DOPS: Drone Optimized Performance Score for Evaluating Real-Time Tomato Ripeness Detection

Ylli Rexhaj[1] , Roni Kasemi [2], Lucas Lammer[3]

***Abstract*— In recent years, deep learning (DL) has emerged as a promising tool to detect ripeness or diseases in different types of plants, which helps farmers monitor crop health and determine the optimal harvest times. However, a significant challenge is the integration of these DL models into drones (UAVs) due to low onboard computing capacity, forcing the images captured by UAV cameras to be transmitted to ground-based processors, introducing delays relying on wireless data transmission that compromise real-time identification and affect the accuracy and efficiency of real-life classification. In this study, we present a new metric called Drone Optimized Performance Score (DOPS) to optimize the performance of real-time Tomato Ripeness Detection, taking into consideration accuracy, frames per second (FPS), and latency. We use a systematic methodology where our research includes an approach in the model training phases and also in the deployment phase of two CNN models, MobileNetV2 and ResNet50, with a main focus on evaluating key performance metrics for classification from drones and integrated cameras. Initially, the lighter model MobileNetV2 proves to be more effective for real-time applications based on DOPS evaluation, but after applying a series of optimizations to ResNet50, which is a resource-intensive model, we can maintain its superior accuracy of 98%, but also outperform MobileNetV2 in DOPS evaluation with higher FPS and lower latency, proving that resource-intensive models can also be optimized for real-world deployment.**

I. INTRODUCTION

Agriculture has long been a vital pillar of society, ensuring both economic sustainability and food security since the beginning of humanity. The Food and Agriculture Organization (FAO) predicts that by 2050, there will be more than 9.73 billion people on the planet, and by 2100, there may be 11.2 billion. As a result, the food sector is under pressure to provide the rising demand for food [1]. To solve these problems and boost production and efficiency, the advancement of Artificial Intelligence (AI) and Machine Vision (MV) are playing a crucial role [2].

Many agricultural practices have been significantly advanced through the integration of AI and Machine Vision in precision agriculture, and according to S. Jinya *et al* [3] one of the new developments is the integration of AI with Unmanned Aerial Vehicles (UAVs) to achieve higher productivity in special, large or untargeted spaces, minimize the cost, and automate the process. Drones or UAVs equipped with high-resolution cameras and sensors have emerged as very valuable tools to capture detailed images and gather important crop data, which, when combined with AI, help farmers monitor crop health and determine the optimal harvest times [4],[5],[6].

Although numerous research studies have been done on the detection of ripeness or diseases of vegetables or fruits using machine vision and deep learning (DL) [7],[8], [9], the key challenge is deploying these DL models on UAVs due to limited onboard computational capacity, requiring the images captured by UAV cameras to be transmitted to ground-based processors for analysis, introducing delays relying on wireless data transmission that can compromise real-time identification and affect the accuracy and efficiency of real-time classification [10], [11], [12]. To tackle the existing problems we present a new metric called DOPS - Drone Optimized Performance Score to optimize the performance of real-time classification of tomato ripeness taking into account accuracy, frames per second (FPS), and latency. Real-time and low-latency classification are crucial in precision agriculture for timely decisions affecting crop health, yield, and resource optimization, enabling targeted interventions and minimizing damage [13]. The research conducts an analysis to compare how a lightweight model MobileNetV2 [14][15] and the resource-intensive Convolutional Neural Network (CNN) model ResNet50 [16][17] perform on two setups: a drone equipped with an onboard camera that captures aerial video and streams it to a ground-based processing unit, and a laptop using its built-in webcam in a controlled indoor environment. The study also investigates how these models should be optimized to achieve higher DOPS. The key contributions of this paper are:

- Introduction of DOPS as a Novel Evaluation Metric: The Drone-Optimized Performance Score is introduced to evaluate model performance on edge-device drones, integrating accuracy, FPS, and latency for real-time applications.
- Benchmarking CNN Architectures for UAV-Based AI: We systematically compare MobileNetV2 and ResNet50, identifying the best-performing architecture for drone-based AI applications.
- Optimization of ResNet50 for Real-Time Performance: ResNet50 undergoes targeted optimizations, including input size reduction, layer freezing, and mixed precision training, improving efficiency for real-time UAV applications without sacrificing accuracy.

The paper proceeds as follows. The related work can be

*This work was not supported by any organization

[1]Ylli Rexhaj is with Faculty of Mechatronics Engineering, University for Business and Technology, 10000 Prishtina, Kosovo ylli.rexhaj@ubt-uni.net

[2]Roni Kasemi is with Faculty of Mechatronics Engineering, University for Business and Technology, 10000 Prishtina, Kosovo roni.kasemi@ubt-uni.net

[3]Lucas Lammer lucas.lammer@gmail.com

https://doi.org/10.34749/3061-0710.2025.7

found in Section 2. Our method for the experimental setup, training phase, and DOPS evaluation measure is described in Section 3. The results of the evaluation phase, the deployment phase in a real-time application, and the model tuning for improved performance are shown in Section 5. A summary of the results and a proposal for further research are presented in Section 6.

II. RELATED WORK

Rejeb *et al.* [18] states that drones are changing the agricultural industry by improving efficiency and operational costs. Drones are used to monitor diseases, reducing pesticide usage and the need for human inspection of the crops. Image and sensor technologies in UAVs (Unmanned Aerial Vehicles) allow farmers to precisely monitor crops and detect diseases early, reducing the need for human labor. However, their study primarily offers a bibliometric overview and does not address practical aspects of deploying affordable, low-cost drones or the feasibility of running algorithms on external computing devices rather than onboard hardware. Rajagopal and Murugan [19] use AI-powered drones to detect diseases in cashew trees. MobileNetV2, a deep learning model, is used to scan photos and pinpoint diseases in their early stages to minimize damage to the trees. Egi *et al.* [20] designed a system that processes drone footage to identify and count tomato flowers and fruits. Their method uses YOLOv5 [21] for object detection and Deep-Sort for tracking. While this system works well for estimating how many fruits and flowers are present, it is focused on counting rather than analyzing ripeness. Hobart *et al.* [22] shows an example of a low-cost drone paired with a consumer-grade RGB camera to detect ripe fruits, demonstrating the potential for affordable solutions in agriculture monitoring with UAVs. While their work focuses on apples, similar approaches can be adapted for other crops, including tomatoes. For tomato ripeness classification specifically, Wang *et al.* [23] introduces a tomato ripeness detection system based on an existing detection framework (RT-DETR) [24], which they adapt to be more efficient. Khan *et al.* [25] introduce a technique that combines CNNs with transformer-based models for tomato ripeness classification. Zhang *et al.* [26] alters YOLOv8 [27] in a different investigation to manage intricate ripeness detecting settings. Although all three models are effective in classifying the ripeness of the crops, they are not developed to run on the lower-end hardware of edge devices. Hernández *et al.* [28] investigates a less compute-intensive method to deal with this by classifying tomato ripening stages using YOLOv3tiny [29]. Their approach tries to find a compromise between accuracy and computational requirements. Therefore, they only use data from a very controlled environment, which makes the model less suited outside of controlled environments.

III. OUR APPROACH: DOPS

Although ripeness detection is very critical, the main objective of this research is to tackle the existing problem of CNN models in real-time applications. This research is structured into two significant phases: the model training phase and the real-time deployment phase. In the initial phase, we concentrate on training the models and assessing the performance of two architectures, MobileNetV2 and ResNet50. We conduct a comparative analysis of their accuracy before testing them in real-world scenarios. The subsequent phase involves deployment, during which we introduce a new metric known as DOPS to evaluate the effectiveness of real-time applications. In this phase, we compare the models using two distinct camera setups: a drone-mounted camera and an integrated laptop camera. This methodology addresses existing challenges related to the classification of wirelessly transmitted frames.

A. Experimental Setup

DJI, in collaboration with Intel, created the compact, reasonably priced DJI Tello, a fully programmable drone that records 720p HD video with its 5-megapixel camera [30]. Due to the restricted processing capability of the drone's onboard processor, direct integration of DL models is not feasible. In this instance, the Tello functions as an aerial imaging tool, gathering visual information and transmitting it in real time to a system on the ground (a laptop in this case). This enables the drone to concentrate on gathering data while the laptop's computing capacity is used to run complex AI models for tasks like classifying the ripeness of tomatoes.

The drone-captured frames are sent in real time to a ground-based laptop with powerful processing capabilities. To better compare the two setups, the laptop also has a 720p HD resolution camera, which allows for a better comparison of how well the two models work with various image sources using the same computational framework.

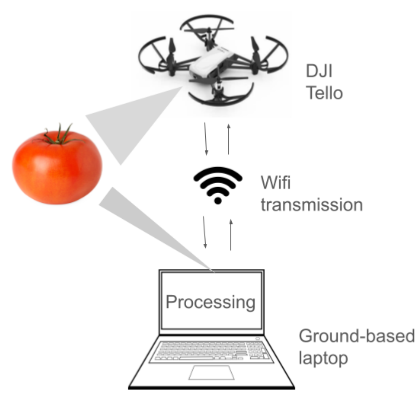

Fig. 1. Experimental Setup

B. Model Training Phase

The models MobileNetV2 and ResNet50 are trained using a small dataset of 711 images, of which 294 are of ripe tomatoes, 302 are of unripe tomatoes, and 115 are for the background to minimize false detection in the background.

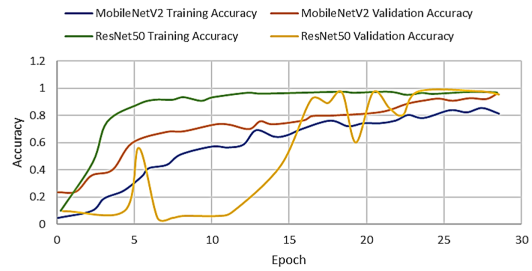

Fig. 2. Training and Validation Accuracy

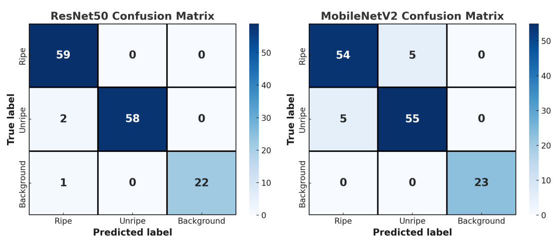

Fig. 3. Comparing Confusion Matrix

Both models are trained for 28 epochs using 569 of the dataset's images for training and 142 for validation. Images of 512x512 pixels are used to train the ResNet50 model, while 224x224 pixel images are used to train the MobileNetV2, which is the primary difference between the models. Validation accuracy and loss metrics are used to moderate the two models' training. As we will discuss later, the model's accuracy and inference latency are impacted by the disparity in using different image size inputs. The training and accuracy curves are illustrated in Figure 2. Starting with a lower initial accuracy, the MobileNetV2 gradually improves, reaching 93% validation accuracy, whereas ResNet50 demonstrates a faster rate of convergence and reaches a higher validation accuracy of 99%. The key difference is that the ResNet50 works better on larger input sizes and contributes to better feature extraction and discrimination between classes.

During training, both models showed a steady decline in training and validation loss. ResNet50 had a lower and more stable validation loss, indicating a strong fit. MobileNetV2 showed more fluctuations, which, while less stable, can help the model avoid sharp, narrow solutions in the loss landscape. Prior work [31] suggests that flatter solutions tend to generalize better to unseen data.

We look at the confusion matrix for both models in Figure 4 and a summary of important performance metrics in Table 1 to further assess how well the two models perform. Because of its deeper architecture and larger input size, ResNet50 performs better than MobileNetV2 across all evaluation metrics, according to the results of training both models. This advantage, however, comes at the expense of higher latency and computational demand, as MobileNetV2, a lightweight model, shows a competitive performance.

Metric	MobileNetV2	ResNet50
Accuracy	0.93	0.99
Macro Avg F1	0.94	0.99
Weighted Avg F1	0.93	0.99

TABLE I
PERFORMANCE COMPARISON

C. DOPS

DOPS is a metric that balances accuracy, FPS (Frames per Second), and latency to evaluate a model's effectiveness in real-time applications. In this case, we use the DOPS metric to compare models in different environments such as laptop vs drone, and also optimize the models for better deployment on real-time applications, where accuracy indicates the classification accuracy of the model, latency (ms) is the time taken for a single inference, including preprocessing and post-processing and FPS measure how many images the model processes per second.

$$\mathrm{DOPS} = \frac{\mathrm{Accuracy} \times \mathrm{FPS}}{\mathrm{Latency}} \tag{1}$$

Different deployment environments prioritize different factors for example in cloud-based applications, accuracy, and FPS are more important as computational resources are not a constraint, but on the other hand in low-power edge devices such as drones, latency, and power consumption are critical, making efficiency a key factor, adding a weight assigned to each parameter based on the real-time application. . However, detailed power consumption analysis is left for future work and will be integrated in subsequent stages of development.

IV. RESULTS

The evaluation of deep learning models for real-time application requires a comprehensive analysis beyond traditional accuracy-base metrics, as introduced in the training phase section. So, using DOPS, we analyze the real-time performance of MobileNetV2 and ResNet50, ensuring the model's reliability and practical deployment constraints, such as FPS and latency, which play a critical role in real-world applications.

A. Deployment Phase

Using two different camera setups, the drone-mounted camera and an integrated laptop camera, the deployment phase concentrates on testing the two models' real-time tomato ripeness detection performance. The drone itself does not perform any onboard processing, instead, it operates as a mobile image acquisition platform, transmitting frames wirelessly in real-time to the laptop from an aerial perspective. To ensure consistency, all image processing is carried out on a laptop equipped with a dedicated GPU. In contrast, the laptop setup eliminates any wireless transmission delay by using the integrated camera to capture frames. In real-time testing, the main difference between the two configurations was the effect of wireless transmission latency.

https://doi.org/10.34749/3061-0710.2025.7

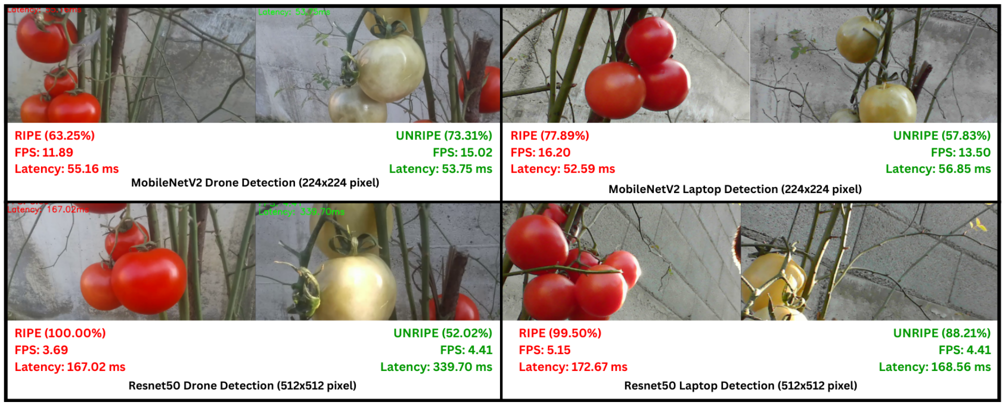

Fig. 4. Real-time FPS and latency acquisition for DOPS evaluation, comparing MobileNetV2 and ResNet50 across drone and laptop setups.

The drone camera wirelessly transmits the frames, as we previously discussed, but this leads to network-induced delays that result in irregular frame drops and slow processing time. Even though the aerial perspective covers a larger range of view, latency has a detrimental impact on real-time classification. In contrast, directly collected frames on the laptop eliminate transmission latency, enabling faster inference and greater FPS. However, because of its relatively limited field of view, the fixed camera proves less effective in monitoring large crop fields.

During the testing phase, both models MobileNetV2 and ResNet50 are tested in real-time in both setups, compared, and evaluated using the DOPS metric, the results of which are shown and compared in section 4 B, from the real-time acquisition. The live performance of both models is shown in Figure 4 with the labels of live measurement results of FPS, latency, and accuracy.

B. DOPS Evaluation

Following the real-time deployment phase, we evaluate MobileNetV2 and ResNet50 in laptop and drone configurations using the DOPS. DOPS provides a balanced assessment that reflects real-time feasibility, which is crucial for agricultural AI applications, in contrast to traditional evaluations that only consider accuracy. The primary objective of this evaluation is to assess each model's performance by combining accuracy, frames per second (FPS), and latency as critical variables. Stated differently, an AI model is considered better when it has a higher DOPS score, meaning it can process frames quickly and with minimal latency while maintaining high classification accuracy, on the other hand, lower DOPS indicates worse real-time performance. A higher DOPS score indicates that the model is well-suited for real-time applications, as speed and precision are crucial in drone-based agricultural monitoring [32].

Despite this, MobileNetV2 is an optimal model for real-time inference, maintaining a strong balance between accuracy and processing speed.

In contrast, the resource-intensive ResNet50 model, known for its superior accuracy, performs well in terms of classification accuracy, with 99% accuracy in both sets. However, as seen in Figure 5, its computational intensity significantly impacts its real-time usability. Table 3 shows that compared to MobileNetV2, the FPS in the laptop setup is only 5.61, and in the drone setup, it is much lower at 3.30. Additionally, ResNet50 has a significantly greater latency, reaching 312.19 ms in drone setup and 173.52 ms in laptop setup. After the evaluation, the DOPS results of ResNet50 are far lower than those of MobileNetV2. Scoring 0.0320 on the laptop and only 0.0105 for the drone setup. This demonstrates that, despite its great accuracy, ResNet50 is not appropriate for real-time drone-based agricultural applications due to its slow inference speed and high latency.

The main factor causing ResNet50 to perform worse than MobileNetv2 in all performance metrics is its large input size (512x512), while MobileNetV2 has an input size of 224x224. This resolution has a significant effect on processing time and computational power, leading to higher latency and a lower frame rate since each image requires more memory and computation with each forward pass. As a result, ResNet50 is ineffective for real-time applications and takes longer to analyze each frame. This leads to lower FPS and greater latency, whereas MobileNetV2's smaller input size enables

Setup	Laptop	Drone
Accuracy	0.93	0.93
Latency	62.26	65.56
FPS	15.44	15.05
DOPS	0.23	0.21

TABLE II
MOBILENETV2 (LAPTOP VS DRONE) DOPS EVALUATION

Setup	Laptop	Drone
Accuracy	0.99	0.99
Latency	173.52	312.19
FPS	5.61	3.30
DOPS	0.032	0.010

TABLE III
RESNET50 (LAPTOP VS DRONE) DOPS EVALUATION

https://doi.org/10.34749/3061-0710.2025.7

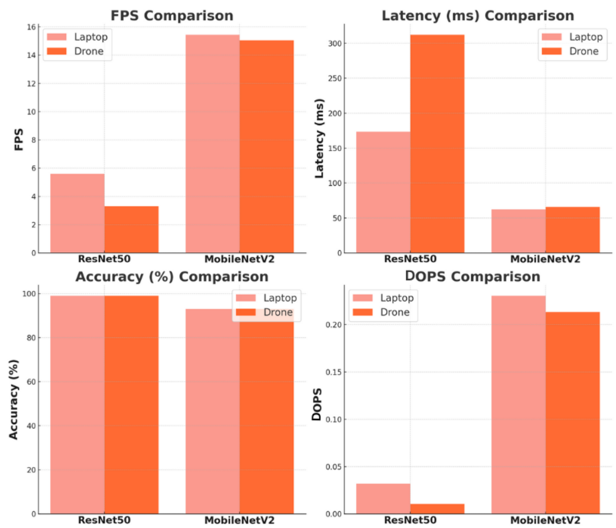

Fig. 5. DOPS Evaluation comparison of MobileNetV2 & ResNet50 (Laptop vs Drone)

faster processing with less memory usage, allowing for smoother and more stable real-time inference. To add to this, ResNet50 is a model with more parameters and convolutional layers than MobileNetV2, which uses fewer computational resources. Wireless transmission affects both models, with ResNet50's larger input size causing heavier data packets and increased network latency. MobileNetV2's smaller input size leads to faster data transfer and less lag, making it more effective in drone-based agricultural monitoring. ResNet50 performs worse in real-time deployment due to its deeper network architecture, larger input shape, and higher computational demands. However, MobileNetV2's ability to balance speed and accuracy makes it more suitable for real-time AI applications, especially in UAV-based agricultural monitoring.

C. Optimizing ResNet50 for Real-Time Use

In this section, we'll maximize ResNet50's effectiveness without sacrificing its high classification accuracy. Our main optimizations include freezing the first 50 layers of ResNet50 and lowering the size of the input image. We also use TensorFlow's automatic mixed precision feature to apply mixed precision training, using FP16 computation whenever feasible. A balance between computational efficiency and retention is achieved by reducing the input size from 512x512 to 256x256, which increases inference speed without significantly affecting classification performance.

Using mixed-precision training optimization not only reduces GPU memory usage but also accelerates training and inference speed, making the model more appropriate for real-time deployment. Freezing the layers allows the model to retain its ability to extract robust features, which results in faster model convergence and decreased processing time per frame. Following training with these adjustments, the model's training performance is displayed in Figure 6. It

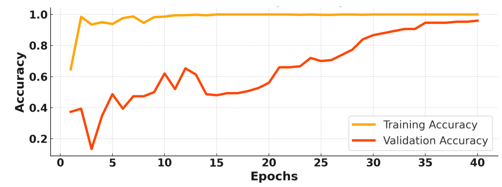

Fig. 6. Optimized ResNet50 Accuracy over Epochs

is evident that the optimized ResNet50 maintains a high training and validation accuracy of 98% throughout the training phase. This time, 30% of the data is split for validation during training, and we observe that only four images are incorrectly classified in Figure 7 in the confusion matrix. But evaluating the model in real-time testing during the deployment phase is the primary objective through the DOPS Evaluation.

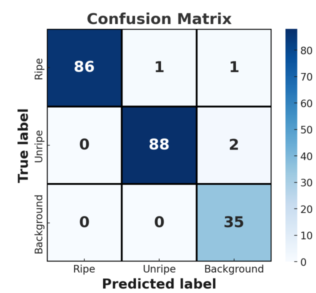

Fig. 7. Confusion Matrix of Optimized ResNet50

Setup	Laptop	Drone
Accuracy	0.98	0.98
FPS	20.93	21.69
Latency (ms)	42.89	46.74
DOPS	0.48	0.45

TABLE IV
DOPS EVALUATION FOR OPTIMIZED RESNET50

The optimized ResNet50 model outperforms MobileNetV2 in real-time inference speed and latency, despite maintaining a high classification accuracy of 98%. The laptop setup achieves an FPS of 20.93, while the drone setup slightly outperforms it at 21.69 FPS. The optimizations increase processing speed without impacting performance. Latency, a key variable of the DOPS evaluation, is substantially reduced compared to the model before the optimization. The laptop setup achieves a latency of 42.89 milliseconds, while the drone setup exhibits a slightly higher latency of 46.74 milliseconds, primarily due to wireless transmission delay. These latency improvements are significant compared to the standard ResNet50 implementation, making the optimized model more viable for real-time classification in drone-based agricultural monitoring. The overall DOPS score confirms the success of these optimizations, with the laptop setup achieving a DOPS score of 0.478, while the drone setup slightly lags at 0.454 due to network-related delays. These

https://doi.org/10.34749/3061-0710.2025.7

results highlight that the optimized ResNet50 successfully balances accuracy and real-time performance, making it a viable solution for UAV-based agricultural classification tasks.

V. CONCLUSIONS

In this work, we introduce a new metric for evaluation the Drone Optimized Performance Score (DOPS), a benchmark for real-time deep learning inference on UAVs. DOPS is a metric that takes into account accuracy, frame rate, and latency, providing a simple yet powerful evaluation for the real-time application of AI models. As a finding initially, MobileNetV2, being a lightweight model, outperforms the heavy resource model ResNet50 in real-time inference speed and, therefore, overall in DOPS. But, after some optimizations such as input size reduction, mixed-precision training, and layer freezing made on the ResNet50, it was able to surpass MobileNetV2, raising the DOPS score from 0.010 to 0.45 and improving the real-time performance while still maintaining a high classification accuracy. In agricultural monitoring, where timely and precise identification of issues like crop stress or insect outbreaks is crucial to preventing yield loss and guaranteeing resource efficiency, this study shows that DOPS is a useful metric for assessing real-time applications. Also, it provides a framework for improving deep learning models' performance on edge devices such as UAVs while still maintaining high accuracy. Future work includes further optimizing deep learning models for UAV-based inference by integrating advanced model compression techniques such as pruning and quantization. These methods will reduce computational overhead while maintaining high classification accuracy. Additionally, deploying the models directly on the edge devices is work that will be implemented to eliminate transmission delays, improving real-time responsiveness. Also, to extend the applicability of DOPS, power consumption will be incorporated as a metric, enabling more energy-efficient AI deployments on battery-powered drones.

REFERENCES

[1] Food and Agriculture Organization of the United Nations (FAO). *The Future of Food and Agriculture: Trends and Challenges*. FAO, Rome, Italy, 2017.

[2] Rosana Cavalcante de Oliveira and Rogério Diogne de Souza e Silva. Artificial intelligence in agriculture: Benefits, challenges, and trends. *Applied Sciences*, 13(13), 2023.

[3] Jinya Su, Xiaoyong Zhu, Shihua Li, and Wen-Hua Chen. Ai meets uavs: A survey on ai empowered uav perception systems for precision agriculture. *Neurocomputing*, 518:242–270, 2023.

[4] Zhao Zhang, Hu Liu, Ce Yang, Yiannis Ampatzidis, Jianfeng Zhou, and Yu Jiang. *Unmanned Aerial Systems in Precision Agriculture*. 05 2022.

[5] Jong-Hwa Park and Dong-Ho Lee. Development of a uav-based multi-sensor deep learning model for predicting napa cabbage fresh weight and determining optimal harvest time, 07 2024.

[6] Juhi Agrawal and Muhammad Yeasir Arafat. Transforming farming: A review of ai-powered uav technologies in precision agriculture. *Drones*, 8(11), 2024.

[7] Andreas Kamilaris and Francesc X. Prenafeta-Boldú. Deep learning in agriculture: A survey. *Computers and Electronics in Agriculture*, 147:70–90, 2018.

[8] Md Habib, Md. Ariful Arif, Sumaita Binte Shorif, Mohammad Shorif Uddin, and Farruk Ahmed. *Machine Vision-Based Fruit and Vegetable Disease Recognition: A Review*, pages 143–157. 03 2021.

[9] Mahmoud Soltani Firouz and Hamed Sardari. Defect detection in fruit and vegetables by using machine vision systems and image processing. *Food Engineering Reviews*, 14, 03 2022.

[10] Alfonso Torres-Rua. Drones in agriculture: an overview of current capabilities and future directions. In *Utah Water Users Workshop, Saint George, UT, USA*, pages 1–9, 2017.

[11] Chenglin Zhang, Xiaoyong Qiang, Qingzhe Lv, Hao Guan, and Fang Huang. Comparative study on real time image data transmission methods for unmanned aerial vehicle. In *IGARSS 2024 - 2024 IEEE International Geoscience and Remote Sensing Symposium*, pages 6609–6612, 2024.

[12] Xu Kang, Bin Song, Jie Guo, Zhijin Qin, and F. Richard Yu. Task-oriented image transmission for scene classification in unmanned aerial systems, 2021.

[13] Tarek Alahmad, Miklós Neményi, and Anikó Nyéki. Applying iot sensors and big data to improve precision crop production: A review. *Agronomy*, 13(10), 2023.

[14] Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, and Liang-Chieh Chen. Mobilenetv2: Inverted residuals and linear bottlenecks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4510–4520, 2018.

[15] Xiaofei Xie, Guodong Zhao, Wei Wei, and Wei Huang. Mobilenetv2 accelerator for power and speed balanced embedded applications. In *2022 IEEE 2nd International Conference on Data Science and Computer Application (ICDSCA)*, pages 134–139, 2022.

[16] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[17] Matthew Wilkerson, Grace Vincent, Zaki Hasnain, Sambit Bhattacharya, and Emily Dunkel. Benchmarking resnet50 for image classification, Dec 2024.

[18] Anis Rejeb, Karim Rejeb, and Harry B. Simske. Drones in agriculture: A review and bibliometric analysis. *Computers and Electronics in Agriculture*, 198:107017, 2022.

[19] R. Rajagopal and S. Murugan. Artificial intelligence-based drone for early disease detection and precision pesticide management in cashew farming. *Agricultural Systems*, 190:103096, 2023.

[20] R. Egi, T. Nakamura, and L. Xu. Drone-computer communication based tomato generative organ counting model using yolo v5 and deep-sort. *IEEE Access*, 10:57834–57845, 2022.

[21] Glenn Jocher. Yolov5 by ultralytics, 2020.

[22] Marius Hobart, Michael Pflanz, Nikos Tsoulias, Cornelia Weltzien, Mia Kopetzky, and Michael Schirrmann. Fruit detection and yield mass estimation from a uav based rgb dense cloud for an apple orchard. *Drones*, 9(1), 2025.

[23] T. Wang, X. Liu, and Y. Zhang. Lightweight tomato ripeness detection algorithm based on the improved rt-detr. *Sensors*, 24(5):1897–1912, 2024.

[24] Wenyu Lv, Shangliang Xu, Yian Zhao, Guanzhong Wang, Jinman Wei, Cheng Cui, Yuning Du, Qingqing Dang, and Yi Liu. Detrs beat yolos on real-time object detection, 2023.

[25] M. Khan, A. Gupta, and P. Ramesh. Tomato maturity recognition with convolutional transformers. *International Journal of Computer Vision*, 132(3):412–430, 2023.

[26] F. Zhang and L. Chen. A method for detecting tomato maturity based on deep learning. *IEEE Transactions on Automation Science and Engineering*, 21(2):334–345, 2024.

[27] Glenn Jocher, Ayush Chaurasia, and Jing Qiu. Ultralytics yolov8, 2023.

[28] J. Hernández, P. Alvarez, and M. Gonzalez. Detection of tomato ripening stages using yolov3-tiny. *Journal of Agricultural Robotics*, 12(4):215–230, 2023.

[29] Pranav Adarsh, Pratibha Rathi, and Manoj Kumar. Yolo v3-tiny: Object detection and recognition using one stage improved model. In *2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS)*, pages 687–694, 2020.

[30] RYZE Robotics. Tello. https://www.ryzerobotics.com/fr/tello.

[31] Nitish Shirish Keskar, Dheevatsa Mudigere, Jorge Nocedal, Mikhail Smelyanskiy, and Ping Tak Peter Tang. On large-batch training for deep learning: Generalization gap and sharp minima. In *International Conference on Learning Representations (ICLR)*, 2017.

[32] Hicham Slimani, Jamal El Mhamdi, and Abdelilah Jilbab. Deep learning structure for real-time crop monitoring based on neural architecture search and uav, Sep 2024.

https://doi.org/10.34749/3061-0710.2025.7