Benchmark Dataset for AI-Driven Palm Tree Detection and Analysis in the UAE

Mina Al-Saad [1], Nour Aburaed [1], Mohammed Q. Alkhatib [1], Mohammed S. Zitouni [1], Saeed Almansoori [2], Hussain Al-Ahmad [1]

[1] College of Engineering and IT, University of Dubai, Dubai, UAE - (minaalsaad, nour.aburaed, mqalkhatib)@ieee.org, (mzitouni, halahmad)@ud.ac.ae

[2] Mohammed Bin Rashid Space Centre, Dubai , UAE - saeed.almansoori@mbrsc.ae

Keywords: Object Detection, Convolutional Neural Networks, Transformers, High Resolution, UAV

Abstract

Automated detection and counting of palm trees is a significant area of research for numerous countries, including the UAE. Currently, all palm trees are counted and monitored manually, a labor-intensive process that demands significant time and effort. However, the UAE has recently made significant advancements in remote sensing technologies, creating an opportunity to integrate space technology with agriculture for the efficient monitoring of palm trees throughout the country. This research paper presents a novel High-Resolution (HR) remote sensing dataset designed for the autonomous detection of palm trees in the UAE. The dataset has been acquired using Unmanned Aerial Vehicles (UAVs) covering various areas within UAE, including Ajman, Dubai, Khorfakkan, and Al-Ain. This paper utilizes the introduced dataset to evaluate the strengths and weaknesses of four object detection neural networks; You Only Look Once (YOLO)-v4 and -v5, Faster Region-based Convolutional Neural Network (FRCNN), and Detection Transformer (DETR). YOLOv5s achieved outstanding performance, with an Average Precision (AP) of 96.6% and Average F1-score (AF) of 95.7%, demonstrating its effectiveness in accurately detecting and localizing palm trees. Moreover, model outputs are effectively integrated into Geographic Information Systems (GIS) for enhanced spatial analysis and monitoring.

1. Introduction

The date palm tree is considered one of the most ancient trees in the Arabian Peninsula, the Middle East, and North Africa. According to Food and Agriculture Organisation of the United Nations (FAO), around 70% of the world’s date palm trees are concentrated in Arab countries. The UAE is home to over 40 million palm trees, with more than 4 million situated in the Al Ain region (Raza et al., 2022). Performing a tree inventory through field-based measurements is a labor-intensive, time-consuming, and prone-to-error process. This is particularly challenging given the vast size of palm plantations, which are difficult to monitor effectively from the ground (Culman et al., 2020). Nowadays, advancements in High-Resolution (HR) satellites and Unmanned Aerial Vehicles (UAVs) technologies have made the automatic detection of date palms from HR remote sensing images one of the most popular methods for detecting and counting palm trees. UAVs have recently gained popularity across various applications in agriculture and forestry due to their flexibility, cost-effectiveness, and capacity to cover large areas and provide ultra-high-resolution images (Gibril et al., 2024). Recent studies indicate that the primary methods for identifying and detecting palm trees mainly include traditional image processing as well as machine learning and Deep Learning (DL) approaches. Despite the significant progress in DL-based object detection methods for remote sensing imagery, a major challenge remains: the lack of publicly available labeled datasets for palm tree detection. This shortcoming restricts the development and evaluation of effective detection models for this specific application.

This study introduces a novel palm tree dataset gathered from various regions in the UAE, including Al Ain, Khorfakkan, Ajman, and Dubai. The dataset includes samples that vary from simple to more complex scenarios. For example, some images depict palm trees that are evenly spaced, while others show trees that are densely packed and overlapping. Additionally, certain images have minimal shadows around the palm trees, whereas others display strong shadows enveloping each tree. Furthermore, while some images exclusively feature palm trees, others include a mix of palm trees and various types of vegetation. The dataset also encompasses palm trees of different sizes to ensure a comprehensive range of samples.

Subsequently, the dataset undergoes preprocessing, labeling, and comprehensive quality assurance checks. The dataset is then used to evaluate the performance of four DL object detection models, highlighting both their strengths and weaknesses. The models are You Only Look Once (YOLO)-v4 and -v5s, Faster Region-based Convolutional Neural Network (FRCNN), and Detection Transformer (DETR). Moreover, the models’ predictions are seamlessly integrated into a Geographic Information System (GIS) software by exporting them in shapefile format.

The remainder of the paper is organized as follows: Section 2 presents the literature review of object detection based on DL approaches in the field of remote sensing, Section 3 explains the study area and dataset collection used in this research work, Section 4 discusses the four object detection models being trained and evaluated, Section 5 illustrates and discusses the results, and finally, Section 6 summarizes and concludes the paper.

2. Literature Review

In the recent decades, numerous methods have been developed for object detection in aerial and satellite images. Several studies in the literature focus on the autonomous counting and detection of palm trees through HR remote sensing imagery, utilizing image processing and machine learning techniques. Almaazmi (AlMaazmi, 2018) used WorldView-3 satellite images to develop an algorithm for detecting and counting palm trees in the UAE. The process involved two main stages: first, palm trees were detected through supervised maximum likelihood

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-X-G-2025-95-2025

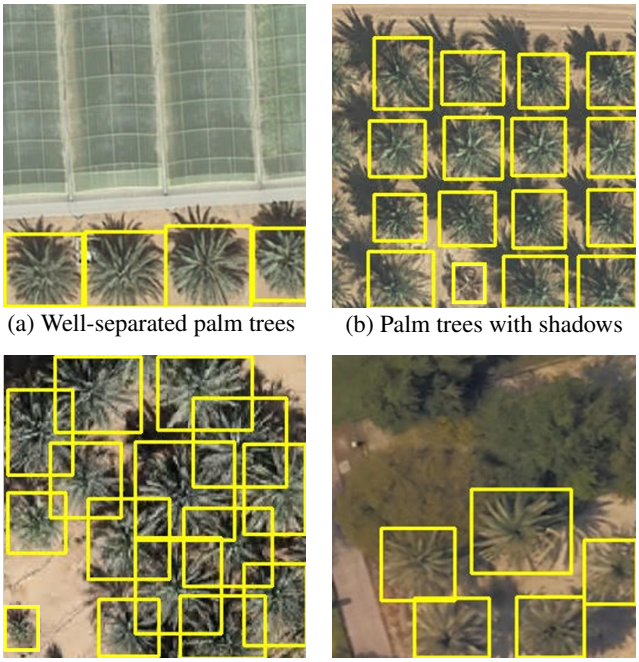
Figure 1. The study area and regions for dataset sample collection.

classification using Red, Blue, Green, and Near-Infrared (NIR) bands; second, the trees were counted by extracting local spatial maxima from NDVI masks. The algorithm was tested in various areas of the Al Ain region, achieving an overall accuracy of 89%. In another research, the authors (Al Mansoori et al., 2018) utilized UAV imagery to identify and count palm trees in the UAE by employing spectral information and morphological operations. They effectively leveraged the Normalized Difference Vegetation Index (NDVI) and the histogram-equalized Y channel from the YCbCr color space. By integrating this data with Canny edge detection and evaluating the roundness of the objects, they successfully classified the objects as palm trees or non-palm trees. The performance of the algorithm was evaluated using precision, recall, and F1-score metrics, which are 97.1%, 95.7%, 96.0%, respectively.

Recently, DL techniques have demonstrated promising results in addressing object detection challenges within the realm of remote sensing imagery (Aljishi et al., 2023). Researchers in (Li et al., 2016) introduced a Convolutional Neural Network (CNN) detection system utilizing a sliding window approach to localize and classify palm trees in Malaysia, achieving an accuracy of 96%. The results demonstrated that the proposed CNN detector outperformed both the local maximum filter and template matching methods. Jintasuttisak (Jintasuttisak et al., 2022) applied the advanced CNN model YOLOv5, available in multiple sizes, to detect date palm trees in aerial images captured by a drone at a height of 122 meters over farmlands in the Northern Emirates, UAE. They compared YOLOv5's performance with other CNN models, such as YOLOv3, YOLOv4, and SSD300, finding that YOLOv5 achieved the highest accuracy, with a mean average precision of 92.34%. This indicates its strong capability in detecting palm trees in both dense and sparse environments. Another research (Al-Saad et al., 2022) showcased the superiority of YOLOv4 architecture, which surpassed FRCNN in detecting date palms using UAV imagery in Al Ain, UAE.

3. Study Area and Dataset Collection

In this study, aerial RGB images of different sizes with a spatial resolution of 10cm were collected from various municipalities across the UAE. These images were divided into patches of 256×256. The dataset includes images from diverse regions, including Ajman, Dubai, Khorfakkan, and Al Ain. Figure 1 illustrates the study area for this research. All palm trees in these images were manually annotated in Pascal VOC, COCO, and YOLO formats. A total of 2,976 images containing 21,935 palm trees were compiled. Out of these, 1,905 images with 13,938 palm trees were designated for model training, while 476 images featuring 3,492 palm trees were reserved for validation during training. To assess the model's performance, 595 images containing 4,505 palm trees were used. Sample images from the dataset are shown in Figure 2. A portion of the dataset for the Al Ain area is publicly available on GitHub at https://github.com/Nour093/Palm-Tree-Dataset. The dataset contains a diverse range of samples as shown in Figure 2, from simple to more complex scenarios. For instance, some images show well-separated palm trees, while others depict densely packed and overlapping trees. Additionally, certain images feature minimal shadowing, whereas others display strong shadows around each tree. In some cases, only palm trees are present, while others include mixed vegetation. Moreover, some palm trees are fully visible, while others are partially cropped or obstructed. Palm trees are labeled if at least 50% of the tree is visible. Shadows and occlusions present a significant challenge for object detection, especially in palm tree detection.



(a) Well-separated palm trees (b) Palm trees with shadows

(c)Overlapping palm trees (d)Trees mixed with other vegetation

Figure 2. Samples from the proposed palm tree dataset.

4. Methodology

This section presents the DL-based object detection architectures employed for training and evaluating the proposed dataset; Figure 3 illustrates the flowchart of the entire process, from preparing the data to training and evaluating the model, and finally integrating predictions with GIS.

4.1 You Only Look Once (YOLO) Series

YOLO is a one-stage convolutional neural network specifically designed for object detection tasks. Ever since its first introduced by Redmon et al. in 2016 (Redmon et al., 2016), the YOLO series has attracted considerable attention for its impressive speed and accuracy in object detection (Zheng et al., 2020). YOLO simplifies the object detection process by approaching it as a direct regression problem. This allows the network to simultaneously predict both bounding boxes and

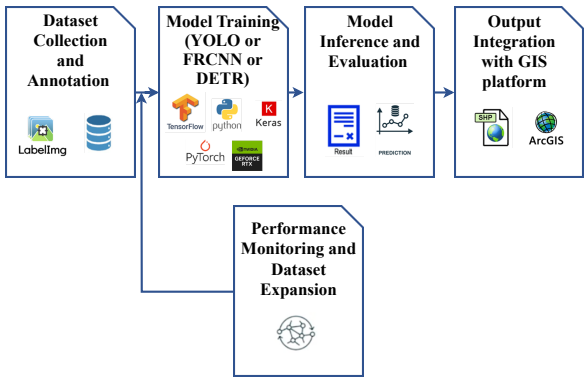


Figure 3. Flowchart of methodology framework.

their class probabilities in a single, unified process. This paper explores two versions of the YOLO series: YOLOv4 and YOLOv5.

4.1.1 YOLOv4: In 2020, YOLOv4 was introduced as an advancement over YOLOv3 (Hussain, 2024, Bochkovskiy et al., 2020), standing out for its remarkable performance and high operational speed. YOLOv4 is composed of four key components: the backbone network, CSPDarkNet53, for feature extraction; the Spatial Pyramid Pooling (SPP) module, which enriches high-level semantic features; and the Path Aggregation Network (PANet), which enables multi-scale feature fusion, improving the network's ability to detect objects at different sizes. Additionally, three YOLO heads are used to predict object locations and classes (Yu, 2022).

4.1.2 YOLOv5: YOLOv5 network architecture consists of four main components: input, backbone, neck, and prediction. There are five versions of the network YOLOv5n, YOLOv5s, YOLOv5m, YOLOv5l, and YOLOv5x, these models share the same architecture but vary in depth and width. YOLOv5s has the smallest depth and width, leading to the highest processing speed. In this paper, YOLOv5s is selected as the base model (Cao et al., 2023).

4.2 Faster-RCNN (FRCNN)

Region-based CNN (RCNN) is a collection of machine-learning architectures specifically created for object detection tasks. It was first introduced in 2014 by Girshick et al. (Girshick et al., 2014) as an answer to the PASCAL VOC Challenge. RCNN consists of three distinct components: feature extraction, bounding box classification, and a model for adjusting the position and size of the bounding boxes. Although RCNN is reliable, it cannot be trained in an end-to-end fashion due to its multi-stage architecture. To overcome this limitation, Fast RCNN (Girshick, 2015) was developed, which consolidates the three components into a single model and incorporates a pooling layer into its architecture. However, Fast RCNN still utilizes the Selective Search method to determine Regions of Interest (ROIs), which can be a time-consuming process. To address this challenge, Faster R-CNN (FRCNN) (Ren et al., 2016) was introduced as an enhanced version of Fast RCNN, featuring a Region Proposal Network (RPN). This innovation enables the model to eliminate the need for Selective Search, resulting in faster training and detection times and ultimately improving overall performance.

4.3 DEtection TRansformer (DETR)

Transformers leverage attention mechanisms to capture global dependencies while filtering out irrelevant information. Initially used in Natural Language Processing (NLP), Transformers were later applied to vision tasks with the introduction of the Vision Transformer (ViT) for image recognition(Dai et al., 2021). In 2020, Facebook researchers extended this concept to object detection with DETR (DEtection TRansformer) (Carion et al., 2020), marking a significant advancement in the field. DETR combines the strengths of CNNs and Transformers through a hybrid architecture that eliminates the need for traditional region proposals and anchor boxes used in CNN-based detectors.

DETR's architecture is composed of three main components: a CNN backbone (such as ResNet) for initial feature extraction, an encoder-decoder Transformer to capture the contextual relationships, and a Feed Forward Network (FFN) for final predictions. The backbone generates feature maps, which the Transformer encoder processes using multi-head self-attention and FFNs to reduce dimensionality while capturing complex dependencies. The decoder, consisting of multiple layers with self-attention and cross-attention mechanisms, refines object queries and generates object candidates. Each decoder layer is followed by FFNs that predict object classes and bounding boxes, allowing the model to simultaneously identify and localize objects across the entire image (Kong et al., 2024).

5. Results and Discussion

In this research, four key architectures (YOLOv4, YOLOv5s, DETR, and FRCNN) were rigorously trained and tested using the carefully curated dataset outlined in Section 3. For a fair and unbiased comparison, all models were trained and evaluated within a consistent and controlled environment. The optimization process employed the widely-used Adam optimizer. The learning rate was empirically set to 10^{-4}, and each model was trained for 200 epochs to allow sufficient time for convergence and to achieve optimal performance.

A series of objective quantitative evaluation metrics were used to assess the effectiveness and performance of the trained models. These metrics, including Average Precision (AP), Average Recall (AR), and Average F1-score (AF), which are defined in Equations 1 - 3 as follows:

$$AP = \sum_{n=1}^{n=N} \left(\frac{TP_n}{TP_n + FP_n} \right) \div N, \tag{1}$$

$$AR = \sum_{n=1}^{n=N} \left(\frac{TP_n}{TP_n + FN_n} \right) \div N, \tag{2}$$

$$AF = \sum_{n=1}^{n=N} \left(2 \times \frac{precision \times recall}{precision + recall} \right) \div N, \tag{3}$$

where n is the index of the image being evaluated, and N is the total number of testing images, which is 595 in this case. All equations rely on determining True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) cases. TP refers to instances where objects are correctly detected as belonging to the target class, while TN represents instances

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-X-G-2025-95-2025 | © Author(s) 2025. CC BY 4.0 License.

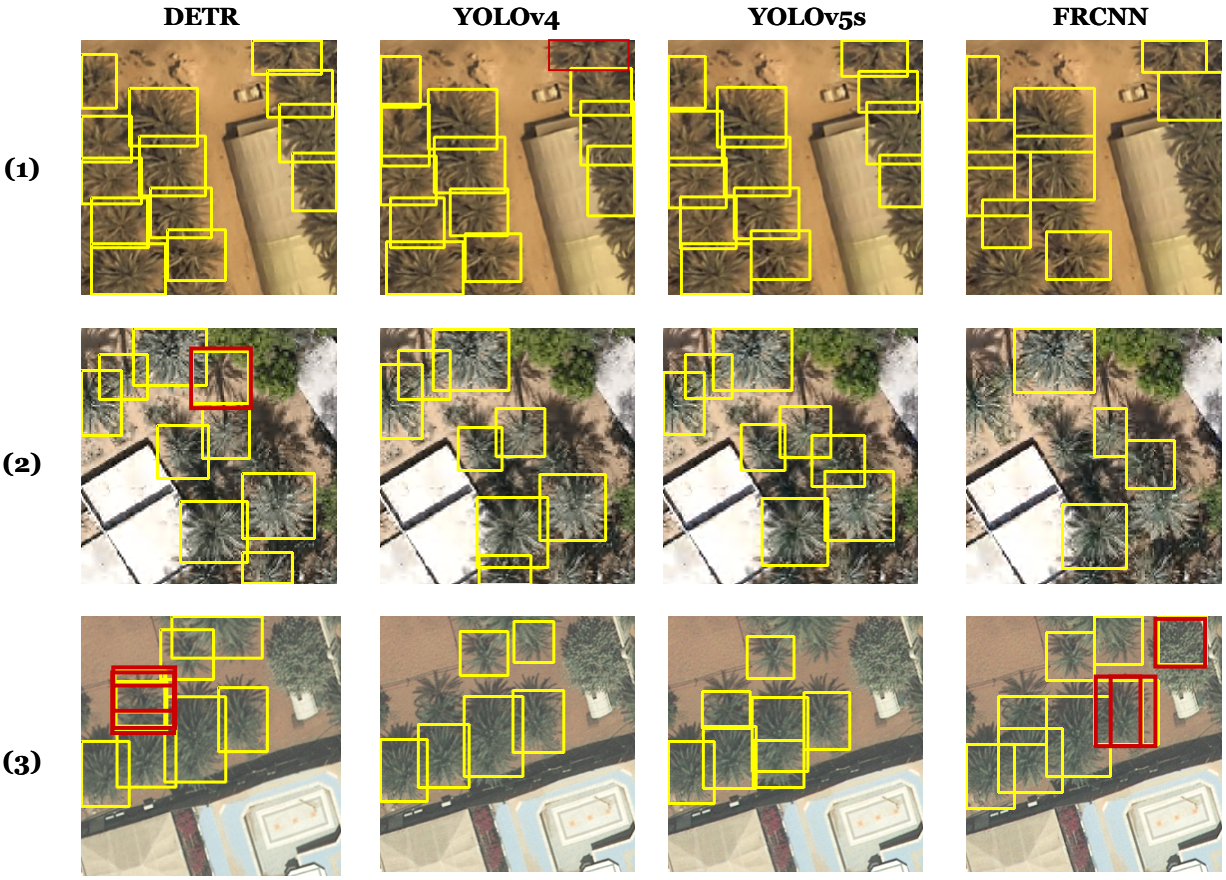


Figure 4. Visual results of palm tree detection produced by YOLOv4, YOLOv5s, DETR, and Faster R-CNN

Table 1. Results summary of each model's performance in terms of Average Precision (AP), Average Recall (AR), and Average F1-score (AF).

Model	AP (%)	AR (%)	AF (%)
YOLOv4	94.09	94.32	94.20
YOLOv5s	**96.60**	94.90	**95.74**
FRCNN	79.54	78.31	78.93
DETR	91.58	**97.21**	93.67

where no objects are present and are correctly classified as such. FP occurs when non-target objects are incorrectly identified as belonging to the target class, and FN represents target objects that the model fails to detect.

Table 1 summarizes all the quantitative evaluation results for all networks. YOLOv5s achieves the highest AP, which indicates that it effectively minimizes FP cases. It also scores the highest AF, which shows its well-balanced performance in accurately identifying and localizing palm trees. YOLOv4 ranks second in both AP and AF. For AR, YOLOv5s takes second place, while DETR leads with the highest AR. This reflects DETR's low rate of FN cases. However, DETR does not perform well in terms of AP and AF. FRCNN scores the lowest in terms of all quantitative metrics. Overall, YOLOv5s demonstrates the best performance among the models, as it scores the highest in two metrics out of three.

Visually examining the results provides insight about the shortcomings and strengths of each network as shown in Figure 4. The observations are consistent with those seen in Table 1. For instance, in Sample 1, YOLOv4 fails to detect a palm tree (indicated in red), while all other models accurately identify it. Conversely, in Sample 2, DETR incorrectly identifies shadows (marked in red) as palm trees, a misclassification not seen in the other three models. Additionally, Sample 3 illustrates duplicate detections (red bounding box) generated by DETR and FRCNN. For the same sample, FRCNN detects non-palm vegetation (marked in red). This issue is not seen in both YOLOv4 and YOLOv5s results. Consistent with the quantitative results, FRCNN demonstrates suboptimal performance on a substantial portion of the test dataset, while YOLOv5s shows the best performance. Duplicate detections and false positive cases could be minimized by fine-tuning the Non-Maximum Suppression (NMS) threshold or by employing more advanced post-processing techniques to effectively handle these instances.

The final result is integrated with GIS by exporting the prediction results to shapefile that shows a point feature in the center of each detected palm tree, representing latitude and longitude of each date palm, as shown in Figure 5.

Furthermore, for testing purposes, we assessed the performance of YOLOv4 in spatial monitoring, with the model successfully detecting $\sim 74\%$ of the trees. This detection rate represents a

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-X-G-2025-95-2025 | © Author(s) 2025. CC BY 4.0 License.

Figure 5. Output integration with GIS. Each detected palm tree is marked with a point shape at the center.

significant improvement in decision-making processes, optimizing resource allocation, and enabling more accurate yield estimation. Integrating the model’s output with GIS could significantly reduce several days of manual labor, benefiting not only farmers but also policymakers and relevant government agencies, empowering them to make more informed and effective decisions.

6. Conclusion

The study presented a new, HR remote sensing dataset acquired via UAVs for automated palm tree detection in the UAE. Four object detection models — YOLOv4, YOLOv5s, FRCNN, and DETR — were trained and tested using this dataset. YOLOv5s achieved the highest performance in AP and AF, which proves its effectiveness in accurately detecting and localizing palm trees. DETR demonstrated strong recall but had limitations with FP cases, while FRCNN underperformed overall. The findings prove YOLOv5s’ suitability for palm tree detection tasks, with model outputs effectively integrated into GIS for enhanced spatial analysis and monitoring. This research supports the UAE’s agricultural monitoring efforts and demonstrates the potential of HR UAV datasets in facilitating advancements in remote sensing applications. Future work can focus on utilizing newer YOLO variants (e.g., YOLOv8) or hybrid approaches that combine multiple models to further boost the detection performance.

References

Al Mansoori, S., Kunhu, A., Al Ahmad, H., 2018. Automatic palm trees detection from multispectral uav data using normalized difference vegetation index and circular hough transform. *High-Performance Computing in Geoscience and Remote Sensing VIII*, 10792, SPIE, 11–19.

Al-Saad, M., Aburaed, N., Al Mansoori, S., Al Ahmad, H., 2022. Autonomous palm tree detection from remote sensing images-uae dataset. *IGARSS 2022-2022 IEEE International Geoscience and Remote Sensing Symposium*, IEEE, 2191–2194.

Aljishi, F., Alharbi, A., De Weck, O., Habib, A., 2023. A hybrid pipeline for date palm tree detection in high-resolution satellite imagery. *IGARSS 2023-2023 IEEE International Geoscience and Remote Sensing Symposium*, IEEE, 6572–6575.

AlMaazmi, A., 2018. Palm trees detecting and counting from high-resolution worldview-3 satellite images in united arab emirates. *Remote Sensing for Agriculture, Ecosystems, and Hydrology XX*, 10783, SPIE, 387–397.

Bochkovskiy, A., Wang, C.-Y., Liao, H.-Y. M., 2020. YOLOv4: Optimal Speed and Accuracy of Object Detection. *ArXiv*, abs/2004.10934. https://api.semanticscholar.org/CorpusID:216080778.

Cao, F., Xing, B., Luo, J., Li, D., Qian, Y., Zhang, C., Bai, H., Zhang, H., 2023. An Efficient Object Detection Algorithm Based on Improved YOLOv5 for High-Spatial-Resolution Remote Sensing Images. *Remote Sensing*, 15(15), 3755.

Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A., Zagoruyko, S., 2020. End-to-end object detection with transformers. *European conference on computer vision*, Springer, 213–229.

Culman, M., Delalieux, S., Van Tricht, K., 2020. Individual palm tree detection using deep learning on RGB imagery to support tree inventory. *Remote Sensing*, 12(21), 3476.

Dai, Z., Cai, B., Lin, Y., Chen, J., 2021. Up-detr: Unsupervised pre-training for object detection with transformers. *Proceedings*

of the IEEE/CVF conference on computer vision and pattern recognition, 1601–1610.

Gibril, M. B. A., Shafri, H. Z. M., Shanableh, A., Al-Ruzouq, R., bin Hashim, S. J., Wayayok, A., Sachit, M. S., 2024. Large-scale assessment of date palm plantations based on UAV remote sensing and multiscale vision transformer. *Remote Sensing Applications: Society and Environment*, 34, 101195.

Girshick, R., 2015. Fast r-cnn. *2015 IEEE International Conference on Computer Vision (ICCV)*, 1440–1448.

Girshick, R., Donahue, J., Darrell, T., Malik, J., 2014. Rich feature hierarchies for accurate object detection and semantic segmentation. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 580–587.

Hussain, M., 2024. Yolov1 to v8: Unveiling each variant–a comprehensive review of yolo. *IEEE Access*, 12, 42816–42833.

Jintasuttisak, T., Edirisinghe, E., Elbattay, A., 2022. Deep neural network based date palm tree detection in drone imagery. *Computers and Electronics in Agriculture*, 192, 106560.

Kong, Y., Shang, X., Jia, S., 2024. Drone-DETR: Efficient Small Object Detection for Remote Sensing Image Using Enhanced RT-DETR Model. *Sensors*, 24(17), 5496.

Li, W., Fu, H., Yu, L., Cracknell, A., 2016. Deep learning based oil palm tree detection and counting for high-resolution remote sensing images. *Remote sensing*, 9(1), 22.

Raza, M., Abdallah, H. A., Abdullah, A., Abu-Jdayil, B., 2022. Date palm surface fibers for green thermal insulation. *Buildings*, 12(6), 866.

Redmon, J., Divvala, S., Girshick, R., Farhadi, A., 2016. You only look once: Unified, real-time object detection. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 779–788.

Ren, S., He, K., Girshick, R., Sun, J., 2016. Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE transactions on pattern analysis and machine intelligence*, 39(6), 1137–1149.

Yu, X., 2022. Remote sensing object detection based on improved yolov4. *2022 IEEE 5th International Conference on Automation, Electronics and Electrical Engineering (AUTEEE)*, IEEE, 570–573.

Zheng, Z., Lei, L., Sun, H., Kuang, G., 2020. A review of remote sensing image object detection algorithms based on deep learning. *2020 IEEE 5th International Conference on Image, Vision and Computing (ICIVC)*, IEEE, 34–43.