drones

Review

A Survey on Vision-Based Anti Unmanned Aerial Vehicles Methods

Bingshu Wang [1,2], Qiang Li [1], Qianchen Mao [1], Jinbao Wang [2], C. L. Philip Chen [3,4], Aihong Shangguan [5,*] and Haosu Zhang [5,*]

[1] The School of Software, Northwestern Polytechnical University, Xi'an 710129, China; wangbingshu@nwpu.edu.cn (B.W.); lq2023@mail.nwpu.edu.cn (Q.L.); mqc@mail.nwpu.edu.cn (Q.M.)
[2] National Engineering Laboratory for Big Data System Computing Technology, Shenzhen University, Shenzhen 518060, China; wangjb@szu.edu.cn
[3] The School of Computer Science and Engineering, South China University of Technology, Guangzhou 510641, China; philipchen@scut.edu.cn
[4] Pazhou Lab, Guangzhou 510335, China
[5] Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an 710119, China
[*] Correspondence: xiner@opt.ac.cn (A.S.); zhanghaosu@opt.ac.cn (H.Z.)

Abstract: The rapid development and widespread application of Unmanned Aerial Vehicles (UAV) have raised significant concerns about safety and privacy, thus requiring powerful anti-UAV systems. This survey provides an overview of anti-UAV detection and tracking methods in recent years. Firstly, we emphasize the key challenges of existing anti-UAV and delve into various detection and tracking methods. It is noteworthy that our study emphasizes the shift toward deep learning to enhance detection accuracy and tracking performance. Secondly, the survey organizes some public datasets, provides effective links, and discusses the characteristics and limitations of each dataset. Next, by analyzing current research trends, we have identified key areas of innovation, including the progress of deep learning techniques in real-time detection and tracking, multi-sensor fusion systems, and the automatic switching mechanisms that adapt to different conditions. Finally, this survey discusses the limitations and future research directions. This paper aims to deepen the understanding of innovations in anti-UAV detection and tracking methods. Hopefully our work can offer a valuable resource for researchers and practitioners involved in anti-UAV research.

Keywords: anti-UAV detection; UAV tracking; anti-UAV systems; anti-UAV datasets

check for updates

Citation: Wang, B.; Li, Q.; Mao, Q.; Wang, J.; Chen, C.L.P.; Shangguan, A.; Zhang, H. A Survey on Vision-Based Anti Unmanned Aerial Vehicles Methods. *Drones* **2024**, *8*, 518. https://doi.org/10.3390/drones8090518

Academic Editor: Anastasios Dimou

Received: 5 August 2024
Revised: 15 September 2024
Accepted: 16 September 2024
Published: 23 September 2024

1. Introduction

In recent years, it has made significant progress for Unmanned Aerial Vehicles (UAV) in intelligence and automation [1–5]. The ease of operation, low cost, and high efficiency have contributed to the widespread application across various domains [6–11]. However, the extensive use of UAV has also raised concerns regarding safety and privacy problems. In urban areas, incidents of unauthorized or excessive UAV flights, as well as their exploitation by illicit actors for criminal activities, have seriously infringed upon individual privacy, endangered public safety, and disrupted social order.

The demand for anti-UAV technology is becoming increasingly important. As shown in Figure 1, Anti-UAV detection and tracking is an early stage of anti-UAV technology, aimed at addressing security issues caused by the widespread use of drones, such as illegal entry into no-fly zones, threats to public safety, and unauthorized surveillance of sensitive targets [12–15]. By accurately detecting and tracking illegal UAV, this process provides support and data for subsequent countermeasures like interference, capture, or destruction. In recent years, many anti-UAV methods [16–19] have been introduced to address the issues caused by the misuse of drones. These methods predominantly rely on physical countermeasures, such as UAV radar detection, radio frequency analysis, and acoustic detection [20,21]. Traditional radar systems exhibit limited effectiveness in detecting small UAV, particularly in complex terrains and urban environments where there are numerous reflections and interferences. The radio frequency and acoustic detection systems are usually low-cost and easy to deploy, but they are very susceptible to electromagnetic and noise interference in urban environments. These methods share a common characteristic: they do not utilize visual information. Visual information holds unique advantages in UAV detection, as it allows for more intuitive and precise identification and tracking by capturing and analyzing the visual features of the UAV. However, the processing of visual information also present several challenges: (1) UAV targets often undergo dramatic scale changes, disappear frequently during flight, and tracking performance is heavily influenced by camera motion [22,23]; (2) in infrared scenes, UAV targets have small scales, low resolution, lack appearance information, and are easily overshadowed by background information; (3) the flying of UAV in complex scenes can lead to target occlusion and unstable flight paths.

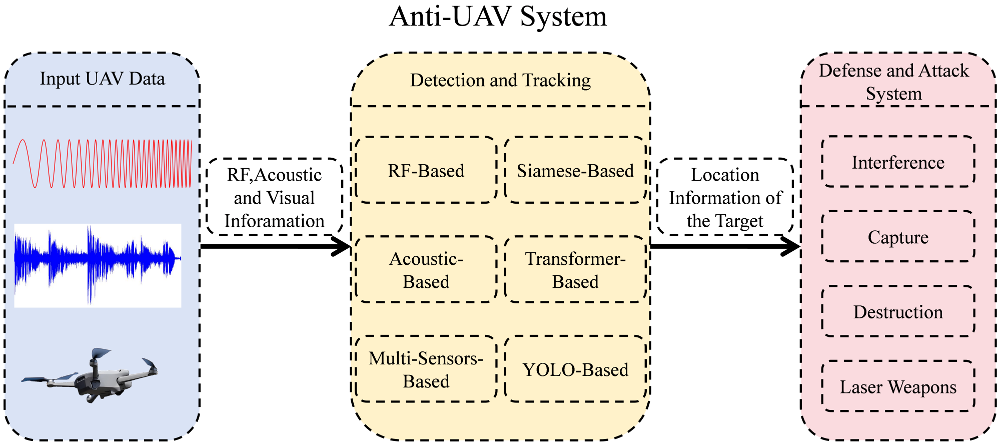

Figure 1. Anti-UAV Systems: The detection and tracking of UAV represent the early stages of anti-UAV technology.

The motivation of this paper is to summarize the advances in the field of anti-UAV detection techniques, aiming to provide reference for researchers and engineers to develop more efficient models. The contributions of this paper are concluded as follows:

- This paper surveys recent methods for anti-UAV detection. We classify the collected methods based on the types of backbones used in the article and the scenarios in which they are applied. The methods are classified by Sensor-Based methods and Vision-Based methods. Typical examples are outlined to illustrate the main thoughts.
- We collect and summarize the public anti-UAV datasets, including RGB images, infrared images and acoustic data. Additionally, the dataset links are also provided so that the readers can access them quickly.
- The advantages and disadvantages of existing anti-UAV methods are analyzed. We give detailed discussions about the limitations of anti-UAV datasets and methods. Meanwhile, five potential directions are suggested for the future research.

This paper is organized as follows. Section 2 provides a detailed overview of the relevant literatures. Section 3 gives the public datasets of anti-UAV detection and tracking. Section 4 summarizes the advantages and disadvantages of anti-UAV detection and tracking methods. Section 5 discusses the limitations of datasets and methods, as well as future research directions. Section 6 concludes the paper.

2. Analysis of Surveyed Literatures

Since the increasing popularity of small drones and the rapid development of deep learning technology, many innovative research papers [24–29] have been proposed in the field of anti-UAV detection and tracking. These scholarly articles not only explore various efficient algorithms but also demonstrate the potentials of utilizing deep learning for UAV identification and tracking. In this section, we describe the statistical data of the state-of-the-art methods.

2.1. Stats of Surveyed Literatures

We mainly gather literatures of anti-UAV methods on Google Scholar, IEEE Xplore, Web of Science, and some academic websites for investigation. The search keywords are "Anti-UAV", "UAV Detection", "UAV Tracking", "Deep learning for anti drone detection". The literatures were surveyed by the end of 31 August 2024. Notably, we select literatures published within the last 5 years, which can represents the advances.

In the collection of several hundred results, we imposed certain restrictions to streamline our selection: (1) The literature must be written in English; (2) Owing to the rapid development of technology, we limited our selection to literature published within the last 5 years; (3) We specifically targeted methodologies based on deep learning, computer vision technologies sensor fusion; (4) Preference was given to literature published in well-known academic journals and conferences, such as Drones, IEEE, CVPR, Science sub-journals, etc.

These primarily included journal articles, conference papers, arXivs, and various competition documents. In this study, we classified them by the country and publication time. Figure 2 summarizes the number of papers from different countries or regions around the world and presents the distribution of publication dates. As shown in the left image, it suggests that the number of literature keeps an increasing trend. It means that the field has been attracting more and more attentions. In the right image, the data reflects that China publishes the most of literatures, with an 68% ratio. This may be beneficial from the increasing investment in drone field.

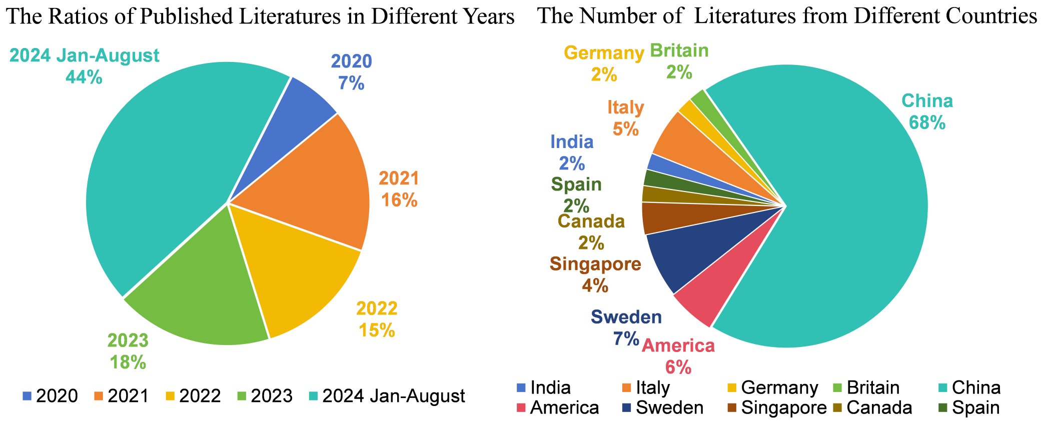

Figure 2. Stats of state-of-the-art methods. The left is based on the year, and the right is based on the author's country or area.

2.2. Method Classification of Surveyed Literatures

Based on our comprehensive analysis of the collected literature, anti-UAV methods have been effectively summarized and categorized, as illustrated in Figure 3 outlines the mainstream approaches. These methods are divided into two categories: Sensor-Based and Vision-Based.

Sensor-Based methods for anti-UAV detection and tracking represent one of the earliest technologies [30] to tackle the challenges posed by unmanned aerial vehicles. For instance, Radio Frequency (RF) and Acoustics play a significant role in anti-UAV systems. However, they face certain challenges in dealing with more complex flying environments. To leverage the complementary advantages of different sensors, these methods are often integrated into multi-sensor systems [31].

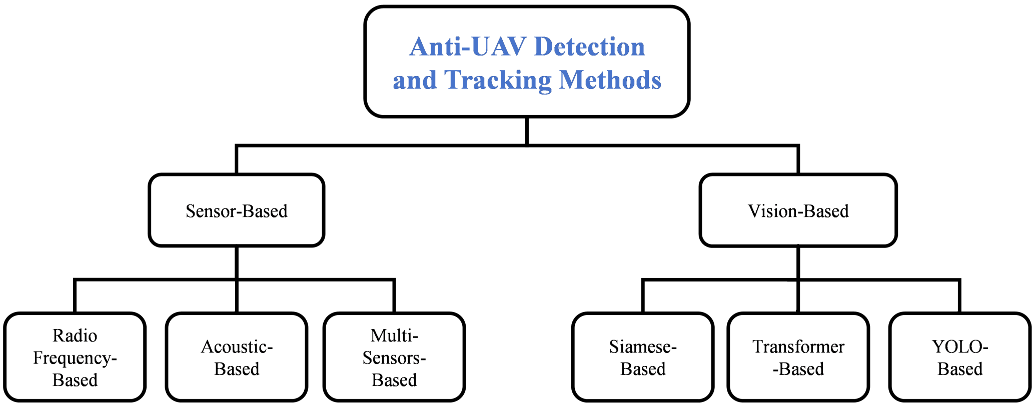

Figure 3. Hierarchical representation of the state-of-the-art methods of anti-UAV detection and tracking.

Vision-Based methods apply to anti-UAV detection and tracking have significantly improved in terms of accuracy and speed [32–38]. We divide it into three categories based on real-time tracking performance, global perception ability, and practicality: Siamese-Based, Transformer-Based, and YOLO-Based.

3. Anti-UAV Detection and Tracking Datasets

To address the challenges of anti-UAV detection, researchers worldwide have proposed numerous datasets to train models. These datasets typically include a vast number of UAV images, video sequences, sensor readings and corresponding environmental background data. This section will provides a detailed description and discussion of these datasets.

Zhao et al. [20] proposed a multi-UAV dataset named DUT anti-UAV, which include a detection dataset and a tracking dataset. The detection dataset contains 5200 images for training, 2600 images for validation and 2200 images for testing. Each image comes with an annotation file detailing the objects. The tracking dataset contains 20 sequences. This dataset consists of RGB images, with a small number of close-range UAV images collected from the internet included in the training set, while the rest are manually collected mid-to-long-range UAV images. Although the dataset is extensive, it has a limited variety of scene types.

Jiang et al. [23] designed a large-scale benchmark dataset that consists of 318 RGB-T video pairs, each containing an RGB video and an infrared video, along with annotation files for each video. The author recorded videos of different types of drones, primarily from DJI and Parrot, flying in the air, which were used to collect tracking data. The recorded videos capture scenes under two different lighting conditions(day and night), two types of light modes (infrared and visible), and different backgrounds (trees, clouds, buildings). Each video is saved as an MP4 file format, with a frame rate of 25 FPS. However, the infrared scene dataset has a lower resolution, and there are fewer scenarios of drones flying at long distances.

Yuan et al. [39] created a dataset named MMAUD, which integrates multiple sensor inputs including vision, radar, and audio arrays. It encompasses various types of drones and noise sequences. The speed, size, and estimated radar cross section of the drones are also modeled with accuracy relative to ground truth values. The dataset simulates real-world scenarios by incorporating the sounds of surrounding heavy machinery, capturing the precise challenges faced when operating near vehicles, and thereby enhancing the applicability. Notably, the dataset is accessible in two formats: rosbag format and file system format, making it versatile for different applications.

Fredrik et al. [31] introduced a multi-sensor dataset that supplements audio datasets with classes for drones, helicopters, and background noise. It is captured at three airports in Sweden and utilizes three different types of drones: the Hubsan H107D+ (a small first-person view drone), the high-performance DJI Phantom 4 Pro and the medium-sized

DJI Flame Wheel (F450) [40]. The dataset includes 90 audio clips, 650 videos, with a total of 203,328 annotated images. The audio data includes recordings of small drones, large helicopters, and background sounds such as birdsong, wind noise, water flow, etc. The IR videos have a resolution of 320 × 256 pixels, and the visible videos 640 × 512 pixels. The maximum sensor-to-target distance for drones in the dataset is 200 m. For ease of annotation, both videos and audio clips are trimmed to 10 s in duration.

The dataset proposed by Vedanshu et al. [41] is mainly composed of quadcopter images. It contains 1847 images, most of which were taken from Kaggle [42] and some of which were taken with smartphone selfies. Self-shot images are taken by considering visual differences in distance (near, middle and far) analysis. The scenes are different between training set and testing set.

Christian [43] collected 500 images of the DJI Phantom 3 quadcopter from Google's image search and numerous screenshots from YouTube videos. In this dataset 350 images are used for training and 150 images for testing. It mainly focuses on individual drone targets at close range, and due to the close capture distance, the scale ratio of the drone is relatively large.

Zheng et al. [44] developed a large-scale dataset named DetFly, comprising over 13,271 images, each with a resolution of 3840 × 2160 pixels. These images are captured by a drone shooting another flying target drone. The dataset notably includes some challenging scenarios, such as strong and weak lighting conditions, motion blur, and partial occlusions, which further enhance the dataset's practicality and the complexity of testing algorithms. Approximately half of the images feature UAV objects that are smaller than 5% of the total image size. Although the image resolution is high, the images contain only a single target and are exclusively of one type of UAV (DJI Mavic), which may limit the dataset's diversity.

Viktor et al. [45] introduced a dataset known as MIDGARD, which is automatically annotated through the Ultra Violet Direction And Ranging (UVDAR) system. This dataset is collected in various environments, such as rural areas, urban landscapes, as well as modern and classical environments. Additionally, it includes many challenging scenarios, such as the disappearance of micro UAV and situations where they are obscured. The dataset also provides annotations, including the position of drones in the images and their bounding boxes, as well as their approximate distances, supporting effective detection and location training for UAV.

The dataset provided by the 3rd Anti-UAV competition [46] consists of infrared images derived from video sequences. It is characterized by its large scale and the richness of its content. This dataset includes scenarios where targets are present within the image, targets are at the image boundaries, and targets disappear, among others. It primarily supports the competition's two tracks: Track 1, anti-UAV tracking (with the target present in the first frame), and Track 2, anti-UAV detection and tracking (with the target's presence unknown in the first frame). However, some images contain textual information from the equipment at the top, which may potentially interfere with model training.

In summary, the overview of the mentioned datasets is presented in Table 1. Some examples from these datasets are shown in Figure 4. Table 2 also provides the corresponding available links. All links have been verified as valid before 28 March 2024.

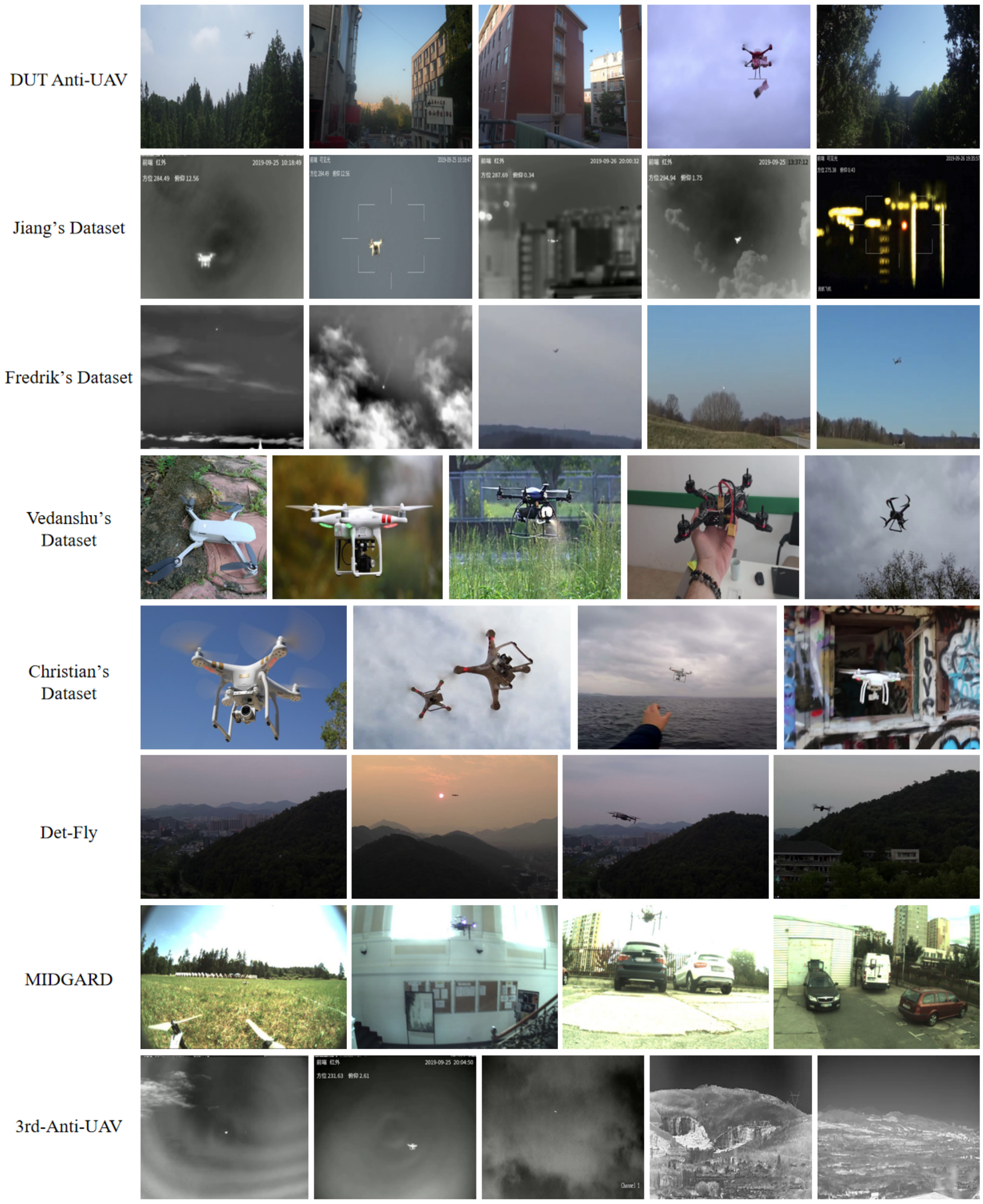

Figure 4. Some samples selected from different datasets in DUT Anti-UAV [20], Jiang's Dataset [23], Fredrik's Dataset [31], Vedanshu's Dataset [41], Christian's Dataset [43], Det-Fly [44], MIDGARD [45], 3rd-Anti-UAV [46]. The Chinese characters are the shooting time and place of the images.

Table 1. Detailed descriptions of some open datasets for anti-UAV detection and tracking.

Dataset Name	Main Characteristics	Image or Video Number	Complexity	Multi-Sensors	Scene	UAV Type
DUT Anti-UAV [20]	Images and videos of various scenes are from around DUT, such as the sky, dark clouds, jungles, high-rise buildings, residential buildings, farmlands and playgrounds	Image 10,000 Video 20	Medium	NO	RGB	Not available
Jiang's Dataset [23]	DJI and Parrot drones are used to captured video from the air. The video records two lighting conditions (day and night), two light modes (infrared and visible light), and a variety of backgrounds (buildings, clouds, trees)	Video_RGB 318 Video_Infrared 318	Large	NO	RGB and Infrared	DJI and Parrot drones

Table 1. *Cont.*

Dataset Name	Main Characteristics	Image or Video Number	Complexity	Multi-Sensors	Scene	UAV Type
MMAUD [39]	This dataset is built by integrating multiple sensing inputs including stereo vision, various Lidar, radar, and audio arrays	Multi-category	Large	Yes	Infrared	Not available
Fredrik's Dataset [31]	This dataset is captured at three airports in Sweden and three different drones are used for the video shooting	Audio 90 Video_RGB 285 Video_Infrared 365	Large	Yes	Infrared	DJI Phantom4 Pro,DJI Flame Wheel and Hubsan H107D+
Vedanshu's Dataset [41]	The majority of the dataset's images are collected from Kaggle and the remaining are captured using a smartphone camera	Image 1874	Medium	NO	RGB	Not available
Drone Detection [43]	Images of the DJI Phantom 3 quadcopter obtained through Google image search and dozens of screenshots from YouTube videos	Image 500	Small	NO	RGB	DJI Phantom 3 quadcopter
Det-Fly [44]	This dataset uses a flying drone (DJI M210) to photograph another flying target drone (DJI Mavic)	Image 13,271	Medium	NO	RGB	DJI M210 and DJI Mavic
MIDGARD [45]	This dataset is automatically generated using relatively Micro-scale Unmanned Aerial Vehicles and positioning sensor	Image 8776	Medium	NO	Infrared	Not available
3rd-Anti-UAV [46]	This dataset consists of single-frame infrared images derived from video sequences	Video Sequence	large	NO	Infrared	Not available

Table 2. Available websites of open-source datasets.

Dataset Name	Available Link	Access Date
DUT Anti-UAV [20]	https://github.com/wangdongdut/DUT-Anti-UAV	22 January 2024
Jiang's Dataset [23]	https://github.com/ucas-vg/Anti-UAV	22 January 2024
MMAUD [39]	https://github.com/ntu-aris/MMAUD	22 January 2024
Fredrik's Dataset [31]	https://github.com/DroneDetectionThesis/Drone-detection-dataset	15 March 2024
Vedanshu's Dataset [41]	https://drive.google.com/drive/folders/1FJ09dOOa-VFMy_tM7UoZGzOA8iYpmaHP	15 March 2024
Drone Detection [43]	https://github.com/creiser/drone-detection	11 April 2024
Det-Fly [44]	https://github.com/Jake-WU/Det-Fly	15 March 2024
MIDGARD [45]	https://mrs.felk.cvut.cz/midgard	21 March 2024
3rd-Anti-UAV [46]	https://anti-uav.github.io	28 March2024

4. Anti-UAV Detection and Tracking Methods

In this section, we reviewed the literatures on anti-UAV detection and tracking from recent years. It starts from early Sensor-Based methods [47–50], such as radio frequency signatures and acoustics. Then, it mainly focus on Vision-Based detection methods [34,51–55]. Various methods have been proposed to address the challenges posed by the dramatic scale changes [56–62] during flight, frequent disappearances, and unstable flight paths of UAV. These methods [63–69] primarily focus on leveraging advanced image processing and deep learning technologies to enhance the accuracy of UAV identification and tracking in complex environments.

4.1. Sensor-Based Methods

The Sensor-Based anti-UAV detection methods mainly rely on non-visual technologies such as radio frequency spectrum monitoring and acoustic signal analysis. These techniques detect the presence of UAV by sensing the physical characteristics of radio and sound waveforms signals [70–73]. Figure 5 shows the general Sensor-Based methods. It involves multiple technologies aimed at addressing UAV threats in different environments.

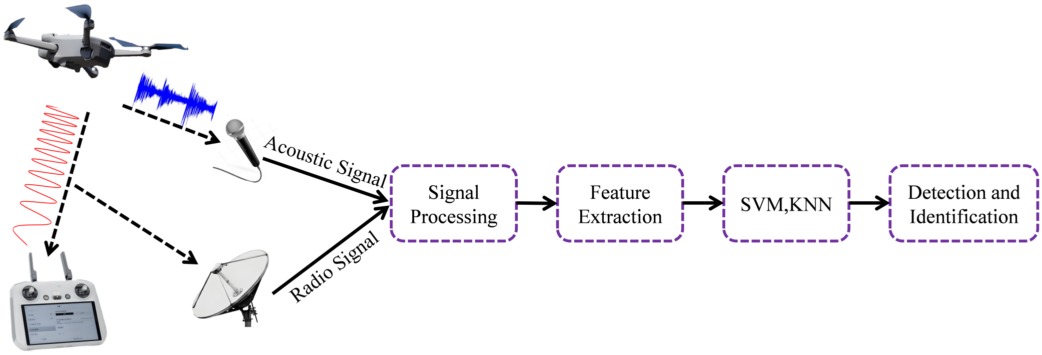

Figure 5. Overview of general Sensor-Based methods. The classifiers can be SVM or KNN.

4.1.1. RF-Based

By monitoring RF signals within a specific area, a large amount of radio signal data is collected. Based on this data, non-relevant interference signals are filtered out by analyzing the technical parameters, frequency characteristics, and specific UAV communication protocols of these signals. Subsequently, precise feature recognition of the UAV communication signals is achieved using machine learning algorithms. In the process of RF-Based UAV detection, features are first extracted from the received RF signals. These extracted features are then compared with a database of UAV radio frequency characteristics to achieve UAV identification.

Xiao et al. [47] proposed a method to detect and identify small UAV using RF signals from UAV downlink communications. This process involves framing and preprocessing continuous time-domain signals, extracting cyclostationary and spectral features. The features extracted are utilized to train Support Vector Machine (SVM) and k-Nearest Neighbors (KNN) classifiers. The performance of the classifiers is tested under various Signal-to-Noise Ratio (SNR) conditions. They also described the features of different micro UAV signals. This method is applicable in more general cases and can be implemented through machine learning approaches, yet it is susceptible to environmental noise interference, such as signals from WiFi and Bluetooth, which may affect the detection and identification of UAV signals.

4.1.2. Acoustic-Based

Although sound waves and electromagnetic waves differ in nature, Acoustic-Based and RF-Based detection methods share certain similarities due to the fundamental properties of their respective waves [30]. Acoustic-Based [74,75] detection methods play a complementary role in anti-UAV technology. These methods leverage the unique sounds produced by a UAV's propellers as they disturb the airflow during flight. The sound characteristics differ depending on the type of UAV, flight speed, and environmental conditions. High-sensitivity microphones are used to capture these sounds, and the system then extracts and analyzes the sound signals to identify the presence of a UAV.

Yang et al. [48] introduced a low-cost UAV detection system that utilizes multiple acoustic nodes and machine learning models to detect and track UAV through the analysis of audio signals. This study employs two types of features: Mel-frequency cepstral coefficients and Short-Time Fourier Transform (STFT), which are trained using SVM and Convolutional Neural Networks (CNN), respectively. Each model was tested individually during the evaluation phase. The results indicated that the STFT-SVM model yielded the best outcomes, demonstrating high detection accuracy and robustness. Unlike optical methods, acoustic detection is not affected by visual obstructions such as fog or darkness, hence offering robustness against visual impairments. However, the system is sensitive to noise. In noisy environments, surrounding sounds might mask the UAV's acoustic features, leading to degraded system performance.

4.1.3. Multi-Sensors-Based

Multi sensor detection methods involve integrating and analyzing data from different types of sensors [76–80] to improve the accuracy and efficiency of detecting and tracking targets. The key steps mainly include sensor selection, data preprocessing, data fusion, pattern recognition, and decision specification. Among them, data fusion [81,82] is a key technology that combines data from multiple sensors into a representative information.

Fredrik et al. [31] explored the process of designing an automatic multi-sensor UAV detection system. The discussion cover various sensors used for UAV detection, including radar, visible light cameras, thermal infrared cameras, microphones, radio frequency scanners, lasers, etc. Even with slightly lower resolution, its performance was comparable to cameras operating within the visible spectrum. The study also examined detector performance as a function of the distance from the sensor to the target. Through multi-sensors fusion, the system's robustness is enhanced beyond that of individual sensors, aiding in the reduction of false positives. However, the fusion of data from multiple sensors requires sophisticated algorithms and computational resources, potentially impacting the real-time capabilities of the detection system.

Xie et al. [83] proposed a framework that integrates RF and visual information. The complementarity of these two modalities alleviates the limitations of single-sensor approaches, which are prone to interference. The authors also introduced a denoising method based on image segmentation (ISD-UNet), providing a novel approach to eliminating noise in RF data. Compared to traditional signal processing denoising techniques, deep learning-based methods can better preserve the local features and spatiotemporal information of RF signals. The authors constructed their own RF-visual dataset across multiple scenarios, but this dataset has not been made publicly available.

The approach based on RF signals utilizes feature engineering and machine learning techniques for effective detection and identification of micro UAV. It is particularly suited for scenarios requiring differentiation among various UAV types. The method is reported to be low-cost, easily scalable, and applicable for acoustic signal detection. The Multi Sensors-based method enhances detection robustness and accuracy by integrating data from various sensors. The advantages of Sensor-Based methods include their ability to operate under various lighting conditions and to penetrate certain obstacles. However, these methods also have limitations; for instance, acoustic detection is susceptible to interference from environmental noise, and radio frequency spectrum monitoring struggles with UAV employing autonomous flight or encrypted communications.

4.2. Vision-Based Methods

In recent years, many novel methods have been proposed based on visual information to address the challenges in anti-UAV in real-world applications [84–89]. Vision-Based approaches are more prevalent in anti-UAV tasks because they offer greater flexibility, higher accuracy, and higher efficiency [22,90–93]. Vision-based algorithms detect and track targets by analyzing drone features such as shape, color, and texture in single-frame images or video sequences, with notable examples including the Siamese network [94–96], Transformer [97,98], and YOLO [99–104] series. These techniques are mainly applied in natural and infrared scene contexts [36,105–108]. Natural scenes refer to visible-light-based environments, while infrared scenes rely on thermal radiation imaging technology, using infrared cameras to detect the thermal radiation characteristics of objects. Since infrared imaging [109,110] does not require external light sources, it has a clear advantage at night. During flight, drones generate significant heat from components like motors and batteries, leading to relatively high thermal radiation, making them more distinguishable in infrared images. The design process of Vision-Based methods is shown in Figure 6. The left visual data, such as images or video sequences, is input and processed by the detection network. Detected targets are then directly marked and presented in the original input.

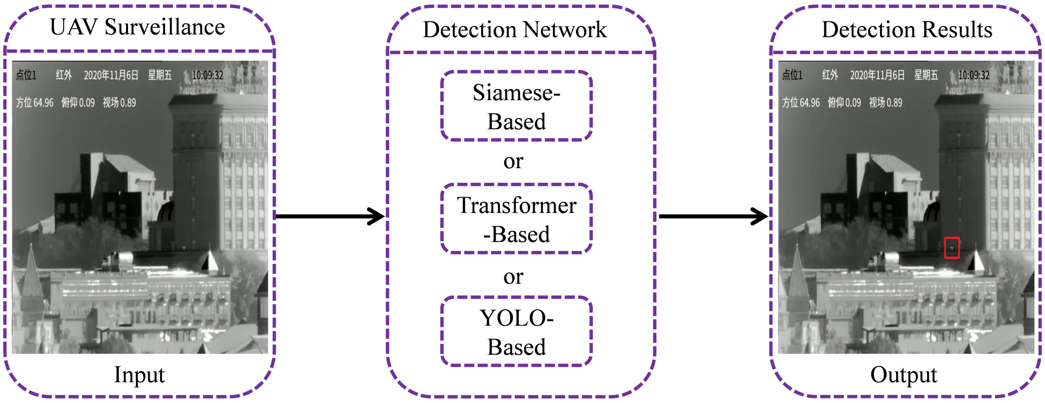
UAV Surveillance
点位1 红外 2020年11月6日 星期五 10:09:32
方位64.96 俯仰0.09 视场0.89
Input
Detection Network
Siamese-Based
or
Transformer-Based
or
YOLO-Based
Detection Results
点位1 红外 2020年11月6日 星期五 10:09:32
方位64.96 俯仰0.09 视场0.89
Output

Figure 6. Overview of general Vision-Based methods. The input can be a video sequence or an image in RGB or infrared scenes. The Chinese characters are the shooting time and place of the images.

4.2.1. Siamese-Based

The Siamese-Based is a deep learning model based on the siamese architecture. Siamese is a common network in single target tracking [111–115]. Typically, it consists of two similar sub-networks, with each sub-network responsible for processing one input data. The similarity between two inputs is evaluated by comparing the outputs of the two sub-networks. Given its excellent performance, Siamese-Based networks have gained significant popularity in the field of target tracking, becoming a popular method for learning the similarity between target templates and corresponding regions in search image in tracking applications [116,117]. Bertinetto et al. [118] first conceptualized the tracking task as a similarity matching problem between the target template and the search region, Siamese has become a focal point of research in the field of target tracking.

Huang et al. [119] designed a Siamese framework composed of template branch and detection branch, named SiamSTA, and applied it to track UAV in infrared videos. Facing the challenges of small scale and rapid movement of UAV in infrared scenarios, SiamSTA integrates local tracking and global detection mechanisms. The local tracking approach employs both spatial and temporal constraints to restrict the position and aspect ratio of candidate proposals generated within the nearby area, thereby suppressing background interference and accurately locating the target. Simultaneously, to address the situation where the target is lost from the local region, a three-stage global re-detection mechanism is introduced. This mechanism utilizes valuable motion features from a global view by employing a change detection-based correlation filter to re-detect the target. Finally, a state-aware switching strategy is adopted to adaptively apply local tracking and global re-detection based on different target states. The spatiotemporal attention mechanism and three-stage re-detection mechanism may heighten the computational complexity of the algorithm and affect tracking speed.

Fang et al. [120] combined the high-speed multi-stage detector YOLOv3 with Siamese networks to create a real-time multi-scale object tracking model for infrared scenes, named SiamYOLO. This model utilizes a global real-time perception mechanism to initially identify candidate targets, and then leverages spatiotemporal information to eliminate distractions and obtain true UAV targets. In the multi-scale fusion operation, a channel attention mechanism is introduced to enhance target-specific features while suppressing non-UAV target features. Additionally, the spatiotemporal information of candidate targets is utilized along with Kalman filtering to accurately locate real UAV targets from their flight trajectories. Their approach is validated through comparative experiments on their self-constructed dataset, demonstrating a balance between tracking accuracy and speed compared to other competitive models.

Huang et al. [121] proposed a new method called Siamese Drone Tracker (SiamDT), which uses a dual semantic feature extraction mechanism (DS-RPN) to generate candidate proposals in order to address the issue of small drone tracking in dynamic backgrounds and cluttered environments. The authors also introduced a branch specifically designed to suppress background interference. Experiments conducted on the self-built Anti-UAV410 dataset demonstrated that SiamDT can effectively distinguish the target from the background, thereby improving the accuracy of drone tracking.

Shi et al. [122] developed a Siamese tracker based on graph attention, termed GASiam, which builds upon SiamR-CNN [123] as its foundation and introduces the following improvements: (1) Enhanced the feature extraction capability for infrared UAV targets by refining the MobileNetV2 [124]. (2) Introduced a graph attention module for local tracking, utilizing local feature matching to enhance the embedding of information between the target template and search region. (3) Designed a switching strategy in conjunction with a three-stage global re-detection function, allowing the network to adaptively choose between local tracking and global re-detection strategies to enhance the speed of re-capturing targets after loss. The effectiveness of the graph attention module and switching strategy depends on appropriate parameter settings and sufficient model training, which may require extensive experimentation to optimize.

Cheng et al. [125] designed a method for long-term anti-UAV target tracking that incorporates a Siamese network and a re-detection module. To address the issues of target repositioning and updating templates in extended duration tracking of UAV targets, a hierarchical discriminator is employed to produce target localization response maps based on the Siamese network's output. Furthermore, reliability criteria are established to assess the confidence of the response maps. When the confidence level of the output is low, the algorithm triggers the re-detection module and refreshes the template. Compared to the strong baseline SiamRPN++ [111], the success rate and accuracy of this method are improved by 13.7% and 16.5% respectively.

Xie et al. [126] proposed a novel spatio-temporal focused Siamese network, named STFTrack. This method adopts a two-stage object tracking framework. Firstly, a siamese backbone based on feature pyramid is constructed, enhancing the feature representation of infrared UAV through cross-scale feature fusion. By integrating template and motion features, the method directs previous anchor boxes towards suspicious regions for adaptive search area selection. This effectively suppresses background interference and generates high-quality candidate targets. Secondly, it proposes an instance-discriminative region-CNN based on metric learning to further refine the focus on infrared UAV within candidate targets. Finally, the method effectively addresses similar distractors and interference arising from thermal cross talk. However, the generalization ability in natural scenes has not been fully validated yet.

4.2.2. Transformer-Based

In recent years, there have been notable improvements in CNN-based algorithms for tracking targets. However, these CNN-based trackers often struggle to maintain the relationship between the template and the spatial search area. Transformers have been successfully applied to target tracking due to their efficient and effective global modeling capabilities [127–130]. The use of self-attention mechanisms in fusion modules fully exploits global context information, adaptively focusing on useful tracking information.

Yu et al. [22] devised a novel unified Transformer-Based Tracker, termed UTTracker. UTTracker comprises four modules: multi-region local tracking, global detection, background correction, and dynamic small object detection. The multi-region local tracking module is used to handle target appearance changes and multi-region search to track targets in multiple proposals. The global detection module addresses the challenge of frequent target disappearance. The combined background correction module aligns the background between adjacent frames to mitigate the impact of camera motion. The dynamic small object detection module is utilized for tracking small objects lacking appearance information. UTTracker achieves robust UAV tracking in infrared mode, and serves as the foundation of the second-place winning entry in the 3rd Anti-UAV Challenge [46].

Tong et al. [131] proposed a spatiotemporal transformer named ST-Trans for detecting low-contrast infrared small targets in complex backgrounds. By introducing the spatiotemporal transformer module, ST-Trans captures the spatiotemporal information across consecutive frames, enhancing the detection performance of small targets in complex backgrounds. This characteristic is particularly suitable for long-range small-scale UAV detection, as these targets are often difficult to discern in single-frame images. In practical applications, it is necessary to balance its computational complexity with real-time requirements.

4.2.3. YOLO-Based

You Only Look Once (YOLO) [132] is a widely used deep learning algorithm due to its classification or regression-based single-stage object detection approach, which is characterized by its simplicity in structure, small model size, and fast computation speed [133–136]. The integration of YOLO algorithm [137–140] and anti-UAV technology is constantly developing, leading to a surge in research outcomes and applications.

YOLOv3-Based: Hu et al. [141] introduced an improved method for UAV detection using YOLOv3 [142], tailored for the anti-UAV field. By leveraging the last four feature maps for multi-scale prediction, the method captures more texture and contour information to detect small targets. Additionally, to reduce computational complexity, the number of anchor boxes is adjusted according to the size of UAV calculated from the four scale feature maps. Experimental results demonstrate that this approach achieves the highest detection accuracy and most accurate UAV bounding boxes while maintaining fast speed.

YOLOv4-Based: Zhou et al. [92] designed a network deployed on UAV edge devices for detecting and tracking invasive drones, named VDTNet. This method uses YOLOv4 as the backbone, and further improves inference speed through model compression techniques. To compensate for the accuracy loss caused by model compression, the authors introduced SPPS and ResNeck modules integrated into the network's Neck, where SPPS replaces the original SPP module with a 7 × 7 pooling. Finally, the proposed method was deployed on the onboard computer of the drone and conducted real-time detection in both ground-to-air and air-to-air test scenarios. On the FL-Drone dataset, the method achieved an mAP of 98% with a latency of 11.6 ms.

YOLOv5-Based: Fardad et al. [143] developed a method for small UAV detection and classification based on YOLOv5 [144], integrating Mosaic data augmentation and Path Aggregation Network [145] architecture to enhance the model's ability to detect small objects. The model is trained using publicly air-to-air datasets merged with challenge datasets to increase the number of small targets and complex backgrounds. The enhancement of detection accuracy is pursued through the amalgamation of Faster R-CNN [146] and Feature Pyramid Network [147]. Experimental results demonstrate that YOLOv5 performed well in the detection challenge.

Vedanshu et al. [41] utilized different versions of YOLO models to detect small UAV, namely YOLOv5 [144] and YOLOv7 [148]. The results indicate that YOLOv5 performs on datasets with color adjustments, while YOLOv7 performs better on RGB datasets. In the analysis of different distances, YOLOv5 exhibits good accuracy in near, medium, and far-distance scenes, particularly in complex backgrounds. Conversely, YOLOv7 shows poorer detection performance in far-distance and complex background scenarios.

YOLOV7-Based: Li et al. [149] created a motion and appearance-based infrared anti-UAV global-local tracking framework to address issues such as UAV frequently appearing, disappearing, unstable flight paths, small sizes, and background interference. GLTF-MA consists of four modules: Periodic Global Detection module, Multi-stage Local Tracking module, Target Disappearance Judgment module, and Boundary Box Refinement module. GLTF-MA achieves optimal performance in the 3rd Anti-UAV Challenge and the anti-UAV benchmark [23], particularly in scenarios involving fast motion and low resolution.

YOLOV8-Based: Huang et al. [150] proposed an improved model called EDGS-YOLOv8, focused on detecting small-sized UAV, based on YOLOv8. In the detection head, they used the DCNv2 deformable convolutional network to handle local details and scale variations of drones. Additionally, they applied Ghost convolution and the C3Ghost module in the neck part to compress the model, reducing its size to 4.23 MB, making it more suitable for deployment on edge devices. Although the model optimizes computational efficiency, the introduction of complex mechanisms such as EMA and DCNv has led to a decrease in FPS. According to comparison experiments, the improved model's FPS dropped from 96.2 FPS (for the baseline YOLOv8n) to 56.2 FPS. Both the model size and FPS decreased after the improvements, making it less efficient than the baseline YOLOv8n.

In summary, Siamese-Based methods are well-suited for real-time single-target tracking, Transformer-Based methods excel in processing complex scenes and capturing long-range dependencies, and YOLO-Based [151,152] methods emphasize speed and real-time capabilities, suitable for rapid detection and tracking. In practical applications, the choice of method should be based on the specific environment, task requirements, and resource constraints.

4.3. Discussions on the Results of Methods

According to the methods used in this paper, the anti-UAV detection and tracking technology is classified. Table 3 classifies some representative methods and experimental results.

4.3.1. Evaluation Metrics

For the different methods, as shown in Table 3, the evaluation metrics used are distinct. To better understand the evaluation criteria for these methods, the evaluation metrics used in Table 3 are listed here: *Accuracy*, *F*1 *score* (*F*1), *Precision*, *mAP*, *Recall*, *Success rate* (*Success*), *tracking average accuracy* (*acc*).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}, \tag{1}$$

$$F1 = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall}, \tag{2}$$

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

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

$$mAP = \frac{1}{N}\sum_{i=1}^{N} AP_i, \tag{5}$$

where TP represents the true positives, TN represents the true negatives, FP represents false positives, and FN represents false negatives. The mean Average Precision mAP), where N denotes the number of classes and AP_i denotes the Average Precision for the i_{th} class.

$$\text{success} = \frac{1}{T}\sum_{t=1}^{T} \delta(IOU_t > T_0), \tag{6}$$

$$acc = \frac{1}{T}\sum_{t=1}^{T} IOU_t \times \delta(v_t > 0) + p_t \times (1 - \delta(v_t > 0)), \tag{7}$$

where IoU_t defines the Intersection over Union (IoU) between the tracked bounding box and its corresponding ground-truth bounding box. T_o denotes the set threshold. v_t is the ground-truth visibility flags of the target, and δ indicates an indicator function. When the target is truly visible, $v_t = 1$ and $\delta(v_t > 0) = 1$. When the target is invisible, $v_t = 0$ and $v_t(v_t > 0) = 0$. p_t denotes the prediction indicator function. When the prediction is empty, p_t = 1, otherwise $p_t = 0$.

Table 3. Different categories of methods and their detailed descriptions.

Category	Methods		Datasets	Results	Experimental Environment
Sensor-Based	RF-Based	Xiao et al. [47]	Mavic Pro, Phantom3 and WiFi signals	With SVM, the *Accuracy* more than 0.90 at − 3 dB SNR; With KNN, the *Accuracy* more than 0.90 at −4 dB SNR	receiver with 200 MHz on 2.4 GHz ISM frequency
	Acoustic-Based	Yang et al. [48]	The six nodes recorded audio data	With SVM, result of training STFT $F1 = 0.787$, With SVM, result of training MFCC $F1 = 0.779$	$C = 10^i$, where $i = 1, 2, ..., 14, 15$; $\gamma = 10^i$, where $i = -15, -14, ..., -2, -1$
	Multi-Sensors-Based	Fredrik et al. [31]	365 infrared videos, 285 visible light videos and an audio dataset	The average of infrared sensor $F1 = 0.7601$, The average of Visible camera $F1 = 0.7849$, The audio sensor $F1 = 0.9323$	camera, sound acquisition device, and ADS-B receiver
		Xie et al. [83]	Self built multiple background condition UAV detection datasets containing visual images and RF signals	$AP = 44.7$, $FPS = 39$	NVIDIA GeForce GTX 3090 Hikvision pan-tilt-zoom (PTZ) dome camera, and TP-8100 five-element array antenna
Vision-Based	Siamese-Based	Huang et al. [119]	2nd Anti-UAV [46]	$Precision = 88.8\%$, $Success = 65.55\%$, average overlap accuracy = 67.30%	Not provide
		Fang et al. [120]	14,700 infrared images self-built	$Precision = 97.6\%$, $Recall = 97.6\%$, $F1 = 0.976$, acc = 70.3%, $FPS = 37.1$	2.40 GHz Intel Xeon Silver 4210R CPU, 3× NVIDIA RTX3090 GPU and PyTorch 1.8.1 with CUDA 11.1
		Huang et al. [121]	410 self built infrared tracking video sequences	$Precision = 68.19\%$	Not provide
		Shi et al. [122]	Jiang's dataset [23] and 163 videos self-built	$Precision = 94.9\%$, $Success = 71.5\%$	4× NVIDIA Geforce RTX 2080 Super cards, Python 3
		Cheng et al. [125]	Jiang's dataset [23]	$Precision = 88.4\%$, $Success = 67.7\%$	NVIDIA RTX 3090 GPU and Pytorch
		Xie et al. [126]	Jiang's dataset [23] and LSOTB-TIR [153]	$Precision = 92.12\%$, $Success = 66.66\%$, $Accuracy = 67.7\%$, $FPS = 12.4$	Intel Core i9-9940X@3.30 GHz, 4× NVIDIA RTX 2080Ti, Python 3.7 and PyTorch 1.10
	Transformer-Based	Tong et al. [131]	Collect and organize data on anti-UAV competitions	$Precision = 88.39\%$	Intel i9-13900K, GeForce RTX 4090 GPU
		Yu et al. [22]	1st and 2nd Anti-UAV [46]	1st TestDEV: $AUC = 77.9\%$, $Precision = 98.0\%$; 2st TestDEV: $AUC = 72.4\%$, $Precision = 93.4\%$	4× NVIDIA RTX 3090 GPU, Python 3.6 and Pytorch 1.7.1
	YOLO-Based	Hu et al. [141]	280 testing images, self-built	$Precision = 89\%$, $mAP = 37.41\%$, $FPS = 56.3$	Intel Xeon E5-2630 v4, NVIDIA GeForce GTX 1080 Ti, 64-bit Ubuntu 16.04 operating system
		Zhou et al. [92]	FL-Drone [154]	$mAP = 98\%$	NVIDIA RTX3090
		Fardad et al. [143]	116,608 images from [44,155]	$mAP = 98\%$, $Recall = 96\%$	4× Tesla V100-SXM2 graphic cards
		Vedanshu et al. [41]	1874 images from Kaggle [42] and self-built	$mAP = 96.7\%$, $Precision = 95\%$, $Recall = 95.6\%$	Not provide
		Li et al. [149]	3rd Anti-UAV [46]	$acc = 49.61\%$	Not provide
		Fang et al. [150]	Det-Fly [44]	$Precision = 0.914$, $Recall = 91.9\%$	NVIDIA A40

4.3.2. Results of the Methods

From Tables 3 and 4, we can see that early anti-UAV detection and tracking methods [120,149,156] need the support of sensors. The advantage of Sensor-Based methods is that they are not affected by field of view occlusion or changes in target scale However, the effectiveness of these methods is susceptible to a variety of factors, including electromagnetic interference, environmental noise, detection distance.

Recent Vision-Based approaches have opened up new paths from a visual perspective, complementing the shortcomings of earlier Sensor-Based methods. Although each classification method is sorted by publication time, it is difficult to evaluate all the methods because these methods use different datasets. From [31,48,120] we get Vision-Based approaches tend to provide higher $F1$ score in anti-UAV approaches compared to methods that rely on sensors. Vision-Based methods can capture rich information about the environment and the target, including shape, color, which helps to improve the accuracy of target detection and recognition. In contrast, the information provided by sensors, such as RF or acoustic, may be limited.In addition, YOLO-Based methods [141] have higher inference speed compared to Siamese-Based methods [120,126]. In the pursuit of real-time object tracking, YOLO-Based methods are preferred.

Table 4. Analysis of the advantages and disadvantages of the classified methods.

Category	Methods	Advantages	Disadvantages
Sensor-Based	RF-Based	Long distance monitoring No need for line of sight	The communication protocol of UAV may undergo periodic changes Many UAV can dynamically switch communication frequencies Prone to interference from devices like WiFi and signal towers
	Acoustic-Based	Low cost Multiple acoustic nodes monitoring	Susceptible to interference from environmental noise Short sound propagation distance limits the listening range.
	Multi-Sensors-Based	High accuracy Has good adaptability to complex environments	High model complexity and large computational load Multimodal data fusion is challenging
Vision-Based	Siamese-Based	Suitable for real-time single target tracking Low computational cost Fast inference speed	Brief occlusion can cause Siamese networks to lose track Sensitive to scale changes
	Transformer-Based	Excel in processing complex scenes Robust against challenging scenarios	Not sensitive to sparse small UAV features Large computational load, insufficient real-time performance
	YOLO-Based	Emphasize speed and real-time capabilities Suitable for rapid detection or tracking Balance between inference speed and accuracy	Sensitive to obstruction by obstacles Easy to be affected by weather conditions

5. Discussions

5.1. Discussions on the Limitations of Datasets

Anti-UAV detection and tracking datasets are crucial foundation for training machine learning models and evaluating the performance of anti-UAV technologies. However, existing public datasets have some limitations.

- Lack of multi-target datasets: In the publicly available datasets we have collected, most are focused on single-target tracking, with minimal interference from surrounding objects (such as birds, balloons, or other flying objects). There is a lack of datasets related to multi-target tracking in complex scenarios. This limitation creates a significant gap between research and actual application needs. With the rapid development of small drone technology, scenarios involving multiple drones working in coordination are expected to become increasingly common in the future. However, existing datasets are insufficient to fully support research and development in these complex scenarios.
- Low resolution and quality: Some datasets suffer from low-resolution images and videos, which can hinder the development and evaluation of high-precision detection and tracking algorithms, especially when identifying small or distant drones. The illumination can be a factor to impact the appearance of drones.
- Scenarios: Although some datasets have a large number of images, many images have similar scenes and mainly focus on specific conditions or environments, such as urban or rural environments, day or night scenes. It limits the generalizability of evaluation techniques in the real world.
- UAV types: Existing datasets may only include a few types of drones, whereas, in reality, there is a wide variety of drones with different appearances, sizes, and flight characteristics. Recently, many bionic drones are being produced and appear in many scenes. It has stronger concealment.

5.2. Discussions on the Limitations of Methods

In the task of anti-UAV, there are some limitations for current methods.

- Insufficiency of uniform assessment rules. Although many methods utilized their own metrics for experimental comparison, it still lacks of uniform assessment rules. Meanwhile, considering many approaches are designed and implemented under different running environments, it faces challenges to provide a fair evaluation of methods.
- Uncertainty of the model's size. Quite a number of literatures do not provide the models' size. This is important for real-time applications because the light-weight models are more likely to be deployed in embedded devices.
- Difficult to achieve the trade-off between performance and accuracy: For rapidly flying UAV, it has large requirements for accurate detection in short time. The algorithms

that achieve high performance always with high complexity, that is, it needs more computation resources.
- Insufficient generalization ability: Due to the lack of diversity in the dataset, the network model may overfit to specific scenarios during training, resulting in insufficient generalization ability. This means that performance may decrease when the model is applied to a new and different environment. Especially in the continuous day-to-night monitoring scenarios, it is difficult for one model to cover all day's surveillance.
- The detection and tracking of UAV swarms remain underdeveloped: Current technologies face significant challenges in handling multi-target recognition, trajectory prediction, and addressing occlusions and interferences in complex environments. Particularly, during UAV swarm flights, the close distance between individuals, along with varying attitudes and speeds, makes it difficult for traditional detection algorithms to maintain efficient and accurate recognition.

5.3. Future Research Directions

The development of anti-UAV is an important means to address the growing threat of UAV, especially in areas such as security monitoring, military defense, and civil aviation. The future research directions for anti-UAV detection and tracking may include the following five aspects.

- Image super-resolution reconstruction: In infrared scenarios, anti-UAV systems often operate at long distances [157] where the image resolution is not only very low but also often encountering many artifacts. Super-resolution techniques enable the recovery of additional details from low-resolution images, making the appearance, shape, and size of UAV clearer. When the drone moves quickly or away from the camera, super-resolution technology can help restore lost image details and maintain tracking continuity. However, image super-resolution usually requires significant computing resources, and algorithms need to be optimized to balance computational efficiency and image quality. Therefore, image super-resolution reconstruction can be considered as a critical technology for small UAV target detection and tracking.
- Autonomous learning capability: As UAV technology becomes increasingly intelligent, UAV can autonomously take countermeasures when detecting interference during flight. For example, they might change communication protocols, switch transmission frequencies to avoid being intercepted, or even dynamically adjust flight strategies. This advancement in intelligence imposes higher demands on anti-UAV detection and tracking algorithms. To effectively address these challenges, anti-UAV detection and tracking algorithms need to possess autonomous learning capabilities and be able to make real-time decisions, thereby adapting to and countering the evolving intelligent behaviors of UAV. This not only requires algorithms to be highly flexible and adaptive but also to maintain effective tracking and countermeasure capabilities in complex environments.
- Integration of multimodal perception techniques: In Section 3, we have discussed in detail the advantages and disadvantages of Sensor-Based and Vision-Based methods. However, these two approaches can play complementary roles in anti-UAV technology. While Sensor-Based methods may be affected by environmental noise, Vision-Based methods can provide additional information to compensate for these interferences. Therefore, combining these two approaches can significantly enhance UAV detection and identification capabilities under various environments and conditions. Although Section 3 has discussed Multi-Sensors-Based methods, these approaches have primarily been explored at the experimental stage, indicating considerable room for improvement in practical applications.
- Countering multi-agent collaborative operations: With the continuous advancement of drone technology, the trends of increasing intelligence and reducing costs

are becoming more evident, leading to more frequent scenarios where multiple intelligent UAV work collaboratively. This collaborative operation mode offers significant advantages in complex tasks; however, it also presents new challenges for anti-drone technology. Existing detection and tracking algorithms may perform well against single targets, but when faced with multiple intelligent UAV operating collaboratively, they may experience decreased accuracy, target loss, and other issues. Therefore, developing anti-drone technologies that can effectively counter multi-agent collaborative operations has become a critical direction for current technological development.

- Anti-interference capability: In practical applications, anti-UAV systems need not only to detect and track UAVs but also to possess strong anti-interference capabilities. It is crucial to accurately distinguish between similar objects such as birds, kites, and balloons, thereby significantly enhancing anti-interference performance and ensuring stable operation in various complex environments.

6. Conclusions

In this paper, we conduct a systematic review of Vision-Based anti-UAV detection and tracking methods. This paper summarizes the mainstream methods for anti-UAV detection tasks, and sheds light on the advances in time. It aims to providing the help to address the issues of unauthorized or excessive UAV flights. A detailed analysis of all the surveyed papers is presented to provide readers a good understanding of the advances in the anti-UAV field.

This review mainly covers several aspects, namely, the datasets, the methods, and the future research. Firstly, nine anti-UAV datasets are collected, coupled with visual images, detailed descriptions and the access links. The datasets are discussed from multi-target, low resolution and quality, scenarios, and UAV types. Readers can have a good understanding of the related datasets and find what they want quickly.

Secondly, we review a series of anti-UAV methods. We propose a hierarchical representation to classify them into Sensor-Based and Vision-Based. Representative approaches are illustrated to show the main thoughts and characteristics. The advantages and disadvantages are also summarized. It can be concluded that Vision-Based methods have become the hot topic. Three categories, Siamese-Based, Transformer-Based, and YOLO-Based, are being attached more importance.

Finally, we highlight five future directions. Researchers and engineers around the world could easily grasp the recent advances of Vision-Based anti-UAV methods and find valuable data for practical applications. This can foster their efficiency largely. It is expected that this review can provide technique support for the applications such as security monitoring, military defense, and civil aviation.

Author Contributions: All authors contributed to the idea for the article. B.W. provided the initial idea and framework regarding this manuscript.; The literature search and analysis were performed by Q.L. and Q.M.; writing, B.W., Q.L., Q.M. and J.W.; investigation, C.L.P.C., A.S. and H.Z.; review and editing, J.W., A.S. and H.Z. All authors participated in the commenting and modification of the paper. All authors have read and agreed to the published version of the manuscript.

Funding: This work is supported by the Basic Research Programs of Taicang, 2023, under number TC2023JC23, and the Open Fund of National Engineering Laboratory for Big Data System Computing Technology, under number SZU-BDSC-OF2024-08. This work is funded in part by the National Natural Science Foundation of China under number 62301558, 92267203, and Youth Innovation Promotion Association Project of Chinese Academy of Sciences under number 2022405. This work is also supported in part by the Science and Technology Major Project of Guangzhou under number 202007030006, and in part by the Program for Guangdong Introducing Innovative and Entrepreneurial Teams under number 2019ZT08X214.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

UAV	Unmanned Aerial Vehicles
RF	Radio Frequency
SVM	Support Vector Machine
CNN	Convolutional Neural Networks
KNN	k-Nearest Neighbors
SNR	Signal to Noise Ratio
YOLO	You Only Look Once
STFT	Short-Time Fourier Transform

References

1. Fan, J.; Yang, X.; Lu, R.; Xie, X.; Li, W. Design and implementation of intelligent inspection and alarm flight system for epidemic prevention. *Drones* **2021**, *5*, 68. [CrossRef]
2. Filkin, T.; Sliusar, N.; Ritzkowski, M.; Huber-Humer, M. Unmanned aerial vehicles for operational monitoring of landfills. *Drones* **2021**, *5*, 125. [CrossRef]
3. McEnroe, P.; Wang, S.; Liyanage, M. A survey on the convergence of edge computing and AI for UAVs: Opportunities and challenges. *IEEE Internet Things J.* **2022**, *9*, 15435–15459. [CrossRef]
4. Wang, Z.; Cao, Z.; Xie, J.; Zhang, W.; He, Z. RF-based drone detection enhancement via a Generalized denoising and interference-removal framework. *IEEE Signal Process. Lett.* **2024**, *31*, 929–933. [CrossRef]
5. Zhou, T.; Xin, B.; Zheng, J.; Zhang, G.; Wang, B. Vehicle detection based on YOLOv7 for drone aerial visible and infrared images. In Proceedings of the 2024 6th International Conference on Image Processing and Machine Vision (IPMV), Macau, China, 12–14 January 2024; pp. 30–35.
6. Shen, Y.; Pan, Z.; Liu, N.; You, X. Performance analysis of legitimate UAV surveillance system with suspicious relay and anti-surveillance technology. *Digit. Commun. Netw.* **2022**, *8*, 853–863. [CrossRef]
7. Lin, N.; Tang, H.; Zhao, L.; Wan, S.; Hawbani, A.; Guizani, M. A PDDQNLP algorithm for energy efficient computation offloading in UAV-assisted MEC. *IEEE Trans. Wirel. Commun.* **2023**, *22*, 8876–8890. [CrossRef]
8. Cheng, N.; Wu, S.; Wang, X.; Yin, Z.; Li, C.; Chen, W.; Chen, F. AI for UAV-assisted IoT applications: A comprehensive review. *IEEE Internet Things J.* **2023**, *10*, 14438–14461. [CrossRef]
9. Zhang, Y.; Zhang, H.; Gao, X.; Zhang, S.; Yang, C. UAV target detection for IoT via enhancing ERP component by brain–computer interface system. *IEEE Internet Things J.* **2023**, *10*, 17243–17253. [CrossRef]
10. Wang, B.; Li, C.; Zou, W.; Zheng, Q. Foreign object detection network for transmission lines from unmanned aerial vehicle images. *Drones* **2024**, *8*, 361. [CrossRef]
11. Saha, B.; Kunze, S.; Poeschl, R. Comparative study of deep learning model architectures for drone detection and classification. In Proceedings of the 2024 IEEE International Mediterranean Conference on Communications and Networking (MeditCom), Madrid, Spain, 12 August 2024; pp. 167–172.
12. Khawaja, W.; Semkin, V.; Ratyal, N.I.; Yaqoob, Q.; Gul, J.; Guvenc, I. Threats from and countermeasures for unmanned aerial and underwater vehicles. *Sensors* **2022**, *22*, 3896. [CrossRef]
13. Zamri, F.N.M.; Gunawan, T.S.; Yusoff, S.H.; Alzahrani, A.A.; Bramantoro, A.; Kartiwi, M. Enhanced small drone detection using optimized YOLOv8 with attention mechanisms. *IEEE Access* **2024**, *12*, 90629–90643. [CrossRef]
14. Elsayed, M.; Reda, M.; Mashaly, A.S.; Amein, A.S. LERFNet: An enlarged effective receptive field backbone network for enhancing visual drone detection. *Vis. Comput.* **2024**, *40*, 1–14. [CrossRef]
15. Kunze, S.; Saha, B. Long short-term memory model for drone detection and classification. In Proceedings of the 2024 4th URSI Atlantic Radio Science Meeting (AT-RASC), Meloneras, Spain, 19–24 May 2024; pp. 1–4.
16. Li, Y.; Fu, M.; Sun, H.; Deng, Z.; Zhang, Y. Radar-based UAV swarm surveillance based on a two-stage wave path difference estimation method. *IEEE Sens. J.* **2022**, *22*, 4268–4280. [CrossRef]
17. Xiao, J.; Chee, J.H.; Feroskhan, M. Real-time multi-drone detection and tracking for pursuit-evasion with parameter search. *IEEE Trans. Intell. Veh.* **2024**, 1–11. [CrossRef]
18. Deng, A.; Han, G.; Zhang, Z.; Chen, D.; Ma, T.; Liu, Z. Cross-parallel attention and efficient match transformer for aerial tracking. *Remote Sens.* **2024**, *16*, 961. [CrossRef]
19. Nguyen, D.D.; Nguyen, D.T.; Le, M.T.; Nguyen, Q.C. FPGA-SoC implementation of YOLOv4 for flying-object detection. *J. Real-Time Image Process.* **2024**, *21*, 63. [CrossRef]
20. Zhao, J.; Zhang, J.; Li, D.; Wang, D. Vision-based anti-uav detection and tracking. *IEEE Trans. Intell. Transp. Syst.* **2022**, *23*, 25323–25334. [CrossRef]
21. Sun, Y.; Li, J.; Wang, L.; Xv, J.; Liu, Y. Deep Learning-based drone acoustic event detection system for microphone arrays. *Multimed. Tools Appl.* **2024**, *83*, 47865–47887. [CrossRef]
22. Yu, Q.; Ma, Y.; He, J.; Yang, D.; Zhang, T. A unified transformer based tracker for anti-uav tracking. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 17–24 June 2023; pp. 3036–3046.

23. Jiang, N.; Wang, K.; Peng, X.; Yu, X.; Wang, Q.; Xing, J.; Li, G.; Guo, G.; Ye, Q.; Jiao, J.; et al. Anti-uav: A large-scale benchmark for vision-based uav tracking. *IEEE Trans. Multimed.* **2021**, *25*, 486–500. [CrossRef]
24. Wang, C.; Wang, T.; Wang, E.; Sun, E.; Luo, Z. Flying small target detection for anti-UAV based on a Gaussian mixture model in a compressive sensing domain. *Sensors* **2019**, *19*, 2168. [CrossRef]
25. Sheu, B.H.; Chiu, C.C.; Lu, W.T.; Huang, C.I.; Chen, W.P. Development of UAV tracing and coordinate detection method using a dual-axis rotary platform for an anti-UAV system. *Appl. Sci.* **2019**, *9*, 2583. [CrossRef]
26. Fang, H.; Li, Z.; Wang, X.; Chang, Y.; Yan, L.; Liao, Z. Differentiated attention guided network over hierarchical and aggregated features for intelligent UAV surveillance. *IEEE Trans. Ind. Inform.* **2023**, *19*, 9909–9920. [CrossRef]
27. Zhu, X.F.; Xu, T.; Zhao, J.; Liu, J.W.; Wang, K.; Wang, G.; Li, J.; Zhang, Z.; Wang, Q.; Jin, L.; et al. Evidential detection and tracking collaboration: New problem, benchmark and algorithm for robust anti-uav system. *arXiv* **2023**, arXiv:2306.15767.
28. Zhao, R.; Li, T.; Li, Y.; Ruan, Y.; Zhang, R. Anchor-free multi-UAV Detection and classification using spectrogram. *IEEE Internet Things J.* **2023**, *11*, 5259–5272. [CrossRef]
29. Zheng, J.; Chen, R.; Yang, T.; Liu, X.; Liu, H.; Su, T.; Wan, L. An efficient strategy for accurate detection and localization of UAV swarms. *IEEE Internet Things J.* **2021**, *8*, 15372–15381. [CrossRef]
30. Yan, X.; Fu, T.; Lin, H.; Xuan, F.; Huang, Y.; Cao, Y.; Hu, H.; Liu, P. UAV detection and tracking in urban environments using passive sensors: A survey. *Appl. Sci.* **2023**, *13*, 11320. [CrossRef]
31. Svanström, F.; Englund, C.; Alonso-Fernandez, F. Real-time drone detection and tracking with visible, thermal and acoustic sensors. In Proceedings of the 2020 25th International Conference on Pattern Recognition (ICPR), Milan, Italy, 10–15 January 2021; pp. 7265–7272.
32. Wang, C.; Shi, Z.; Meng, L.; Wang, J.; Wang, T.; Gao, Q.; Wang, E. Anti-occlusion UAV tracking algorithm with a low-altitude complex background by integrating attention mechanism. *Drones* **2022**, *6*, 149. [CrossRef]
33. Sun, L.; Zhang, J.; Yang, Z.; Fan, B. A motion-aware siamese framework for unmanned aerial vehicle tracking. *Drones* **2023**, *7*, 153. [CrossRef]
34. Li, S.; Gao, J.; Li, L.; Wang, G.; Wang, Y.; Yang, X. Dual-branch approach for tracking UAVs with the infrared and inverted infrared image. In Proceedings of the 2022 7th International Conference on Intelligent Computing and Signal Processing (ICSP), Xi'an, China, 15–17 April 2022; pp. 1803–1806.
35. Zhao, J.; Zhang, X.; Zhang, P. A unified approach for tracking UAVs in infrared. In Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW), Montreal, BC, Canada, 11–17 October 2021; pp. 1213–1222.
36. Zhang, J.; Lin, Y.; Zhou, X.; Shi, P.; Zhu, X.; Zeng, D. Precision in pursuit: A multi-consistency joint approach for infrared anti-UAV tracking. *Vis. Comput.* **2024**, *40*, 1–16. [CrossRef]
37. Ojdanić, D.; Naverschnigg, C.; Sinn, A.; Zelinskyi, D.; Schitter, G. Parallel architecture for low latency UAV detection and tracking using robotic telescopes. *IEEE Trans. Aerosp. Electron. Syst.* **2024**, *60*, 5515–5524. [CrossRef]
38. Soni, V.; Shah, D.; Joshi, J.; Gite, S.; Pradhan, B.; Alamri, A. Introducing AOD 4: A dataset for air borne object detection. *Data Brief* **2024**, *56*, 110801. [CrossRef] [PubMed]
39. Yuan, S.; Yang, Y.; Nguyen, T.H.; Nguyen, T.M.; Yang, J.; Liu, F.; Li, J.; Wang, H.; Xie, L. MMAUD: A comprehensive multi-modal anti-UAV dataset for modern miniature drone threats. *arXiv* **2024**, arXiv:2402.03706.
40. Svanström, F.; Alonso-Fernandez, F.; Englund, C. A dataset for multi-sensor drone detection. *Data Brief* **2021**, *39*, 107521. [CrossRef] [PubMed]
41. Dewangan, V.; Saxena, A.; Thakur, R.; Tripathi, S. Application of image processing techniques for uav detection using deep learning and distance-wise analysis. *Drones* **2023**, *7*, 174. [CrossRef]
42. Kaggle. Available online: https://www.kaggle.com/datasets/dasmehdixtr/drone-dataset-uav (accessed on 15 March 2024).
43. Drone Detection. Available online: https://github.com/creiser/drone-detection (accessed on 11 April 2024).
44. Zheng, Y.; Chen, Z.; Lv, D.; Li, Z.; Lan, Z.; Zhao, S. Air-to-air visual detection of micro-uavs: An experimental evaluation of deep learning. *IEEE Robot. Autom. Lett.* **2021**, *6*, 1020–1027. [CrossRef]
45. Walter, V.; Vrba, M.; Saska, M. On training datasets for machine learning-based visual relative localization of micro-scale UAVs. In Proceedings of the 2020 IEEE International Conference on Robotics and Automation (ICRA), Paris, France, 31 May–31 August 2020; pp. 10674–10680.
46. 3rd-Anti-UAV. Available online: https://anti-uav.github.io/ (accessed on 15 March 2024).
47. Xiao, Y.; Zhang, X. Micro-UAV detection and identification based on radio frequency signature. In Proceedings of the 2019 6th International Conference on Systems and Informatics (ICSAI), Shanghai, China, 2–4 November 2019; pp. 1056–1062.
48. Yang, B.; Matson, E.T.; Smith, A.H.; Dietz, J.E.; Gallagher, J.C. UAV detection system with multiple acoustic nodes using machine learning models. In Proceedings of the 2019 Third IEEE International Conference on Robotic Computing (IRC), Naples, Italy, 25–27 February 2019; pp. 493–498.
49. Rahman, M.H.; Sejan, M.A.S.; Aziz, M.A.; Tabassum, R.; Baik, J.I.; Song, H.K. A comprehensive survey of unmanned aerial vehicles detection and classification using machine learning approach: Challenges, solutions, and future directions. *Remote Sens.* **2024**, *16*, 879. [CrossRef]
50. Moore, E.G. Radar detection, tracking and identification for UAV sense and avoid applications. Master's Thesis, University of Denver, Denver, CO, USA, 2019; p. 13808039.

51. Fang, H.; Ding, L.; Wang, L.; Chang, Y.; Yan, L.; Han, J. Infrared small UAV target detection based on depthwise separable residual dense network and multiscale feature fusion. *IEEE Trans. Instrum. Meas.* **2022**, *71*, 1–20. [CrossRef]
52. Yan, B.; Peng, H.; Wu, K.; Wang, D.; Fu, J.; Lu, H. Lighttrack: Finding lightweight neural networks for object tracking via one-shot architecture search. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 20–25 June 2021; pp. 15180–15189.
53. Blatter, P.; Kanakis, M.; Danelljan, M.; Van Gool, L. Efficient visual tracking with exemplar transformers. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), Waikoloa, HI, USA, 2–7 January 2023; pp. 1571–1581.
54. Cheng, Q.; Wang, Y.; He, W.; Bai, Y. Lightweight air-to-air unmanned aerial vehicle target detection model. *Sci. Rep.* **2024**, *14*, 1–18. [CrossRef]
55. Zhao, J.; Li, J.; Jin, L.; Chu, J.; Zhang, Z.; Wang, J.; Xia, J.; Wang, K.; Liu, Y.; Gulshad, S.; et al. The 3rd C workshop & challenge: Methods and results. *arXiv* **2023**, arXiv:2305.07290.
56. Chen, C.P.; Li, H.; Wei, Y.; Xia, T.; Tang, Y.Y. A local contrast method for small infrared target detection. *IEEE Trans. Geosci. Remote Sens.* **2013**, *52*, 574–581. [CrossRef]
57. Huang, L.; Zhao, X.; Huang, K. Globaltrack: A simple and strong baseline for long-term tracking. *AAAI Conf. Artif. Intell.* **2020**, *34*, 11037–11044. [CrossRef]
58. Yan, B.; Peng, H.; Fu, J.; Wang, D.; Lu, H. Learning spatio-temporal transformer for visual tracking. In Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, QC, Canada, 10–17 October 2021; pp. 10428–10437.
59. Deng, T.; Zhou, Y.; Wu, W.; Li, M.; Huang, J.; Liu, S.; Song, Y.; Zuo, H.; Wang, Y.; Yue, Y.; et al. Multi-modal UAV detection, classification and tracking algorithm–technical report for CVPR 2024 UG2 challenge. *arXiv* **2024**, arXiv:2405.16464.
60. Svanström, F.; Alonso-Fernandez, F.; Englund, C. Drone detection and tracking in real-time by fusion of different sensing modalities. *Drones* **2022**, *6*, 317. [CrossRef]
61. You, J.; Ye, Z.; Gu, J.; Pu, J. UAV-Pose: A dual capture network algorithm for low altitude UAV attitude detection and tracking. *IEEE Access* **2023**, *11*, 129144–129155. [CrossRef]
62. Schlack, T.; Pawlowski, L.; Ashok, A. Hybrid event and frame-based system for target detection, tracking, and identification. *Unconv. Imaging Sens. Adapt. Opt.* **2023**, *12693*, 303–311.
63. Song, H.; Wu, Y.; Zhou, G. Design of bio-inspired binocular UAV detection system based on improved STC algorithm of scale transformation and occlusion detection. *Int. J. Micro Air Veh.* **2021**, *13*, 17568293211004846. [CrossRef]
64. Huang, B.; Dou, Z.; Chen, J.; Li, J.; Shen, N.; Wang, Y.; Xu, T. Searching region-free and template-free siamese network for tracking drones in TIR videos. *IEEE Trans. Geosci. Remote Sens.* **2023**, *62*, 5000315. [CrossRef]
65. Wang, C.; Meng, L.; Gao, Q.; Wang, J.; Wang, T.; Liu, X.; Du, F.; Wang, L.; Wang, E. A lightweight UAV swarm detection method integrated attention mechanism. *Drones* **2022**, *7*, 13. [CrossRef]
66. Lee, H.; Cho, S.; Shin, H.; Kim, S.; Shim, D.H. Small airborne object recognition with image processing for feature extraction. *Int. J. Aeronaut. Space Sci.* **2024**, *25*, 1–15. [CrossRef]
67. Lee, E. Drone classification with motion and appearance feature using convolutional neural networks. Master's Thesis, Purdue University, West Lafayette, IN, USA, 2020; p. 30503377.
68. Yin, X.; Jin, R.; Lin, D. Efficient air-to-air drone detection with composite multi-dimensional attention. In Proceedings of the 2024 IEEE 18th International Conference on Control & Automation (ICCA), Reykjavík, Iceland, 18–21 June 2024; pp. 725–730.
69. Ghazlane, Y.; Gmira, M.; Medromi, H. Development Of a vision-based anti-drone identification friend or foe model to recognize birds and drones using deep learning. *Appl. Artif. Intell.* **2024**, *38*, 2318672. [CrossRef]
70. Swinney, C.J.; Woods, J.C. Low-cost raspberry-pi-based UAS detection and classification system using machine learning. *Aerospace* **2022**, *9*, 738. [CrossRef]
71. Lu, S.; Wang, W.; Zhang, M.; Li, B.; Han, Y.; Sun, D. Detect the video recording act of UAV through spectrum recognition. In Proceedings of the 2022 IEEE International Conference on Artificial Intelligence and Computer Applications (ICAICA), Dalian, China, 24–26 June 2022; pp. 559–564.
72. He, Z.; Huang, J.; Qian, G. UAV detection and identification based on radio frequency using transfer learning. In Proceedings of the 2022 IEEE 8th International Conference on Computer and Communications (ICCC), Chengdu, China, 9–12 December 2022; pp. 1812–1817.
73. Cai, Z.; Wang, Y.; Jiang, Q.; Gui, G.; Sha, J. Toward intelligent lightweight and efficient UAV identification with RF fingerprinting. *IEEE Internet Things J.* **2024**, *11*, 26329–26339. [CrossRef]
74. Uddin, Z.; Qamar, A.; Alharbi, A.G.; Orakzai, F.A.; Ahmad, A. Detection of multiple drones in a time-varying scenario using acoustic signals. *Sustainability* **2022**, *14*, 4041. [CrossRef]
75. Guo, J.; Ahmad, I.; Chang, K. Classification, positioning, and tracking of drones by HMM using acoustic circular microphone array beamforming. *EURASIP J. Wirel. Commun. Netw.* **2020**, *2020*, 1–19. [CrossRef]
76. Shi, X.; Yang, C.; Xie, W.; Liang, C.; Shi, Z.; Chen, J. Anti-drone system with multiple surveillance technologies: Architecture, implementation, and challenges. *IEEE Commun. Mag.* **2018**, *56*, 68–74. [CrossRef]
77. Koulouris, C.; Dimitrios, P.; Al-Darraji, I.; Tsaramirsis, G.; Tamimi, H. A comparative study of unauthorized drone detection techniques. In Proceedings of the 2023 9th International Conference on Information Technology Trends (ITT), Dubai, United Arab Emirates, 24–25 May 2023; pp. 32–37.

78. Xing, Z.; Hu, S.; Ding, R.; Yan, T.; Xiong, X.; Wei, X. Multi-sensor dynamic scheduling for defending UAV swarms with Fresnel zone under complex terrain. *ISA Trans.* **2024**, *153*, 57–69. [CrossRef]
79. Chen, H. A benchmark with multi-sensor fusion for UAV detection and distance estimation. Master's Thesis, State University of New York at Buffalo, Getzville, NY, USA, 2022.
80. Li, S. Applying multi agent system to track uav movement. Master's Thesis, Purdue University, West Lafayette, IN, USA, 2019.
81. Samaras, S.; Diamantidou, E.; Ataloglou, D.; Sakellariou, N.; Vafeiadis, A.; Magoulianitis, V.; Lalas, A.; Dimou, A.; Zarpalas, D.; Votis, K.; et al. Deep learning on multi sensor data for counter UAV applications—A systematic review. *Sensors* **2019**, *19*, 4837. [CrossRef]
82. Jouaber, S.; Bonnabel, S.; Velasco-Forero, S.; Pilte, M. Nnakf: A neural network adapted kalman filter for target tracking. In Proceedings of the 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Toronto, ON, Canada, 6–11 June 2021; pp. 4075–4079.
83. Xie, W.; Wan, Y.; Wu, G.; Li, Y.; Zhou, F.; Wu, Q. A RF-visual directional fusion framework for precise UAV positioning. *IEEE Internet Things J.* **2024**, 1. [CrossRef]
84. Cao, B.; Yao, H.; Zhu, P.; Hu, Q. Visible and clear: Finding tiny objects in difference map. *arXiv* **2024**, arXiv:2405.11276.
85. Singh, P.; Gupta, K.; Jain, A.K.; Jain, A.; Jain, A. Vision-based UAV detection in complex backgrounds and rainy conditions. In Proceedings of the 2024 2nd International Conference on Disruptive Technologies (ICDT), Greater Noida, India, 15–16 March 2024; pp. 1097–1102.
86. Wu, T.; Duan, H.; Zeng, Z. Biological eagle eye-based correlation filter learning for fast UAV tracking. *IEEE Trans. Instrum. Meas.* **2024**, *73*, 7506412. [CrossRef]
87. Zhou, Y.; Jiang, Y.; Yang, Z.; Li, X.; Sun, W.; Zhen, H.; Wang, Y. UAV image detection based on multi-scale spatial attention mechanism with hybrid dilated convolution. In Proceedings of the 2024 3rd International Conference on Image Processing and Media Computing (ICIPMC), Hefei, China, 17–19 May 2024; pp. 279–284.
88. Wang, G.; Yang, X.; Li, L.; Gao, K.; Gao, J.; Zhang, J.y.; Xing, D.j.; Wang, Y.z. Tiny drone object detection in videos guided by the bio-inspired magnocellular computation model. *Appl. Soft Comput.* **2024**, *163*, 111892. [CrossRef]
89. Munir, A.; Siddiqui, A.J.; Anwar, S. Investigation of UAV detection in images with complex backgrounds and rainy artifacts. In Proceedings of the 2024 IEEE/CVF Winter Conference on Applications of Computer Vision Workshops (WACVW), Waikoloa, HI, USA, 1–6 January 2024; pp. 232–241.
90. Unlu, E.; Zenou, E.; Riviere, N.; Dupouy, P.E. Deep learning-based strategies for the detection and tracking of drones using several cameras. *IPSJ Trans. Comput. Vis. Appl.* **2019**, *11*, 1–13. [CrossRef]
91. Al-lQubaydhi, N.; Alenezi, A.; Alanazi, T.; Senyor, A.; Alanezi, N.; Alotaibi, B.; Alotaibi, M.; Razaque, A.; Hariri, S. Deep learning for unmanned aerial vehicles detection: A review. *Comput. Sci. Rev.* **2024**, *51*, 100614. [CrossRef]
92. Zhou, X.; Yang, G.; Chen, Y.; Li, L.; Chen, B.M. VDTNet: A high-performance visual network for detecting and tracking of intruding drones. *IEEE Trans. Intell. Transp. Syst.* **2024**, 25, 9828–9839. [CrossRef]
93. Kassab, M.; Zitar, R.A.; Barbaresco, F.; Seghrouchni, A.E.F. Drone detection with improved precision in traditional machine learning and less complexity in single shot detectors. *IEEE Trans. Aerosp. Electron. Syst.* **2024**, *60*, 3847–3859. [CrossRef]
94. Zhang, Z.; Jin, L.; Li, S.; Xia, J.; Wang, J.; Li, Z.; Zhu, Z.; Yang, W.; Zhang, P.; Zhao, J.; et al. Modality meets long-term tracker: A siamese dual fusion framework for tracking UAV. In Proceedings of the 2023 IEEE International Conference on Image Processing (ICIP), Kuala Lumpur, Malaysia, 8–11 October 2023; pp. 1975–1979.
95. Xing, D. UAV surveillance with deep learning using visual, thermal and acoustic sensors. Ph.D. Thesis, New York University Tandon School of Engineering, Brooklyn, NY, USA, 2023.
96. Xie, W.; Zhang, Y.; Hui, T.; Zhang, J.; Lei, J.; Li, Y. FoRA: Low-rank adaptation model beyond multimodal siamese network. *arXiv* **2024**, arXiv:2407.16129.
97. Elleuch, I.; Pourranjbar, A.; Kaddoum, G. Leveraging transformer models for anti-jamming in heavily attacked UAV environments. *IEEE Open J. Commun. Soc.* **2024**, 5, 5337–5347. [CrossRef]
98. Rebbapragada, S.V.; Panda, P.; Balasubramanian, V.N. C2FDrone: Coarse-to-fine drone-to-drone detection using vision transformer networks. *arXiv* **2024**, arXiv:2404.19276.
99. Zeng, H.; Li, J.; Qu, L. Lightweight low-altitude UAV object detection based on improved YOLOv5s. *Int. J. Adv. Network, Monit. Control.* **2024**, *9*, 87–99. [CrossRef]
100. AlDosari, K.; Osman, A.; Elharrouss, O.; Al-Maadeed, S.; Chaari, M.Z. Drone-type-set: Drone types detection benchmark for drone detection and tracking. In Proceedings of the 2024 International Conference on Intelligent Systems and Computer Vision (ISCV), Fez, Morocco, 8–10 May 2024; pp. 1–7.
101. Bo, C.; Wei, Y.; Wang, X.; Shi, Z.; Xiao, Y. Vision-based anti-UAV detection based on YOLOv7-GS in complex backgrounds. *Drones* **2024**, *8*, 331. [CrossRef]
102. Wang, C.; Meng, L.; Gao, Q.; Wang, T.; Wang, J.; Wang, L. A target sensing and visual tracking method for countering unmanned aerial vehicle swarm. *IEEE Sens. J.* **2024**, 1. [CrossRef]
103. Sun, S.; Mo, B.; Xu, J.; Li, D.; Zhao, J.; Han, S. Multi-YOLOv8: An infrared moving small object detection model based on YOLOv8 for air vehicle. *Neurocomputing* **2024**, *588*, 127685. [CrossRef]
104. He, X.; Fan, K.; Xu, Z. Uav identification based on improved YOLOv7 under foggy condition. *Signal Image Video Process.* **2024**, *18*, 6173–6183. [CrossRef]

105. Fang, H.; Wu, C.; Wang, X.; Zhou, F.; Chang, Y.; Yan, L. Online infrared UAV target tracking with enhanced context-awareness and pixel-wise attention modulation. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 5005417. [CrossRef]
106. Jiang, W.; Pan, H.; Wang, Y.; Li, Y.; Lin, Y.; Bi, F. A multi-level cross-attention image registration method for visible and infrared small unmanned aerial vehicle targets via image style transfer. *Remote Sens.* **2024**, *16*, 2880. [CrossRef]
107. Noor, A.; Li, K.; Tovar, E.; Zhang, P.; Wei, B. Fusion flow-enhanced graph pooling residual networks for unmanned aerial vehicles surveillance in day and night dual visions. *Eng. Appl. Artif. Intell.* **2024**, *136*, 108959. [CrossRef]
108. Nair, A.K.; Sahoo, J.; Raj, E.D. A lightweight FL-based UAV detection model using thermal images. In Proceedings of the 7th International Conference on Networking, Intelligent Systems and Security (NISS), Meknes, Morocco, 18–19 April 2024; pp. 1–5.
109. Xu, B.; Hou, R.; Bei, J.; Ren, T.; Wu, G. Jointly modeling association and motion cues for robust infrared UAV tracking. *Vis. Comput.* **2024**, *40*, 1–12. [CrossRef]
110. Li, Q.; Mao, Q.; Liu, W.; Wang, J.; Wang, W.; Wang, B. Local information guided global integration for infrared small target detection. In Proceedings of the ICASSP 2024—2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Seoul, Republic of Korea, 14–19 April 2024; pp. 4425–4429.
111. Li, B.; Wu, W.; Wang, Q.; Zhang, F.; Xing, J.; Yan, J. Siamrpn++: Evolution of siamese visual tracking with very deep networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019; pp. 4282–4291.
112. Hu, W.; Wang, Q.; Zhang, L.; Bertinetto, L.; Torr, P.H. Siammask: A framework for fast online object tracking and segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2023**, *45*, 3072–3089.
113. Xu, Y.; Wang, Z.; Li, Z.; Yuan, Y.; Yu, G. Siamfc++: Towards robust and accurate visual tracking with target estimation guidelines. *AAAI Conf. Artif. Intell.* **2020**, *34*, 12549–12556. [CrossRef]
114. Danelljan, M.; Bhat, G.; Khan, F.S.; Felsberg, M. Atom: Accurate tracking by overlap maximization. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019; pp. 4660–4669.
115. Chen, J.; Huang, B.; Li, J.; Wang, Y.; Ren, M.; Xu, T. Learning spatio-temporal attention based siamese network for tracking UAVs in the wild. *Remote Sens.* **2022**, *14*, 1797. [CrossRef]
116. Yin, W.; Ye, Z.; Peng, Y.; Liu, W. A review of visible single target tracking based on Siamese networks. In Proceedings of the 2023 4th International Conference on Electronic Communication and Artificial Intelligence (ICECAI), Guangzhou, China, 12–14 May 2023; pp. 282–289.
117. Feng, M.; Su, J. RGBT tracking: A comprehensive review. *Inf. Fusion* **2024**, *110*, 102492. [CrossRef]
118. Bertinetto, L.; Valmadre, J.; Henriques, J.F.; Vedaldi, A.; Torr, P.H. Fully-convolutional siamese networks for object tracking. In Proceedings of the Computer Vision—ECCV 2016 Workshops, Amsterdam, The Netherlands, 8–16 October 2016; pp. 850–865.
119. Huang, B.; Chen, J.; Xu, T.; Wang, Y.; Jiang, S.; Wang, Y.; Wang, L.; Li, J. Siamsta: Spatio-temporal attention based siamese tracker for tracking uavs. In Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW), Montreal, BC, Canada, 11–17 October 2021; pp. 1204–1212.
120. Fang, H.; Wang, X.; Liao, Z.; Chang, Y.; Yan, L. A real-time anti-distractor infrared UAV tracker with channel feature refinement module. In Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision Workshops, Montreal, BC, Canada, 11–17 October 2021; pp. 1240–1248.
121. Huang, B.; Li, J.; Chen, J.; Wang, G.; Zhao, J.; Xu, T. Anti-uav410: A thermal infrared benchmark and customized scheme for tracking drones in the wild. *IEEE Trans. Pattern Anal. Mach. Intell.* **2023**, *46*, 2852–2865. [CrossRef] [PubMed]
122. Shi, X.; Zhang, Y.; Shi, Z.; Zhang, Y. Gasiam: Graph attention based siamese tracker for infrared anti-uav. In Proceedings of the 2022 3rd International Conference on Computer Vision, Image and Deep Learning & International Conference on Computer Engineering and Applications (CVIDL & ICCEA), Changchun, China, 20–22 May 2022; pp. 986–993.
123. Voigtlaender, P.; Luiten, J.; Torr, P.H.; Leibe, B. Siam r-cnn: Visual tracking by re-detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 6578–6588.
124. Sandler, M.; Howard, A.; Zhu, M.; Zhmoginov, A.; Chen, L.C. Mobilenetv2: Inverted residuals and linear bottlenecks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 4510–4520.
125. Cheng, F.; Liang, Z.; Peng, G.; Liu, S.; Li, S.; Ji, M. An anti-UAV long-term tracking method with hybrid attention mechanism and hierarchical discriminator. *Sensors* **2022**, *22*, 3701. [CrossRef] [PubMed]
126. Xie, X.; Xi, J.; Yang, X.; Lu, R.; Xia, W. Stftrack: Spatio-temporal-focused siamese network for infrared uav tracking. *Drones* **2023**, *7*, 296. [CrossRef]
127. Zhang, Z.; Lu, X.; Cao, G.; Yang, Y.; Jiao, L.; Liu, F. ViT-YOLO: Transformer-based YOLO for object detection. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, BC, Canada, 11–17 October 2021; pp. 2799–2808.
128. Gao, S.; Zhou, C.; Ma, C.; Wang, X.; Yuan, J. Aiatrack: Attention in attention for transformer visual tracking. In Proceedings of the European Conference on Computer Vision, Tel Aviv, Israel, 23–27 October 2022; pp. 146–164.
129. Mayer, C.; Danelljan, M.; Bhat, G.; Paul, M.; Paudel, D.P.; Yu, F.; Van Gool, L. Transforming model prediction for tracking. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, Louisiana, USA, 19–24 June 2022; pp. 8731–8740.

130. Lin, L.; Fan, H.; Zhang, Z.; Xu, Y.; Ling, H. Swintrack: A simple and strong baseline for transformer tracking. *Adv. Neural Inf. Process. Syst.* **2022**, *35*, 16743–16754.
131. Tong, X.; Zuo, Z.; Su, S.; Wei, J.; Sun, X.; Wu, P.; Zhao, Z. ST-Trans: Spatial-temporal transformer for infrared small target detection in sequential images. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 5001819. [CrossRef]
132. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You only look once: Unified, real-time object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788.
133. Stefenon, S.F.; Singh, G.; Souza, B.J.; Freire, R.Z.; Yow, K.C. Optimized hybrid YOLOu-Quasi-ProtoPNet for insulators classification. *IET Gener. Transm. Distrib.* **2023**, *17*, 3501–3511. [CrossRef]
134. Souza, B.J.; Stefenon, S.F.; Singh, G.; Freire, R.Z. Hybrid-YOLO for classification of insulators defects in transmission lines based on UAV. *Int. J. Electr. Power Energy Syst.* **2023**, *148*, 108982. [CrossRef]
135. Stefenon, S.F.; Seman, L.O.; Klaar, A.C.R.; Ovejero, R.G.; Leithardt, V.R.Q. Hypertuned-YOLO for interpretable distribution power grid fault location based on EigenCAM. *Ain Shams Eng. J.* **2024**, *15*, 102722. [CrossRef]
136. Xiao, H.; Wang, B.; Zheng, J.; Liu, L.; Chen, C.P. A fine-grained detector of face mask wearing status based on improved YOLOX. *IEEE Trans. Artif. Intell.* **2023**, *5*, 1816–1830. [CrossRef]
137. Ajakwe, S.O.; Ihekoronye, V.U.; Kim, D.S.; Lee, J.M. DRONET: Multi-tasking framework for real-time industrial facility aerial surveillance and safety. *Drones* **2022**, *6*, 46. [CrossRef]
138. Wang, J.; Hongjun, W.; Liu, J.; Zhou, R.; Chen, C.; Liu, C. Fast and accurate detection of UAV objects based on mobile-YOLO network. In Proceedings of the 2022 14th International Conference on Wireless Communications and Signal Processing (WCSP), Nanjing, China, 1–3 November 2022; pp. 01–05.
139. Cheng, Q.; Li, J.; Du, J.; Li, S. Anti-UAV detection method based on local-global feature focusing module. In Proceedings of the 2024 IEEE 7th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC), Chongqing, China, 15–17 March 2024; Volume 7, pp. 1413–1418.
140. Tu, X.; Zhang, C.; Zhuang, H.; Liu, S.; Li, R. Fast drone detection with optimized feature capture and modeling algorithms. *IEEE Access* **2024**, 12, 108374–108388. [CrossRef]
141. Hu, Y.; Wu, X.; Zheng, G.; Liu, X. Object detection of UAV for anti-UAV based on improved YOLOv3. In Proceedings of the 2019 Chinese Control Conference (CCC), Guangzhou, China, 27–30 July 2019; pp. 8386–8390.
142. Redmon, J.; Farhadi, A. YOLOv3: An incremental improvement. *arXiv* **2018**, arXiv:1804.02767.
143. Dadboud, F.; Patel, V.; Mehta, V.; Bolic, M.; Mantegh, I. Single-stage uav detection and classification with YOLOv5: Mosaic data augmentation and panet. In Proceedings of the 2021 17th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), Washington, DC, USA, 16–19 November 2021; pp. 1–8.
144. Jocher, G.; Stoken, A.; Borovec, J.; Chaurasia, A.; Changyu, L.; Hogan, A.; Hajek, J.; Diaconu, L.; Kwon, Y.; Defretin, Y.; et al. ultralytics/yolov5: V5. 0-YOLOv5-P6 1280 models, AWS, Supervise. ly and YouTube integrations. *Zenodo* **2021**.
145. Liu, S.; Qi, L.; Qin, H.; Shi, J.; Jia, J. Path aggregation network for instance segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 8759–8768.
146. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster r-cnn: Towards real-time object detection with region proposal networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 1137–1149. [CrossRef]
147. Lin, T.Y.; Dollár, P.; Girshick, R.; He, K.; Hariharan, B.; Belongie, S. Feature pyramid networks for object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 2117–2125.
148. Wang, C.Y.; Bochkovskiy, A.; Liao, H.Y.M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 18–22 June 2023; pp. 7464–7475.
149. Li, Y.; Yuan, D.; Sun, M.; Wang, H.; Liu, X.; Liu, J. A global-local tracking framework driven by both motion and appearance for infrared anti-uav. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 17–24 June 2023; pp. 3026–3035.
150. Fang, A.; Feng, S.; Liang, B.; Jiang, J. Real-time detection of unauthorized unmanned aerial vehicles using SEB-YOLOv8s. *Sensors* **2024**, *24*, 3915. [CrossRef]
151. Wang, C.Y.; Yeh, I.H.; Liao, H.Y.M. YOLOv9: Learning what you want to learn using programmable gradient information. *arXiv* **2024**, arXiv:2402.13616.
152. Wang, C.; He, W.; Nie, Y.; Guo, J.; Liu, C.; Wang, Y.; Han, K. Gold-YOLO: Efficient object detector via gather-and-distribute mechanism. *arXiv* **2024**, arXiv:2309.11331.
153. Liu, Q.; Li, X.; He, Z.; Li, C.; Li, J.; Zhou, Z.; Yuan, D.; Li, J.; Yang, K.; Fan, N.; et al. LSOTB-TIR: A large-scale high-diversity thermal infrared object tracking benchmark. In Proceedings of the 28th ACM International Conference on Multimedia, Seattle, WA, USA, 12–16 October 2020; pp. 3847–3856.
154. Rozantsev, A.; Lepetit, V.; Fua, P. Detecting flying objects using a single moving camera. *IEEE Trans. Pattern Anal. Mach. Intell.* **2016**, *39*, 879–892. [CrossRef]
155. Coluccia, A.; Fascista, A.; Schumann, A.; Sommer, L.; Dimou, A.; Zarpalas, D.; Méndez, M.; De la Iglesia, D.; González, I.; Mercier, J.P.; et al. Drone vs. bird detection: Deep learning algorithms and results from a grand challenge. *Sensors* **2021**, *21*, 2824. [CrossRef]

156. Coluccia, A.; Fascista, A.; Schumann, A.; Sommer, L.; Dimou, A.; Zarpalas, D.; Akyon, F.C.; Eryuksel, O.; Ozfuttu, K.A.; Altinuc, S.O.; et al. Drone-vs-bird detection challenge at IEEE AVSS2021. In Proceedings of the 2021 17th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), Washington, DC, USA, 16–19 November 2021; pp. 1–8.
157. Xi, Y.; Zhou, Z.; Jiang, Y.; Zhang, L.; Li, Y.; Wang, Z.; Tan, F.; Hou, Q. Infrared moving small target detection based on spatial-temporal local contrast under slow-moving cloud background. *Infrared Phys. Technol.* **2023**, *134*, 104877. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.