Computers and Electronics in Agriculture 192 (2022) 106560

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

journal homepage: www.elsevier.com/locate/compag

Deep neural network based date palm tree detection in drone imagery

Thani Jintasuttisak [a,*], Eran Edirisinghe [b], Ali Elbattay [c]

[a] Department of Computer Science, Loughborough University, United Kingdom
[b] School of Computing and Mathematics, Keele University, United Kingdom
[c] International Center for Biosaline Agriculture, Dubai, United Arab Emirates

ARTICLE INFO

Keywords:
Convolutional Neural Networks
Date palm tree
Object detection
Drone imagery
YOLO-V5

ABSTRACT

Date palm trees are an important economic crop in the Arabian Peninsula, Middle East, and North Africa. Counting the numbers and determining the locations of date palm trees are important for predicting the date production and plantation management. In this paper, we exploit the effective use of the state-of-the-art CNN, YOLO-V5, in detecting date palm trees in images captured by a camera onboard of a drone flying 122 m above farmlands in the Northern Emirates of the United Arab Emirates (UAE). In the dataset preparation process, we randomly selected 125 captured images and divided them into three datasets: training (60%), validation (20%), and testing (20%). The images of date palm trees in the training and validation datasets were manually annotated and those in the training dataset were used to train the four sub-versions of YOLO-V5 CNNs. The validation dataset was used during the training process to assess how well the network was performing during training. Finally, the images in the test dataset were used to evaluate the performance of the trained models. The results of using YOLO-V5 for date palm tree detection in drone imagery are compared with those obtainable with other popular CNN architectures, YOLO-V3, YOLO-V4, and SSD300, both quantitatively and qualitatively. The results show that for the amount of training data used, YOLO-V5m (medium depth) model records the highest accuracy, resulting in a mean average precision of 92.34%. Further it provides the ability to detect and localize date palm trees of different sizes, in crowded, overlapped environments and areas where the date palm tree distribution is sparse. Therefore, it is concluded that the method can be a useful component of an automated plantation management system and help forecast the quantities of date production and condition monitoring of the date palm trees.

1. Introduction

Date palm tree (Phoenix dactylifera L.) is considered one of the oldest fruit trees in the Arabian Peninsula, Middle East, and North Africa. It is an important fruit crop cultivated in countries of the above regions and a key category of agricultural production, characterized by their arid climate. The region contributes to approximately 90% of worldwide date fruit production. The commonly used parts of the date palm trees are its fruits, bark, and leaves. The fruits of date palm tree can be considered as a food that provides high nutritional values with many potential health benefits (Ali et al., 2012). The other parts of date palm trees can be processed into a variety of commercial products such as cosmetics, building materials, and paper (Loutfy, 2010). Therefore, surveying date palm trees including counting their numbers, determining their locations, pattern and distribution is of crucial importance for predicting and forecasting the production volumes and for the purpose of plantation management.

Aerial images of large-cultivated areas of date palm trees can be captured by either satellite imaging or by using an aircraft, manned or unmanned. In the case of using satellite imaging, challenges exist due to the cloud base (Adam et al., 2018), which will make the date palm trees difficult to detect due to the poor captured quality of images. Manned aircrafts are not suitable for crop monitoring due to the costs and the potential environmental pollution (Colefax et al., 2018). An Unmanned Aerial Vehicle (UAV) is the best alternative. Drones are a subset of UAVs and they are typically much smaller, lightweight, and cheap. A drone typically contains a very high-resolution camera that can capture medium to high quality images at a wide angle, depending on the altitude of flying.

In the very recent past, drones have enabled a large variety of applications such as surveillance (Singh et al., 2018; Saqib, 2018), search and rescue (Sampedro et al., 2019; Mishra et al., 2020), and precision

* Corresponding author.
E-mail addresses: T.Jintasuttisak@lboro.ac.uk (T. Jintasuttisak), e.edirisinghe@keele.ac.uk (E. Edirisinghe), a.elbattay@biosaline.org.ae (A. Elbattay).

https://doi.org/10.1016/j.compag.2021.106560
Received 21 April 2021; Received in revised form 29 August 2021; Accepted 7 November 2021
Available online 30 November 2021

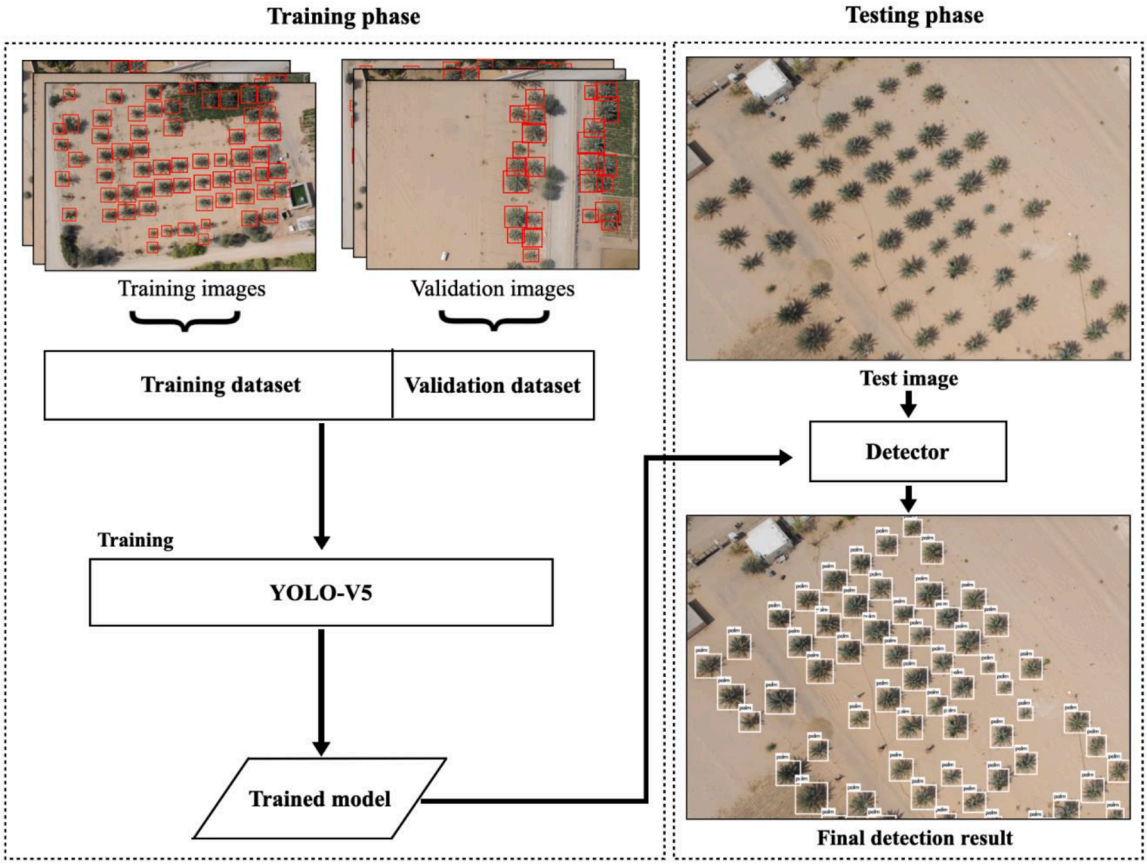
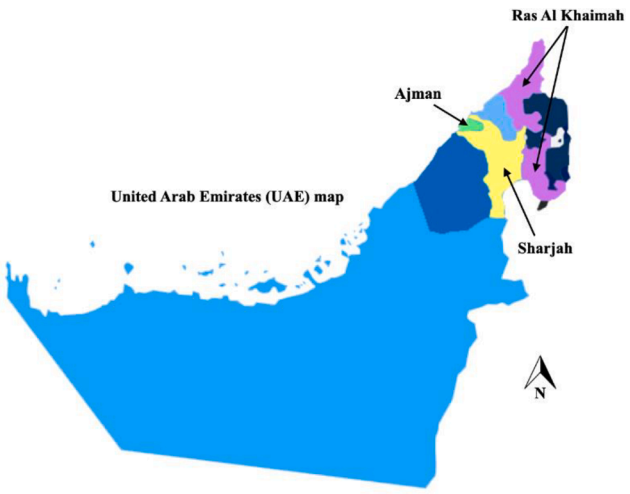


Fig. 1. The workflow of the proposed method.

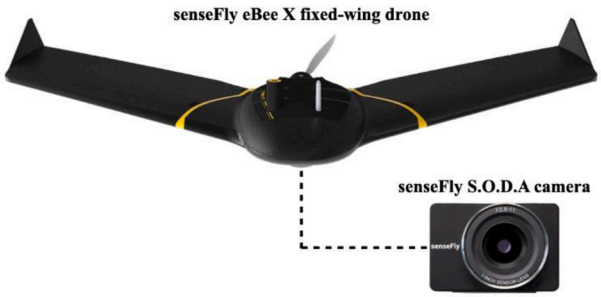
agriculture (Puri et al., 2017; Kulbacki et al., 2018; Ore et al., 2020). They are also relatively low cost as compared to all other options of capturing aerial video footage of large areas, easy to fly and have the flexibility of flying at different altitudes. However, regulations and laws of various countries govern the minimum altitude and areas in which drones can be flown to capture important video footage for agriculture or any other needs. The object appearances at the minimum altitude that drones can fly are often small and occluded by other objects. This will make the task of object detection and recognition based on computer vision, more challenging.

Traditional machine learning based object detection methods have been used in detecting date palm trees and other species of palm trees in images captured by UAVs. These approaches consist of three stages: image pre-processing, feature extraction, and classification. In 2014, Bazi et al. (2014) used SIFT feature extraction to extract a set of key-points of date palm trees. Subsequently, feature vectors were extracted at each key-point which were then entered into an Extreme Learning Machine Classifier to detect date palm trees. The reported accuracy of the date palm tree detection method was 91.11%. In 2016, Manandhar et al. (2016) exploited Circular Autocorrelation of Polar Shape (CAPS) matrix as the feature extractor with a Support Vector Machine (SVM) classifier to detect oil palm trees based on shape features. The reported average detection accuracy of the method was 84%. However, the feature extractors used in the traditional object detection methods are based on handcrafted features decided by humans. It is difficult to achieve the robustness of feature representation. The classification accuracy will depend heavily on the feature set and the classifier used, which is turn is dependent on the dataset, that makes the effective manual design of such systems, very difficult.

Deep learning is an alternative approach to solving the aforementioned problem. Convolutional Neural Networks (CNNs) are one of the deep learning algorithms that have the capability to extract millions of high-level features of objects that can then be used effectively for object detection and classification. In 2017, Li et al. (2017) used LeNet convolutional neural network to extract and learn the features of oil palm trees in high-resolution images captured by the QuickBird satellite. After the LeNet CNN (Li et al., 2017) was trained with the training images, the trained model was used to predict the oil palm trees in the test-image dataset. The method achieved a detection accuracy of 96%. In 2018, Zortea et al. (2018) proposed an idea of combining the results of two CNNs to detect different sizes of oil palm trees in images captured by drone. The first and second CNNs were trained with the sample images of oil palm trees size of 32 $\times$ 32 and 64 $\times$ 64 pixels. The detection probabilities of both models were averaged and estimated as a confidence score of each detected oil palm tree. The method was experimented with three different sizes of oil palm trees and achieved an average detection accuracy about 95%. You-Only-Look-Once (YOLO) is a popular, one-stage, object detection approach that uses CNNs as a backbone. YOLO was proposed by Redmon et al. (2016). In the design of YOLO, the authors aimed to reduce the detection time of the other popular CNN-based object detection method such as R-CNN (Girshick et al., 2014), Fast R-CNN (Girshick, 2015), and Faster R-CNN (Ren et al., 2015). YOLO provides good performance in terms of speed and accuracy and it can be applied in the real-time object detection application. Since the initial publication of YOLO in (Redmon et al., 2016) it has evolved, resulting in more effective and efficient versions, as YOLO-V1 (YOLO) (Redmon et al., 2016), YOLO-V2 (YOLO9000) (Redmon and Farhadi, 2017), YOLO-V3 (Redmon and Farhadi, 2018), YOLO-V4 (Bochkovskiy et al., 2020), and YOLO-V5 (Nelson and Jacob, 2020). The first four versions of YOLO have been widely used and applied in many applications such as medicine (Nie et al., 2019; Yao et al., 2020), remote sensing (Liu et al., 2018; Ma et al., 2020), transportation (Zhang et al., 2020; Dursa and Tune, 2020), and agriculture (Yueju et al., 2018; Tian et al., 2019; Liu, 2020).



(a) The study areas



(b) The UAV platform

Fig. 2. The study areas and the UAV platform used in this research.

In this paper, we exploit the use of YOLO-V5, the latest version of YOLO, in detecting date palm trees in images captured by a drone. YOLO-V5 has been demonstrated to work well in other application areas for detecting small objects (Chen et al., 2021). It should be noted that pre-trained models of YOLO-V5 exist for detecting small objects, but they are not suitable for the detection of date palm trees as such models have been obtained by training the neural networks on images captured by humans or by human installed CCTV cameras. The view angle of such cameras is very different to the aerial view captured by a camera on-board a drone. Therefore, one important aspect of the proposed approach should be, re-training whatever model selected to detect objects, from an aerial viewpoint. In the experimental results presented in this paper, we compare the palm-tree detection accuracy results obtainable by using four different sub-versions of YOLO-V5, with the results obtainable from the other state-of-the-art one-stage object detection methods, including: YOLO-V3, YOLO-V4, and the Single-Shot-MultiBox-Detector (SSD300) (Liu et al., 2016).

The remainder of this paper is organized as follows. Section 2 describes the method used to detect date palm trees in drone imagery. The experimental results and detailed analysis of the results are provided in section 3. Finally, section 4 concludes with insights to further research.

2. Methodology

The workflow of the proposed date palm-tree detection system is illustrated in Fig. 1. It consists of two main phases: the training phase (comprising training and validation) and testing phase. Prior to commencing the training, the captured dataset has to be prepared. The details of each phase, i.e., data preparation, training, validation, and testing are presented in the following sub-sections.

2.1. Data capture and preparation

In the research conducted in this paper, we used a senseFly eBee X, fixed-wing drone with an onboard senseFly S.O.D.A camera to acquire imagery. With permission from the civil aviation authorities, the drone was set to fly at a fixed altitude (122 m) above selected farmlands in the Northern Emirates of the United Arab Emirates (UAE), including Sharjah, Ras Al Khaimah (RAK), and Ajman. Fig. 2. shows the study areas and the UAV platform used in this research. The data samples were captured and provided to us by Falcon Eye Drones, Ltd., Dubai.

The missions were planned by using eMotion mission planner software with a horizontal and vertical overlaps of 70% and 40%, respectively. The drone has cruising speed of 40–110 km/h, wind resistance limit of up to 46 km/h, with a belly linear landing and manual hard lunching and covers up to 5 sq.km on a single battery (with extension). The camera onboard of the drone is a professional photogrammetric camera with a RGB lens of focal points 2.8–11, 10.6 mm (35 mm equivalent: 29 mm), resolution of 5,472 × 3,648 pixels (3:2), exposure compensation ± 2.0 (1/3 increments), a global shutter 1/30 – 1/2000 s, and ISO range 125 – 6400. The ground sampling distance at 122 m (400 ft) is 2.5 cm/pixel. All images were captured during January and February 2019, on cloud free days, between 9:00 a.m. and 1:00p.m. (GMT + 3). Fig. 3 shows two typical examples of the raw images captured by the drone, with the drone camera looking vertically downwards from the drone. The given example consists of date palm trees, cultivated areas (rectangular small crop areas) and other trees which are predominantly positioned along the hedges/wind-breaks. It is noted that the date palm trees are of different sizes, sparsely and non-uniformly distributed on the image. There are some date palm trees that appear to be dead or have only few leaves and it is noted that we have ignored dealing with these date palm trees, within the context of the research conducted and presented in this paper.

A large number of images were captured during the planned flights. We randomly selected 125 images and divided them into three datasets: training (60%), validation (20%), and testing (20%). Therefore, the number of such images in the training, validation, and testing datasets were 75, 25, and 25, respectively. To begin the training phase, all date palm trees in the training and validation datasets were annotated with bounding boxes, using the LabelImg (Tzutalin, 2015) tool and were given a label as “palm”. Table 1 tabulates the number of date palm trees labelled in each image dataset.

The labelled data of the training dataset are used to train YOLO-V5 CNNs. The classification is binary, i.e. a single object, a date palm tree is detected from its background. Similarly, labelled date palm trees of the validation dataset are used during the training process to assess how well the network is performing. Finally, the test dataset is used for evaluating the performance of the palm tree detection models.

When labelling the data for training and validation, in enclosing the date palm trees within rectangles, the rectangles may contain date palm trees that are different sizes, that may overlap with or be occluded by other date palm trees or objects, and might have different backgrounds (i.e., sand, underlying crops, etc.). It is important to have rectangles of image pixels having the above variations captured for testing and training as this will enable the effective training of the CNN and the effective use of the trained model for date palm tree detection, subsequently.

2.2. Yolo-V5

2.2.1. The network architecture

Within the context of the proposed research, the four sub-versions of

Fig. 3. Two examples of the captured drone images.

Table 1
Number of labelled date palm trees in each dataset.

Dataset	Number of labelled date palm trees
Training	5,940
Validation	1,392
Testing	1,465

YOLO-V5 network, including YOLO-V5s (small), YOLO-V5m (medium), YOLO-V5l (large), and YOLO-V5x (extra-large), were trained by using the dataset presented in section 2.1. Each version of the YOLO-V5 network has different model depths, but they have been designed based on the same network structure that consists of three main parts: the backbone, the neck, and the head. In the naming of the different YOLO-V5 sub-versions, s, m, l, and × indicate increasing depth of the network architecture adopted. Fig. 4 illustrates the network architecture of YOLO-V5s, which is used as a core basic structure of all sub-versions of YOLO-V5. The model backbone of YOLO-V5 is used to extract essential features from a given input image. It is designed based on the Cross Stage Partial Network (CSPNet) (Wang et al., 2020) aimed at extracting high level features, while maintaining high accuracy and reducing the processing time of the models. The model neck is mainly used to collect feature maps from different stages of the model backbone for generating feature pyramids. In YOLO-V5, the Path Aggregation Network (PANet) (Shu Liu et al., 2018) has been adopted within the model neck to obtain the feature pyramids. The important role of feature pyramids is to help the model identify and match same type of objects, that are of different sizes and scales. Finally, the model head is used to perform as the final object detection part of YOLO-V5. The model head of YOLO-V5, has been designed as the same as that of YOLO-V3 and YOLO-V4. The model head applies anchor boxes on the final feature maps and generates the final output vectors, with objectness scores, class

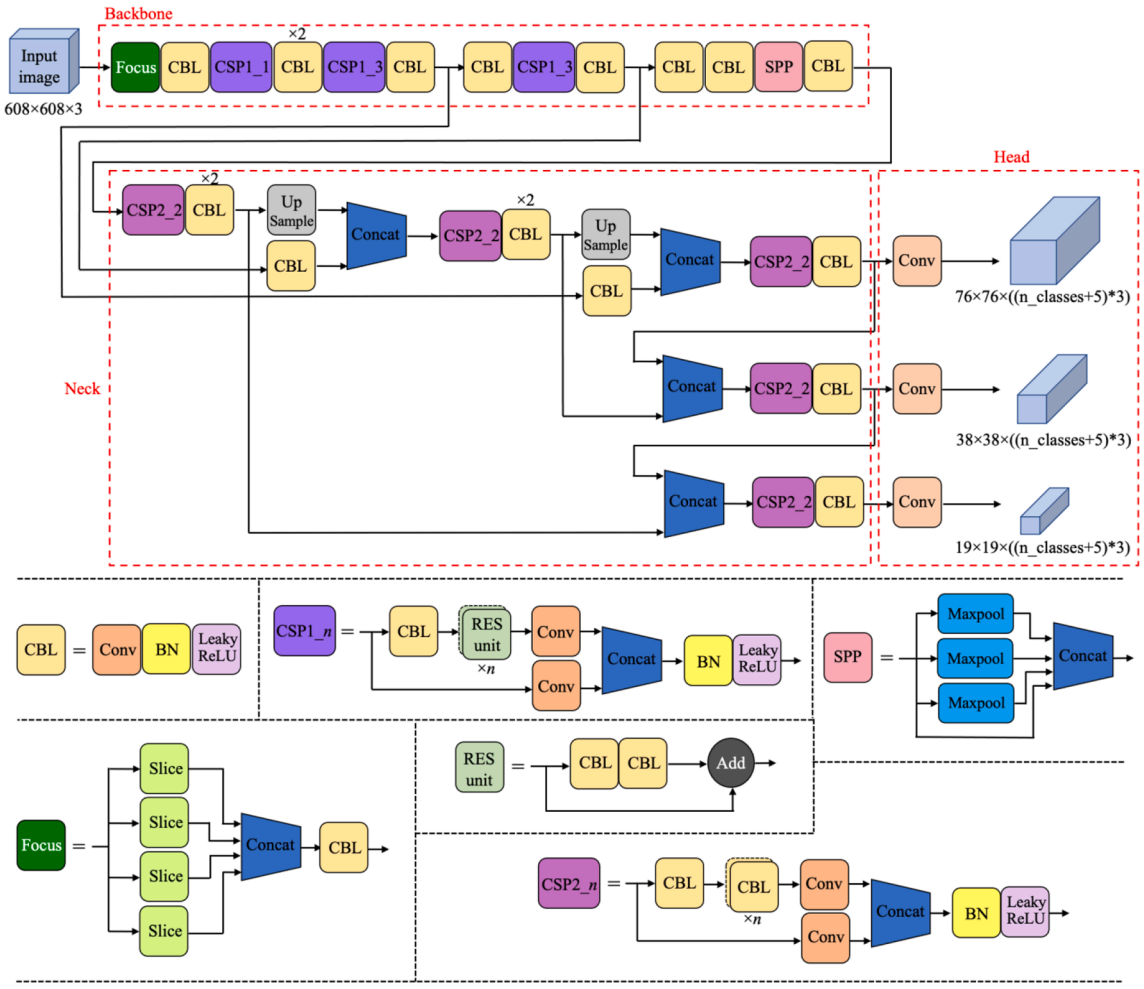


Fig. 4. The network architecture of YOLO-V5s.

Table 2
The CSP modules used in different versions of YOLO-V5 networks.

CSP Modules	YOLO-V5s	YOLO-V5m	YOLO-V5l	YOLO-V5x
CSP1	CSP1_1	CSP1_2	CSP1_3	CSP1_4
CSP1	CSP1_3	CSP1_6	CSP1_9	CSP1_12
CSP1	CSP1_3	CSP1_6	CSP1_9	CSP1_12
CSP2	CSP2_2	CSP2_4	CSP2_6	CSP2_8
CSP2	CSP2_2	CSP2_4	CSP2_6	CSP2_8
CSP2	CSP2_2	CSP2_4	CSP2_6	CSP2_8
CSP2	CSP2_2	CSP2_4	CSP2_6	CSP2_8
CSP2	CSP2_2	CSP2_4	CSP2_6	CSP2_8

probabilities, and coordinates of the bounding boxes that encloses the detected objects.

As shown in Fig. 4, YOLO-V5 has many key components/modules within each part of its network described above, such as Focus, CBL (Convolution, Batch Normalization, and Leaky-ReLU), CSP (Cross-Stage-Partial connections), and SPP (Spatial Pyramid Pooling) modules. The Focus module is a module used for dividing the input image into four parallel slices, to subsequently create feature maps using CBL module. The CBL module is a basic module that uses a convolution operation combined with batch normalization and a leaky-ReLU activation function for feature extraction. The CSP module is a module designed based on CSPNet which is used to enhance the learning ability of the model. There are two types of CSP modules within YOLO-V5, namely CSP1, and CSP2. The CSP1 and CSP2 modules are utilised in the backbone and the neck sections of YOLO-V5 networks. In Fig. 4, a CSP module with a notation of CSP1_n contains a CBL module and n RES (residual) units. A module with notation CSP2_n contains $n + 1$ CBL modules. Both CSP1_n and CSP2_n modules are operated under the same operation by dividing the input feature map into two parts and then fusing the cross-level features, where n indicates the number of RES units and CBL modules in CSP1 and CSP2, respectively. More RES units and CBL modules result in networks with a deeper architecture. Table 2 shows the use of CSP modules in the four sub-versions of YOLO-V5 networks. The SPP module is a module used in the model backbone for mixing and pooling spatial features (He et al., 2015). It down-samples the input features through three parallel max pooling layers and then concatenates to its initial features.

2.2.2. *Training the network*

During the training process, we used data augmentation to increase the diversity of the training image dataset by using different transformations such as scaling, cropping, rotation and color space adjustments. These four additional transformations increased the number of images available for training (75), five-fold, making the number of images available for training 375. Therefore, the total number of date palm trees used for training, for e.g., is 5 × 5,940. In addition, we adopted mosaicking as a data augmentation technique. The mosaic data augmentation is a data augmentation that was initially introduced with YOLO-V4 and can also be used with YOLO-V5 to help a model learn how to recognize objects outside their normal context. The mosaic data augmentation generates a new training image by combining randomly cropped parts of four different training images, and stitching them to one image (see Fig. 5). This process significantly further increases the

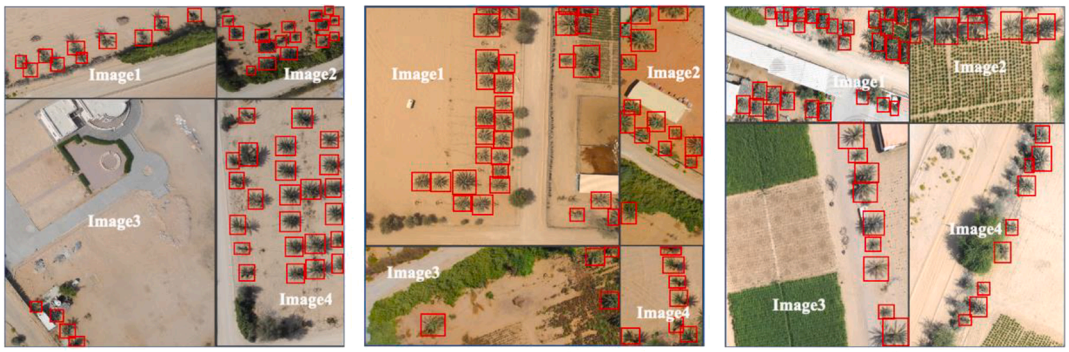


Fig. 5. Mosaic data augmentation.

size of data used in training.

To save time and data needed for training a network from scratch to reach a level of operational accuracy, we use transfer learning, by using pre-trained weights obtained by training the network on COCO dataset (Lin et al., 2014). However, the pre-trained weights obtained from training on COCO dataset are the weights trained for recognizing objects belonging to 80 classes that do not include date palm trees and the default anchor boxes of YOLO-V5 are decided based on the distribution of the bounding boxes of the objects in the COCO dataset. Therefore, the distribution of the bounding box sizes and locations of objects between using the COCO dataset and the date palm tree dataset are different. In the case of using a custom dataset, YOLO-V5 has a built-in function to adjust the preset anchor to optimize the bounding boxes for the date palm tree dataset. YOLO-V5 adopts auto-learning bounding box anchors, to learn and adjust the initial anchor boxes that were decided based on the COCO dataset, optimizing using the date palm tree dataset, by using K-means learning algorithm to adjust to the group size of date palm trees and using a genetic learning algorithm to optimize the process. We trained each sub-version of the YOLO-V5 network for 300 epochs and controlled the learning process of the networks with the same set of the hyper-parameters as follows: batch size = 8, momentum = 0.9, decay = 0.0005, and learning rate = 0.001.

2.2.3. Date palm tree detection

After finishing the training process, the trained models of each sub-version YOLO-V5 network, was used for date palm tree detection in the test dataset. The test images of size 5,472 × 3,648 pixels were, automatically resized to 608 × 608 pixels by the network, while keeping the original aspect ratio of the original input images by using image padding technique. Then, the square images were processed by the trained networks to extract the features and generating three scales of feature maps, of size 76 × 76, 38 × 38, and 19 × 19 for making predictions. The prediction at the feature maps size of 76 × 76 is used to detect date palm trees of relatively small size and the prediction at the feature maps size of 38 × 38 and 19 × 19 are used to detect palm trees of relatively medium and large size. At each scale of prediction, each cell of feature maps, predicts three bounding boxes by using three anchor box scales that are automatically learnt in the training process. Each bounding box contains x and y coordinates, width, height, confidence score, and class probability. The (x, y) coordinates represent the center of the box. The confidence score reflects the level of confidence of a bounding box containing a date palm tree. If the confidence score of a bounding box is very low or zero, it means that a bounding box does not contain any object. The bounding boxes that have low confidence scores can be removed by setting a threshold value in order to keep only the bounding boxes that are most likely to contain the object of interest thus reducing false positives. Finally, we applied Non-Maximum Suppression (NMS) to suppress the redundant and overlapping bounding boxes and rescaled the final detection results of the square image to the original image size. Fig. 6 illustrates the key process of YOLO-V5 object detection and shows how YOLO-V5 handles the bounding boxes.

3. Experimental results

In this section, we compare the performance of the proposed YOLO-V5 based approach (of all sub-versions) to date palm tree detection, with the performance of the state-of-the-art CNN based one-stage object detection methods, including YOLO-V3 (Redmon and Farhadi, 2018), YOLO-V4 (Bochkovskiy et al., 2020), and SSD300 (Liu et al., 2016). Each method was trained with the captured drone image training dataset by using the COCO pre-trained weights and the same set of hyper-parameters that we used in training YOLO-V5 networks in section 2.2.2. The networks were trained until their validation loss was no longer changed and they were tested on identical test datasets. The performance was compared both quantitatively and qualitatively. All experiments were conducted on a PC with Intel Core i5-10400F CPU, NVIDIA GeForce GTX-1080Ti GPU (11 GB GPU memory), and 16 GB of RAM. The operating system used by the PC was Ubuntu 16.04.

3.1. Quantitative performance comparison

In order to conduct a quantitative performance comparison, the metric Intersection over Union (IoU) was used as an indicator to determine whether the detection of each date palm tree is correct. The IoU can be calculated by using the intersection of area between the predicted bounding box and the ground truth and dividing it by the area of their union, as shown in Eq. (1), where B_{prd} is the predicted bounding box, and B_{gt} is the ground-truth bounding box. Fig. 7 illustrates the calculation of the IoU between a predicted bounding box and ground truth bounding box.

$$IoU = \frac{area(B_{prd} \cap B_{gt})}{area(B_{prd} \cup B_{gt})} \tag{1}$$

In this research, we considered the predicted bounding box that has the IoU value greater than or equal to 0.5 as a correct prediction of date palm tree. The number of correct predictions can be used to calculate precision and recall rate for evaluating the performance of the compared models. The precision is defined as the proportion of the correctly detected objects in all detected objects and the recall is referred to the proportion of the correctly detected objects in all ground truth objects. The equations for calculating the precision and recall rate are defined as follows:

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

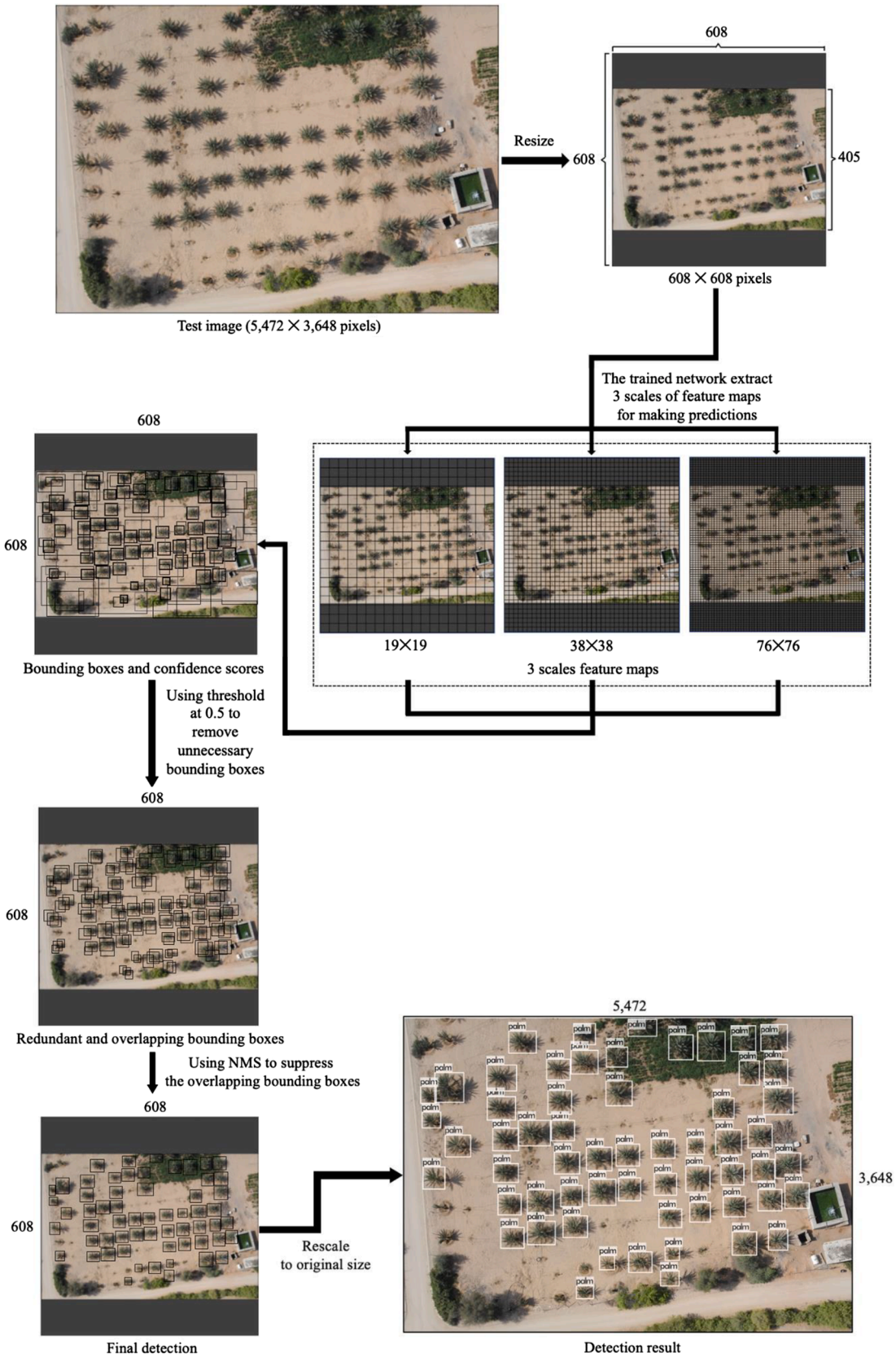


Fig. 6. The key processes of YOLO-V5 object detection method.

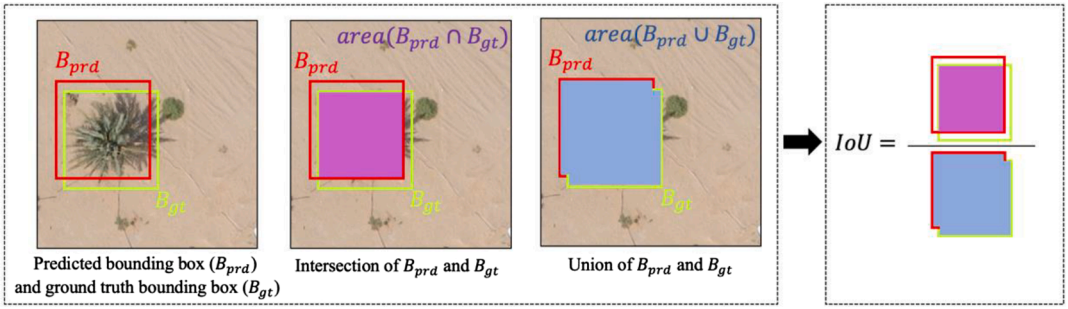


Fig. 7. The visualization of IoU calculation between predicted and ground-truth bounding boxes.

Table 3
Precision, recall, mAP, and average detection time per image of the seven methods at an IoU threshold of 0.5.

Model	Precision	Recall	mAP@0.5 IoU	Average detection time per image (milli-second)
SSD300	0.59	0.58	58.55%	24.31
YOLO-V3	0.95	0.85	88.31%	26.87
YOLO-V4	0.86	0.86	89.73%	29.84
YOLO-V5s	0.91	0.90	90.40%	11.33
YOLO-V5m	0.91	0.92	92.34%	16.42
YOLO-V5l	0.92	0.90	90.51%	23.34
YOLO-V5x	0.92	0.91	91.02%	35.12

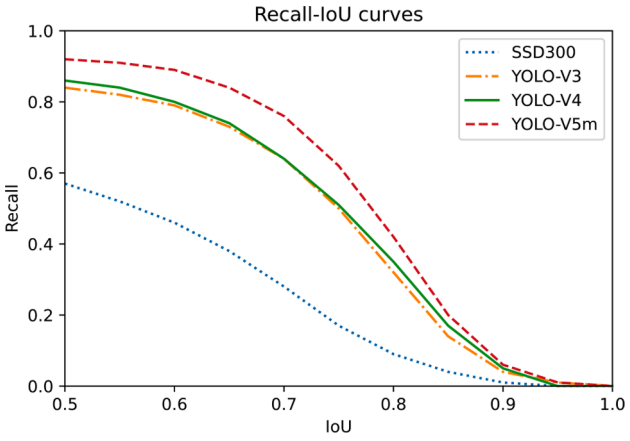


Fig. 8. Recall-IoU curves of SSD300, YOLO-V3, YOLO-V4, and YOLO-V5m.

where *TP* (True Positive) is the number of the date palm trees correctly detected, *FP* (False Positive) is the number of other objects detected as date palm trees, and *FN* (False Negative) is the number date palm trees that are not detected/missed.

Apart from the precision and recall rate, we also used the Mean Average Precision (mAP) and calculated the average detection time per image to measure the average precision (AP) and the deployment speed of the models. The mAP was calculated by averaging 11 precision values, when recall varies in a range between 0 and 1 with 0.1 interval, at an IoU threshold of 0.5, as shown in Eq. (4), where *P* is the precision, and *R* is the recall.

$$mAP = \frac{1}{11} \sum_{R \in \{0, 0.1, 0.2, \cdots, 1\}} P(R) \tag{4}$$

Table 4
The training times of SSD300, YOLO-V3, YOLO-V4, and the four sub-versions of YOLO-V5.

DNN Architecture	Training Time (hours)
SSD300	2.42
YOLO-V3	2.67
YOLO-V4	2.92
YOLO-V5s	0.84
YOLO-V5m	1.59
YOLO-V5l	2.35
YOLO-V5x	3.85

Table 3 summarizes the quantitative results of all seven models. It can be seen that the mAP values of all YOLO-V5 based models exceed 90%, and the best mAP is 92.34% when using the YOLO-V5m based model. It is noted that if more data was used in training it would be likely for the models of YPLO-V5l or YOLO-V5x to have the best performance that would however come at a higher deployment time. However, the mAP values of SSD300, YOLO-V3, and YOLO-V4 are 58.55%, 88.31%, and 89.73%, respectively. In terms of detection speed, the performance of SSD300, YOLO-V3, and YOLO-V4 are slower than YOLO-V5s, YOLO-V5m, and YOLO-V5l. YOLO-V5x seems to have the highest average detection time per image as it has the highest number of layers in its model and is hence the model with the deepest architecture.

Among the four models of YOLO-V5, we selected the model of YOLO-V5m that gave the highest mAP value for further performance comparison with SSD300, YOLO-V3, and YOLO-V4. We calculated the recall values of the four models under different IoU thresholds in the range 0.5 to 1.0. Fig. 8 plots the recall values of the four models under the different IoU thresholds. It can be observed that the overall recall values of YOLO-V5m is the highest among the recall values of the four models. SSD300 demonstrates the lowest overall recall values. Meanwhile, the overall recall values of YOLO-V3 and YOLO-V4 are quite similar. Plotting the recall-IoU curves of the four models in Fig. 8 also shows that with the IoU thresholds increasing, the recall values of SSD300, YOLO-V3, and YOLO-V4 drop more sharply than that of YOLO-V5m. In summary, YOLO-V5m demonstrates the best recall values under the various IoU thresholds tested.

Table 4 below presents a comparison of training times when using each neural network. It is noted that both SSD300 and YOLO (V3, V4 and V5l) have similar training times. This is due to their network's architectures being similar in complexity. It is also seen that YOLO-V5s has a significantly lower training time as compared to YOLO-V5x. This is expected as YOLO-V5s is an architecture that has far less level of the network that compared with YOLO-V5x.

3.2. Visual performance comparison

In terms of visual performance comparison, the detection results of YOLO-V5m were used to represent the performance of YOLO-V5 and compared to the detection results of SSD300, YOLO-V3, and YOLO-V4.

(a) SSD300 (c) YOLO-V4

(b) YOLO-V3 (d) YOLO-V5m

Fig. 9. The results of date palm tree detection in drone imagery using the four CNNs, SSD300, YOLO-V3, YOLO-V4, and YOLO-V5m.

(a) SSD300 (c) YOLO-V4

(b) YOLO-V3 (d) YOLO-V5m

Fig. 10. The results of date palm tree detection in drone imagery using the four CNNs, SSD300, YOLO-V3, YOLO-V4, and YOLO-V5m.

Both YOLO and SSD300 are popular one-stage CNN-based object detection methods, with YOLO-V5 being the latest of the YOLO network versions. Figs. 9 and 10 show the date palm tree detection results in drone imagery of the four models. It can be seen that YOLO-V5m achieves very good detection results even when the date palm trees are overlapped with each other and some parts of palm trees are occluded by other objects such as hedges, and other trees. The date palm trees of different sizes and date palm trees that are grown amongst small crop areas are also detected. SSD300 can detect the date palm trees, when the date palm trees are non-overlapping, but it could not detect any of the overlapped date palm trees. Moreover, SSD300 also detected other objects as date palm tree. YOLO-V3 and YOLO-V4 achieve better results than SSD300. It can detect sparse date palm trees and some of overlapped date palm trees but it could not detect the date palm trees that are occluded by the hedges. This demonstrates that the YOLO-V5m model provides an enhanced ability to detect crowded and overlapped date palm trees and the date palm trees that are planted very closely to the hedges and other trees, as compared with what was achievable with SSD300, YOLO-V3, and YOLO-V4.

4. Conclusion

In this paper, we have proposed the use of the latest version of YOLO, namely YOLO-V5, to detect date palm trees in drone imagery. Our experiments were conducted on 125 images of size 5,472 × 3,648 pixels captured by a fixed-wing drone flying at a fixed altitude (122 m) above farmlands in the Northern Emirates of the UAE, where different sizes of date palm trees exist, with different overlaps, occlusions and backgrounds. The date palm trees were also sparsely and non-uniformly distributed within the terrain. Approximately 60% of images were used for training and the rest of the images were used for validation and testing. The performance of the use of four sub-version of YOLO-V5 network were compared with the performance of the use of three other alternative state-of-the-art CNN based one-stage object detection methods, namely SSD300, YOLO-V3, and YOLO-V4, both in terms of date palm tree detection precision and average time consumed for testing an image. The mAP values of the YOLOv5 models all exceed 90%, and the maximum is 92.34% in YOLO-V5m, while the mAP values of the SSD300, YOLO-V3, and YOLO-V4 models are just 58.55%, 88.31%, and 89.73%, respectively. The detection speed of YOLO-V5s, YOLO-V5m, and YOLO-V5l models are faster than SSD300, YOLO-V3, and YOLO-V4 models. However, YOLO-V5x has the highest average detection per image amongst the compared models because it has the highest number of layers in its network model. We provide subjective experimental results to illustrate the impact of performance of each CNN on date palm tree detection confirming the superiority of performance of YOLO-V5.

CRediT authorship contribution statement

Thani Jintasuttisak: Conceptualization, Methodology, Software, Validation, Investigation, Data curation, Writing – original draft. **Eran Edirisinghe:** Supervision, Writing – review & editing. **Ali Elbattay:** Resources.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

The authors would like to thank Falcon Eye Drones Ltd., Dubai, United Arab Emirate (UAE), for capturing data, making the data available for our research and their involvement in problem definition and system requirement analysis.

References

Ali, Amanat, Mostafa Waly, M., Essa, Mohamed, Devaranjan, Sankar, 2012. 26 Nutritional and medicinal. Dates: Prod., Proces., Food Med. Values 361.
Loutfy, I., 2010. EI-Juhany, "Degradation of date palm trees and date production in arab countries: causes and potential rehabilitation". Aust. J. Basic Appl. Sci. 4 (8), 3998–4010.
Adam, F., Monks, M., Esch, T., Datcu, M., 2018. Cloud removal in high resolution multispectral satellite imagery: comparing three approaches. Multidisciplinary Digital Publ. Institu. Proc. 2 (7), 353.
Colefax, Andrew P., Butcher, Paul A., Kelaher, Brendan P., 2018. The potential for unmanned aerial vehicles (UAVs) to conduct marine fauna surveys in place of manned aircraft. ICES J. Marine Sci. 75 (1 (January/February)), 1–8.
Singh, A., Patil, D., Omkar, S.N., 2018. Eye in the sky: Real-time drone surveillance system (dss) for violent individuals identification using scatternet hybrid deep learning network. In: *Proceeding of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 1629–1637.
Saqib, M., Khan, Sultan Daud, Sharma, Nabin, Scully-Power, Paul, Butcher, Paul, Colefax, Andrew, Blumenstein, Michael, 2018. Real-time drone surveillance and population estimation of marine animals from aerial imagery. In: *2018 International Conference on Image and Vision Computing New Zealand*, pp. 1–6.
Sampedro, C., Rodriguez-Ramos, A., Bavle, H., Carrio, A., de la Puente, P., Campoy, P., 2019. A fully-autonomous aerial robot for search and rescue applications in indoor environments using learning-based techniques. J. Intell. Rob. Syst. 95 (2), 601–627.
Mishra, B., Garg, D., Narang, P., Mishra, V., 2020. Drone-surveillance for search and rescue in natural disaster. Comput. Commun. 156, 1–10.
Puri, V., Nayyar, A., Raja, L., 2017. Agriculture drones: a modern breakthrough in precision agriculture. J. Statistics Manag. Syst. 20 (4), 507–518.
Kulbacki, Marek, Segen, Jakub, Kniec, Wojciech, Klempouse, Rayszard, Kluwak, Konrad, Nikodem, Jan, Kulbacka, Julita, Serester, Andrea, 2018. Survey of drones for agriculture automation from planting to harvest. In: *2018 IEEE 22nd International Conference on* Intelligent Engineering Systems (INES), pp. 353–358.
Ore, G., Alcantara, M.S., Goes, J.A., Oliveira, L.P., Yepes, J., Teruel, B., Castro, V., Bins, L. S., Castro, F., Luebeck, D., Moreira, L.F., Gabrielli, L.H., Hernandez-Figueroa, H.E., 2020. Crop growth monitoring with drone-borne DInSAR. Remote Sensing 12 (4), 615–632.
Bazi, Y., Malek, S., Alajlan, N., AlHichri, H., 2014. An automatic approach for palm tree counting in uav images. In: *2014 IEEE Geoscience and Remote Sensing Symposium*, pp. 537–540.
Manandhar, A., Hoegner, L., Stilla, U., 2016. Palm tree detection using circular autocorrelation of polar shape matrix. ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Inform. Sci. 3, 465–472.
Li, W., Fu, H., Yu, Le, Cracknell, A., Fu, Haohuan, Yu, Le, Cracknell, Arthur, 2017. Deep learning based oil palm tree detection and counting for high-resolution remote sensing images. Remote Sensing 9 (1), 22. https://doi.org/10.3390/rs9010022.
Zortea, M., Nery, M., Ruga, B., Carvalho, L.B., Bastos, A.C., 2018. Oil-palm tree detection in aerial images combining deep learning classifiers. In: IGARSS 2018 IEEE International Geoscience and Remote Sensing Symposium, pp. 657–660.
Redmon, J., Divvala, S., Girshick, R., Farhadi, A., 2016. You only look once: Unified, real-time object detection. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 779–788.
Girshick, R., Donahue, J., Darrell, T., Malik, J., 2014. Rich feature hierarchies for accurate object detection and semantic segmentation. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 580–587.
Girshick, R., 2015. Fast r-cnn. In: *Proceedings of the IEEE international conference on computer vision*, pp. 1440–1448.
Ren, S., He, K., Girshick, R., Sun, J., 2015. Faster r-cnn: Towards real-time object detection with region proposal networks. In: *Advances in neural information processing systems*, pp. 91–99.
Redmon, J., Farhadi, A., 2017. Yolo9000: better, faster, stronger. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 7263–7271.
Joseph Redmon and Ali Farhadi. 2018. "Yolov3: An incremental improvement," *arXiv preprintarXiv:1804.02767*.
Nie, Y., Sommella, P., O'Nils, M., Liguori, C., Lundgren, J., 2019. Automatic detection of Melanoma with Yolo deep convolutional neural networks. In: 2019 E-Health and Bioengineering Conference (EHP), pp. 1–4.
Yao, S., Chen, Y., Tian, X., Jiang, R., Ma, S., 2020. An improved algorithm for detecting Pneumonia based on Yolov3. Appl. Sci. 10 (5), 1818–1833.
Liu, W., Ma, L., Chen, H.e., 2018. Arbitrary-oriented ship detection framework in optical remote sensing images. IEEE Geosci. Remote Sensing Lett. 15 (6), 937–941.
Ma, H., Liu, Y., Ren, Y., Jingzian, Y.u., 2020. Detection of collapsed buildings in post-earthquake remote sensing images based on the improved YOLOv3. Remote Sensing 12 (1), 44–62.
Zhang, H., Qin, L., Li, J., Guo, Y., Zhou, Y.a., Zhang, J., Zhi, X.u., 2020. Real-time detection method for small traffic signs based on Yolov3. IEEE Access 8, 64145–64156.
Bedada Bekele Dursa, and Kula Kekeba Tune. 2020. "Developing traffic congestion detection model using deep learning approach: a case study of Addis Ababa city road," https://doi.org/10.21203/rs.3.rs-113234/v1.
Yueju, Xue, Ning, Huang, ShuQin, Tu, Liang, Mao, AQing, Yang, XunMu, Zhu, XiaoFan, Yang, PengFei, Chen, 2018. Immature mango detection based on improve Yolov2. Trans. Chinese Soci. Agric. Eng. 34 (7), 173–179.
Tian, Yunong, Yang, Guodong, Wang, Zhe, Wang, Hao, Li, En, Liang, Zize, 2019. Apple detection during different growth stages in orchards using the improved Yolo-v3 model. Comput. Electron. Agric. 157, 417–426.

Liu, G., Nouaze, Joseph Christian, Touko Mbouembe, Philippe Lyonel, Kim, Jae Ho, 2020. Yolo-Tomato: A robust algorithm for tomato detection base on YOLOv3. Sensors 20 (7), 2145–2164.

Chen, Yuwen, Zhang, Chao, Qiao, Tengfei, Xiong, Jianlin, Liu, Bin, 2021. Ship detection in optical sensing images based on YOLOv5. In: *Twelfth International Conference on Graphics and Image Processing*, p. 117200E.

Liu, Wei, Anguelov, Dragomir, Erhan, Dumitru, Szegedy, Christian, Reed, Scott, Cheng-Yang, Fu, CBerg, Alexander, 2016. Ssd: Single shot multibox detector. In: European conference on computer vision. Springer, pp. 21–37.

Wang, Chien-Yao, Liao, Hong-Yuan Mark, Wu, Yueh-Hua, Chen, Ping-Yang, Hsieh, Jun-Wei, Yeh, I-Hau, 2020. CSPNet: A new backbone that can enhance learning capability of CNN. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops, pp. 390–391.

Liu, Shu, Qi, Lu, Qin, Haifang, Shi, Jianping, Jia, Jiaya, 2018. Path aggregation network for instance segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 8759–8768.

He, K., Zhang, X., Ren, S., Sub, J., 2015. Spatial pyramid pooling in deep convolutional networks for visual recognition. IEEE Trans. Pattern Anal. Mach. Intell. 37 (9), 1904–1916.

Lin, Tsung-Yi, Maire, Michael, Belongie, Serge, Hays, James, Perona, Pietro, Ramanan, Deva, Dollar, Piotr, Lawrence Zitnick, C., 2014. Common objects in context. In: European conference on computer vision. Springer, pp. 740–755.

Joseph Nelson and Jacob Solawetz. 2020. YOLOv5 is here: state-of-the-art object detection at 140 fps, accessed: 2020-11-12. https://blog.roboflow.com/yolov5-is-here/.

Tzutalin. 2015. LabelImg, accessed: 2020-10-15. https://github.com/tzutalin/labelImg.

Bochkovskiy, A., Wang, C. Y., and Liao, H. Y. M. 2020. "Yolov4: optimal speed and accuracy of object detection," *arXiv preprint arXiv:2004.10934*.