Journal of Al-Qadisiyah for Computer Science and Mathematics Vol.18.(1) 2026,pp.Comp 281–301

Available online at www.qu.edu.iq/journalcm
JOURNAL OF AL-QADISIYAH FOR COMPUTER SCIENCE AND MATHEMATICS
ISSN:2521-3504(online) ISSN:2074-0204(print)

Vision-Based UAV Detection Methods Using Deep Learning: A Review

Firas Mohammed Kadhum [a,*], ***Abbas Abdulazeez Abdulhameed*** [a], ***Jameelah Harbi Saud*** [a]

[a]*Department of Computer Science, College of Science, Mustansiriyah University, Baghdad, Iraq. Email: firas.mohammed@uomustansiriyah.edu.iq, abbasabdulazeez@uomustansiriyah.edu.iq, dr.jameelahharbi@uomustansiriyah.edu.iq*

ARTICLE INFO

Article history:
Received: 19 /12/2025
Rrevised form: 26 /01/2026
Accepted : 28 /01/2026
Available online: 30 /03/2026

Keywords: UAV Detection, UAV Threats, UAV Datasets, Deep Learning, Computer Vision.

ABSTRACT

Unmanned Aerial Vehicles (UAVs), or drones are increasingly used for civilian and military purposes. But their misuse raises serious concerns related to privacy, safety and security making them double-edged weapons. Consequently, there is an urgent need for effective UAV detection systems to mitigate threats posed by unauthorized UAV operations over restricted territories. With rapid advances in deep learning and computer vision, vision-based UAV detection systems have achieved notable progress. However, the existing reviews often lack systematic algorithmic analysis and clear summarization of trends and limitations. Therefore, this review aims to consolidate and summarize recent vision-based UAV detection methods using deep learning, focusing on convolutional neural network (CNN)-based models and to provide actionable directions for future research. Firstly, this study presents the evolution of UAV detection, key challenges and the pros and cons of the technologies used. Next it presents a summary of the recent advances in UAV detection methods utilizing one- or two-stage detectors only; the literature shows a strong dominance of YOLO-based architectures due to their favorable accuracy–speed trade-off and suitability for real-time deployment. It further summarizes commonly reported evaluation metrics (e.g., precision, recall and F1-score). Finally, it systematically reviews public UAV datasets and their characteristics highlighting persistent dataset limitations, including limited diversity in altitude, weather, illumination and environment which contributes to a comprehensive understanding of their characteristics and applicability.

MSC..

https://doi.org/10.29304/jqcsm.2026.18.12504

1. Introduction

Unmanned aerial vehicle (UAV), remotely piloted aircraft system (RPAS), and unmanned aircraft system (UAS) are various terminologies for what are popularly referred to as drones [1][2]. In recent years UAVs have gained popularity across different sectors including civilian and military applications due to their compact size, flexibility and simple control. UAVs, such as those for package delivery, aerial photography and videography and several other activities, are present in urban regions, as illustrated in Figure 1. While operating in low airspace up to 400 feet above ground level they also support search and rescue operations and are utilized for agricultural purposes and the inspection of vital infrastructure in rural regions. Although developments in UAV technology have generated a wide array of applications and conferred meaningful benefits to society broadly, they concurrently raise profound concerns related to potential misuse, security and privacy. For example, UAVs may pose threats to critical

*Corresponding author: *Firas Mohammed Kadhum*

Email addresses: *firas.mohammed@uomustansiriyah.edu.iq*

Communicated by ‘sub etitor’

infrastructure including power stations, airports or government facilities. Also the operation of UAVs near densely populated public events or gatherings may raise safety issues.
To stop or prevent the aforementioned threats, effective and robust UAV detection systems are now crucial. Several studies have recently looked at methods to detect and identify different types of UAVs using a variety of technological developments, including audio, radio frequency (RF), image, video, and radar. Based on these technologies, many traditional methods exist for detecting and identifying malicious UAVs; however, most of these methods have not attained an adequate prevention rate upon detection of the UAVs.

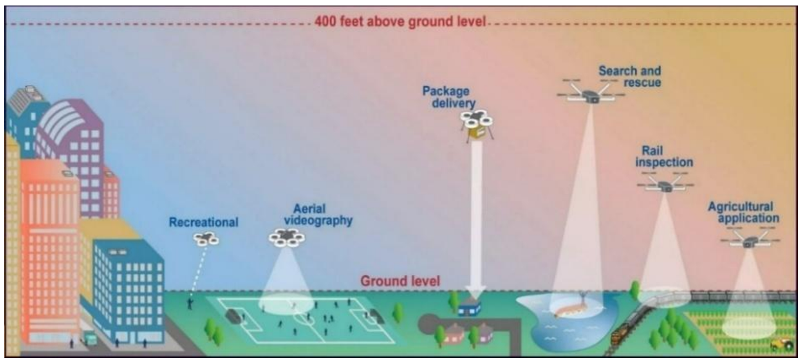

Fig. 1 - Some of the civilian and commercial uses of UAV [3].

In the past few years, many applications and tasks within computer vision, such as object detection [4][5], image segmentation [6][7], medical image analysis [8][9], and more, have been significantly influenced by the rising benefits of machine learning (ML) and deep learning (DL) techniques. As a result, the scientific community became increasingly interested in UAV detection and tracking following the emergence of technologies that employ DL techniques. Some of the rising advantages of DL for detecting UAVs are automatic feature learning, less intensive computing, data efficiency, and capabilities of end-to-end learning. However, there are some disadvantages to DL, including low performance on more sophisticated detection tasks and the requirement for a significant amount of labeled data for training, which can be difficult or costly to collect.
Recently, many review studies on object detection systems have been conducted. Most of them focus on using visual data (images or videos) for object detection and classification, especially in applications such as vehicle detection [10], forestry [11], crop disease identification [12], and emergency rescue operations [13], and a lot more. In contrast, this review study focuses on the vision-based detection of UAVs, or drones, that employ deep learning techniques, with particular emphasis on CNN-based models. Table 1 compares our contributions to several UAV detection review papers published in recent years.

Table 1 - Contributions of this review with other studies.

Ref	Year	Contributions
[14]	2020	Survey on object detection techniques that utilize deep learning and analyzed the current advancements in these techniques applied to low-altitude UAV datasets.The main focus of the study is on low-altitude UAV datasets, as they have contributed less to the literature compared to standard or remote-sensing-based datasets.
[15]	2021	Survey on drone detection and neutralization, with less focus on detection and classification techniques. It studies non-military anti-drone systems in detail.
[16]	2022	Review of existing UAV detection methodologies, highlighting their advantages and drawbacks. In addition to evaluating the trade-offs between performance metrics such as accuracy and detection range, the authors evaluate several detection technologies such as radar, acoustic, RF, and visual detection.
[17]	2023	A survey on various techniques for detecting and tracking UAVs in urban areas along with their technical classifications. It emphasized both the advantages and disadvantages of DL-based anti-UAV systems as it evaluated their effectiveness.
[18]	2024	Review of the current deep learning-based UAV detection methods. The study discusses UAV types, cybersecurity concerns, safety, privacy, and unethical behavior
[19]	2024	Survey on the UAV detection and classification methods as well as a comprehensive review of challenges, solutions, and future research directions utilizing machine learning.
Ours	2025	Review of vision-based UAV detection methodologies utilizing deep learning techniques. The main contributions are outlined as follows:(1) identify and describe the categories of UAVs. (2) present and discuss the most recent UAV-related incidents and threat categories. (3) present the UAV detection by highlighting the main challenges and the pros and cons of used technologies. (4) summarize and analyze recent vision-based methods for UAV detection employing deep learning with focused on CNN-based models. (5) offer the publicly available UAV datasets and evaluation metrics as well as some future research directions.

The remaining sections of this study are structured in the following order: Section 2 provides a concise overview of UAV detecting developments. Existing UAV detection technologies and challenges are discussed in Section 3, while Section 4 discusses vision-based UAV detection methodologies. A summary of the relevant datasets and evaluation metrics is provided in section 5, and the final section concludes the study and draws directions for future works. The review structure is shown in Figure 2.

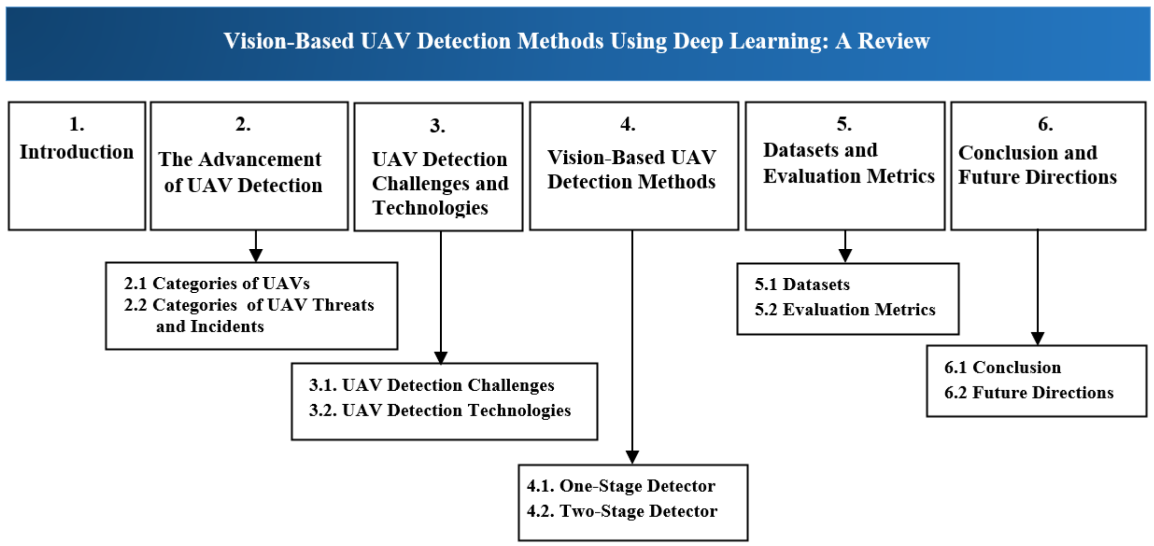

Fig. 2 - Overall structure of this review.

2. The Advancement of UAV Detection

The widespread use of remote sensing has made it possible for everyone to get data for any object on the surface of the Earth. The rapid advancement of object detection technology and the increasing popularity of UAVs have made it easier to collect data from UAV aerial photography (or UAV-based images). In consequence, aerial photography is now a widely used technique for gathering data. The UAVs are also in a state of rapid development owing to their cheap nature and their capacity to take pictures easily. They could be installed with high-resolution cameras and other sensors to provide the information about the object in a better and more precise form. Without extra expenditures and risking deaths, UAV platforms are able to detect objects in high-altitude or hazardous environments, unlike orbital and other aerial sensing gathering techniques. Therefore, many commercial and scientific research projects make use of UAV-based image acquisition systems. Table 2 shows how UAV-based images that are utilized in the process of object detection are categorized according to flight altitude and the areas where they can be used at various altitudes.

Table 2 - Comparison of UAV-based Images Categories.

Category	Altitude	Representation
Eye-level	0-5 m	Ground-level views.
Low and medium	5-120 m	The gap between most industrial and commercial uses.
High	>120 m	High-altitude data collection often requires particular authorization.

Nowadays, UAV detection can be divided into two scenarios of application. Ground-to-air is the first, which detects UAVs in flight using ground cameras that are typically stationary or moving slowly. The second scenario, air-to-air, involves a flying UAV detecting other flying UAVs by its onboard cameras. There are several categories of UAVs that must be detected and identified. In the next subsection, these categories will be reviewed.

2.1 Categories of UAVs

Various categories can be used to classify UAVs according to their size, design, payload, and flight characteristics. These categories were developed to meet a variety of demands and objectives [20]. Table 3 encompasses some of common categories of UAVs.

Table 3 - Categories of UAVs.

Category	Description
Fixed-wing UAVs:	Are designed for flight similarly to traditional aircraft utilizing a singular set of wings for stability and lift. Compared to other UAVs they are usually bigger and have a longer flying duration and range which makes them ideal for long-distance operations.
Rotary-wing UAVs:	Are designed for flight similar to helicopters utilizing rotors for stability and lift. Rotary-wing UAVs are better than fixed-wing UAVs for tasks that need vertical takeoff and landing like aerial photography and infrastructure inspections, because they are more flexible and agile.
Hybrid UAVs:	Are combined the attributes of rotary-wing and fixed-wing UAVs to provide their respective benefits on a single platform. They can therefore operate like traditional airplanes on long-distance trips but they can also switch to a helicopter-like mode during vertical takeoff and landing.
Micro UAVs:	Are lightweight, small and portable, which makes them ideal for jobs requiring mobility and stealth, such reconnaissance and surveillance operations.
Heavy-lift UAVs:	Are made to transport massive loads, such equipment and goods. They are ideal for jobs requiring the transfer of heavy goods since they are usually bigger and more robust than other kinds of UAVs.

Since each of the aforementioned UAV categories has unique advantages and disadvantages the category selected will rely on the specific demands of the mission at hand. For instance, a delivery mission that necessitates the transportation of substantial payloads may be well-suited to a heavy-lift UAV. Meanwhile for surveillance missions requiring mobility and stealth a micro-UAV may be a good selection. Even though these diverse categories of UAVs have varied operational benefits, their growing use in both civilian and restricted airspace has led to a number of safety and security incidents. Incidents such as collisions and near-collisions with airplanes, unauthorized surveillance and smuggling activities illustrate how the misuse or breakdown of UAVs can turn into serious threats in a variety of scenarios.

2.2 Categories of UAV Threats and Incidents

In recent years, the use of UAVs has rapidly expanded from military to civilian purposes. Also, as the use of UAVs expands across various fields, it becomes more crucial to be able to detect them properly to stop or prevent illegal or unsafe UAV activities. To show the importance of a UAV detection system, this subsection briefly addresses some threat categories in the context of specific incidents, as seen in Figure 3.

Fig. 3 - Main categories of threats caused by UAVs.

- **Physical Attacks:** These belong to the first category of potential threats, as they employ UAVs for loading explosives, biological, and chemical weapons. Such explosives can be employed for threatening various targets, including specific people, governmental or business organizations, and entire countries [21].

- **Smuggling:** It is regarded as the second most common threat category for UAVs. Drug smuggling by UAV has become a significant concern for border patrol and prison workers. In addition, there are cases in which weapons and illegal materials are smuggled outside of the area that is protected by ground-based security [21].
- **Espionage:** UAVs equipped with high-resolution cameras can be utilized to spy on individuals, businesses, and governmental entities remotely. But this concern may pose a threat, such as espionage or privacy invasion, despite the privacy claims that are made [22].
- **Collision or Near Collision**: Releasing a UAV, whether intentionally or unintentionally, near airports, stadiums, or other critical infrastructures, UAV incursions can halt them, cause financial losses, and cause safety risks [22].
- **Cyber Attacks**: Perhaps the most insidious threat is the UAV used as a cyber-weapon. These "flying computing devices" can hover close to a server room or land on a roof to connect air-gapped network [21].

A list of some of the recent UAV incidents is outlined in Table 4.

Table 4 - Some of the recent UAV-related incidents reported by the global media.

Date and Location	Sector	Threat Category	Incident Details
10 April 2023, Duhok, Iraq	Military and Special Forces	Physical Attacks	A villager found an unexploded fixed-wing UAV carrying explosives close to the Duhok dam. The UAV has been handed over to security forces for further investigation [23].
27 January 2025, Baton Rouge, Louisiana, USA	Prisons	Smuggling	Five persons were taken into custody for smuggling mobile phones and drugs into Pollock Federal Prison using UAVs [24].
12 March 2025, Baghdad, Iraq	Law Enforcement Agencies and First Responders	Espionage	A surveillance UAV, identified as an RTK300, encountered a crash in proximity to the First Battalion of the Federal Police facility. [25].
14 January 2025, Isle of Wight, England, UK	Prisons	Collision or Near Collision	A UAV that was thought to be carrying contraband crashed and was taken away close to the jail at HMP Isle of Wight [26].

The aforementioned incidents emphasize the necessity for UAV detection systems to halt and prevent undesirable UAV-related activities. The objective of this detection system is to detect and track UAVs, as well as determine their position, category, direction, velocity, and other characteristics. As a result, businesses and researchers have recently exhibited considerable interest in this field. The challenges in detecting UAVs have increased due to the quick development of UAV technology, which includes their diverse sizes, speeds, and altitudes of flight; these challenges are discussed later. Furthermore, UAVs are less detectable than traditional big-air targets and may easily breach no-fly zones. Consequently, it becomes more challenging to detect suspicious UAVs.

3. UAV Detection Challenges and Technologies

Generally, in the subsequent sections the most significant challenges will be covered along with technologies that are used for detecting UAVs.

3.1. UAV Detection Challenges

There are several important challenges with the UAV detection systems, for which researchers and business experts are working to find a solution. Here are some of these challenges:

- **Speed and size variations:** UAVs come in a variety of sizes, ranging from small to large. Most small UAVs operate at speeds of 15 m/s or less. Conversely, the maximum speed of large UAVs may reach 100 m/s [27]. The diversity in size and speed impacts UAV detection tasks since different UAVs may have different flight characteristics and forms.
- **Similarity with other aerial objects:** It is challenging to differentiate UAVs from other aerial objects because of their noticeable similarity to other airborne objects like birds and airplanes. Consequently, an UAV detection system needs to be precise enough to distinguish UAVs from other aerial objects while also being quick enough to identify UAVs in video frames. A drone vs. bird challenge is held annually since this problem is so urgent [28].

- **Diversity of detection ranges and altitudes:** The region from which a UAV may be commanded remotely is referred to as its range. Smaller UAVs have a range of only a few meters, but larger ones may go hundreds of kilometers. Aerial platforms might therefore be classified as either high altitude or low altitude. Low-altitude platforms are typically employed for malicious purposes because of their low cost and rapid deployment. So, the traditional defense systems may be ineffective or even dangerous due to potential adverse effects because malicious UAVs typically fly at low altitudes [28].
- **Environmental factors:** The efficiency of UAV detection systems can be affected by different environmental factors including diverse weather conditions (e.g. humidity, rain, fog or wind) and the environments of urban life (characterized by numerous obstacles such as buildings and vegetation). In UAV detection such circumstances might reduce the accuracy and reliability of sensors within radar-based or vision-based detection systems leading to errors like false positives or false negatives. As a result developing and enhancing detection and classification methods for various environmental circumstances is still a challenge [29].
- **Real-time processing:** It represents one of the biggest challenges. Due to their rapid speed and unpredictable flight patterns and behaviors, UAVs are extremely challenging to detect and accurately identify. As a result the UAV detection systems become more challenging because of the dynamic nature of UAVs. So, to detect and react quickly to UAVs, it would be necessary to process data in real time. To ensure real-time processing capabilities this calls for improvements to the hardware settings and algorithms, possibly utilizing edge computing, parallel computing or hardware acceleration [28].

All highlighted challenges raised motivate more research into developing novel UAV detection methods. Furthermore addressing these challenges can ensure the safety, security, and privacy of individuals as well as vital infrastructure in general.

3.2. UAV Detection Technologies

To address the aforementioned challenges the academic and industrial communities have been investigating a range of UAV detection technologies. Generally speaking, there are essentially four categories of UAV detection technologies: radar-based, RF-based, vision-based, and acoustic-based. Moreover, many researchers have begun focusing on hybrid sensor-based UAV detection techniques; however, these techniques are still in their infancy stages in terms of development and deployment. Table 5 presents a summary of these technologies used to enable UAV detection, as well as the advantages and disadvantages of each.

Table 5- Comparison of different categories of UAV detection technologies.

UAV Detection Technology	Description	Detection Range	Advantages	Disadvantages
Radar-based	UAVs are detected using their radar signature	1-20 km	▪ Less affected by weather ▪ Long-range detection	▪ High-cost ▪ Vulnerable to obstacles ▪ Complexity of deployment
RF-based	UAVs are detected by capturing the RF signals used for communications	3-50 km	▪ Obstacle-free ▪ Long-range detection ▪ Cost-effective ▪ Capability to differentiate various types of UAVs	▪ Lacking the ability to identify autonomous UAVs ▪ Potentially hackable ▪ Interfering with other RF signals
Vision-based	UAVs are detected by capturing visual data using optical cameras	0.5-3 km	▪ Visual confirmation ▪ Cost-effective ▪ Non-intrusive ▪ Simple configuration	▪ Limited detection range ▪ Vulnerable to obstacles ▪ High affected by weather
Acoustic-based	UAVs are detected using their unique sound signatures	< 0.2 km	▪ Works in low-visibility environments ▪ Cost-effective ▪ Quick deployment	▪ Low-range detection ▪ Vulnerability to wind condition
Hybrid Sensor-based	UAVs are detected using fusion methods of two or more of the above-mentioned technologies			

The above technologies are motivational in detecting objects, but the traditional technologies mentioned above mainly rely on the electromagnetic signal such as acoustic, RF, and radar, which are less intuitive as compared to that of object detection using the visual sensors (e.g., cameras that record images or videos), which provides one with a better perspective of the group information. This technology has several benefits, including the ability to record consecutive images of objects in real time, cost-effectiveness, rapid detection speed, and not being affected by low-altitude clutter, and more [30]. Other than the rapid progress of computer vision, many researchers are now considering and developing methods utilizing visual data (images or videos) for the purpose of UAV detection. The following sections of this review describe these methods and the datasets used in this context.

4. Vision-based UAV Detection Methods

The basic idea of a vision-based detection system involves utilizing camera sensors to acquire visual data of UAVs, including images and videos, and subsequently employing computer vision algorithms for object detection to identify UAVs within that data. Image acquisition is the process that captures visual data from objects, including three main steps: first, the object of interest reflects energy (e.g., a UAV); second, an optical system concentrates this energy; and third, the sensor in a camera quantifies the energy's amount. Visual images are often used for semantic segmentation and object detection due to their high resolution [19]. However, there are drawbacks to employing visual images as well, such as cluttered backgrounds, obscured areas, and changeable light. Moreover, detecting UAVs in visual imagery is usually challenging because of several factors, including the relatively small size of the UAV, potential confusion with birds, and occlusion, as mentioned before. These factors necessitate the development of comprehensive and efficient object detection methods. The development of object detection was rapid as convolutional neural networks (CNNs) gained popularity in 2012 [31]. Object detection methods usually utilize deep neural networks in order to remove high-level features from inputted images and subsequently classify them. These methods perform at high speed and with high accuracy. This advancement in DL technology has enabled faster and more accurate object detection methods utilizing visual data. Consequently, in recent years, DL-based object detection methods has become widely adopted in vision-based object detection research, particularly for UAV detection. These models fall into two main categories: models based on CNNs and models based on transformers, as shown in Figure 4.

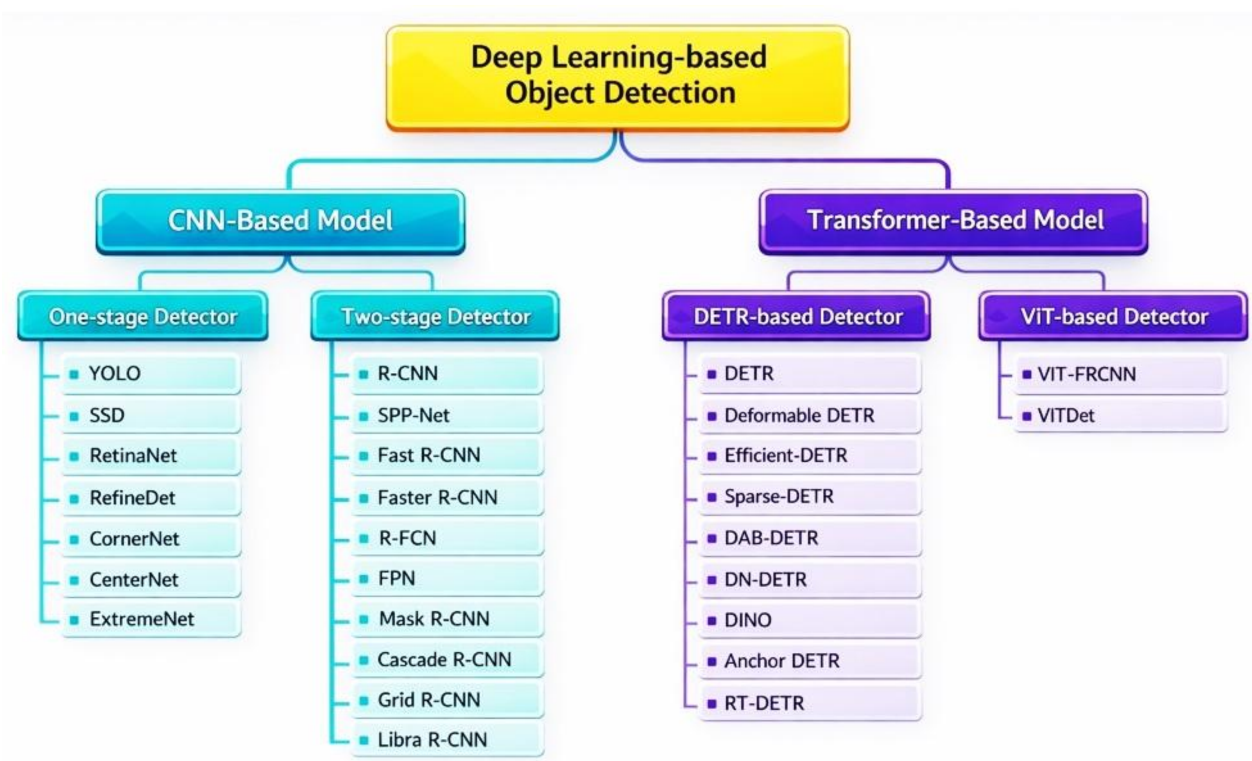

Fig. 4 – Diagram of deep learning-based methods for object detection.

CNN-based models benefit from CNN's hierarchical structure which increases their computational performance and learning capability. As a result, they would be more adaptive to complex datasets than the traditional object detection methods would be. Figure 5 illustrates an example of how to use the CNN-based model for basic UAV detection with visual data.

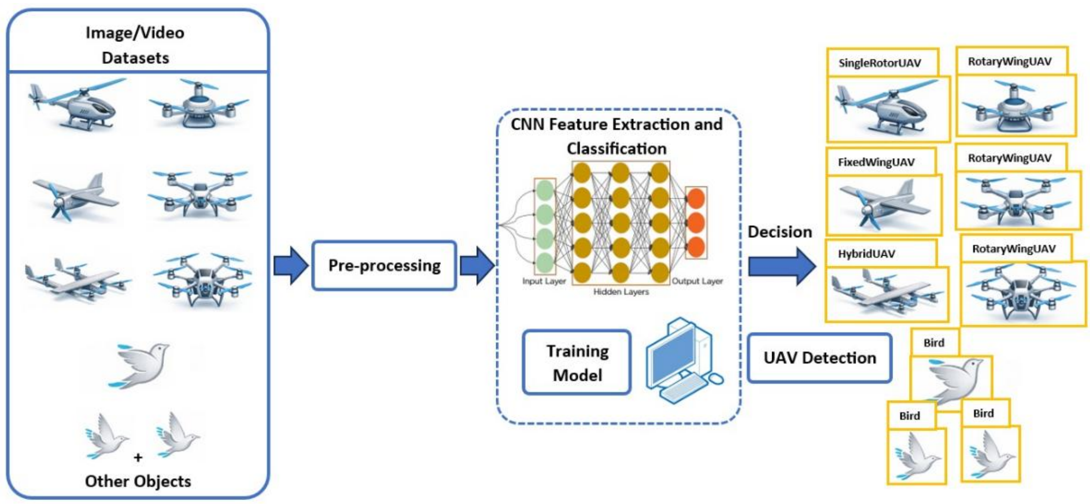

Fig. 5 - UAV detection mechanism using visual data.

The CNN-based models can be divided into two major types: one-stage detectors (sometimes referred to as single-stage detectors) and two-stage detectors (sometimes referred to as multi-stage detectors). Two-stage detectors only process region proposals during the second stage. Their focus includes boundary object detection, localization inaccuracies, and multiscale object detection. The initial stage of these algorithms generates proposal regions, followed by classification and regression of the contents within the regions of interest in the subsequent stage. These algorithms demonstrate high accuracy; however, they exhibit slow performance. A few well-known two-stage detectors are region-based CNN (R-CNN) [32], Faster R-CNN [33], Mask R-CNN [34], Grid R-CNN [35], etc. As previously stated, because two-stage detectors require two stages to do detection, a large number of proposal regions in the first stage generates, resulting in a lot of processing and slow performance, thus making them not suitable for real-time applications. These issues are resolved by one-stage detectors. One-stage object detection algorithms are quicker because they eliminate the proposal region process and generate location and category information immediately from the image. However, since there are so many densely generated region proposals, these detectors are prone to erroneous detection when localizing and detecting small objects. There are numerous examples of one-stage detectors, including You Only Look Once (YOLO)[36], Single Shot Detector (SSD)[37], RetinaNet[38], etc.

The Transformer-based models leverage the self-attention process for establishing global dependencies between the various points in sequence [39] that let the model calculate the attentional weights between positions dynamically. These models can be broadly categorized into DEtection TRansformer (DETR)-based models [40] and Vision Transformer (ViT)-based models [41]. The DETR series model streamlines object detection into an end-to-end framework by leveraging a Transformer encoder–decoder. DETR-based, unlike the CNN-based model, eliminates the non-maximum suppression (NMS) procedure, which is a hand-designed component for dealing with redundant boxes. On the other hand, the ViT series model split the entire images into fixed-size patches, with each patch containing a part of the image area. After that, the pixel values for each patch are stacked into a vector, and then these vectors are used as the model's input.

The following two subsections will only discuss the CNN-based models used in UAV detection methods with one- and two-stage detectors.

4.1. One-Stage Detector

A one-stage detector-based UAV detection method predicts the location and classifies UAVs from videos or images using deep learning techniques. To determine the location of the UAV and categorize it, these algorithms typically only require processing the image once. Consequently, these algorithms exhibit high speed and can be employed in scenarios where real-time performance is crucial. Many studies have concentrated on the development of systems that can detect UAVs by utilizing a variety of object detection methods that are based on one-stage detectors. Table 6 summarizes the studies on vision-based methods of detecting UAVs utilizing learned features with one-stage detectors. We compared the research works based on the tasks, base models, datasets, results, and limitations.

Q. Shi and J. Li (2020) [42], in their study using the YOLOv4, addressed low-altitude UAV detection. Performance comparisons were made between the YOLOv4 detection results and the SSD and YOLOv3 models. According to this study, the YOLOv4 network had better mean Average Precision (mAP) and detection speed than the YOLOv3 and SSD networks. YOLOv4 effectively detected and classified three distinct UAV types (DJI Inspire, DJI Phantom, and XIRO Xplorer), achieving an impressive 89.32% mAP. Due to the unavailability of a public low-altitude UAV dataset, a customized database was built that includes 1,540 images captured from both eye-level and elevated angles in various flight conditions. The dataset was then added with 556 UAV images gathered from the internet.

In [43], Liu et al. (2021) optimized the UAV detection under high-security areas through modifying the YOLOv4 architecture to the challenges related to small target size and fast movement of the UAV. The authors simplified the model by cutting down on convolutional channels and shortcut layers, thus reducing the model architecture; by doing so the authors achieved higher computational efficiency without lowering its accuracy. The resulting pruned YOLOv4 model had a memory footprint of 15.1 MB, an mAP of 90.5% and a processing rate of 69 FPS. Training of the models used a self-built dataset of 10,000 UAV images and a different set of 2,000 images was used to test the models.

F. Samadzadegan and colleagues (2022) [44] used visible imagery to develop a deep convolutional neural network that can detect UAVs. The presented model was trained on publicly available images and videos of multirotor and helicopter UAVs as well as on the representatives of different bird species. Because of the morphological and behavioral resemblances between birds and UAVs, there were high rates of confusion; however, the model had a classification score of 83% and a mAP of 84%, which effectively differentiated birds and UAVs. Moreover, the model will be able to detect the presence of multirotor and helicopter UAVs instantly and decide if they are present in a particular area or not.

In reference [45], Y. Lv et al. (2022) presented a method that combines background difference and a lightweight network to increase UAV detection speed and accuracy in images with high resolution. To enhance feature extraction and accuracy the method incorporates advanced features including the Ghost module and the Simplified Attention Module (SimAM) attention mechanism. These developments aim to maximize efficiency and performance in deep learning models including those used for UAV detection. Their method balanced efficiency and precision for high-resolution UAV detection achieving some gains in detection speed and accuracy based on the publicly available Drone-vs-Bird dataset [46]. According to the experiment's results it performed appropriately with mAP of 97.50% precision of 97.30% and recall of 95.50%.

X. Zhai et al. (2023), in [47], optimized YOLOv8 for detecting tiny UAVs. In an attempt to improve the detection of small targets the proposed model incorporates a high-resolution detection head and eliminates redundant layers and the large-target detection head. This reduces network parameters and increases detection speed. A Space-to-Depth (SPD) layer is added in the extraction of features in order to maintain fine-grained details, followed by a non-strided convolutional layer (SPD-Conv). Furthermore the neck module contains a Global Attention Mechanism (GAM) to aid in the best feature fusion. The optimized model based on TIB-Net [48] can do better than the baseline by 11.9% in precision, 15.2% in recall, and 9% in mAP. It also uses 59.9% fewer parameters and 57.9% less model size. These quantitative improvements make the model a good choice for use in engineering and in real-world UAV surveillance systems.

In [49], C. Wang et al. (2023) proposed a lightweight YOLOX-based model that detects the presence of a swarm of UAVs. In order to concentrate on important information and improve the process of feature extraction the researchers incorporated the Convolutional Block Attention Module (CBAM) along with the Squeeze-and-Excitation (SE) module. MixUp and Mosaic Augmentation are employed to augment the data. Using UAVSwarm [50] dataset, experimental findings show that the designed model has a mAP of 82.32% with a model size of merely 3.85 MB surpassing the baseline model by around 2%.

J. Liu and colleagues (2024) [51], in their study, introduced the Global-Local YOLOMotion (GL-YOMO) algorithm in their study. This algorithm enhances the accuracy and stability of small UAV target detection by integrating the YOLO framework with multi-frame motion detection. The Ghost module multi-scale feature fusion and attention methods improve YOLO efficiency. Another motion detector technique is template-matching-based which is developed to facilitate the detection of small UAVs. The proposed system uses a collaborative global-local technique to optimize precision and efficiency. Experimental evaluations on both a self-built dataset and the Drone-vs-Bird

[46, 52] dataset indicate that GL-YOMO is highly helpful in detection tasks in the UAVs and improves accuracy and robustness in detection.

In [53], M. Huang et al. (2024) designed the EDGS-YOLOv8 a lightweight YOLOv8-based model for detecting UAVs in real time. The key improvements include structural changes that utilize ghost convolution in the neck to minimize model size the inclusion of Efficient Multi-scale Attention (EMA) and the enhancement of the detecting head with Deformable Convolutional Net v2 (DCNv2). When evaluated on the DUT Anti-UAV [54] and Det-Fly [55] datasets the suggested model achieved a 0.971 AP on the DUT Anti-UAV dataset, outperforming YOLOv8n by 3.1% mAP while remaining compact at 4.23 MB. The study findings and methodologies mentioned here are critical for developing lightweight UAV models and boosting target detection accuracy.

Y. Gao et al. (2024) [56] address the difficulty of long-range UAV detection by introducing a lightweight model based on YOLOv11. This model incorporates the Haar Wavelet Downsampling (HWD) module in the backbone to reduce feature loss during the process of downsampling while minimizing parameter usage. Additionally a lighter cross-feature fusion structure is then used in place of the original neck portion to enhance detection of small targets and improve adaptability to varying scales. To further improve detection accuracy for small UAVs the large-scale detection head is removed and replaced with a dedicated small-scale detection head. Experimental evaluation on the DUT Anti-UAV [54] dataset shows improvements over the YOLOv11 baseline, with gains of 4% in precision, 4.5% in recall and 4.1% in mAP and the parameters decreased by 38.4%. The proposed model is suitable for real-time UAV detection due to its high detection performance and efficiency.

In study [57], H. Hao et al. (2025) introduced ATA-YOLOv8 a high-precision model that overcomes existing algorithms' limited precision, recall and device processing power dependence which limit small UAV target detection. This model is evaluated versus YOLOv8n utilizing two datasets Det-Fly [55] and MOT-FLY [58]. The experimental results show that ATA-YOLOv8's mAP on the MOT-FLY and Det-FLY datasets is 94.9% and 96.4% respectively. This is a 25% and 5.9% increase over YOLOv8n's mAP and the model size stays the same at 5.1 MB. The proposed model can detect air-to-air UAVs in real time on edge computing devices because it is small size, high accuracy and speed.

Q. Zhang et al. (2025), in [59], presented BRA-YOLOv10, a UAV detection technique designed for difficult scenarios. This model uses Bi-Level Routing Attention (BRA) to cut down on background noise and bring out important target features during feature extraction which makes the detection more accurate overall. A Small Target Detection Layer (STDL) has been included to facilitate improved detection of small UAVs. The SimCSPSPPF module is used to retain low-level features and make the system more stable in cluttered scenes. Finally the model is trained and tested using TIB-Net [48], Det-Fly [55] and UAVfly [60]. The experiments demonstrate that BRA-YOLOv10 outperforms current models with 98.9% precision, 92.3% recall and 96.5% mAP50. This represents improvements of 2.5%, 2.9% and 1.7% over YOLOv10. These results indicate that the model works well for UAV applications in challenging situations and is effective at detecting small UAVs.

In [61], a recent study by S. Zhou and colleagues (2025) developed a lightweight model for detecting UAVs called LAMS-YOLO utilizing linear attention mechanisms as well as adaptive downsampling process. This model employs a CPU-optimized structure incorporating depthwise separable convolutions and efficient activation functions to substantially reduce the number of parameters. The detection accuracy in challenging environments is improved by a new linear attention module that uses an LSTM-inspired gating structure to improve semantic feature extraction. Also there is an improved loss function for bounding boxes that helps make localization more accurate. The model is trained and validate on a self-built dataset of 5288 training images (1438 birds and 3850 UAVs) and 1700 validation images (348 birds and 1352 UAVs). Experiments show that LAMS-YOLO is better than YOLOv11n since it raises mAP by 3.89% and lowers the number of parameters by 9.35%. These enhancements provide significant assistance for real-time UAV surveillance.

Table 6- Comparison summary of One-Stage detector-based UAV detection methods.

Ref	Year	Tasks	Base Model(s)	Dataset(s) Used	Main Results				Limitations
[42]	2020	UAV detection at low altitude	YOLOv3, YOLOv4 and SSD	Self-built		YOLOv3	YOLOv4	SSD	Limited ability to detect various types of UAVs.
					Accuracy:	84.14%	89.32%	79.52%	
					Recall:	89.27%	92.48%	85.31%	

					mAP:	84.14%	89.32%	76.84%	
[43]	2021	Small UAV detection	YOLOv4	Self-built	Precision:	74.2%			The model was too big to utilize on an edge device because it was 15.1 MB in size. The issues posed by cluttered backgrounds, as well as the similarities between UAVs and other objects like birds, were not addressed.
					Recall:	93.1%			
					mAP:	93.6%			
					F1-score:	82.6%			
[44]	2022	UAVs and birds detection	YOLOv4	Self-built	Accuracy:	83%			Slow inference method may detect and identify only two UAV categories, such as multirotor and helicopter.
					Recall:	84%			
					mAP:	84%			
					F1-score:	83%			
[45]	2022	Small UAV detection	YOLOv5	Drone-vs-Bird [46]	Precision:	97.30 %			Limited detection accuracy in more complex backgrounds.
					Recall:	95.50 %			
					mAP:	97.60 %			
[47]	2023	Tiny UAV detection	YOLOv8	TIB-Net [48]	Precision:	93.30%			Inability to detect UAVs at long distances and poor generalizability.
					Recall:	93.30%			
					mAP:	95.10%			
[49]	2023	UAV swarm detection	YOLOX	UAVSwarm [50]	mAP:	82.32%			Slow inference speed and complexity of the model
[51]	2024	Small UAV detection	YOLO	Drone-vs-Bird [46,52]	Precision:		88.60%		Limited detection accuracy in more complex environments.
					Recall:		80.90%		
					mAP:		76.30%		
					F1-score:		84.60%		
[53]	2024	Small UAV detection	YOLOv8	DUT Anti-UAV[54] and Det-Fly [55]	Using DUT Anti-UAV				Limited detection accuracy for small UAV detection at long range.
					mAP:	97.10%			
					Using Det-Fly				
					mAP:	93.40%			
[56]	2024	Small UAV detection	YOLOv11	DUT Anti-UAV[54]	Precision:	96.90%			Complexity of the model and limited generalization capability across different environments.
					Recall:	89.80%			
					mAP:	95.20%			
[57]	2025	Air-to-Air Small UAV detection	YOLOv8n	Det-Fly [55] and MOT-FLY[58]	Using MOT-FLY				Limited applicability for detecting various UAV categories.
					Precision:	96.40%			
					Recall:	87.60%			
					mAP:	94.90%			
					Using Det-Fly				
					Precision:	98%			
					Recall:	92.40%			
					mAP:	96.40%			
[59]	2025	Small UAV detection	YOLOv10	TIB-Net [48], Det-Fly [55], and UAVfly[60]	Using TIB-Net				Complexity of the model and limited applicability for detecting various UAV categories.
					Precision:	92.60%			
					Recall:	93.10%			
					mAP:	96.50%			
					Using Det-Fly				
					Precision:	97.10%			

					Recall:	96.50%	
					mAP:	98.90%	
					Using UAVfly		
					Precision:	99.60%	
					Recall:	99.30%	
					mAP:	99.50%	
[61]	2025	UAV detection	YOLOv11	Self-built	Precision:	96.10%	Limited capacity to generalize in different environments and poor detection accuracy for small UAV detection at long range.
					Recall:	89.30%	
					mAP:	93.40%	
					F1-score:	92.58%	

4.2. Two-Stage Detector

The two-stage detector-based UAV detection methods propose regions of interest (ROI) and then perform classification to identify categories. These methods are more accurate than the one-stage detector-based methods because they locate and classify the regions of interest in the first stage. However, because of additional regions and stage processing, the two-stage detector-based method requires more time to infer than the one-stage method. Table 7 provides a comparative summary of studies on vision-based UAV detection methods with learned features using two-stage detectors. It presents each study's tasks, base model, datasets utilized, limitations, and results achieved.

A. Coluccia and colleagues (2021) [46] used the YOLOv3, YOLOv5, Faster R-CNN and Cascade R-CNN to detect and differentiate between UAVs and birds. Three teams—Gradiant, Eagledrone and Alexis—employed different deep learning strategies. Using the Drone-vs-Bird [46] dataset Team Gradiant used Cascade R-CNN to get the highest overall performance of 80%. This study found that the most difficult video sequences are acquired by very distant UAVs and moving cameras. Also the models' performance changes according to the UAV's visibility size and shape.

In [55], Y. Zheng et al. (2021) offered a novel dataset called Det-Fly [55], comprising over thirteen thousand images of a moving target UAV. The researchers employed the Det-Fly [55] and MIDGARD [62] datasets to evaluate eight deep learning algorithms: SSD, YOLOv3, RetinaNet, FPN, RefineDet, Faster R-CNN, Cascade R-CNN and Grid R-CNN. They evaluated the algorithms under various challenging conditions, such as motion blur, partial occlusion and varying lighting, to assess the key features of the images in terms of detection effectiveness. Alongside emphasizing the primary challenges related to air-to-air UAV detection the study's findings also offer opportunities for further algorithmic enhancements.

B. K. Isaac-Medina et al. (2021) [63], in their study, employed a dataset of visual and thermal images utilizing four deep learning architectures: Faster R-CNN, SSD, YOLOv3 and DETR, for object detection. This study utilized the Drone-vs-Bird [46], Anti-UAV [64] and MAV-VID [65] datasets for training and testing purposes. The findings indicated that all of the networks under study were capable of accurately detecting small UAVs. Furthermore, YOLOv3 showed the highest overall accuracy, while Faster R-CNN continuously showed the highest mAP for detecting small UAVs.

In reference [50], C. Wang and colleagues (2022) introduced an innovative dataset called UAVSwarm [50] for UAV detection and Multiple Object Tracking (MOT). The authors used the Faster R-CNN and YOLOX algorithms to check the dataset for detection and the GNMOT and ByteTrack algorithms to check it for UAV swarm MOT tasks. The experimental results showed that YOLOX did better than Faster R-CNN on the UAVSwarm dataset with 83.68% mAP versus 50.75%.

J. Zhao et al. (2022) [54] presented a novel dataset DUT Anti-UAV [54], comprising two subsets for detection and tracking. The researchers used both one-stage detectors, like YOLOX and SSD, as well as two-stage detectors, like Faster R-CNN, Cascade R-CNN, and ATSS. They utilized their dataset to evaluate various UAV detection and tracking techniques across diverse scenarios. Experimental results indicate that two-stage detector-based models typically demonstrate greater accuracy, whereas one-stage models excel in speed. This study presented a basic and straightforward fusion approach based on detectors and trackers. Furthermore, it provided a comparison of the tracking outcomes of the combinations of 8 trackers and 14 detectors. According to the results of a number of experiments their fusion approach has the potential to significantly improve tracking performance across all trackers.

A. S. Mubarak et al. (2022) [66], in their work, used a two-stage detector-based method that included a Mask R-CNN with two backbones, namely ResNet-50 and MobileNet. The two networks' performances were compared after the model was trained using the Kaggle [67] dataset that contained over 1000 UAV images. Experimental results showed that Mask R-CNN with ResNet-50 performed better than Mask R-CNN with the MobileNet backbone, and the results obtained were a mAP of 96.50% and a recall of 96.10%. While this method gives excellent detection accuracy its detection speed is relatively slow. In contrast, one-stage detectors provide superior speed performance.

Table 7- Comparison summary of Two-Stage detector-based UAV detection methods.

Ref	Year	Tasks	Base Model(s)	Dataset(s) Used	Main Results	Limitations
[46]	2021	UAV and Bird detection	Cascade R-CNN	Drone-vs-Bird [46]	Detection results of Cascade R-CNN were 80.0%	The model has a heavier backbone, leading to the increased model size and slower speed.
[55]	2021	Air-to-Air Small UAV detection	Cascade R-CNN, Faster R-CNN, and Grid R-CNN	Det-Fly [55] and MIDGARD [62]	Using Det-Fly Cascade R-CNN / Faster R-CNN / Grid R-CNN AP: 79.40% / 70.50% / 82.40% Using MIDGARD AP: 89.40% / 89.10% / 90.10%	High architecture complexity and limited applicability for detecting various UAV types.
[63]	2021	Small UAV detection	Faster R-CNN	Drone-vs-Bird [46], Anti-UAV-RGB [64] and MAV-VID [65]	Using Anti-UAV RGB AP: 98.20% Using MAV-VID AP: 97.80% Using Drone-vs-Bird AP: 63.20%	Limited capability for long-distance detection when UAVs seem tiny.
[50]	2022	UAV swarm detection	Faster R-CNN	UAVSwarm [50]	mAP: 50.75%	Slow inference speed and high computation.
[54]	2022	UAV detection and tracking	Faster R-CNN and Cascade R-CNN	DUT Anti-UAV [54]	Faster R-CNN / Cascade R-CNN mAP: 65.30% / 68.30%	High architecture complexity and slow detction speed.
[66]	2022	UAV detection	Mask R-CNN	Kaggle [67]	Recall: 96.10% mAP: 96.50%	High architecture complexity, slow detction speed, and limited generalizability.

5. Datasets and Evaluation Metrics

5.1. Datasets

In computer vision the dataset is a vital element in the development of a robust model. This is because the training of models and the maintenance of resilience rely on the availability of datasets. Furthermore, the availability of high-quality datasets is an important factor in the advancement of research on UAV detection. These datasets must consider the variety of weather and lighting conditions, environments, UAV types and operational situations that the UAV detection systems may encounter. Table 8 provides a summary of the relevant information from a variety of publicly available datasets that are employed for UAV detection. The subsequent subsection reviews these datasets. Figure 6 displays some example images from these databases.

Table 8 - Summarized information of popular UAV datasets.

Dataset Name	Content; Resolution; Size	Enviroment	Conditions	UAV Type	Used by	Year	Links+
MIDGARD [62]	8,776 images; 752 × 480; 3.3 GB	Fields, Hills, Forest, Urban, Indoor	Daytime, Dawn, Nighttime	1	[55]	2020	https://mrs.fel.cvut.cz/midgard
TIB-Net [48]	2,860 images; 1920 × 1080; 616 MB	University	Daytime, Dawn, Nighttime.	3	[47,59]	2020	https://github.com/kyn0v/TIB-Net
MAV-VID [65]	40,232 images, 64 videos; 1920 × 1080; 10.8 GB	NA	NA	1	[63]	2020	https://bitbucket.org/alejodosr/mav-vid-dataset/downloads/
Drone-vs-Bird [46,52]	104,760 images, 77 videos; 720 × 576 to 3840 × 2160; 7.1 GB	Sky, Urban, Semi-urban, Forest, Sea	Daytime, Dawn, Nighttime; Weather	8	[45,46,51,63]	2021	https://github.com/wosdetc/challenge
Det-Fly [55]	13,271 images; 3840 × 2160; 3.1 GB	Sky, Urban, Mountain, Field	Daytime, Dawn, Nighttime	1	[53,55,57,59]	2021	https://github.com/Jake-WU/Det-Fly
Anti-UAV (RGBT) [64]	186,494 images, 318 videos; 1920 × 1080 (RGB), 640 × 512 (IR); 6.25G	Sky, Urban, Mountain	Daytime, Nighttime; Weather	6	[63]	2021	https://github.com/ZhaoJ9014/Anti-UAV
DUT Anti-UAV [54]	10,000 images, 20 videos; 160 × 240 to 3744 × 5616; 4.81 GB	Sky, Jungles, Urban, Filed	Daytime, Dawn, Nighttime; Weather	35	[53,54,56]	2022	https://github.com/wangdongdut/DUT-Anti-UAV
UAVSwarm [50]	12,598 images, 72 videos; 446 × 270 to 1919 × 1079; 1.82 GB	Sky, Filed	Daytime, Dawn, Nighttime	19	[49,50]	2022	https://github.com/UAVSwarm/UAVSwarm-dataset
MOT-FLY [58]	11,186 images; 1080 × 1920; 2.64 GB	Sky,Village, Flat, Urban	Daytime, Dawn, Nighttime	3	[57]	2023	https://github.com/CZC-123/MOT-FLY
UAVfly [60]	10,281 images; 1280 × 720; 1.88 GB	Sky, Urban, Suburban, Field, Desert, Lake, Mountain	Daytime, Dawn, Nighttime	3	[59]	2024	https://github.com/lucien22588/UAVfly

+ The links to the datasets were accessed on 20 October 2025

MIDGARD [62] includes 8,776 images, capturing both indoor and outdoor environments. The outdoor scenes were collected from diverse locations, such as forests, meadows, fields, and urban areas, while the indoor data features complex backgrounds and lighting conditions. The dataset focuses mostly on UAVs and does not include any samples of confusing aerial objects such as birds or balloons, thus reducing the ability of trained models to handle false positives in cluttered scenarios. Also, the images are frequently captured at medium to close ranges; hence, there are fewer long-range, small UAV examples in it.

TIB-Net [48] contains 2,860 images of various UAV types, including fixed-wing and multi-rotor UAVs. The images are captured by a fixed camera on the ground, which is approximately 500 meters from the UAV in the air. The captured images have a resolution of 1920 × 1080 pixels. The scenes cover a variety of daytime and nighttime lighting conditions. 75% of the collected data is used for training, and the rest is used for testing. TIB-Net is a viable choice for detecting small UAVs; however, it doesn't have as many different types of UAVs and various environments as larger databases do.

MAV-VID [65] includes 64 videos taken in different settings on a single UAV. The videos were taken by handheld mobile devices, ground-based cameras, and other UAVs. It consists of 53 videos (29,500 images) for training and 11 videos (10,732 images) for validation with an average object size of 136 x 77 pixels. In evaluating the accuracy of detection and tracking systems in crowded environments it is essential to incorporate UAV occlusion scenarios into the dataset. Unfortunately this dataset does not presently satisfy that criterion.

Drone-vs-Bird [46, 52] focuses on long-range UAV detection among confusing objects like birds making it one of the most challenging datasets available. The dataset consists of 77 videos of UAVs (104,760 images). The UAV images in this dataset are often surrounded by small objects like birds and are captured at long distances. The average UAV image size is 34 x 23 pixels. It requires models that are able to differentiate between UAVs and birds which is of critical importance in scenarios where vision-based techniques frequently produce false positives. Finding UAVs in this dataset is extremely challenging which shows how limited current detection methods are especially when it comes to scale variance.

Det-Fly [55] is an air-to-air dataset including 13,271 high-resolution images (3840 x 2160 pixels) of a micro UAV (DJI Mavic) captured by a camera-equipped flying UAV (DJI M210). This dataset encompasses various scenarios including diverse lighting conditions, relative distances and viewing angles. But it doesn't take into account different weather conditions and it only covers one type of UAV.

Anti-UAV (RGBT) [64] consists of 318 RGB-T video pairs with one thermal video and one RGB video in each pair. A total of 160 videos are utilized to make up the training set, 91 videos are utilized to make up the testing set and the remaining videos are utilized to make up the validation set. In total, there are 186,494 images. This dataset focuses on the detection and tracking of UAVs across six different models. The data was collected in two distinct light modes (infrared and visible), under two lighting conditions (day and night), and against a variety of backgrounds. Infrared data inclusion in the Anti-UAV gives it a great advantage in developing powerful models that will enable it to function in the low-visibility environments, which include foggy weather or night. However there is an issue: the RGB and infrared cameras are out of sync in time and space and there's only one shot area.

DUT Anti-UAV [54] was designed to aid in both detection and tracking missions. This dataset has 10,000 images for the detection missions, with 5,200 images in the training set, 2,200 images in the testing set, and 2,600 images in the validation set. It also has 20 videos for applying single- and multi-UAV tracking missions in challenging scenarios. Due to this dataset's focus on small UAVs against complex backgrounds, it can be challenging to evaluate for UAV detection and tracking. However, a limited set of tracking scenarios, a lack of variety in UAV behaviors, and long-term tracking are some of DUT Anti-UAV's limitations.

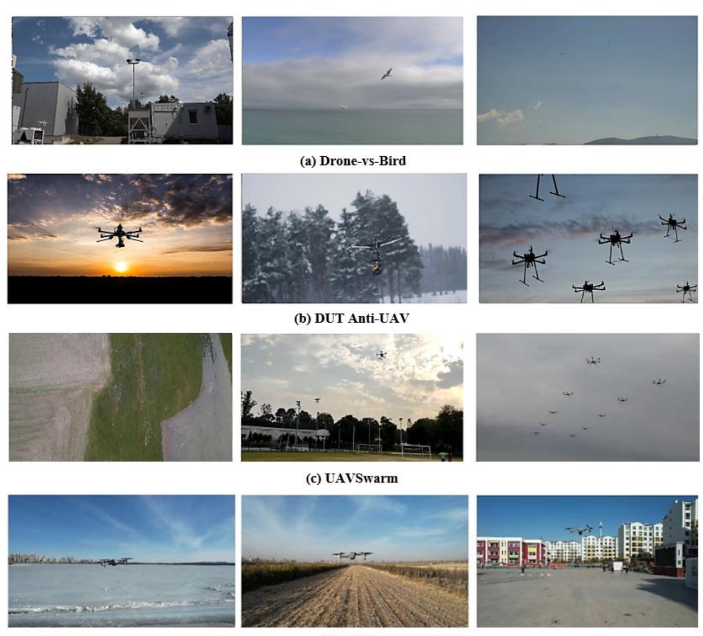

Fig. 6 - Some such images are in (a) Drone-vs-Bird, (b) DUT Anti-UAV, (c) UAVSwarm, and (d) UAVfly.

UAVSwarm [50] is designed for detecting and tracking swarm UAVs. There are 12,598 images in total each with 3 to 23 UAVs. They are divided into 72 sequences, with each sequence having between 58 and 705 images. The dataset is divided into 36 training sequences with 6,844 images and 36 testing sequences with 5,754 images. It contains images from various ground and sky background views that were taken with both static and moving cameras. Because there are so many different types of UAVs and interactions are so complicated it is the best option to evaluate how robust tracking models are. Although the dataset includes a wide range of UAV scenarios it falls short in terms of diverse environmental conditions.
MOT-FLY [58] includes 11,000 images of three different types of UAVs. For robust algorithm training and testing it provides a variety of backgrounds, lighting conditions and object sizes. It addresses issues that frequently arise in UAV tracking scenarios like motion blur, partial occlusion and various angles of view.
UAVfly [60] is a dataset of 10,281 images that were made for the purpose of air-to-air UAV detection. The images have a resolution of 1280 × 720 pixels. It encompasses a wide range of geographical environments. Three DJI AIR2 UAVs were used to gather data at three different times of the day. The data was gathered in challenging scenarios such as changing light conditions, blurring and partial occlusion.

5.2 Evaluation Metrics

This subsection presents the metrics that are commonly used to evaluate the detection models' performance:

- **Confusion Matrix:** A matrix of size n x n (n corresponds to the number of classes) that indicates the model's accuracy. For example, the UAVs are regarded as belonging to the positive class, whilst the birds are regarded as belonging to the negative class. Hence, the information regarding the true class of the intended objects is shown by the columns in this matrix. In contrast, the rows in this matrix include information regarding the predicted class that was determined by the model. Figure 7 provides an example of a 2x2 confusion matrix, which is also known as a binary confusion matrix. In this matrix,

  True Positive (TP): A correct detection of a target object (e.g., the model correctly detected the UAV as a UAV).
  False Positive (FP): An incorrect detection of a non-target object or an inaccurate detection of an actual object (e.g., the model detected and misclassified the bird as a UAV).
  False Negative (FN): An incorrect detection of a target object (e.g., a UAV is detected, but the model incorrectly classified it as a bird).
  True Negative (TN): A correct detection of a non-target object (e.g., the model correctly detected the bird as a bird).

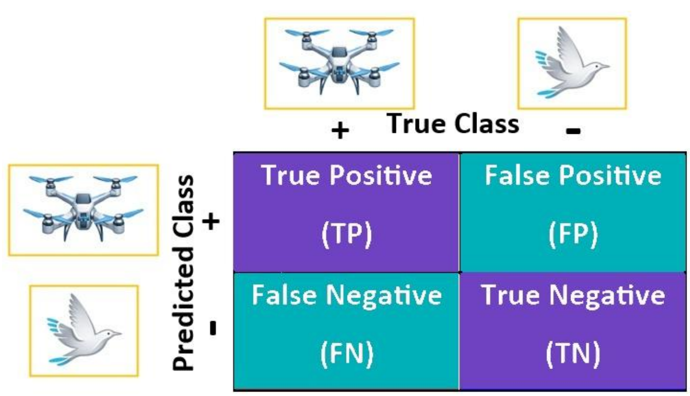

Fig. 7- Confusion Matrix.

- **Intersection over Union (IoU):** According to the aforementioned definitions, it is necessary to clarify the terms "correct detection" and "incorrect detection". The IoU is a commonly used metric of doing this. This metric takes its cue from the Jaccard Index [68], a similarity coefficient between two sets of data. As illustrated in Figure 8, the IoU calculates the intersection and union ratio of the predicted bounding box Bp and the ground-truth bounding box Bgt in the object detection field, represented by eq. 1.

$$J(B\text{p}\,,B\text{gt}\,) = \text{IoU} = \frac{area\ (B\text{p}\cap B\text{gt})}{area\ (B\text{p}\cup B\text{gt})} \tag{1}$$

$$\text{IoU} = \frac{\textit{area of intersection}}{\textit{area of union}} =$$

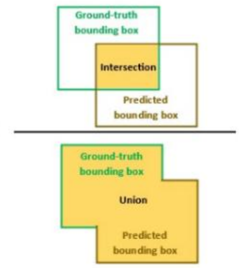

Fig. 8 - Intersection over Union.

The IoU value is a numerical value between 0 and 1, with 0 indicating no overlap between the boxes and 1 representing complete overlap or identical areas. When two bounding boxes have higher IoU values, it means they are more similar or in agreement. The detection can be classified as correct or incorrect by comparing the IoU value with a given threshold t. Thus, if IoU ≥ t, the detection is considered as correct; otherwise, i.e., if IoU < t, the detection is considered as incorrect.

- **Accuracy:** One of the standard measures used to evaluate performance. It is defined as the ratio of all predictions that are predicted correctly by the model, as shown in eq. 2.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{2}$$

A higher accuracy score means that the model is making more accurate predictions which makes it more reliable in detecting objects like UAVs. Detection accuracy is often measured with more specific metrics such as precision, recall, F-measure and others.

- **Precision:** Refers to the ability of a model to identify only relevant objects. As represented by eq. 3, it is defined as the ratio of correct positive predictions.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{3}$$

A high precision means that the model is good at avoiding false positives, which is when it wrongly identifies an object as a UAV.

- **Recall:** Represents the model's ability to find all relevant objects (all ground-truth bounding boxes). It is defined by eq. 4 as the proportion of correct positive predictions among all ground truths that currently exist.

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{4}$$

A high recall indicates that the model effectively reduces false negatives successfully detecting UAVs that are actually present.

- **F-Measure (F1-score):** A balanced metric that considers both recall and precision. The harmonic mean of precision and recall is calculated to get a single value that shows how well the model works overall. Equation (5) shows it.

$$\text{F1\_score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{5}$$

A higher F1-score indicates a better balance between precision and recall. Therefore, in the context of UAV detection, depending on the specific application, a balance between precision and recall may be necessary. For example, in a critical infrastructure security scenario, a higher recall might be prioritized, even if it means accepting more false positives, while in a commercial UAV delivery operation, a higher precision might be more important to minimize unnecessary delays and errors.

- **Average Precision (AP):** is determined by calculating the area right below the precision-recall curve. This is a frequently utilized metric for evaluating the accuracy of detection models [68]. The following equation is used to find AP:

$$\text{AP} = \sum_{n} (R_n - R_{n-1})\, P_n \tag{6}$$

The value of AP is high when both the precision and recall are high, and its value is low when one of the precision or recall is low.

- **Mean Average Precision (mAP):** The accuracy of object detectors across all classes in a specific database is measured using this metric, which is simply the average of the AP scores over all classes, as shown in eq. 7.

$$\text{mAP} = \frac{1}{N} \sum_{i=1}^{N} AP_i , \tag{7}$$

where N is the total number of classes being evaluated, and APi is the AP in the ith class [68].

It is worth noting here that the mAP score is a value that increases as the model's detection performance improves, meaning a higher mAP score indicates more accurate results.

- **Frames Per Second (FPS):** refers to the number of frames of video (or images) that the detection model can process per second. It measures the model's speed and efficiency, indicating how quickly it can process incoming images and generate detection results. A higher FPS is preferred for real-time applications in which the model needs to keep up with the video feed.
- **Parameters (Params):** Indicate the number of learnable parameters in a model. Adding more parameters to a model makes it more knowledgeable and accurate but it also makes the model more complicated, which can slow it down and make it more demanding in terms of storage and computation.
- **Floating-point Operations (FLOPs):** This metric is used to quantify the total number of floating-point operations per second a model executes in inference time. The knowledge of FLOPs is crucial to evaluate the efficiency of models, to compare models of different architectures, and to evaluate whether a model is efficient on a particular hardware platform, whether it is a powerful cloud server or a resource-constrained edge device. In general, a model with higher FLOPs means that it is more accurate and complex but slower. While a lower FLOPs means that the model will be simpler and faster, it might be less accurate.

6. Conclusion And Future Directions

6.1. Conclusion

There are advantages and disadvantages to UAV technology advancement at breakneck speed particularly when it comes to managing low-altitude airspace. In recent years the illegal operation of UAVs has posed a severe threat to personal privacy and public security. Therefore, the employing of vision-based UAV detection systems offers a promising solution to detect the rogue UAVs as well as to safely integrate UAVs into the airspace. However, there exist great challenges that still remain to be solved, such as the detectability of fast and small UAVs in a reliable way, multi-UAV scenarios, real-time performance, diversity of detection ranges and altitude cases, and detection with various environmental conditions. To tackle these challenges, ongoing research and development of UAV detection systems are necessary for airspace security to counteract the threats. To address these rising concerns, researchers and engineers working on UAV detection systems will be able to benefit from this article paper, as it aggregates the facts and insights obtained from different research efforts.
This study gives an in-depth, detailed review of vision-based UAV detection methods utilizing deep learning techniques, focused on CNN-based models. In general, these methods can be broadly classified into one-stage detectors and two-stage detectors. The earlier two-stage detectors are accurate but not fast, and the later one-stage detectors are fast with relatively high accuracy. Furthermore, this review paper provides an analysis and summary of the pros and cons of both proposed and existing methods for addressing issues related to UAV detection

technology, thereby offering researchers a comprehensive understanding of various methodologies. Lastly, it provides a structured review of popular datasets and evaluation metrics, which will help with understanding and analysis of their features and suitability.

6.2. Future Directions

Many directions can be proposed in the future work. The following is a set of these directions:

- To guarantee real-time processing capabilities, optimize algorithms and hardware configurations potentially utilizing edge computing or graphics processing unit acceleration. As a result, visual data may be processed in real time enabling faster and more accurate UAV detection and tracking.
- Investigate transformer-based models, particularly DETR-based models, which are increasingly used for small-object detection in cluttered scenes due to their ability to leverage self-attention for global context modeling, overcoming the limitations of local receptive fields in CNNs.
- Develop systems to detect and track fleets of UAVs (UAV swarms) in real-time through distributing computation across edge devices and onboard units to achieve low latency and high throughput.
- Design and curate large and comprehensive datasets of various categories of UAVs and challenging scenarios. In order to address the challenge of the lack or the scarcity of diverse and well-annotated UAV datasets that is hampering the training process of DL-based models. In addition, propose a standardized benchmarking framework with diverse datasets and evaluation models for UAV detection and tracking promoting fair comparisons and reproducibility.
- Develop hybrid sensor-based techniques for UAV detecting and tracking by exploring multimodal sensor fusion strategies that combine visual data together with other sensor modalities (e.g. RF or radar) for more robust and reliable UAV detection and tracking systems.

References

[1] A. Bertolini, “Artificial Intelligence and civil law: liability rules for drones,” *European Parliament Think Tank*, Study IPOL_STU(2018)608848, May 2018. [Online]. Available: https://www.europarl.europa.eu/thinktank/en/document/IPOL_STU(2018)608848. [Accessed: Sep. 28, 2025].

[2] O. D. H. Hani and A. A. Abdulhameed, "Drone network and the cyber attacks: A survey study," in *AIP Conference Proceedings*, vol. 3264, no. 1, p. 020012, Mar. 2025, AIP Publishing LLC. doi: https://doi.org/10.1063/5.0258870.

[3] U.S. Government Accountability Office, “Drones: Actions Needed to Better Support Remote Identification in the National Airspace,” GAO-24-106158, Jun. 6, 2024. [Online]. Available: https://www.gao.gov/products/gao-24-106158. [Accessed: Sep. 29, 2025].

[4] Z.-Q. Zhao, P. Zheng, S.-T. Xu, and X. Wu, “Object Detection With Deep Learning: A Review,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 30, no. 11, pp. 3212–3232, Nov. 2019, doi: https://doi.org/10.1109/tnnls.2018.2876865.

[5] A. Sharba and H. Kanaan, “Improving Tiny Object Detection in Aerial Images with Yolov5,” *Journal of Engineering and Sustainable Development*, vol. 29, no. 1, pp. 57–67, Jan. 2025, doi: https://doi.org/10.31272/jeasd.2682 .

[6] J. Cheng, H. Li, D. Li, S. Hua, and V. S. Sheng, “A Survey on Image Semantic Segmentation Using Deep Learning Techniques,” *Computers, Materials & Continua*, vol. 74, no. 1, pp. 1941–1957, 2023, doi: https://doi.org/10.32604/cmc.2023.032757.

[7] L. Khrissi, N. El Akkad, H. Satori, and K. Satori, “Clustering method and sine cosine algorithm for image segmentation,” *Evolutionary Intelligence*, Jan. 2021, doi: https://doi.org/10.1007/s12065-020-00544-z.

[8] A. A. Maryoosh and Saeid Pashazadeh, “Leukemia Detection Using Machine Learning Algorithms: Current Trends and Future Directions (Literature Survey),” *Mustansiriyah Journal of Pure and Applied Sciences*, vol. 2, no. 3, pp. 33–49, Jul. 2024, doi: https://doi.org/10.47831/mjpas.v2i3.76 .

[9] Alaa Hussein Abdulaal *et al.*, “Hybrid CNN and RNN Model for Histopathological Sub-Image Classification in Breast Cancer Analysis Using Self-Learning,” *Journal of Engineering and Sustainable Development*, vol. 29, no. 3, pp. 310–320, May 2025, doi: https://doi.org/10.31272/jeasd.2746 .

[10] S. Srivastava, S. Narayan, and S. Mittal, “A survey of deep learning techniques for vehicle detection from UAV images,” *Journal of Systems Architecture*, vol. 117, p. 102152, Aug. 2021, doi: https://doi.org/10.1016/j.sysarc.2021.102152.

[11] Y. Diez, S. Kentsch, M. Fukuda, M. L. L. Caceres, K. Moritake, and M. Cabezas, “Deep Learning in Forestry Using UAV-Acquired RGB Data: A Practical Review,” *Remote Sensing*, vol. 13, no. 14, p. 2837, Jul. 2021, doi: https://doi.org/10.3390/rs13142837.

[12] T. B. Shahi, C.-Y. Xu, A. Neupane, and W. Guo, “Recent Advances in Crop Disease Detection Using UAV and Deep Learning Techniques,” *Remote Sensing*, vol. 15, no. 9, p. 2450, Jan. 2023, doi: https://doi.org/10.3390/rs15092450.

[13] P. Tang, J. Li, and H. Sun, “A Review of Electric UAV Visual Detection and Navigation Technologies for Emergency Rescue Missions,” *Sustainability*, vol. 16, no. 5, pp. 2105–2105, Mar. 2024, doi: https://doi.org/10.3390/su16052105.

[14] P. Mittal, R. Singh, and A. Sharma, “Deep learning-based object detection in low-altitude UAV datasets: A survey,” *Image and Vision Computing*, vol. 104, p. 104046, Dec. 2020, doi: https://doi.org/10.1016/j.imavis.2020.104046.

[15] S. Park, H. T. Kim, S. Lee, H. Joo, and H. Kim, "Survey on Anti-Drone Systems: Components, Designs, and Challenges," *IEEE Access*, vol. 9, pp. 42635–42659, 2021, doi: https://doi.org/10.1109/access.2021.3065926.

[16] M. A. Khan, H. Menouar, A. Eldeeb, A. Abu-Dayya, and F. D. Salim, "On the Detection of Unauthorized Drones—Techniques and Future Perspectives: A Review," *IEEE Sensors Journal*, vol. 22, no. 12, pp. 11439–11455, Jun. 2022, doi: https://doi.org/10.1109/JSEN.2022.3171293.

[17] X. Yan *et al.*, "UAV Detection and Tracking in Urban Environments Using Passive Sensors: A Survey," *Applied Sciences*, vol. 13, no. 20, p. 11320, Jan. 2023, doi: https://doi.org/10.3390/app132011320.

[18] Nader Al-lQubaydhi *et al.*, "Deep learning for unmanned aerial vehicles detection: A review," *Computer science review*, vol. 51, pp. 100614–100614, Feb. 2024, doi: https://doi.org/10.1016/j.cosrev.2023.100614.

[19] M. H. Rahman, M. A. S. Sejan, M. A. Aziz, R. Tabassum, J. I. Baik, and H. K. Song, "A Comprehensive Survey of Unmanned Aerial Vehicles Detection and Classification Using Machine Learning Approach: Challenges, Solutions, and Future Directions," *Remote sensing (Basel)*, vol. 16, no. 5, pp. 879–879, Mar. 2024, doi: https://doi.org/10.3390/rs16050879.

[20] A. A. Laghari, A. K. Jumani, R. A. Laghari, H. Li, S. Karim, and A. A. Khan, "Unmanned aerial vehicles advances in object detection and communication security review," *Cognitive Robotics*, vol. 4, pp. 128–141, Aug. 2024, doi: https://doi.org/10.1016/j.cogr.2024.07.002.

[21] Safaa Dafrallah and M. A. Akhloufi, "Malicious UAV detection using various modalities," *Drone systems and applications*, Mar. 2024, doi: https://doi.org/10.1139/dsa-2023-0049 .

[22] J. Moore, "Counter drone tactics: Which drones are a real threat, and which aren't?," *IFSEC Global*, Feb. 2021. [Online]. Available: https://www.ifsecglobal.com/drones/counter-drone-tactics-which-drones-are-a-real-threat-and-which-arent/. [Accessed: Oct. 5, 2025].

[23] N. Ali, "Unexploded drone found near Duhok dam," *Rudaw*, Apr. 10, 2023. [Online]. Available: https://www.rudaw.net/english/kurdistan/100420231. [Accessed: Oct. 5, 2025].

[24] A. Greenstein, "Five arrested for using drones to smuggle contraband into federal prison," *Louisiana Radio Network*, Jan. 27, 2025. [Online]. Available: https://louisianaradionetwork.com/2025/01/27/five-arrested-for-using-drones-to-smuggle-contraband-into-federal-prison/. [Accessed: Oct. 6, 2025].

[25] Shafaq News, "Drone crashes near security facility in Baghdad," *Shafaq News*, Mar. 12, 2025. [Online]. Available: https://shafaq.com/en/Iraq/Drone-crashes-near-security-facility-in-Baghdad [Accessed: Oct. 6, 2025].

[26] D. Toogood, "Crashed drone seized after failed prison smuggling attempt," *Island Echo*, Jan. 15, 2025. [Online]. Available: https://www.islandecho.co.uk/drone-crashed-and-seized-after-failed-prison-smuggling-attempt/. [Accessed: Oct. 6, 2025].

[27] F.-L. Chiper, A. Martian, C. Vladeanu, I. Marghescu, R. Craciunescu, and O. Fratu, "Drone detection and defense systems: Survey and a software-defined radio-based solution," *Sensors*, vol. 22, no. 4, p. 1453, Feb. 2022, doi: https://doi.org/10.3390/s22041453.

[28] S. A. H. Mohsan, N. Q. H. Othman, Y. Li, M. H. Alsharif, and M. A. Khan, "Unmanned aerial vehicles (UAVs): Practical aspects, applications, open challenges, security issues, and future trends," *Intelligent Service Robotics*, vol. 16, no. 1, pp. 1–21, Jan. 2023, doi: https://doi.org/10.1007/s11370-022-00452-4.

[29] U. Seidaliyeva, D. Akhmetov, L. Ilipbayeva, and E. T. Matson, "Real-time and accurate drone detection in a video with a static background," *Sensors*, vol. 20, no. 14, p. 3856, Jul. 2020, doi: https://doi.org/10.3390/s20143856.

[30] J. Dai, L. Wu, and P. Wang, "Overview of UAV Target Detection Algorithms Based on Deep Learning," *2021 IEEE 2nd International Conference on Information Technology, Big Data and Artificial Intelligence (ICIBA)*, pp. 736–745, Dec. 2021, doi: https://doi.org/10.1109/iciba52610.2021.9688305 .

[31] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, Jun. 2017, https://doi.org/10.1145/3065386 .

[32] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2014, pp. 580–587. doi: https://doi.org/10.1109/CVPR.2014.81.

[33] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017. doi: https://doi.org/10.1109/TPAMI.2016.2577031.

[34] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2980–2988. doi: https://doi.org/10.1109/ICCV.2017.322.

[35] X. Lu, B. Li, Y. Yue, Q. Li, and J. Yan, "Grid R-CNN," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 7363–7372. doi: https://doi.org/10.1109/CVPR.2019.00754.

[36] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 779–788. doi: https://doi.org/10.1109/CVPR.2016.91.

[37] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "SSD: Single Shot MultiBox Detector," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Sep. 2016, pp. 21–37. doi: https://doi.org/10.1007/978-3-319-46448-0_2.

[38] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar, "Focal loss for dense object detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1–1, 2018, doi: https://doi.org/10.1109/tpami.2018.2858826.

[39] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention Is All You Need," *arXiv preprint* arXiv:1706.03762, Jun. 2017, doi: 10.48550/arXiv.1706.03762 .

[40] N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov, and S. Zagoruyko, "End-to-End Object Detection with Transformers," in *Computer Vision – ECCV 2020*, A. Vedaldi, H. Bischof, T. Brox, and J.-M. Frahm, Eds., *Lecture Notes in Computer Science*, vol. 12346. Cham, Switzerland: Springer, 2020, pp. 213–229, doi: 10.1007/978-3-030-58452-8_13 .

[41] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Usunier, and N. Houlsby, "An Image Is Worth 16x16 Words: Transformers for Image Recognition at Scale," *arXiv preprint* arXiv:2010.11929, 2020, doi: 10.48550/arXiv.2010.11929.

[42] Q. Shi and J. Li, "Objects Detection of UAV for Anti-UAV Based on YOLOv4," *2020 IEEE 2nd International Conference on Civil Aviation Safety and Information Technology (ICCASIT*, pp. 1048–1052, Oct. 2020, doi: https://doi.org/10.1109/iccasit50869.2020.9368788.

[43] H. Liu, K. Fan, Q. Ouyang, and N. Li, "Real-Time Small Drones Detection Based on Pruned YOLOv4," *Sensors*, vol. 21, no. 10, p. 3374, May 2021, doi: https://doi.org/10.3390/s21103374.

[44] F. Samadzadegan, F. Dadrass Javan, F. Ashtari Mahini, and M. Gholamshahi, "Detection and Recognition of Drones Based on a Deep Convolutional Neural Network Using Visible Imagery," *Aerospace*, vol. 9, no. 1, p. 31, Jan. 2022, doi: https://doi.org/10.3390/aerospace9010031 .

[45] Y. Lv, Z. Ai, M. Chen, X. Gong, Y. Wang, and Z. Lu, "High-Resolution Drone Detection Based on Background Difference and SAG-YOLOv5s," *Sensors*, vol. 22, no. 15, p. 5825, Aug. 2022, doi: https://doi.org/10.3390/s22155825.

[46] A. Coluccia *et al.*, "Drone vs. Bird Detection: Deep Learning Algorithms and Results from a Grand Challenge," *Sensors*, vol. 21, no. 8, p. 2824, Jan. 2021, doi: https://doi.org/10.3390/s21082824 .

[47] X. Zhai, Z. Huang, T. Li, H. Liu, and S. Wang, "YOLO-Drone: An Optimized YOLOv8 Network for Tiny UAV Object Detection," *Electronics*, vol. 12, no. 17, p. 3664, Jan. 2023, doi: https://doi.org/10.3390/electronics12173664 .

[48] H. Sun, J. Yang, J. Shen, D. Liang, L. Ning-Zhong, and H. Zhou, "TIB-Net: Drone Detection Network With Tiny Iterative Backbone," *IEEE Access*, vol. 8, pp. 130697–130707, 2020, doi: https://doi.org/10.1109/access.2020.3009518.

[49] C. Wang *et al.*, "A Lightweight Uav Swarm Detection Method Integrated Attention Mechanism," *Drones (Basel)*, vol. 7, no. 1, pp. 13–13, Dec. 2023, doi: https://doi.org/10.3390/drones7010013 .

[50] C. Wang, Y. Su, J. Wang, T. Wang, and Q. Gao, "UAVSwarm Dataset: An Unmanned Aerial Vehicle Swarm Dataset for Multiple Object Tracking," *Remote Sensing*, vol. 14, no. 11, p. 2601, May 2022, doi: https://doi.org/10.3390/rs14112601 .

[51] J. Liu, L. Plotegher, E. Roura, J. Cristino, and S. He, "Real-Time Detection for Small UAVs: Combining YOLO and Multi-frame Motion Analysis," *arXiv.org*, 2024. https://arxiv.org/abs/2411.02582 .

[52] A. Coluccia, Alessio Fascista, L. Sommer, A. Schumann, A. Dimou, and Dimitrios Zarpalas, "The Drone-vs-Bird Detection Grand Challenge at ICASSP 2023: A Review of Methods and Results," *IEEE Open Journal of Signal Processing*, vol. 5, pp. 766–779, Jan. 2024, doi: https://doi.org/10.1109/ojsp.2024.3379073 .

[53] M. Huang, W. Mi, and Y. Wang, "EDGS-YOLOv8: An Improved YOLOv8 Lightweight UAV Detection Model," *Drones*, vol. 8, no. 7, p. 337, Jul. 2024, doi: https://doi.org/10.3390/drones8070337 .

[54] J. Zhao, J. Zhang, D. Li, and D. Wang, "Vision-Based Anti-UAV Detection and Tracking," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 12, pp. 25323–25334, May 2022, doi: https://doi.org/10.1109/tits.2022.3177627 .

[55] Y. Zheng, Z. Chen, D. Lv, Z. Li, Z. Lan, and S. Zhao, "Air-to-Air Visual Detection of Micro-UAVs: An Experimental Evaluation of Deep Learning," *IEEE Robotics and Automation Letters*, vol. 6, no. 2, pp. 1020–1027, Apr. 2021, doi: https://doi.org/10.1109/lra.2021.3056059 .

[56] Y. Gao, Y. Xin, H. Yang, and Y. Wang, "A Lightweight Anti-Unmanned Aerial Vehicle Detection Method Based on Improved YOLOv11," *Drones*, vol. 9, no. 1, p. 11, Dec. 2024, doi: https://doi.org/10.3390/drones9010011 .

[57] H. Hao *et al.*, "A High Performance Air-to-Air Unmanned Aerial Vehicle Target Detection Model," *Drones*, vol. 9, no. 2, p. 154, Feb. 2025, doi: https://doi.org/10.3390/drones9020154 .

[58] Z. Chu, T. Song, R. Jin, and T. Jiang, "An Experimental Evaluation Based on New Air-to-Air Multi-UAV Tracking Dataset," *2023 IEEE International Conference on Unmanned Systems (ICUS)*, pp. 671–676, Oct. 2023, doi: https://doi.org/10.1109/icus58632.2023.10318461 .

[59] Q. Zhang *et al.*, "BRA-YOLOv10: UAV Small Target Detection Based on YOLOv10," *Drones*, vol. 9, no. 3, p. 159, Feb. 2025, doi: https://doi.org/10.3390/drones9030159 .

[60] Q. Cheng, Y. Wang, W. He, and Y. Bai, "Lightweight air-to-air unmanned aerial vehicle target detection model," *Scientific Reports*, vol. 14, no. 1, p. 2609, Jan. 2024, doi: https://doi.org/10.1038/s41598-024-53181-2 .

[61] S. Zhou *et al.*, "A Lightweight Drone Detection Method Integrated into a Linear Attention Mechanism Based on Improved YOLOv11," *Remote Sensing*, vol. 17, no. 4, p. 705, Feb. 2025, doi: https://doi.org/10.3390/rs17040705 .

[62] V. Walter, M. Vrba, and M. Saska, "On training datasets for machine learning-based visual relative localization of micro-scale UAVs," pp. 10674–10680, May 2020, doi: https://doi.org/10.1109/icra40945.2020.9196947 .

[63] B. K. Isaac-Medina, M. Poyser, D. Organisciak, C. G. Willcocks, T. P. Breckon, and H. P. H. Shum, "Unmanned Aerial Vehicle Visual Detection and Tracking using Deep Neural Networks: A Performance Benchmark," in *Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, Montreal, QC, Canada, Oct. 2021, pp. 1223–1232. doi: 10.1109/ICCVW54120.2021.00142.

[64] N. Jiang *et al.*, "Anti-UAV: A Large Multi-Modal Benchmark for UAV Tracking," *arXiv.org*, 2021. https://arxiv.org/abs/2101.08466 .

[65] A. Rodriguez-Ramos, J. Rodriguez-Vazquez, C. Sampedro, and P. Campoy, "Adaptive Inattentional Framework for Video Object Detection With Reward-Conditional Training," *IEEE Access*, vol. 8, pp. 124451–124466, Jan. 2020, doi: https://doi.org/10.1109/access.2020.3006191 .

[66] A. S. Mubarak, M. Vubangsi, F. Al-Turjman, Z. S. Ameen, A. S. Mahfudh, and S. Alturjman, "Computer Vision Based Drone Detection Using Mask R-CNN," *2022 International Conference on Artificial Intelligence in Everything (AIE)*, pp. 540–543, Aug. 2022, doi: https://doi.org/10.1109/aie57029.2022.00108 .

[67] M. Ozel, Drone, "Drone Dataset," Kaggle, [Online]. Available: https://www.kaggle.com/dasmehdixtr/drone-dataset-uav. [Accessed: Oct. 12, 2025].

[68] R. Padilla, S. L. Netto, and E. A. da Silva, "A Survey on Performance Metrics for Object-Detection Algorithms," in *Proc. 2020 Int. Conf. Systems, Signals and Image Processing (IWSSIP)*, Rio de Janeiro, Brazil, Jul. 2020, pp. 237–242, doi: https://doi.org/10.1109/iwssip48289.2020.9145130.