applied sciences

MDPI

Review

Target Detection and Recognition for Traffic Congestion in Smart Cities Using Deep Learning-Enabled UAVs: A Review and Analysis

Sundas Iftikhar [1,†], Muhammad Asim [2,3,*,†], Zuping Zhang [1,†], Ammar Muthanna [4,5,†], Junhong Chen [3,6,†], Mohammed El-Affendi [2,†], Ahmed Sedik [7,8,†] and Ahmed A. Abd El-Latif [2,9,*,†]

1 School of Computer Science and Engineering, Central South University, Changsha 410083, China
2 EIAS Data Science Lab, College of Computer and Information Sciences, Prince Sultan University, Riyadh 11586, Saudi Arabia
3 School of Computer Science and Technology, Guangdong University of Technology, Guangzhou 510006, China
4 Department of Applied Probability and Informatics, Peoples' Friendship University of Russia (RUDN University), Miklukho-Maklaya, 117198 Moscow, Russia
5 Department of Telecommunication Networks and Data Transmission, The Bonch-Bruevich Saint-Petersburg State University of Telecommunications, 193232 Saint Petersburg, Russia
6 Expertise Centre for Digital Media, Hasselt University, 3500 Hasselt, Belgium
7 Smart Systems Engineering Laboratory, College of Engineering, Prince Sultan University, Riyadh 11586, Saudi Arabia
8 Department of the Robotics and Intelligent Machines, Kafrelsheikh University, Kafrelsheikh 33511, Egypt
9 Department of Mathematics and Computer Science, Faculty of Science, Menoufia University, Shebin El-Koom 32511, Egypt
* Correspondence: asimpk@gdut.edu.cn (M.A.); aabdellatif@psu.edu.sa (A.A.A.E.-L.)
† These authors contributed equally to this work.

check for updates

Citation: Iftikhar, S.; Asim, M.; Zhang, Z.; Muthanna, A.; Chen, J.; El-Affendi, M.; Sedik, A.; Abd El-Latif, A.A. Target Detection and Recognition for Traffic Congestion in Smart Cities Using Deep Learning-Enabled UAVs: A Review and Analysis. *Appl. Sci.* **2023**, *13*, 3995. https://doi.org/10.3390/app13063995

Academic Editors: Muhammad Babar, Saleem Iqbal and Aftab Khan

Received: 30 January 2023
Revised: 22 February 2023
Accepted: 1 March 2023
Published: 21 March 2023

Abstract: In smart cities, target detection is one of the major issues in order to avoid traffic congestion. It is also one of the key topics for military, traffic, civilian, sports, and numerous other applications. In daily life, target detection is one of the challenging and serious tasks in traffic congestion due to various factors such as background motion, small recipient size, unclear object characteristics, and drastic occlusion. For target examination, unmanned aerial vehicles (UAVs) are becoming an engaging solution due to their mobility, low cost, wide field of view, accessibility of trained manipulators, a low threat to people's lives, and ease to use. Because of these benefits along with good tracking effectiveness and resolution, UAVs have received much attention in transportation technology for tracking and analyzing targets. However, objects in UAV images are usually small, so after a neural estimation, a large quantity of detailed knowledge about the objects may be missed, which results in a deficient performance of actual recognition models. To tackle these issues, many deep learning (DL)-based approaches have been proposed. In this review paper, we study an end-to-end target detection paradigm based on different DL approaches, which includes one-stage and two-stage detectors from UAV images to observe the target in traffic congestion under complex circumstances. Moreover, we also analyze the evaluation work to enhance the accuracy, reduce the computational cost, and optimize the design. Furthermore, we also provided the comparison and differences of various technologies for target detection followed by future research trends.

Keywords: unmanned aerial vehicles; target detection; traffic congestion; deep learning; YOLO versions; faster R-CNN; cascade R-CNN

1. Introduction

In smart cities, the intelligent transportation network has gained much attention in computer vision in order to avoid traffic congestion and accidents. Traffic congestion occurs when the number of traffic increases and the speed of the object becomes slow. It causes

several disadvantages such as fuel consumption, time loss, mental pressure, and aggravates air pollution. Accurate and timely target detection in intelligent transportation networks may reduce traffic congestion. Target detection is utilized to find out where targets are located in the area and which group they are associated with such as pedestrians [1], vehicles [2], etc. Many researchers tried to handle the target detection problem in various fields such as solving the target detection in the form of pedestrians in autonomous vehicles under different sub-issues such as occlusion, large and small-size objects, and complex backgrounds with the use of various strategies [3,4]. However, in this review paper, we study target detection through UAVs.

UAVs, which are also called "Drones" or "Unmanned Aircraft Systems", are an airship that works automatically through remote controls and sensors without any human aviator, crew, or rider. There are four major types of UAVs as shown in Figure 1. Among multi-rotor drones/UAVs, the quadcopter is much more famous due to its approachability and virtuous camera command as compared to other types of drones/UAVs, it is easy to use and can work in a restricted region.

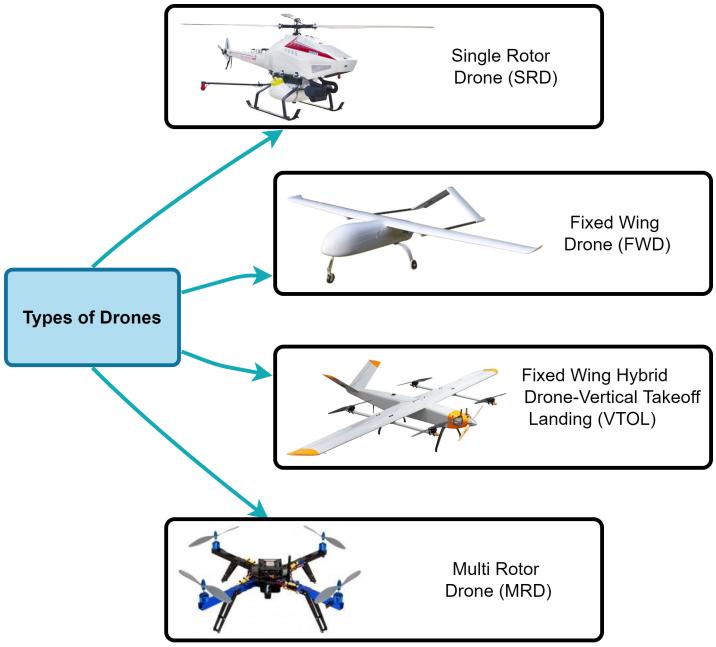

Figure 1. Types of UAVs/drones.

UAVs have been integrated in different fields, for example, computer vision [5], search and rescue operations [6], and communication systems [7–9]. UAVs play an important role in military and non-military applications such as surveillance, filmmaking, attack, cultivation, scientific analysis, cargo transport, and many more due to low cost, mobility, ease to use, and accessibility [10]. The advantages of drones/UAVs lie in rescuing time and investment, boosting the reliability of fact measurement, increasing the security of data logging, improving the performance of complex scenarios, and making the investigation more systematic. Despite all its advantages, UAVs also face several significant challenges in terms of full deployment, which are given below:

- In terms of heavy mobility, UAVs have insufficient ability to mark traffic motion evaluation at the intersection due to their dimensions, underneath stimulation, and braking capabilities.
- In dense urban regions, UAV delivery can be impractical due to a limited utmost payload (mass and volume), comparably low-range, insufficient low-elevation airspace, insubstantial sensor's ability, and limited battery capacity, which can cause difficulties and security risks.

- It has insufficient auto-navigation expertise, transmission, and energy problems to detect different kinds of disasters such as floods, fire situations, and air tragedies. It has insufficient facts about patrolling regions and is unable to examine and track the informant of pollution.
- Moreover, target detection in the form of pedestrians, vehicles, and traffic signs is one of the critical issues of UAVs in traffic congestion due to complicated environments, small and heavy recipient size, drastic occlusion, weather and lightning variations, and unclear object characteristics such as appearance and color as demonstrated in Figure 2.

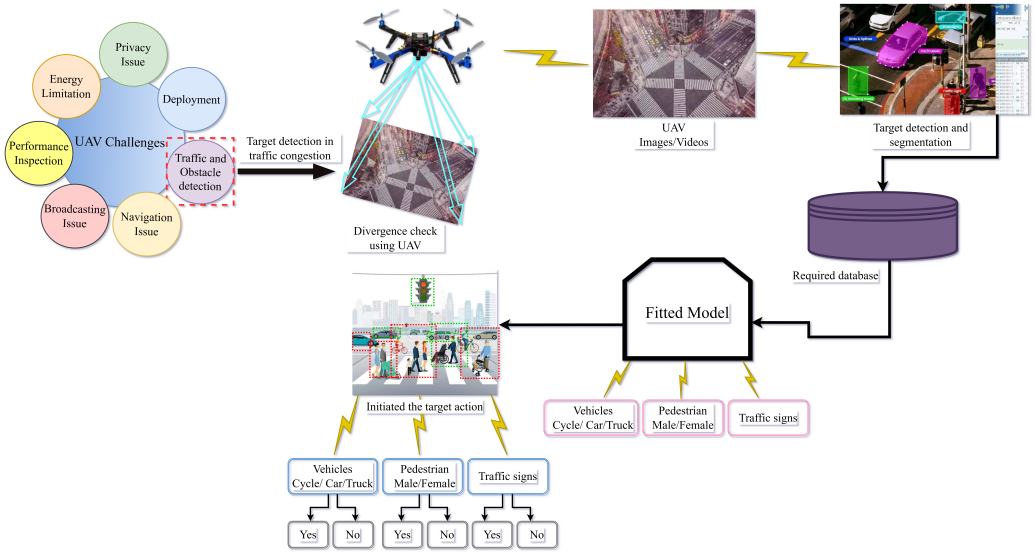

Figure 2. Typical challenges of UAVs with basic framework of target detection in traffic jams through UAV.

To overcome the above-mentioned issues in UAV-based target detection, researchers have proposed many DL-based approaches. Recent DL approaches in the area of target detection undergoing UAVs have progressed rapidly by virtue of enormous performance. DL is the class of artificial intelligence and machine learning and has the expertise to gather, learn, and analyze a massive quantity of data. DL uses both graph and transfiguration techniques to create multi-level learning models.

In this article, we review the DL approach to address problems of target detection in traffic congestion. In the recent years, many researchers have tried to sort out this issue with the help of different approaches. To handle the tracking and detection issues of the moving target, there have been several tasks based on UAV-initiated cameras using traditional approaches [11,12]. In 2016, Zhang et al. [13] presented the pixel-based adaptive segmenter algorithm for target detection. In [14,15], fast fourier transform and kernel function were employed on a discriminative correlation filter-based detective to perform the complex computation in the frequency domain rather than in the spatial domain which optimized and enhanced the performance of the detecting model. In 2018, the Kanade–Lucas–Tomasi tracking approach was presented by Ke et al. for target detection tracking [16]. However, traditional approaches are less precise due to bad generalization accomplishment.

For vigorous tracking achievement, DeepSort was used which further omitted the DL properties and Kalman filter [17]. Apart from that, in image depth feature screening, CNN unifies the feature abstraction, selection, and categorization which are superior than traditional techniques. Its effectiveness is higher, so the scalability and correctness of the target detection are also much better than traditional techniques.

Furthermore, different object detection approaches have been presented based on CNN, for instance, Region-based Convolutional Neural Network (RCNN) and Single Shot MultiBox Detector (SSD) to perceive the small targets and perform at low elevations. However, their performances are still not good enough in target detection in terms of accuracy [18–20]. In contrast, for better performance, the single-stage detector techniques such as YOLO v3 and YOLO v4 were presented by Redmon and Boboc [21] and Alexey et al. [22], which are better in target detection and permit excellent real-time achievement and high precision. But it prerequisites a notable computational scheme. Besides this, in [21,23,24], UAV-based detection and tracking methods have been presented, which merge the illusion, KCF (kernel correlation filter) based detectors, and YOLOv3. Moreover, many review books and articles have been found to detect targets in traffic congestion. For example, Butilă [25] presented a review on the applications of UAVs associated with traffic attention. Srivastava et al. [26] studied important parameters to foreground the small-size issues. Osco et al. [27] reviewed both UAV remote sensing images and DL in a review paper. Alzahrani et al. [28] studied a review based on an extant UAV-assisted system. Kanistras et al. [29] monitored the traffic based on the UAV system, while Outay et al. [30] reviewed the vision processing techniques. Park et al. [31] reviewed the information and communication technology approaches based on DL application to analyze the target in traffic in real-time using UAVs. Zhang et al. [32] presented correlation filtering and other tracking algorithms to solve the target occlusion problem in a review paper. From the above body of work, it is worth noting that all the above-mentioned works consider either only UAV-based target detection or one specific kind of DL for target detection. None of the above-mentioned reviews considered applications of DL and UAVs in target detection at the same time. The purpose of this review paper is to discuss different issues rather than one specific issue using different methods in contrast to other review papers and provide the future direction related to more preferable methods to handle those issues in the upcoming years. It also presents a brief overview of DL-based approaches for UAV-based target detection.

The main contribution of this review paper is to further address the target detection problem under various circumstances with the help of DL approaches instead of exploring one specific DL technique for target detection entirely and inspect the best execution performance model built on current research. The statistics of this paper are taken from different sources, for instance, conference proceedings, journals, and workshops to offer readers a glimpse of the common relationship between DL techniques and target detection through UAVs in traffic congestion at an advance level.

The rest of the paper is organized as follows: Section 2 narrates the concept and research tasks affiliated with UAVs classifying the area of target detection in traffic congestion along with major metrics and crucial issues. Next, Section 3 introduces DL approaches used for handling crucial issues in target detection. Section 4 inspects the existing implementation of DL techniques in target detection. Section 5 puts forward a comprehensive discussion on utilization of DL techniques in target detection and outlines the limitations to denominate future research trends. Finally, Section 6 concludes this review paper. For more clarity, the organization of the paper is presented in Figure 3.

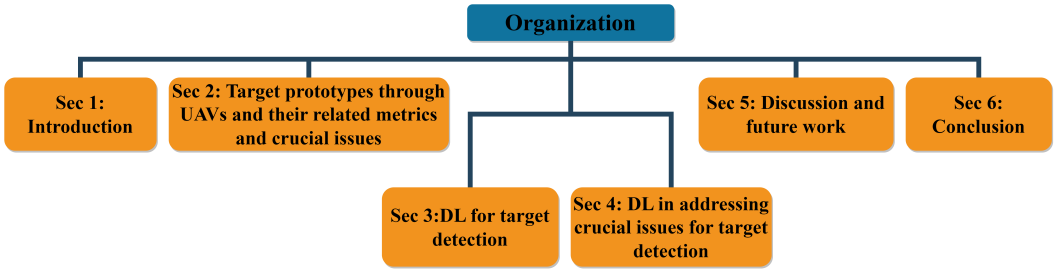

Figure 3. Paper Organization.

2. Target Detection Prototypes and Their Related Metrics and Crucial Issues

This section reviews research tasks that use UAV videos and images to improve target detection in traffic congestion and management. The following subsections give detailed studies based on different methods and metrics.

A. *Target Detection Prototypes*

The dominant features of UAVs in target detection are elaborated in Table 1.

Table 1. UAV features in target detection.

Features	Merit
Human Workload	Moderate
Financial Resources	Prevent
Mobility	High
Efficiency	High
Battery Timing	Low
Service Access	Multiple Access
Validity	High
Speed of Detection	High
Frequency	Low
Adaptability	Centralized
Safety	Less Secure
Resolution	High
Application Range	Vast
Transportability	Flexible
Mobility	High/Strong

1. *Mobile UAV Trajectories Based on Road Traffic Monitoring Approach:* Within a town, due to the insufficient amount of UAVs, it is not appropriate to mark a comprehensive tracking of all the targets in terms of vehicles, pedestrians, and motorbikes. To handle this issue, a mobile UAV-based road traffic monitoring system was proposed by Elloumi et al. [33] in 2019.
    - Advantages: This mode aims to track the detection rate and the statistics of restrained vehicles to overcome accidents and overspeeding. Moreover, it initiates UAV trajectories to observe targets within a town from a long distance to avoid traffic congestion.
    - Limitation: It is necessary to enhance the detection rate of an isolated vehicle of a congested event at a low speed by sharing the facts related to dispersion through UAVs.
2. *DeepSort Approach:* For electric vehicles and target detection in metropolitan surroundings, a DeepSORT approach based on DL has been formulated by Liu and Zhang [34] in 2021. This approach combines YOLOv4, various tracking methods, and fuses the target detection web to reduce the state estimation of the pursued target in the non-uniform tactic and realized the target position through UAVs.
    - Advantages: The combination of the presented model aims to significantly enhance the performance, robustness, and positioning of numerous targets' perception and tracking in complicated metropolitan surroundings.
    - Limitation: Still, the performance is deficient when the UAVs are fluttering at a high elevation, which may cause the problem in detecting the tiny bulk of the ground substance.

3. *SAHER System Approach Based on UAVs:* In traffic congestion, road misadventures are caused due to energy and coverage issues. To handle this issue, Ali et al. [35] proposed a SAHER system based on UAVs using the 5G data processing in 2020.
    - Advantages: In real-time scenarios, this approach detects swiftness and alternative traffic interruptions to overcome the number of crashes.
    - Limitation: The proportion of tragedies and wounds is still extremely high.
4. *Traditional and UAVs Vision Approach:* In 2022, Cheng et al. [36] presented a model based on traditional and UAV approaches which mainly include YOLO 3, Mean-Shift, Gaussian background difference approach, and Kalman filter algorithms to observe the unauthorized behavior of the target.
    - Advantages: The aim of this approach was to compare the results of UAVs and traditional approaches on the four features: manual time, divergence results, recognition speed, and accuracy. Therein, the target detection based on the UAV approach performs better as compared to the traditional approach. This is because, in the traditional approach, the computational cost is low and poor in robustness. Furthermore, it cannot fulfill the substantial application demands in real-time detection. The comparison results of UAVs and traditional approaches based on four features are demonstrated in Figure 4.
    - Limitation: Still, some conflicts arise. If the target detection results are not enough, then it will create errors in the detection outcomes and may affect future observations.

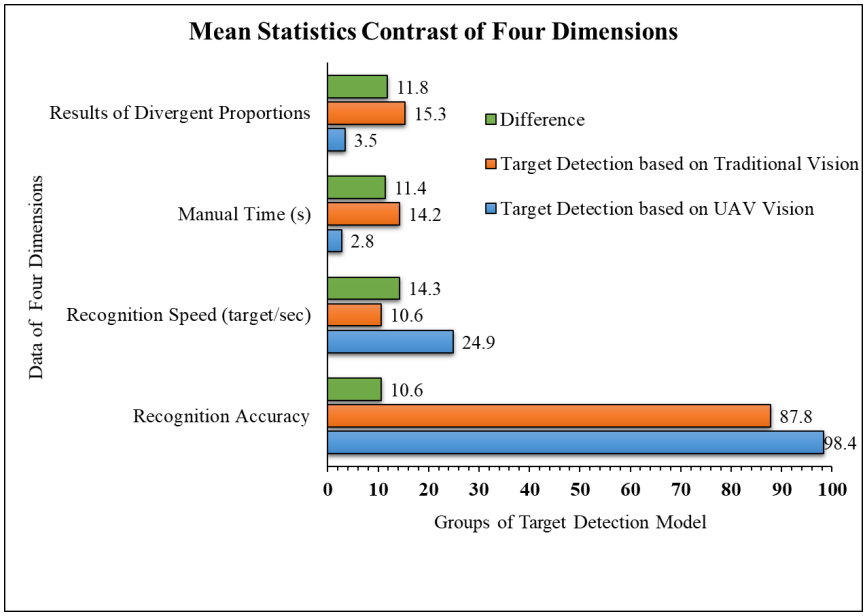

Figure 4. Average comparison of target detection based on UAVs and traditional vision.

B. *Metrics*

In the recent years, various researchers concentrated on the miscellaneous features of UAVs for target detection, which include cost, safety, privacy, etc. A few dominant metrics are shown in Figure 5 and reviewed below.

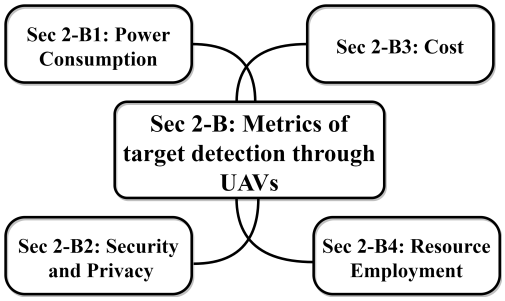

Figure 5. Metrics of target detection through UAVs.

1. *Power Consumption:* For UAVs, it is difficult to detect small targets under different weather conditions due to the restricted battery ability. Although it is not possible to extend the size of the battery because doing this will extend the mass of the UAVs which is one more critical consideration. Many research reviews have described the battery demand for the UAVs through the wireless power transfer (WPT) approach [37,38]. Through an expressive connection of WPT, the charging of UAVs can be performed to increase the flight time and dimension for the target inspection, observation, and other surveillance assignments. Moreover, it can also overcome the many restrictions of the current inspection approaches, for instance, costly tasks and dangerous functions. In [39], the researchers charged the battery of the drone and carried out the testing of presented schemes for various distances and unbalanced cartography with the help of magnetic resonance coupling WPT. Besides this, machine learning and DL algorithms for drones also provide a better solution in terms of energy consumption for data collection and processing in compute-hungry realms [40].
2. *Security and Privacy:* Security and privacy are important key factors while detecting targets through UAVs. This is because sometimes the attackers liberate all the accumulated information of UAVs through scareware, viruses, and keystrokes with the help of computer streaming-assigned software. Hacking all the data will lead to false detection of traffic congestion, convey corrupt statistics which misguide the ground command stations, and may be used in illegal activities, for instance, using the stolen data against military action. To save the system from such kind of hacking and lay out the correct information about the required recipient, a model was presented named "Privacy by Design" in [41] which provides a solution for security and privacy violations. Recently, blockchain, machine learning, DL, and watermarking have performed an important role to secure UAV applications. These approaches aim to supply reliable, safe, and accurate information and secure programmatically updated facts. Further details of these approaches are presented in the research article [42].
3. *Cost:* To identify the targets, it is compulsory to develop low-cost UAV detectors. Light mass and low cost are the features of UAVs to ratify the quality inspection with extreme temporal and multidimensional resolutions without endangering the lives of humans. In the recent years, traditional approaches have been used to detect the targets through UAVs, but traditional approaches such as scale-invariant feature transform features [43] have some drawbacks such as high computational cost, bulky deployment, and unpredictable risks and cannot identify the targets in real-time scenarios. In contrast to traditional approaches, DL approaches, such as R-FCN [44], have considerable computational cost and meet real-time demands. However, there is a requirement to better stabilize the dummy complexity and perception reliability and to validate the dummy with unified data from multiple sources [45].
4. *Resource Employment:* For humans, it is exhausting for the crew to find the targets only from aerial graphics which consumes extensive human resources and time. The technology of intelligent automatic target recognition and monitoring can empower UAVs to become more competent in rescue, tragedy response, stuff transportation, target crushing, enemy research, and other tasks, which can tremendously decrease the consumption of mankind's resources and additionally stimulate the evolution

of the UAVs' domain. Moreover, the implementation is done based on the different algorithms concerning resource utilization and time complexity such as the start of Faster R-CNN which consumes extreme computing resources. Therefore, in 2016, Jifeng Da et al. [44] presented an R-FCN approach to overcome this issue. Traditional standard approaches for screening graphic content will lead to omissions and authorization miscalculations. Therefore, it is impossible to entirely depend on the resources of humans for capturing, displaying, and processing large-scale video statistics. To gather large-scale video statistics in real-time through UAVs, they can be screened by DL [46] and big data algorithms to modify the traditional approaches of target detection from a weak standard manner to a smart real-time structured one and produce useful facts for the users which fulfill their demands, save manpower and data resources, overcome the cost of monitoring, and upgrade the efficiency.

C. *Issues in Target Detection*

To tackle the target detection challenges and upgrade the above-stated metrics through UAVs, it is necessary to sort out the following problems.

1. *Small-Size Objects:* In real applications, the height of the shot is high, the size of the target is much smaller than that of the image, and the target has defective properties; then, the target suffers a particular degree of distortion overwhelmed by the angle of the shot and the correlative movement between the UAVs and target leads to a target which is substantially changing in the background, etc. Besides this, some datasets such as MS COCO characterized small targets due to limited discrete features which remit the missing and several false target detections [47].
2. *Target Occlusion:* Target occlusion will occur due to target blockage and the effect of illumination surrounding which sorely validates the tracking and identification of targets. Further, false and missed detection is also caused due to occlusion. Many researchers tried to solve the target occlusion issue with the help of different methods based on various occlusion conditions. Some of the progressed methods based on DL are discussed in the next section.
3. *Joint Issues:* An increased number of executions are multi-perspective. For example, some jobs are concurrently time-sensitive and have extensive resources, such as multi-scale appearance, spot, missed recipient and victim recognition [48,49], and enhancement of realistic applications [50]. Scientists began to tackle various combined problems in target detection which are called joint issues.

3. DL for Target Detection

DL has currently demonstrated excellent results in solving numerous robotic functions in the area of awareness, planning, segmentation, and management. It has an excellent ability to learn characterization from composite data obtained in the real-world context which makes it ideal for several types of autonomous applications. We have concentrated on three categories of DL approaches utilized in target detection for traffic congestion, i.e., one-stage detectors and two-stage detectors which are demonstrated in Figure 6.

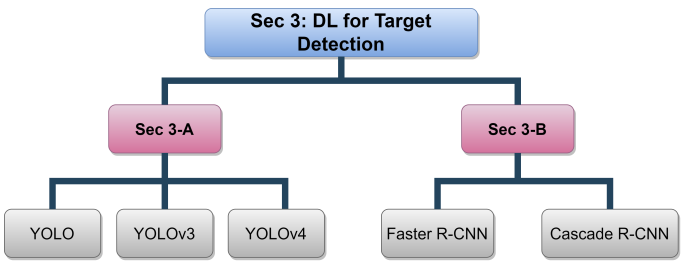

Figure 6. The main prototype of the DL family utilized for target detection.

A. *One-Stage Detectors*

The one-stage detector is also known as a regression-based approach which directly computes the object's correlatives and the class probability and then produces the outcomes after an isolated detection and enormously increases the detection speed. Some frequently used one-stage detectors for target detection through UAVs are stated below.

- YOLO was introduced by Joseph Redmon [21] in 2015. The major aim of this approach is to detect small objects and compute their fast speed. Through an artificial neural network ANN, this algorithm takes out the image attributes and then utilizes the regression algorithm to execute the image detection effect. With the help of a neural network, it can instantly extract the classification and locality of the bounding box. As a backbone, Darknet-19 and GooleNet is used in the training network, while confidence loss is utilized as a loss function. The grid segment is answerable for target detection. This algorithm has vigorous generalization capabilities because it can understand the highly versatile features to delegate to other regions.
- YOLOv3 was presented by Joseph Redmon and Farhadi [21] in 2018 which is the updated version of YOLO. As a backbone, this version uses the Darknet53 classifier and utilizes a multi-scale indicator. Feature extraction is carried out with the help of Darknet-53. There are 53 convolutional layers to train ImageNet. The feature bounding is downward-sampled because convolutional layers are a two-step process and, at three various dimensions, it executes detection. Meanwhile, to verify the normalization of the input in intense layers with the help of convolutional layers, batch normalization is illustrated. In contrast to Darknet-19, Darknet53 shows superior accuracy. Besides this, to overcome the over-fitting, Leaky RELU can be utilized. Through extra convolutional layers, this version seized the low measure feature which boosts the small targets and other issues as well as enhanced its speed. Moreover, by contrasting the prediction outcomes with the actual merit of the sample class, the loss value is gained, and the framework variables are updated with the help of a back propagation design to get the boost network prototype for target detection. The loss function is the aggregate of three distinct losses which are: (i) confidence loss, (ii) classification loss, and (iii) location regression loss as shown in Equation (1).

$$L = \lambda_1 Loss_{conf} + \lambda_2 Loss_{cla} + \lambda_3 Loss_{loc} \quad (1)$$

- YOLOv4 is the latest version of the YOLO group, which was presented by Alexey et al. [22] in April 2020. This version is the improved version of YOLOv3 and is more marvelous than YOLOv3. This model is categorized into three parts: (i) backbone grid, (ii) neck grid, and (iii) head grid. The CSPDarnet53 classifier is utilized as a backbone grid which is the combination of Darknet53 and CSPNet [51]. With extra modules, convolutional and bath normalization layers are attached after the backbone grid. Further, the Mish activation function and spatial pyramid pooling are manipulated to enhance the correctness of the feature output and generalization capacity of the network [52]. Moreover, the main advantage of these layers is to enhance the difficult multi-target depiction experiment. In the neck grid, to lessen the information trajectory for the various detectors, a path aggregation network and FPN (feature pyramid network) are operated as a parameter assembling approach [53]. The head grid still employs the head grid of YOLOv3. Additionally, GIoU (Generalized Intersection over Union) is used as a loss function to improve the evaluation consequences of the target and optimized the model based on various factors such as illumination situation, height, dimension of the object, occlusion gradation, etc. Moreover, GIoU measures the intersection proportion between ground truth and bounding boxes of prior mount and prediction mount. The mathematical equation of the loss function is expressed in Equation (2) as:

$$L_{GIoU} = 1 - IoU + \frac{\left|P - X \cup Y^{gt}\right|}{\left|P\right|} \quad (2)$$

where P represents the small-scale box that immerses the ground truth and the predicted bounding boxes which determine the speculated ground target.

B. *Two-Stage Detectors*

The two-stage detectors are also called regional proposal-based approaches. In this approach, detection and categorization are achieved by extracting the applicant areas on the attribute map and accomplishing DL. Some frequently used one-stage detectors for target detection through UAVs are explained in the following.

- In 2015, Ross Girshick et al. [54] presented a Faster R-CNN which is another famous target detector. Faster R-CNN symbolizes a "region-based Convolution neural networks" which works better on real images. These real images are employed to the area of UAVs images. It can predict locality and classification of numerous bounding boxes at the same time. Its main advantage over other similar models of this algorithm is its high accuracy. In the beginning, the Faster-RCNN algorithm introduces a regional proposal network (RPN) network. In the congruent classification, the target candidates were specified in a similar classification and allocated in similar networks to execute the outstanding detection consequences during training [19]. FPN is an attribute that integrates various levels to make the final regression and classification more efficient during the employment of attributes [55].
- Cascade R-CNN was presented in 2018 which is the repeated form of the Faster R-CNN [56]. It is a cascaded structure that consists of numerous repeated structures and is linked sequentially [19,57]. This algorithm is composed of three segments which are: (1) feature extraction unit, (2) RPN unit, and (3) multi-stage cascade identification unit. The Cascade R-CNN framework is a multistage augmentation that trains the continuous detectors at various IOU thresholds [56]. For the next phase of training, the boundary frames built by the R-CNN scene are used as input. By utilizing a variety of special repressors, the Cascade R-CNN detects high-quality recognition by eliminating noisy identified boundary frames while maintaining useful, adjacent, and optimistic examples.

4. DL in Addressing Crucial Issues For Target Detection

A. *One-Stage Detectors in Target Detection*

In this subsection, we review the useful work of one-stage detectors in resolving the issues of target detection through UAVs in traffic congestion, which are outlined in Table 2.

Table 2. One-stage detectors to tackle the issues in target detection.

Issue	One-Stage Detectors	Measure	Prototype	Reference
Small-Size Objects	YOLO	computing cost, time complexity, low efficiency, high computing resources	Fusion of DL and traditional model matching for multi-target detection	[58]
	YOLOv3	computing cost, resource utilization	YOLO-GCC and Traffic-DQN	[59]
	YOLOv3	computing cost, safety and privacy, resource utilization	TAU and DeepSORT	[60]
	YOLOv3	Resource usage and time complexity	Tiny YOLOv3	[61]
	YOLOv3	computing cost, computing power	SiamMask target tracking technique	[62]
	YOLOv4	computing cost	KCF tracking method and average peak-to-correlation energy scheme	[63]

Table 2. *Cont.*

Issue	One-Stage Detectors	Measure	Prototype	Reference
Target Occlusion	YOLOv3	computing cost, processing speed	K-means++ algorithm, Soft-NMS algorithm and data augmentation technique	[64]
Complex Background	YOLO	computing cost, resource utilization, computing power	CSP BoT method, data augmentation, and Adaptive Spatial Feature Fusion approach	[65]
Joint Issues	YOLOv4	cost	K-Means Clustering, Data Enhancement Approach	[66]
	YOLOv4	computing cost, boost the efficiency, material resources	RFB, ULSAM and Soft-NMS design	[67]
	YOLOv4	computing cost, safety and stability	YOLOD based on YOLOv4	[68]

1. *One-Stage Detectors for Small-Size Objects:* In 2016, YOLO [18] performed real-time target detection to simultaneously predict the probability of location reliability and all target classes. In 2021, Sun et al. [69] proposed the YOLO approach because this approach is easy and simple. The target detection issue has been completely resolved by regression. The images captured from the UAVs have high-resolution characteristics. For the target, recognition detection used the VGG16 network as a backbone formation and the optimization scheme utilized the adaptive moment estimation Adam [58] in the training process which aims to hasten the speed of the prototype convergence. When data were deficient in an inadequate feature extraction network, they introduced transfer learning to enhance the accuracy of the training recognition estimation. Moreover, YOLO merges the target locality forecast and grouping forecast into an isolated neural network prototype to attain quick target recognition and detection with high reliability. In the target detection, the detection rate of the YOLO network is extremely high with 69% and the detection speed is 40 FPS/s, apart from the detection reliability which is lower as compared to other DL networks. Li et al. [59] proposed a global context cross (YOLO-GCC) model which signifies the design of YOLOv3 and GCNet to handle the blurring and fuzzy features of small objects. To extract several multi-dimensional feature maps, this model utilized the DarkNet-53 [21] as a backbone. The global context attention segment was attached as the latest backbone with DarkNet-53 and called GC-DarkNet to extract further accurate and compelling features. The H-Swish activation function was used to decrease the computing cost. In addition, an approach of intelligent traffic signal planning called Traffic Deep-Q Network (Traffic-DQN) is introduced which is based on deep reinforcement learning, taking the advantage of traffic flow facts gained from the YOLO-GCC and is used as the basis for transportation planning. The Traffic-DQN system shows apparent benefit in convergence velocity, and each diagnostic indicator is better than the corresponding one in the other approaches. The experiment testing was performed on four familiar UAVs datasets: (i) the UCAS-AOD dataset [70], (ii) the VisDrone2019 dataset [71], (iii) the TSD-MAX dataset [72], and (iv) the UA-DETRAC dataset [73] which comprises different classes such as car, bus, van, and others. The exploratory results show that the potential of the proposed method to identify small flow factors is clear and it is better than the YOLOv3 algorithm. Moreover, with small targets and mixed backgrounds, the position of the bounding box is more precise which is very essential for target detection in UAV images. In 2021, Benjdira et al. [60] proposed a Traffic Analysis from UAVs (TAU) approach to detect all of the existing targets inside one assembly and generated a UAV image-based

dataset which is divided into five groupings such as pedestrian, motorcycle/bicycle, car, truck, and bus. However, to further pursue the detecting target, an online multi-object tracking approach called DeepSORT was utilized [74]. This approach decreases the time consumption, guarantees safety on the highway, and somehow reduces the computing cost. However, YOLOv3 still entails high extraction cost. In addition, in the current genre, the TAU approach has a few limitations which are:

- The incoherence of the metrics of the x and y axis when the pixel indicator is a multiple of the height and breadth of the rectifying frames.
- Due to the high resolution, it is unsatisfactory to pass it online.

So, it is necessary to solve these limitations with some new generic algorithms in the future. In 2020, Feng et al. [75] presented a design composed of four segments: (i) vehicle detection, (ii) background registration, (iii) trajectory construction and compensation, and (iv) trajectory denoising to draw remarkably well the orbit of highway users which include pedestrian, motor vehicles (MV) and non-motor vehicles (NMV) such as bicycles, tricycles, and motorcycles and to track the small target trajectories. The YOLOv3 algorithm was applied in the first segment for detection accuracy and to get the target bounding boxes at this stage. To gain the image locomotion in the second step, the Shi–Tomasi corner attribute is utilized. This approach is an extremely popular corner detector and is extensively used due to its high correctness and fast speed for numerous real-time clarification applications and manipulated for monitoring and tracking of the target characteristics [76,77]. Trajectory construction and compensation is the third step of this design which has three main phases: (i) data correlation, (ii) trajectory classification, and (iii) trajectory compensation. The purpose of these phases is to configure the irregular vehicle trajectories formed on the basis of the perception of the speed restraint, contest the smashed trajectories, and implement the assembly tasks to rectify the omitted components. Moreover, the ensemble empirical mode decomposition approach is employed in the last step to remove noise and errors from arbitrary and unbalanced signals and enhance the trajectory reliability [78]. The exploratory outcome reveals that the presented design attains high accuracy in trajectory abstraction and detection. Figure 7 demonstrates the recall outcome of target classes in three test videos recorded by high-depiction cameras; similarly, Figure 8 represents the precision outcome of the target classes in the same test record videos.

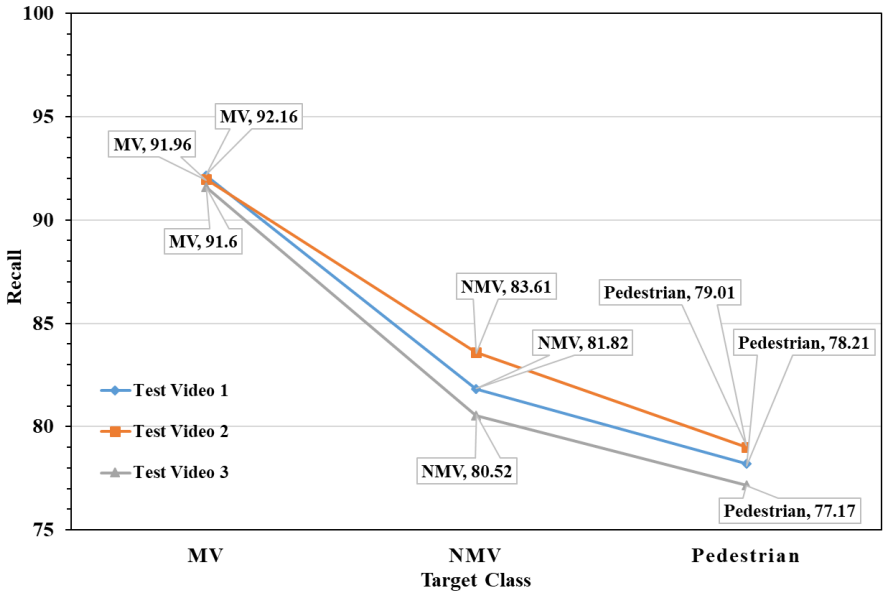

Figure 7. Recall calculation of target classes on three captured test datasets.

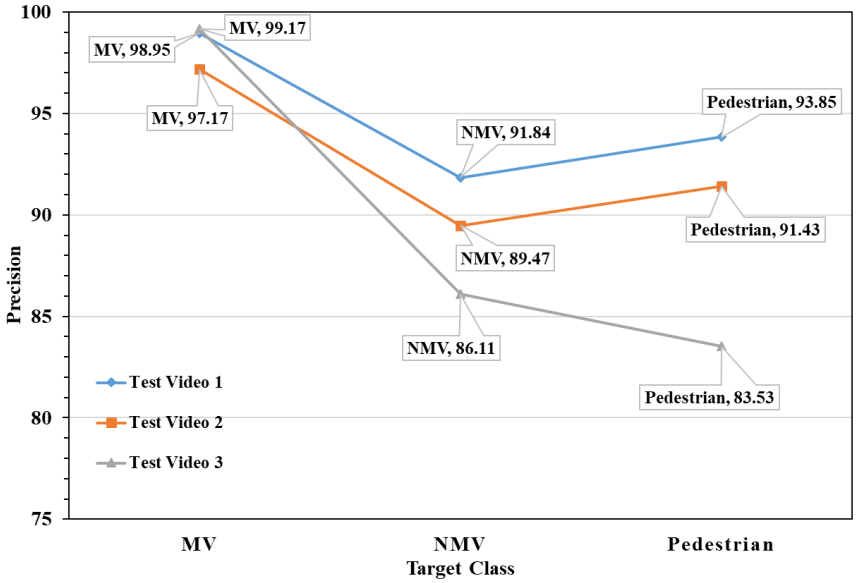

Figure 8. Precision calculation of target classes on three captured test datasets.

2. In 2022, Tian et al. [63] presented a YOLOv4-based approach to detect small targets in terms of cars and pedestrians on the VisDrone dataset. To enhance the performance of the presented model, the KCF algorithm and average peak-to-correlation energy scheme were utilized to stabilize the model and track small targets of long interspace.
One-Stage Detectors for Target Occlusion: In 2020, Luo et al. [64] presented a YOLOv3, soft non-maximum suppression (Soft-NMS), and K-means++ framework to handle the moderately occluded targets in the region of the UAV portrayal. The K-means++ method is used in the YOLOv3 algorithm to optimize the series of the first recognition box and improve the AP estimate of the network [79]. Later, the Soft-NMS method was executed to solve the issue of multi-box crushing by NMS to improve the AP estimate of the network [80]. During the training operation, overfitting occurs due to some training feature samples. To lessen this issue, data augmentation was implemented which comprises of color oscillation, arbitrary rotation, and image flip. For validation, three generic datasets were selected with various image characteristics to enhance the network. Based on the experimental results, it is observed that the upgraded YOLOv3 method achieved high accuracy and a fast detection rate. The results of the three datasets in terms of average precision (AP), precision, and recall metrics are presented in Figure 9.

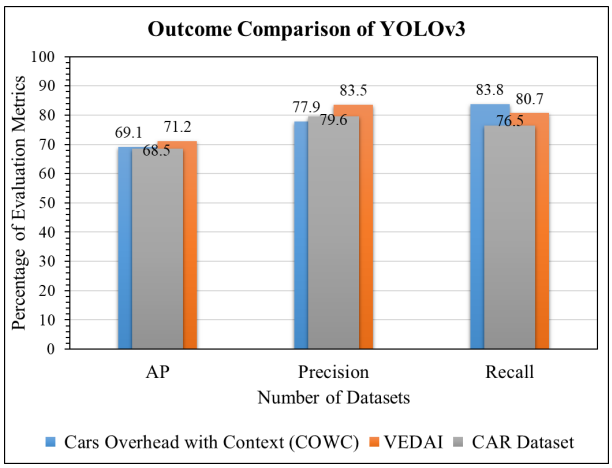

Figure 9. Result of YOLOv3 approach on different datasets.

3. *One-Stage Detectors for Complex background:* In [65], Feng et al. presented a gradient classification prophecy branch in the head network of YOLO to produce angular data and utilize the circular rectify class to overcome gradient classification loss and detect the target under complex circumstances. Moreover, to improve the UAV images, they implemented the data augmentation approach which consists of rotation, arbitrary flip, translation, and HSV augmentation. Then, they presented the cross-stage partial bottleneck transformer (CSP BoT) segment which is a hybrid approach that uses the multi-head self-attention process convolutions to encapsulate the latent broad spatial correlation of the target in the UAV images and improve the critical information. Finally, they adopted the general characteristics at various resolutions and predicted the spatial disparity in ambiguity by the weighted cross-scale interconnection. The adaptive spatial feature fusion-Head block was presented. The enormous experimental outcomes on UACS-AOD and UAV-ROD datasets show the presented model's dominance, low design complexity, and cost-effectiveness.
4. *One-Stage Detectors for Joint Issues:* Sun et al. [66] proposed a YOLOv4 approach based on K-means clustering to recognize the multi-resolution detection scheme. The drone collected data from a low-elevation aerial viewpoint which comprises of various data such as heights, size, and positions. Moreover, they applied a data enhancement approach to improve the robustness of the target model. This approach has two main parts: one is color transformation such as brightness, tinge, and contrast and the other one is geometric transformation such as rotating, arbitrarily clipping, flipping, and splicing. A Darknet scheme is used in the training process. Mean average precision (mAP), recall, and precision are evaluation metrics used for the assessment. During the experiment, it was observed that the target model's accuracy reaches up to 95% while on the same dataset, the accuracy reaches up to 96% in cloudy weather. Moreover, it is observed that there is a need to improve the detection of the model in terms of a dark target, for example, black vehicles. With respect to the occlusion issue, this approach performs the best, but there appear to be some conflicts in the huge sector such as it shows some error that needs to be minimized.
   In 2021, Tan et al. [67] presented a model named YOLOv4_Drone which contains the YOLOv4 algorithm based on UAV images. The detection correctness of the isolated YOLOv4 algorithm is almost low and causes errors. Therefore, to enhance the abstraction of the small targets, the receptive field block (RFB) segment [81] is included in the feature extraction phase of YOLOv4 to test the feature map and withdraw the features of various scales. The RFB segment was added to the target detection prototype and termed YOLOv4_r. There is no replication of the gradient statistics in the system optimization because this approach provides high validity while decreasing the computational complexity. In the UAV images, to solve the issue of small targets and complex backgrounds, the ultra-lightweight subspace attention mechanism (ULSAM) has been incorporated into the YOLOv4_r segment [82]. This segment derives a feature map with various attention functions for the respective feature map to represent the multi-scale function. A ULSAM segment connected to the target detection approach is called YOLOv4_u. Moreover, the soft non-maximum suppression (Soft-NMS) approach is utilized to reduce the missed target which is caused by the occlusion [80]. This is because this approach deletes the lower count frames of the two close targets if there is a huge overlap. Additionally, it remarkably decreases the statistics of detection frames. In the experimental testing, the VisDrone dataset is utilized. The various datasets of 14 areas of China based on lightning and weather situations are collected which comprise 10 kinds of targets as outlined in Figure 10. Moreover, mAP is used as an evaluation metric. From the testing, it is observed that the YOLOv4_Drone target detection approach achieved a high accuracy of 45.67% in all the weather conditions as compared to previous target detection approaches such as RetinaNet which achieved 35.95% accuracy, SMPNet achieved 35.98% accuracy, while DPNetV3 achieved 37.37%. Moreover, it is observed that

isolated YOLOv4 gained 40.99% accuracy which is 5% less than the YOLOv4_Drone target detection approach. Still, the presented approach is not abundantly stable due to a moderate runtime as compared to the other target detection paradigm.

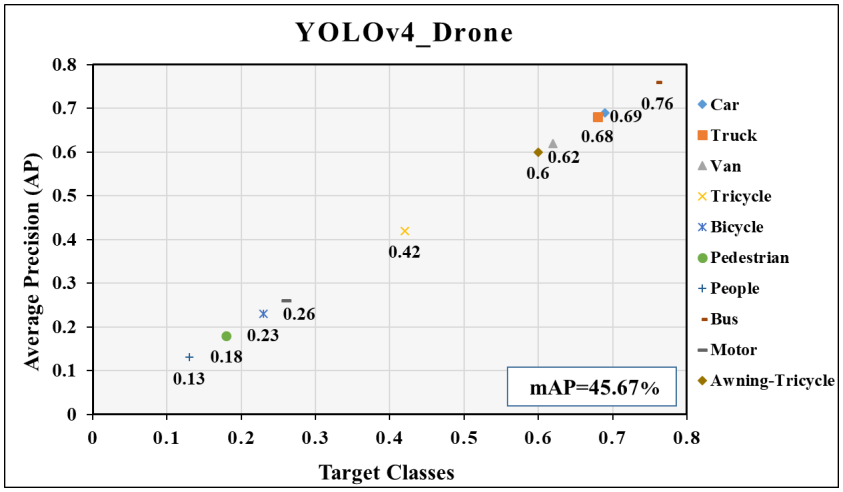

Figure 10. Average precision of YOLOv4_Drone approach based on different target classes.

In 2022, Luo et al. [68] presented a YOLO-DRONE (YOLOD) model of UAV images which were upgraded on the foundation of YOLOv4 to handle the small-size objects and clustered background. To decrease the complexity of the model and obtain the best detection consequences, different activation functions were used as a backbone which comprise of Mish [83] and HardSwish [84] activation functions. To enhance the location consequences, they stimulated the convergence, and summed up the loss of the bounding box regression EIOU loss function [85]. Moreover, they utilized the pyramid pooling module in the replacement of the SPP segment and compared the model with the YOLOv4 algorithm. The aim of using the pyramid pooling module is to enhance the receptive field and performance of the detection model [86]. To boost the multi-scale feature fusion and detect the targets on various scales at the end of the model, an adaptive spatial feature fusion segment was introduced [87]. The testing was done with the help of three different datasets which included forklift, VEDAI, and PASCAL VOC datasets, where the forklift is the first known dataset deployed on UAV images. During the investigation, it was observed that the presented YOLOD model achieves higher accuracy on all datasets and is good enough for complex backgrounds and small targets as compared to YOLOv4, as shown in Figure 11. However, it is still necessary to boost the performance of the model when the number of images is expanded.

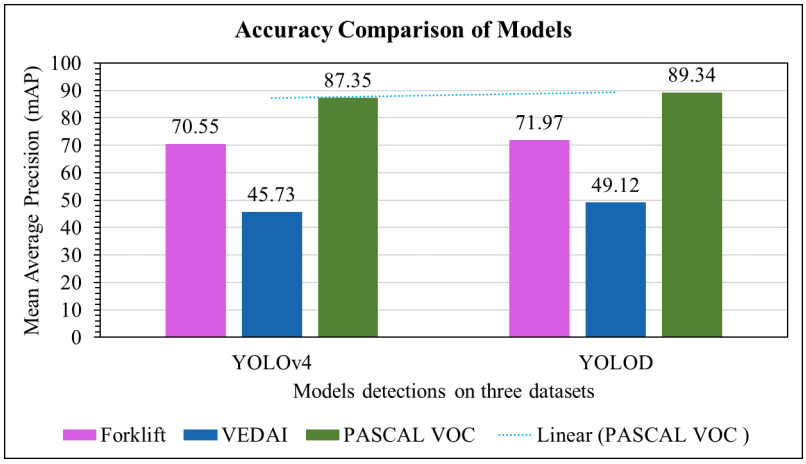

Figure 11. mAP of the target model based on three datasets.

B. *Two-Stage Detectors in Target Detection*

This subsection presents relative work on the adoption of two-stage detectors in resolving the issues of target detection through UAVs in traffic congestion. The basic work is summarized in Table 3.

Table 3. Two-stage detectors to tackle the issues in target detection.

Issue	Two-Stage Detectors	Measure	Prototype	Reference
Small-Size Objects	Faster R-CNN	Computing complexity, computing cost	cross-channel attributes, RoIAlign algorithm, loss function and data augmentation approach	[88]
	Faster R-CNN	processing efficiency, computing cost	GC-faster-RCNN, cluster approach, Resnext50 feature extractor and channel attention mechanism	[89]
	Faster R-CNN	Computing complexity	Utilized dataset of wide-range "object detection in Aerial images (DOTA)" and Annotation framework	[90]
	Cascade R-CNN	Computing time and cost	ECascade R-CNN network	[91]
Target Occlusion	Faster R-CNN	computing cost, processing speed	RPN and Multi-layer feature fusion	[92]
Complex Background	Faster R-CNN	Computing cost and complexity,	HOG+SVM, Faster R-CNN, ViBe and YOLOv3	[93]
Joint Issues	Faster R-CNN	computing cost and complexity	MS-Faster R-CNN and DeepSORT	[94]
	Cascade R-CNN	computing cost, efficiency, resource utilization	superclass detection design, Regression Confidence and Loss Function Improvements	[95]
	Cascade R-CNN	computational efficiency, safety and privacy	Faster R-CNN design, SSD and Cascade R-CNN + FPN	[96]

1. *Two-Stage Detectors for Small-Size Objects:* Zhu et al. [88] presented a faster R-CNN-based approach to detect the small-size target when the quantity of the corresponding anchor is deficient or when the targets are adjacent or lightly overlapped. This is because an insufficient quantity of the corresponding anchor increases the calculation complexity of the network. Therefore, the proposed model is divided into five phases:
   (a) As a feature extractor, ResNet101 is utilized to reduce the incline dispersion issue while reinforcing the system depth [97].
   (b) The RPN is demonstrated to produce the anchors of different sizes and dimensions.
   (c) By attaining the feature interconnection and cross-channel integration and minimizing the feature portray channels, a convolution layer of 1×1 is used.
   (d) The RoIAlign algorithm is used to prevent the loss of margin pixels. These pixels help to track and differentiate between small targets and adjacent or lightly overlapped targets. It also determines the misconfiguration generated by the RoIPool.
   (e) To analyze the targets and environmental situations through the image domain and purify the target domain bounding boxes, the classification and regression system is utilized.

Experimental evaluation was performed on the COCO dataset, where image flick augmentation is deployed in a horizontal position to achieve a model accuracy of 79.77%. To detect small-size objects according to different weather conditions, considered complex spots, highways, and roads, in 2022, Cheng et al. [89] presented a GC-faster-RCNN where GC is called "Group Convolution" which is gained by boosting the Faster R-CNN algorithm and various models. This includes a cluster approach to examine the datasets and, in the replacement of real feature extraction, the lattice Resnext50 is employed. Moreover, to enhance the statistics of features depicting into the network, an output attention mechanism of the enhanced channel is unified. During the testing, it was observed that the detection accuracy slightly increased by 94.8% while the speed of detection was slow. Besides this, small target detection is not fitted with a very deep network structure. Further, this approach enhances the computing cost when a huge number of categorized datasets are used. In the target detection work, to handle the multi-scale issue from the UAV images, in 2021, an ECascade-RCNN target detection framework was presented by Lin et al. [91], which is the improved version of Cascade-RCNN. As a backbone, Trident-FPN is employed to extract the attributes and to boost the execution of the detectors, a modern attention mechanism scheme is presented. In addition, the K-means technique has been used to create anchors to refine the model detection and attain the best regression precision. On the Visdrone dataset, testing was performed and during the investigation, it was observed that the model achieved better accuracy in huge and small-size objects.

2. *Two-Stage Detectors for Target Occlusion:* Due to occlusion, missed detection occurs. Therefore, it is necessary to reduce this issue. In 2020, Wang et al. [92] proposed a Faster R-CNN algorithm to handle this issue. The presented approach uses different anchor fusions to choose the maximum anchor number and scale, which enhances the network perception value. Further boosting the value of network perception, they added a multi-layer feature fusion. The experimental testing was performed based on datasets of UAV images where AP, precision, and recall are employed as evaluation indicators. The results of these indicators are demonstrated in Figure 12. However, it is observed that the variety of scenarios manipulated in this approach is restricted, and changing the focus of the UAVs may cause miss-identification.

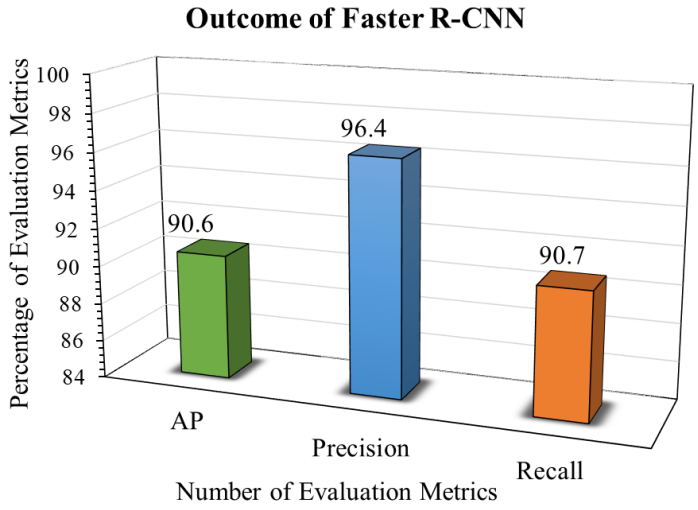

Figure 12. Result of evaluation metrics for target detection based on UAV dataset.

3. *Two-Stage Detectors for Complex Background:* Liu et al. [93] presented a framework composed of four different models such as (i) Faster R-CNN, (ii) YOLOv3 model, (iii) histogram of oriented gradients (HOG) + support vector machine (SVM) algorithms which are a form of machine learning [98], and (iv) for the background difference

technique, they chose the visual background extractor (ViBe) method. In video progression, ViBe is a dominant stochastic approach for predicting the background [99]. This proposed model handles the complicated background under various factors, for instance, wind, dozens of small and large-scale moving targets, unlimited speed, and substantial scenes. From the analysis, it is observed that the ViBe and HOG+SVM approaches do not yield satisfactory results due to restricted coherent and feature trajectory perception while YOLOv3 and Faster R-CNN perform best as compared to the other two approaches. In contrast to YOLOv3, Faster R-CNN yields the best accuracy in terms of recall and precision. However, it seems that Faster R-CNN has a crucial problem due to excessive hardware demand when implemented in the actual landscape.

4. *Two-Stage Detectors for Joint Issues:* In 2021, Avola et al. [94] presented an MS-Faster R-CNN where MS stands for multi-stream which has three stages: in the stated frame, the multi-stream CNN abstract attributes at multiple scales from the target in the first step by manipulating its inherent architectural model. Second, under the Faster-CNN method, the bounding boxes close to the target are obtained using the extracted attributes map where the backbone produces CNN capability so that the area of the affinity group layer and region proposal network can turn the output of the classifier into the necessary bounding boxes. In the last stage, when the targets can be detected accurately by MS Faster-R-CNN inside of the graphical cascade, the DeepSORT [17] technique is used to attain the real-time monitoring abilities from the UAVs perception. The evaluation was performed on four datasets which include: (i) UAV20L [100], (ii) UMCD [101], (iii) UAV123 [100], and UAVDT [102]. The data recorded from the UAVs comprise different features such as weather situations, small-size objects, lightning variations, huge occlusion, partial and full occlusion, background group, and low rectification. It is observed that the model is slightly improved but still the design is not fast and the detection speed is limited. In 2022, Huang et al. [95] presented an improved innovation in accordance with the Cascade R-CNN network where the framework of superclass detection is developed to supply the best precise region of interest for the consequential detector to enhance the recognition of the equivalent group. The final dependence can reflect the superior characteristics of the detection outcome with the help of regression dependence and virtually enhance the area accuracy. Simultaneously, to boost the detection outcome of the framework in the inspection of small-size targets in complex backgrounds, heavy target occlusion is used, and to lessen the phenomenon of the false alarms, the loss function is used. Moreover, this approach aims to improve the accuracy and speed of the target. The testing was performed on the VisDrone dataset which consists of 10 kinds of target classes as shown in Figure 13.

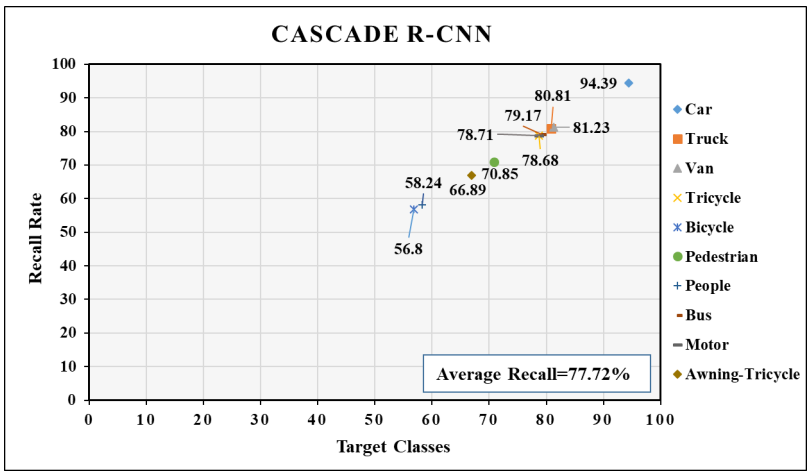

Figure 13. Recall rate of Cascade R-CNN approach based on different target classes.

5. Discussion and Future Research Trends

Current studies demonstrate that scientists show much concern for the modern use of DL approaches to address problems in target detection through UAVs, for instance, small-size objects, target occlusion, complex backgrounds, and joint issues. However, to overcome these problems, there are even now obvious questions for scientists. The data in this section reflect DL-existing conditions for target detection and indicate future fact-finding directions.

A. *Discussion*

Figure 14 shows the total percentage division of DL approaches listed in Figure 6 to solve the problems for target detection. Figure 15 describes all problems, DL approaches, and the aggregate of the articles related to each problem in target detection. In addition, we specified the major research findings obtained from Section 4.

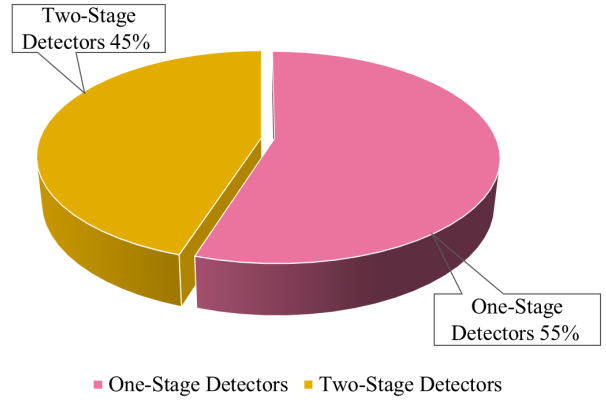

Figure 14. Comprehensive percentage division of DL approaches for handling problems in target detection.

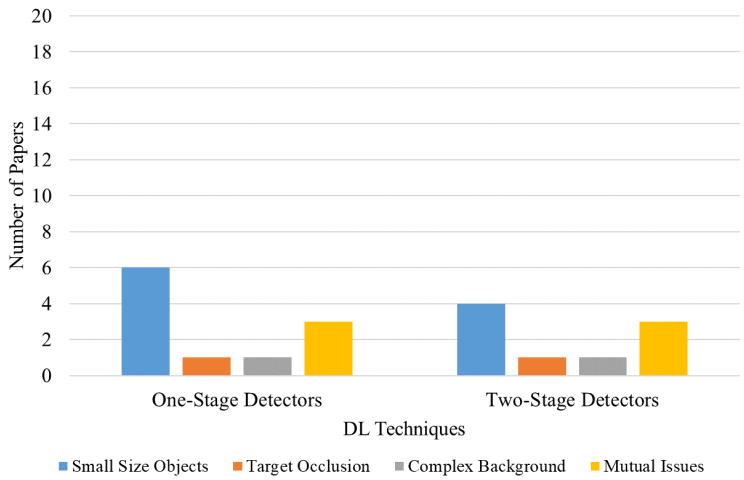

Figure 15. Applicability of DL approaches for handling problems in target detection.

1. *One-Stage Detectors:* In 55% of the papers, the one-stage detectors have been manipulated in contrast to other DL approaches as presented in Figure 14. From Figure 16, we can notice that small-size objects are a trendy problem tackled by one-stage detectors and YOLOv3 is commonly utilized as one-stage detector in target detection. Figure 16 also presents the classification of one-stage detectors in problems related to target detection.

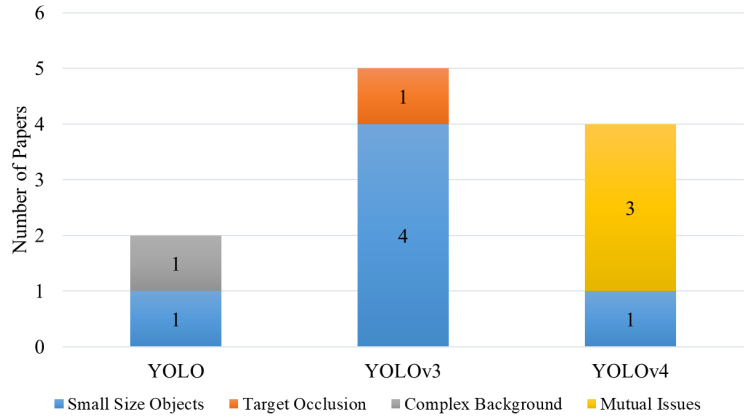

Figure 16. Classification of One-Stage Detectors in DL.

2. *Two-Stage Detectors:* From Figure 14, it is observed that 45% of the papers used two-stage detectors to tackle the problems in target detection. Two-stage detectors have been commonly employed for small-size objects and joint issues as demonstrated in Figure 17. The classification of two-stage detectors for problems related to target detection is shown in Figure 17.

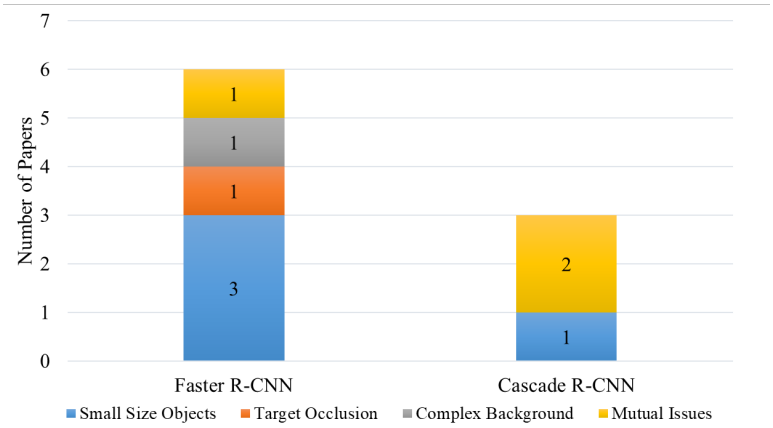

Figure 17. Classification of Two-Stage Detectors in DL.

B. *Remarks*

Based on the previous consideration, we want to give further details to figure out what type of DL approaches are good for resolving what types of problems:

- In Faster R-CNN, the information about small targets will moderately disappear as the network progresses. Therefore, for very deep network structures, small target detections are not good enough.
- The detection speed of the two-stage detectors is slow as compared to one-stage detectors which is the major defect in target detection. On the other hand, it is observed that YOLOv3 achieved high accuracy and a fast detection rate as compared to other YOLO versions for small-size objects but still possesses high computing complexity in real-world performance.
- One-stage detector performs best in the occlusion issue while in the occlusion of large regions, the model still shows some misconception.
- In the YOLO algorithm [18], small targets in the form of groups are hard to handle because the generalization capacities of the model are poor, and loss function issues easily cause prominent positioning miscalculations.
- Faster R-CNN performs best compared to YOLOv3 in the complex background but it is noticed that the excessive hardware need is a serious issue. So, according to the

particular complex situations, how to execute the Faster R-CNN on moderate-size hardware with wide prospects is also another issue.

- Overall, it is also observed that it is crucial to obtain real-time identification in two-stage detectors because the accuracy is high while the calculation amount is still huge.
- For joint issues, the accuracy of the YOLOv4 algorithm is improved in contrast to other approaches but its speed is moderately decreased.

C. *Future Research Trends*

- UAV's vision-based approaches for the target detection are highly remarkable in contrast to traditional approaches. Through UAV's target detection, the UAV vision-based approaches, recognition and detection encryption are important parts of it. So, for the future direction, it is necessary to perceive more appropriate recognition and detection encryption. Moreover, it is necessary to enhance the UAV vision-based model accuracy for the desirable outcome and to overcome the existing conflicts.
- To reduce the number of false alarms during the tiny target detection for railing broadcasting of surveillance, it is demanded to deploy the various visual sensors and different perspective stereo visions with the help of multi-sensor collaboration.
- In terms of power consumption, it is compulsory to propose efficient approaches and present the advanced function offloading methods which can discover the light-leverage batteries that can notably increase the flight duration of UAVs, endure the prolonged distance, and decrease the total power consumption to carry out the predefined functions in the future.
- In the future, it is mandatory to optimize the proposed algorithms by the use of laser facts and depth maps etc., to record more productive image characteristics and detect small and dense obstacles.
- The YOLOv4_Drone approach yields a 5% to 15% higher detection accuracy as compared to other models such as CenterNet which has 29.85% detection accuracy but in terms of real-time implementation and speed, it is required to improve the YOLOv4_Drone approach in the future.
- The bounding box associated with the pedestrian category is extracted from the concluding measurement in the TAU approach and has some limitations in the current design. In the future, it is obligatory to replace the current design of YOLOv3 with the latest version of YOLOv4 and an online DeepSORT target tracker with a multi-object tracker to solve these issues with efficient results.
- In the area of aerial photography, developing an extensive and versatile dataset is a major challenge. So, it is obligatory to gather high-standard datasets to ease this challenge. In addition, researchers are required to generate effective and surprisingly automated approaches to classify the training data.
- In the future, it is necessary to merge the one-stage and two-stage detectors to achieve the best outcome because both detectors have their own benefits. For instance, one-stage algorithms are fast while two-stage algorithms have influential accuracy.
- In contrast to YOLO-based algorithms, the detection speed of Faster R-CNN still requires to be upgraded and gradually put forward in upcoming research.
- To train a framework of high quality, in the future, there is a prerequisite to change the portable datasets of high quality into vast datasets.
- To discuss every feature regarding the employment of UAVs in target detection, we plan to extend the study, involve more work, and compute the latest approach in the future.
- There are also some other YOLO versions such as YOLOv5 which were introduced in May 2020 after two months of the YOLOv4 version. Recently, a few papers have been published based on YOLOv5 on some custom datasets due to efficient memory in the training process. Meanwhile, the other versions YOLOv6, YOLOv7, and YOLOv8 are still in the improvement phase in terms of training speed and computational cost. Therefore, no paper has been published yet based on those versions. In the coming years, our entire focus will be on these versions to handle different issues.

6. Conclusions

In this study, we discussed the implementation of DL approaches to four crucial problems in target detection through UAVs: small-size objects, target occlusion, complex background, and joint issues. We inaugurated primitive ideas of DL along with crucial problems, and metrics and then concentrated on two groups of DL approaches utilized in target detection: one-stage detectors and two-stage detectors. Subsequently, different designed methods depending on DL approaches were examined in the framework of target detection based on UAVs. Further, the data about the current status quo of DL for target detection were presented on the basis of the study gathered in this review article. We observed that one-stage detectors were vigorously utilized in target detection due to their fast detection rate while two-stage detectors are not extensively used in target detection due to their low detection rate. In the end, we suggested some remarkable challenges and upcoming research directions for DL and target detection.

Author Contributions: All authors have equally contributed. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the EIAS Data Science Lab, College of Computer and Information Sciences, Prince Sultan University, Riyadh, Saudi Arabia. Also, the studies at St. Petersburg State University of Telecommunications. M.A. Bonch-Bruevich was supported by the Ministry of Science and High Education of the Russian Federation by the grant 075-15-2022-1137.

Data Availability Statement: Not appliable.

Acknowledgments: The authors would like to acknowledge the support of Prince Sultan University for paying the Article Processing Charges (APC) of this publication.

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

Abbreviations

The following abbreviations are used in this manuscript:

UAVs	Unmanned Aerial Vehicles
IMU	Inertial Measurement Unit
GPS	Global Positioning System
DL	Deep Learning
RCNN	Region-Based Convolutional Neural Network
SSD	Single Shot MultiBox Detector
WPT	Wireless Power Transfer
GIoU	Generalized Intersection over Union
RPN	Region Proposal Network Network
YOLO-GCC	You Only Live Once-Global Context Cross
TAU	Traffic Analysis from UAVs
MV	Motor Vehicles
NMV	Non-Motor Vehicles
KCF	Kernel Correlation Filter
Soft-NMS	Soft Non-Maximum Suppression
CSP BoT	Cross-Stage Partial Bottleneck Transformer
mAP	Mean Average Precision
RFB	Receptive Field Block
ULSAM	Ultra-Lightweight Subspace Attention Mechanism
GC	Group Convolution
HOG	Histogram of Oriented Gradients
SVM	Support Vector Machines
ViBe	Visual Background Extractor
MS	Multi-Stream
FPN	feature pyramid network

References

1. Tian, Y.; Luo, P.; Wang, X.; Tang, X. Pedestrian detection aided by deep learning semantic tasks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 5079–5087.
2. Zhou, Y.; Liu, L.; Shao, L.; Mellor, M. DAVE: A unified framework for fast vehicle detection and annotation. In Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 11–14 October 2016; Springer: Cham, Swizterland, 2016; pp. 278–293.
3. Iftikhar, S.; Asim, M.; Zhang, Z.; El-Latif, A.A.A. Advance generalization technique through 3D CNN to overcome the false positives pedestrian in autonomous vehicles. *Telecommun. Syst.* **2022**, *80*, 545–557. [CrossRef]
4. Iftikhar, S.; Zhang, Z.; Asim, M.; Muthanna, A.; Koucheryavy, A.; Abd El-Latif, A.A. Deep Learning-Based Pedestrian Detection in Autonomous Vehicles: Substantial Issues and Challenges. *Electronics* **2022**, *11*, 3551. [CrossRef]
5. Kazim, M.; Azar, A.T.; Koubaa, A.; Zaidi, A. Disturbance-Rejection-Based Optimized Robust Adaptive Controllers for UAVs. *IEEE Syst. J.* **2021**, *15*, 3097–3108 . [CrossRef]
6. Alotaibi, E.T.; Alqefari, S.S.; Koubaa, A. LSAR: Multi-UAV Collaboration for Search and Rescue Missions. *IEEE Access* **2019**, *7*, 55817–55832. [CrossRef]
7. Asim, M.; ELAffendi, M.; El-Latif, A.A.A. Multi-IRS and Multi-UAV-Assisted MEC System for 5G/6G Networks: Efficient Joint Trajectory Optimization and Passive Beamforming Framework. *IEEE Trans. Intell. Transp. Syst.* **2022**, 1–12. [CrossRef]
8. Mustafa Hilal, A.; Jaber, S.; Alzahrani, J.S.; Elkamchouchi, D.H.; Eltahir, M.M.; Almasoud, A.S.; Motwakel, A.; Zamani, A.S.; Yaseen, I. Optimal Deep Learning Enabled Communication System for Unmanned Aerial Vehicles. *Comput. Syst. Sci. Eng.* **2022**, *45* , 030132. [CrossRef]
9. Khan, M.A.; Kumar, N.; Mohsan, S.A.H.; Khan, W.U.; Nasralla, M.M.; Alsharif, M.H.; Żywiołek, J.; Ullah, I. Swarm of UAVs for Network Management in 6G: A Technical Review. *IEEE Trans. Netw. Serv. Manag.* **2022**, *20*, 741–761. [CrossRef]
10. Asim, M.; Wang, Y.; Wang, K.; Huang, P.Q. A Review on Computational Intelligence Techniques in Cloud and Edge Computing. *IEEE Trans. Emerg. Top. Comput. Intell.* **2020**, *4*, 742–763. [CrossRef]
11. Rozantsev, A.; Lepetit, V.; Fua, P. Flying objects detection from a single moving camera. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 4128–4136.
12. 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]
13. Zhang, Y.; Zhao, C.; Chen, A.; Qi, X. Vehicle detection in urban traffic scenes using the pixel-based adaptive segmenter with confidence measurement. *J. Intell. Fuzzy Syst.* **2016**, *31*, 1609–1620. [CrossRef]
14. Bolme, D.S.; Beveridge, J.R.; Draper, B.A.; Lui, Y.M. Visual object tracking using adaptive correlation filters. In Proceedings of the 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Francisco, CA, USA, 13–18 June 2010; IEEE: New York, NY, USA, 2010; pp. 2544–2550.
15. Henriques, J.F.; Caseiro, R.; Martins, P.; Batista, J. Exploiting the circulant structure of tracking-by-detection with kernels. In Proceedings of the European Conference on Computer Vision, Florence, Italy, 7–13 October, 2012; Springer: Cham, Switzerland, 2012; pp. 702–715.
16. Ke, R.; Li, Z.; Tang, J.; Pan, Z.; Wang, Y. Real-time traffic flow parameter estimation from UAV video based on ensemble classifier and optical flow. *IEEE Trans. Intell. Transp. Syst.* **2018**, *20*, 54–64. [CrossRef]
17. Wojke, N.; Bewley, A.; Paulus, D. Simple online and realtime tracking with a deep association metric. In Proceedings of the 2017 IEEE International Conference on Image Processing (ICIP), Beijing, China, 17–20 September 2017; IEEE: New York, NY, USA, 2017; pp. 3645–3649.
18. 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.
19. 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] [PubMed]
20. Radovic, M.; Adarkwa, O.; Wang, Q. Object recognition in aerial images using convolutional neural networks. *J. Imaging* **2017**, *3*, 21. [CrossRef]
21. Redmon, J.; Farhadi, A. Yolov3: An incremental improvement. *arXiv* **2018**, arXiv:1804.02767.
22. Bochkovskiy, A.; Wang, C.Y.; Liao, H.Y.M. Yolov4: Optimal speed and accuracy of object detection. *arXiv* **2020**, arXiv:2004.10934.
23. Saribas, H.; Uzun, B.; Benligiray, B.; Eker, O.; Cevikalp, H. A hybrid method for tracking of objects by UAVs. In Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Long Beach, CA, USA, 16–17 June 2019.
24. Henruques, J.; Caseiro, R. High speed tracking with kemelized correlation filters. *IEEE Trans. Pattern Anal. Mach. Intell.* **2015**, *37*, 583–596. [CrossRef]
25. Butilă, E.V.; Boboc, R.G. Urban Traffic Monitoring and Analysis Using Unmanned Aerial Vehicles (UAVs): A Systematic Literature Review. *Remote Sens.* **2022**, *14*, 620. [CrossRef]
26. Srivastava, S.; Narayan, S.; Mittal, S. A survey of deep learning techniques for vehicle detection from UAV images. *J. Syst. Archit.* **2021**, *117*, 102152. [CrossRef]
27. Osco, L.P.; Junior, J.M.; Ramos, A.P.M.; de Castro Jorge, L.A.; Fatholahi, S.N.; de Andrade Silva, J.; Matsubara, E.T.; Pistori, H.; Gonçalves, W.N.; Li, J. A review on deep learning in UAV remote sensing. *Int. J. Appl. Earth Obs. Geoinf.* **2021**, *102*, 102456. [CrossRef]

28. Alzahrani, B.; Oubbati, O.S.; Barnawi, A.; Atiquzzaman, M.; Alghazzawi, D. UAV assistance paradigm: State-of-the-art in applications and challenges. *J. Netw. Comput. Appl.* **2020**, *166*, 102706. [CrossRef]
29. Kanistras, K.; Martins, G.; Rutherford, M.J.; Valavanis, K.P. A survey of unmanned aerial vehicles (UAVs) for traffic monitoring. In Proceedings of the 2013 International Conference on Unmanned Aircraft Systems (ICUAS), Atlanta, GA, USA, 28–31 May 2013; IEEE: New York, NY, USA, 2013; pp. 221–234.
30. Outay, F.; Mengash, H.A.; Adnan, M. Applications of unmanned aerial vehicle (UAV) in road safety, traffic and highway infrastructure management: Recent advances and challenges. *Transp. Res. Part A Policy Pract.* **2020**, *141*, 116–129. [CrossRef]
31. Park, H.; Byun, S.; Lee, H. Application of deep learning method for real-time traffic analysis using UAV. *J. Korean Soc. Surv. Geod. Photogramm. Cartogr.* **2020**, *38*, 353–361.
32. Zhang, S.; Zheng, K.; Huaiyuan, S. Analysis of the Occlusion Interference Problem in Target Tracking. *Math. Probl. Eng.* **2022**, *2022*, 4605111. [CrossRef]
33. Elloumi, M.; Dhaou, R.; Escrig, B.; Idoudi, H.; Saidane, L.A. Monitoring road traffic with a UAV-based system. In Proceedings of the 2018 IEEE Wireless Communications and Networking Conference (WCNC), Barcelona, Spain, 15–18 April 2018; IEEE: New York, NY, USA, 2018; pp. 1–6.
34. Liu, X.; Zhang, Z. A vision-based target detection, tracking, and positioning algorithm for unmanned aerial vehicle. *Wirel. Commun. Mob. Comput.* **2021**, *2021*, 5565589. [CrossRef]
35. Khan, N.A.; Jhanjhi, N.; Brohi, S.N.; Usmani, R.S.A.; Nayyar, A. Smart traffic monitoring system using unmanned aerial vehicles (UAVs). *Comput. Commun.* **2020**, *157*, 434–443. [CrossRef]
36. Cheng, S.; Qin, J.; Chen, Y.; Li, M. Moving Target Detection Technology Based on UAV Vision. *Wirel. Commun. Mob. Comput.* **2022**, *2022*, 5443237. [CrossRef]
37. Campi, T.; Cruciani, S.; Feliziani, M. Wireless power transfer technology applied to an autonomous electric UAV with a small secondary coil. *Energies* **2018**, *11*, 352. [CrossRef]
38. Mohsan, S.A.H.; Othman, N.Q.H.; Khan, M.A.; Amjad, H.; Żywiołek, J. A Comprehensive Review of Micro UAV Charging Techniques. *Micromachines* **2022**, *13*, 977. [CrossRef]
39. Campi, T.; Dionisi, F.; Cruciani, S.; De Santis, V.; Feliziani, M.; Maradei, F. Magnetic field levels in drones equipped with wireless power transfer technology. In Proceedings of the e 2016 Asia-Pacific International Symposium on Electromagnetic Compatibility (APEMC), Shenzhen, China, 18–21 May 2016; Volume 1, pp. 544–547.
40. Trihinas, D.; Agathocleous, M.; Avogian, K.; Katakis, I. Flockai: A testing suite for ml-driven drone applications. *Future Internet* **2021**, *13*, 317. [CrossRef]
41. Vattapparamban, E.; Güvenç, I.; Yurekli, A.I.; Akkaya, K.; Uluağaç, S. Drones for smart cities: Issues in cybersecurity, privacy, and public safety. In Proceedings of the 2016 international Wireless Communications and Mobile Computing Conference (IWCMC), Paphos, Cyprus, 5–9 September 2016; IEEE: New York, NY, USA, 2016; pp. 216–221.
42. Syed, F.; Gupta, S.K.; Hamood Alsamhi, S.; Rashid, M.; Liu, X. A survey on recent optimal techniques for securing unmanned aerial vehicles applications. *Trans. Emerg. Telecommun. Technol.* **2021**, *32*, e4133.
43. Choi, J.Y.; Sung, K.S.; Yang, Y.K. Multiple vehicles detection and tracking based on scale-invariant feature transform. In Proceedings of the 2007 IEEE Intelligent Transportation Systems Conference, Bellevue, WA, USA, 30 September–3 October 2007; IEEE: New York, NY, USA, 2007; pp. 528–533.
44. Dai, J.; Li, Y.; He, K.; Sun, J. R-fcn: Object detection via region-based fully convolutional networks. *Adv. Neural Inf. Process. Syst.* **2016**, *29*.
45. Dkabrowski, P.S.; Specht, C.; Specht, M.; Burdziakowski, P.; Makar, A.; Lewicka, O. Integration of multi-source geospatial data from GNSS receivers, terrestrial laser scanners, and unmanned aerial vehicles. *Can. J. Remote Sens.* **2021**, *47*, 621–634. [CrossRef]
46. Han, R.; Zhang, C. Big Data Analysis on Economical Urban Traffic in Beijing: Organize overlapping transportation though the underground diameter line of Beijing railway hub. In Proceedings of the 2019 IEEE 4th International Conference on Cloud Computing and Big Data Analysis (ICCCBDA), Chengdu, China, 12–15 April 2019; IEEE: New York, NY, USA, 2019; pp. 269–273.
47. Lin, T.Y.; Maire, M.; Belongie, S.; Hays, J.; Perona, P.; Ramanan, D.; Dollár, P.; Zitnick, C.L. Microsoft coco: Common objects in context. In Proceedings of the European Conference on Computer Vision, Zurich, Switzerland, 6–12 September 2014; Springer: Cham, Switzerland, 2014; pp. 740–755.
48. Lif, P.; Näsström, F.; Tolt, G.; Hedström, J.; Allvar, J. Visual and IR-based target detection from unmanned aerial vehicle. In Proceedings of the International Conference on Human Interface and the Management of Information, Vancouver, BC, Canada, 9–14 July 2017; Springer: Cham, Switzerland, 2017; pp. 136–144.
49. Jawaharlalnehru, A.; Sambandham, T.; Sekar, V.; Ravikumar, D.; Loganathan, V.; Kannadasan, R.; Khan, A.A.; Wechtaisong, C.; Haq, M.A.; Alhussen, A.; et al. Target Object Detection from Unmanned Aerial Vehicle (UAV) Images Based on Improved YOLO Algorithm. *Electronics* **2022**, *11*, 2343. [CrossRef]
50. Ren, X.; Sun, M.; Jiang, C.; Liu, L.; Huang, W. An augmented reality Geo-registration method for ground target localization from a low-cost UAV platform. *Sensors* **2018**, *18*, 3739. [CrossRef]
51. Wang, C.Y.; Liao, H.Y.M.; Wu, Y.H.; Chen, P.Y.; Hsieh, J.W.; Yeh, I.H. CSPNet: A new backbone that can enhance learning capability of CNN. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Seattle, WA, USA, 14–19 June 2020; pp. 390–391.

52. He, K.; Zhang, X.; Ren, S.; Sun, J. Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2015**, *37*, 1904–1916. [CrossRef] [PubMed]
53. 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.
54. Girshick, R. Fast r-cnn. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 1440–1448.
55. 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.
56. Cai, Z.; Vasconcelos, N. Cascade r-cnn: Delving into high quality object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Portland, OR, USA, 23–28 June 2018; pp. 6154–6162.
57. Cai, Z.; Vasconcelos, N. Cascade R-CNN: High quality object detection and instance segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2019**, *43*, 1483–1498. [CrossRef]
58. Sun, L.; Chen, J.; Feng, D.; Xing, M. Parallel ensemble deep learning for real-time remote sensing video multi-Target detection. *Remote Sens.* **2021**, *13*, 4377. [CrossRef]
59. Li, Y.; Chen, Y.; Yuan, S.; Liu, J.; Zhao, X.; Yang, Y.; Liu, Y. Vehicle detection from road image sequences for intelligent traffic scheduling. *Comput. Electr. Eng.* **2021**, *95*, 107406. [CrossRef]
60. Benjdira, B.; Koubaa, A.; Azar, A.T.; Khan, Z.; Ammar, A.; Boulila, W. TAU: A framework for video-based traffic analytics leveraging artificial intelligence and unmanned aerial systems. *Eng. Appl. Artif. Intell.* **2022**, *114*, 105095. [CrossRef]
61. Ajaz, A.; Salar, A.; Jamal, T.; Khan, A.U. Small Object Detection using Deep Learning. *arXiv* **2022**, arXiv:2201.03243.
62. Li, X.; Wang, F.; Xu, A.; Zhang, G. UAV Aerial Photography Target Detection and Tracking Based on Deep Learning. In Proceedings of the 5th China Aeronautical Science and Technology Conference, Wuzhen, China, 27–29 May 2022; Springer: Cham, Switzerland, 2022; pp. 426–438.
63. Tian, X.; Jia, Y.; Luo, X.; Yin, J. Small Target Recognition and Tracking Based on UAV Platform. *Sensors* **2022**, *22*, 6579. [CrossRef] [PubMed]
64. Luo, X.; Tian, X.; Zhang, H.; Hou, W.; Leng, G.; Xu, W.; Jia, H.; He, X.; Wang, M.; Zhang, J. Fast automatic vehicle detection in uav images using convolutional neural networks. *Remote Sens.* **2020**, *12*, 1994. [CrossRef]
65. Feng, J.; Yi, C. Lightweight Detection Network for Arbitrary-Oriented Vehicles in UAV Imagery via Global Attentive Relation and Multi-Path Fusion. *Drones* **2022**, *6*, 108. [CrossRef]
66. Sun, H.; Xing, G. A YOLOv4-based vehicle detection method from UAV Videos. In Proceedings of the 2021 33rd Chinese Control and Decision Conference (CCDC), Kunming, China, 22–24 May 2021; IEEE: New York, NY, USA, 2021; pp. 3082–3087.
67. Tan, L.; Lv, X.; Lian, X.; Wang, G. YOLOv4_Drone: UAV image target detection based on an improved YOLOv4 algorithm. *Comput. Electr. Eng.* **2021**, *93*, 107261. [CrossRef]
68. Luo, X.; Wu, Y.; Zhao, L. YOLOD: A Target Detection Method for UAV Aerial Imagery. *Remote Sens.* **2022**, *14*, 3240. [CrossRef]
69. Kingma Diederik, P.; Adam, J.B. A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980.
70. Zhu, H.; Chen, X.; Dai, W.; Fu, K.; Ye, Q.; Jiao, J. Orientation robust object detection in aerial images using deep convolutional neural network. In Proceedings of the 2015 IEEE International Conference on Image Processing (ICIP), Quebec City, QC, Canada, 27–30 September 2015; IEEE: New York, NY, USA, 2015; pp. 3735–3739.
71. Zhu, P.; Wen, L.; Du, D.; Bian, X.; Fan, H.; Hu, Q.; Ling, H. Detection and tracking meet drones challenge. *IEEE Trans. Pattern Anal. Mach. Intell.* **2021**, *44*, 7380–7399. [CrossRef]
72. Available online: http://trafficdata.xjtu.edu.cn/index.do (accessed on 28 February 2023).
73. Wen, L.; Du, D.; Cai, Z.; Lei, Z.; Chang, M.C.; Qi, H.; Lim, J.; Yang, M.H.; Lyu, S. UA-DETRAC: A new benchmark and protocol for multi-object detection and tracking. *Comput. Vis. Image Underst.* **2020**, *193*, 102907. [CrossRef]
74. Ciaparrone, G.; Sánchez, F.L.; Tabik, S.; Troiano, L.; Tagliaferri, R.; Herrera, F. Deep learning in video multi-object tracking: A survey. *Neurocomputing* **2020**, *381*, 61–88. [CrossRef]
75. Feng, R.; Fan, C.; Li, Z.; Chen, X. Mixed road user trajectory extraction from moving aerial videos based on convolution neural network detection. *IEEE Access* **2020**, *8*, 43508–43519. [CrossRef]
76. Ramakrishnan, N.; Wu, M.; Lam, S.K.; Srikanthan, T. Automated thresholding for low-complexity corner detection. In Proceedings of the 2014 NASA/ESA Conference on Adaptive Hardware and Systems (AHS), Leicester, UK, 14–17 July 2014; IEEE: New York, NY, USA, 2014; pp. 97–103.
77. Luo, Y.; Liang, Y.; Ke, R.; Luo, X. Traffic flow parameter estimation from satellite video data based on optical flow. In Proceedings of the Transportation Research Board 97th Annual Meeting, Washington, DC, USA, 7–11 January 2018 .
78. Huang, N.E.; Shen, Z.; Long, S.R.; Wu, M.C.; Shih, H.H.; Zheng, Q.; Yen, N.C.; Tung, C.C.; Liu, H.H. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.* **1998**, *454*, 903–995. [CrossRef]
79. Shah, S.; Singh, M. Comparison of a time efficient modified K-mean algorithm with K-mean and K-medoid algorithm. In Proceedings of the 2012 International Conference on Communication Systems and Network Technologies, Rajkot, India, 11–13 May 2012; IEEE: New York, NY, USA, 2012; pp. 435–437.
80. Bodla, N.; Singh, B.; Chellappa, R.; Davis, L.S. Soft-NMS–Improving object detection with one line of code. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 5561–5569.

81. Liu, S.; Huang, D. Receptive field block net for accurate and fast object detection. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–11 September 2018; pp. 385–400.
82. Saini, R.; Jha, N.K.; Das, B.; Mittal, S.; Mohan, C.K. Ulsam: Ultra-lightweight subspace attention module for compact convolutional neural networks. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, Snowmass Village, CO, USA, 1–5 March 2020; pp. 1627–1636.
83. Misra, D. Mish: A self regularized non-monotonic activation function. *arXiv* **2019**, arXiv:1908.08681.
84. Howard, A.; Sandler, M.; Chu, G.; Chen, L.C.; Chen, B.; Tan, M.; Wang, W.; Zhu, Y.; Pang, R.; Vasudevan, V.; et al. Searching for mobilenetv3. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 1314–1324.
85. Zhang, Y.F.; Ren, W.; Zhang, Z.; Jia, Z.; Wang, L.; Tan, T. Focal and efficient IOU loss for accurate bounding box regression. *Neurocomputing* **2022**, *506*, 146–157. [CrossRef]
86. Wang, X.; Peng, Y.; Lu, L.; Lu, Z.; Bagheri, M.; Summers, R.M. Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 2097–2106.
87. Liu, S.; Huang, D.; Wang, Y. Learning spatial fusion for single-shot object detection. *arXiv* **2019**, arXiv:1911.09516.
88. Zhu, H.; Qi, Y.; Shi, H.; Li, N.; Zhou, H. Human detection under UAV: An improved faster R-CNN approach. In Proceedings of the 2018 5th International Conference on Systems and Informatics (ICSAI), Nanjing, China, 10–12 November 2018; IEEE: New York, NY, USA, 2018; pp. 367–372.
89. Cheng, J.; Liu, Y.; Li, G.; Li, J.; Peng, J.; Hong, J. An Efficient Detection Approach for Unmanned Aerial Vehicle (UAV) Small Targets Based on Group Convolution. *Appl. Sci.* **2022**, *12*, 5402. [CrossRef]
90. Xia, G.S.; Bai, X.; Ding, J.; Zhu, Z.; Belongie, S.; Luo, J.; Datcu, M.; Pelillo, M.; Zhang, L. DOTA: A large-scale dataset for object detection in aerial images. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 3974–3983.
91. Lin, Q.; Ding, Y.; Xu, H.; Lin, W.; Li, J.; Xie, X. ECASCADE-RCNN: Enhanced cascade RCNN for multi-scale object detection in UAV images. In Proceedings of the 2021 7th International Conference on Automation, Robotics and Applications (ICARA), Prague, Czech Republic, 4–6 February 2021; IEEE: New York, NY, USA, 2021; pp. 268–272.
92. Wang, M.; Luo, X.; Wang, X.; Tian, X. Research on Vehicle Detection Based on Faster R-CNN for UAV Images. In Proceedings of the IGARSS 2020—2020 IEEE International Geoscience and Remote Sensing Symposium, Waikoloa, HI, USA, 26 September–2 October 2020; IEEE: New York, NY, USA, 2020; pp. 1177–1180.
93. Liu, S.; Liu, H.; Shi, W.; Wang, S.; Shi, M.; Wang, L.; Mao, T. Performance Analysis of Vehicle Detection Algorithm in Aerial Traffic Videos. In Proceedings of the 2019 International Conference on Virtual Reality and Visualization (ICVRV), Hong Kong, China, 18–19 November 2019; IEEE: New York, NY, USA, 2019; pp. 59–64.
94. Avola, D.; Cinque, L.; Diko, A.; Fagioli, A.; Foresti, G.L.; Mecca, A.; Pannone, D.; Piciarelli, C. MS-Faster R-CNN: Multi-stream backbone for improved Faster R-CNN object detection and aerial tracking from UAV images. *Remote Sens.* **2021**, *13*, 1670. [CrossRef]
95. Huang, H.; Li, L.; Ma, H. An Improved Cascade R-CNN-Based Target Detection Algorithm for UAV Aerial Images. In Proceedings of the 2022 7th International Conference on Image, Vision and Computing (ICIVC), Xi'an, China, 26–28 July 2022; IEEE: New York, NY, USA, 2022; pp. 232–237.
96. Youssef, Y.; Elshenawy, M. Automatic vehicle counting and tracking in aerial video feeds using cascade region-based convolutional neural networks and feature pyramid networks. *Transp. Res. Rec.* **2021**, *2675*, 304–317. [CrossRef]
97. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
98. Dalal, N.; Triggs, B. Histograms of oriented gradients for human detection. In Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), San Diego, CA, USA, 20–25 June 2005; IEEE: New York, NY, USA, 2005; Volume 1, pp. 886–893.
99. Barnich, O.; Van Droogenbroeck, M. ViBe: A powerful random technique to estimate the background in video sequences. In Proceedings of the 2009 IEEE International Conference on Acoustics, Speech and Signal Processing, Taipei, Taiwan, 19–24 April 2009; IEEE: New York, NY, USA, 2009; pp. 945–948.
100. Mueller, M.; Smith, N.; Ghanem, B. A benchmark and simulator for uav tracking. In Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 8–16 October 2016; Springer: Cham, Switzerland, 2016; pp. 445–461.
101. Avola, D.; Cinque, L.; Foresti, G.L.; Martinel, N.; Pannone, D.; Piciarelli, C. A UAV video dataset for mosaicking and change detection from low-altitude flights. *IEEE Trans. Syst. Man Cybern. Syst.* **2018**, *50*, 2139–2149. [CrossRef]
102. Yu, H.; Li, G.; Zhang, W.; Huang, Q.; Du, D.; Tian, Q.; Sebe, N. The unmanned aerial vehicle benchmark: Object detection, tracking and baseline. *Int. J. Comput. Vis.* **2020**, *128*, 1141–1159. [CrossRef]