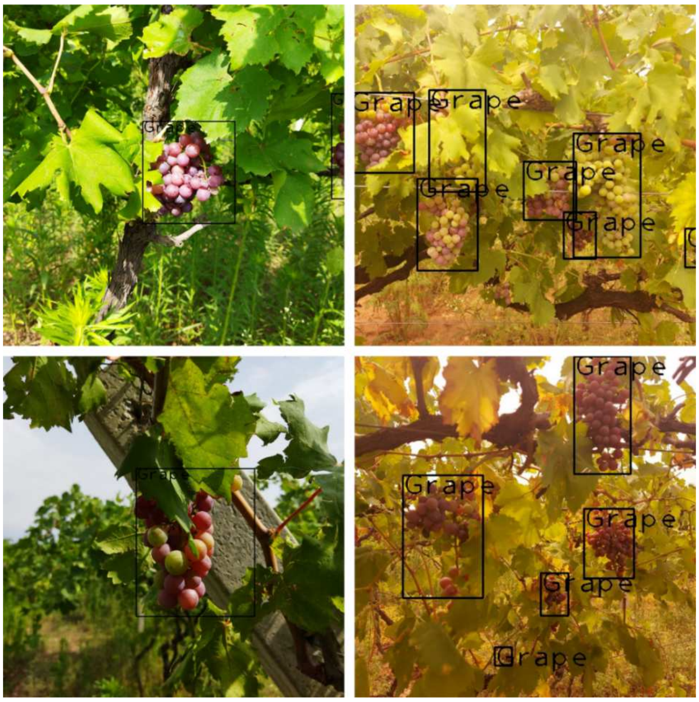


Fig. 7. Output Results after bounding box being drawn

REFERENCES

[1] D. Trinklein, "Grapes: A Brief History," University of Missouri, Missouri, 2013.

[2] M. Imran, A. Rauf, A. Imran, M. Nadeem, Z. Ahmad, M. Atif, M. Awais, M. Sami, M. Imran, Z. Fatima, and A. B. Waqar, "Health Benefits of Grapes Polyphenols," Journal of Environmental and Agricultural Sciences, vol. 10, pp. 40-51, 2017.

[3] N. Clara Eli-Chukwu, "Applications of Artificial Intelligence in Agriculture: A Review," Engineering, Technology & Applied Science Research, vol. 9, no. 4, pp. 4377-4383, 2019.

[4] O. Sambucci, Julian M. Alston, "Grape in the World Economy," in The Grape Genome, 2013.

[5] O. B. Shahar, and A. Shapiro, Y.Edan, and Ron Berenstein, "Grape clusters and foliage detection algorithms," Intel Serv Robotics, no. 3, pp. 233-243, 2010.

[6] A. Kicherer, L. Klingbeil, A. Milioto, and Laura Zabawa, "Detection of Single Grapevine Berries in Images Using Fully Convolutional Neural Networks," in CVF Conference on Computer Vision and Pattern Recognition Workshops, IEEE, 2019.

[7] B. Millan, Maria-Paz Diago, J. Tradaguila, and Arturo Aquino, "Automated early yield prediction in vineyards from the on-the-go image," Computers and Electronics in Agriculture, vol. 144, pp. 26-36, 2018.

[8] E. Kelman, and Raphael Linker, "Apple detection in nighttime tree images using the geometry of light," Computers and Electronics in Agriculture, vol. 114, pp. 154-162, 2015.

[9] James P. Underwood, and S. Bargoti, "Image Segmentation for Fruit Detection and Yield Estimation in Apple Orchards," Journal of Field Robotics, 2016.

[10] T. Huang, Z. Li, S. Liu, T. Hong, and H. Huang, "Design of Citrus Fruit Detection System Based on Mobile Platform and Edge Computer Device," Sensors, 2021.

[11] H. S. Ahn, M. Nejati, J. Bell, H. Williams, B. A. MacDonald, and J. Y. Lim, "Deep Neural Network Based Real-time Kiwi Fruit Flower Detection in an Orchard Environment," in Australasian conference on robotics and automation, 2019.

[12] F. Borne, J. Sarron, E. Faye, and P. Borianne, "Deep Mangoes: from fruit detection to cultivar identification in color images of mango trees," in International Conference on Digital Image and Signal Processing, 2019.

[13] A. Vardar, W. S. Lee, and F. Kurtulmus, "A. Immature peach detection in color images acquired in natural illumination conditions using statistical classifiers and neural network," Precision Agri, vol. 15, pp. 57-59, 2014.

[14] K. Hee, R. Girshick, J. Sun, and S. Ren, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, no. 6, pp. 1137-1149, 2017.

[15] S. Divvala, R. Girshick, A. Farhadi, and J. Redmon, "You Only Look Once: Unified, Real-Time Object Detection," in IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.

[16] Anjana K. Nellithimaru, and George A. Kantor, "ROLS : Robust Object-Level SLAM for Grape Counting," in 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2019.

[17] M. Tresanchez, D. Martinez, J. Moreno, E. Clotet and D. Font, "Vineyard Yield Estimation Based on the Analysis of High Resolution Images Obtained with Artificial Illumination at Night," Sensors, vol. 15, no. 4, pp. 8284-8301, 2015.

[18] Diego G. Aguilera, Pablo R. Gonzalvez, David H. Lopez, Mónica H. Huerta, "Vineyard yield estimation by automatic 3D bunch modelling in field conditions," Computers and Electronics in Agriculture, vol. 110, pp. 17-26, 2015.

[19] Y. Bengigo, A. Courville, and I. Goodfellow, Deep Learning, MIT Press, 2016.

[20] F. Chollet, Deep learning with python, New York: Manning Publication, 2017.

[21] A. Farhadi, and J. Redmone, "YOLOv3: An Incremental Improvement," in Computer Vision & Pattern Recognation (CVPR), 2018.

[22] Thiago T. Santos, Leonardo L. de Souza, Andreza A. doc Santos, and S. Avila, "Grape detection, segmentation, and tracking using deep neural networks and three-dimensional association," Computer and Electronics in Agriculture, vol. 170, 2020.

[23] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, Cheng Y. Fu, and Alexander C. Berg, "SSD: Single Shot MultiBox Detector," in Springer, 2016.

978-1-6654-6472-7/23/$31.00 © 2023 IEEE