This is a peer-reviewed, author's accepted manuscript of the following conference proceedings output: Al-Saad, M., Aburaed, N., Mansoori, S. A., & Ahm

Autonomous Palm Tree Detection from Remote Sensing Images - UAE Dataset

Mina Al-Saad*, Nour Aburaed†, Saeed Al Mansoori‡, Hussain Al Ahmad§
*†§ College of Engineering and IT, University of Dubai, UAE
‡Mohammed Bin Rashid Space Centre, Dubai, UAE
Email: *minaalsaad,†nour.aburaed@ieee.org

***Abstract*—Autonomous detection and counting of palm trees is a research field of interest to various countries around the world, including the UAE. Automating this task saves effort and resources by minimizing human intervention and reducing potential errors in counting. This paper introduces a new High Resolution (HR) remote sensing dataset for autonomous detection of palm trees in the UAE. The dataset is collected using Unmanned Aerial Vehicles (UAV), and it is labeled properly in PASCAL VOC and YOLO formats after pre-processing and visually inspecting its quality. A comparative evaluation between Faster-RCNN and YOLOv4 networks is then conducted to observe the usability of the dataset in addition to the strengths and weaknesses of each network. The dataset is publicly available at https://github.com/NourO93/Palm-Tree-Dataset.**

***Index Terms*—Remote Sensing, Object Detection, Convolutional Neural Networks, YOLOv4, FRCNN**

I. Introduction

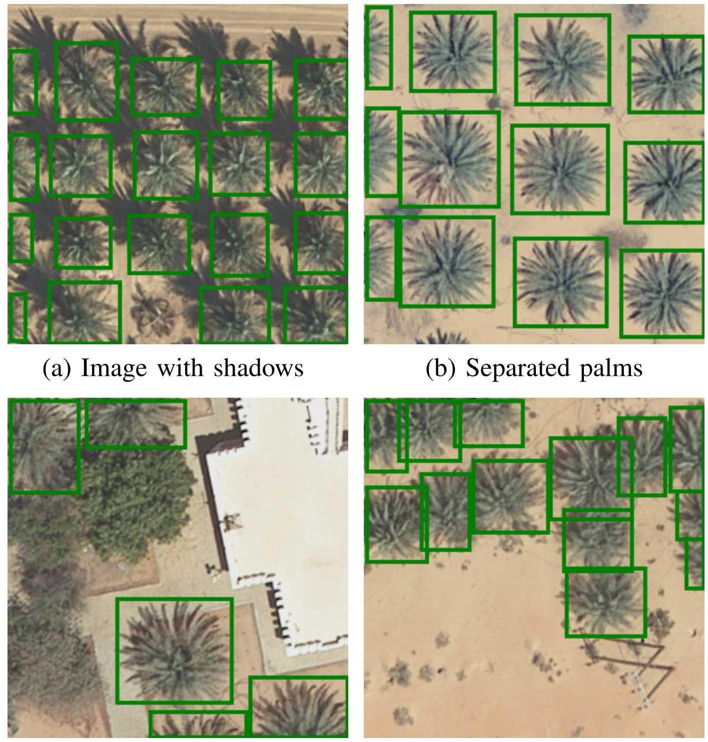
Palm trees are an important cultural and heritage symbol in middle eastern countries, including KSA, Egypt, Oman, Algeria, and UAE. According to FAOStatistical Database, those are some of the countries that dominate world dates trade [1]. The UAE was in the top 6 countries that produced the biggest annual production of fresh dates for the year 2019. The UAE relies on dates as one of its major non-oil exports. Therefore, counting and observing the number of palm trees in the UAE is a crucial task for monitoring production growth and determining plantation layouts. As of 2018, the number of palm trees across the UAE exceeded 40 million. All palm trees are counted and monitored manually; a daunting task that requires tremendous time and effort. Recently, the UAE has been prospering in the field of remote sensing, which introduces the opportunity of combining space technology with agriculture to monitor the palm trees in the UAE. There are several approaches in the literature related to the autonomous counting and monitoring of palm trees from High Resolution (HR) remote sensing imagery. For example, Srestasathiern et al. [2] conducted a study in Thailand related to counting palm trees from satellite images, where they used QuickBird and WorldView-2 imagery and manually cropped plantation areas. Then, they computed vegetation index, detected local peaks, and selected features to determine whether the detected object is a palm tree or not. The performance of their approach was assessed using precision, recall, and F-score, which all scored a value of 98.1% on average. Another example can be seen in [3], where the authors detect and count palm trees in the UAE from Unmanned Aerial Vehicles (UAV) using spectral information and morphological operations. The authors make efficient use of NDVI with histogram equalized Y channel extracted from YCbCr color domain, and then use this information in conjunction with Canny edge detection and measuring the roundness of the detected object to classify it into palm trees or otherwise. The algorithm is assessed using precision, recall, and F-score, which are 97.1%, 95.7%, 96.0%, respectively. The formerly discussed examples are considered as traditional approaches. The shortcoming of such approaches is that they work on a specific type of images or palm trees only, which means that they are incapable of generalizing on a wider sample of data. This obstacle can be overcome using Convolutional Neural Networks (CNNs). CNNs are a subset of Artificial Neural Networks (ANNs), which are in turn a subset of Machine Learning and Artificial Intelligence (AI). CNNs have proven their efficiency in various image processing tasks, including classification and object detection. Some of the most widely used CNNs for object detection include one-stage algorithms, such as You Only Look Once (YOLO) networks and their variations, and two-stage algorithms, such as Region-based CNN (RCNN) and its variations. Researchers attempted to adapt CNNs as a solution to detect palm trees from remote sensing imagery. Zheng et al. [4] utilized Faster-RCNN (FRCNN) with pre-processing steps that boost the performance and accuracy of the network. The authors evaluate the performance of the network on images captured by QuickBird using precision, recall, and F-score, which are on average 95.8%, 95.3%, and 95.5%, respectively. A similar study was conducted in Thailand by Yarak et al. [5] using UAV images.

The goal of this paper is to automate counting palm trees in the UAE by applying image processing and AI techniques to HR remote sensing imagery captured by UAV. The first step and one of the main contributions of this study is to collect a dataset of palm trees from various farms in Al Ain - UAE. Second, the dataset is pre-processed, labelled, and inspected for quality. Then, two CNNs are tested on the dataset to observe their performance, advantages, and shortcomings. Finally, the results are reported and analyzed. The newly created dataset will be shared publicly with the scientific community. the rest of the paper is organized as follows: Section II showcases the collected dataset, Section III explains the methodology, Section IV illustrates and discusses the results, and finally, Section V summarizes and concludes the paper.

II. Dataset

For this research study, aerial RGB imagery provided by Al Ain Municipality are used, which are of size 256×256 with spatial resolution of 1-meter. The images cover Al Ain City; it is located approximately between latitude $24°03'$ and $24°22'$ North and longitude $55°28'$ and $55°53'$ East [6]. All palm trees are labelled manually [7] in PASCAL VOC and YOLO formats. The total number of collected images is 1209 with 10975 palm trees. 1078 of these images with 9883 palm trees are used for training the models. While 131 images with 1092 palm trees are used to test and evaluate the performance of the models. Samples of the dataset can be seen in Fig. 1.

The dataset contains samples that range from simple cases to more complicated ones. For instance, some images contain well-separated palm trees, while others contain crowded overlapping palm trees. Additionally, some images have minimal shadow, while others contain intense shadow around each palm tree. Also, some images contain palm trees only, while others are mixed with vegetation. Finally, some palm trees appear fully, while others are either cropped or obstructed. If a palm tree is visible by at least 50%, then it is labeled as such. The presence of shadows and occlusions is a challenge in the area of object detection generally and palm tree detection particularly. Furthermore, palm trees can appear in different shapes and sizes. For instance, the appearance of young palm trees is different than fully-grown ones. Such samples are included in the dataset, as it is also a challenge that needs to be addressed. The 131 testing images were handpicked from the dataset to contain a wide variety of simple cases and complicated ones in order to fairly evaluate the resilience of each tested network.



(a) Image with shadows (b) Separated palms

(c) Mixed with other vegetation (d) Overlapping palm trees

Fig. 1: Samples from palm tree dataset

III. Methodology

The following subsections describe one-stage and two-stage object detection architectures to train and evaluate the proposed dataset, which are YOLOv4 and Faster-RCNN (FR-CNN).

A. *YOLOv4*

YOLO is a one-stage CNN that performs object detection tasks and it was introduced for the first time by Redmon et al. in 2016 [8]. YOLO can perform both object identification and classification at the same time, and it addresses the task of object detection as a direct regression problem by spatially separating the bounding boxes and their corresponding class probabilities, which are predicted using a single network. An improved version of YOLO known as YOLOv2 [9] was released in 2017. A number of improvements were made on top of the original YOLO architecture, such as batch normalization to avoid overfitting, higher resolution classifier, anchor boxes for bounding boxes prediction, direct location prediction, dimension cluster, and multi-scale training, all of which significantly improve the detection accuracy. YOLOv3 was released in 2018 [10] and it has incremental improvements over its prior versions. It uses upsampling and concatenation of feature layers with earlier feature layers, which preserve fine-grained features. Furthermore, it uses three different scales to predict bounding boxes, which in turn makes the model more efficient in detecting objects with different scales in an image. After YOLOv3 came YOLOv4, which is developed by Bochkovskiy et al. [11]. It is optimized to perform parallel computing tasks where only single GPU is used for training the object detection model. The network is composed of three main blocks: backbone, neck, and head blocks. YOLOv4 outperforms other existing YOLO family detectors in terms of speed and performance due to several improvements, such as introducing bag-of freebies, bag-of-specials, and mosaic augmentation methods during the training stage of the object detection [12].

B. *Faster-RCNN*

Region-Based CNN (RCNN) are a group of machine learning models particularly developed for performing object detection tasks. It was first introduced in 2015 by Girshick et al. [13] as an answer to PASCAL VOC Challenge. RCNN mainly consists of three separate models: feature extraction, bounding box classification, and a model for regressing and fine-tuning the position and size of the bounding box. Despite the reliability of RCNN, it cannot be trained in end-to-end manner, as it is a multi-stage model. A new model known as Fast RCNN is presented in [14] to address this problem. It combines the three models into a single model and also employs the pooling layer within the model architecture. However, Fast RCNN uses the Selective Search method to find the Regions of Interest (ROIs), which is a time consuming process. To solve this problem, the authors in [15], [16] presented a Faster-RCNN (FRCNN) network, which is an upgraded version of Fast RCNN, through introducing Region Proposal Network (RPN). RPN is utilized to avoid Selective Search method, thus, it speeds up the training and detection processes, which in turns improves the overall performance of the model.

IV. Results and Analysis

FRCNN and YOLOv4 models were both developed, trained, and tested using Python tensorflow (2.3.1) and keras (2.1.0) libraries with Tensorflow backend using NVIDIA Quadro P6000-24GB GPU and Intel(R) 12 core Processor CPU with 380GB RAM. Both models were trained in the same environment using the same parameters to ensure a fair comparison. The training parameters are summarized in Table I.

TABLE I: Training Parameters.

Parameter	Value
Epochs	150
Optimization Function	Adam
Learning rate	1×10^{-3}
Shuffle	True

The results are calculated by comparing the predicted bounding boxes to the groundtruth bounding boxes for each palm tree in each image. If the percentage of Intersection over Union (IoU) between both boxes exceeds 50%, it is considered as a True Positive (TP) case. Otherwise, it is considered as a False Positive (FP) case. If the same palm tree is predicted twice, then it is considered as TP the first time, and FP the second time. In the case where a palm tree exists in the groundtruth label but the network was not able to predict it, it is considered as False Negative (FN) case. Finally, if an image does not contain palm trees and the network does not predict any bounding boxes, it is considered as a True Negative (TN) case. Based on the aforementioned cases, both FRCNN and YOLOv4 are evaluated over all testing images according to four quantitative metrics; Average Accuracy (AA), Average Precision (AP), Average Recall (AR), Average F-score (AF). These metrics are computed according to the following equations:

$$AA = \sum_{n=1}^{n=N} \left(\frac{TP_n + TN_n}{TP_n + FP_n + TN_n + FN_n} \right) \div N, \quad (1)$$

$$AP = \sum_{n=1}^{n=N} \left(\frac{TP_n}{TP_n + FP_n} \right) \div N, \quad (2)$$

$$AR = \sum_{n=1}^{n=N} \left(\frac{TP_n}{TP_n + FN_n} \right) \div N, \quad (3)$$

$$AF = \sum_{n=1}^{n=N} \left(2 \times \frac{precision \times recall}{precision + recall} \right) \div N, \quad (4)$$

where n is the index of the image being evaluated, and N is the total number of testing images, which is 131 in this case. In addition to the quantitative metrics, the average detection time for each network is calculated after repeating the experiments 10 times.

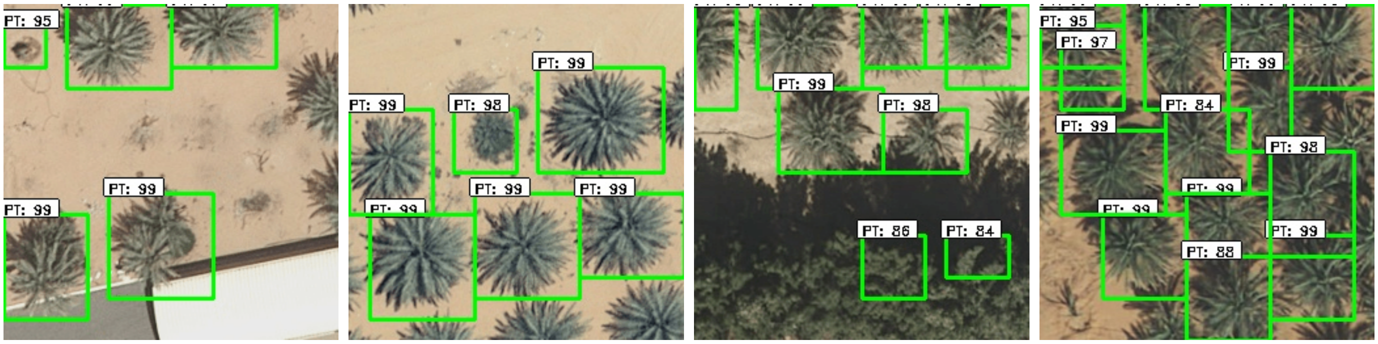
TABLE II: Comparisons between YOLOv4 and FRCNN performance on the proposed dataset.

CNN	Metric				
	AA (%)	AP (%)	AR (%)	AF (%)	Time (s)
YOLOv4	**96.22**	**99.02**	**97.16**	**97.93**	**10.48± 1.25**
FRCNN	66.12	84.24	80.72	83.63	68.94 ± 1.97

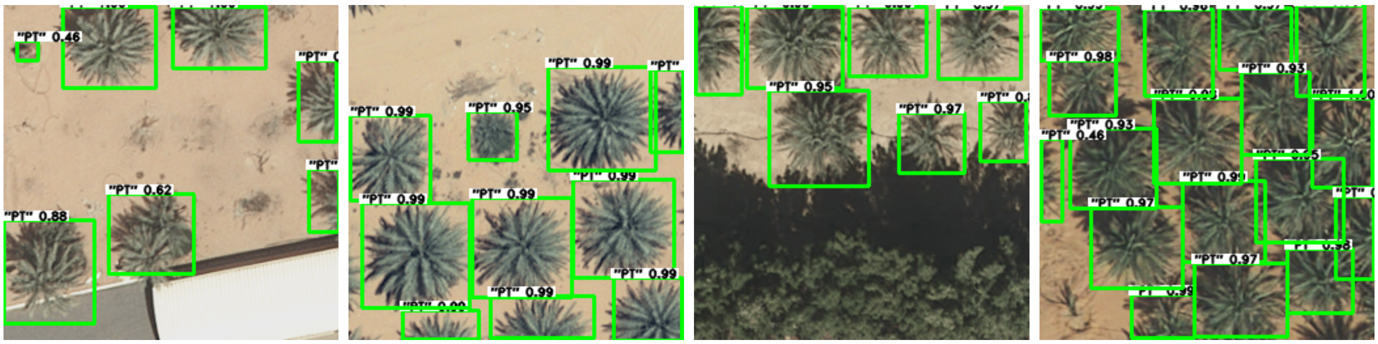
The results of FRCNN and YOLOv4 are summarized in Table II. It is evident that YOLOv4 outperforms FRCNN in terms of all quantitative metrics and it is also faster by 58s. Visually inspecting the outcome provides insight about the shortcomings and strengths of each network. For instance, in Figure 2a, FRCNN was unable to detect the cropped palm trees near the border of the image, whereas YOLOv4 was able to predict them, as seen in Figure 2e. A similar case can be seen in Figures 2b and 2f. Furthermore, FRCNN often falsely detects other vegetation as palm trees, as seen in Figure 2c. YOLOv4 does not show this pattern of false detection, as demonstrated by the sample in Figure 2g. Finally, in the cases where palm trees are crowded and shadows are present, FRCNN can miss palm trees or detect the same palm tree more than once, as seen in Figure 2d, whereas YOLOv4 successfully detects all palm trees without fail, as seen in Figure 2h. It is worth mentioning that both networks are able to detect young palm trees that have newly sprouted, as seen in the upper left corner of Figures 2a and 2e. Overall, YOLOv4 shows more resilience in various scenarios compared to FRCNN.

V. Conclusion

In this paper, a new HR remote sensing dataset is presented for autonomous detection of palm trees in Al Ain - UAE. The proposed dataset is labeled in two formats; PASCAL VOC and YOLO after pre-processing and visually inspecting its quality. Two state-of-the-art FRCNN and YOLOv4 networks are trained and tested to demonstrate the usability of the dataset in detecting the palm trees from remote sensing imagery. Additionally the strengths and weaknesses of each network were discussed. It is evident from the results that YOLOv4 is capable of detecting palm trees even in complex environments where shadows and occlusions exist. As a future work, the dataset will be extended further to include more samples, and



(a) FRCNN prediction sample 1 (b) FRCNN prediction sample 2 (c) FRCNN prediction sample 3 (d) FRCNN prediction sample 4



(e) YOLOv4 prediction sample 1 (f) YOLOv4 prediction sample 2 (g) YOLOv4 prediction sample 3 (h) YOLOv4 prediction sample 4

Fig. 2: Prediction results using FRCNN and Yolov4.

it will be tested on other architectures to explore the possibility of boosting the accuracy.

Acknowledgement

The authors would like to express gratitude to Al Ain City Municipality, Abu Dhabi City Municipality, Al Dhafra Region Municipality, and Department of Municipalities and Transport for providing the images used in this study. The authors would also like to thank Mohammed Bin Rashid Space Centre (MBRSC) for providing the suitable facilities to complete this project.

References

[1] "10 world's biggest countries in date production of 2019," May 2021.

[2] P. Srestasathiern and P. Rakwatin, "Oil palm tree detection with high resolution multi-spectral satellite imagery," *Remote Sensing*, vol. 6, no. 10, pp. 9749–9774, 2014.

[3] S. Al Mansoori, A. Panthakkan, and H. Al Ahmad, "Automatic palm trees detection from multispectral UAV data using normalized difference vegetation index and circular Hough transform," in *High-Performance Computing in Geoscience and Remote Sensing VIII*, Bormin Huang, Sebastián López, and Zhensen Wu, Eds. International Society for Optics and Photonics, 2018, vol. 10792, pp. 11 – 19, SPIE.

[4] J. Zheng, W. Li, M. Xia, R. Dong, H. Fu, and S. Yuan, "Large-scale oil palm tree detection from high-resolution remote sensing images using faster-rcnn," in *IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium*, 2019, pp. 1422–1425.

[5] K. Yarak, A. Witayangkurn, K. Kritiyutanont, C. Arunplod, and R. Shibasaki, "Oil palm tree detection and health classification on high-resolution imagery using deep learning," *Agriculture*, vol. 11, no. 2, 2021.

[6] M. M. Yagoub, "Geographic information systems (gis) application for health: Case of al ain (uae)," *International Journal of Geoinformatics*, vol. 7, no. 1, pp. 21, 2011.

[7] Tzutalin, "Labelimg," https://github.com/tzutalin/labelImg, 2015.

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

[9] J. Redmon and A. Farhadi, "Yolo9000: better, faster, stronger," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 7263–7271.

[10] A. Farhadi and J. Redmon, "Yolov3: An incremental improvement," in *Computer Vision and Pattern Recognition*. Springer Berlin/Heidelberg, Germany, 2018, pp. 1804–2767.

[11] A. Bochkovskiy, C. Wang, and H. M. Liao, "Yolov4: Optimal speed and accuracy of object detection," *arXiv preprint arXiv:2004.10934*, 2020.

[12] C. Gao, Q. Cai, and S. Ming, "Yolov4 object detection algorithm with efficient channel attention mechanism," in *2020 5th International Conference on Mechanical, Control and Computer Engineering (ICMCCE)*. IEEE, 2020, pp. 1764–1770.

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

[14] R. Girshick, "Fast r-cnn," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 1440–1448.

[15] N. Aburaed, M. Al-Saad, M. C. El Rai, S. Al Mansoori, H. Al-Ahmad, and S. Marshall, "Autonomous object detection in satellite images using wfrcnn," in *2020 IEEE India Geoscience and Remote Sensing Symposium (InGARSS)*. IEEE, 2020, pp. 106–109.

[16] M. Al-Saad, N. Aburaed, A. Panthakkan, S. Al Mansoori, H. Al Ahmad, and S. Marshall, "Airbus ship detection from satellite imagery using frequency domain learning," in *Image and Signal Processing for Remote Sensing XXVII*. SPIE, 2021, vol. 11862, pp. 267–273.