International Journal of Cognitive Computing in Engineering 2 (2021) 215–228

Contents lists available at ScienceDirect

International Journal of Cognitive Computing in Engineering

journal homepage: https://www.keaipublishing.com/en/journals/international-journal-of-cognitive-computing-in-engineering/

A review on object detection in unmanned aerial vehicle surveillance

Anitha Ramachandran[a], Arun Kumar Sangaiah[a,*]

[a] School of Computing Science and Engineering, Vellore Institute of Technology, Vellore, India.

ARTICLE INFO

Keywords:
Object detection
UAV
Drone
Deep learning
Precision agriculture
Security

ABSTRACT

Purpose: Computer vision in drones has gained a lot of attention from artificial intelligence researchers. Providing intelligence to drones will resolve many real-time problems. Computer vision tasks such as object detection, object tracking, and object counting are significant tasks for monitoring specified environments. However, factors such as altitude, camera angle, occlusion, and motion blur make it a more challenging task.
Methodology: In this paper, a detailed literature review has been conducted focusing on object detection and tracking using UAVs concerning different applications. This study summarizes the findings of existing research papers and identifies the research gaps.
Contribution: Object detection methods applied in UAV images are classified and elaborated. UAV datasets specific to object detection tasks are listed. Existing research works in different applications are summarized. Finally, a secure onboard processing system on a robust object detection framework in precision agriculture is proposed to mitigate identified research gaps.

1. Introduction

Computer vision is already achieved remarkable achievements due to advancements in deep learning algorithms, hardware requirements, and the availability of datasets. Object detection is the most investigation task done by the researchers as it has many applications. The objective of object detection is to detect objects from a certain category (for example human beings, dogs, vehicles, motorcycles, or cats) in a picture and if any, return the area and extent of each instance of objects. It constitutes the backbone for solving complex and high-level computer vision tasks such as object tracking, segmentation, event detection, image captioning, scene comprehension, crowd monitoring, and activity recognition. Researchers started to overcome the challenge of developing general object detection systems that can detect categories of objects which match human ones. The field of object detection has been greatly enhanced. However, object detection in drone applications is yet to research. From surveillance to agriculture, all applications need accurate object detection to serve their purpose.

Drones proliferate mostly in all real-time applications like surveillance (Sien, Lim & Au, 2019), delivery services (Hwang & Kim, 2019; Hwang, Kim & Lee, 2020), traffic monitoring (Kyrkou, Plastiras, Venieris, Theocharides & Bouganis, 2018), agriculture (Nuijten, Kooistra & De Deyn, 2019), disaster management (Kyrkou & Theocharides, 2020), and marine surveillance (Kim et al., 2015). Recently, Amazon received federal approval to use drones for its delivery purpose. Hii, Courtney & Royall (2019) studied drone transportation exclusive for medicines and suggest that it is feasible. Precision agriculture is expected to see significant growth compared to other applications, as the use of UAV is becoming one of the most crucial parts in managing farm tasks. Precision agriculture is a collection of methods for tracking crops, collecting data, and carrying out informed crop management tasks such as the optimal water supply, the selection of effective pesticides. UAV can assist farmers in a vast range of activities such as planning and analyzing crop plantation, farm monitoring to assess crop growth and health. The benefits of using aerial services in agriculture favored the expansion of the use of fertilizers from the air in 1940 to other activities, such as the top dressing. Drone manufacturing company called DJI developed drones by integrating sensors especially for protecting crops from insects or weeds.

Although drones are popular now, their history is long years back. It can be categorized based on flying speed, stabilizing its position, hovering, fixed or loitering, and flying environment. Different types of UAVs have different characteristics (Fig. 1). Single rotor UAV can carry heavy payloads but it has mechanical complexity which leads to high cost. Multirotor UAV is a common type as its usage applies to both professionals and common people. It can hover or move along the given target. Fixed-wing UAVs are high in flying speed and can carry heavy payloads but it needs a runway to takeoff and landing. Hybrid UAV is an improved version of the fixed-wing but it is still under development. Their autonomy standard may be other criteria for classifying the UAVs apart from their aerodynamic concepts. When the pilot gives references to each aircraft actuator, the pilot may be graded into a tele operated one. Tele-commanded vehicles display the second degree of autonomy. In this situation, the aircraft depends on an onboard automatic controller

* Corresponding author.
E-mail address: arunkumarsangaiah@gmail.com (A.K. Sangaiah).

https://doi.org/10.1016/j.ijcce.2021.11.005
Received 16 August 2021; Received in revised form 26 November 2021; Accepted 29 November 2021
Available online 4 December 2021

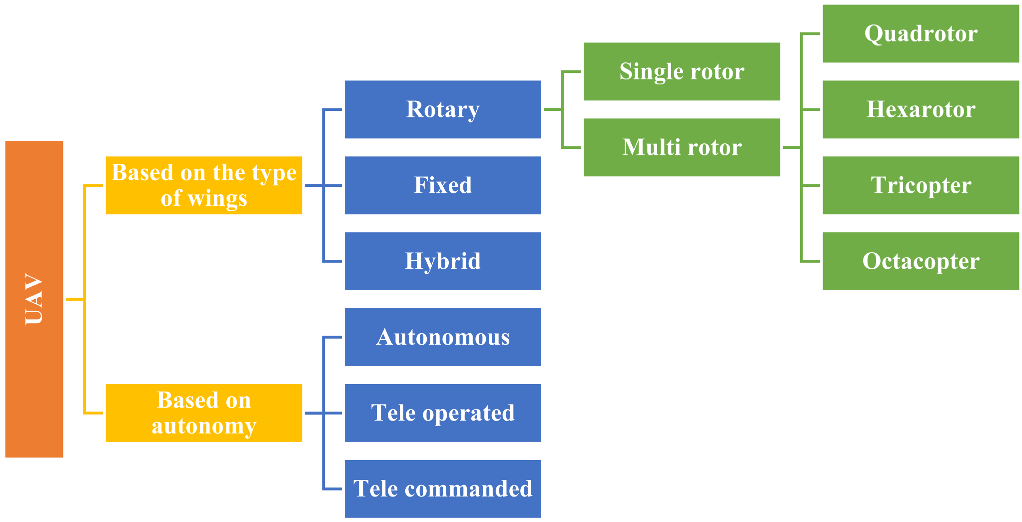

Fig 1. Classification of UAV Types.

for safe flight. The ground operator can provide the onboard controller with velocity and orientation commands. Without human interference, autonomous vehicles can execute some sort of flight plan. Most commercial units that depend on GPS onboard enable an ordered list of 3D points called waypoints to be identified.

In drone videos, the camera is placed at a higher altitude and the scene contains more contextual information. However, changes in viewpoint, and changes in scale make the object detection task in drones more challenging than traditional object detection. In a traffic surveillance scenario, drones record traffic from a bird's eye view. The advantage of this to 100 m in height to record vehicle traffic. Object detection from a bird's eye perspective is more challenging than the front-parallel view for the following aspects,

- The dynamic transition of moving objects.
- Changes in aspect ratio and image scale.
- Abrupt camera motion.
- Severe perspective distortion.
- Motion blur.
- The high density of objects.
- Complex background.

In addition to these difficulties, the object detection studies in aerial view are also confronted by the biased dataset problem. To avoid this problem, the dataset must be annotated which reflects real-world applications. Therefore, it is not unusual that the object detection models learned from traditional images are not appropriate for aerial images.

1.1. Research motivation

Drone surveillance has higher mobility and large surveillance scope in contrast to fixed cameras. It has imperfections such as unstable background, low resolution, and illumination changes. There is a high demand for intelligent drones in real-world applications. However, object detection in drone images or videos is different from traditional object detection. The size varies in aerial photos from object instances. Not only because of spatial sensor resolutions but also because of the size differences within the same type of object. Many tiny instances of objects are crowded in aerial photos, for example, cars and vehicles in the parking place and ships at a port. Besides, the frequencies of instances are unbalanced in aerial photos, for example, a small size image may have 100 instances whereas a large size image may have only a limited amount of instances. Objects in aerial photographs also appear in arbitrary directions. There are also several instances of the large aspect ratio. Hence, this work aims to discuss the state-of-the-art approaches in object detection in drone images and explore its real-time applications.

1.2. Research contribution

The review is designed to provide a systematic overview of current studies by addressing all object detection methods and applications in drone images. The main contribution of this study can be summarized as follows:

- Categorize existing studies based on the approach namely traditional image processing or deep learning and real-time applications.
- Provide a list of drone datasets with related details to assist more researchers in this field.
- Proposed secure onboard processing for robust object detection framework.

The remaining paper is structured as follows: Section 2 discusses existing survey papers on object detection in drone data. Research questions and identified research gaps are given in Section 3. For each research question, Section 4 describes in-depth the solution. Secure onboard processing for robust object detection framework is proposed in Section 5 and the paper concludes with Section 6.

2. Related works

Researchers have worked out several drone studies. Some of them focused on applications and some of them focused on methods applied in drone research. Carrio, Sampedro, Rodriguez-Ramos & Campoy (2017) performed an extensive literature review on deep learning methods applied to UAVs. Deep learning algorithms used in the context of drones are explained in detail. Also, the situations where these algorithms are applied such as scene classification, object recognition, drone movement control, planning, and situational awareness are elaborated. Belmonte, Morales & Fernández-Caballero (2019) have done systematic mapping related to computer vision on UAVs. Overall the review is subdivided into four categories based on UAV vision namely control, navigation, sense-avoid, and tracking or guidance. Kanellakis & Nikolakopoulos (2017) presented a detailed literature review on UAVs concerning its current development and future trends. UAV types, sensors, vision-based navigation, and obstacle discussion are discussed. Yao, Qin & Chen (2019) reviewed the concept of drones in remote sensing applications. UAV sensors are explained in detail and remote sensing analysis is elaborated. Different types of applications are also discussed. Shakhatreh et al. (2019) have done a detailed survey on drones in civil infrastructure applications. It is expected that intelligent drones can increase the market value of up to $45 billion in civil infrastructure applications. The main advantage will be reduced risk and less cost. Drone usage in civil infrastructure has many challenges such as charging, swarming, collision avoidance, network, and security. The authors addressed these challenges and gave insights to address these challenges. Many of the existing surveys focused either on specific applications or methods. (Boursianis et al., 2020) reviewed the internet of things and UAV from an agriculture perspective. IoT sensors, networks, applications used in agriculture are explained in detail. del Cerro, Cruz Ulloa, Barrientos & de León Rivas (2021) presented a survey about the use of UAV in agriculture. Sensors used in UAV for smart farming are classified and explained in detail. Chandra, Desai, Guo & Balasubramanian (2020) reviewed the deep learning methods used for plant phenotyping in smart agriculture. There is a lack of survey work dedicated to general object detection in drone images or videos. Hence, in this paper, we provide a brief survey on the state of the art methods and applications related to object detection tasks in drone images or videos.

3. Research methodology

RQ1. What are all the latest research strategies to address drone surveillance challenges?

Motivation: It is intended to explore existing solutions to the problem of drone object detection offered by researchers to develop a proposed model to address the constraints of existing works.

RQ2. How did researchers provide object detection solutions for drone images in various applications?

Motivation: The purpose is to research the applications where object detection in drones is considered and how the researchers provide solutions for it.

RQ3. Are there appropriate drone datasets available?

Motivation: The objective is to find and explore the available drone datasets. Hence, the researchers can get a list of drone datasets to work with it.

RQ4. What performance metrics are focused primarily on object detection in drone images or videos?

Motivation: To obtain the performance metrics used for object detection and to know how the model for object detection in drones is evaluated.

3.1. Identified research gaps

Despite many achievements, the deep learning methods, drone technologies and combination of both present many challenges that are yet to resolve.

(1) There are many unanswered questions there in the deep learning area such as why some network architecture performs better than the other and the lack of geometry understanding in the objective function. Developing an efficient unsupervised deep learning algorithm is significant research as labeling large data is an expensive task.
(2) The literature review reveals that the researchers use either deep learning or traditional image processing methods for drone object detection. Deep learning algorithms are good at both feature extraction and classification which makes it a better option. However, the drone has limitations in weight, size, and power consumption that make onboard processing a difficult task. It is more challenging when large image data has to be transmitted but there is less availability of bandwidth. These challenges motivate researchers to develop more efficient deep learning architectures.
(3) In real-time applications, the use of the internet in UAV opens up both security and privacy risk. There is a need for dedicated security measures for UAV applications.

4. Solutions to research questions

This section addresses each research question in detail.

RQ1. What are all the latest research strategies to address object detection challenges in drones?

Solution: The object detection methods are classified into two namely, traditional image processing methods and deep learning methods (Fig 2).

4.1. Traditional image processing methods

4.1.1. Background subtraction

Background subtraction is a popular method applied in most traditional image processing tasks. It focuses on revealing foreground objects by subtracting the pixels of background scenes. There are three phases in this process namely, background initialization, background maintenance, and background pixel classification. The initial step is to calculate the temporal frame difference (Eq. (1)). The frame difference at a specific time t is,

$$FD_t = |p_t(x, y) - \ p_{t-1}(x, y) \tag{1}$$

Where $p_t(x, y)$is the pixel value of t frame and $p_{t-1}(x, y)$ is the pixel value of the t-1 frame. A significant motion of pixels is considered if it reaches the threshold T and is referred to as a foreground (Eq. (2)).

$$FG_t = 1 \qquad \begin{cases} 1 & if\ FD_t > T \\ 0 & otherwise \end{cases} \tag{2}$$

If there is no large movement, the pixels are considered to be stable or a background pixel. The Background difference frame can be formulated as (Eq. (3)),

$$BD_t = |p_t(x, y) - \ BM_{t-1}(x, y) \tag{3}$$

Where $B_{t-1}(x, y)$ is the background model. Finally, the pixels are classified as foreground or background using the following equation (Eq. (4)),

$$FG_t = \begin{cases} 1 & if\ D_t(x, y) \geq T \\ 0 & otherwise \end{cases} \tag{4}$$

4.1.2. Continuously adaptive mean shift tracking (CAMShift)

CAMShift is a significant object tracking method based on colors. It is primarily focused on a mean-shift tracking method and was originally proposed to monitor human-faces in a user interface framework. This approach has the benefit of adjusting the search window. The mean-shift method is a step-by-step technique that selects a search window and gives us a history of the position, type, shape, and size of the object. The window's center of mass is determined and converges with the center of the window. The steps are repeated before the window ends. The algorithm for CAMShift is calculated using the steps below:

(1) Select the initial position of the search window.
(2) Execute mean-shift
   a Calculate the mean position in the search window.
   b The search window is placed based on the mean position determined in the previous step.
   c Repeat 2.a and 2.b before convergence is reached.
(3) Set the size of the search window which equals the zero moment function described in Step 2.

Search window size is calculated as in Eq. (5),

$$s = 2 * \sqrt{\frac{m_{00}}{256}} \tag{5}$$

Where, m_{00} is the zeroth moment.

Dramouskas, Perikos & Hatzilygeroudis (2012) aim to detect and track a given object in an indoor environment. The overall framework of this method is that initially, the drone camera searches for the desired object concerning the size and color of the object. If no object is detected,

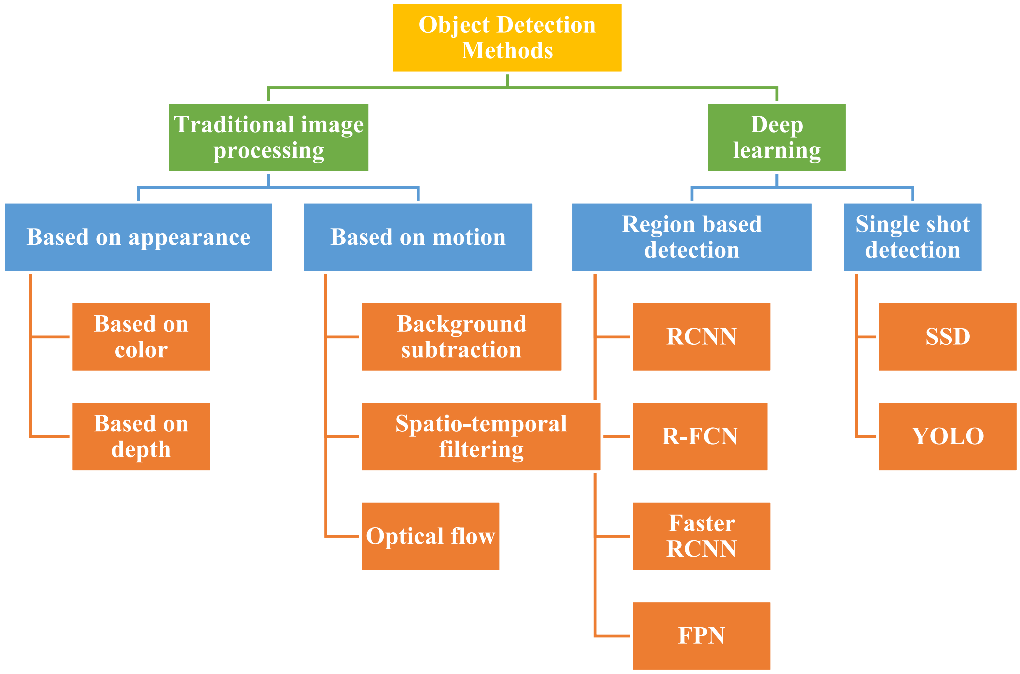

Fig 2. Classification of object detection methods.

then the drone rotates and moves to the next location. The image captured by the drone is BGR color format which is later converted to HSV color format. In this stage, the objects which don't match with the color of the desired object are all rejected. The result is a binary image. Finally, the position and the angle of the object are calculated to track it. Kamate & Yilmazer (2015) proposed an object detection and tracking model for surveillance purposes. The model aims to find out unusual events occurring on drone videos. The video sequences are captured by Phantom UAV. Object detection is performed based on the adaptive background subtraction method. Object tracking is performed using Lucas Kanade and Cam Shift methods. Fradi, Bracco, Canino & Dugelay (2018) proposed a color-based detection framework for individual humans. Parrot AR drone is used in the experiment. The recorded video includes both indoor and outdoor environments, varying light conditions which makes the model robust. Kadouf & Mustafah (2013) develop a drone that transmits video in real-time to the base station. A color-based algorithm named the HSV model is implemented to detect objects. The resultant output is given to the control system to track the object. Teutsch & Kruger (2012) proposed a three-layer framework to detect and track multi objects. The first layer separates the UAV images as stationary or moving with the help of local features. In the second layer, the algorithms such as moving feature clustering and segmentation based on appearance are applied for object hypothesis. Finally, in the third layer objects are tracked using the Kalman filter. This framework achieves better accuracy in terms of detection and tracking. Saif, Prabuwono & Mahayuddin (2015), Saif, Prabuwono & Mahayuddin (2014) proposed a feature extraction algorithm for moving object detection based on moments rather than color, corner, or edges.

4.2. Deep learning methods

Object detection algorithms based on deep learning approaches in UAV images have witnessed several innovations in recent times (Fig. 3). The main challenge for the object detection algorithms is view point variations in images as the dataset captured from top view angle. It is also difficult to learn the features from different angle and it is not transferable. Data preprocessing is the most important step in deep learning

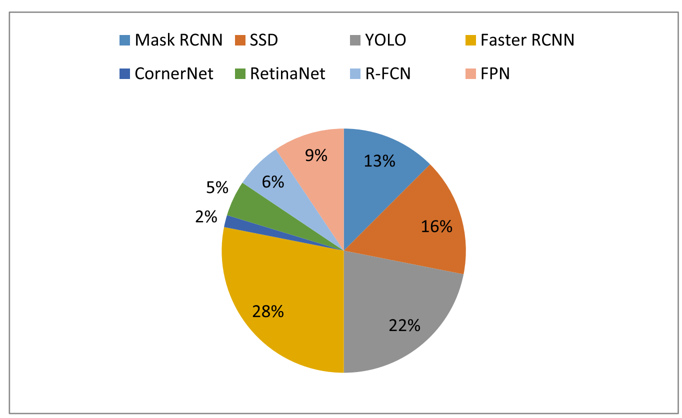

Fig. 3. Relative number of deep learning papers published in the UAV domain.

based classification. Wang et al. (2021) discussed advanced preprocessing methods which can be applied in deep learning based classification.

The deep learning methods for object detection can be classified into types namely, region-based and single-shot detection.

4.2.1. Region-based detection

Region based detection identify the objects in two stages. Each stage generates regions of interest and classify them in a deep learning network. Region-based detection algorithms use the regional proposal approach to establish regions of interest (ROI) for object detection. In Selective Search (SS), the model starts with each pixel as a group of its own. First, it measures the texture for each category and combines the next two. But to avoid a single field, it tends to group a smaller group first. The model continues to merge regions until all are merged. Previous object detection methods learn different tasks such as localization, classification and prediction of bounding boxes using a unified CNN network. Later, RCNN shows a great improvement in object detection challenge. RCNN calculates object location and crops them. Finally each object are classified using a deep learning network. To reduce the computational time of training, Faster R- CNN is proposed.

4.2.1.1. Faster R-CNN. The Faster R-CNN network (Ren, He, Girshick & Sun, 2016) consists of a feature extraction network that is usually a pre-trained CNN. This is followed by the two subnetworks that can be trained. The first is the Region Proposal Network, which, as its name implies, is used to generate object proposals and the second is used to predict the actual class of the object. The main difference in Faster R-CNN is RPN is added in after the last convolutional layer. This mechanism helps to produce region proposals without a selective search approach. Next to this step, ROI pooling, the upstream classifier, and the bounding box regressor are attached consequently. Faster R-CNN performs better when compared to other region-based detectors and it has several advantages. Some of them are as follows:

- Quality of detection in terms of mAP is higher than the RCNN and fast RCNN.
- Both classification and regression training was performed in a single stage.
- Memory storage is not required for feature extraction.
- Region proposal Network designs ROI which makes the system to run in real time frame rates.

The most significant challenge of computer vision is to detect objects in multiple scales. Feature Pyramid Network (FPN) is able to generate multi scale feature representation at high resolution levels which are semantically strong. Han, Shen & Liu (2016) proposed a framework based on an embedded system named Deep Drone. It is used for both detection and tracking. The convolution neural network is implemented for object detection. The proposed system is implemented on both desktop and embedded GPU. This work achieves fast and accurate results. Wang, Cheng, Liu & Uzochukwu (2018) studied different convolutional neural network-based algorithms on the Stanford drone dataset. Subash, Srinu, Siddhartha, Harsha & Akkala (2020) implement mask RCNN for object detection. Ryze Tello drone is used to capture images and videos.

4.2.2. Single shot detection

Object detection algorithms like region based detection achieved remarkable success in accuracy. However, the speed rate is inefficient. Single shot detection methods has high speed and less memory requirement compared to region based detection approaches. Algorithms based on the single-shot detection method take only a single shot to recognize many objects using the multi-box in an image. It is substantially faster in speed and high accuracy because it avoids bounding box proposals such as those used in RCNN. Also, it includes a gradually decreasing convolution filter to predict object classes and offsets in locations.

Rohan, Rabah & Kim (2019) proposed a deep learning-based model to perform single class detection in drone images. Parrot AR Drone 2 is used to capture images and image analysis is computed in PC. The drone and PC are connected via WiFi. The authors applied SSD (Single Shot Detector) to detect objects as they consider region-based algorithms are computationally heavy. The output of SSD is given to the PID controller to track the object. The PID controller considers a 3D plane that is all x,y, and z-axis. The approach reaches better accuracy and better computational time which makes it suitable for real-time applications. Budiharto et al. (2018) proposed the MobileNet SSD object detection model. The model achieves better results and can be implemented in real-time situations.

4.2.2.2. YOLO. Yolo Redmon, Divvala, Girshick & Farhadi (2016) implemented a new strategy to object detection. The feature extraction module and object localization module are combined into a unique monolithic entity. Besides, the heads of localization and classification are also combined. This single-stage architecture results in quick inference time. Together with the other detectors based on MobileNet, this new strategy has brought the vision of edge devices ever close to existence. The concept behind YOLO is this: there are no separate classification or detection modules that should be synchronized with each other and no repeated region proposals as in region-based detectors. The unique monolithic network performs all tasks such as feature extraction, boundary box regression, and classification. It contains only one output layer with different features.

Hossain & Lee (2019) experimented implementation of deep learning algorithms in different processing environments. They tested state-of-the-art deep learning algorithms in onboard GPU processing, onboard GPU constrained processing, and off-board GPU processing. The authors proposed the DeepSort algorithm for target tracking. According to the experiments carried out, Jetson AGX Xavier GPU is more powerful and Jetson Xavier GPU is feasible for a small deep learning model. The authors suggested that Yolo v2 tiny version is good at computation speed but it is not able to detect the far objects. Wang & Zhang (2020) studied the effects of tuning hyperparameter in one of the pretrained model DenseNet 201. Radovic, Adarkwa & Wang (2017) aim to detect airplanes from drone images. Yolo algorithm is used for detection and it shows promising results in this work. (Tijtgat et al., 2017b) also used the Yolo network to detect hazard symbols in drone images. Lee, Wang, Crandall, Sabanovic & Fox (2017) move all the computation to the cloud to detect objects in real-time. Small navigation and low-level object detection tasks are kept on board. Experiments show this work is suitable for multiple object detection in real-time. Vandersteegen, Beeck & Goedeme (2019) proposed a multi dataset learning approach for faster and accurate object detection. This strategy outperforms the transfer learning approach when implemented with the YOLO algorithm.

RQ 3. How did researchers provide object detection solutions for drone images in various applications?

Solution: Object detection has a range of uses in drone images or videos. To solve the problems within it, researchers focus on particular applications. Some of the works are summarized in Table 1. These works are classified and further described.

4.3. Applications

4.3.1. Disaster response and recovery

Disaster information is significant for effective disaster management. Data collection through drones is affordable. Simultaneously, it produces high-resolution images. The main challenge in disaster management using drones is to process a large amount of image data and map the object of interest in real-time. Pi, Nath & Behzadan (2020) proposed an aerial video dataset named Volan2018. The authors implemented the YOLO algorithm for object detection. Transfer learning is applied to train the learning model. First, the YOLO model is pretrained with the COCO dataset and then trained with the Volan2018 dataset. The authors also studied the effect of height, balanced data, and weights. This work achieved better results in less time. Kyrkou & Theocharides (2020) proposed an aerial database of the disaster which includes many types of disaster images taken in birds-eye view. The authors also proposed an EmergencyNet object detection model based on a convolutional neural network. The proposed EmergencyNet model has shown good accuracy and faster result when compared to existing methods. Lygouras et al. (2019) proposed a human detection system for search and rescue operations in disaster management. The author used Yolo v3 deep learning algorithm to detect humans that run on board. The work focused on detecting swimming humans on water. The work implements both global navigation satellite systems and computer vision techniques to achieve efficiency. WiSAR mission focuses on detecting and tracking the lost people in remote wild areas. Drones can be highly useful in this mission as they can fly, detect, and track objects. We cannot expect stable communication in remote wild areas. So, drones need to be more intelligent and autonomy such that it is less dependent on communication and human effort. It is a great challenge to design the drone of full autonomy in the aspect of perception. Al-Kaff, Gómez-Silva, Moreno, de la Escalera & Armingol (2019) proposed a human detection and tracking algorithm in wild areas. The human detection algorithm is based on the color and the depth present in drone images. Convolutional

Table 1
Summary of existing works on object detection in drone images or videos.

S.no	Author & year	Objective	Methodology	Approach	Dataset	Advantage	Disadvantage	Metric
1	Boonpook et al. (2018)	Building detection	SegNet	Deep learning	Own data, Inria Aerial Image Labeling, ISPRS Potsdam semantic labeling	The buildings in drone images are correctly detected, despite the complex characteristics, multiple design patterns, and different designs	Difficult to classify small buildings	mIoU, Accuracy
2	(Dramouskas et al. (2012)	Detect and track objects	Converting image to HSV	Traditional image processing method	Own data	Detects and tracks an individual object in real-time	The color feature is only considered as a prominent element. It is possible to consider object size and depth.	–
3	Chang et al. (2018)	Detect and track drones	Calculate object angle and position	Traditional image processing method	Own data	A good estimation of the drone's location which matches GPS	Accuracy is not calculated	Estimation error, CDF of the estimation error
4	Al-Kaff et al. (2019)	Detect and track humans	TDOA estimation algorithm and Kalman filter	Traditional image processing method	ARMOT dataset	Robust to complex and unconstrained scenarios	Large variations of the people visual representations affect false negative	Multi-object tracking accuracy
5	Hong et al. (2019)	Bird detection	Human shape validation filter, Colour, and depth-based detection	Deep learning	Own dataset	Good accuracy	Difficult to detect smaller birds	
6	Kamate & Yilmazer (2015)	Object detection and tracking	Faster RCNN and YOLO	Traditional image processing	Own data	Less computational cost	Not compared with other methods	Reference velocity,
7	Wyder et al. (2019)	Drone detection	Background subtraction, optical flow	Deep learning	Synthetic data	Works on GPS denied environment	Accuracy and detection speed can be improved	Velocity estimation, distance estimation
8	Fradi et al. (2018)	Person detection	Tiny YOLO	Traditional image processing	Own data	Good fps rate	Considers fewer features	Precision, Recall, IoU
9	Shao et al. (2020)	Cattle detection and counting	Color segmentation and matching	Deep learning	Own data	A decrease in FP and FN in cattle crowded scenes.	Less performance on fast-moving cattle	–
10	Rohan et al. (2019)	Person detection and tracking	YOLO v2	Deep learning	Own data	Less computational time	Sometimes targeted object is missed	Fps, accuracy
11	Pi et al. (2020)	Object detection	SSD	Deep learning	Volan2018	Analysis of different combinations of YOLO model, parameters and report the best.	Only hurricane type of disaster is considered	Average tracking accuracy, average tracking error
12	Schumann et al. (2017)	Drone detection	Yolo	Deep learning	Own dataset	Good accuracy	Small objects are difficult to classify.	Precision, Recall, F measure, IoU
13	Sien et al. (2019)	Human activity recognition	CNN	Deep learning	KTH dataset	The pipeline method improves accuracy and speed	Images from the new camera angle fail to recognize human activity	IoU,
14	Hossain & Lee (2019)	Multi-object detection	SSD and LSTM	Deep learning	Own dataset	Good Performance achieved on embedded device	Yolo unable to detect small objects at a far distance	Sensitivity,
15	Gonzalez-Trejo & Mercado-Ravell (2020)	Human detection and counting	Yolo, SSD, and RCNN	Deep learning	Own dataset	Good performance in counting	Need more crowd patterns to improve accuracy	Precision, FPS
16	Rivas et al. (2018)	Cattle detection and counting	Faster RCNN	Deep learning	Own dataset	Better accuracy	Counting error occurs when the same animal crosses several times	mAP, IoU, Precision, Recall, F1 score
17	Mishra, Garg, Narang & Mishra (2020)	Human action detection	CNN	Deep learning	Okutama dataset	Detects multi actions with good performance	Waving hand action is focused more than the other actions.	Confusion matrix, false-positive and miss rate
18	Kyrkou et al. (2018)	Vehicle detection	Modified SSD	Deep learning	Own dataset	Faster compared to Yolo	The little decrease in sensitivity and precision scores	Accuracy, Loss

(*continued on next page*)

Table 1 (*continued*)

S.no	Author & year	Objective	Methodology	Approach	Dataset	Advantage	Disadvantage	Metric
19	Kyrkou & Theocharides (2020)	Object detection	DroNet	Deep learning	Own dataset	Multi-resolution images can be processed	Accuracy can be improved	CPU and Memory usage, FPS
20	Singh et al. (2018)	Violent activity detection	EmergencyNet	deep learning and Traditional image processing methods	AVI dataset	Good accuracy	Processing speed varies based on the number of instances in a frame	Crowd count
21	Wang et al. (2018)	Object detection	Scatternet Hybrid Deep Learning	Deep learning	Stanford drone dataset	Different algorithms are evaluated	Usage of the pre-trained network	Accuracy,
22	Budiharto et al. (2018)	Object detection	Different CNN architectures	Deep Learning	Own data	Real-time experiment	The image from the drone is similar to a CCTV image	Cattle count
23	Cai et al. (2020)	Object detection	MobileNet and SSD	Deep learning	UAVDT, CARPK, and PUCPR+	Achieved better accuracy with the usage of background and foreground attention model	Multi objects are not considered	IoU, Precision and recall, mAP
24	Perera et al. (2018)	Human detection and pose estimation	GANet	Deep learning	Randomcollected from many public datasets	Ideal accuracy	Training images are clutter-free background which is not suitable for the real-time scenario	IoU, Sensitivity, precision, FPS
25	Yong & Yeong (2018)	Human detection	RCNN and HOG	Deep learning	Own dataset	Good performance	Not suitable for the real-time scenario	FPS, Mean F1 score,
26	Wittstruck et al. (2020)	Hokkaido Pumpkin fruit detection	Random forest	Traditional image processing and machine learning	Own dataset	Robust to different light conditions	Hidden fruits are difficult to detect	Accuracy
27	Chen et al. (2019))	Strawberry detection and count	Faster RCNN	Deep learning	Own dataset	Good accuracy in occluded images	Immature fruits are difficult to detect as it appears like dead flowers	Accuracy, FP, TP
28	Yuan & Choi (2021)	Apple flower bud detection	Yolo v4	Deep learning	Own dataset	Determination of heating requirement based on detection is feasible	Average classification performance	Accuracy, AP, mAP
29	Chen et al. (2021)	Pest detection	Tiny Yolo v3	Deep learning	Own dataset	Optimized performance in both terms FPS and mAP	Accuracy is less when the image resolution is low	mAP, FPS
30	Osco et al. (2021)	Plant detection	CNN	Deep learning	Own dataset	Detects high dense crop plantations with good accuracy	It is challenging to detect both plants and the rows simultaneously	Precision, Recall, F measure

Pose Machine extract human skeleton pose. This proposed framework is implemented in simulation datasets also to prove the robustness of the algorithm.

4.3.2. Surveillance

Sien et al. (2019) proposed a hybrid drone surveillance system to monitor human activity. Human activity recognition is highly helpful to monitor and find out any abnormal activities for security concerns. The spatial feature is extracted by SSD whereas the temporal feature is extracted by LSTM. The model is implemented on the KTH dataset. The computation is performed both offline and online google cloud platform. This work achieves good results in streaming data. The main limitation is that the model depends on the angle position of humans. If the novel angle of humans is detected, then the model is failed to detect humans in the video. (Tijtgat et al., 2017a) performs object detection on drone videos using Yolo V2 deep learning algorithm on the Nvidia Jetson TX2 GPU platform. The author proposes a warning system based on the result of the object detection model. Gonzalez-Trejo & Mercado-Ravell (2020) studied human detection and counting using deep learning algorithms in drone videos. The authors experimented with two approaches namely, VGG19 with density map generation and Faster RCNN with ResNet as the backbone. The study concludes that VGG19 performs better than Faster RCNN when the crowd is dense. Singh, Patil & Omkar, (2018) proposed a framework which detects human who involved in violent activities. As a first step, the FPN network detects humans in the drone images. Next, the proposed SHDL model estimates the human pose. Finally, SVM will classify whether the activity is violent or not. Perera, Al-Naji, Law & Chahl (2018) implemented a Region-based Convolutional Neural Network to detect humans and Histogram Oriented Gradients to estimate human pose. The main advantage of this method is that it gives better accuracy in different distorted videos. Yong & Yeong (2018) proposes a human detection model using MobileNet and SSD to avoid illegal entries and activities in forest areas. In this work, a 3D Solo drone is used to capture videos in the forest area. AlDahoul, Sabri & Mansoor (2018) proposed a deep model that includes pre-trained CNN, supervised CNN, and HELM to detect humans in drone videos. The main advantage of this work is that this model is robust to positions, color, height, and orientations.

4.3.3. Bird detection

Human health depends on ecosystems. An abnormal change in ecosystems can affect human life. Hence, monitoring birds is essential to study their habitats and populations. Especially, migrant birds should be monitored as it may be the reason for spreading animal diseases. Traditional methods of monitoring birds are dropping count, line-transect count, total ground count, and aerial count. Among all methods, the aerial count method is considered to be a significant method as it covers all the areas where human access is difficult. However, aerial photography using large aircraft is expensive. To solve this problem, drones are used for aerial photography to monitor birds. Hong, Han, Kim, Lee & Kim (2019) capture aerial images of birds using commercial drones named k-mapper at Shihwa lake and Yeongjong island. Deep learning methods namely, Faster RCNN, RetinaNet, SSD, R-FCN, and YOLO are implemented and compared for accuracy. In this study, the boundary box for labeled images is predetermined. Hence it would be less adaptive for the study with different datasets.

4.3.4. Cattle detection

Grazing cattle management includes tasks such as cattle detection and cattle counting. Farmers need to know the location of cattle where it gazes. Monitoring cattle manually and counting is a hectic task. Hence, an automatic cattle management system solves this problem. Shao et al. (2020) used DJI Phantom 4 to capture images of cattle located in Japan. The author uses Yolo v2 for cattle detection and cattle counting is also implemented by merging the results of detection. Cattle counting is considered a great challenge as there is difficult to count both moving and standing cattle. Rivas, Chamoso, González-Briones & Corchado (2018) proposed a system to detect and count the cattle in real-time. The system comprises a multirotor drone, a Ground Control System, and software to show the results of the detection model. The convolutional neural network is used for cattle detection and counting.

4.3.5. Traffic monitoring

Traffic monitoring on roads includes tasks such as vehicle detection, counting, and tracking. The application of drones in traffic monitoring eases up following traffic regulations. Earlier researchers used traditional image processing methods such as background subtraction and Haar cascade classifier for vehicle detection. Kyrkou et al. (2018) proposed the DroNet which is based on the convolutional neural network for vehicle detection in traffic roads. The authors used a DJI Matrice100 drone and all the processing is on board. The proposed DroNet achieves better results in less time. Chen et al. (2016)) proposed a traffic monitoring system using fog computing. DJI Phantom 3 is used to monitor the vehicles and the information is sent to one laptop which acts as a fog computing node. Nexus 9 tablet is also used for extracting information from drones and stream video in real-time. Multiple vehicle tracking was achieved in this work. Cai et al. (2020) proposed the guided attention network based on features. Different features are fused using background attention methods. Foreground attention is implemented to localize the object. This proposed method is applied to three different datasets and showed promising results.

4.3.6. Civil industry

Building detection can be useful in emergencies. Boonpook et al. (2018) proposed a deep learning-based approach to detect buildings near the river bank. Deep architecture named SegNet is used for semantic segmentation. The image datasets are collected in the area of China by using the drone DJI Phantom 4 pro. These datasets are annotated and partitioned into training, validation, and testing sets. SegNet has an encoder, decoder, and pixel-based classification layer. To avoid overfitting, the authors used data augmentation and early stopping methods. The model is trained in standard desktop with NVIDIA Quadro K620 GPU. This method achieves better accuracy but the main limitation is that it is unable to detect small buildings.

4.3.7. Drone autonomous navigation

Palossi, Conti & Benini (2019) proposed a visual navigation engine that works in an indoor environment for surveillance purposes. The Dronet algorithm is proposed for the classification task. It is based on the convolution neural network and built upon the CrazyFile 2.0 nano quadrotor. Faza & Darma (2020) implemented a single shot detector algorithm to detect the object and this result is given as feedback to the drone controller to track it. Yu et al. (2018)) proposed landmark detection to assist drones for autonomous landing. A deep convolutional neural network named SqueezeNet is implemented for landmark detection. To prove the efficiency of the proposed work, the authors carried four different experiments. First, the landmark detection model is evaluated. Second, the robustness of the model is evaluated under different illumination conditions. Third, the robustness of the model is evaluated under different background situations. Finally, the efficiency of the model is evaluated. The proposed framework is tested under both real and simulated worlds.

4.3.8. Amateur drone detection

It is familiar that the drone has many real-time applications due to its flying ability and small size with onboard computation power. Hence, there is an increasing number of drone development and usage by common people. This threatens people with privacy and security issues. To ensure protection from amateur drones, many researchers work on detecting and tracking drones to deactivate them. State-of-the-art architectures for amateur drone monitoring are discussed by Kaleem & Rehmani, (2018). Ding et al. (2018) proposed a framework based on the cognitive internet of things named Dragnet for amateur drone surveillance. The main idea is to make surveillance devices to think, learn, and understand the world by themselves. Chang, Yang, Wu, Shi & Shi (2018) develop drone detection and tracking algorithm based on acoustic arrays. Time Difference Of Arrival (TDOA) algorithm is proposed to localize the drone and tracking of the drone is performed by the Kalman filter algorithm. Unlu, Zenou, Riviere & Dupouy (2019)) designed a system including both day and thermal cameras with high resolutions. The authors used the YOLO algorithm for drone detection and CNN for tracking it. Wyder et al. (2019) propose a UAV system to detect, track, and neutralize other illegal drones in GPS denied environments. A tiny YOLO deep learning algorithm is used for drone detection. It is trained in both real and simulated datasets. The simulated dataset is created using the AirSim simulator. The visual tracking system runs onboard Jetson TX2. This proposed system has the great advantage that it works even in the absence of GPS. Schumann, Sommer, Klatte, Schuchert & Beyerer (2017) proposed a deep learning-based framework for drone detection. The dataset is created in such a way that it includes drones and other flying objects. This framework has two modules where the first module is to find out the region where the drone is present and the second module is to classify whether the object is a drone or a bird. Unlu, Zenou, Riviere & Dupouy (2019) proposed a system of using the multi-camera to detect amateur drones. The image frames from the zoomed camera and wide-angle camera are combined. The Yolo deep learning algorithm is used for drone detection. Initial detection and detection on zoomed plane performs simultaneously which reduces the cost of resources. Unlu, Niehaus, Chirita, Evangeliou & Tzes (2019) proposed a system based on the PTZ camera to monitor amateur drones. Initially, a border is calculated using optical flow and HOG methods. Deep learning pre-trained model ResNet is trained offline to detect drones within the calculated border. Wu, Xie, Shi, Shao & Shi (2018) improves the Yolo network and implements it to detect intruder drones. Sun et al. (2020) constructed a drone detection framework named Tiny Iterative Backbone (TIB net). The main advantage of this method is that it can detect small drones.

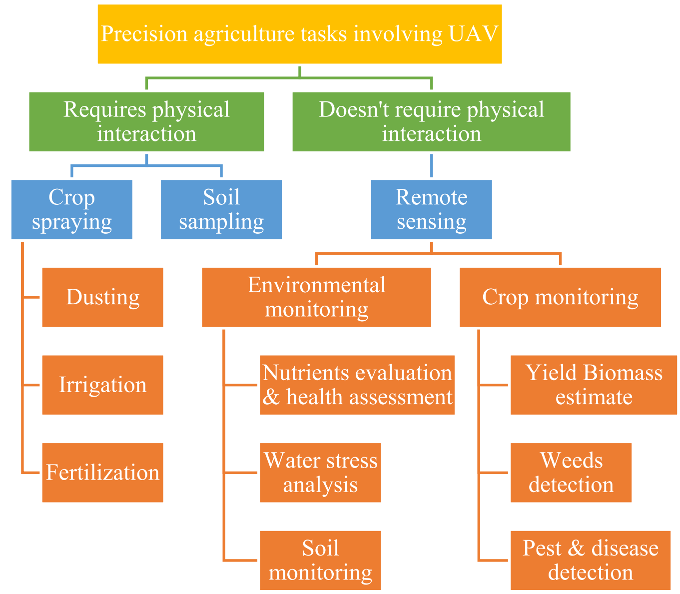

Fig 4. Precision agriculture tasks involving UAV.

4.3.9. Precision agriculture

The use of drone systems has increased in agricultural production, especially because drones and simple RGB cameras are becoming more affordable. High-resolution data from UAV becomes an important source in many agriculture tasks such as plant identification, plant tracking, and yield estimation (Fig. 4). Wittstruck, Kühling, Trautz, Kohlbrecher & Jarmer (2020) proposed a methodology based on traditional image processing to detect Hokkaido Pumpkin fruit and to estimate the size and weight of the fruit. DJI Phantom 4 drone with an onboard RGB camera is used to collect the UAV image data. The drone flight height

Fig 5. Some example of UAV images.

is chosen as 46 m above ground level. The proposed methodology consists of five steps namely, preprocessing, classification, morphological filtering, evaluation of single fruit identification, and fruit quantification. In the first stage, the collected UAV images are used to create an orthophoto mosaic. Image stitching and geo-referencing are also done as a part of pre-processing step. Random forest is used for fruit classification. To overcome misclassifications, morphological filters are applied. The proposed methodology is robust to various light conditions.

Chen et al. (2019) proposed a strawberry flower detection system to predict yield. A small UAV from DJI equipped with an RGB camera is used to take close-ground photographs of two varieties of strawberries. For the identification and count of flowers, mature strawberries, and immature strawberries, a Faster RCNN network is selected. Yield estimation is necessary to forecast stock volume, prevent stock ruptures, and schedule harvest operations. The study Apolo-Apolo, Martínez-Guanter, Egea, Raja & Pérez-Ruiz (2020) developed an automatic method for the identification, count, and estimation of citrus fruit size using deep learning techniques for individual trees. The encouraging results show that this measuring method can be used before harvesting for size discrimination, the authors further added. The authors concluded that the methodology can be used to estimate the yield of other fruit types. Kalantar, Edan, Gur & Klapp (2020) proposed a deep learning system based on RetinaNet to detect and estimate melons individually from top-view UAV images. DJI Phantom 4 Pro-drone is used to collect melon images. The UAV flight altitude is set to 15 m above the ground. A total mean precision score of 0.92 and an F1 score of more than 0.9 is achieved in the detection process. In the calculation, the weight of a single melon measured by the MAPE index is 16%. In this study Yuan & Choi (2021), the authors presented an evaluation technique of frost protection heating requirements in an apple orchard using thermal and RGB cameras mounted on UAV. Image stitching algorithm based on BRISK feature is also developed. Yolo v4 network is implemented to detect different stages of the apple flower bud. Chen et al. (2021) proposed a pest identification system. The authors used two drones namely, a detection drone and an agriculture drone. Detection drone capture images of pest which is given as input to object detection module. Tiny YOLO v3 is implemented in embedded GPU to recognize pests in real-time. Agriculture drone is used to spray optimum level of pesticide. UAV plays an important role in monitoring trees. However, due to imbalance and other issues, accuracy is still a major research challenge. In this study Zheng et al. (2021), an approach to detect oil palm trees and monitoring their growth is proposed. Faster RCNN and Refined Pyramid Features are combined to monitor individual tree growth. Weed infestation is considered a major threat to crop yield in agriculture fields. It requires appropriate weed mapping to analyze weed spatial distribution and implement accurate herbicide applications. A simple UAV was used in this study Rozenberg, Kent & Blank (2021) to survey 11 Allium cepa L (dry onion) commercial fields to analyze the classification of weeds and investigate the spatial pattern of weeds. To model and analyze the precision of weed mapping at various altitudes, they were resampled to a coarser spatial resolution. Overall, 176 maps of weeds were produced and assessed. Maximum Likelihood and Support Vector Machine algorithms are tested using object based image analyses.

RQ 4. Are there appropriate drone datasets available?

4.4. Datasets

Solution: Deep learning has emerged as a successful range of technologies to the latest demands for intelligent UAV operations, due to its outstanding potential of learning high-level features from the given raw data. In recent years, data-based analysis played a major role in research. Large datasets like ImageNet and MSCOCO played a key role in the image classification research area. A dataset similar to ImageNet and MSCOCO in terms of both the number of images and comprehensive annotation is not available in drone object detection, which is one of the main barriers for drone object detection research, particularly

Table 2
Summary of the aerial dataset for drone object detection and tracking.

S.No	Dataset	Author & year	Link	Objects	Number of samples	Resolution
1	UAVid	Lyu, Vosselman, Xia, Yilmaz & Yang (2020)	https://uavid.nl/	Car, Building, Low Vegetation, Human, Tree, Background clutter, and Road	300	*4096×2160 or 3840×2160*
2	Stanford Drone dataset	Robicquet, Sadeghian, Alahi & Savarese (2016)	https://cvgl.stanford.edu/projects/uav_data/	Skateboard, cart, Pedestrian, Bicyclist, bus, and car	929.5k	*1400×1904*
3	Okutama-Action dataset	Barekatain et al. (2017)	http://okutama-action.org/	Human actions	77.4k	*3840×2160*
4	Mini-drone	Bonetto, Korshunov, Ramponi & Ebrahimi (2015)	https://www.epfl.ch/labs/mmspg/downloads/mini-drone/	Human actions in the parking area	23.3 k	*1920×1080*
5	Drone-Action dataset	Perera, Law & Chahl (2019)	https://asankagp.github.io/droneaction/	Human actions	66.9k	*1920×1080*
6	VisDrone	Zhu et al. (2020)	http://aiskyeye.com/	Vehicles	10,209	*2000×1500*
7	UAVDT	Yu et al. (2020))	https://sites.google.com/site/daviddo0323/projects/uavdt	Vehicles	80k	*1080×540*
8	UAV20L, UAV123	Mueller, Smith & Ghanem (2016)	https://cemse.kaust.edu.sa/ivul/uav123	Person, car, and building	58.7k 110 k	*1280×720*
9	CARPK, PUCPR+	Hsieh, Lin & Hsu 2017	https://lafi.github.io/LPN/	Car	1448	*1280×720*
10	The UAV-Gesture dataset	Perera, Law & Chahl (2018)	https://asankagp.github.io/uavgesture/	Human actions	37.2 k	*1920×1080*
11	VEDAI512	Razakarivony & Jurie (2016)	https://downloads.greyc.fr/vedai/	Vehicles	1.2k	*1024×1024*
12	Drone face	Hsu & Chen (2017)	https://hjhsu.github.io/DroneFace/	Human face	620	*3680×2760*
13	Drone surf	Kalra et al. (2019)	http://www.iab-rubric.org/resources/dronesurf.html	Human face	411.5 k	*1280×720*
14	Deepweeds	Olsen et al. (2019)	https://github.com/AlexOlsen/DeepWeeds	Weeds	17.5 k	*1920×1200*
15	Agriculture-Vision	Chiu et al. (2020)	https://www.agriculture-vision.com/dataset	Agriculture patterns	94.9k	*512×512*

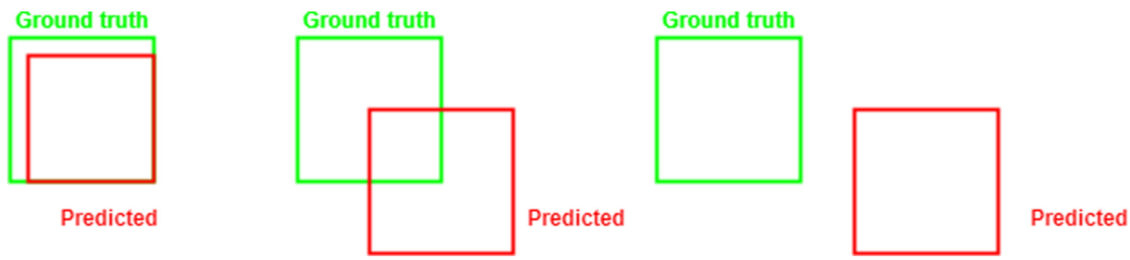

Fig 6. Boundary box based on IoU values. (a) Good IoU value (b) Average IoU value (c) Poor IoV value.

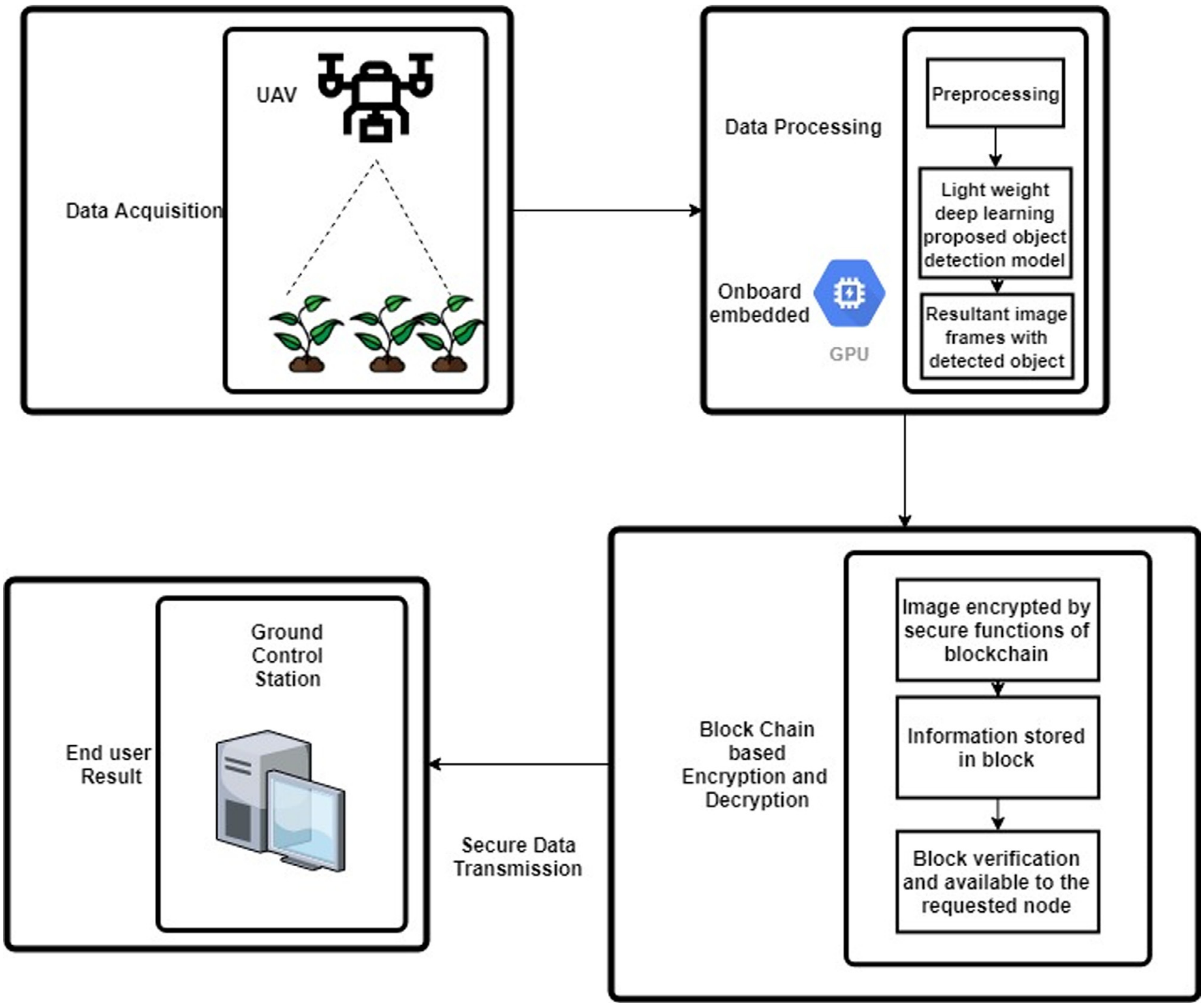

Fig 7. Proposed framework.

to develop deep learning algorithms. Some example of UAV images is given (Fig. 5). It is clear from the literature that; drone object detection has many real-time applications. Hence, a largescale benchmark dataset dedicated to the drone object detection research problem is crucial for advancing research in this field. Some of the available datasets are listed in Table 2.

RQ 5. What performance metrics are focused primarily on object detection in drone images or videos?

4.5. Evaluation metrics

Solution :

4.5.1. Intersection over union(IoU)

IoU is computed as the area of the intersection divided by the area of the union of two boxes namely prediction and ground truth boxes (Fig. 6). Hence, IoU (Eq. (6)) must be ≥ 0 and ≤ 1.

$$IoU = \frac{Area\ of\ Intersection}{Area\ of\ Union} \tag{6}$$

The object model should have an IoU value ~1.

4.5.2. Precision

Precision can be computed as the ratio of true predictions and the total predictions (Eq. (7)).

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive} \tag{7}$$

4.5.3. Recall

Recall can be computed as a ratio of true predictions and the actual ground truth positives (Eq. (8)).

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative} \tag{8}$$

4.5.4. F1 – score

F1 score is finding the harmonic mean of both precision value and recall value (Eq. (9)).

$$F1\ score\ = 2 * \frac{Precision * Recall}{Precision + Recall} \tag{9}$$

4.5.5. Mean average precision(mAP)

Average Precision (AP) is estimating the area under the precision-recall curve. The Precision recall curve is drawn concerning IoU threshold values. Mean average precision is finding the mean of AP over all available classes.

5. Proposed framework

The implementation of UAV and IoT in agriculture increases the risk of data privacy violations and cyber-attacks. Security concerns are related to the technical advancement of hacking in which adversaries seek to take advantage of the dignity of IoT and UAV. Such attacks allow the ability to remotely connect to on-site sensors and UAVs and control them. For instance, exploits that can ruin the whole field of crops, farmland flooding, the spraying of pesticides using UAV, etc. can cause insecure consumption and deterioration of the economy. Such attacks are mostly organized and recognized as agro-terrorism. To mitigate the identified research gaps, secured onboard processing for a robust object detection framework is proposed (Fig. 7). Initially, UAV is responsible for data acquisition, the images are pre-processed onboard an embedded GPU platform. A lightweight deep learning model is implemented for crop detection. The resultant images are securely transmitted to the ground control station using a blockchain-based encryption method. Implementing the proposed framework is our future work.

6. Conclusion and future work

Drone object detection is an important area of research as it has many applications in real-time. In this review paper, existing research works are studied. The works are categorized according to their applications and methods. This paper explores deep learning and traditional image processing methods used for drone object detection. Dataset and evaluation metrics are also discussed. The literature reveals that deep learning algorithms perform better than traditional image processing methods. Although the drones are small in size and restrictions in power, it is important to design an efficient deep learning algorithm for drone object detection. Variations viewpoint is another significant challenge to be addressed by the object detection algorithms. As a future work, we aimed to investigate object detection using UAV in agriculture application. Precision agriculture deals to monitor the fields precisely by collecting and analyzing the data. Acquiring aerial images using UAV are less expensive than the satellites. Object detection in UAV aerial images in precision farming is an important research area to be investigated. We proposed a secure onboard object detection framework in precision agriculture in this paper and implementing it will be our future work.

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.

References

AlDahoul, N., Sabri, A. Q., & Mansoor, A. M. (2018). Real-time human detection for aerial captured video sequences via deep models. *Hindawi, 2018*, 15.
Al-Kaff, A., Gómez-Silva, M., Moreno, F., de la Escalera, A., & Armingol, J. (2019). An Appearance-based tracking algorithm for aerial search and rescue purposes. *Multidisciplinary Digital Publishing Institute, 19*(3), 652. 10.3390/s19030652.
Apolo-Apolo, O. E., Martínez-Guanter, J., Egea, G., Raja, P., & Pérez-Ruiz, M. (2020). Deep learning techniques for estimation of the yield and size of citrus fruits using a UAV. *European Journal of Agronomy, 115*, Article 126030. 10.1016/j.eja.2020.126030.
Barekatain, M., Martí, M., Shih, H.-. F., Murray, S., Nakayama, K., Matsuo, Y., et al. (2017). Okutama-action: an aerial view video dataset for concurrent human action detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops* (pp. 28–35).
Belmonte, L. M., Morales, R., & Fernández-Caballero, A. (2019). Computer vision in autonomous unmanned aerial vehicles-a systematic mapping study. *Multidisciplinary Digital Publishing Institute, 9*(15), 3196. 10.3390/app9153196.
Bonetto, M., Korshunov, P., Ramponi, G., & Ebrahimi, T. (2015). Privacy in mini-drone based video surveillance. In *Proceedings of the 2015 11th IEEE international conference and workshops on automatic face and gesture recognition (FG): 4* (pp. 1–6).
Boonpook, W., Tan, Y., Ye, Y., Torteeka, P., Torsri, K., & Dong, S. (2018). A deep learning approach on building detection from unmanned aerial vehicle-based images in riverbank monitoring. *Multidisciplinary Digital Publishing Institute, 18*(11), 3921.
Boursianis, A. D., Papadopoulou, M. S., Diamantoulakis, P., Liopa-Tsakalidi, A., Barouchas, P., Salahas, G., & Goudos, S. K. (2020). Internet of things (IoT) and agricultural unmanned aerial vehicles (UAVs) in smart farming: a comprehensive review. *Internet of Things*, 100187.
Budiharto, W., Gunawan, A. A. S., Suroso, J. S., Chowanda, A., Patrik, A., & Utama, G. (2018). Fast object detection for quadcopter drone using deep learning. In *Proceedings of the 2018* 3rd international conference on computer and communication systems (ICCCS) (pp. 192–195). 10.1109/CCOMS.2018.8463284.
Cai, Y., Du, D., Zhang, L., Wen, L., Wang, W., Wu, Y., et al. (2020). Guided attention network for object detection and counting on drones. In *Proceedings of the 28th ACM international conference on multimedia* (pp. 709–717).
Carrio, A., Sampedro, C., Rodriguez-Ramos, A., & Campoy, P. (2017). A review of deep learning methods and applications for unmanned aerial vehicles. *Hindawi, 2017*, 1–13. 10.1155/2017/3296874.
Chandra, A. L., Desai, S. V., Guo, W., & Balasubramanian, V. N. (2020). Computer vision with deep learning for plant phenotyping in agriculture: A survey. *Advanced Computing and Communications*. 10.34048/ACC.2020.1.F1.
Chang, X., Yang, C., Wu, J., Shi, X., & Shi, Z. (2018). A surveillance system for drone localization and tracking using acoustic arrays. In *Proceedings of the 2018 IEEE 10th sensor array and multichannel signal processing workshop (SAM)* (pp. 573–577). 10.1109/SAM.2018.8448409.
Chen, C. J., Huang, Y. Y., Li, Y. S., Chen, Y. C., Chang, C. Y., & Huang, Y. M. (2021). Identification of fruit tree pests with deep learning on embedded drone to achieve accurate pesticide spraying. *IEEE Access : Practical Innovations, Open Solutions, 9*, 21986–21997. 10.1109/ACCESS.2021.3056082.
Chen, N., Chen, Y., You, Y., Ling, H., Liang, P., & Zimmermann, R. (2016). Dynamic urban surveillance video stream processing using fog computing. In *Proceedings of the 2016 IEEE second* international conference on multimedia big data (BigMM) (pp. 105–112).
Chen, Y., Lee, W. S., Gan, H., Peres, N., Fraisse, C., Zhang, Y., et al. (2019). Strawberry yield prediction based on a deep neural network using high-resolution aerial orthoimages. *Remote Sensing, 11*(13), 1584. 10.3390/rs11131584.
Chiu, M.T., .Xu, X., Wei, Y., Huang, Z., Schwing, A., Brunner, R. et al. (2020). Agriculture-vision: A large aerial image database for agricultural pattern analysis. ArXiv:2001.01306 [Cs, Eess]. http://arxiv.org/abs/2001.01306
del Cerro, J., Cruz Ulloa, C., Barrientos, A., & de León Rivas, J. (2021). Unmanned aerial vehicles in agriculture: A survey. *Agronomy, 11*(2), 203. 10.3390/agronomy11020203.
Ding, G., Wu, Q., Zhang, L., Lin, Y., Tsiftsis, T. A., & Yao, Y. D. (2018). An amateur drone surveillance system based on cognitive internet of things. *IEEE, 56*(1), 29–35.
Dramouskas, I., Perikos, I., & Hatzilygeroudis, I. (2012). A method for performing efficient real-time object tracing for drones. 6(2), 55.
Faza, A., & Darma, S. (2020). Implementation of single shot detector for object finding in drone platform. *Journal of Physics: Conference Series, 1528*, Article 012005. 10.1088/1742-6596/1528/1/012005.
Fradi, H., Bracco, L., Canino, F., & Dugelay, J. L. (2018). Autonomous person detection and tracking framework using unmanned aerial vehicles (UAVs). In *Proceedings of the 2018 26th European Signal Processing Conference (EUSIPCO)* (pp. 1047–1051). 10.23919/EUSIPCO.2018.8553010.
Gonzalez-Trejo, J., & Mercado-Ravell, D. (2020). Dense crowds detection and surveillance with drones using density maps. ArXiv:2003.08766 [Cs]. http://arxiv.org/abs/2003.08766
Han, S., Shen, W., & Liu, Z. (2016). Deep drone: object detection and tracking for smart drones on embedded system. 8.
Hii, M. S. Y., Courtney, P., & Royall, P. G. (2019). An evaluation of the delivery of medicines using drones. *Multidisciplinary Digital Publishing Institute, 3*(3), 52.

Hong, S. J., Han, Y., Kim, S. Y., Lee, A. Y., & Kim, G. (2019). Application of deep-learning methods to bird detection using unmanned aerial vehicle imagery. *Multidisciplinary Digital Publishing Institute, 19*(7), 1651. 10.3390/s19071651.

Hossain, S., & Lee, D. (2019). Deep learning-based real-time multiple-object detection and tracking from aerial imagery via a flying robot with GPU-based embedded devices. *Multidisciplinary Digital Publishing Institute, 19*(15), 3371. 10.3390/s19153371.

Hsieh, M. R., Lin, Y. L., & Hsu, W. H. (2017). Drone-based object counting by spatially regularized regional proposal network. In *Proceedings of the 2017* IEEE International Conference on Computer Vision (ICCV) (pp. 4165–4173). 10.1109/ICCV.2017.446.

Hsu, H. J., & Chen, K. T. (2017). Drone face: An open dataset for drone research. In *Proceedings of the 8th ACM on multimedia systems conference* (pp. 187–192). 10.1145/3083187.3083214.

Hwang, J., & Kim, H. (2019). Consequences of a green image of drone food delivery services: The moderating role of gender and age. *Wiley Online Library, 28*(5), 872–884. 10.1002/bse.2289.

Hwang, J., Kim, J. J., & Lee, K. W. (2020). Investigating consumer innovativeness in the context of drone food delivery services: Its impact on attitude and behavioral intentions. *Elsevier*, Article 120433. 10.1016/j.techfore.2020.120433.

Kadouf, H. H. A., & Mustafah, Y. M. (2013). Colour-based object detection and tracking for autonomous quadrotor UAV. *IOP Publishing, 53*(1), Article 012086. 10.1088/1757-899X/53/1/012086.

Kalantar, A., Edan, Y., Gur, A., & Klapp, I. (2020). A deep learning system for single and overall weight estimation of melons using unmanned aerial vehicle images. *Computers and Electronics in Agriculture, 178*, Article 105748. 10.1016/j.compag.2020.105748.

Kaleem, Z., & Rehmani, M. H. (2018). Amateur drone monitoring: State-of-the-art architectures, key enabling technologies, and future research directions. *IEEE Wireless Communications, 25*(2), 150–159. 10.1109/MWC.2018.1700152.

Kalra, I., Singh, M., Nagpal, S., Singh, R., Vatsa, M., & Sujit, P. B. (2019). Drone SURF: Benchmark dataset for drone-based face recognition. In *Proceedings of the 2019* 14th IEEE international conference on automatic face & gesture *recognition (FG 2019)* (pp. 1–7).

Kamate, S., & Yilmazer, N. (2015). Application of object detection and tracking techniques for unmanned aerial vehicles. *Elsevier, 61*, 436–441. 10.1016/j.procs.2015.09.183.

Kanellakis, C., & Nikolakopoulos, G. (2017). Survey on computer vision for UAVs: Current developments and Trends. *Journal of Intelligent & Robotic Systems, 87*(1), 141–168. 10.1007/s10846-017-0483-z.

Kim, H., Kim, D., Jung, S., Koo, J., Shin, J. U., & Myung, H. (2015). *October). Development of a UAV-type jellyfish monitoring system using deep learning* (pp. 495–497). IEEE.

Kyrkou, C., Plastiras, G., Venieris, S., Theocharides, T., & Bouganis, C.-. S. (2018). DroNet: Efficient convolutional neural network detector for real-time UAV applications. In *Proceedings of the 2018 design, automation & test in Europe conference & exhibition (DATE). 2018 design, automation & test in Europe conference & exhibition (DATE)*. 10.23919/DATE.2018.8342149.

Kyrkou, C., & Theocharides, T. (2020). Emergency net: Efficient aerial image classification for drone-based emergency monitoring using atrous convolutional feature fusion. *IEEE, 13*, 1687–1699.

Lee, J., Wang, J., Crandall, D., Sabanovic, S., & Fox, G. (2017). Real-time, cloud-based object detection for unmanned aerial vehicles. In *Proceedings of the 2017 first IEEE international conference on robotic computing (IRC)* (pp. 36–43). 10.1109/IRC.2017.77.

Lygouras, E., Santavas, N., Taitzoglou, A., Tarchanidis, K., Mitropoulos, A., & Gasteratos, A. (2019). Unsupervised human detection with an embedded vision system on a fully autonomous UAV for search and rescue operations. *Multidisciplinary Digital Publishing Institute, 19*(16), 3542.

Lyu, Y., Vosselman, G., Xia, G. S., Yilmaz, A., & Yang, M. Y. (2020). UA Vid: A semantic segmentation dataset for UAV imagery. *ISPRS Journal of Photogrammetry and Remote Sensing, 165*, 108–119. 10.1016/j.isprsjprs.2020.05.009.

Mishra, B., Garg, D., Narang, P., & Mishra, V. (2020). Drone-surveillance for search and rescue in natural disaster. *Elsevier, 156*, 1–10. 10.1016/j.comcom.2020.03.012.

Mueller, M., Smith, N., & Ghanem, B. (2016). A benchmark and simulator for UAV tracking. *Computer Vision - ECCV 2016, 9905*, 445–461. 10.1007/978-3-319-46448-0_27.

Nuijten, R. J. G., Kooistra, L., & De Deyn, G. B. (2019). Using unmanned aerial systems (UAS) and object-based image analysis (OBIA) for measuring plant-soil feedback effects on crop productivity. *Multidisciplinary Digital Publishing Institute, 3*(3), 54. 10.3390/drones3030054.

Olsen, A., Konovalov, D. A., Philippa, B., Ridd, P., Wood, J. C., Johns, J., et al. (2019). Deep weeds: A multiclass weed species image dataset for deep learning. *Scientific Reports, 9*(1), 2058. 10.1038/s41598-018-38343-3.

Osco, L. P., de Arruda, M. dos S., Gonçalves, D. N., Dias, A., Batistoti, J., de Souza, M., et al. (2021). A CNN Approach to simultaneously count plants and detect plantation-rows from UAV imagery. *ISPRS Journal of Photogrammetry and Remote Sensing, 174*, 1–17. 10.1016/j.isprsjprs.2021.01.024.

Palossi, D., Conti, F., & Benini, L. (2019). An open source and open hardware deep learning-powered visual navigation engine for autonomous nano-UAVs. In *Proceedings of the 2019 15th international conference* on distributed computing in sensor systems (DCOSS) (pp. 604–611).

Perera, A. G., Al-Naji, A., Law, Y. W., & Chahl, J. (2018a). Human detection and motion analysis from a quadrotor UAV. In Proceedings of the *IOP Conference series: Materials science and engineering: 405*. 10.1088/1757-899X/405/1/012003.

Perera, A. G., Law, Y. W., Chahl, J., Leal-Taixé, L., & Roth, S. (2018b). UAV-GESTURE: A dataset for UAV control and gesture recognition. In *Proceedings of the european conference on computer vision (ECCV)* (pp. 117–128). Springer International Publishing. (Vol. 11130). 10.1007/978-3-030-11012-3_9.

Perera, A. G., Law, Y. W., & Chahl, J. (2019). Drone-action: an outdoor recorded drone video dataset for action recognition. *Multidisciplinary Digital Publishing Institute, 3*(4), 82. 10.3390/drones3040082.

Pi, Y., Nath, N. D., & Behzadan, A. H. (2020). Convolutional neural networks for object detection in aerial imagery for disaster response and recovery. *Elsevier, 43*, Article 101009. 10.1016/j.aei.2019.101009.

Radovic, M., Adarkwa, O., & Wang, Q. (2017). Object recognition in aerial images using convolutional neural networks. *Journal of Imaging, 3*(2), 21. 10.3390/jimaging3020021.

Razakarivony, S., & Jurie, F. (2016). Vehicle detection in aerial imagery: A small target detection benchmark. *Elsevier, 34*, 187–203. 10.1016/j.jvcir.2015.11.002.

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).

Ren, S., He, K., Girshick, R., & Sun, J. (2016). Faster r-CNN: Towards real-time object detection with region proposal networks. *IEEE, 39*(6), 1137–1149.

Rivas, A., Chamoso, P., González-Briones, A., & Corchado, J. M. (2018). Detection of cattle using drones and convolutional neural networks. *Multidisciplinary Digital Publishing Institute, 18*(7), 2048. 10.3390/s18072048.

Robicquet, A., Sadeghian, A., Alahi, A., & Savarese, S. (2016). Learning social etiquette: Human trajectory understanding in crowded scenes. In Proceedings of the European conference on computer vision (pp. 549–565).

Rohan, A., Rabah, M., & Kim, S. H. (2019). Convolutional neural network-based real-time object detection and tracking for parrot AR drone 2. *IEEE, 7*, 69575–69584. 10.1109/ACCESS.2019.2919332.

Rozenberg, G., Kent, R., & Blank, L. (2021). Consumer-grade UAV utilized for detecting and analyzing late-season weed spatial distribution patterns in commercial onion fields. *Precision Agriculture*. 10.1007/s11119-021-09786-y.

Saif, A. F. M. S., Prabuwono, A. S., & Mahayuddin, Z. R. (2014). Moving object detection using dynamic motion modelling from UAV aerial images. *The Scientific World Journal, 2014*, 1–12. 10.1155/2014/890619.

Saif, A. F. M. S., Prabuwono, A. S., & Mahayuddin, Z. R. (2015). Moment feature based fast feature extraction algorithm for moving object detection using aerial images. *PloS One*, 11.

Schumann, A., Sommer, L., Klatte, J., Schuchert, T., & Beyerer, J. (2017). Deep cross-domain flying object classification for robust UAV detection. In Proceedings of the *2017* 14th IEEE international conference on advanced video and signal based surveillance (AVSS) (pp. 1–6). 10.1109/AVSS.2017.8078558.

Shakhatreh, H., Sawalmeh, A. H., Al-Fuqaha, A., Dou, Z., Almaita, E., Khalil, I., et al. (2019). Unmanned aerial vehicles (UAVs): A survey on civil applications and key research challenges. *IEEE, 7*, 48572–48634.

Shao, W., Kawakami, R., Yoshihashi, R., You, S., Kawase, H., & Naemura, T. (2020). Cattle detection and counting in UAV images based on convolutional neural networks. *Taylor & Francis, 41*(1), 31–52. 10.1080/01431161.2019.1624858.

Sien, J. P. T., Lim, K. H., & Au, P. I. (2019). Deep learning in gait recognition for drone surveillance system. *IOP Conference Series: Materials Science and Engineering, 495*, Article 012031. 10.1088/1757-899X/495/1/012031.

Singh, A., Patil, D., & Omkar, S. N. (2018). Eye in the sky: Real-time drone surveillance system (DSS) for violent individuals identification using Scatter Net Hybrid Deep Learning Network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops* (pp. 1629–1637).

Subash, K. V. V., Srinu, M. V., Siddhartha, M., Harsha, N. S., & Akkala, P. (2020). Object detection using Ryze Tello drone with help of mask-RCNN. In Proceedings of the 20*20 2nd international conference on innovative mechanisms for industry applications (ICIMIA)* (pp. 484–490).

Sun, H., Yang, J., Shen, J., Liang, D., Ning-Zhong, L., & Zhou, H. (2020). TIB-net: Drone detection network with tiny iterative backbone. *IEEE, 8*, 130697–130707.

Teutsch, M., & Kruger, W. (2012). Detection, segmentation, and tracking of moving objects in UAV videos. In Proceedings of the *2012* IEEE ninth international conference on advanced video and signal-based surveillance (pp. 313–318). 10.1109/AVSS.2012.36.

Tijtgat, N., Ranst, W. V., Goedeme, T., Volckaert, B., & Turck, F. D. (2017a). Embedded real-time object detection for a UAV warning system. In Proceedings of the Proceedings of the IEEE international conference on computer vision workshops (pp. 2110–2118).

Tijtgat, N., Volckaert, B., & De Turck, F. (2017b). Real-time hazard symbol detection and localization using UAV imagery. In Proceedings of the *2017 IEEE 86th vehicular technology conference (VTC-Fall)* (pp. 1–5). 10.1109/VTCFall.2017.8288259.

Unlu, E., Zenou, E., Riviere, N., & Dupouy, P. E. (2019a). Deep learning-based strategies for the detection and tracking of drones using several cameras. *Springer, 11*(1), 7. 10.1186/s41074-019-0059-x.

Unlu, E., Zenou, E., Riviere, N., & Dupouy, P. E. (2019b). An autonomous drone surveillance and tracking architecture. *Electronic Imaging, 2019*(15) 35-1–35-37.. 10.2352/ISSN.2470-1173.2019.15.AVM-035.

Unlu, H. U., Niehaus, P. S., Chirita, D., Evangeliou, N., & Tzes, A. (2019c). Deep learning-based visual tracking of UAVs using a PTZ camera system. In Proceedings of the IECON 2019 - 45th annual conference of the IEEE industrial electronics society (pp. 638–644). 10.1109/IECON.2019.8927731.

Vandersteegen, M., Beeck, K. V., & Goedeme, T. (2019). Super accurate low latency object detection on a surveillance UAV. In Proceedings of the *2019 16th international* conference on machine vision applications (MVA) (pp. 1–6).

Wang, S. H., & Zhang, Y. D. (2020). Dense Net-201-based deep neural network with composite learning factor and precomputation for multiple sclerosis classification. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM), 16*, 1–19 (2s).

Wang, S., Celebi, M. E., Zhang, Y. D., Yu, X., Lu, S., Yao, X., et al. (2021). Advances in data preprocessing for biomedical data fusion: An overview of the methods, challenges, and prospects. *Information Fusion*, 3171–3175.

Wang, X., Cheng, P., Liu, X., & Uzochukwu, B. (2018). Fast and accurate, convolutional neural network based approach for object detection from UAV. In Proceedings of the IECON 2018-44th annual conference of the IEEE industrial electronics society (pp. 3171–3175).

Wittstruck, L., Kühling, I., Trautz, D., Kohlbrecher, M., & Jarmer, T. (2020). UAV-based RGB imagery for Hokkaido pumpkin (Cucurbita max.) detection and yield estimation. *Sensors, 21*(1), 118. 10.3390/s21010118.

Wu, M., Xie, W., Shi, X., Shao, P., & Shi, Z. (2018). Real-time drone detection using deep learning approach. In Proceedings of the international conference on machine learning and intelligent communications (pp. 22–32).

Wyder, P. M., Chen, Y. S., Lasrado, A. J., Pelles, R. J., Kwiatkowski, R., Comas, E. O. A., et al. (2019). Autonomous drone hunter operating by deep learning and all-onboard computations in GPS-denied environments. *Public Library of Science San Francisco, CA USA, 14*(11), Article e0225092. 10.1371/journal.pone.0225092.

Yao, H., Qin, R., & Chen, X. (2019). Unmanned aerial vehicle for remote sensing applications-a review. *Multidisciplinary Digital Publishing Institute, 11*(12), 1443. 10.3390/rs11121443.

Yong, S. P., & Yeong, Y. C. (2018). Human object detection in forest with deep learning based on drone's vision. In Proceedings of the *2018 4th international conference on computer and information sciences (ICCOINS)* (pp. 1–5).

Yu, H., Li, G., Zhang, W., Huang, Q., Du, D., Tian, Q., et al. (2020). The unmanned aerial vehicle benchmark: Object detection, tracking and baseline. *Springer, 128*(5), 1141–1159. 10.1007/s11263-019-01266-1.

Yu, L., Luo, C., Yu, X., Jiang, X., Yang, E., Luo, C., et al. (2018). Deep learning for vision-based micro aerial vehicle autonomous landing. *International Journal of Micro Air Vehicles, 10*(2), 171–185. 10.1177/1756829318757470.

Yuan, W., & Choi, D. (2021). UAV-Based heating requirement determination for frost management in apple orchard. *Remote Sensing, 13*(2), 273. 10.3390/rs13020273.

Zheng, J., Fu, H., Li, W., Wu, W., Yu, L., Yuan, S., et al. (2021). Growing status observation for oil palm trees using unmanned aerial vehicle (UAV) images. *ISPRS Journal of Photogrammetry and Remote Sensing, 173*, 95–121. 10.1016/j.isprsjprs.2021.01.008.

Zhu, P., Wen, L., Du, D., Bian, X., Hu, Q., & Ling, H. (2020). Vision meets drones: Past, present and future. ArXiv:2001.06303 [Cs]. http://arxiv.org/abs/2001.06303.