

RESEARCH

Open Access



Enhancing oil palm segmentation model with GAN-based augmentation

Qi Bin Kwong^{1*}, Yee Thung Kon¹, Wan Rusydiah W. Rusik¹, Mohd Nor Azizi Shabudin¹, Shahirah Shazana A. Rahman¹, Harikrishna Kulaveerasingam¹ and David Ross Appleton¹

*Correspondence:

Qi Bin Kwong
kwong.qi.bin@gmail.com
¹SD Guthrie Research Sdn Bhd,
Banting, Selangor Darul Ehsan,
Malaysia

Abstract

In digital agriculture, accurate crop detection is fundamental to developing automated systems for efficient plantation management. For oil palm, the main challenge lies in developing robust models that perform well in different environmental conditions. This study addresses the feasibility of using GAN augmentation methods to improve palm detection models. For this purpose, drone images of young palms (<5 year-old) from eight different estates were collected, annotated, and used to build a baseline detection model based on DETR. StyleGAN2 was trained on the extracted palms and then used to generate a series of synthetic palms, which were then inserted into tiles representing different environments. CycleGAN networks were trained for bidirectional translation between synthetic and real tiles, subsequently utilized to augment the authenticity of synthetic tiles. Both synthetic and real tiles were used to train the GAN-based detection model. The baseline model achieved precision and recall values of 95.8% and 97.2%. The GAN-based model achieved comparable result, with precision and recall values of 98.5% and 98.6%. In the challenge dataset 1 consisting older palms (>5 year-old), both models also achieved similar accuracies, with baseline model achieving precision and recall of 93.1% and 99.4%, and GAN-based model achieving 95.7% and 99.4%. As for the challenge dataset 2 consisting of storm affected palms, the baseline model achieved precision of 100% but recall was only 13%. The GAN-based model achieved a significantly better result, with a precision and recall values of 98.7% and 95.3%. This result demonstrates that images generated by GANs have the potential to enhance the accuracies of palm detection models.

Keywords Oil palm segmentation, GAN, Object detection, Object segmentation, Data augmentation, Vision transformer, Detectron, Phenotyping

Introduction

As the world's population is poised to reach unprecedented levels in the coming decades [1], ensuring food security for this rapidly growing populace becomes a global imperative. Addressing this challenge necessitates a holistic and sustainable approach to agriculture. In recent years, digital agriculture has emerged as a transformative force, reshaping our conventional methods. This surge in automation, driven by innovations

such as computer vision, the Internet of Things, and robotics, is causing a paradigm shift in the agriculture sector. Simultaneously, these technologies have also enabled automated phenotyping, which accurately assesses plant traits. This plays a key role in improving breeding selection and advancing precision breeding. These automated practices hold the potential to substantially boost agricultural yield per unit area.

Despite the potential of oil palm to yield up to 10 tonnes oil per hectare per year (t/ha/yr), the global average productivity has plateaued around 3 t/ha/yr. Unfortunately, progress in closing this gap has been sluggish for many years. Genomic selection initiatives have shown great promise in addressing this issue. This is particularly true for yield component traits with high genetic heritability such as shell or fruit mesocarp thickness, as various researchers have shown [2, 3]. However, when it comes to complex traits like total oil yield [4] and height [5], environmental factors account for a significant 60% or more of the variation, adding complexity to the improvement efforts. Furthermore, phenotyping for complex traits tends to be slow and labour-intensive, a significant challenge given labour shortages [6]. In response to these challenges, the integration of digital agriculture and automated phenotyping has emerged as a pivotal solution. Among the most cost-effective tools in this transformative journey are the usage of drones.

The utilization of drone technology in agriculture is an emerging and continuously evolving practice. Its applications covers various aspects, including crop classification, pest and disease detection [7–9] and phenotyping tasks such as height measurements [10]. While recent advancements in computer vision, especially convolutional neural networks (CNNs), have facilitated the development of highly accurate and automated agricultural object detection models [10–12], the persistent challenge lies in constructing models that can robustly generalize across diverse scenarios. This challenge is amplified by the vast diversity of environmental conditions encountered in real-world settings. Within the oil palm industry, the accurate detection and segmentation of palms across diverse age groups, sizes, and environmental conditions remains a significant challenge, impeding the widespread implementation of automated monitoring and management systems. Moreover, the manual annotation of extensive image datasets demands substantial resources in terms of time and labour.

The introduction of generative adversarial network (GAN) [13] offers an intriguing prospect for addressing these challenges. GAN provides a pathway to augment existing image datasets and generate new data [14, 15]. In agriculture, researchers have proposed GANs as a solution to mitigate overfitting and improve CNN classification networks, with successful applications in identifying diseases in crops such as tomatoes and grapes [16–18]. Given the relative novelty of this approach in agricultural context, the purpose of this study is to assess the influence of GAN on both detection and segmentation accuracies, with a particular emphasis on oil palm.

Methodology

Data collection and processing

To ensure a wide representation of diverse terrains, topographies, and other environmental factors in oil palm plantations, eight estates owned by SD Guthrie across Malaysia were selected using a stratified random sampling method. These estates contained immature or young palms (<5 years old, before canopy overlap) with a planting density of approximately 180 palms per hectare. For this study, the DJI Mavic 2 Pro drone,

equipped with a Hasselblad camera with an F2.8 EQV 28 mm lens, was employed for mapping purposes. The flight altitude was set at 80 m to capture detailed imagery. Flights were scheduled between 8 AM and 11 AM, and 2 PM and 5 PM on clear days with low wind conditions to capture varied lighting conditions while minimizing shadow effects. As for the drone settings, image overlap was set at 80%, sidelap at 60%, and a flight speed of 5 m/s was maintained to balance efficiency with image quality. Additional drone settings, including optimal camera parameters such as exposure, sharpness, and precise GPS accuracy, were kept at default.

The collected images were uploaded to our customized WebODM [19] server. After image processing, the stitched orthophotos were separated into individual tiles of 640×640 using `gdal_retile.py` script from GDAL library [20]. From this step, 7755 generated tiles were selected and annotated using LabelMe [21]. After manual inspection and quality checking, 6499 high-quality tiles were selected. Of these tiles, 5168 tiles were randomly assigned to the training set, while the remaining 1331 tiles were used as the validation set.

Two additional independent estates were chosen, and the images acquired followed the same tiling and processing procedures as previously described. From these estates, a total of 100 tiles were selected and designated as test/challenge set 1. This dataset consists primarily of palms older than 5 years. In addition, a separate challenge set 2 was assembled from an estate impacted by a destructive storm, resulting in the generation of 100 tiles for evaluation.

Detection and segmentation model

The palm detection and segmentation models were built with Detection Transformer (DETR) [22] on top of the Detectron 2 framework [23]. The model backbone architecture used was the “ResNeXt50_32×4D” [24], an extension of the ResNet architecture [25], featuring 50 network layers, 32 cardinality levels and width of 4 (Fig. 1).

For the transformer-based object detection training, the initial learning rate was set at 1e-4, batch size 16, weight decay at 1e-4 and learning rate drop at 50. Most of these parameters were determined through trial and error using grid search, while the batch size was optimized based on the available computing memory. The encoding and decoding layers were both kept at 6, embedding size at 256, dropout at 0.1, number of attention heads at 8 and number of query slots at 100 [22]. Training was stopped when both the train and validation loss converged. The segmentation head of the network was trained separately using the frozen weight from the previous training. The same

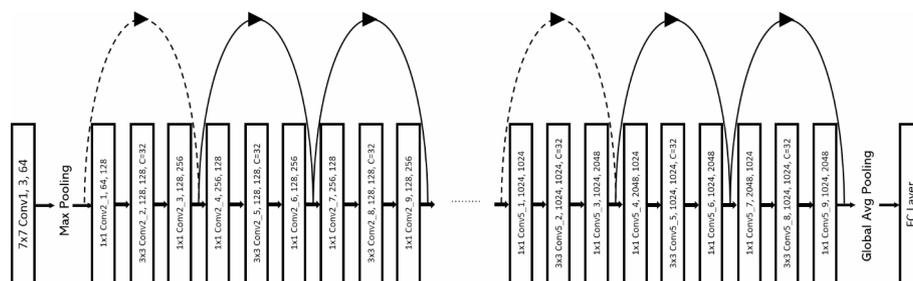


Fig. 1 Representative backbone architecture for ResNeXt50_32×4D. The convolutional layers were labelled as kernel size, convolutional layer name, in channel, out channel, cardinality level. The dotted curve arrows represent skip connection with dimension correction (convolutional residual block), whereas the solid curved arrows represent skip connection without dimension correction (identity residual block)

parameters were used except for the learning rate drop, set at 20 and batch size at 4. The model trained was labelled as “baseline palm model”. The training was carried out on a Google Cloud Platform Virtual Machine with a single NVIDIA Tesla A100 GPU, 85GB RAM and 12 CPU.

During each validation step, Common Objects in Context (COCO) [26] evaluator function was used to assess the model quality/accuracy. The COCO evaluation metrics used in this study were mean Average Precision (mAP) (at Intersection over Union (IoU) of 0.50:0.95) [26], Average Precision (AP) (at IoU of 0.50) [27] and mean Average Recall (mAR) (at IoU 0.50:0.95), both for maximum detections of 100 and for all areas. Besides COCO metrics, a simpler and more practical metric, known as palm count precision/recall, which was based on precision and recall at a detection score of 0.9, was also calculated manually.

GAN-based augmentation

From the training dataset, individual palms were segmented out from the tiles and placed into the center of 256×256 pixel sized images with black backgrounds using a Python script. The images were manually reviewed, and those depicting complete and clear palm features were selected. 1,444 images were selected from this step. These images served as the dataset for training the Generative Adversarial Network (GAN) generator and discriminator from scratch. The GAN architecture used in this step was Style-based Generative Adversarial Network 2 (StyleGAN2) [28] with adaptive discriminator augmentation [29], implemented in Pytorch [30, 31]. The “king” parameter was set at 25,000 [29], learning rate at 0.0025 while the batch size was set at 64 with a single GPU. The other parameters were kept at default. The training process was stopped after the FID (Fréchet inception distance) score plateaued and no longer showed improvement on TensorBoard [32]. Using the resulting model, approximately 200,000 synthetic palm images were generated. Accompanying each of these synthetic palms, automated palm segmentations were generated in JSON [33] format using a customized Python script.

37 random drone orthophotos from diverse global locations were retrieved from OpenAerialMap [34] and subsequently partitioned into individual tiles. From this pool, a total of 20,225 tiles were selected, alongside an additional 29,775 background control tiles generated from vacant field images. This combined dataset of 50,000 tiles served as the background dataset for the subsequent phase of the study. Employing a custom Python script, four synthetic palm images were randomly inserted into each background tile, ensuring no overlap, and simultaneously generating the corresponding segmentation JSON file.

Cycle-Consistent Generative Adversarial Network (CycleGAN) [35] was utilized to enhance the realism of synthetic tiles. The dataset of this step comprised of 50,000 synthetic tiles and 13,422 real (unannotated drone-captured) tiles. To facilitate model training and evaluation, both the synthetic and real tiles were divided into training and validation sets. In this process, 90% of the tiles from each category were designated for the training set, while the remaining 10% were set aside for validation. In the context of the A-to-B direction, the synthetic tiles were utilized as training dataset A, and the real tiles were employed as validation dataset B. Conversely, for the B-to-A direction, the real tiles constituted training dataset B, and the synthetic tiles served as validation dataset

A. The selected mode for the GAN was “lsgan” with the discriminator network (net_D) kept as “basic”. Conversely, the generator network (net_G) implemented was “UNET_128”. The learning rate was set at 0.0002, batch size at 20, decay epoch at 10 and loading size at 640. These parameters were determined through iterative trial and error experiments, ultimately selecting the configuration that produced the best balance of image quality, training stability, and computational efficiency for the augmentation task. Training was stopped when the loss values for generator losses and the discriminator losses all stabilized. Attainment of acceptable level of image quality was another condition. Utilizing the final generator model, all the synthetic tiles were transformed to closely resemble the real drone tiles.

33,746 good quality transformed tiles were combined with the original 5,168 tiles and used as the new training set to build the new palm detection and segmentation models using the same network architecture and method as before (Supplementary Fig. 1). The resulting model, known as the “GAN Palm Detector” was also evaluated on the final test/challenge datasets.

The performance difference between the GAN palm model and the baseline model was quantified. Using the “numpy” and “scipy” Python library, 95% confidence intervals for palm count precision, recall and F1 score metrics were calculated across all challenge datasets. To quantify the uncertainty in our performance metrics, we implemented a bootstrap resampling procedure. 1000 bootstrap samples were generated, each comprising 30 randomly selected individuals from the original dataset, using sampling with replacement. The resulting confidence intervals were then analysed to determine the statistical significance of the observed performance improvements.

Software used

All steps in the methodology were carried out using Python 3.10.13.

Result

StyleGAN2-ADA network training was completed after 1500 epochs, and the final plateaued FID score was 16.82. Sample synthetic palm images can be found in Fig. 2.

As for CycleGAN, training was stopped after 50 epochs. The generator A had a loss of 0.166, and discriminator A's loss was 0.266, while for generator B and discriminator B, the losses were 0.196 each. An example of synthetic tile before and after CycleGAN transformation can be found in Fig. 3.

Tables 1 and 2 present the COCO-evaluated model performance for both the baseline and GAN-based models, focusing on detection and segmentation respectively. Additionally, Fig. 4 illustrates the CIs calculated for palm count precision/recall metrics. The result shows that the performance of both GAN palm model with the baseline palm model were comparable for both the validation set and the challenge set 1. The GAN palm model outperformed the baseline palm model for challenge dataset 2. As illustrated in Fig. 4, the baseline model achieved 100.0% precision but only 13.0% recall (F1 score 0.23), indicating that while it rarely misidentified palms, it failed to detect most of them. In contrast, the GAN-based model achieved both high precision (98.7%) and high recall (95.3%), resulting in a significantly higher F1 score of 0.97. Representative examples of the baseline and GAN-based model's performance in detecting and segmenting palms were shown in Figs. 5, 6 and 7.

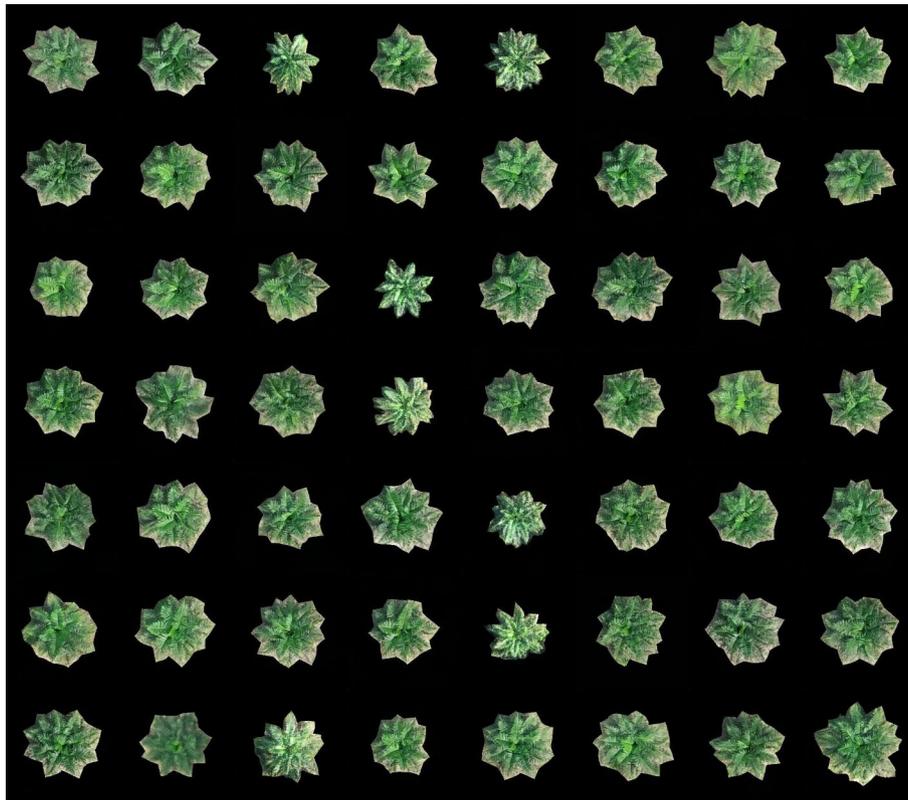


Fig. 2 Sample GAN-generated palm images



Fig. 3 **A)** Synthetic palm tile before CycleGAN transformation. **B)** Synthetic palm tile after CycleGAN transformation

Table 1 COCO evaluation table for palm detection models

	Baseline Palm Model	GAN Palm Model	Baseline Palm Model (Test Set 1)	GAN Palm Model (Test Set 1)	Baseline Palm Model (Test Set 2)	GAN Palm Model (Test Set 2)
Mean Average Precision (mAP)	0.628	0.758	0.468	0.462	0.025	0.512
Average Precision (AP)	0.966	0.972	0.922	0.927	0.053	0.927
Mean Average Recall (mAR)	0.702	0.785	0.585	0.587	0.083	0.604

Table 2 COCO evaluation table for palm segmentation models

	Baseline Palm Model	GAN Palm Model	Baseline Palm Model (Test Set 1)	GAN Palm Model (Test Set 1)	Baseline Palm Model (Test Set 2)	GAN Palm Model (Test Set 2)
Mean Average Precision (mAP)	0.502	0.625	0.401	0.398	0.032	0.413
Average Precision (AP)	0.958	0.963	0.878	0.876	0.065	0.937
Mean Average Recall (mAR)	0.569	0.647	0.484	0.483	0.143	0.495

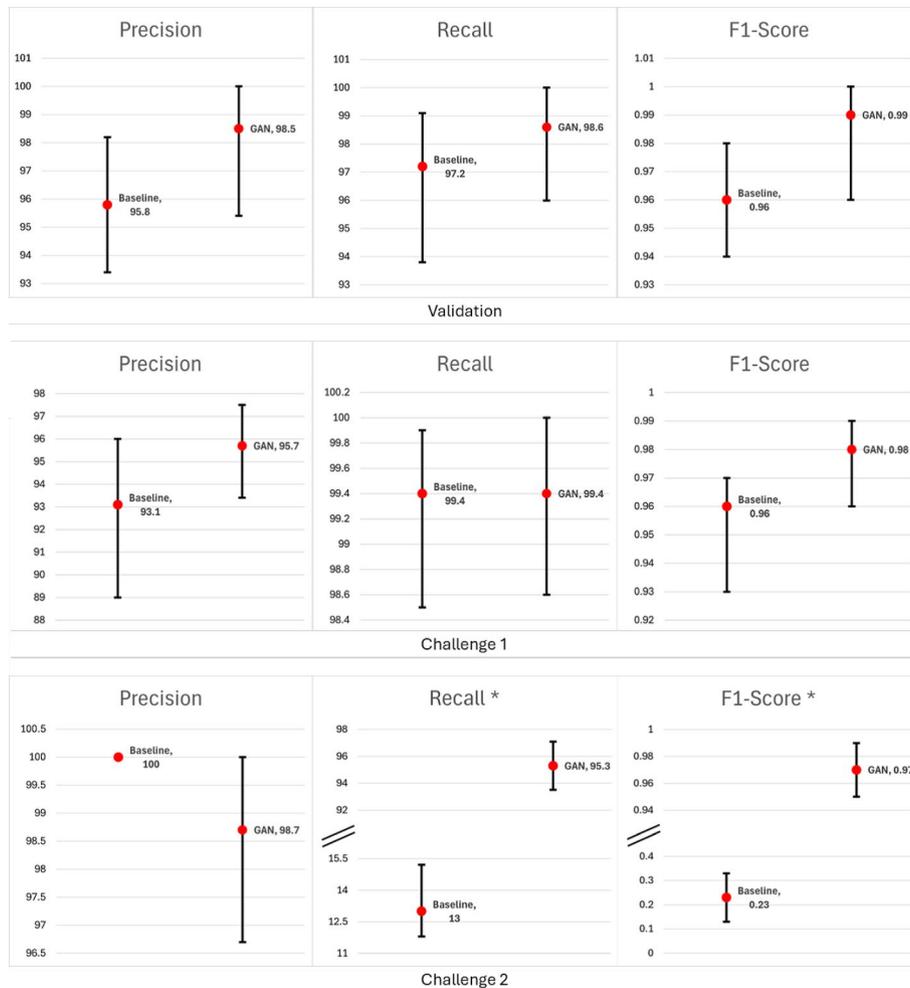


Fig. 4 Precision, Recall, and F1 Scores: Baseline vs. GAN palm models across validation and challenge datasets. Asterisks (*) denote statistically significant differences between models

Discussion

The introduction of CNN-based object detection models has spurred advancements in agricultural automation. Notably, these models have found application in tasks such as automated weed identification in crops [36] and the detection of plant diseases [37]. In the context of oil palm industry, CNN-based models are being explored for automated phenotyping and drone or satellite-based palm detection and counting [38–41]. This current study builds upon the application of CNNs in palm detection and extends the concept into palm segmentation. Here, segmentation is specifically referred to as instance segmentation, focusing on pixels representing individual palms, instead of panoptic and

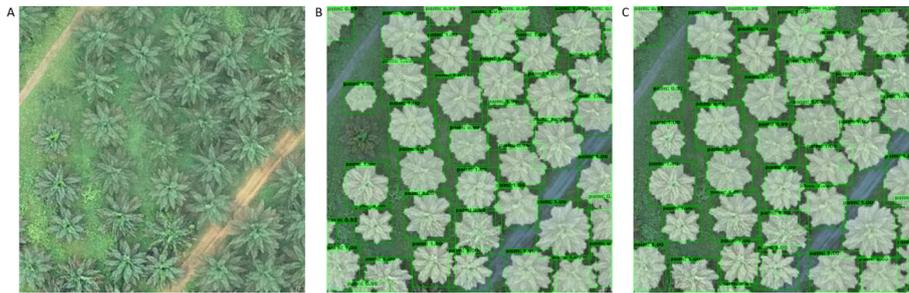


Fig. 5 Comparison of **A**) raw tile **B**) baseline palm model and **C**) GAN palm model for validation dataset. Both models performed almost equally in detecting palms

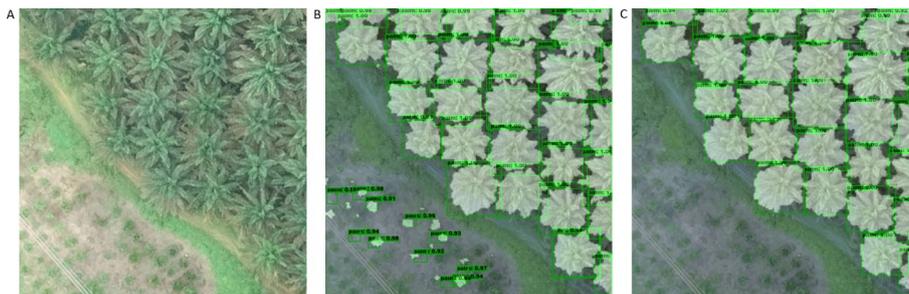


Fig. 6 Comparison of **A**) raw tile **B**) baseline palm model and **C**) GAN palm model for challenge dataset 1. The baseline model mistakenly detected some of the low-resolution shrubs as palms



Fig. 7 Comparison of **A**) raw tile **B**) baseline palm model and **C**) GAN palm model for challenge dataset 2. The baseline model was not able to detect storm-affected palms. Comparatively, the GAN palm model was even able to detect the fallen palm in the middle

semantic segmentation [42]. However, a significant challenge lies in the generalizability of these palm CNN models across diverse environments. Unlike crops grown in controlled environments like greenhouses, oil palms are cultivated in open fields, exposed to numerous unpredictable factors. Environmental elements such as weather conditions (rain, wind, and fog) and varying lighting conditions influenced by sunlight and time of day are known to impact drone image quality [43]. Additionally, factors such as drone camera type and flight altitudes have been identified as contributors to variations in image quality [44]. The intricacies of image processing and stitching further complicate this issue [45, 46]. Hence, developing a generalizable model necessitates a diverse representation of palm images. Rather than manually addressing every conceivable scenario, GAN [13] offer a potential solution to mitigate these issues.

GAN based background switch has been proposed as an augmentation method for object detection [47]. This study shares a similar augmentation principle as the

referenced publication, with the object of interest being inserted into a new background image. However, one major difference is that the palms being used were generated via StyleGAN2 [28, 29]. StyleGAN2 comprises a generator and a discriminator; the generator produces synthetic images, while the discriminator evaluates and distinguishes them from real images. An essential feature of StyleGAN2 is its ability to independently manipulate high and fine-level details in images, known as style-mixing. It also introduces stochastic variation, adding randomness for greater diversity in synthetic image generation (Supplementary Fig. 2). Leveraging these capabilities, StyleGAN2, along with its predecessor StyleGAN, had been used in generating highly realistic human faces [48], aerial imageries [49], medical images [50], and microstructural images of alloys [51]. Given StyleGAN2's exceptional ability in generating high-quality, diverse images with fine-grained control over style attributes, it was selected in this study to generate realistic synthetic palms. These synthetic palms expanded the dataset and introduced a wide range of variations in appearance, thereby potentially improving the robustness and generalization capability of the palm detection model. In this study, the synthetic palms were generated onto an empty background image, and the annotation masks - essential for pixel-wise class classification, were automatically derived from object boundaries.

While the synthetic palms introduced variability in individual palm characteristics during model development, accurately representing the full spectrum of background and environmental conditions across all possible plantation scenarios remained a challenge. To enhance the models' robustness in diverse settings, we incorporated additional environmental variations using a random selection of drone images from various global flight missions, sourced from OpenAerialMap [34]. This approach aimed to expose the models to a wider range of real-world contexts. However, imprinting individual palms onto these background tiles presented challenges including color inconsistency, resulting in an artificial appearance. To address this, CycleGAN [35] was employed. Known for its image-to-image translation capability, CycleGAN maps images from one domain to another while preserving content. Beyond artistic style transfer, CycleGAN has found applications in X-ray image augmentation [52] and laser-visible image translation [53]. In our case, CycleGAN was used mainly to enhance the realism of synthetic tiles by harmonizing color and lighting conditions, effectively bridging the gap between synthetic and real tiles. While it's acknowledged that a combination of manual mixing techniques can potentially substitute GAN for this purpose [54], this avenue was not explored here.

Data augmentation has been proven valuable in image classification [55] and its extension into object detection has been explored, albeit with a slightly lower impact on accuracy [56]. In addition to conventional augmentations such as random flipping which were done for both datasets in this study, the synthetic tiles generated through GAN-based augmentation were used together with the real tiles to develop the GAN-based palm detector. The DETR framework [22] implemented on top of the Detectron2 object detection framework [23] was used to build the palm detectors used in this study. DETR integrates a transformer encoder-decoder with a CNN backbone. The CNN backbone used for feature extraction was a variation of the ResNeXt architecture [24], which was built upon ResNet [25] with the introduction of the "cardinality" concept. Nevertheless, both these architectures shared similarities in that they were both based on residual learning, which involves the use of bottleneck blocks and skip connections. The cardinality feature of the ResNeXt architecture, which divides the input channels into

multiple groups and perform separate convolutions for each group, helps the model capture diverse features and learn different aspects of the training images in parallel.

The transformer following the CNN backbone, consists of an encoder and decoder [22], is used for global context reasoning. The DETR architecture incorporates a set-based global loss with bipartite matching, enabling pairwise and parallel decoding of object embeddings and simultaneous prediction of object coordinates and class labels. DETR's versatility has been demonstrated across various applications, including medical object and drone-based insulator defect detection [57, 58]. Given the typically structured and dense arrangement of palms in plantations, and that replanting is usually conducted on entire fields, DETR's ability to capture global context and relationships between objects positions it as a particularly fitting choice for our application.

In their respective validation datasets, both the baseline and the GAN-based palm detectors demonstrated strong performance across all detection, segmentation and counting tasks. The GAN-based palm detector achieved impressive precision and recall values, reaching up to 98.5% and 98.6% respectively. These results were comparable, and in some cases, slightly superior to reported values for similar tasks involving various agricultural crops or plants [59, 60]. The mAP and AP values also stood on par with the findings of other relevant research works [61–64]. It is noteworthy, however, that the single-class focus of this study—oil palm—likely contributed to these high-performance metrics.

Upon applying the models to challenge dataset 1, in general a slight decrease in palm detection accuracies was observed. This can be attributed to the dataset's characteristics, which include older palms and overlapping canopies, posing increased challenges for detection. Compared with detection, the segmentation accuracies declined slightly more. This decline can be attributed to the presence of dense canopies, which cast shadows and obstruct the visibility of individual palm canopies in the surrounding area. This limitation hinders effective observation and delineation of the palm canopies during the segmentation process. Comparing palm count accuracies between the baseline and GAN models revealed that the GAN-based model demonstrates a lower susceptibility to false positives. Though not reflected in the mAP and AP metrics, the baseline model displayed a slight inclination to mistakenly identify indistinct shrubs, which loosely resemble palm seedlings from top view, as palms (as illustrated in Fig. 6). It is important to note that the validation and challenge set tiles used in this study were predominantly from our plantations, with most tiles exclusively featuring palms. Instances where tiles contained both shrubs and palms were rare, and tiles exclusively featuring shrubs lacked annotations and resulted in their exclusion from COCO evaluations. Consequently, many of the falsely detected palms on these tiles could only be accounted for in palm count metric. While the precision improvement may not appear statistically significant within the confines of our current dataset, it's important to consider the potential impact in more diverse plantation settings. The GAN-based model's enhanced ability to differentiate palms from other vegetation types suggests improved generalizability, which could prove particularly valuable when applied to plantations with more varied flora. This generalizability can be attributed to the diverse background objects present in the drone background tiles sourced and processed from OpenAerialMap [34].

The challenge dataset 2, collected from a plantation affected by a destructive storm, represented an extreme scenario that served as a test for unforeseen and severe

plantation circumstances. The palms found here were also >5-year-old, similar to challenge dataset 1. The mAP, AP and mAR of the baseline palm model were all lower than 0.15. Though achieving count precision of 100%, the recall was only 13%, indicating that all detected palms were correct; however, the model missed a substantial number of actual palms. As a direct comparison, the GAN-based palm detector was capable of all detecting, segmenting, and counting the palms, achieving comparable accuracies with challenge dataset 1 and its validation set. In many cases (as shown in Fig. 7), the model was even able to detect fallen palms. From the top view, an individual palm canopy appears radial and almost symmetrical. The formation of the canopy is driven by the sequential growth and arrangement of fronds in a spiral pattern, known as phyllotaxis [65, 66]. These fronds are usually packed in an organized spiral, which contributes to the vertical and horizontal dimensions of the canopy. As for the individual fronds of an oil palm, they have a fan-like shape with a central axis and radiating leaflets. While some StyleGAN2-generated synthetic palms resembled the phenotypic outcome of these biological patterns, others did not and generated asymmetrical canopies with irregular fronds. In the challenge dataset 2, the storm has, to a certain extent, altered the canopy shapes and orientation of the fronds. It is likely that the training dataset used to construct baseline model inadequately represented these structural changes. On the other hand, the GAN-based training dataset exhibited a broader range of “possible” palm canopy structures in real-life, thereby achieving high accuracies. The storm-affected palm dataset exemplified an extreme instance of palm canopy variations; typically, distortions to palm canopies in a plantation are usually not as severe. Nevertheless, the results demonstrate that a detection or segmentation model constructed using GAN-generated synthetic tiles in conjunction with raw tiles exhibits superior generalizability and robustness compared to a model relying solely on raw tiles.

While both the GAN-based and baseline models showcase effective performance across various age groups, with the GAN-based model demonstrating the ability to detect palms with canopy distortions, it's important to acknowledge the limitations of our study. Our research focused on a specific set of conditions and did not explore several potentially influential factors. These include variations in drone flight altitudes, oil palm planting densities, drone camera models, and camera settings. Specifically, our study utilized a DJI Mavic 2 Pro, a consumer-grade drone. This choice was driven by its cost-effectiveness and widespread accessibility, aligning with the objectives of large-scale plantation digitalization. Flight altitude was optimized at 80 m above ground level to balance coverage area, battery efficiency, and image resolution. This configuration resulted in a ground sampling distance sufficient for palm detection but slightly limited the quality of the fine canopy details captured. Furthermore, this study primarily concentrated on the *Elaeis guineensis* species of oil palm. Although the developed methods may have some applicability to *Elaeis oleifera*, the extent of this cross-species effectiveness was not directly evaluated in this research.

Future research directions could include expanding the model's capability to accurately detect palms across all age groups and more locations. In addition, the quality of the GAN generators could be significantly enhanced through the use of higher-resolution imagery. Future studies could employ professional-grade drones equipped with advanced sensors, operating at lower altitudes to capture finer ground sampling distances. These improvements would yield more detailed palm canopy feature

representations, potentially improving the accuracy of the GAN generators and the subsequent detection and segmentation models. While this study focuses on oil palms, the challenge of CNN models not generalizing well in crop or weed detection has been highlighted by various researchers [67, 68]. One of the promising solution involves the use of a modified CNN architecture [69]. The GAN-based image augmentation approach presented in this study could be easily integrated as part of a comprehensive solution to improve model generalizability across diverse agricultural contexts.

The GAN-based models developed in this study have demonstrated robust performance in detecting and segmenting palms across a range of previously unseen scenarios, including diverse palm canopy variations and novel environmental contexts. This capability opens up several immediate practical applications in oil palm plantation management. They can be operationalized to automate counting of young palm (<5-year-old) and to phenotype their canopy growth, enabling the identification of abnormal palms. Furthermore, when integrated with hyperspectral or multispectral imagery from drones or satellites, these models can facilitate accurate plant health assessment and early detection of diseased palms [70–72]. This automation forms the foundation of an integrated digital agriculture solution. When coupled with drone-based systems for precise fertilizer and insecticide application [73, 74], it creates a comprehensive approach to plantation management. The implementation of these technologies has the potential to advance the oil palm industry towards a new era of digital agriculture characterized by enhanced automation and precise phenotypic measurement.

Abbreviations

AP	Average Precision
CI	Confidence Interval
CNN	Convolutional neural network
COCO	Common Objects in Context
CycleGAN	Cycle-Consistent Generative Adversarial Network
DETR	Detection Transformer
FID	Fréchet inception distance
GAN	Generative adversarial network
IoU	Intersection over Union
mAP	Mean Average Precision
mAR	Average Recall
StyleGAN2	Style-based Generative Adversarial Network architecture 2
t/ha/yr	tonnes oil per hectare per year

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s40537-024-00990-x>.

Supplementary Material 1

Supplementary Material 2

Acknowledgements

All authors thank the employees of SD Guthrie Research & Upstream Malaysia for their assistance in data collection. Also, they thank the editors and reviewers for their attention to the paper.

Author contributions

QBK was involved in conception and design of the work. QBK, YTK, WRWR, MNAS & SSAR were involved in data acquisition and analysis. QBK, WRWR, MNAS, DRA & HK were involved in data interpretation. QBK, YTK & SSAR were involved in the development of new software/scripts used in the work. QBK drafted the work and all authors substantively revised and approved it.

Funding

This project was funded by SD Guthrie Research Sdn Bhd.

Data availability

The palm GAN generator model has been uploaded and deployed at https://huggingface.co/spaces/qbin85/fake_palm_generator. The image dataset has been uploaded to https://drive.google.com/file/d/1blbmHL_

br4SWwl0g7AaYPI4KEWLUbtm. Other information used during the current study are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 4 January 2024 / Accepted: 26 August 2024

Published online: 08 September 2024

References

1. Lee R. The outlook for population growth. *Science*. 2011;333(6042):569–73.
2. Cros D, Bocs S, Riou V, Ortega-Abboud E, Tisne S, Argout X, et al. Genomic preselection with genotyping-by-sequencing increases performance of commercial oil palm hybrid crosses. *BMC Genomics*. 2017;18(1):839.
3. Kwong QB, Teh CK, Ong AL, Heng HY, Lee HL, Mohamed M, et al. Development and validation of a high-density SNP genotyping array for African Oil Palm. *Mol Plant*. 2016;9(8):1132–41.
4. Kwong QB, Ong AL, Teh CK, Chew FT, Tammi M, Mayes S, et al. Genomic selection in commercial perennial crops: Applicability and Improvement in Oil Palm (*Elaeis guineensis* Jacq). *Sci Rep*. 2017;7(1):2872.
5. Garzón-Martínez GAO-GJAMLPB, Barrero S, Lopez-Cruz LS. Marco; Enciso-Rodríguez, Felix E. Genomic selection for morphological and yield-related traits using genome-wide SNPs in oil palm. *Mol Breeding*. 2022.
6. Crowley MZ. Foreign Labor Shortages in the Malaysian Palm Oil Industry: Impacts and Recommendations. *Research Paper in Economics*. 2020.
7. Inoue Y. Satellite- and drone-based remote sensing of crops and soils for smart farming – a review. *Soil Sci Plant Nutr*. 2020;66(6):798–810.
8. Rejeb A, Abdollahi A, Rejeb K, Treiblmaier H. Drones in agriculture: a review and bibliometric analysis. *Comput Electron Agric*. 2022;198:107017.
9. Kalischuk M, Paret ML, Freeman JH, Raj D, Da Silva S, Eubanks S, et al. An improved crop scouting technique incorporating unmanned aerial vehicle-assisted multispectral crop imaging into conventional scouting practice for Gummy Stem Blight in Watermelon. *Plant Dis*. 2019;103(7):1642–50.
10. Volpato L, Pinto F, Gonzalez-Perez L, Thompson IG, Borem A, Reynolds M, et al. High Throughput Field phenotyping for Plant Height using UAV-Based RGB Imagery in wheat breeding lines: feasibility and validation. *Front Plant Sci*. 2021;12:591587.
11. Chen J, Zhou H, Hu H, Song Y, Gifu D, Li Y, et al. Research on agricultural monitoring system based on convolutional neural network. *Future Generation Comput Syst*. 2018;88:271–8.
12. Lu J, Tan L, Jiang H. Review on Convolutional Neural Network (CNN) Applied to Plant Leaf Disease classification. *Agriculture*. 2021;11(8).
13. Goodfellow IP-A, Mirza J, Xu M, Warde-Farley B, Ozair D, Courville S. Aaron; Bengio, Yoshua. Generative adversarial nets. *Advances in neural information processing systems* 2014. pp. 2672–80.
14. Motamed S, Rogalla P, Khalvati F. Data augmentation using Generative Adversarial Networks (GANs) for GAN-based detection of Pneumonia and COVID-19 in chest X-ray images. *Inf Med Unlocked*. 2021;27:100779.
15. Sandfort V, Yan K, Pickhardt PJ, Summers RM. Data augmentation using generative adversarial networks (CycleGAN) to improve generalizability in CT segmentation tasks. *Sci Rep*. 2019;9(1):16884.
16. Guerrero-Ibañez A, Reyes-Muñoz A. Monitoring Tomato Leaf Disease through Convolutional neural networks. *Electronics*. 2023;12(1).
17. Li M, Zhou G, Chen A, Yi J, Lu C, He M, et al. FWDGAN-based data augmentation for tomato leaf disease identification. *Comput Electron Agric*. 2022;194:106779.
18. Jin H, Li Y, Qi J, Feng J, Tian D, Mu W. GrapeGAN: unsupervised image enhancement for improved grape leaf disease recognition. *Comput Electron Agric*. 2022;198:107055.
19. OpenDroneMap A. ODM – a command line toolkit to generate maps, point clouds, 3D models and DEMs from drone, balloon or kite images <https://github.com/OpenDroneMap/ODM2020>.
20. GDAL Oc. GDAL/OGR Geospatial Data Abstraction software Library. 2023.
21. Torralba A, Russell BC, Yuen J, LabelMe. Online Image Annotation and Applications. *Proceedings of the IEEE*. 2010;98(8):1467–84.
22. Carion N, Massa F, Synnaeve G, Usunier N, Kirillov A, Zagoruyko S, editors. End-to-end object detection with transformers. *Computer vision – ECCV 2020*. Cham: Springer International Publishing; 2020.
23. Wu Y, Kirillov A, Massa F, Lo W-Y, Girshick R. Detectron2. 2019.
24. Xie S, Girshick R, Dollár P, Tu Z, He K, editors. Aggregated Residual Transformations for Deep Neural Networks. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2017.
25. He K, Zhang X, Ren S, Sun J, editors. Deep Residual Learning for Image Recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2016.
26. Lin T-Y, Maire M, Belongie S, Hays J, Perona P, Ramanan D, et al. editors. Microsoft COCO: Common Objects in Context. *Computer Vision – ECCV 2014; 2014*; Cham: Springer International Publishing.

27. Everingham M, Van Gool L, Williams CKI, Winn J, Zisserman A. The Pascal Visual object classes (VOC) challenge. *Int J Comput Vision*. 2010;88(2):303–38.
28. Karras T, Laine S, Aittala M, Hellsten J, Lehtinen J, Aila T, editors. Analyzing and Improving the Image Quality of StyleGAN. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2020.
29. Karras T, Aittala M, Hellsten J, Laine S, Lehtinen J, Aila T. Training Generative Adversarial Networks with Limited Data. *Advances in Neural Information Processing Systems* 2020.
30. Paszke A, Gross S, Chintala S, Chanan G, Yang E, DeVito Z, et al. editors. Automatic differentiation in PyTorch. NIPS 2017 Workshop; 2017.
31. Paszke A, Gross S, Massa F, Lerer A, Bradbury J, Chanan G et al. PyTorch: An Imperative Style, High-Performance Deep Learning Library. *Advances in Neural Information Processing Systems* 32: Curran Associates, Inc.; 2019. pp. 8024-35.
32. Abadi M, Barham P, Chen J, Chen Z, Davis A, Dean J et al. TensorFlow: A System for Large-Scale Machine Learning on Heterogeneous Distributed Systems. *Proceedings of the 12th USENIX Conference on Operating Systems Design and Implementation*. USA: USENIX Association; 2016. pp. 265–83.
33. Pezoa F, Reutter JL, Suarez F, Ugarte M, Vrgoč D, editors. Foundations of JSON schema. *Proceedings of the 25th International Conference on World Wide Web: International World Wide Web Conferences Steering Committee*.
34. OpenAerialMap A. OpenAerialMap <https://openaerialmap.org/2023>.
35. Zhu JY, Park T, Isola P, Efros AA, editors. Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks. 2017 IEEE International Conference on Computer Vision (ICCV); 2017.
36. Hashemi-Beni L, Gebrehiwot A, Karimodini A, Shahbazi A, Dorbu F. Deep convolutional neural networks for weeds and crops discrimination from UAS Imagery. *Front Remote Sens*. 2022;3.
37. Boulent J, Foucher S, Theau J, St-Charles PL. Convolutional neural networks for the Automatic Identification of Plant diseases. *Front Plant Sci*. 2019;10:941.
38. Freudenberg M, Nölke N, Agostini A, Urban K, Wörgötter F, Kleinn C. Large Scale Palm Tree Detection in High Resolution Satellite images using U-Net. *Remote Sens*. 2019;11(3).
39. Li W, Fu H, Yu L, editors. Deep convolutional neural network based large-scale oil palm tree detection for high-resolution remote sensing images. 2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS); 2017.
40. Kipli K, Osman S, Joseph A, Zen H, Awang Salleh DNSD, Lit A, et al. Deep learning applications for oil palm tree detection and counting. *Smart Agricultural Technol*. 2023;5:100241.
41. Kwong QB, Wong YC, Lee PL, Sahaini MS, Kon YT, Kulaveerasingam H, et al. Automated stomata detection in oil palm with convolutional neural network. *Sci Rep*. 2021;11(1):15210.
42. Chuang Y, Zhang S, Zhao X. Deep learning-based panoptic segmentation: recent advances and perspectives. *IET Image Processing*; 2023.
43. Puliti S, Ørka HO, Gobakken T, Næsset E. Inventory of small forest areas using an unmanned aerial system. *Remote Sens*. 2015;7(8):9632–54.
44. Domingo D, Ørka HO, Næsset E, Kachamba D, Gobakken T. Effects of UAV Image Resolution, Camera Type, and Image Overlap on Accuracy of Biomass predictions in a Tropical Woodland. *Remote Sens*. 2019;11(8).
45. Duan H, Liu Y, Huang H, Wang Z, Zhao H. Image Stitching Algorithm for drones based on SURF-GHT. *IOP Conf Series: Mater Sci Eng*. 2019;569(5):052025.
46. Bouchekara HREH, Sadiq BO, O Zakariyya S, Sha'aban YA, Shahriar MS, Isah MM. SIFT-CNN Pipeline in Livestock Management: a Drone Image Stitching Algorithm. *Drones*. 2023;7(1).
47. Hedayati H, McGuinness BJ, Cree MJ, Perrone JA, editors. Generalization Approach for CNN-based Object Detection in Unconstrained Outdoor Environments. 2019 International Conference on Image and Vision Computing New Zealand (IVCNZ); 2019.
48. Meira N, Silva M, Bianchi A, Rabelo R. Generating Synthetic Faces for Data Augmentation with StyleGAN2-ADA. *International Conference on Enterprise Information Systems* 2023. pp. 649 – 55.
49. Yates M, Hart G, Houghton R, Torres MT, Pound M. Evaluation of synthetic aerial imagery using unconditional generative adversarial networks. *ISPRS J Photogrammetry Remote Sens*. 2022;190:231–51.
50. Tariq U, Qureshi R, Zafar A, Aftab D, Wu J, Alam T, et al. editors. Brain Tumor Synthetic Data Generation with adaptive StyleGANs. Cham: Springer Nature Switzerland: Artificial Intelligence and Cognitive Science; 2023.
51. Lambard G, Yamazaki K, Demura M. Generation of highly realistic microstructural images of alloys from limited data with a style-based generative adversarial network. *Sci Rep*. 2023;13(1):566.
52. Bargshady G, Zhou X, Barua PD, Gururajan R, Li Y, Acharya UR. Application of CycleGAN and transfer learning techniques for automated detection of COVID-19 using X-ray images. *Pattern Recognit Lett*. 2022;153:67–74.
53. Qin M, Fan Y, Guo H, Wang M. Application of Improved CycleGAN in laser-visible Face Image translation. *Sensors*. 2022;22(11).
54. Wyawahare M, Ekbote N, Pimperkhede S, Deshpande A, Bapat P, Aphale I, editors. Comparison of image blending using cycle GAN and Traditional Approach. Singapore: Springer Nature Singapore: Pervasive Computing and Social Networking; 2023.
55. Shorten C, Khoshgoftaar TM. A survey on Image Data Augmentation for Deep Learning. *J Big Data*. 2019;6(1):60.
56. Zoph B, Cubuk ED, Ghiasi G, Lin T-Y, Shlens J, Le QV, editors. Learning Data Augmentation strategies for object detection. *Computer vision – ECCV 2020*. Cham: Springer International Publishing; 2020.
57. Ickler MK, Baumgartner M, Roy S, Wald T, Maier-Hein KH, editors. Taming Detection Transformers for Medical Object Detection. *Bildverarbeitung für die Medizin 2023*; 2023; Wiesbaden: Springer Fachmedien Wiesbaden.
58. Cheng Y, Liu D. An image-based Deep Learning Approach with Improved DETR for Power line insulator defect detection. *J Sens*. 2022;2022:6703864.
59. Zhao W, Yamada W, Li T, Digman M, Runge T. Augmenting crop detection for Precision Agriculture with Deep Visual transfer Learning—A case study of Bale Detection. *Remote Sens*. 2021;13(1).
60. Morales G, Kemper G, Sevillano G, Arteaga D, Ortega I, Telles J. Automatic segmentation of *Mauritia flexuosa* in unmanned aerial vehicle (UAV) Imagery using deep learning. *Forests*. 2018;9(12).
61. Cai Z, Vasconcelos N, Cascade R-CNN, editors. Delving Into High Quality Object Detection. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2018.

62. Ren S, He K, Girshick R, Sun J, Faster R-CNN. Towards real-time object detection with region proposal networks. In: Cortes C, Lawrence N, Lee D, Sugiyama M, Garnett R, editors. *Advances in neural information Processing systems*. Curran Associates, Inc.; 2015.
63. Cao D, Chen Z, Gao L. An improved object detection algorithm based on multi-scaled and deformable convolutional neural networks. *Human-centric Comput Inform Sci*. 2020;10(1):14.
64. Zhao L, Li S. Object detection Algorithm based on improved YOLOv3. *Electronics*. 2020;9(3).
65. Aholoukpè HNS, Dubos B, Deleporte P, Flori A, Amadji LG, Chotte J-L, et al. Allometric equations for estimating oil palm stem biomass in the ecological context of Benin, West Africa. *Trees*. 2018;32(6):1669–80.
66. Thomas RL, Chan KW, Easau PT. Phyllotaxis in the Oil Palm: arrangement of fronds on the trunk of mature palms. *Ann Botany*. 1969;33(5):1001–8.
67. Wang A, Zhang W, Wei X. A review on weed detection using ground-based machine vision and image processing techniques. *Comput Electron Agric*. 2019;158:226–40.
68. Lottes P, Behley J, Milioto A, Stachniss C. Fully Convolutional Networks with Sequential Information for Robust Crop and Weed Detection in Precision Farming. *IEEE Rob Autom Lett*. 2018;3:2870–7.
69. Albattah W, Javed A, Nawaz M, Masood M, Albahli S. Artificial Intelligence-based Drone System for Multiclass Plant Disease Detection Using an improved efficient convolutional neural network. *Front Plant Sci*. 2022;13:808380.
70. Abbas A, Zhang Z, Zheng H, Alami MM, Alrefaei AF, Abbas Q et al. Drones in Plant Disease Assessment, efficient monitoring, and detection: a Way Forward to Smart Agriculture. *Agronomy*. 2023;13(6).
71. Abdulridha J, Ampatzidis Y, Roberts P, Kakarla SC. Detecting powdery mildew disease in squash at different stages using UAV-based hyperspectral imaging and artificial intelligence. *Biosyst Eng*. 2020;197:135–48.
72. Chin R, Catal C, Kassahun A. Plant disease detection using drones in precision agriculture. *Precision Agric*. 2023;24(5):1663–82.
73. Khan S, Tufail M, Khan MT, Khan ZA, Iqbal J, Wasim A. Real-time recognition of spraying area for UAV sprayers using a deep learning approach. *PLoS ONE*. 2021;16(4):e0249436.
74. Hafeez A, Husain MA, Singh SP, Chauhan A, Khan MT, Kumar N, et al. Implementation of drone technology for farm monitoring & pesticide spraying: a review. *Inform Process Agric*. 2023;10(2):192–203.

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.