



OBJECT DETECTION USING UNMANNED AERIAL VEHICLE: A REVIEW

Shri Varun B G^{1,*}, Nireesh Jayarajan², Tamilselvan Ganesan³

¹Student, Department of Automobile Engineering, PSG College of Technology, Coimbatore, TamilNadu, India

²Assistant Professor, Department of Automobile Engineering, PSG College of Technology, Coimbatore, TamilNadu
India

³Research Scholar, Department of Automobile Engineering, PSG College of Technology, Coimbatore, TamilNadu
India

*Corresponding Author

ABSTRACT

Artificial intelligence researchers are very interested in computer vision in drones. Many real-time problems will be solved by providing intelligence to drones. Object identification, tracking, and counting are all important computer vision tasks for monitoring certain situations. However, altitude, camera angle, occlusion, and motion blur make it a more difficult endeavour. Methodology: A thorough literature evaluation concentrating on object recognition and tracking utilising UAVs in various applications has been undertaken in this work. This study reviews previous research publications' conclusions and indicates research needs. The object detection algorithms used in UAV photos are categorised and further upon. There is a collection of UAV datasets for object detection tasks. Existing research studies in various applications have been summarised. Finally, in precision agriculture, a secure on board processing system based on a strong object detection framework is presented to address identified research needs.

KEYWORDS: UAV, Object Detection, Deep Learning, Precision Agriculture.

1. INTRODUCTION

Computer vision has already made amazing advances as a result of advances in deep learning techniques, system requirements, and dataset availability. The most common research activity performed by researchers is object detection since it has so many uses. The goal of object detection is to recognise things from a certain category (for example, humans, dogs, automobiles, motorbikes, or cats) in a photograph and, if found, report the area and extent of each occurrence of objects. It serves as the foundation for complicated and high-level computer vision tasks including object tracking, segmentation, event detection, picture captioning, scene understanding, crowd monitoring, and activity recognition [1]–[4]. Researchers began to address the difficulty of developing universal object identification systems capable of detecting categories of things similar to human ones. Object detecting technology has advanced significantly. However, object identification in drone applications remains a work in progress. From surveillance to agriculture, all applications require excellent object detection to function properly.

Precision agriculture is predicted to expand significantly faster than other applications, as the usage of unmanned aerial vehicles (UAVs) becomes one of the most important components of managing agricultural activities. It is a set of strategies for tracking crops, gathering data, and performing educated crop management activities such as optimal water delivery and pesticide choices [5], [6]. Farm monitoring to evaluate crop growth and health, as well as planning and assessing agricultural plantations, are just a few of the many tasks that UAV may help farmers with. The advantages of employing airborne services in agriculture supported the spread of aerial fertiliser use in 1940 to other operations such as top dressing. Single-rotor UAVs may carry large payloads, but their mechanical complexity results in expensive costs. Multirotor UAVs are popular because they can be used by both experts and laypeople. It can hover or move along the supplied destination. Although fixed-wing UAVs may fly quickly and carry large payloads, they require a runway for take-off and landing. The hybrid UAV is an enhanced version of the fixed-wing, however it is still in the works. In addition to their aerodynamic designs, the UAVs may also be classified according to their autonomy standard. When the pilot provides references to each aircraft actuator, he or she may be classed as tele operated.

In drone videos, the camera is mounted at a greater height, and the scene has more contextual information. However, variations in view point and size make object recognition in drones more difficult than standard object detection [7], [8]. Drones capture traffic from the air in a traffic surveillance scenario. This has the advantage of being able to capture vehicle traffic from a height of 100 metres. For the following reasons, detecting objects from a bird's eye viewpoint is more difficult than from a front-parallel view.

- The dynamic movement of moving things.
- Modifications to the aspect ratio and picture scale.
- Rapid camera movement.
- Extremely severe perspective distortion.
- Blurred motion.
- The high object density.
- A complicated backstory.

In addition to these challenges, object identification investigations in aerial view face the biased dataset problem. To circumvent this issue, the dataset should be annotated to represent real-world applications. As a result, it is fairly uncommon for object identification algorithms learnt from standard photos to be inappropriate for aerial images.

2. TYPES OF DRONES

Drones are becoming more and more popular nowadays due to its wide range of applications like defence, agriculture, photography, deliver products etc. It's because of its convenience and quality of data that it provide from very challenging areas like mountains, forests. These flying machines are classified based on the size, flying time, range, area on application and number of rotor. The most obvious classification are made from the rotor blades as shown in the figure 1.

(a) Fixed Wing Drone

Fixed Wing Drones are the drone which doesn't have any propeller. Its fly by its own due to its forward motion (thrust) which is generated by the motor or the conventional IC Engines. In addition to that the fixed wing drone operates in very less power than any other drone because of the elimination of rotor. Due to that the UAV can operate on other power sources like solar power and wireless power sources like lasers. It is very efficient among other drones [9]–[11]. This type of drone have the ability to take a more payload as it does and rotating parts to carry. But for agricultural purposes this type of drone is useless because it needs a separate runway for its flight to take off, and also the control of this UAV is also very difficult for precise operations like agriculture.

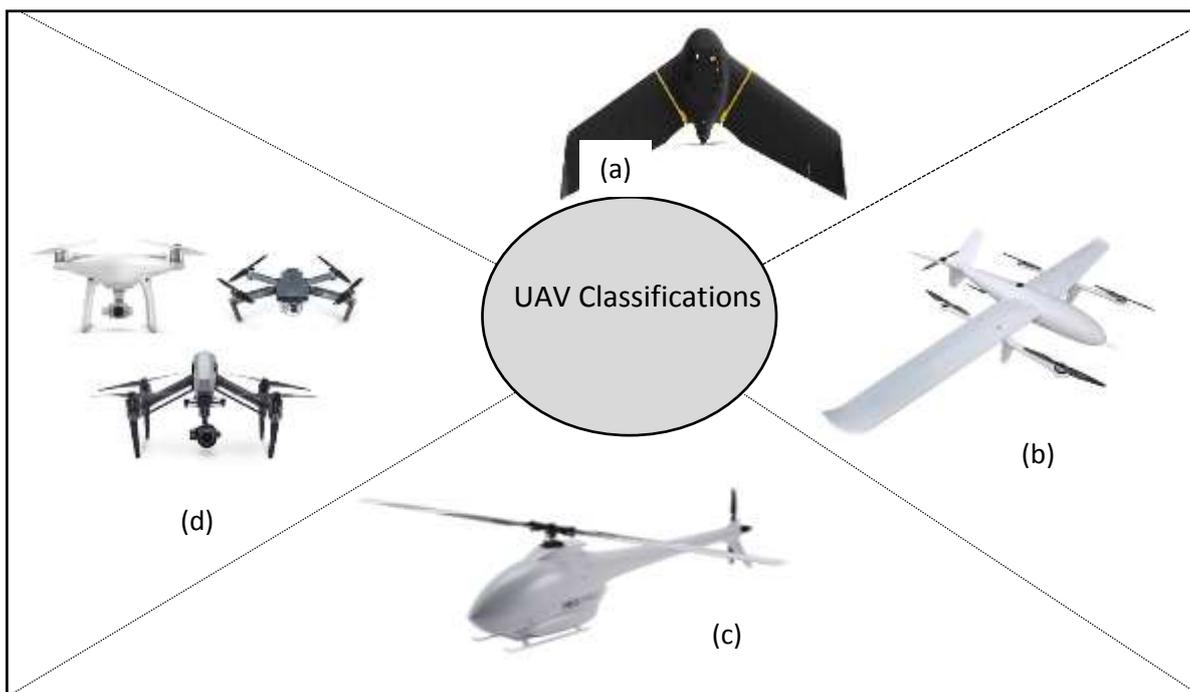


Figure 1. (a) Fixed Wing Drone (b) Fixed wing VTOL Drone (c) Single Rotor Drone (d) Multi Rotor Drone

(b) Fixed Wing VTOL Drone

Fixed wing VTOL drones are hybrid type drone because it have the benefits of both fixed wing drone (good flight time) and the rotor drone (take-off and landing). This type of drones are new to the market. These type of drones are very costly as compare to multi rotor drone. Large manufacture companies are using this type of drone for field monitoring. The Amazon Prime delivery project is based on this type of drones.

(c) Single Rotor Drone

Single rotor drones are actually a helicopter but unmanned and it is controlled by a Ground Control System (GCS). This drone consist of single rotor and the top for take-off. The tail end having a rotor which is used for the control of Drone. It has



longer range than the multi rotor drone but less than the fixed wing drone. It is used for military applications. These type of drones are not popular like multi rotor drones. Because of its high complexity and high operational risks.

(d) Multi-rotor Drone

Multi-rotor Drones are the one which has two or more rotors for the operation. This type of UAV are having the same rotor as that of other in terms of pitch. So that it generates a necessary lift force to tackle the gravity and the drag which is acting on the drone. One of the greatest advantages of this type of drone is the greatest manoeuvrability at a very precise rate. So that's make I one of the most suitable systems for the agriculture [12]–[14]. This system enables them to fly in areas which other drones can't even imagine to go over there. Anyhow they can't operate for long hours due to the efficiency of the drone. These types of UAVs are used in various applications especially in precise agriculture.

3. OBJECT DETECTION METHODS

The object detection methods are classified into two namely, traditional image processing methods and deep learning methods

3.1. Traditional Image Processing

Background removal is a widely used technique in most traditional image processing jobs. It concentrates on highlighting foreground items by removing pixels from background scenes [4], [15]. This procedure comprises three steps: background start-up, background upkeep, and background pixel categorization. The first step is to compute the temporal frame difference (Eq. (1)).

$$FD_t = [P_t(x, y) - P_{t-1}(x, y)] \tag{1}$$

Where $P_t(x, y)$ represents the pixel value of the t frame and $P_{t-1}(x, y)$ represents the pixel value of the t-1 frame. If the motion of pixels approaches the threshold T, it is regarded important and is referred to as a foreground (Eq. (2)).

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

If there is no significant movement, the pixels are called steady or background pixels. The background difference frame is denoted by (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 by applying the equation (Eq. (4))

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

CAMShift is an important color-based object tracking system. It was initially suggested to monitor human faces in a user interface framework, using a mean-shift tracking technique as its main emphasis. This method has the advantage of allowing you to customise the search window. The mean-shift approach is a step-by-step procedure that picks a search window and provides us with a history of the object's position, type, shape, and size [16], [17]. The centre of mass of the window is established and converges with the centre of the window. Before the window closes, the processes are repeated. The following stages are used to calculate the CAMShift algorithm:

- 1) Choose the search window's initial location.
- 2) Perform a mean-shift
 - a. Determine the average position in the search window.
 - b. The search window is positioned using the mean position determined in the preceding phase.
 - c. Repeat steps 2.a and 2.b until convergence is attained.
- 3) Set the search window size to equal the zero moment function mentioned in Step 2.
 Search window size is calculated using (Eq. (5))

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

Where, m_{00} is the Zeroth moment.

The drone captures images in BGR colour format, which is then transformed to HSV colour format. Objects that do not match the hue of the intended object are rejected at this step. The outcome is a binary picture. Finally, the object's location and angle are determined in order to track it. In [18] suggested a surveillance object identification and tracking methodology. The algorithm is designed to detect unexpected events in drone recordings. Phantom UAV captures the video sequences. Adaptive background subtraction is used to accomplish object detection. Cam Shift and Lucas Kanade techniques are used for object

tracking. In [19] suggested an individual human colour-based detection approach. The first layer uses local cues to classify UAV photos as fixed or moving. For object hypothesis, methods such as moving feature clustering and appearance-based segmentation are used in the second layer. Finally, the Kalman filter is used to track objects in the third layer. This approach achieves higher detection and tracking accuracy. In [8] author suggested a feature extraction approach based on moments rather than colour, corner, or edge for moving object recognition.

3.2 Deep Learning Methods

Object recognition methods based on deep learning approaches in UAV photos have seen a number of advancements in recent years. The key problem for object recognition algorithms is the variance in view point in pictures due to the dataset being recorded from a top view perspective. It's also tough to understand the characteristics from many perspectives, and it's not transportable. The most critical phase in deep learning is data preparation. In [20] addressed improved pre-processing strategies for deep learning-based classification. Deep learning algorithms for object detection are categorised into two types: region-based detection and single-shot detection.

3.2.1 Region Based Detection

Objects are identified in two steps using region-based detection. Each level produces and classifies areas of interest in a deep learning network. The regional proposal technique is used by region-based detection algorithms to define areas of interest (ROI) for item detection [20]–[22]. The model of Selective Search (SS) begins with each pixel as a separate group. It begins by measuring the texture of each category and then combines the following two. To prevent a single field, it is best to group a smaller group first. The model will continue to combine areas until they are entirely integrated. Using a unified CNN network, previous object identification algorithms learned diverse tasks such as localization, classification, and prediction of bounding boxes. Later, in the object detection task, RCNN exhibits a significant improvement as shown in figure 2. RCNN determines the position of objects and crops them. Finally, each item is categorised with the help of a deep learning. Faster R- CNN has been suggested to shorten training's computational time.

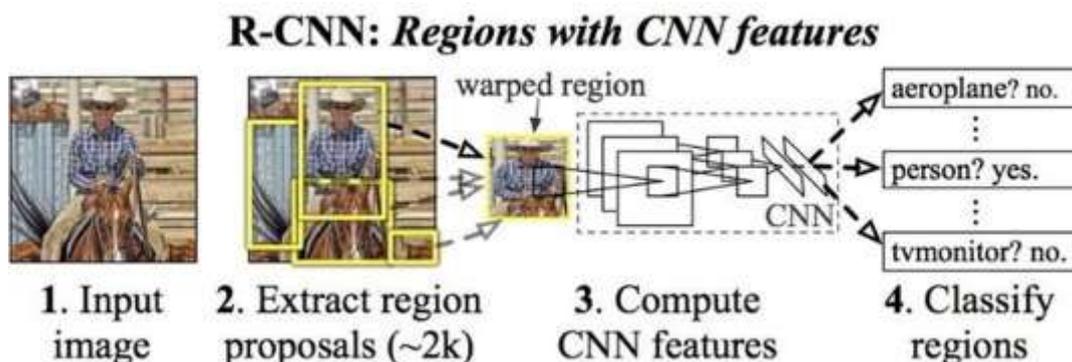


Figure 2. Region Based Object Detection

3.2.2 Faster R-CNN

The Faster R-CNN network is built on a feature extraction network, which is often a pre-trained CNN. Following that are the two trainable sub networks as shown in figure 3. The Region Proposal Network, as its name suggests, is used in the first to create object proposals, while the Actual Class Prediction Network is utilised in the second to forecast the actual class of the item. The primary distinction between Faster R-CNN and traditional R-CNN is that RPN is introduced after the final convolutional layer [6], [23]–[25]. This process aids in the generation of region ideas without the use of a selective search approach. The upstream classifier, the bounding box regressor, and ROI pooling are subsequently attached after this stage. R-CNN that is faster outperforms other region-based detectors. The most difficult difficulty in computer vision is detecting things at different sizes. The Feature Pyramid Network (FPN) can construct semantically strong multi scale feature representations at high resolution levels.

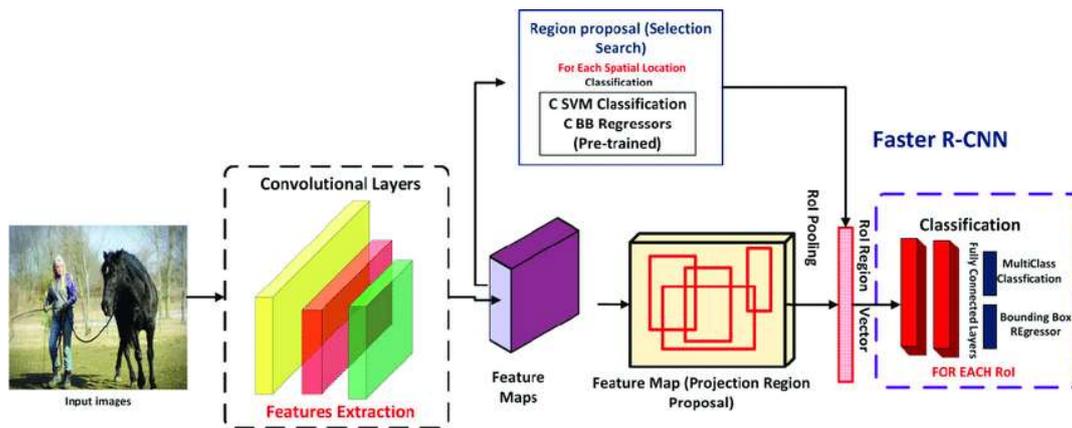


Figure 3. Methods Involved in Faster R-CNN

3.3 Applications

3.3.1. Disaster Response and Recovery

Disaster data is critical for efficient disaster management. Drone data collecting is inexpensive. It simultaneously generates high-resolution photos. The biggest difficulty in employing drones for disaster management is processing a significant volume of visual data and quickly mapping the target area. In [26]–[28] proposed the Volan2018 aerial video dataset. The authors used the YOLO method to detect objects. The learning model is trained via transfer learning. The YOLO model is first pre-trained with the COCO dataset before being trained with the Volan2018 dataset. Height, balanced data, and weights were also investigated by the writers. This project produced better outcomes in less time [16], [29].

3.3.2. Surveillance

Sien et al. [1] (2019) suggested a hybrid drone surveillance system to track human behaviour. Human activity identification is extremely useful for monitoring and detecting unusual behaviours for security issues. SSD extracts the spatial data, while LSTM extracts the temporal feature. The model is run on the KTH dataset. Google Cloud Platform is used for the computing both online and offline. This technique yields promising results with streaming data. The key limitation is that the model is dependent on human angle position. When humans are seen from a fresh viewpoint, the model has failed to recognise them in the video.

3.3.3. Bird Detection

The health of humans is dependent on ecosystems. An aberrant alteration in ecosystems can have an impact on human life. As a result, monitoring birds is critical for studying their habitats and population [30]. Migrant birds, in particular, should be observed since they may be the source of animal disease transmission. Traditional techniques of bird monitoring include dropping counts, line-transect counts, total ground counts, and aerial counts as shown in figure 4 (a). The aerial count method is regarded as an important approach since it covers all locations where human access is impossible. However, aerial photography with large planes is costly. Drones are utilised for airborne photography and bird monitoring to solve this issue.



Figure 4. (a) Bird Detection. (b) Cattle Detection. (c) Construction Applications (d) Agricultural Drone

3.3.4. Cattle Detection

Cattle identification and counting are two examples of grazing cattle management duties as shown in figure 4 (b). Farmers must be aware of the position of cattle where it gazes. Manually monitoring and counting animals is a time-consuming process. As a result, an autonomous livestock management system overcomes this issue. In [7] captured photographs of livestock in Japan using DJI. Using Yolo v2 for livestock detection, the author additionally implements cattle counting by combining the detection's findings. Livestock counting is considered a challenging task since it is difficult to count both moving and stationary cattle. Rivas (2018) presented a real-time detection and counting system for cattle. The system consists of a multi rotor drone, a Ground Control System, and software that displays the detection model's results.

3.3.5. Civil Industry



Building detection can be beneficial in an emergency. In (2018) [31] suggested a deep learning-based method for detecting structures along the river's edge. For semantic segmentation, a deep architecture known as SegNet is employed. The picture datasets were gathered in China using the DJI drone. These datasets are annotated and divided into three groups: training, validation, and testing. SegNet includes an encoder, decoder, and classification layer based on pixels. The authors employed early halting and data augmentation techniques to prevent over fitting. The model was trained on a regular desktop computer. This approach yields higher precision, but it is incapable of detecting tiny buildings as shown in figure 4 (c).

3.3.6. Precision Agriculture

The usage of drone systems in agricultural productivity has expanded, owing to the lower cost of drones and simple RGB cameras. UAV data with high resolution has become a significant source for various agricultural activities such as plant identification, plant tracking, yield calculation and spraying as shown in figure 4 (d). In 2020 [32] suggested a method for detecting Hokkaido Pumpkin fruit and estimating its size and weight based on classical image analysis. The UAV image data is collected using a drone with an inbuilt RGB camera. The drone flying height is set at 46 metres above ground. The suggested technique consists of five steps: pre-processing, identification, structural filtering, single fruit identification evaluation, and fruit quantification. The acquired UAV photos are utilised to construct an orthophoto mosaic in the first stage. As part of the pre-processing procedure, image stitching and geo-referencing are also performed. Random forest is used to classify fruits. Morphological filters are used to correct misclassifications. The suggested approach is resistant to changing lighting conditions.

4. CONCLUSION

Drone object identification is an essential topic of research since it has various real-time applications. Existing research efforts are examined in this review study. The works are classified based on their applications and techniques. This research investigates the use of deep learning and classical image processing approaches for drone object recognition. Despite the fact that drones are tiny and have limited power, it is critical to create an efficient deep learning system for drone object recognition. Another key problem that the object detection algorithms must solve is the viewpoint of variations. Precision agriculture is concerned with carefully monitoring fields through data collection and analysis. UAVs are less costly than satellites for acquiring aerial photographs. It is crucial to look at object recognition in UAV aerial photos for precision farming.

REFERENCES

1. J. P. T. Sien, K. H. Lim, and P. I. Au, "Deep Learning in Gait Recognition for Drone Surveillance System," in *IOP Conference Series: Materials Science and Engineering*, 2019, vol. 495, no. 1. doi: 10.1088/1757-899X/495/1/012031.
2. J. Hwang and H. Kim, "Consequences of a green image of drone food delivery services: The moderating role of gender and age," *Bus. Strateg. Environ.*, vol. 28, no. 5, 2019, doi: 10.1002/bse.2289.
3. J. Hwang, J. J. Kim, and K. W. Lee, "Investigating consumer innovativeness in the context of drone food delivery services: Its impact on attitude and behavioral intentions," *Technol. Forecast. Soc. Change*, vol. 163, 2021, doi: 10.1016/j.techfore.2020.120433.
4. R. J. G. Nuijten, L. Kooistra, and G. B. De Deyn, "Using unmanned aerial systems (Uas) and object-based image analysis (obia) for measuring plant-soil feedback effects on crop productivity," *Drones*, vol. 3, no. 3, 2019, doi: 10.3390/drones3030054.
5. M. S. Y. Hii, P. Courtney, and P. G. Royall, "An evaluation of the delivery of medicines using drones," *Drones*, vol. 3, no. 3, 2019, doi: 10.3390/drones3030052.
6. H. Kim, D. Kim, S. Jung, J. Koo, J. U. Shin, and H. Myung, "Development of a UAV-type jellyfish monitoring system using deep learning," 2015. doi: 10.1109/URAI.2015.7358813.
7. A. Rivas, P. Chamoso, A. González-Briones, and J. M. Corchado, "Detection of cattle using drones and convolutional neural networks," *Sensors (Switzerland)*, vol. 18, no. 7, 2018, doi: 10.3390/s18072048.
8. A. F. M. S. Saif, A. S. Prabuwo, and Z. R. Mahayuddin, "Moving object detection using dynamic motion modelling from UAV aerial images," *Sci. World J.*, vol. 2014, 2014, doi: 10.1155/2014/890619.
9. S. Sohail, S. Nasim, and N. H. Khan, "Modeling, controlling and stability of UAV Quad Copter," 2017. doi: 10.1109/ICIEECT.2017.7916559.
10. N. Chen, Y. Chen, Y. You, H. Ling, P. Liang, and R. Zimmermann, "Dynamic urban surveillance video stream processing using fog computing," 2016. doi: 10.1109/BigMM.2016.53.
11. A. F. M. S. Saif, A. S. Prabuwo, and Z. R. Mahayuddin, "Moment feature based fast feature extraction algorithm for moving object detection using aerial images," *PLoS One*, vol. 10, no. 6, 2015, doi: 10.1371/journal.pone.0126212.
12. F. G. Costa, J. Ueyama, T. Braun, G. Pessin, F. S. Osorio, and P. A. Vargas, "The use of unmanned aerial vehicles and wireless sensor network in agricultural applications," 2012. doi: 10.1109/IGARSS.2012.6352477.
13. M. Teutsch and W. Krüger, "Detection, segmentation, and tracking of moving objects in UAV videos," 2012. doi: 10.1109/AVSS.2012.36.
14. A. Olsen et al., "DeepWeeds: A Multiclass Weed Species Image Dataset for Deep Learning," *Sci. Rep.*, vol. 9, no. 1, 2019, doi: 10.1038/s41598-018-38343-3.
15. S. Razakarivony and F. Jurie, "Vehicle detection in aerial imagery: A small target detection benchmark," *J. Vis. Commun. Image Represent.*, vol. 34, 2016, doi: 10.1016/j.jvcir.2015.11.002.
16. E. Lygouras, N. Santavas, A. Taitzoglou, K. Tarchanidis, A. Mitropoulos, and A. Gasteratos, "Unsupervised human detection with an embedded vision system on a fully autonomous UAV for search and rescue operations," *Sensors (Switzerland)*, vol. 19, no. 16, 2019, doi: 10.3390/s19163542.
17. M. Mueller, N. Smith, and B. Ghanem, "A benchmark and simulator for UAV tracking," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2016, vol. 9905 LNCS.



- doi: 10.1007/978-3-319-46448-0_27.
18. S. Kamate and N. Yilmazer, "Application of Object Detection and Tracking Techniques for Unmanned Aerial Vehicles," in *Procedia Computer Science*, 2015, vol. 61. doi: 10.1016/j.procs.2015.09.183.
 19. H. Fradi, L. Bracco, F. Canino, and J. L. Dugelay, "Autonomous person detection and tracking framework using unmanned aerial vehicles (UAVs)," in *European Signal Processing Conference, 2018*, vol. 2018–September. doi: 10.23919/EUSIPCO.2018.8553010.
 20. X. Wang, P. Cheng, X. Liu, and B. Uzochukwu, "Fast and accurate, convolutional neural network based approach for object detection from UAV," 2018. doi: 10.1109/IECON.2018.8592805.
 21. S. Wang et al., "Advances in data preprocessing for bio-medical data fusion: An overview of the methods, challenges, and prospects," *Information Fusion*, vol. 76. 2021. doi: 10.1016/j.inffus.2021.07.001.
 22. S. H. Wang and Y. D. Zhang, "DenseNet-201-Based Deep Neural Network with Composite Learning Factor and Precomputation for Multiple Sclerosis Classification," *ACM Trans. Multimed. Comput. Commun. Appl.*, vol. 16, no. 2s, 2020, doi: 10.1145/3341095.
 23. Y. Chen et al., "Strawberry yield prediction based on a deep neural network using high-resolution aerial orthoimages," *Remote Sens.*, vol. 11, no. 13, 2019, doi: 10.3390/rs11131584.
 24. J. Lee, J. Wang, D. Crandall, S. Sabanovic, and G. Fox, "Real-time, cloud-based object detection for unmanned aerial vehicles," 2017. doi: 10.1109/IRC.2017.77.
 25. K. V. V. Subash, M. V. Srinu, M. R. V. Siddhartha, N. C. S. Harsha, and P. Akkala, "Object Detection using Ryze Tello Drone with Help of Mask-RCNN," 2020. doi: 10.1109/ICIMIA48430.2020.9074881.
 26. Y. Pi, N. D. Nath, and A. H. Behzadan, "Convolutional neural networks for object detection in aerial imagery for disaster response and recovery," *Adv. Eng. Informatics*, vol. 43, 2020, doi: 10.1016/j.aei.2019.101009.
 27. C. Kyrkou, G. Plastiras, T. Theocharides, S. I. Venieris, and C. S. Bouganis, "DroNet: Efficient convolutional neural network detector for real-time UAV applications," in *Proceedings of the 2018 Design, Automation and Test in Europe Conference and Exhibition, DATE 2018*, 2018, vol. 2018–January. doi: 10.23919/DATE.2018.8342149.
 28. C. Kyrkou and T. Theocharides, "EmergencyNet: Efficient Aerial Image Classification for Drone-Based Emergency Monitoring Using Atrous Convolutional Feature Fusion," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 13, 2020, doi: 10.1109/JSTARS.2020.2969809.
 29. A. Al-Kaff, M. J. Gómez-Silva, F. Miguel Moreno, A. De La Escalera, and J. M. Armingol, "An appearance-based tracking algorithm for aerial search and rescue purposes," *Sensors (Switzerland)*, vol. 19, no. 3, 2019, doi: 10.3390/s19030652.
 30. S. J. Hong, Y. Han, S. Y. Kim, A. Y. Lee, and G. Kim, "Application of deep-learning methods to bird detection using unmanned aerial vehicle imagery," *Sensors (Switzerland)*, vol. 19, no. 7, 2019, doi: 10.3390/s19071651.
 31. W. Boonpook, Y. Tan, Y. Ye, P. Torteeka, K. Torsri, and S. Dong, "A deep learning approach on building detection from unmanned aerial vehicle-based images in riverbank monitoring," *Sensors (Switzerland)*, vol. 18, no. 11, 2018, doi: 10.3390/s18113921.
 32. L. Wittstruck, I. Kühling, D. Trautz, M. Kohlbrecher, and T. Jarmer, "UAV-based RGB imagery for hokkaido pumpkin (*Cucurbita max.*) detection and yield estimation," *Sensors (Switzerland)*, vol. 21, no. 1, 2021, doi: 10.3390/s21010118.