Uluslararası İleri Doğa Bilimleri ve Mühendislik Araştırmaları Dergisi
Sayı 7, S. 39-45, 7, 2023
© Telif hakkı IJANSER'e aittir
Araştırma Makalesi

IJANSER

https://as-proceeding.com/index.php/ijanser
ISSN: 2980-0811

International Journal of Advanced Natural Sciences and Engineering Researches
Volume 7, pp. 39-45, 7, 2023
Copyright © 2023 IJANSER
Research Article

Kiwi Fruit Detection with Deep Learning Methods

Esra Gungor Ulutas[1,*] and Cemil ALTIN [1]

[1]*Department of Electrical and Electronics Engineering, Yozgat Bozok University, Turkey*
[1]*Department of Electrical and Electronics Engineering, Yozgat Bozok University, Turkey*

[*]*(esra.ulutas@bozok.edu.tr) Email of the corresponding author*

(Received: 28 July 2023, Accepted: 09 August 2023)

(1st International Conference on Modern and Advanced Research ICMAR 2023, July 29-31, 2023)

.**ATIF/REFERENCE:** Ulutas, E. G. & Altın, C. (2023). Kiwi Fruit Detection with Deep Learning Methods. *International Journal of Advanced Natural Sciences and Engineering Researches*, 7(7), 39-45.

Abstract – The automatic detection of kiwifruit in orchards is a challenging task due to the similarity between the fruit and the complex backgrounds formed by branches and stems. Moreover, the traditional method of hand-picking kiwifruit heavily relies on human labor and affects the overall yield. This study focuses on the fast and accurate detection of kiwifruit in natural orchard environments, which is crucial for yield estimation and cost reduction. Two deep learning methods, Faster Region-based Convolutional Neural Network (Faster R-CNN) and Mask Region-based Convolutional Neural Network (Mask R-CNN), are utilized for kiwifruit detection, and their results are compared. The study begins with obtaining images of kiwi trees from the Güngör farm in Samsun Çarşamba and creating an original dataset. Preprocessing techniques are applied to improve the dataset, followed by detection using the Faster R-CNN method. Different pre-trained architectures like SqueezeNet and MobileNetV3 are used, achieving average precision (mAP) values of 87.4% and 88.8%, respectively. In the second part of the study, kiwifruit images are processed using the ResNet50-based Mask R-CNN method, which achieves a higher mAP value of 98.48%. The experimental results demonstrate the applicability and effectiveness of the proposed deep learning models for real-time kiwifruit detection in orchards. Accurate kiwifruit detection allows farmers to optimize yield prediction, reduce costs, and improve productivity. The application of Faster R-CNN and Mask R-CNN in this study showcases their potential for enhancing the efficiency and accuracy of kiwifruit detection in orchard environments.

Keywords – Faster R-CNN, Mask R-CNN, Fruit Detection, Kiwifruit

I. Introduction

The kiwifruit, also known as "Yang Tao" and alternatively referred to as "Chinese gooseberry" and "monkey peach" in English, is a fruit. A kiwifruit variety successfully cultivated by New Zealander Hayward Wright in 1927 is now known as "Hayward." By the 1960s, the "Hayward" kiwifruit had become a standard variety, accounting for approximately 90% of global kiwifruit production [1]. In Turkey, kiwifruit is also grown as an agriculturally important product, and kiwifruit orchards hold a significant place in the agricultural sector.

Kiwifruit production activities began in Turkey in 1988. Initially, adaptation and demonstration gardens were established by the Atatürk Central Horticultural Research Institute in Yalova, and studies determined that the Black Sea, Marmara, and Aegean coastal regions were suitable for kiwifruit cultivation. Among these regions, the Central and Eastern Black Sea region were found to be more suitable for the ecological requirements of the plant and economically viable for kiwifruit cultivation [2]. Nowadays, kiwifruit is grown in the Mediterranean, Aegean, Black Sea, and Marmara regions, with a total of 29,902 hectares yielding 61,920 tons of produce [2].

Applications of precision agriculture and deep learning methods accelerate developments in the agricultural sector, enabling more efficient, sustainable, and profitable harvests. The use of these technologies provides farmers with a competitive advantage and supports the growth of the respective agricultural products in the sector. Especially with the use of technologies such as sensors, drones, and satellite imaging, monitoring and managing critical parameters like plant health, soil moisture levels, and disease detection in various plant species and orchards becomes more accessible [3]–[6]. The utilization of modern technologies like precision agriculture and deep learning methods is increasingly growing in kiwifruit cultivation [7]–[9]. In recent years, there has been an increasing interest in using deep learning techniques for yield prediction in agriculture [10], [11]. One such study is conducted by Zhou et al., titled "Real-Time Kiwifruit Detection in Fruit Orchards using Deep Learning on Android Smartphones for Yield Estimation" [12]. In this study, the authors proposed a method for real-time kiwifruit detection in fruit orchards using deep learning on Android smartphones. The method accurately detected and counted kiwifruit in images captured by the smartphone camera, enabling real-time yield estimation. The findings revealed that MobileNetV2, quantized MobileNetV2, InceptionV3, and quantized InceptionV3 achieved true detection rates (TDR) of 90.8%, 89.7%, 87.6%, and 72.8%, respectively.

Fu et al. proposed a method for recognizing clustered kiwifruit in field images using convolutional neural networks [13]. The aim of the method was to improve kiwifruit detection and recognition accuracy in field images. The CNN architecture for kiwifruit training was optimized using batch normalization, which reduced training time and improved convergence. The proposed model achieved an overall recognition rate of 89.29%, and it takes an average of only 0.27 seconds to recognize a fruit. In their other proposed work, Fu et al. developed a kiwifruit detection system based on deep convolutional neural networks, achieving a recognition rate of 92.3% per image with a processing time of 0.274 seconds [14]. This suggested technique can detect individual kiwifruits in clusters, promising accurate yield mapping and multi-armed robotic harvesting. Song et al. developed a machine vision system for kiwifruit harvesting in China, using the Faster R-CNN model implemented by VGG16 [15]. The proposed method achieved higher detection mean average precision than ZFNet, providing robust support for a harvest robot capable of working throughout the dense season. The VGG16 model attained an average accuracy of 87.61% and effectively detected kiwifruit images collected under various timing and lighting conditions.

Liu et al. proposed a novel method for fruit detection utilizing RGB-D sensors and deep convolutional neural networks, which demonstrated improved accuracy, speed, and reliability potential [16]. The method involved aligning and fusing RGB and NIR images using two different fusion techniques, resulting in an average precision of up to 90.7%. Yu et al. utilized a Mask R-CNN model with ResNet50 and FPN architecture for fruit detection in a strawberry harvesting robot, achieving high precision and recall rates [30]. The proposed method demonstrated enhanced universality and robustness in a non-structured environment, particularly for overlapping and occluded fruits under varying lighting conditions. In another study, single fruits in images were detected using multi-modal input data [31]. By introducing a deep learning approach called "Deep Orange" and incorporating HSV (Hue, Saturation, Value) data, the accuracy for segmenting fruits into parts increased from 0.8947 to 0.9753 when using only RGB data. The primary aim of this study is to investigate the effectiveness and applicability of deep learning techniques for kiwifruit detection. The investigation employs two different deep learning methods: faster region-based convolutional neural network (R-CNN) and mask region-based convolutional neural network (Mask R-CNN). The selection of these methods is based on the unique advantages of each model. Specifically, Faster R-CNN provides a fast and accurate approach for detecting and classifying candidate regions, whereas the more advanced

Mask R-CNN detects and classifies candidate regions at the pixel level. Both of these models are widely used and validated in the object recognition and segmentation domains. Their particular success in recognizing objects in images with complex backgrounds makes them suitable choices for kiwifruit detection.

In terms of implementation, the Faster R-CNN method uses network architectures such as SqueezeNet and MobileNet, whereas the Mask R-CNN method for kiwifruit detection is based on the ResNet50 backbone. The different backbone architectures used (MobileNetV3, SqueezeNet, and ResNet-50) contributed to the diversity of this study and demonstrated their efficacy across different application scenarios and datasets. The dataset used in the study, deep learning techniques, and evaluation parameters are briefly explained in that order.

II. Materials and Method

A. Dataset

The image data was collected by a smartphone (a Samsung A51) under different lighting conditions in a fruit orchard. The Samsung A51 smartphone is equipped with a primary camera with a 48-megapixel sensor and f/2.0 aperture, an ultra-wide-angle camera with a 12-megapixel sensor and f/2.2 aperture, a depth camera with a 5-megapixel sensor and f/2.2 aperture, and a macro camera with a 5-megapixel sensor and f/2.4 aperture. These systems can record 4K videos at a frame rate of 30 frames per second.

The smartphone was held by an experimenter using a selfie stick, and the kiwifruit's shadow was positioned approximately 1 meter below. The smartphone's rear camera was fixed on the selfie stick, facing towards the canopy, to capture the images. RGB (Red, Green, and Blue) images were taken during the harvest season in August 2021 in Samsun (Çarşamba Ovası, Güngör Çiftliği). Table 1 shows the number of tags in the dataset. The dataset contains a total of 66 original images, as shown in Figure 1.

Table 1. Dataset's total number of tags

Models	Backbone	Number of tags
Mask R-CNN	Resnet50	5361 (polygon)
Faster R-CNN	SqueezeNet	2925 (rectangle)
	MobileNetV3	

Fig. 1 Samples of kiwi images found in the dataset

B. Image preprocessing

In computer vision, preprocessing plays a crucial role in increasing the accuracy and reliability of deep learning models. It involves applying various techniques to the input data before feeding it to the models to enhance data quality and extract relevant features. Effective preprocessing can significantly impact the performance and results of recognition tasks. In this study, the kiwi images obtained were resized to 512x512 dimensions during the preprocessing step. The images in the dataset were labeled through the "V7labs" platform.

C. Object Detection

Object detection is a powerful deep learning method used to identify and locate objects in images and video sequences [17]. It allows for the counting of objects in an image and provides precise information about their locations using bounding boxes and masks. In this study, the obtained images for kiwi detection are analyzed. Object detection algorithms, especially Faster R-CNN and Mask R-CNN, are utilized to accurately identify and locate kiwi fruits in the images. The goal is to detect kiwi fruits in the images with precision and reliability using these object detection techniques.

D. Faster R-CNN

The Faster R-CNN model has revolutionized the field of object detection by combining the strengths of the Region Proposal Network (RPN) and Fast R-CNN models [18], [19]. This innovative approach introduced a direct connection of the RPN to the Fast R-CNN architecture by incorporating RPN into the sampling layer. The integration of RPN into the Fast R-CNN framework has significantly improved object detection tasks. Before the development of Faster R-CNN, the Fast R-CNN model required a considerable amount of time to find all candidate boxes, posing challenges for efficient object detection. However, the Faster R-CNN method has greatly improved the speed and efficiency of object detection. By integrating RPN, the process of detecting candidate regions has been accelerated, leading to faster and more accurate detection.

E. Mask R-CNN

The Mask R-CNN framework is built upon the Faster R-CNN object detection model and has gained significant popularity in the field of object detection [20]. It offers an additional instance segmentation task that allows precise pixel-level identification of objects within an image. Mask R-CNN follows a two-stage architecture. In the first stage, a Region Proposal Network (RPN) is used to generate region proposals for objects present in the image [19]. The RPN utilizes a convolutional neural network (CNN) to propose regions that potentially contain an object. Candidate objects are extracted from these proposed regions, and the bounding boxes are classified.

In the second stage, features are extracted from the candidate boxes using a technique called Region of Interest (ROI) pooling. The RPN produces two separate outputs: the class of the object to which the candidate belongs and the size and position of the bounding box surrounding the object. Due to the variable bounding boxes generated by the RPN, handling images with different input sizes can be challenging. To overcome this, a fixed input size is determined, and ROI pooling is used to align the ROI features to a fixed size.

III. Results

This section presents the experimental results of the proposed methods for kiwifruit detection. In this study, kiwifruits were detected using two different deep learning methods: Faster R-CNN and Mask R-CNN. These two methods employ different approaches and architectures for kiwifruit detection.

The server's powerful hardware configuration, including a large memory capacity and a high-performance processor, along with the advanced NVIDIA GeForce RTX 3090 Ti graphics card, provided the computational capabilities necessary for the efficient execution of the deep learning algorithms. Python 3, along with the Jupyter Notebook framework, facilitated the implementation and analysis of the experimental results. These resources enabled the accurate and efficient detection of kiwifruits using the Faster R-CNN and Mask R-CNN methods in the study.

Faster R CNN model

In this study, the results are given comparatively by using SqueezeNet and MobileNetV3 backbones in the Faster R-CNN model for kiwi fruit detection. Faster R-CNN model configuration is given in Table 2. Kiwi fruit detection with Faster R-CNN (SqueezeNet and MobileNetV3) results are given in Figs. 2 and 3, respectively. The mAP and training loss graphs are given in Figs. 4–7, respectively.

Table 2. Faster R-CNN model configuration

Parameters	Value	Value
Num workers	4	4
Momentum	0.9	0.9
Nesterov	True	True
Learning rate	0.00001	0.00001
Epoch	500	500
Optimizer	SGD	SGD
Batch Size	2	2
Backbone	squeezenet	mobilenetv3
İmg_size	640	640
Num_classes	1+1 (kiwi) + (__background__)	1+1 (kiwi) + (__background__)
device	cuda	cuda
no_mosaic	True	True
Use_train_aug	False	False
cosine_annealing	False	False
Weights	Null	Null
Resume_training	False	False

Fig. 2 Faster RCNN with SqueezeNet backbone model kiwi fruit detection results

Fig. 3 Faster R-CNN with MobileNetV3 backbone model kiwi fruit detection results

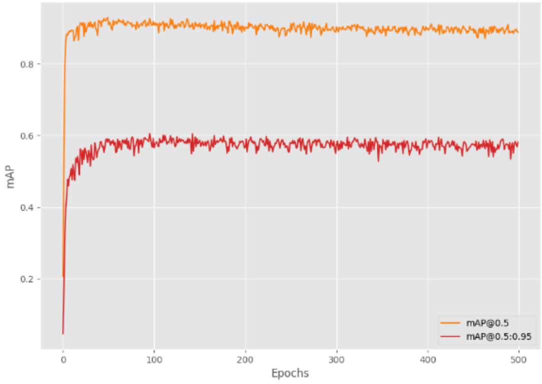

Fig. 4 Faster R-CNN with SqueezeNet backbone model mAP results

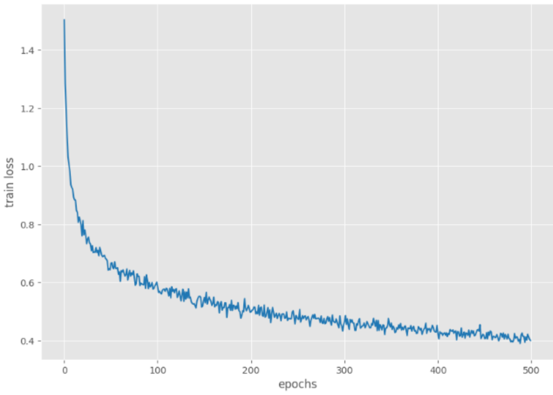

Fig. 5 Faster R-CNN with MobileNetV3 backbone model training loss graph

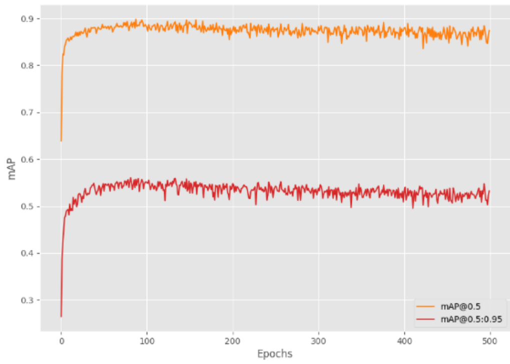

Fig. 6 Faster R-CNN with MobileNetV3 backbone model mAP results

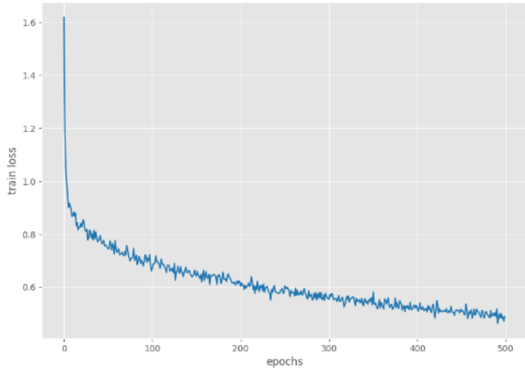

Fig. 7 Faster R-CNN with MobileNetV3 backbone model training loss plot

Mask R CNN model

The backbone of the network trained in this section is ResNet50. In the training phase, a total of 500 iterations are performed, and the initial learning rate is set at 0.0001. Mask R-CNN model configuration parameters are given in Table 2. Mask R-CNN model kiwi fruit detection results are given in Fig. 8.

Table 2 Mask R-CNN with ResNet50 backbone model configuration

Parametre	Value
Images_Per_Gpu	1
Steps_Per_Epoch	100
Valıdatıon_Steps	100
Backbone	ResNet50
Num_Classes	1 + 1
Image_Mın_Dım	2048
Image_Max_Dım	2048
Learnıng_Rate	0.0001
Learnıng_Momentum	0.9

Fig. 8 Mask R-CNN model kiwi fruit detection results

The Mask R-CNN model successfully performed the object detection task using the ResNet-50 backbone structure, and the mAP (Mean Average

Precision) value it obtained was measured at 98.48%. This high mAP value shows that the model detects objects quite successfully, and the object predictions are in high agreement with the real labels. Mask R-CNN with ResNet50 backbone model loss graph is given in Fig. 9. ResNet-50 is an architecture based on the backbone structure of deep neural networks and is frequently used, especially in the field of computer vision.

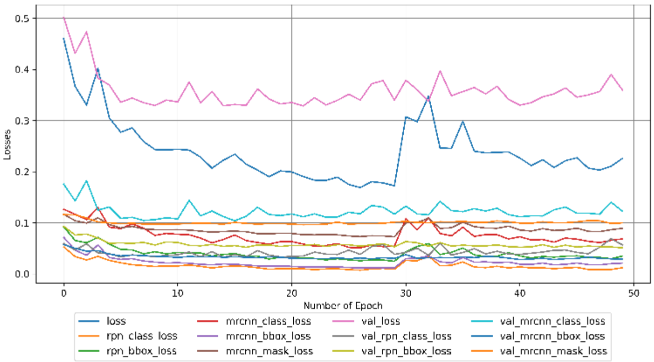

Fig. 9 Mask R-CNN with ResNet50 backbone model loss graph

Table 3 Mask R-CNN and Faster R-CNN mAP results

Model	Backbone	mAP (%)
Mask R-CNN	ResNet-50	**98.48**
Faster R-CNN	MobileNetV3	88.8
	Squeezenet	**87.4**

According to Table 3, the performance of three different object detection models was evaluated using the mAP (Mean Average Precision) metric: The Mask R-CNN model uses the ResNet-50 backbone structure, and the mAP value it obtained was determined to be 98.48%. This result shows that the model detects objects quite successfully, and the object predictions are in high agreement with the real labels. A high mAP value indicates that the model successfully distinguishes various object classes. The Faster R-CNN model achieved 88.8% mAP using the MobileNetV3 backbone structure. This value indicates that the model has good performance in the object detection task but has a slightly lower level of accuracy compared to ResNet-50. The Faster R-CNN model using the SqueezeNet backbone structure achieved a mAP value of 87.4%. This result shows that the model has a slightly higher level of accuracy than the MobileNetV3 backbone structure.

IV. DISCUSSION

In this study, the feasibility and effectiveness of deep learning methods for automatic kiwifruit detection in orchards have been investigated. Kiwifruit detection is a challenging task due to the complex background and similarity between fruits and branches. Traditional hand-picking methods, being dependent on human labor, may negatively impact productivity. Therefore, the application of deep learning methods in agricultural automation holds significant potential for yield estimation and cost reduction. Two different deep learning methods, Faster R-CNN and Mask R-CNN, are used for kiwifruit detection, and their results were compared. The Faster R-CNN model was evaluated using MobileNetV3 and SqueezeNet backbone architectures, achieving mAP (mean Average Precision) values of 88.8% and 87.4%, respectively. However, the Mask R-CNN model, with a more advanced backbone architecture, ResNet-50, achieved the highest accuracy with a mAP value of 98.48%. The results demonstrate that the ResNet-50 based Mask R-CNN model is the most suitable and effective method for kiwifruit detection. This model outperforms other methods in terms of mAP and provides more precise object detection, which can significantly improve agricultural production processes by optimizing yield estimation, reducing costs, and enhancing productivity. Additionally, other methods used in this study, such as the Faster R-CNN model with MobileNetV3 and SqueezeNet backbone architectures, also yielded promising results. This indicates that these methods can be effectively utilized in different datasets and application scenarios. The study highlights the potential of deep learning methods for enhancing efficiency and accuracy in agricultural automation.

V. CONCLUSION

This study investigates the effectiveness and applicability of deep learning methods for automatic kiwifruit detection in orchards. Utilizing the ResNet-50 backbone architecture, the Mask R-CNN model achieved the highest accuracy, with a mAP value of 98.48%. On the other hand, the Faster R-CNN model using MobileNetV3 and SqueezeNet backbone architectures obtained mAP values of 88.8% and 87.4%, respectively. These results indicate that the ResNet-50-based Mask R-CNN model is the most suitable and effective method for kiwifruit detection, demonstrating the significant potential of deep learning methods in yield estimation and cost reduction in agricultural automation. The use of these methods can enable farmers to enhance productivity, optimize yield estimations, and consequently improve agricultural production processes. While the current study provides valuable insights, future studies are encouraged to delve into further optimization of existing models or even explore alternative deep learning architectures to improve both accuracy and computational efficiency. In conclusion, the use of deep learning techniques in kiwifruit detection has significant potential for agricultural automation, yield estimation, and cost reduction. Continued advancements in this area can further contribute to optimizing agricultural production processes and increasing overall efficiency.

ACKNOWLEDGMENT

REFERENCES

[1] Warrington, Ian J., and Graham C. Weston. "Kiwifruit: science and management." (No Title) (1990).

[2] Şahin, Nusret. Ordu ilindeki kivi üretim alanlarında görülen toprak kökenli fungusların ve patojenisitelerinin belirlenmesi. MS thesis. Fen Bilimleri Enstitüsü, 2019.

[3] Mkonyi, L., Rubanga, D., Richard, M., Zekeya, N., Sawahiko, S., Maiseli, B., & Machuve, D. Early identification of Tuta absoluta in tomato plants using deep learning. Scientific African, 10, e00590. (2020).

[4] Hasan, Mosin, Bhavesh Tanawala, and Krina J. Patel. "Deep learning precision farming: Tomato leaf disease detection by transfer learning." Proceedings of 2nd international conference on advanced computing and software engineering (ICACSE). 2019.

[5] Ulutaş, Hasan, and Veysel Aslantaş. "Design of Efficient Methods for the Detection of Tomato Leaf Disease Utilizing Proposed Ensemble CNN Model." Electronics 12.4 (2023): 827.

[6] Razfar, Najmeh, et al. "Weed detection in soybean crops using custom lightweight deep learning models." Journal of Agriculture and Food Research 8 (2022): 100308.

[7] Xia, Yi, Minh Nguyen, and Wei Qi Yan. "A real-time kiwifruit detection based on improved YOLOv7." International Conference on Image and Vision Computing New Zealand. Cham: Springer Nature Switzerland, 2022.

[8] Li, Guo, et al. "Real-time detection of kiwifruit flower and bud simultaneously in orchard using YOLOv4 for robotic pollination." Computers and Electronics in Agriculture 193 (2022): 106641.

[9] Lim, JongYoon, et al. "Deep neural network based real-time kiwi fruit flower detection in an orchard environment." arXiv preprint arXiv:2006.04343 (2020).

[10] Xiang, Qian, et al. "Fruit image classification based on Mobilenetv2 with transfer learning technique." Proceedings of the 3rd international conference on computer science and application engineering. 2019.

[11] Santos, T. T., de Souza, L. L., dos Santos, A. A., & Avila, S. Grape detection, segmentation, and tracking using deep neural networks and three-dimensional association. Computers and Electronics in Agriculture, 170, 105247. (2020).

[12] Zhou, Z., Song, Z., Fu, L., Gao, F., Li, R., & Cui, Y."Real-time kiwifruit detection in orchard using deep learning on Android™ smartphones for yield estimation." Computers and Electronics in Agriculture 179 (2020): 105856.

[13] Fu, Longsheng, et al Fu, L., Feng, Y., Elkamil, T., Liu, Z., Li, R., & Cui, Y. "Image recognition method of multi-cluster kiwifruit in field based on convolutional neural networks." Transactions of the Chinese Society of Agricultural Engineering 34.2 (2018): 205-211.

[14] Fu, L., Feng, Y., Majeed, Y., Zhang, X., Zhang, J., Karkee, M., & Zhang, Q. "Kiwifruit detection in field images using Faster R-CNN with ZFNet." IFAC-PapersOnLine 51.17 (2018): 45-50.

[15] Song, Z., Fu, L., Wu, J., Liu, Z., Li, R., & Cui, Y. "Kiwifruit detection in field images using Faster R-CNN with VGG16." IFAC-PapersOnLine 52.30 (2019): 76-81.

[16] Liu, Z., Wu, J., Fu, L., Majeed, Y., Feng, Y., Li, R., & Cui, Y. Improved kiwifruit detection using pre-trained VGG16 with RGB and NIR information fusion. IEEE access, 8, 2327-2336. (2019).

[17] Pathak, Ajeet Ram, Manjusha Pandey, and Siddharth Rautaray. "Application of deep learning for object detection." Procedia computer science 132 (2018): 1706-1717.

[18] Girshick, Ross. "Fast r-cnn." Proceedings of the IEEE international conference on computer vision. 2015.

[19] Sahin, M. E., Ulutas, H., Yuce, E., & Erkoc, M. F. (2023). Detection and classification of COVID-19 by using faster R-CNN and mask R-CNN on CT images. Neural Computing and Applications, 35(18), 13597-13611.

[20] He, Kaiming, et al. "Mask r-cnn." Proceedings of the IEEE international conference on computer vision. 2017.