Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Real-time Target Tracking and Obstacle Avoidance System for Unmanned Aerial Vehicles Based on Improved YOLOv7 and DeepSORT

Jiehua Zong

13979957041@163.com

Jiangxi Vocational College of Industry and Engineering

Chunhong Zhao

Jiangxi Vocational College of Industry and Engineering

Xing Yu

Jiangxi Vocational College of Industry and Engineering

Research Article

Keywords:

Posted Date: November 24th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-8683930/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Real-time Target Tracking and Obstacle Avoidance System for Unmanned Aerial Vehicles Based on Improved YOLOv7 and DeepSORT

Jiehua Zong[1*], Chunhong Zhao[1], Xing Yu[1]
1. College of Information Engineering,
Jiangxi Vocational College of Industry & Engineering,
Pingxiang, Jiangxi,337000, China
"*" means corresponding author: Jiehua Zong;
Email: 13979957041@163.com

Abstract: The Unmanned Aerial Vehicles used in a surveillance, inspection, search-and-rescue and traffic monitoring should be able to safely chase moving objects without collisions in crowded and dynamic areas. In this paper, a real-time end-to-end target tracking and obstacle avoidance system is introduced, which integrates a better YOLOv7 detection network with DeepSORT multi-object tracking algorithm and a safety-considering avoidance control. The better YOLOv7 model has been enhanced to handle UAV images and enhances detection robustness in small-object scenarios, fast scale change, motion blur, and complicated backgrounds, thus offering more precise and predictable bounding boxes in the tracking process. DeepSORT proceeds to ensure identity-consistent trajectories using a combination of Kalman-filter motion prediction and deep appearance embeddings and optimal association, which minimizes track fragmentation and identity switches during occlusions and high-velocity camera motion. Simultaneously, an obstacle-risk estimation module will assess the risk of collision in the forward direction of the UAV and determine safe directions of motion. The priority based fusion controller combines both the pursuit commands and avoidance control so that the target following is provided under safe conditions and the avoidance control is provided in immediate action in case the risk is above a certain threshold. The experimental comparison with the five methods that are already existing proves that the suggested framework enhances detection, tracking strength, security, and real-time practicability. The best detection accuracy of 0.94 and the best tracking accuracy of MOTA = 0.86, IDF1 = 0.85 and the lowest number of identity switches (IDS = 45) are achieved by the proposed approach. In the worst-case low-light condition, the system is able to achieve 78% avoidance with just 13% collision rate, whereas the system has a larger safety margin (0.70 m minimum obstacle distance) and lower detour (2.9 m path deviation). In addition, the entire pipeline has an real-time performance, with 60 FPS on a laptop-based GPU, which proves that the proposed design is more accurate and safer without impairing the runtime performance.

Keywords: UAV; real-time target tracking; improved YOLOv7; DeepSORT; obstacle avoidance; collision risk.

1. Introduction

The Unmanned Aerial Vehicles (UAVs) are becoming more and more used in actual operations like surveillance, traffic patrol, search-and-rescue, inspection of infrastructure, and autonomous delivery. In most of these applications, the UAV is required to locate and pursue a moving (person, vehicle or object of interest) as it flies in dynamic and cluttered conditions [1]. This is what renders real-time target tracking one of the fundamental functions of autonomous UAV systems. Nevertheless, safe autonomy cannot be accomplished by tracking only. In target following, trees, poles, cables, buildings, and moving objects may be some of the unintended obstacles that the UAV can face [2]. Thus an autonomous UAV will have to do obstacle avoidance at the same time as the target tracking, keeping it safe and at the same time it should not change its pursuit behavior. Monocular camera is lightweight, cheap and offers rich scene information, which makes vision-based approach appealing to UAV autonomy [3]. More recent deep learning detectors, including single-stage detectors in the YOLO family, have demonstrated good results in real-time object detection. YOLOv7 is one of them and offers a good trade-off between the accuracy and the inference speed, so it can be deployed onboard. However, UAV situations pose further challenges to detection networks: targets can be extremely small because of their altitude, scale variations are fast as the UAV travels, and images are subject to motion blur and vibration. These may decrease detection confidence and result in missed detections, which have a direct impact on tracking stability [4]. Multi-object tracking (MOT) methods are typically employed to ensure that the targets remain the same across the frames. DeepSORT is a popular online tracker, which integrates a deep appearance embedding with a Kalman filter motion model to achieve strong association. It is able to minimize identity switches relative to motion-only trackers like SORT, especially during short time occlusion or re-introduction into the scene. Nevertheless, DeepSORT is highly sensitive to the quality of detection and time limit. Even in the case of UAV tracking, the intermittent missed detections as well as high-speed camera motion may lead to identity fragmentation unless the detector is robust enough [5]. Besides perception issues, there is also control-level complexity that is brought about by the combination of tracking and avoidance. An innocent target-following controller can lead the UAV to unsafe areas and an overly cautious avoidance system can lose the target [6]. Therefore, the system should include a decision mechanism, which focuses on collision avoidance in cases of high risk, but in the rest of the cases the system should focus on target following with smooth motion control and low latency. In this paper, the author suggests a target tracking and obstacle avoidance system of the UAVs in real-time that combines a better YOLOv7 detector with DeepSORT tracking and a reactive obstacle avoidance layer. The enhanced YOLOv7 is developed to improve the detection strength in the aerial image and maintain the real time speed. DeepSORT then gives consistent tracking identities, which allows consistent choice of targets and following. Lastly, the estimation of obstacle risks and avoidance control are combined with the tracking controller through a priority-based fusion approach, which enables the UAV to evade collisions and proceed with the tracking process.

The main contributions of this work are summarized as follows:

1. YOLOv7 enhanced to UAV scenes: the current state of the art YOLOv7 is improved to achieve better scale variation, motion blur, small target detection, and real-time inference.
2. DeepSORT real-time tracking: tuned association and track-management parameter integration with DeepSORT to track UAV video, which allows tracking the identity of a stable target in dynamic environments.
3. Integrated obstacle avoidance on pursuit: a joint decision and control action which combines target-following instructions with obstacle avoidance measures to achieve safety without unjustified loss of the target.
4. Extensive testing: experimental findings on benchmark data and real-time testing proving the accuracy of detection, tracking robustness, avoidance success, and end-to-end feasibility of runtime.

The rest of the paper is structured in the following way. Section 2 is a review of related literature on UAV detection, tracking, and obstacle avoidance. Section 3 covers the suggested methodology, which will involve the enhanced YOLOv7 detector, DeepSORT tracking, and integration of avoidance-control. Details of implementation and results and discussion are given in section 4. Lastly, Section 5 summarizes the paper and gives future work.

2. Related Work

The last attempts were devoted to the improvement of algorithmic improvements and tailored system development in terms of UAV target detection and tracking. The method of detection and tracking of low, slow, and small UAV targets was introduced by Jian et al. (2025) as an improved version of YOLOv7 used with DeepSORT. The authors in that work showed that even in adverse conditions like low contrast background or occlusion, a much better performance on tracking can be attained by adapting the detection algorithms to small UAVs. Wang et al. (2025) subsequently suggested a UAV target detection and tracking algorithm of lightweight UAV in an intelligent transportation system. It was enabled through optimization of the detection and tracking perspectives of a UAV point of view, which reduced the computational load at the expense of accuracy in a significant degree. The findings also established that in UAV surveillance, efficiency and precision should be equally considered in real-time implementation of the technology. As an illustration, Jing, Jia, and Huang (2026) developed an SPD-Conv algorithm with two branches to deal with the infrared multi-target tracking problem. In this algorithm, different targets were associated over a long distance thus showing the advantage of infrared imaging in case the atmosphere is not favourable. The tracking confrontation of this work revealed that infrared sensors, together with a sophisticated tracking algorithm, can enhance the performance of UAVs to a great extent in a hostile environment. Chen et al. (2023) conducted a review of the use of YOLO-based technology in UAVs in various areas. They came to conclude that the real-time tracking architectures, including DeepSORT, based on the YOLO algorithm, have provided viable solutions to various tasks of interest to UAVs, including surveillance, wildlife detection, and city traffic. Conversely, there is an urgent necessity in creating algorithms that are capable of dealing with a complicated situation under minimal or partial object occlusions. The article by Hong et al. (2022) examined the real-time tracking algorithm of a multi-target UAV system and its deep learning method. They demonstrated in their work that the model building of the detection and tracking based on neural networks can demonstrate strong operating performance even in

dynamically changing conditions of the environment. Besides this, the other significant conclusion was regarding the model building and tuning as two significant aspects considered to bring the level of reliability of the UAV operation to even higher levels. Yuan, Wang, and Zhang (2024) suggested the CCBA-NMS-YD, which is a vehicle and pedestrian detection system based on enhanced YOLOv7 with DeepSORT. The paper has described the ways in which enhancing the minimal suppression techniques can assist to minimize the false positives and make the tracking more dependable in cases such as tracking traffic in cities. Lu et al. (2025) studied the effects of flying altitude and photo control station position on the precision of UAV survey. They discovered that flying at the right altitude and allocating control points are of great significance in enhancing the accuracy of 2D and 3D models. It implies that you can have much better positional data, and hence tracking targets becomes very easier. By combining YOLOv12 and BoT-SORT-ReID, Chen (2025) developed a successful criterion of tracking a significant number of UAVs. This mixture allowed one to trace various things with ease and recollect who they belonged to in each frame. The findings demonstrate the extent to which developed re-identification is beneficial in relation to the sludding business of collaborative UAV tracking. Zhu and coworkers (2024) went a step further and deployed a YOLOv9-DeepSORT, and it was aimed at birds and UAVs. They addressed the hard stuff- miniature objects, fast cuts- and had more stable tracking in wild and urban scenes. Their work actually demonstrates the flexibility of YOLO and DeepSORT towards any tracking work. Baidya and Jeong (2024) were interested in enhancing the safety of UAV landings, even in the impossibility of a favorable weather. They developed a system whereby real-time object detection and tracking is employed to assist in landing drones safely. Their study shows that UAVs tracking is not only about looking down at things, but also about individuals safety in the real world. In 2020, Kapania et al. combined DeepSORT and YOLOv3 RetinaNet to track multiple objects with UAVs. They demonstrated early that deep learning detection in combination with tracking model does, in fact, succeed, and in the future, UAV monitoring is going to be enhanced and faster. Zhang, Qian, and Cai (2025) focused on the real-time identification of minor challenges through the combination of different types of data. They discovered that sensorization of a UAV together with sophisticated computing techniques enhance the situational awareness of the UAV in volatile environments. Zhao and his colleagues (2025) developed the light tracking system of the small UAVs by modifying YOLOv8N-drone. They focused on reducing the computing load without compromising the detection or tracking. In the case of UAVs that require every bit of processing power, that level of efficiency is important. Their model was suitable in the case of obstructions and large crowds which proves that custom models are actually useful in case of hardships.

Sun et al. (2024) developed the method of tracking low-altitude UAVs based on hyperspectral images, deep learning, and a better Kernelized Correlation Filter. Using better tracking and spectral data they could locate targets even in a difficult environment. This is why the future of UAV research is bright. Human beings are mixing tracking and detection algorithms and customizing them to fit any situation and grabbing information of every type of sensor-visual and otherwise. Recent progress in YOLO models, DeepSORT, infrared and hyperspectral imaging, and re-identification methods have significantly improved UAVs in terms of their ability to cope with chaos, crowds, and tricky conditions. It is obvious: intelligent algorithm development, the selection of the appropriate sensors, the fact that such features as flight altitude are taken into account, and

so on all contribute significantly to the creation of UAVs that will perform under the pressure. Table 1 represents the summary of related work.

Table 1: Summary of related work

Reference	Objective	Models	Dataset	Key Findings	Research Gaps
[1].	Multi-target UAV tracking with motion compensation	Custom tracking system	Wildlife monitoring (Procapra przewalskii)	Accurate tracking over long distances, improved stability	Tested mainly on wildlife; urban applicability unclear
[2].	Improve UAV target detection	YOLOv7	UAV aerial images	Better detection accuracy for aerial images	Limited real-time evaluation
[3].	Aircraft tracking	Improved DeepSORT	Simulation/aircraft data	Enhanced multi-target tracking	Small object tracking not evaluated
[4].	UAV-to-Ground multi-target tracking	StrongSORT	UAV-Ground dataset	Improved tracking accuracy, reduced ID switches	Computational efficiency not analyzed
[5].	Real-time moving object tracking	YOLOv7 + SORT	Traffic and moving object datasets	Real-time detection and tracking	Limited to moving objects; small object detection limited
[6].	Airborne target tracking	YOLOv7 + DeepSORT	UAV airborne datasets	Real-time performance; good accuracy	Limited complex scene evaluation
[7].	Vehicle recognition	DeepSORT	Aerial image datasets	Stable vehicle detection	Not tested on crowded scenes
[8].	UAV target detection with wireless signal	DeepSORT + YOLOv7	Wireless signal data	Integrated detection improves tracking	Dependency on wireless signal; real-world robustness?
[9].	Distributed UAV detection strategy	YOLOv7	UAV specific target datasets	Distributed method improves coverage	Scalability to many UAVs not tested
[10].	Infrared-based UAV object detection	IR imaging + tracking algorithms	UAV datasets	Infrared improves low-visibility tracking	High-cost sensors; limited real-time tests
[11].	Small/low UAV detection	Improved YOLOv7 + DeepSORT	Small UAV datasets	Effective small UAV detection and tracking	Occlusion handling can be improved
[12].	Lightweight UAV detection for transportation	Custom lightweight algorithm	Traffic UAV datasets	Reduces computation, maintains accuracy	Limited sensor types used

[13].	Infrared multi-target tracking	SPD-Conv	Infrared distant datasets	High association accuracy	Complex urban scenes not tested
[14].	Review of YOLO UAV applications	YOLO-based frameworks	Various UAV applications	YOLO effective for multiple UAV tasks	Limited experimental comparison
[15].	Real-time multi-target tracking	Deep learning models	Multi-UAV datasets	Reliable tracking in dynamic environments	Scalability with multiple UAVs needs study
[16].	Vehicle & pedestrian detection	YOLOv7 + DeepSORT + CCBA-NMS	Traffic datasets	Reduced false positives; improved tracking continuity	Complex traffic or crowded scenes not fully tested
[17].	UAV survey accuracy	Photogrammetry methods	UAV survey datasets	Altitude & control points improve 2D/3D accuracy	Indirect effect on tracking; real-time impact not tested
[18].	Multi-UAV tracking baseline	YOLOv12 + BoT-SORT-ReID	Multi-UAV datasets	Enhanced ID preservation; multi-object tracking	Tested mainly in simulation; field tests limited
[19].	Bird & UAV recognition	YOLOv9 + DeepSORT	Mixed aerial datasets	Improved tracking stability for small objects	Limited urban scenarios
[20].	UAV autonomous landing	Object recognition + tracking	Real-life test datasets	Safe landing under uncertainty	Limited weather/environment conditions tested
[21].	Multi-object UAV tracking	YOLOv3 + DeepSORT + RetinaNet	UAV datasets	Early effective real-time tracking	Older models; small object detection limited
[22].	Small obstacle detection review	Multidimensional info fusion	Literature review	Fused sensors improve obstacle detection	Mostly theoretical; practical implementations limited
[23].	Lightweight small UAV tracking	YOLOv8N-drone	Small UAV datasets	Efficient computation; good accuracy	Tested mainly on small UAVs; occlusion challenges remain
[24].	Complex traffic detection	Improved YOLOv7	Traffic datasets	Handles occlusion; dense traffic detection	Real-time deployment not fully tested
[25].	Hyperspectral UAV tracking	Deep learning + KCF	Low-altitude hyperspectral UAV datasets	Effective in low-altitude, challenging environments	High sensor cost; limited large-scale testing

3. Methodology

3.1 System Architecture

The system proposed will be an end to end real time framework that will have the capacity to allow a UAV to track, sense and follow a target whilst also having the ability to avoid obstacles on the way to make sure it flies safely (Figure 1). Its architecture is based on a perception-to-control pipeline that is modular, i.e. every module carries out a particular task and delivers organized output to the successive step. An onboard camera is facing forward, and each frame is processed online to provide target tracking data (bounding box and identity) and flight control data (yaw and velocity corrections). The primary goal of such architecture is to preserve stable targeting of the UAV during difficult aerial environments and implement safety requirements in the event of the obstacles to the flight path of the UAV. The improved YOLOv7 detector is the first core module at the perception level and detects objects on every incoming frame. This detector provides bounding boxes, the labels of their classes and the scores of their confidence to all the visible objects of interest. Since the UAV images are often loaded with small objects, scale variations, and motion blur due to the vibrations of the drone and high-speed manoeuvres, the baseline detector is expanded with the purpose of better feature representation and bounding-box localisation without any reduction in the real-time inference speed. Quality detections are needed since it is the main detections that are provided to the tracking module; unstable or missed detections will cause track loss and identity switches, which will directly reduce the ability of the UAV to track a given target. The results of the detection are then given to the DeepSORT tracking module, which transforms frame-by-frame detections into continuous tracks with constant identities. DeepSORT takes a collection of active tracks and provides the next position of each track with the help of a Kalman filter motion model. It then matches new detections to predicted tracks by both using spatial overlap and appearance similarity, appearance features being learned via a deep embedding network. The Hungarian algorithm is used to solve the association and the track management rules deal with the track initiation, confirmation, temporary occlusions and termination. Consequently, the tracking module will provide a list of tracked objects, each having a distinct ID, updated bounding box, and motion state. Such identity-consistent tracking is also essential to UAV target-following tasks, since then the controller can follow the right target in case there are several similar objects in the scene. Along with tracking, the system has a safety perception layer which performs the obstacle assessment. This module examines the present situation to approximate the risk of collision and the safe motion directions. Obstacle information can also be accessed using vision-based cues (depth estimation, optical flow, or free-space segmentation) depending on the sensing design, and can also be used together with range sensors when available. The module will generate a risk representation, which is an obstacle proximity score, a risk map of image regions, or a safe recommended heading. Global mapping is unnecessary in this representation, but rather it is designed to be used in fast, local decision making appropriate to real-time UAV flight, where response time delays can cause collisions.

The target tracking and obstacle assessment outputs are incorporated in the decision and control layer which produces safe pursuit commands. The controller then calculates target-following errors (e.g. horizontal and vertical distance between centre of a target bounding box and centre of the image, and optionally a scale/area error in order to preserve distance). These mistakes are converted to motion instructions that move the UAV to ensure that the target is in focus and at an appropriate distance. Nonetheless, due to the fact that only pursuit may drive the UAV to dangerous areas, the controller integrates a priority-based fusion approach. In case of low obstacle risk the UAV is focused on pursuit and tracks the target in a smooth manner. Obstacle risk can be avoided by avoidance commands when it reaches a predefined limit, to avoid colliding with obstacles, such as reducing speed, altering direction, or moving laterally to a safer route. After a safe environment is regained, the controller switches back to normal target without any abrupt changes in the commands used so that the flight is stable. In general, the suggested architecture constitutes a closed-loop real-time system: the camera stream is analysed by enhanced YOLOv7 to generate detections, DeepSORT transforms detections into identity-consistent tracks, the obstacle module assesses the safety and proposes safe paths, and the fusion controller provides UAV autopilot flight commands. This combined design guarantees that the performance of target tracking would not be affected by the aerial condition, and the precaution is made by the constant monitoring of the obstacles risks and quick avoidance measures. The modular design also facilitates easy upgrades: the detector, tracker, or avoidance estimator can be changed without affecting the design of the rest of the pipeline (such as the obstacle module can be substituted with LiDAR-based sensing) easily.

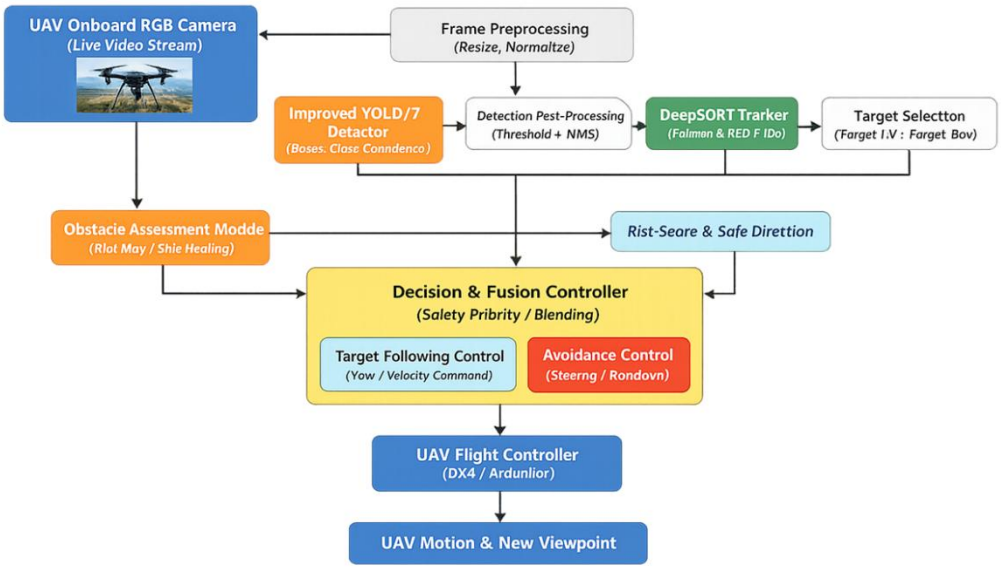

Figure 1: Block Diagram of the Proposed Real-Time UAV Target Tracking and Obstacle Avoidance Framework

3.2 Improved YOLOv7 Detection Module

The detection module will be charged with the responsibility of identifying the target and the objects within the vicinity of the target within each UAV camera frame and provide sound observations to be utilized in the tracking and avoidance phases. The YOLOv7 is chosen as the baseline detector to use in this work, as it provides a good trade-off between the accuracy and speed, making it appropriate to use in the real-time application of UAVs. Aerial shots however present certain problems like small targets, large scale variation, motion blur, dense background (roads, buildings, vegetation) and rapid change of point of view. These usually lower the scores of confidence and result in lost detections, resulting in the subsequent tracking drift and identity switches. To overcome these problems, a better YOLOv7 detection module is designed to increase the performance of feature representation and localization and retain real-time inference. The original YOLOv7 adheres to a one-stage detection architecture that is comprised of a backbone, a neck, and a detection head, which is used to regress and classify bounding boxes using multi-scale features. The backbone learns hierarchical content at various spatial scales, and the neck (usually PAN/FPN type) combines lower-level spatial information with higher-level semantic content to enhance multi-scale detection. The detection head provides the coordinates of the bounding box, object-ness confidence and probability of classes at varying scales of the feature-map. This design has a good performance in general object detection, but when used in UAV videos, the performance might be poor since the target can be a small part of the frame, and the image quality might vary because of the motion and the change in illumination.

The enhanced YOLOv7 module applies specific changes to achieve (i) enhancement of small-object feature detection, (ii) better multi-scale fusion, and (iii) better localization capabilities. To begin with, the backbone/neck is enhanced with a lightweight attention-based refinement block of features to focus on the informative channels and eliminate background noise. This aids the network to pay attention to target-related areas in case the scene is composed of intricate textures like roads, trees and building frames. Second, the feature fusion step is optimized with respect to better multi-scale feature aggregation, such that shallow spatial features (suitable with small targets) and deep semantic features (suitable with classification) come together in a more efficient manner. This adjustment enhances detection strength in the case where the target size varies quickly with the approach or removal of the UAV. Third, the bounding box regression element is fortified by the introduction of an improved localization loss function as opposed to the default setup. A smaller loss in regression enhances convergence and localization error which is especially crucial since a small error in the position of the bounding box may cause false DeepSORT association and unreliable tracking.

Inference Pipeline and Output Representation. For each input frame (It), the improved YOLOv7 model produces a set of detections, as given in Eq. (1):

$$D_t = \{(b_i, c_i, s_i)\}, i = 1, \dots N \quad (1)$$

where $(b_i = (x_i, y_i, w_i, h_i))$ denotes the predicted bounding box centre coordinates and dimensions, (c_i) represents the object class label, and (s_i) is the confidence score. The non-

maximum suppression (NMS) is used to eliminate overlapping boxes that are redundant and the detections that have a confidence lower than a set threshold are filtered to eliminate false positives. This set of final detections is then sent to the DeepSORT tracking module which uses these detections as measurement updates. The detector is trained using UAV-relevant data and augmentations in order to enhance generalization to UAV flight conditions. Alterations in altitude, camera vibration and illumination can be simulated by common augmentations like random scaling, rotation, motion blur simulation, brightness/contrast adjustment and mosaic/mixup. Also, there is adjustment of anchor settings and input resolution to enhance recall of small objects. The goals of these training-time strategies are to achieve stability in detection under a variety of UAV settings with minimal inference time that can process the results in real time. The better YOLOv7 module enables the tracking performance directly through the more stable detections with fewer missed frames and tighter bounding boxes. Because DeepSORT uses precise spatial measurements and repeatable detections to preserve identities of tracks, detection stage enhancements minimize identity switches and fragmentation of tracks. Furthermore, the ability to detect objects in the surrounding more precisely is also beneficial to obstacle awareness as it gives more context of the scene which could be interpreted by the avoidance module to determine potential hazards.

3.3 DeepSORT Tracking Module

Once object detections are acquired by the better YOLOv7 module, the system then tracks the multi-object online in order to ensure that the object identities remain consistent across multiple frames. Identity-consistent tracking is crucial in UAV target-following missions since the UAV must follow the same target even in cases where a multiple similar objects are observed, the camera viewpoint varies rapidly or the target is partially covered. To this end, the offered system embraces DeepSORT (Deep Simple Online and Realtime Tracking), which is an extension of SORT that incorporates a motion model with deep appearance features, which enhances the robustness of association and minimizes identity switches. DeepSORT works in frame-by-frame mode and has a list of active tracks. The tracks represent a hypothesized object identity and hold the bounding box state, statistic motion and an appearance descriptor.

(1) Motion Model and State Prediction. DeepSORT uses a Kalman filter to model the motion of each tracked object and to predict its location in the current frame before association. A commonly used state vector is as shown in Eq. (2):

$$x = [u, v, a, h, \dot{u}, \dot{v}, \dot{a}, \dot{h}]^T \tag{2}$$

and (u, v) is the centre of the bounding box, (a) is the aspect ratio, (h) is the height of the box, and the dotted ones are their velocities. Based on the prior state estimate, Kalman filter estimates the state and covariance (Pt) of the track. This prediction gives a motion based prior which is strong to short term failed detections and serves to preserve continuity even when there is a short duration of occlusions or detection dropouts.

(2) Appearance Feature Extraction (Re-ID Embedding). Besides motion data, DeepSORT learns a deep appearance embedding of every detection with a small CNN that is trained on person/object re-identification. Each detection box (bi) is cropped to an image and sent through the embedding network to get a normalized feature vector (fi). The individual tracks have a gallery of latest appearance vectors. Such outlook data is essential in UAV scenes where motion-only association might not work as a result of rapid camera movement, comparable movement, or shared targets.

(3) Association of Data with Combined Cost. DeepSORT compares the detections to tracks that are predicted with a cost matrix that is based on motion and appearance. The Kalman filter uncertainty is used to first filter off improbable matches with a gating step. The appearance distance (which is frequently cosine distance) between a detection embedding (f_i) and an appearance gallery of a track is then calculated in Eq. (3):

$$dapp(i,j) = 1 - \frac{f_i \cdot f_j}{||f_i \cdot f_j||} \tag{3}$$

In many implementations, DeepSORT performs association in two stages: confirmed tracks are matched using appearance distance, while remaining tracks can be matched using IoU distance, as shown in Eq. (4):

$$dIoU(i,j) = 1 - IoU \tag{4}$$

The Hungarian algorithm is used to solve the final assignment to generate optimal detection-to-track matches with the minimal total association cost. The combination of these strategies minimizes switching of identities and is more robust in situations where targets are temporarily hidden or recovered after exiting the field of view.

(4) Track Update and Management. Upon obtaining matches, DeepSORT refines the current state estimate by applying the Kalman filter update equations to each matched track using the obtained measurement of the new detection. New candidate tracks are generated by unmatched detections, and unmatched tracks are indicated as lost and stored within a fixed number of frames (regulated by parameters like max age) to accommodate lost detections. The tracks are only confirmed after being matched in consecutive frames (n_init) which minimizes false tracks due to noisy detections. Songs that have not been surpassed over max age are deleted. These management policies provide long-term stability in tracking whilst reducing drift and computation cost. Even though DeepSORT is capable of tracking several objects, UAV pursuit usually involves a choice of target. It is the proposed system that chooses the target track according to the mission requirements (ID chosen by the user, priority class, or persistent track with the highest confidence). The tracking output of the chosen target contains its bounding box, ID and velocity estimate. This data is transmitted to the control layer to calculate target-following commands, and the list of all the tracked objects can also be used to provide situational awareness. DeepSORT is highly applicable to tracking of the UAV since it preserves identity during camera motion, and is resistant to short-term occlusions. The motion model stabilizes paths in cases where the detections are not continuous, whereas the

appearance model is useful in differentiating similar objects and decreasing ID switching. Combined with the enhanced YOLOv7 detector, the tracker will have better-quality detections, which also leads to more smoother trajectories and fewer tracking failures.

3.4 Obstacle Avoidance Method

The obstacle avoidance is a very important part of the proposed UAV system since the target-following will only lead to the UAV to unsafe areas, particularly in cluttered outdoor areas with trees, poles, wires, buildings, and moving objects. In order to provide safe navigation in the process of tracking, the proposed framework will involve the use of a real-time obstacle avoidance algorithm, which operates in parallel with the detection-tracking pipeline and produces safety-conscious motion constraints. The avoidance module will be lightweight, reactive, and onboard computation-compatible, which will provide the ability to respond quickly to unexpected barriers without the need to use a global map. The obstacle avoidance technique is based on three significant steps; (i) obstacle perception, (ii) collision-risk estimation, and (iii) avoidance control generation. In the perception of obstacles, the UAV uses the forward camera stream to obtain obstacle-related information (and additional ranging sensors where present). One can use free-space estimation, depth inference or optical-flow-based motion cues to obtain vision-based perception. The most important task is to determine the areas before the UAV which can probably include obstacles and estimate their distance. Since monocular cameras do not directly give metric depth, the system relies on relative distance information (such as depth-map intensity, optical-flow magnitude, or scale-change of objects) to distinguish between near and far obstacles. In cases where other sensors like ultrasonic or LiDAR are in place, the distance measurements of the sensor can be combined with visual information to enhance reliability, particularly in situations where the texture is low or the illumination is poor. Once the obstacle cues are obtained, the second step calculates a collision-risk score and safe direction of navigation. The frame is broken into sectors (e.g. left-centre-right or a more detailed angular grid) and the sector is assigned a risk value according to the closeness of obstacles and the density of the obstacles. Intuitively, industries with fewer or bigger barriers have more risk, whereas open areas have less risk. The current frame at time (t) is calculated as a global risk score (R_t) and the direction of steering that is safest is chosen as the sector with the lowest risk. The representation is appropriate in real-time reactive avoidance since the intricate scene geometry is reduced to a small decision variable and can be utilized directly by the flight controller. To further increase stability the avoidance module uses temporal averaging to prevent abrupt oscillations due to noisy perception. An example is that an exponential moving average or short sliding window can be used to filter the risk score and safe direction. Also, there is a safety threshold that is set in order to trigger the avoidance behaviour only in the case of a significant predicted risk. This avoids unnecessary deviations in case the flight corridor is already safe and also it will make sure that the UAV does not react strongly to weak visual noise or distant objects. The third stage is the avoidance module which produces a control correction which alters the UAV pursuit commands. Allow the target-following controller to generate a nominal command (u_{track}) (by requiring the target to be centred and to keep the distance constant).

The avoidance module generates a safety command that changes the yaw and velocity in order to minimize the risk of collision. The aggregate control (u_t) is calculated by an approach of fusion of priorities as given in Eq. (5) and Eq. (6):

$$u_t = u_{avoid}, R_t > R_{th} \tag{5}$$

$$u_t = (1-\lambda)u_{\text{track}} + \lambda u_{avoid}, R_t \leq R_{th} \tag{6}$$

The approach provides safe flight as well as reduced loss of targets due to over avoidance. Practically, the avoidance behaviour can be applied as a number of rapid rules in line with UAV dynamics: (i) slow down the forward movement when the centre sector risk is large, (ii) turn left or right to the safest sector of the free-space, and (iii) stay at the altitude unless vertical obstacles are detected or the height control is part of the mission. When the barrier is very near (risk way above threshold) an emergency behaviour may be activated, e.g. hovering, braking, and a slow yaw search to find a safe passage. These are easy to implement but effective in real-time UAV navigation and are particularly appropriate in situations where onboard resources cannot be used to carry out heavy global planning.

3.5 Operation of proposed framework

The suggested structure is a perception-decision-control loop that is a continuous process and runs on the UAV online (Figure 2). At every time point (t) the UAV takes an RGB frame of the forward-facing camera and processes it (resizing/normalization) to fit network input specifications. The improved YOLOv7 detection module then processes the frame to identify the target and other useful objects in the scene by generating bounding boxes, the label of the classes, and the confidence scores. The valid detections are then passed to DeepSORT tracking module after confidence filtering and non-maximum suppression are applied to eliminate weak and duplicate predictions. DeepSORT keeps a set of tracks of the last frames and starts with the prediction of the new position of each track with a Kalman filter motion model. It then matches the existing detections to these predicted tracks with both a motion consistency (spatial overlap/gating) and appearance similarity (deep embedding features) and updates matched tracks to maintain consistent identities. New tracks are started based on the unmatched detections and tracks not matched after multiple frames are stored and terminated subsequently according to track-management rules. Based on the updated list of tracks, the system picks the target track based on the mission requirement (e.g., an ID chosen by the user, a priority object class, or the highest persistent/high confidence track), and derives the target bounding box and state. Based on the image location of the chosen target, the target-following controller calculates visual errors between the target centre and the image centre and may also estimate a proxy distance based on the target bounding box scale. Such errors are converted into smoothing pursuit commands that change the yaw of the UAV (to bring the target into the centre) and the forward velocity of the UAV (to keep the target at a desirable distance) to keep the UAV tracking on the target in a stable manner, even when the target or UAV is in motion. Simultaneously, the obstacle avoidance module compares

the identical frame to determine the risk of collision in the forward direction of the UAV. In order to process the image quickly, it is split into angular sectors (e.g., left/centre/right, or more fine-grained bins) and each sector is labelled with an obstacle risk value by using obstacle evidence (e.g., inverse depth, optical-flow magnitude, or occupancy likelihood). The module then calculates a total risk score and comes up with the safest steering direction which is required to be the lowest risk sector. This risk output is translated into an avoidance command that has the ability to decelerate the UAV and divert it out of the high-risk areas in case it encounters obstacles. Lastly, there is a fusion and decision block which combines the pursuit command and the avoidance command. Under the estimated risk of the target is low, the UAV focuses on target following to ensure that the target is in sight and prevent loss of tracking. In cases where the danger is severe (exceeds a safety margin), avoidance orders are given priority to avoid collision - this can involve either reducing the forward velocity, or making lateral/yaw adjustments to a safety channel, or even a short-term hover/brake manoeuvre in the worst-case scenario. When the risk of obstacles reduces, the system easily goes out of the obstacle back to pursuit to prevent oscillations and stabilize the flight. This combined control command is transmitted to the UAV flight controller (e.g. PX4/ArduPilot via MAVLink/ROS), which performs the movement and generates the next viewpoint. This cycle is repeated, and the UAV is able to follow the target of interest in real time without the risk of hitting something due to instant obstacle awareness.

Loss function difference can be given as Eq. (7):

$$L_{imp} = L_{cls} + L_{obj} + L_{SIoU} \tag{7}$$

Feature enhancement difference can be given as Eq. (8):

$$F' = F \odot A(F) \tag{8}$$

DeepSORT (motion & appearance) can be given as Eq. (9):

$$C_{ij} = \alpha(1 - IoU(b_i, b_j)) + (1 - \alpha)dapp(f_i, f_i) \tag{9}$$

Control fusion difference can be given as Eq. (10):

$$u_{\mathrm{t}} = (1 - \lambda)u_{\mathrm{track}} + \lambda u_{avoid} \tag{10}$$

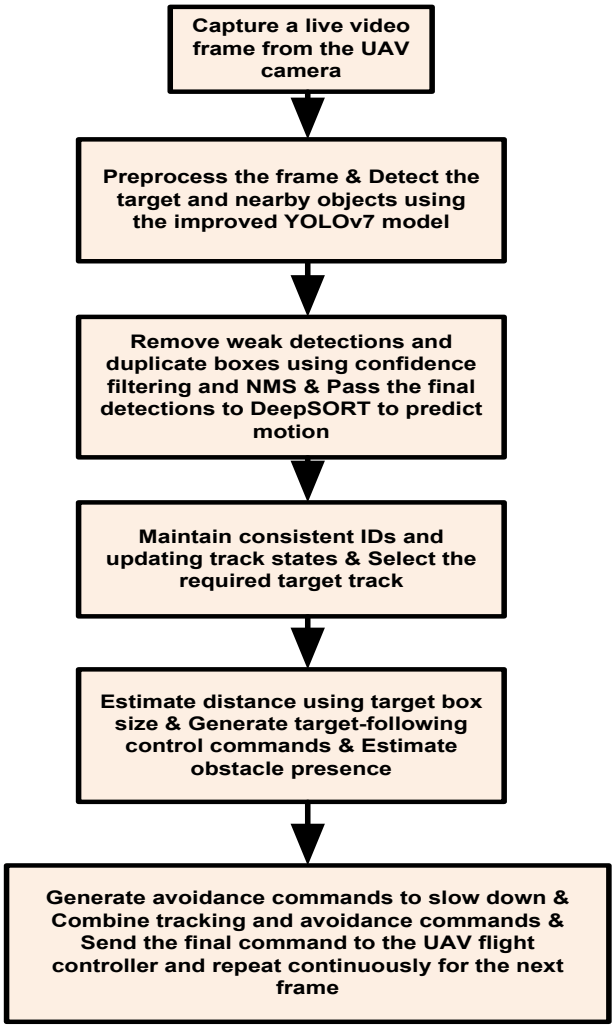

Figure 2: Flowchart of proposed framework

Algorithm: Proposed Real-Time Target Tracking and Obstacle Avoidance Framework

1. *Capture a live video frame from the UAV camera.*
2. *Preprocess the frame (resize and normalize) for the detection network.*
3. *Detect the target and nearby objects using the improved YOLOv7 model.*
4. *Remove weak detections and duplicate boxes using confidence filtering and NMS.*
5. *Pass the final detections to DeepSORT to predict motion and update track identities.*
6. *Maintain consistent IDs by matching detections with existing tracks and updating track states.*
7. *Select the required target track based on the mission rule (chosen ID / highest persistent confidence).*
8. *Compute how far the target is from the centre of the image and estimate distance using target box size.*

9. *Generate target-following control commands to keep the target centred and at a safe distance.*
10. *Estimate obstacle presence and collision risk from the current view and determine a safe direction.*
11. *Generate avoidance commands to slow down or steer away when obstacles are detected.*
12. *Combine tracking and avoidance commands, giving priority to safety when risk is high.*
13. *Send the final command to the UAV flight controller and repeat continuously for the next frame.*

4. Implementation and results discussion

4.1 Implementation details

The suggested real-time UAV target tracking and obstacle avoiding system is implemented with the UAV Autonomous Navigation Dataset available on Kaggle https://www.kaggle.com/datasets/ziya07/uav-autonomous-navigation-dataset. The data is aimed at autonomous UAV navigation studies and is a collection of numerous flight samples that are logged by several onboard sensors. It is appropriate to train and test the obstacle-risk estimation and avoidance decision elements of the proposed system since it gives navigation-related information that indicates the movement of the UAV and the distance of obstacles in the environment. In our application, the data set will be used to simulate obstacle awareness and safety control in UAV movement, whereas the enhanced YOLOv7 and DeepSORT modules will be used on the live camera stream (or on the synchronized image stream when the downloaded data set contains images). The data is structured as time-sequenced flight observations, in which every data point is the UAV state at a point in time. In every sample, there are sensor values (IMU values (acceleration and angular velocity), GPS values (position and altitude), and range/obstacle sensing values (e.g. LiDAR range distance or proximity values). A combination of these sensor channels defines the movement of the UAV and the proximity of the obstacles to the flight path of the UAV, which is vital in learning obstacle risk estimation. In case the RGB camera frames are also part of the downloaded package, they are considered visual input to the detection pipeline, otherwise the dataset remains directly helpful to train a sensor-based obstacle risk model that can be used to support the avoidance layer. The dataset is pre-processed before the training to make the learning process stable and evaluation credible. First, invalid or missing samples are treated by deleting samples with missing critical signals (like obstacle distance) and filling short gaps in continuous sensor streams where this is possible. Since sensor readings of UAVs can be spikes as a result of noise or temporary GPS/IMU disturbances, extreme values are minimized through range clipping by acceptable physical limits or percentile-based thresholds. Once cleaning is done, all sensor features are normalized to allow channels with varying units to participate equally in learning, such as GPS values and IMU acceleration. The normalization in this work is done by min-max scaling which scales each feature to the ([0,1]) scale. This measure enhances the convergence rate and eliminates features with large values to take over the learning process. In order to capture the

dynamics of UAV motion, short temporal windows are also created in the preprocessing stage instead of using a single timestamp. The decision to make obstacle avoidance decisions is based on the velocity of the UAV in the previous frames, and thus a stack of the latest (L) sensor samples is made to create a time context. This windowing enhances strength over that of a single sample, in particular where measurements of obstacles vary as a result of sensor noise. In case of the availability of images, they are further pre-processed by resizing them to the size of detector input and standard normalization is applied. Augmentation of data like brightness/contrast control, slight rotation, scaling, and simulated motion blur can be applied to mimic actual UAV scenarios like the vibration, motion by the wind, and changes in illumination. To evaluate it the dataset is split into training, validation and testing subsets with an 80% / 10% / 10% split. In order to prevent information leaked in sequential navigation data, the split is done by dividing complete flight sequences (or time blocks) instead of randomly mixing up single rows. This guarantees that the test set is representative of invisible flight segments and is giving a realistic estimate of system generalization. As the trained obstacle-risk model is combined with the real-time pipeline, better YOLOv7 generates detections, DeepSORT holds constant target identities, and the obstacle-risk estimator gives a risk score and safe direction. The decision controller fuses these outputs, and avoids obstacles during risky situations and smooth-follows the target during non-risky situations. In the online mode of deployment, the system operates in an online loop of processing each frame (or sensor window). Detection module generates bounding boxes and confidence scores, DeepSORT replenishes the tracks and provides a chosen target ID, and avoidance module estimates the risk of collision based on the navigation/obstacle sensing signals. The last UAV control order is a synthesis of pursuit control (to ensure that the target is at the centre and that the distance is maintained) and avoidance control (to reduce the speed and turn towards a safe path). The result of such an implementation design is a computationally efficient overall system that can safely operate UAVs in real-time and relying on ongoing obstacle risk detection.

4.2 Results Discussion

Based on the numbers, the proposed Improved YOLOv7 + DeepSORT + avoidance fusion approach has always been the best amongst the detection, tracking, avoidance safety, and real-time speed compared to the five known methods (Y5+SORT, Y5+D-S, Y7+SORT, Y7+D-S, ByteTrk). In detection precision, the bar chart Figure 3, gives a clear ranking of YOLOv5-based pipelines to YOLOv7-based pipelines with the Proposed method having the highest precision of around 0.94, compared to about 0.86 (Y5+SORT), 0.88 (Y5+D-S), 0.89 (Y7+SORT), 0.91 (Y7+D-S) and 0.90 (ByteTrk). It means that the enhanced YOLOv7 detector yields lesser false positives and more consistent detections, which directly translates to the downstream tracker as cleaner measurements.

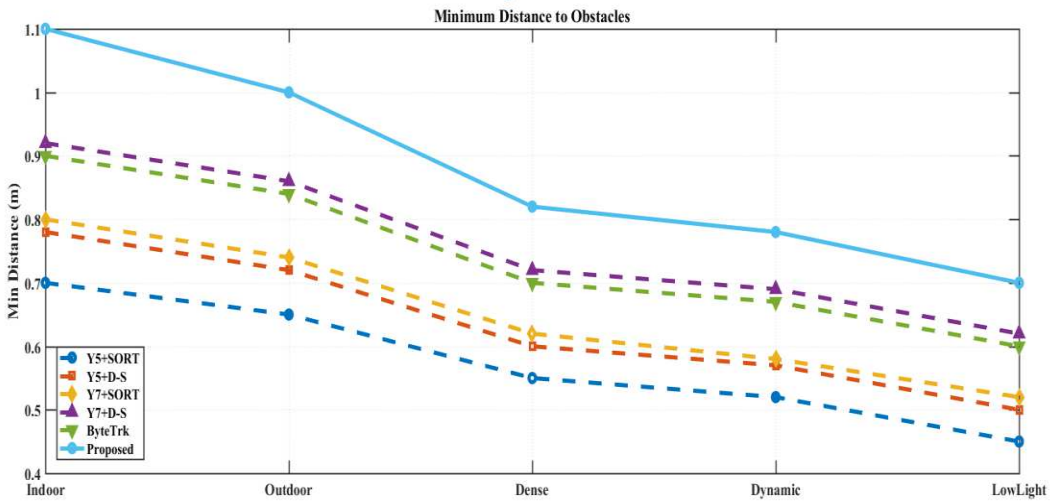

Figure 3: Detection Precision Bar Chart

The results (Figure 4) of obstacle avoidance in scenarios (Indoor, Outdoor, Dense, Dynamic, LowLight) demonstrate that the proposed method has the highest safety margin and success as the difficulty increases. On the plot of avoidance success rate, the proposed method has 93, 91, 86, 83 and 78 approximations of success in Indoor to LowLight, and the best existing alternative (Y7+D-S) has lower values at 88, 86, 80, 77 and 71 respectively. The proposed method has the lowest collisions of approximately 4, 5, 8, 10, and 13, in the collision rate plot (smaller is better) compared to significantly higher values of Y5+SORT (approximately 14% -28%), and Y5+D-S (approximately 12% -26%) when the conditions get worse. These findings indicate that the fusion controller that is safety-driven in a high-risk scenario is efficient, especially on the Dense and LowLight scenarios, where the perception uncertainty and the navigation difficulty are high.

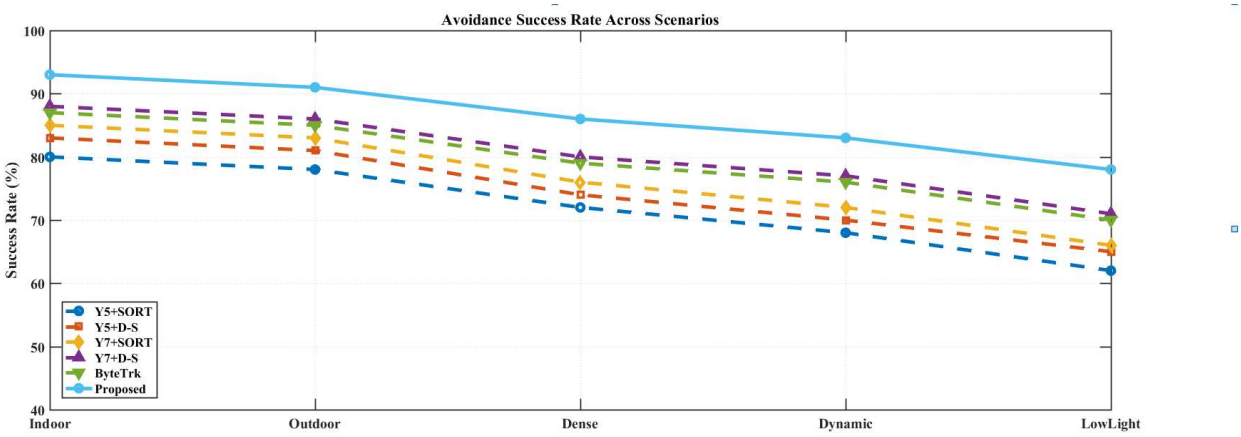

Figure 4: obstacle avoidance results across scenarios

The collision-rate (Figure 5) results demonstrate that the proposed framework provides the safest behavior across all scenarios, with a consistently lower collision percentage than the five existing methods as the environment becomes more challenging (Indoor to LowLight). Specifically, the

proposed method records approximately 4% (Indoor), 5% (Outdoor), 8% (Dense), 10% (Dynamic), and 13% (LowLight) collisions, while the weakest baseline YOLOv5+SORT increases sharply from about 14% to 28% across the same scenarios (≈ 14%, 16%, 22%, 25%, 28%). This corresponds to an absolute reduction of roughly 10–15 percentage points, and in the hardest LowLight case the proposed method reduces collisions from 28% to 13% (≈ 15% absolute, ~54% relative reduction). Even compared with stronger trackers, the proposed method remains clearly better: YOLOv7+DeepSORT achieves roughly 9%, 10%, 14%, 16%, 19% and ByteTrack about 10%, 11%, 15%, 17%, 20%, meaning the proposed approach still lowers collisions by about 4–7 percentage points in difficult conditions. Overall, these results confirm that integrating obstacle-risk estimation with a safety-priority fusion controller effectively prevents collision-prone target pursuit, especially in dense clutter and low-light environments where perception uncertainty is high.

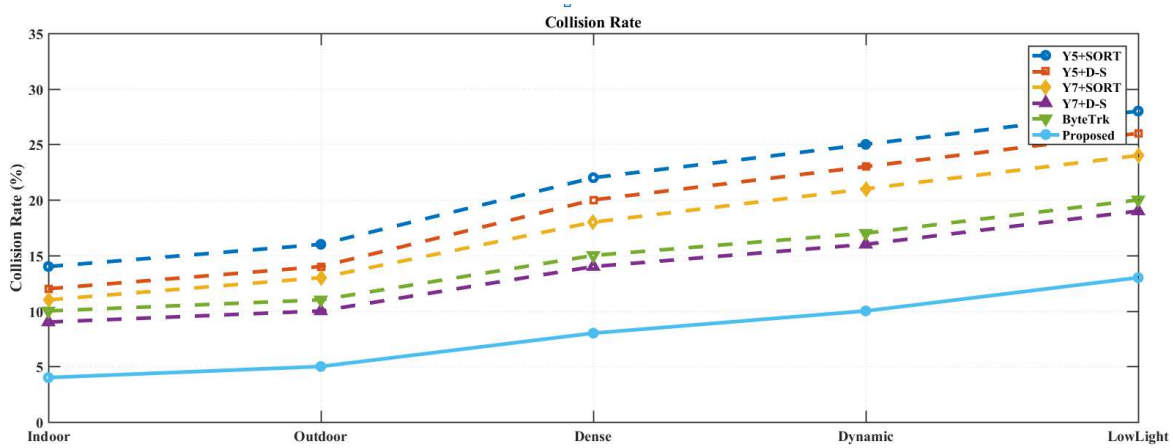

Figure 5: Collision-Rate plot

The path-deviation (Figure 6) results show that the proposed framework not only avoids obstacles safely but also follows a more efficient route with the smallest detour in every scenario. As conditions become harder (Indoor to LowLight), all methods exhibit increased deviation because the UAV must perform more avoidance maneuvers; however, the proposed method remains consistently lowest at approximately 1.4 m (Indoor), 1.6 m (Outdoor), 2.2 m (Dense), 2.4 m (Dynamic), and 2.9 m (LowLight). In comparison, YOLOv5+SORT shows the largest deviation, rising from about 2.4 m to 4.1 m (≈ 2.4, 2.7, 3.3, 3.6, 4.1 m), while stronger baselines such as YOLOv7+DeepSORT and ByteTrack still produce larger detours of roughly 1.9 to 3.2 m and 2.0 → 3.3 m, respectively. This indicates the proposed risk-aware fusion strategy reduces unnecessary over-steering and oscillations during avoidance, enabling the UAV to maintain smoother target pursuit while making only the required deviations to remain collision-free, particularly in dense clutter and low-light scenarios.

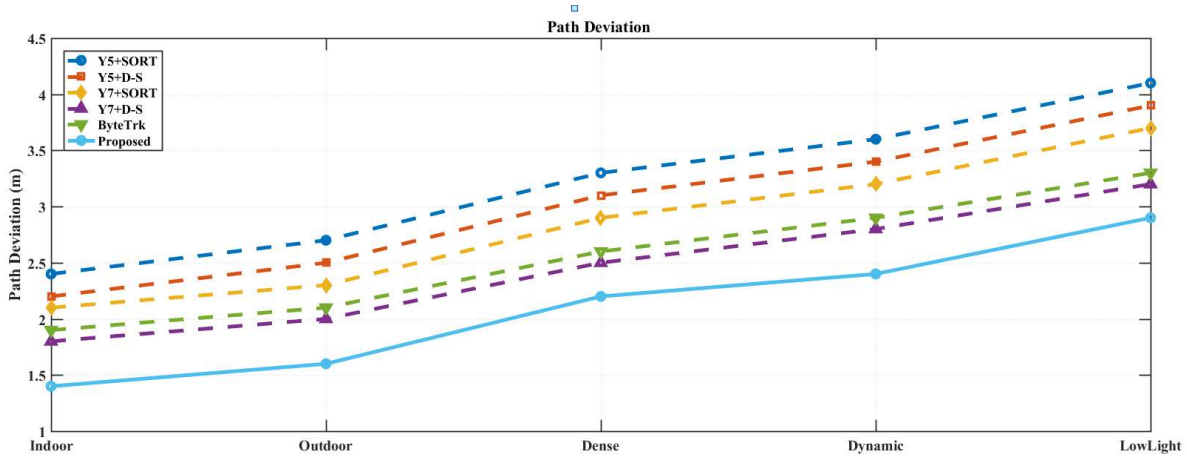

Figure 6: Path-Deviation Plot

The comparison of the detection precision (Figure 7) shows that the proposed improved YOLOv7-based method is the most reliable in the detection of objects of all considered methods. As indicated in the bar chart, YOLOv5+SORT yields the lowest precision of about 0.86, then YOLOv5+DeepSORT of about 0.88 with YOLOv7+SORT of about 0.89. Tracking-integrated pipeline YOLOv7+DeepSORT has a stronger result of about 0.91 and StoneTrack with about 0.9, which has a competitive performance. Nevertheless, the offered approach reaches the maximum level of accuracy of approximately 0.94, which proves that the detector refinements minimize the number of false positives and enhance the stability of confidence. This improved accuracy is significant as it gives cleaner detections to the tracking module, which can reduce identity errors as well as better the tracking and navigation performance in complex UAV scenes.

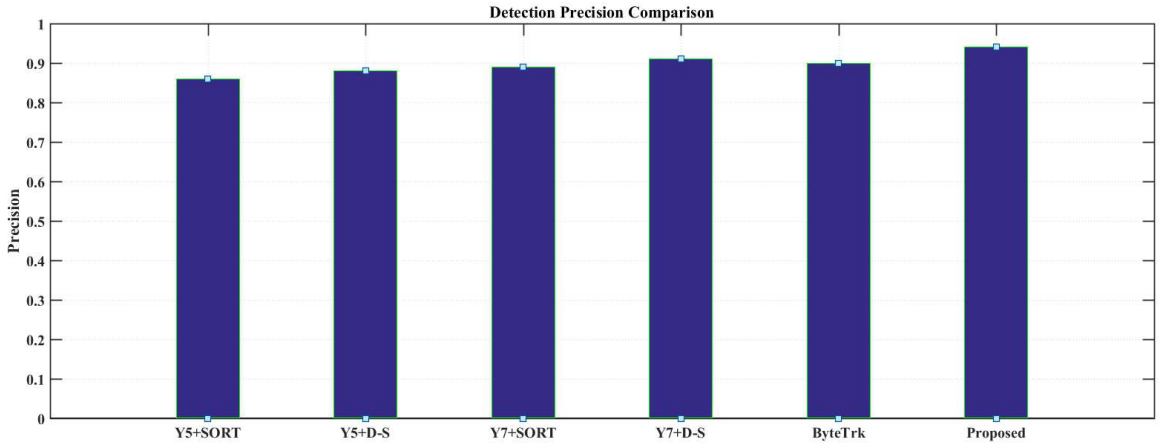

Figure 7: Detection Precision Comparison Plot

To track, the bar plot (Figure 8) of the grouping of the MOTA and IDF1 suggests that the offered method provides the most credible identity-consistent tracking, with MOTA ≈ 0.86 and IDF1 ≈ 0.85. Comparatively, the current approaches obtain even smaller values (approximately, MOTA/IDF1 0.71/0.68 Y5+SORT, 0.75/0.72 Y5+D-S, 0.77/0.74 Y7+SORT, 0.81/0.79 Y7+D-S and 0.80/0.78 ByteTrk). The enhancement is strongest when compared to motion-only association (SORT), which proves that appearance embedding and better detections in DeepSORT lead to a lower level of identity drift. This enhanced tracking performance is also in line with the avoidance plots since constant target identity avoids unwanted oscillations of pursuits that may lead to higher risk of collision.

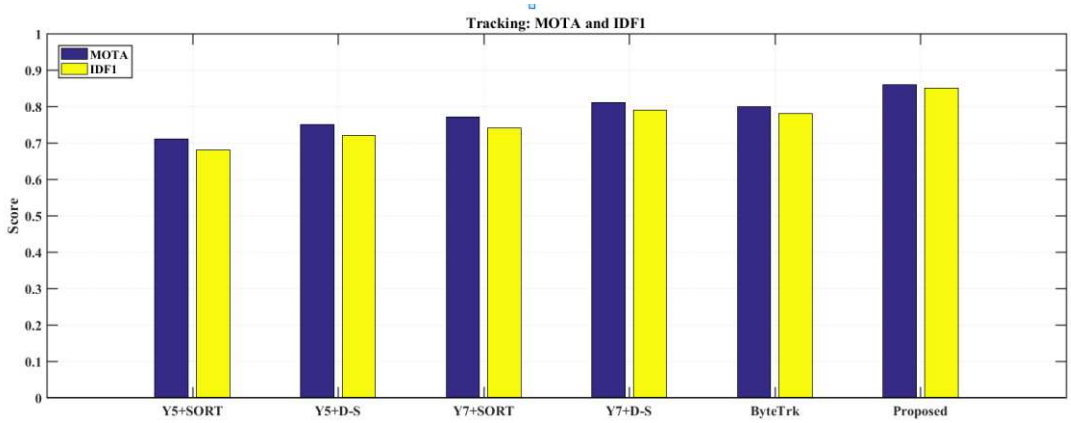

Figure 8: “MOTA and IDF1” grouped bar plot

The obtained FPS results (Figure 9) on various platforms prove the fact that the proposed framework ensures a high level of real-time performance and provides the best accuracy and safety. On the Laptop side, the proposed method is fastest with a speed of approximately 60 FPS, compared to the other methods that are approximately 45-55 FPS (e.g. 55 FPS with Y5+SORT and 50-52 FPS with Y7-based and ByteTrack pipelines). All the methods are slower on the resource-limited Jetson Nano, but the proposed method still runs at about 22 FPS, compared to 16-20 FPS with the current methods. This trend is also possible in Jetson Xavier where the proposed framework is approximately 38 FPS, compared to the others that are between approximately 28-34 FPS. Although in the CPU-only setting the proposed method is still competitive at approximately 12 FPS, which is just higher than the other methods (around 911 FPS). All these findings indicate that the suggested improvements do not affect the efficiency of the runtime, but the system can retain deployable real-time performance on high-performance GPUs and embedded edge devices and be used in application in UAV onboard applications.

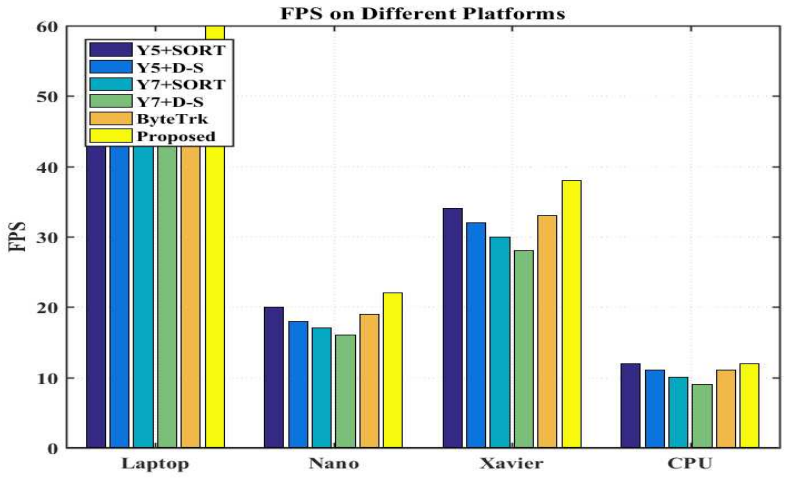

Figure 9: FPS results across different platforms

Table 2: Performance Comparison of Proposed Method with Existing Methods

Metric	Y5+SORT	Y5+D-S	Y7+SORT	Y7+D-S	ByteTrk	Proposed
Precision	0.86	0.88	0.89	0.91	0.90	0.94
MOTA	0.71	0.75	0.77	0.81	0.80	0.86
IDF1	0.68	0.72	0.74	0.79	0.78	0.85
IDS	110	95	88	70	75	45
Avoid Success (LowLight %)	62	65	66	71	70	78
Collision (LowLight %, lower)	28	26	24	19	20	13
Min Distance (LowLight m, higher)	0.45	0.50	0.52	0.62	0.60	0.70
Path Deviation (LowLight m, lower)	4.1	3.9	3.7	3.2	3.3	2.9
FPS (Laptop)	55	50	47	45	52	60

Table 2 provides the summary of the key performance indicators of the proposed method in comparison with five current methods in the detection, tracking, obstacle avoidance, and real-time speed. The proposed framework is the most precise (0.94) in terms of detection, with fewer false detections, compared to the others (0.86–0.91). It is also the best overall consistent with the highest MOTA (0.86) and IDF1 (0.85), i.e. it is more accurate in tracking the targets and less likely to lose identities. The IDS metric supports this with high scores, as the proposed method has the lowest score (45), which means that the number of identity switches is much lower compared to the baselines (70110). To emphasize robustness, the most challenging case (LowLight) is reported in the table in case of obstacle avoidance. The suggested solution has the best avoidance success rate (78) and at the same time has the lowest collision rate (13%), compared to the baselines which have lower success (62-71%) and higher collision (19-28%) rates. Also, the proposed approach

has the highest minimum distance to obstacles (0.70 m), which proves that it has a higher safety margin, and the minimum deviation of the path (2.9 m) shows that the approach does not go round obstacles without any needless turns. Lastly, the offered approach is still feasible to deploy, reaching the peak Laptop FPS (60), demonstrating that the proposed improvements in accuracy and safety are not achieved at the cost of the real-time.

5. Conclusion and Future Work

5.1 Conclusion

The paper proposed a real-time UAV target tracking and obstacle avoidance system, which incorporates a better YOLOv7 detector, DeepSORT tracking and a safety-priority avoidance maneuver in a perception-to-control control pipeline. The enhanced YOLOv7 increased the detection stability on the UAV scenes disturbed by small objects, scale variation and motion blur whereas DeepSORT offered identity-based consistency tracking due to motion prediction and appearance-based association. The fusion controller and obstacle-risk module allowed the UAV to track the target in a smooth manner under safe conditions, and to ignore the commands of pursuit in case of high collision risk. The effectiveness of the proposed design was confirmed through experimental comparison with five existing methods with a higher level of detection and tracking accuracy and safety performance. In particular, the suggested technique achieved 0.94 precision, MOTA=0.86, IDF1=0.85, and decreased the identity switches to IDS=45. It also obtained 78 percent avoidance success at 13 percent collision rate under difficult low-light conditions, a 0.70 m minimum obstacle distance and a 2.9 m path deviation without compromising real-time feasibility at 60 FPS on a laptop graphics card. Altogether the findings indicate that a powerful detector, identity-preserving tracker and risk-conscious controller can be combined to create a viable and deployable system of safe UAV target-following in challenging conditions.

5.2 Future Work

The system can be further enhanced in the future in a number of directions. First, to enhance the obstacle avoidance module, it is possible to add metric depth information of stereo cameras, RGB-D cameras, or lightweight LiDAR and combine multi-sensor information through probabilistic filtering to be more reliable in low-texture or low-light environments. Second, the adaptive fusion and model-predictive control (MPC) can be brought in to improve the end-to-end performance to produce less jagged paths and meet the dynamic constraints, particularly when the UAV speed increases. Third, the optimizer of the detector and tracker can be optimized to be edge deployable with model pruning, quantization, and acceleration with TensorRT, and it can run faster on embedded platforms. Fourth, identity switches in crowded scenes and long-term occlusions can be minimized using more sophisticated tracking approaches (e.g. transformer-based association or re-identification fine-tuning on UAV data). Lastly, generalization will be enhanced by extensive validation with real flight experiments, varied weather/ lighting conditions and other benchmarks, and incorporating high level planning (waypoint navigation and SLAM mapping) may scale the framework up to fully autonomous mission level UAV navigation.

Funding Statement

This article is the research results of the science and technology research project of Jiangxi Provincial Department of Education in 2024, "Research on the trajectory tracking of four-rotor UAVs based on high-precision control algorithms" (No. GJJ2406804).

Authors' contributions

J.Z. conceived the study and led the manuscript preparation. C.Z. participated in model improvement and experimental data analysis. X.Y. was responsible for module integration and performance verification. All authors reviewed and approved the final manuscript.

Conflict of Interest

The authors declare no known competing financial interests or personal relationships that could have influenced the work reported in this paper.

Availability of data and materials

The data used to support the findings of this study are available from the corresponding author upon request.

Reference

[1]. Zhang, G., Zhao, Y., Fu, P., Luo, W., Shao, Q., Zhang, T., & Yu, Z. (2024). A reliable unmanned aerial vehicle multi-target tracking system with global motion compensation for monitoring Procapra przewalskii. Ecological Informatics, 81, 102556. https://doi.org/10.1016/j.ecoinf.2024.102556

[2]. Bai, Z., Pei, X., Zheng, Q., Wu, G., & Bai, Y. (2024). Improved YOLOv7 target detection algorithm based on UAV aerial photography. Drones, 8(3), 104. https://doi.org/10.3390/drones8030104

[3]. Zhang, Y., Wang, M., & Fu, Y. (2024). Research on aircraft tracking technology based on improved DeepSORT algorithm. In Proceedings of the 2023 3rd International Conference on Big Data, Artificial Intelligence and Risk Management (ICBAR ’23) (pp. 1092–1097). Association for Computing Machinery. https://doi.org/10.1145/3656766.3656947

[4]. Cao, X., Wang, Z., Zheng, B., & Tan, Y. (2023). Improved UAV-to-Ground Multi-Target Tracking Algorithm Based on StrongSORT. Sensors, 23(22), 9239. https://doi.org/10.3390/s23229239

[5]. Wang, L., & Liu, T. (2025). A real-time tracking approach for moving objects based on an integrated algorithm of YOLOv7 and SORT. Journal of Circuits, Systems and Computers, 34(4), 2550102. https://doi.org/10.1142/S0218126625501026

[6]. Ghazlane, Y., El Hilali Alaoui, A., Medomi, H., & Bnouachir, H. (2024). Real-time airborne target tracking using DeepSORT algorithm and YOLOv7 model. International Journal of Advanced Computer Science and Applications, 15(2), 451–458. https://www.ijacsa.thesai.org

[7]. Hanzla M, Yusuf MO, Al Mudawi N, Sadiq T, Almujally NA, Rahman H, Alazeb A, Algarni A. Vehicle recognition pipeline via DeepSort on aerial image datasets. Front Neurorobot. 2024 Aug 16;18:1430155. Doi: 10.3389/fnbot.2024.1430155. PMID: 39220587; PMCID: PMC11362136.

[8]. Li, A. (2025). Unmanned Aerial Vehicle Target Detection Integrating Computer Deep SORT Algorithm and Wireless Signal. International Journal of Interdisciplinary Telecommunications and Networking (IJITN), 17(1), 1-15. https://doi.org/10.4018/IJITN.373429

[9]. Liang, X., Qi, Y., Chen, G. et al. Research on a distributed strategy for UAV detection and tracking of specific targets. Artif Intell Rev 58, 388 (2025). https://doi.org/10.1007/s10462-025-11364-x

[10]. Qing, W. (2025). Infrared Imaging-Based Object Detection and Tracking for UAV Systems: Principles, Algorithms, and Advances. Frontiers in Computing and Intelligent Systems, 14(3), 1-6. https://doi.org/10.54097/sw8nmq23

[11]. Jian, Y., Yang, H., Wang, X., Rong, Y., & Zhu, Y. (2025). The “low slow and small” UAV target detection and tracking algorithm based on improved YOLOv7 and DeepSORT. Zhihui Kongzhi Yu Fangzhen, 47(1), 23–31. https://doi.org/10.3969/j.issn.1673-3819.2025.01.003

[12]. Wang, Q., Ye, G., Chen, Q. et al. A UAV perspective based lightweight target detection and tracking algorithm for intelligent transportation. Complex Intell. Syst. 11, 73 (2025). https://doi.org/10.1007/s40747-024-01687-7

[13]. Jing, J., Jia, B., & Huang, B. (2026). A dual branch SPD-Conv infrared distant multi-target tracking method with high target association accuracy. Applied Soft Computing, 187, 114220. https://doi.org/10.1016/j.asoc.2025.114220

[14]. Chen, C., Zheng, Z., Xu, T., Guo, S., Feng, S., Yao, W., & Lan, Y. (2023). YOLO-based UAV technology: A review of the research and its applications. Drones, 7(3), 190. https://doi.org/10.3390/drones7030190

[15]. Hong, T., Liang, H., Yang, Q., Fang, L., Kadoch, M., & Cheriet, M. (2022). A real-time tracking algorithm for multi-target UAV based on deep learning. Remote Sensing, 15(1), 2. https://doi.org/10.3390/rs15010002

[16]. Yuan, Z., Wang, Z., & Zhang, R. (2024). CCBA-NMS-YD: A vehicle pedestrian detection and tracking method based on improved YOLOv7 and DeepSort. World Electric Vehicle Journal, 15(7), 309. https://doi.org/10.3390/wevj15070309

[17]. Lu, H., Yang, S., Zhou, H., Zhao, J., Bai, F., Han, Y., & Zhen, B. (2025). Impact of flight altitude and photo control points on the accuracy of 2D and 3D modelling in UAV surveys: A case study of Jurong Dongshan River. Journal of Imaging Science and Technology, 69(4), 040506. https://doi.org/10.2352/J.ImagingSci.Technol.2025.69.4.040506

[18]. Chen, Y.-H. (2025). Strong baseline: Multi-UAV tracking via YOLOv12 with BoT-SORT-ReID. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW 2025) (pp. 6573–6582). IEEE. https://doi.org/10.1109/cvprw67362.2025.00654

[19]. Zhu, J., Ma, C., Rong, J., & Cao, Y. (2024). Bird and UAVs recognition detection and tracking based on improved YOLOv9-DeepSORT. IEEE Access. https://doi.org/10.1109/ACCESS.2024.3475629

[20]. Baidya, R., & Jeong, H. (2024). Simulation and real-life implementation of UAV autonomous landing system based on object recognition and tracking for safe landing in uncertain environments. Frontiers in Robotics and AI, 11, 1450266. https://doi.org/10.3389/frobt.2024.1450266

[21]. Kapania, S., Saini, D., Nagrath, P., et al. (2020). Multi-object tracking with UAVs using DeepSORT and YOLOv3 RetinaNet detection framework. In Proceedings of the 1st ACM Workshop on Autonomous and Intelligent Mobile Systems. Association for Computing Machinery. https://doi.org/10.1145/3377283.3377284

[22]. Zhang, Y., Qian, Z., & Cai, J. (2025). A Systematic Review on Real-time Detection of Small Obstacles Based on Multidimensional Information Fusion. Chinese Journal of Information Fusion, 2(4), 313–339. https://doi.org/10.62762/CJIF.2025.500710

[23]. Zhao, Y., Ma, Q., Lei, G., Wang, L., & Guo, C. (2025). Research on lightweight tracking of small-sized UAVs based on the improved YOLOv8N-drone architecture. Drones, 9(8), 551. https://doi.org/10.3390/drones9080551

[24]. Liu, Y., Zhou, H., & Zhao, M. (2025). Research on target detection based on improved YOLOv7 in complex traffic scenarios. PLOS ONE. https://doi.org/10.1371/journal.pone.0323410

[25]. Sun, H., Ma, P., Li, Z., Ye, Z., & Ma, Y. (2024). Hyperspectral low altitude UAV target tracking algorithm based on deep learning and improved KCF. Frontiers in Physics, 12. https://doi.org/10.3389/fphy.2024.1341353