2023 International Conference on Electrical, Electronics, Communication and Computers (ELEXCOM) | 979-8-3503-0511-1/23/$31.00 ©2023 IEEE | DOI: 10.1109/ELEXCOM58812.2023.10370061

Hardware-based Implementation of Target Tracking in Unmanned Aerial Vehicles (UAVs)

Rishi Agarwal
Department of Electrical and Electronics Engineering
Birla Institute of Technology and Science, Pilani
Rajasthan, India
f20210973@pilani.bits-pilani.ac.in

Sathwik Gundala
Department of Electrical and Electronics Engineering
Birla Institute of Technology and Science, Pilani
Rajasthan, India
f20201997@pilani.bits-pilani.ac.in

G S S Chalapathi
Department of Electrical and Electronics Engineering
Birla Institute of Technology and Science, Pilani
Rajasthan, India
gssc@pilani.bits-pilani.ac.in

***Abstract*—Unmanned Aerial Vehicles (UAVs) have gained significant attention in various fields, including surveillance, search and rescue, and monitoring applications. One important application for UAVs is target tracking, which requires detecting and tracking a specific object of interest in real time. This paper surveys work done so far in the area of target-tracking in UAVs. It then presents a comprehensive hardware-based framework for target tracking in UAVs. This work utilizes the State-of-the-Art YOLOv8 (You Only Look Once) algorithm for target detection, an efficient high-speed target tracking model, and a Proportional Derivative (PD) control algorithm for precise drone movement control. YOLOv8 provides fast, accurate, and real-time detection of the object of interest, allowing the UAV to detect and identify the target object quickly and reliably. Subsequently, a robust tracking algorithm tracks the identified object across consecutive frames, ensuring accurate localization and trajectory estimation. Furthermore, a PD control algorithm is integrated into the system to enable precise and smooth drone movement. The proposed framework is integrated and used for target tracking in UAVs. Further, this framework is implemented on a UAV. The results demonstrate the effectiveness and robustness of the proposed framework, showcasing its potential for real-world applications.**

***Index Terms*—UAVs, YOLOv8, Multimedia in UAV, Target Tracking, Object tracking**

I. Introduction

Unmanned Aerial Vehicles (UAVs), commonly referred to as drones, have revolutionized numerous fields of technology and science, providing unprecedented opportunities for applications spanning from aerial photography to search-and-rescue missions. In recent years, using UAVs for target tracking has attracted considerable interest due to its potential to transform surveillance, security, and monitoring practices. Target tracking entails locating and perpetually monitoring an object of interest as it moves within a specified environment, making it a vital component in various fields, including military operations, law enforcement, disaster management, and wildlife conservation. Integrating UAVs into target tracking provides several advantages over conventional methods. They provide an unrivaled advantage for real-time monitoring of targets due to their adaptability, versatility, and ability to access remote and hazardous locations. In addition, the development of sophisticated imaging sensors, such as high-resolution cameras and thermal sensors, enables UAVs to capture detailed information and track targets with greater precision and accuracy. This research paper intends to investigate the challenges and possible future orientations of UAV-based target tracking.

The evolution of UAV technology has substantially contributed to the enhancement of target-tracking capabilities. The miniaturization of sensors, enhancements in computational power, and advancements in autonomous flight algorithms have enabled UAVs to operate autonomously, making them ideal for target-tracking tasks. In addition, incorporating artificial intelligence and machine learning techniques enables UAVs to actively monitor and predict the movements of targets, making them more effective in complex and dynamic environments. In general, the altitude and position of a UAV are controlled by controlling the up and down movement, forward and backward movement, and yaw, pitch, and roll angle changes, but the current trends show the focus to be on intelligent sensor technologies and autonomous control.

This paper investigates the cutting-edge methodologies, algorithms, and technologies employed in UAV-based target tracking. The paper will also discuss future directions and additional research and development opportunities. The key contributions of the paper are:

- Implementation of State-of-the-art algorithms: YOLOv8 for target detection and DroTrack for target tracking.
- Integration for a Proportional Derivative control algorithm.
- Deployment of the above-proposed framework on a hardware platform.

Finally, the results of our implementation are presented towards the end and are intended to facilitate the widespread adoption and deployment of unmanned aerial vehicles (UAVs) in various domains where accurate and reliable target tracking is crucial. The motivation for our work was based on developing a lightweight model for target-tracking and chasing, which

Corresponding Author: G S S Chalapathi. email:gssc@pilani.bits-pilani.ac.in

can be deployed on low-cost UAVs with low computational power.

II. Related Works

A notable amount of research work has been conducted in the domain of UAVs in recent years. The problem of target tracking in UAVs is generally divided into two parts: Acquisition of Target Information and Information Processing and Controlling Algorithms to keep the drone in the vicinity of our object of interest. An Image-Based Visual Serving (IBVS) system has been discussed in [1] to track a ground target from a flying UAV. They implemented a tracking system consisting of two deep neural networks: an approaching network, in which the relative distance between the UAV and the target is estimated, and the UAV is driven the required distance, and a chasing network, in which the orientation of the tracked target is used to resolve the ambiguity between the yawing and lateral movements of the UAV and consecutively aligning the UAV with the tracked target heading to ease the tracking task. Deep detection networks formed their backbone.

An improved YOLOv5 algorithm combined with the DeepSort Tracking algorithm was proposed in [2] to accurately detect and track intruding UAVs for 'black flight' and 'indiscriminate flight.' They added a convolutional block attention module (CBAM) to the Neck module of the YOLOv5s network to enhance the extraction of network features. [3] discussed the necessity of using fast object detection in vision target tracking in UAVs and proposed the use of YOLOv3-Tiny, which is an object detection method based on deep learning. It optimizes the network structure based on YOLOv3 and reduces the output of one scale. They use a custom-trained YOLOv3 Tiny model based on human images, and then the trained model is followed by a control signal to the flight controller for target tracking.

An implementation method of a quadrotor UAV target tracking system based on OpenMV was proposed in [4] to solve the problems with complex structure and low resource utilization of traditional target tracking UAVs. The system mainly depended on monocular vision. It includes a flight control system, a visual guidance system, a six-axis motion sensor, and magnetometer serving as a pose measurement module, and an ultrasonic sensor for the altitude measurement system. A centroid-based tracking algorithm was used for calculating and resolving the target. A model for faster object detection was proposed in [5]. The model performs lightweight processing in the YOLOv4 object detection model for faster detection speed. They use the CA attention mechanism module to replace the original SE attention mechanism to build a new model with stronger object detection capabilities. With numerous target tracking and processing applications, [6] studied the applications for a social cause and proposed a model, Ocularone, to enhance the lifestyle of Visually Impaired People (VIP) through navigational assistance and situation awareness. The model enhanced the safety and autonomy of the user with the help of onboard sensors, edge accelerators, and two-way communication using gestures and audio prompts. They used Tello Drones and Jetson Nano modules as early prototypes for validation.

Due to the complex motion of drones, having multiple degrees of freedom in a three-dimensional space causes high uncertainty in target detection and tracking scenarios. Most existing object-tracking frameworks have tackled common challenges such as cluttered backgrounds and occlusion. Still, the problem of complex motion in drones posed a problem in tracking. To tackle this problem, [7] proposed a model, DroTrack, a high-speed visual single-object tracking framework for drone-captured video sequences. They discover the dependency between the object's representation and their motion geometry and nullify the uncertainty problem, leading to inaccurate prediction in target location and fuzziness in scale estimations. The greatest obstacle is the lack of data available to train online deep neural networks to recognize objects that have only been observed in a single example. Training a deep CNN for object tracking is a potential solution [8] [9]. The absence of annotated data, however, hinders the training of deep CNN. By transferring the learned feature hierarchies to online tracking, training an offline CNN using a large set of videos with tracking ground truths can solve this problem. Graph Convolutional Networks are additionally proposed for object detection [10]. The primary challenge of these methods is executing stochastic gradient descent online to adjust the network's weights. This necessitates high computational speed, which renders tracking unreliable, particularly in the case of drones.

In recent decades, numerous studies have focused on the collaborative framework, path computation efficiency, and path quality. Typical algorithms include the Model Predicted Control (MPC) method, the Voronoi method, intelligent algorithms (such as Genetic Algorithms and Particle Swarm Optimisation), the Rapidly-exploring Random Tree (RRT) [11] method, and the Artificial Potential Field (APF) method. By incorporating the feedback mechanism into the optimal fuzzy reasoning method, the so-called Feedback Based Compositional Rule of Inference (FBCRI) [12] is proposed to address the cooperative path planning problem. Zhang et al. [13]. present the Cooperative and Geometric Learning Algorithm (CGLA) to address collision avoidance and information sharing issues. The scalability of these methods is a challenge making deep learning-based approaches preferable for hardware implementations on various levels. Inspired by these works, we propose a model which integrates the state-of-the-art YOLO model with the DroTrack model to achieve a lightweight target-tracking system in UAVs.

III. System Description

Our proposed framework has been tested extensively to find its limits and capability. The hardware used in the paper is the DJI Tello drone (Fig. 2) featuring a 14-core Intel Processor, a 5-megapixel 720p camera, and a feature for backend programming with Python. Tello SDK and the built-in APIs can make the drone perform various functions. A MacBook Air equipped with M2 chip was used for processing

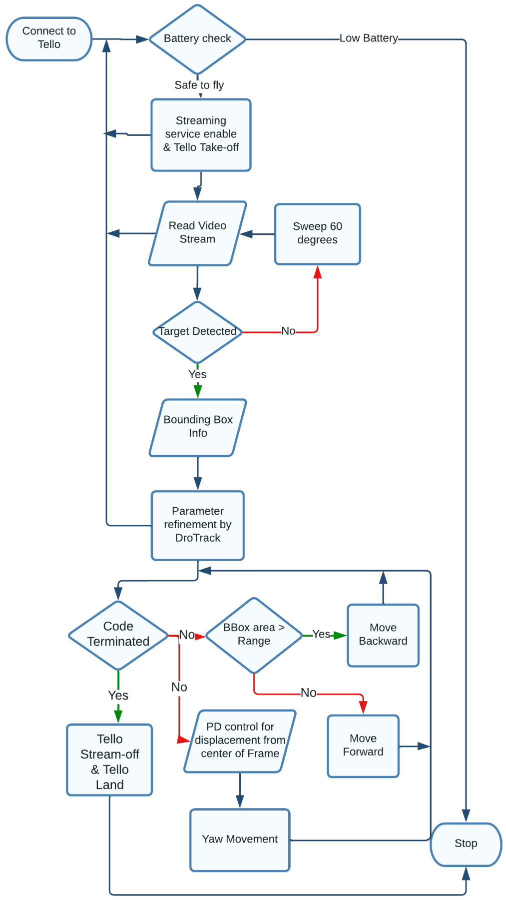

Fig. 1. Workflow of the system

Fig. 2. DJI Tello drone

and to establish communication with the drone. The processing for the whole system was performed locally on the M2 Air, and instructions were sent to the Unmanned Aerial Vehicle (UAV). The Tello SDK uses a local port to establish communication over 2.4 GHz 802.11n Wi-Fi with the aircraft over User Datagram Protocol (UDP). The drone was calibrated by connecting a smartphone to the Drone's Wi-Fi and following the Tello drone Application's on-screen instructions. The system can be

Fig. 3. Calibration of Tello drone via Smartphone App

understood in three parts: a) Target Detection, b) Refinement of parameters, c) An appropriate Control Algorithm. The overall flow of the system can be understood from Fig. 1

A. Target Detection

We use the Ultralytics You-Only-Look-One Version 8(YOLOv8) [14], which forms the basis of the target detection module of our system. YOLOv8, a single-stage object detection algorithm, is a cutting-edge, start-of-the-art (SOTA) model that builds upon the success of the previous YOLO versions but with enhanced performance and flexibility. Partition processing forms the basis of the YOLO algorithm. An input image (or a frame from a video stream) is provided to the algorithm which performs further processing. The working of a YOLO object detection algorithm [15] can be explained as follows :

1) **Grid Division:** Input frame is divided into grid cells depending on image size and the last convolution layer of the network and object detection is performed individually.
2) **Feature Extraction:** A Convolutional Neural Network (CNN) is used to derive features from each cell. A large image dataset is used for pre-training the CNN to acquire features that can be applied to object detection.
3) **Objectness Score:** The logistic regression function predicts the probability of the presence of an object of interest in the cell, which is known as Objectness Score.
4) **Class Probability:** Each cell is assigned a class of object and probability by YOLO. The conditional probability of an object class is calculated using a SoftMax function.
5) **Bounding Box:** For each cell object prediction, a bounding box is also predicted that encloses the object of interest, which is anticipated in relation to the size of the cell and is represented in the form of height, width, and center coordinates.
6) **Non-Maximum Suppression (NMS):** After identifying multiple objects in the image each with own bounding box, YOLO performs Non-Maximum supression to reduce redundancies. The final list of bounding boxes with their related class and confidence score serve as the output.

Fig. 4 portrays the architecture of YOLOv8 [17] . It contains a series of updates and new convolutions compared to the previous YOLO versions in the architecture. One of the major upgrades includes the Anchor-free detection approach, in which an object detection model can directly predict the

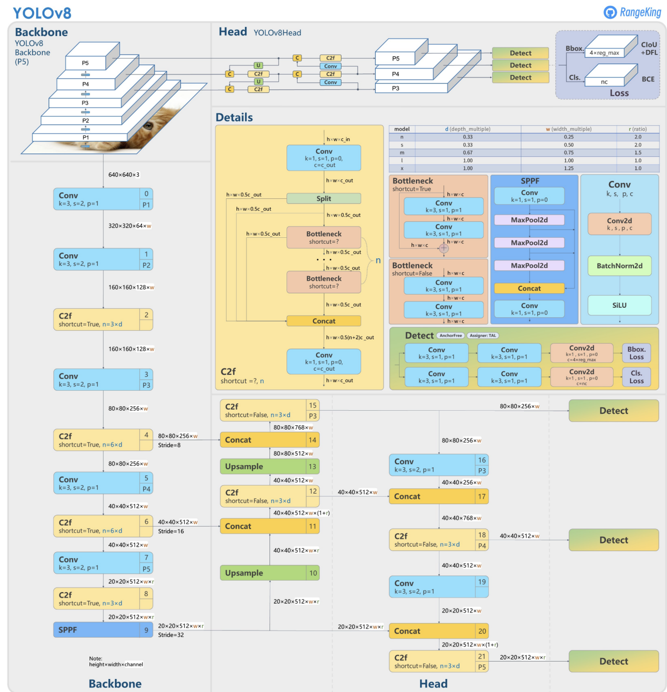

Fig. 4. Yolov8 Architecture as proposed in [17]

center of the object instead of providing an offset from a known anchor box. This makes the model more flexible and efficient because it no longer requires manual specifications of the anchor boxes, which are typically challenging to choose and led to suboptimal results in previous YOLO models such as v1 and v2. YOLO operates simultaneously on an entire image, making it extremely quick and efficient. This distinguishes YOLO from other object detection algorithms, which employ a computationally costly sliding window approach. YOLOv8 employs a CNN with two major components: the Backbone and the Head. The backbone is comprised of a modified version of the CSPDarknet 53 architecture, which consists of 53 convolutional layers and uses cross-stage partial connections to enhance the flow of information between layers. Multiple convolutional layers are followed by a succession of fully connected layers that predict bounding boxes, objectness scores, and class probabilities for the objects detected in an image.

B. Refining parameters with DroTrack

The DroTrack model proposed in [7] has been thoroughly analyzed to comprehend its operation. The model serves as a framework for high-speed object tracking and tackles the problem of the complex motion of drones leading to uncertainty very efficiently. The model performs the following processing on the received frames: 1) Adaptive Center Detection, 2) Optical Flow Tracking, 3) Location Relative Correction, 4) Fuzzy C Means Segmentation, 5) CNN-based Segment Selection, 6) Relative Angular Scaling followed by 7) Final Location Prediction and scaling. These can also be seen in Fig. 5

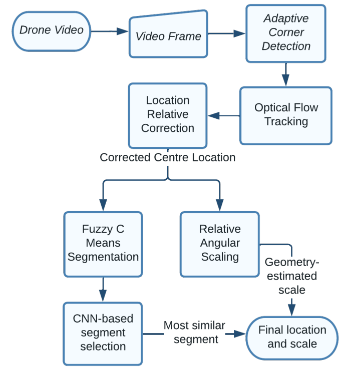

Fig. 5. Process steps involved in DroTrack Model [7]

C. Control Algorithm

To achieve faster tracking times, the overall complexity of the flight software needs to be optimized. By adding additional headroom for hardware-intensive tasks such as detection and tracking, we are limited to using light control algorithms with some reduced maneuverability. One of the simplest control algorithms we can use is a PD (proportionate and derivative) controller. Upon receiving the refined bounding box and center coordinates, we incorporate a Proportional Derivative (PD) control algorithm, which aims to keep the drone in the vicinity of the target operation via further processing. By incorporating feedback from the target tracking system, the PD controller adjusts the drone's control inputs based on the error between the desired and actual target position, resulting in improved tracking performance and enhanced stability. Due to the simplicity of the control algorithm, the drone would not perform complex aerial maneuvers but tries to trace the path of the tracking object. The Drone is equipped with a safety system that performs a self-landing of the drone in case of a low battery.

IV. Results and Discussion

The proposed system was tested first on a laptop webcam. and then deployed on a Tello drone which came with successful results. The working of the proposed system is as follows:

A. Detection of a single person via drone camera

The performance of various YOLO versions and models was tested on the host computer, and the latest, state-of-the-art YOLO version 8 was used for our system based on the performance and flexibility encountered. Fig. 6 portray the difference in performance among different model sizes on the COCO dataset of YOLOv8 [14]. The most commonly used metrics for comparison are: mean Average Precision (mAP), Inferencing speed (in fps), and the compute cost based on the model size in FLOPs and params. YOLOv8 comes with 5 pre-trained models, namely, nano (v8n), small(v8s), medium(v8m), large(v8l) and extra large(v8x), which can be used with great flexibility. After running and testing all the above models, we achieved better performance and response in

Model	size (pixels)	mAP^{val} 50-95	Speed CPU ONNX (ms)	Speed A100 TensorRT (ms)	params (M)	FLOPs (B)
YOLOv8n	640	37.3	80.4	0.99	3.2	8.7
YOLOv8s	640	44.9	128.4	1.20	11.2	28.6
YOLOv8m	640	50.2	234.7	1.83	25.9	78.9
YOLOv8l	640	52.9	375.2	2.39	43.7	165.2
YOLOv8x	640	53.9	479.1	3.53	68.2	257.8

Fig. 6. Comparison of YOLO models of different sizes [14]

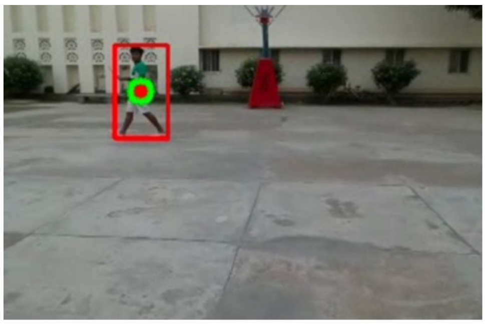

Fig. 7. Person being detected in the frame

the nano model of YOLOv8, even with some reduced accuracy, because in high-speed target-tracking systems, it is essential for the algorithms to respond faster. We use a pre-trained YOLOv8n model with PyTorch and feed it with the live drone video stream using OpenCV libraries and DJITelloPy APIs. The object detection classes were restricted to 1, i.e., person, which is our object of interest in this scenario.

Initially, if the model encounters zero detections, the drone is made to perform a 360-degree sweep at a fixed altitude iteratively until a detection takes place. Upon receiving detections, we obtain the bounding box and center coordinates of the highest confidence detection, which is then used for further processing in the next steps.

B. Parameter refining with DroTrack

The bounding box and center coordinates of the detection with the highest confidence scores obtained in the previous step are now passed as parameters in the DroTrack model, which performs extensive processing on them to obtain refined bounding box and center coordinates. These values are then passed on to the next step.

C. Controlling the drone with a control algorithm

We implement a custom Proportional Derivative (PD) algorithm to maintain the drone's proximity to the target by keeping the bounding box's area within a specified range. At the beginning of the operation, the drone's altitude is set at an optimal value. Constants for the PD algorithm have been fine-tuned and set following iterative testing on Tello. An error variable is employed to calculate the error in the current detection by determining the displacement in center

Fig. 8. Drone moved towards the person successfully

coordinates from the screen's center. Consequently, the pace is transmitted to the drone.

In the first frame, i.e., Fig. 7, the drone detects a person in the frame, and with the help of the control algorithm, the drone is provided with some yaw and forward velocity. In the second frame, i.e., Fig. 8 the drone has successfully moved toward the person. Similarly, the system is tested in a different environment with changed listing conditions, as seen in Fig. 9 and Fig. 10. The drone performed optimally in bright and normal lighting conditions but had some difficulties in dark lighting conditions because the drone is equipped with a Visual Positioning System (VPS), which looks for patterns on the ground to help position the drone, which could not be done in low-light conditions. The range in which the drone performed target tracking optimally was 20 meters, after which the data transmission over Wi-Fi posed a problem, and the drone could not respond to the control algorithm as efficiently as when the drone was near the host computer.

The inference time for person detection obtained from the YOLOv8n model during the execution of code was plotted against the code running time and can be seen in Fig. 11. It was found to be 45ms. The execution time from the DroTrack model to process each frame received was plotted against the code running time and can be seen in Fig. 12.

V. Conclusions and Future Work

This paper uses YOLOV8 for object detection using UAVs. YOLOv8 algorithm proved to be an accurate and efficient method for target detection from aerial footage of the UAVs.

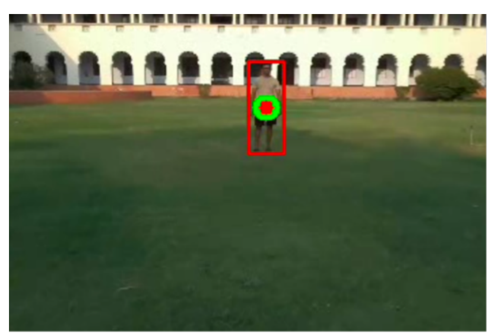

Fig. 9. Person being detected in the frame

Fig. 10. Drone moved towards the person successfully

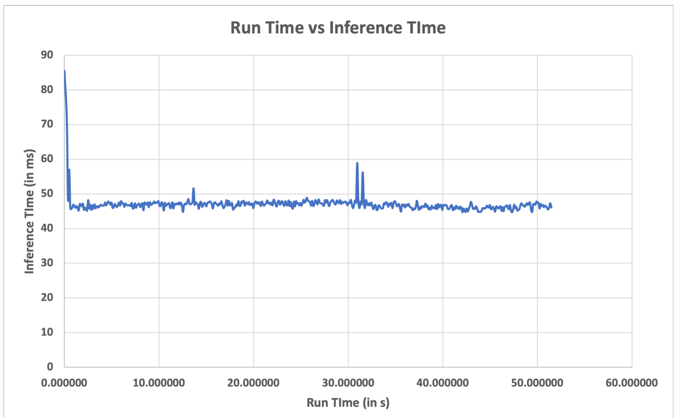

Fig. 11. Run Time vs Inference Time

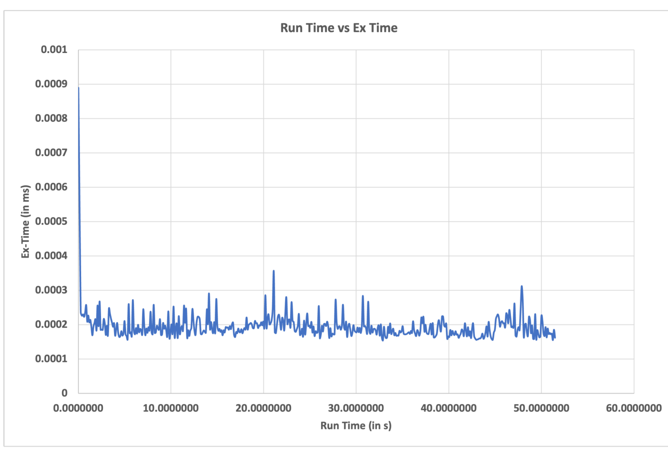

Fig. 12. Run Time vs DroTrack execution Time

Further DroTrack algorithm was used for target tracking in the UAV. A PD algorithm is used to control the movement of the UAV to keep the UAV in the vicinity of the target. This paper integrates these three algorithms and presents their hardware implementation for the target tracking application in a UAV. It also presents the detection time while tracking the target using the above-mentioned algorithms.

Future work in this domain includes the development of a robust obstacle avoidance system that prevents the UAV from collision with unknown objects. Another direction for further research includes modifying this system to incorporate functionality to deal effectively with cases when multiple individuals come together in the UAV's frame. This can be achieved by training the model on a custom dataset that highlights the characteristics of our object of interest.

REFERENCES

[1] M. A. Kassab, A. Maher, F. Elkazzaz and Z. Baochang, "UAV Target Tracking By Detection via Deep Neural Networks," 2019 IEEE International Conference on Multimedia and Expo (ICME), Shanghai, China, 2019, pp. 139-144, doi: 10.1109/ICME.2019.00032.

[2] H. Yang and Y. Ge, "Research on Detecting and Tracking Algorithm of UAV Intrusion Based on YOLOv5+DeepSort," 2022 3rd International Conference on Computer Vision, Image and Deep Learning and International Conference on Computer Engineering and Applications (CVIDL and ICCEA), Changchun, China, 2022, pp. 1209-1212, doi: 10.1109/CVIDLICCEA56201.2022.9823976.

[3] G. Liu et al., "The Development of a UAV Target Tracking System Based on YOLOv3-Tiny Object Detection Algorithm," 2021 IEEE International Conference on Robotics and Biomimetics (ROBIO), Sanya, China, 2021, pp. 1636-1641, doi: 10.1109/ROBIO54168.2021.9739612.

[4] Y. Shao et al., "Research on Target Tracking System of Quadrotor UAV Based on Monocular Vision," 2019 Chinese Automation Congress (CAC), Hangzhou, China, 2019, pp. 4772-4775, doi: 10.1109/CAC48633.2019.8996417.

[5] R. Niu, Y. Qu and Z. Wang, "UAV Detection Based on Improved YOLOv4 Object Detection Model," 2021 2nd International Conference on Big Data and Artificial Intelligence and Software Engineering (ICBASE), Zhuhai, China, 2021, pp. 25-29, doi: 10.1109/ICBASE53849.2021.00012.

[6] Suman Raj, Swapnil Padhi, and Yogesh Simmhan. 2023. Ocularone: Exploring Drones-based Assistive Technologies for the Visually Impaired. In Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems (CHI EA '23). Association for Computing Machinery, New York, NY, USA, Article 220, 1–9. https://doi.org/10.1145/3544549.3585863

[7] A. Hamdi, F. Salim, and D. Y. Kim. 2020. DroTrack: High-speed Drone-based Object Tracking Under Uncertainty. In 2020 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE). IEEE Press, 1–8. https://doi.org/10.1109/FUZZ48607.2020.9177571

[8] Naiyan Wang, Siyi Li, Abhinav Gupta, and Dit-Yan Yeung. Transferring rich feature hierarchies for robust visual tracking. CoRR, abs/1501.04587, 2015.

[9] Hyeonseob Nam and Bohyung Han. Learning multi-domain convolutional neural networks for visual tracking. CoRR, abs/1510.07945, 2015.

[10] Junyu Gao, Tianzhu Zhang, and Changsheng Xu. Graph convolutional tracking. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 4649–4659, 2019.

[11] J-W Lee, B. Walker, K. Cohen, Path planning of unmanned aerial vehicles in a dynamic environment, in: Infotech@Aerospace 2011, Mar. 2011, AIAA paper: 2011-1654.

[12] J. Wilburm, M.G. Perhinschi, B. Wilburn, O. Karas, Development of a modified Voronoi algorithm for UAV path planning and obstacle avoidance, in: AIAA Guidance, Navigation, and Control Conference, Aug. 2012, AIAA paper: 2012-4904.

[13] B.C. Zhang, W.Q. Liu, Z.L. Mao, et al., Cooperative and geometric learning algorithm (CGLA) for path planning of UAVs with limited information, Automatica 50 (3) (2014) 809–820.

[14] Ultralytics, "Home," Ultralytics YOLOv8 Docs, https://docs.ultralytics.com/ (accessed Jun. 15, 2023).

[15] A. Mehra, "Evolution of yolo object detection model from V5 to V8," Labellerr, https://www.labellerr.com/blog/evolution-of-yolo-object-detection-model-from-v5-to-v8/ (accessed Jun. 15, 2023).

[16] A. Mehra, "Understanding yolov8 architecture, applications and features," Labellerr, https://www.labellerr.com/blog/understanding-yolov8-architecture-applications-features/ (accessed Jun. 15, 2023)

[17] R. King, "Brief summary of Yolov8 Model Structure · issue no. 189 · Ultralytics/ultralytics,"GitHub,https://github.com/ultralytics/ultralytics/issues/189 (accessed Jun. 15, 2023).