

# Real-Time Airborne Target Tracking using DeepSort Algorithm and Yolov7 Model

Yasmine Ghazlane<sup>1</sup>, Ahmed El Hilali Alaoui<sup>2</sup>, Hicham Medomi<sup>3</sup>, Hajar Bnouachir<sup>4</sup>

School of Digital Engineering and Artificial Intelligence, Euromed Research Center, Euromed University, Fes, 30110, Morocco<sup>1,2</sup>  
Research Foundation for Development and Innovation in Science and Engineering (FRDISI), Casablanca, 16469, Morocco<sup>1,3,4</sup>  
Engineering Research Laboratory (LRI), System Architecture Team (EAS) National and High School of Electricity and Mechanic (ENSEM), Hassan II University, Casablanca, Morocco<sup>3,4</sup>

**Abstract**—In light of the explosive growth of drones, it is more critical than ever to strengthen and secure aerial security and privacy. Drones are used maliciously by exploiting some gaps in artificial intelligence and cybersecurity. Airborne target detection and tracking tasks have gained paramount importance in various domains, encompassing surveillance, security, and traffic management. As airspace security systems aiming to regulate drone activities, anti-drones leverage mostly artificial intelligence and computer vision advances in the used detection and tracking models to perform effectively and accurately airborne target detection, identification, and tracking. The reliability of the anti-drone systems relies mostly on the ability of the incorporated models to satisfy an optimal compromise between speed and performance in terms of inference speed and used detection evaluation metrics since the system should recognize the targets effectively and rapidly to take appropriate actions regarding the target. This research article explores the efficacy of DeepSort algorithm coupled with YOLOv7 model in detecting and tracking five distinct airborne targets namely, drones, birds, airplanes, daytime frames, and buildings across diverse contexts. The used DeepSort and Yolov7 models aim to be used in anti-drone systems to detect and track the most encountered airborne targets to reinforce airspace safety and security. The study conducts a comparative analysis of tracking performance under different scenarios to evaluate the algorithm's versatility, robustness, and accuracy. The experimental results show the effectiveness of the proposed approach.

**Keywords**—Real-time detection; target tracking; anti-drone; Artificial Intelligence; Computer Vision

## I. INTRODUCTION

The importance of addressing aerial privacy and security issues associated with drones is growing steadily, emphasizing the critical importance and need for reliable target detection and tracking models for effective airspace security systems, such as anti-drone systems.

The deployment of an anti-drone system, which is known also as a counter-drone system depends mostly on the performance of the detection and tracking modules to detect and identify the most encountered airborne targets to avoid triggering false alarms [1], [2]. It is important to note that the detection and identification tasks are very crucial for the success of the anti-drone process and mainly to avoid neutralizing friendly airborne targets such as birds. Thus, the anti-drone system should recognize the target properly without

confusion that could cause weighty damage during the interception phase [3]. Airborne target detection and tracking is crucial in numerous applications, including defense, civilian security, and urban planning. An anti-drone system's effectiveness depends significantly on the detection of the encountered airborne targets [3], [4], [5]. It is important that an anti-drone recognizes and distinguishes between the main types of airborne targets. They share the same airspace and altitudes; mostly the low altitude airspace up to 32 000ft as an upper limit. Due to their similarity, recognizing flying targets at this altitude becomes a real challenge, which increases the probability of false detection. To reinforce and improve the anti-drone process, there is a need to develop suitable detection and tracking models able to meet the requirements and the existing needs. There are several challenges related mainly to the complexity of recognizing effectively and rapidly drones and other airborne targets present in the sky sharing many characteristics with drones that mislead the system. Further, the research tasks related to tracking multiple airborne targets have not been thoroughly studied in the existing literature.

In this study, we aim to develop an advanced detection and tracking model that can identify and track the most common targets in the sky. Computational experiments are conducted through the training, validation, and testing of a model on real-world data. This model is practical based on computational results. Therefore, integrating DeepSort with YOLOv7 is promising due to its real-time object detection capabilities and tracking precision. In addition, using DeepSort for tracking coupled with YOLOv7 for detection offers a significant approach compared to individual applications of these models. This research article aims to contribute to advancing airborne target tracking technology, evaluating the DeepSort with YOLOv7 algorithm's effectiveness in tracking diverse targets under varying scenarios. In the following, Section II provides a summary of the research studies on airborne target detection and tracking, whereas the solving methodology and experimental setup are presented in Section III. The experimental details and results are highlighted in Section IV. We conclude by discussing the advantages and limitations of the proposed model in Section V.

## II. LITERATURE REVIEW

This section delineates existing methods for airborne target tracking, emphasizing the advancements and limitations. It

covers various algorithms and their applications in tracking airborne targets.

The rise of Artificial Intelligence (AI) has improved many conventional applications, systems and tasks in several domains such as, autonomous cars, smart cities, smartphones, smart truck distribution [6] and pandemic detection [7].

Recently, the field of airborne targets detection and tracking has witnessed significant advancements, driven by the outstanding rise of drones across various sectors. Several research studies have addressed the challenges and risks posed by the proliferation of drones, necessitating the development of robust anti-drone systems capable of accurate detection and tracking.

In study [8], the authors present the small drone tracking results using the radar-based range estimation, as well as the receding horizon tracking model of unauthorized drones through the use of the receding horizon maximization technique and the fisher information matrix predictive model. Combining these two approaches yields the best localization results. The tracking approach proposed in study [9] uses a time difference of arrival estimation algorithm based on Gauss priori probability density functions with Kalman filters. The combination of these models achieved good results for drone tracking. Also, the paper in [10] has proposed a tracking model which analyses the generated acoustic signatures of the drones using beamforming algorithm. The conducted experiments show that depending on the type of the drones, they can be tracked up to 250 meters. Further, a radar tracking and detection method based on phase-interferometry and joint range-Doppler-azimuth processing is presented in study [11]. All of the extracted features from the developed model are used to classify drones. Another drone position tracking proposed in study [12] uses the received signal strength indication (RSSI) signals to estimate the distance and angle of the target to track the aerial target. It uses the estimated distance and angle to gradually track the target through the incorporation of CDQA and ADCA algorithms. The implementation of the proposed approach is not implemented on real environment.

The combination of DeepSort and Yolo has been used in different detection and tracking applications. The authors in study [13] have used a combination of YOLOv3 and RetinaNet for generating detections in each frame along with DeepSort algorithm to track multiple objects from a drone-mounted camera. Comparing the results of the experiment with the existing state-of-the-art models, the detection and tracking combination shows competitive performances on VisDrone 2018 dataset. Similarly, the paper [14] has proposed to use Yolov4 to detect and localize vehicles within the restricted zone along with DeepSort to track them to reinforce aerial surveillance.

To the best of our knowledge, there has been no study that has utilized DeepSort in conjunction with YOLOv7 to detect and track multiple airborne targets specifically for deployments in anti-drone systems.

However, the existing studies on airborne tracking have shown a limited exploration of the specific research aspect

addressed in this study, specially tracking the most common airborne targets. The majority of existing research in this domain has predominantly concentrated on drone detection only, with comparatively less emphasis on the comprehensive investigation of the aspect central to our research. Motivated by these limitations and the need for a comprehensive tracking approach, this paper proposes a novel DeepSort algorithm integrated with the YOLOv7 detection model. By combining the real-time detection capabilities of YOLOv7 with the robust object association and tracking features of DeepSort, our proposed methodology aims to address the existing gaps and improve the state-of-the-art in airborne targets detection and tracking.

### III. SOLVING METHODOLOGY

In this study, we propose the use of DeepSort algorithm with the YOLOv7 model for detecting and tracking the most encountered airborne targets. This methodology section outlines the dataset used, model training process, hyperparameters, and evaluation metrics.

#### A. Data Collection

Our developed detection and tracking models are trained on the most encountered airborne targets in the sky which anti-drone systems should recognize rapidly and effectively without causing false alarms. We have gathered a diverse dataset containing video and images with labeled bounding boxes around the targets of interest. The airborne targets in the dataset comprise drones, birds, airplanes, daytime frames, and buildings, which reflects the complexity of real-world detection and tracking scenarios. Indeed, we have trained our detection on five airborne target classes, namely drones, birds, airplanes, dayframes, and buildings. The images are collected mainly from [15], [16], [17] and annotated according to the Yolo format: object class, x, y, width, and height in the corresponding annotation text files. Following, we have used videos from [18] to perform the tracking process. The provided videos highlight mostly drones in different contexts and under different conditions. Furthermore, the used dataset ensures variability in lighting conditions, backgrounds, sizes, orientations, and occlusions to improve the used algorithm's performance and robustness.

#### B. DeepSort Model

Deep Simple Online and Real-time Tracking (DeepSort) is primarily a tracking model that works in conjunction with single-shot and two-shot object detection models, such as You Only Look Once (YOLO) to track targets and objects in real-time across frames in a video sequence [19]. The DeepSort is used for multiple target tracking in videos. It combines two main components: a detection model (like YOLO, SSD, or Faster R-CNN) that identifies the targets in each frame from the video, and a tracking algorithm that maintains the identity of these objects across frames and follows closely the motion of the targets. Initially, the detection model identifies the class of the target and generates bounding boxes around these targets, providing also their corresponding positions and labels. Following, DeepSort extracts relevant features from the detected bounding boxes, which represent the visual characteristics and appearance of the targets. The numerical representation of these features is typically created by a neural

network that performs the detection and tracking. Furthermore, using these extracted features, DeepSort associates objects detected in different frames. The DeepSort algorithm associates detections across frames. Matching detections and maintaining consistent identities through the consecutive frames is achieved by minimizing costs, such as the Hungarian algorithm. Also, DeepSort uses a prediction mechanism to maintain track of occluded or temporarily undetected objects until they reappear. As a result, temporary occlusions are less likely to occur. For each target identified in the video, DeepSort produces a continuous set of tracks related to their motion. In addition, the generated tracks contain the unique identity of targets across frames, allowing comprehensive analysis of object movement. In this research study, using Yolov7 as input, The DeepSORT algorithm perform tracking of different airborne targets. In situations with dynamic aerial movement and occlusions, DeepSort's capacity to associate and track objects across frames by utilizing appearance features and motion information is essential for preserving identities.

Therefore, in an anti-drone system, YOLOv7 operates as the initial detection module, processing input data to identify potential airborne targets. Detected objects, along with their bounding box coordinates and confidence scores, are then passed to DeepSort for tracking. DeepSort associates detections across frames, maintaining tracks for each identified target. The integrated system continuously analyzes the behavior of airborne targets, enabling real-time monitoring and threat assessment.

### C. Experimental Setup

The experiments are conducted on a local machine using a NVIDIA Quadro P4000, an Intel(R) Xeon(R) W-2155@ 3.30GHz with 32GB of main memory, and Windows® as the operating system. As well as that, we have used Pytorch version 1.13.1 along with Cuda 11.6 and Cudnn v8.8 for running the Yolo model. Furthermore, selecting appropriate hyperparameters is crucial for the training and tuning of the developed model. Our process of tuning hyperparameters entails carefully adjusting their corresponding values to find optimal rates. Table I shows the hyperparameters used during training.

TABLE I. THE HYPERPARAMETERS USED

Hyperparameter	Value	Hyperparameter	Value
Epochs	150	Learning rate (lr)	0.01
Batch size	16	Weight decay	0.005
Image size	640×640	Momentum	0.937
Warmup epochs	3	Scale	0.5
Left and right flip	0.5	Translate	0.1

### D. Evaluation Metrics

Evaluation metrics for our detection models include the following assessment metrics: Recall (R), Precision (P), Mean Average Precision (mAP), F1 score, and Frames Per Second (FPS). This allows a better evaluation of the developed models for multi-class detection since it utilizes verified and missed detection samples associated with the detection of each class target, such as False Positives (FP), False Negatives (FN), True

Positives (TP) and True Negatives (TN). When it comes to P, the relevant detection results are considered, while recall is the total number of correct detections. Equations for determining these evaluation metrics are shown below:

$$R = \frac{\sum TP}{\sum TP + FN} \quad (1)$$

$$P = \frac{\sum TP}{\sum TP + FP} \quad (2)$$

For each category, the mean Average Precision represents the overall area under the precision-recall curve.

$$mAP = \frac{1}{N} \sum_{i=1}^N AP_n \quad (3)$$

where,  $N$  is the number of target classes and  $AP_n$  is the mean mean average precision for each class.

$$FPS = \frac{\sum_{i=1}^k I}{Inferencetime} \quad (4)$$

where, I is the total number of images used in the inference phase.

## IV. RESULTS AND DISCUSSION

Using the tracking algorithm DeepSort and our detection model Yolov7, we present the results of the experiments conducted in this section. A detailed analysis of the quantitative metrics follows, providing insights into the strengths of the proposed approach.

### A. Detection Performance

As first step, the detection model is run to recognize and identify the airborne targets efficiently and accurately. We have compared several single-shot object detector algorithms [20] to select the suitable model that satisfies the speed performance compromise required for anti-drone deployment. The efficacy of our developed model has been demonstrated in Table II, which details a selection of the experimental results of the most used models, based on detection performance confidence scores and inference speed. Therefore, it is shown that Yolov7 model surpasses the other models based on the provided results of detection metrics and inference speed. Also, the Yolov7 model reaches high accuracy and fast speed, comprises between speed performances. In addition, we have completed the model analysis by comparing the detection performance with respect to the precision and recall of each class in Table III. Indeed, it is shown that the model effectively detects the targets reaching high rates of the used evaluation metrics.

Fig. 1 shows the detection performance of the selected Yolov7 model as well as the generated loss and the behavior at each epoch during the training. The training and validation behaviors of the model are described in detail with respect to the aforementioned metrics; recall, precision, mAP@0.5 and

mAP@0.5–0.95, as well as objects, classes and boxes losses. The curves converge to a fixed threshold after training for 150 epochs. Additionally, the model has demonstrated both optimal performance and high generalization ability without bias, variance, or overfitting or underfitting. Training and validation curves have similar behaviors with no gaps between them, and they converge at the same time ( $\approx 75$  epochs). This model proves its effectiveness by continuously recording loss, precision, recall, and mAP metrics during training and validation processes.

As shown in Fig. 2 to Fig. 5, we have generated the evolution of the R, P and F1 curves with respect to the confidence score to provide deeper insight into the model. As can be seen from the curves, birds, drones, and airplane targets have similar and close detection behaviors, except for the building and dayframe classes. The precision, recall, and F1 curves show that the detection performances of all categories are above 90%. The precision-recall curve (see Fig. 3) shows that the model has a 96.8% mAP (area under the curve), which corresponds to a 96.8% precision-recall rate. Also, the precision-recall curve shows that the threshold performance metrics for bird is 0.997, drone 0.973, dayframe is 0.994, airplane 0.959 and building is 0.989. This percentage value also indicates whether the model is able to detect targets while guaranteeing a satisfactory recall and precision rate. As shown in F1-confidence curve (see Fig. 5), the confidence score is set at 0.467, which is important since starting from this point, the metrics are optimized and the performance balance is achieved. Additionally, the F1 curve shows the weighted harmonic of precision and recall, as well as the optimized confidence threshold of 0.467, which is highly required to perform an accurate, real-time detection. Other evaluation criteria such as confusion matrices, inference times, and real-time detection images have emphasized the model's performance. The confusion matrix of our model is shown in Fig. 6. There are

five target classes with true positives located along the diagonal in dark blue. According to the true positive values, the proposed model is very effective and efficient at detecting and identifying the types of drones. To make the results more intuitive, we have integrated visualization about the detection performance of the model. Fig. 7 shows the detection of the airborne targets using unseen random images that confirm the model's ability to detect the aforementioned target classes.

TABLE II. EXPERIMENTAL RESULTS OF DETECTION MODELS

Model	Precision	Recall	mAP@0.5	mAP@0.5-0.95	Inference time (ms)
Yolov5	0.916	0.902	0.923	0.701	65.3
Yolov6	0.834	0.87	0.891	0.64	74
Yolov8	0.897	0.868	0.921	0.709	46
Yolov7-d6	0.957	0.928	0.966	0.712	38.5
Yolov7-w6	0.957	0.942	0.968	0.711	29.5
<b>Yolov7</b>	<b>0.973</b>	<b>0.957</b>	<b>0.982</b>	<b>0.753</b>	<b>27.5</b>

TABLE III. PERFORMANCE OF THE MODEL WITH RESPECT TO EACH TARGET CLASS

Target	Precision	Recall	mAP@0.5	mAP@0.5-0.95
all	0.973	0.957	0.982	<b>0.753</b>
Bird	0.997	0.997	0.997	0.85
Drone	0.956	0.958	0.973	0.549
Dayframe	0.995	0.966	0.994	0.79
Airplane	0.931	0.959	0.959	0.776
Building	0.989	0.912	0.989	0.798

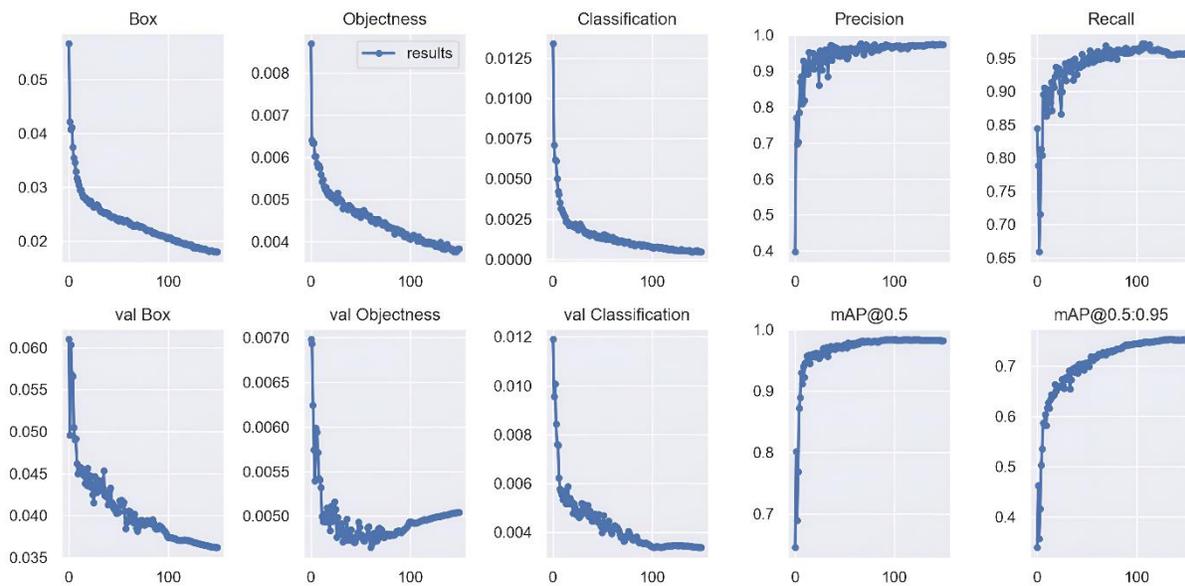


Fig. 1. Performance behavior of the improved model.

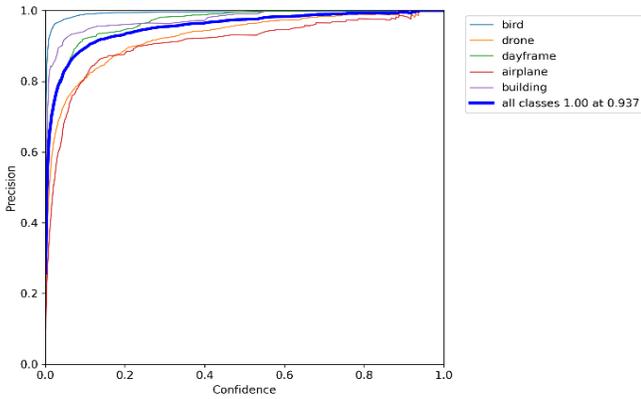


Fig. 2. Evolution of the performance of YOLOv7 model with respect to Precision evolution curve over the target classes.

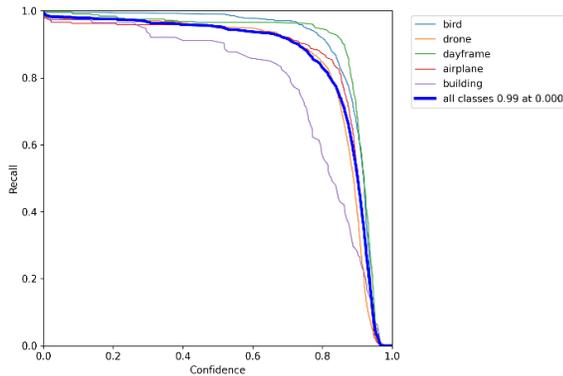


Fig. 3. Evolution of the performance of the YOLOv7 model with respect to Precision- Recall evolution curve over the target classes.

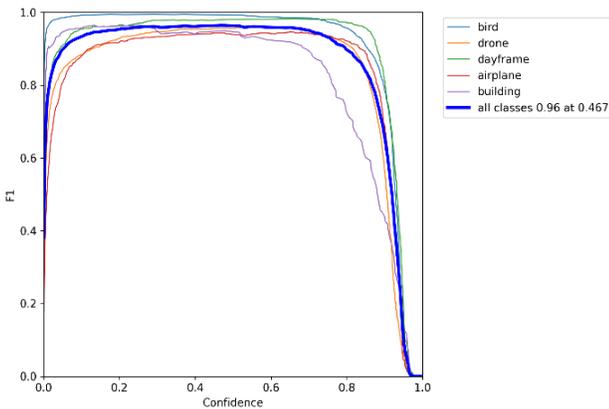


Fig. 4. Evolution of the performance of YOLOv7 model with respect to Recall evolution curve over the target classes.

Furthermore, the developed model is capable of detecting the most encountered targets in real-time. The model has average inference of 27.5 ms, 1.1 ms for Non Max Suppression (NMS) process and an ability to infer images in 0.002 seconds

per image. Additionally, the FPS metric also determines the capacity of the model to process a set of images per second, which is dependent on the model's performance. Considering that the model is tested on 1179 images, it reaches a frame rate of 42, 8 FPS. This model represents the optimal performance speed compromise, as well as being qualitative and quantitative. Additionally, the model's performance was evaluated using unseen images containing airborne targets that were barely visible to the human eye under complex conditions, mainly due to the altitudes and distances with respect to the observer.

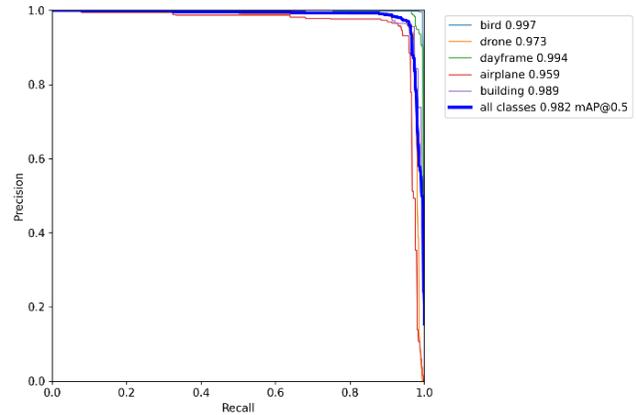


Fig. 5. Evolution of the performance of YOLOv7 model with respect to F1 evolution curve over the target classes.

In view of the provided results, we confirm that our proposed model has the high inference time and the best precision, recall mAP@50 and mAP@50-95, thus outperforming the other models proposed in the literature.

Additionally, we have used performance and speed

Evaluation metrics that is suitable for our requirements and constraints to assess effectively the tested models during the development of our final YOLOv7 model.

### B. DeepSort Tracking Results

Using the selected model YOLOv7, we have performed the tracking on different videos to assess the tracking ability. As part of the current study, DeepSort is used to track airborne targets for anti-drone deployment. It uses the patterns learned from the pre-trained YOLOv7 detection model and later combines that with temporal information to predict associated targets' trajectories. The system keeps track of all objects by mapping their unique identifiers [21]. We have deployed the DeepSort on real-time videos that contain drones in different contexts, environments and times of the day to assess its ability to keep the target within the field of view. Fig. 8(a) and Fig. 8(b) represents a selection the tracking deployed on real-time video sequences. It is shown that model efficiently identifies and tracks the airborne targets, mainly drones and dayframes presents in the videos.

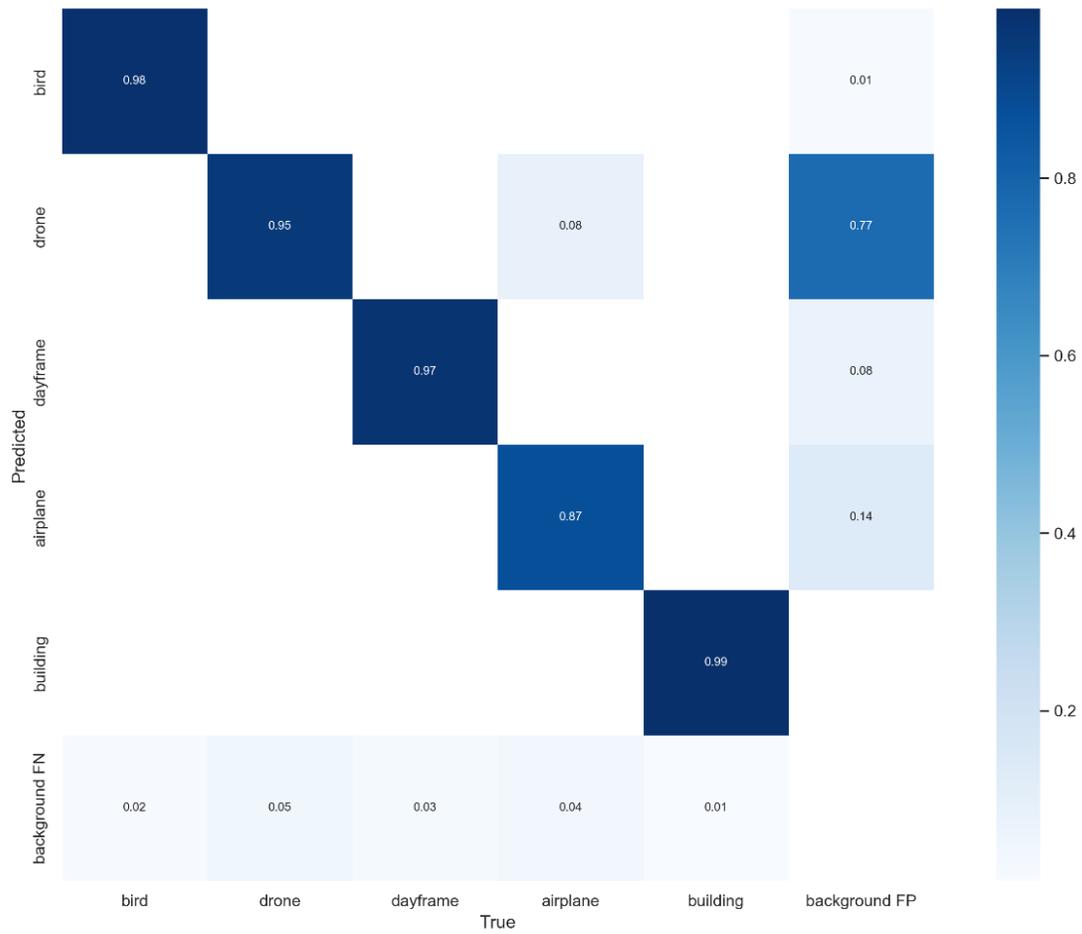


Fig. 6. Confusion matrix.

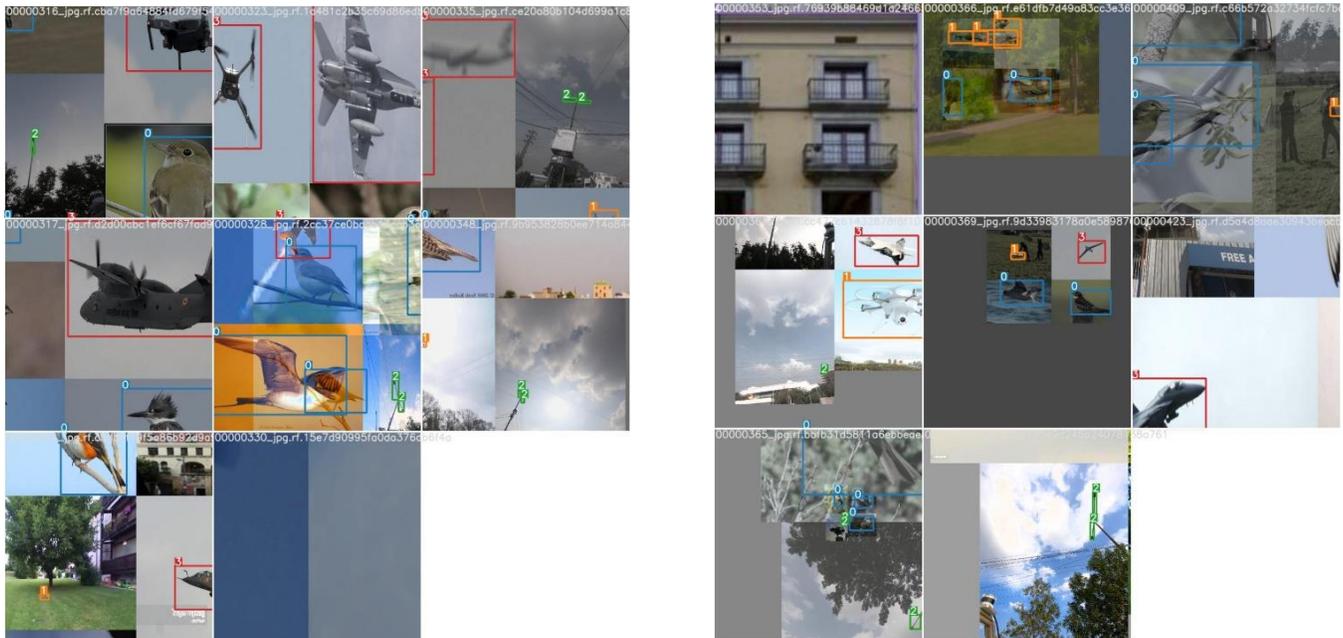


Fig. 7. Detection results of Yolov7 model.

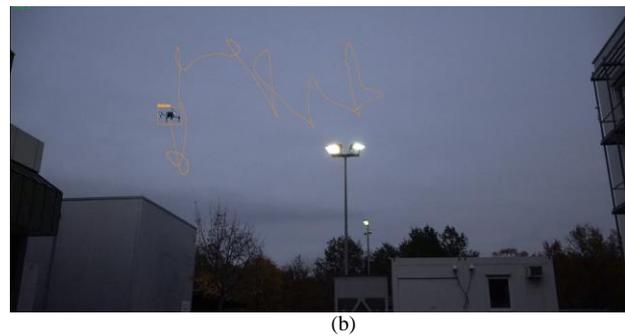
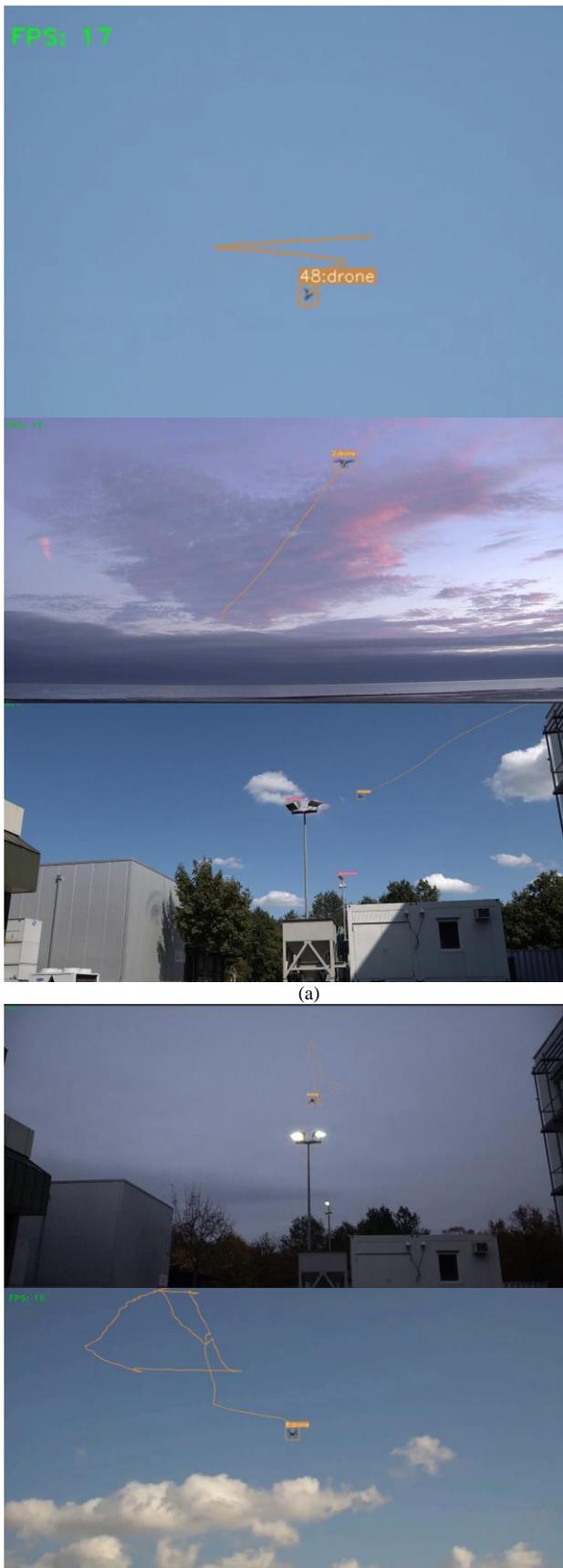


Fig. 8. (a) Captures from the tracking on different video sequences, (b) Captures from the tracking on different video sequences.

We demonstrate the efficacy of DeepSort in our experiments in tandem with the pre-trained YOLOv7 model for tracking the selected airborne targets. With robust tracking accuracy across diverse scenarios and ability to handle challenges such as target occlusions and rapid movements, the integrated system demonstrated the ability to perform effectively in challenging scenarios.

In comparison with the state-of-of the-art papers [8], [9], [10], [11], [12], the proposed tracking approach outperforms the other proposed models based on the ability of our model to detect and identify five airborne targets using a varied and diversified dataset and also the high tracking performance to track the detected targets while generating bounding box around it and drawing its motion line across the successive frames. Further, the proposed model is able to detect, identify and track multi-airborne targets at different views, capture angles and environment which enhance significantly the overall performance. Therefore, the proposed DeepSort algorithm with Yolov7 provides the best compromise between the performance and speed and thus satisfying the anti-drone requirements and challenges.

## V. CONCLUSION AND PERSPECTIVES

Detecting and tracking airborne targets represent important task for the effectiveness of an anti-drone process. Our paper presents a real-time model for identifying and tracking common airborne targets. Based on experimental results, the models have 42.8 FPS detection speed, 0.957 recall, 0.973 precision, 0.732, 0.982 map@0.50–0.95 and 0.753 map@0.50–0.95. In comparison with various benchmark instances recently published in the literature, the proposed model provides a high detection rate and fast inference times. Therefore, the combination of DeepSort algorithm and Yolov7 model provides high detection and tracking performance tested on real-time videos. The conducted experiments showed satisfactory results since the targets are detected and tracked rapidly and effectively across the successive frames of the videos. It suggests potential enhancements and future research directions to improve the algorithm's efficacy. In future work, we are going to collect a larger video dataset including also airplanes, and birds in the same sequences to improve further the tracking process.

REFERENCES

- [1] G. Yasmine, G. Maha, and M. Hicham, "Survey on current anti-drone systems: process, technologies, and algorithms," *International Journal of System of Systems Engineering*, vol. 12, no. 3, pp. 235–270, Jan. 2022, doi: 10.1504/IJSSE.2022.125947.
- [2] Y. Ghazlane, M. Gmira, and H. Medromi, "Development of a vision-based anti-drone identification friend or foe model to recognize birds and drones using deep learning," *Applied Artificial Intelligence*, doi: 10.1080/08839514.2024.2318672.
- [3] G. Yasmine, G. Maha, and M. Hicham, "Anti-drone systems: An attention based improved YOLOv7 model for a real-time detection and identification of multi-airborne target," *Intelligent Systems with Applications*, vol. 20, p. 200296, Nov. 2023, doi: 10.1016/j.iswa.2023.200296.
- [4] V. Gopal, "Developing an Effective Anti-Drone System for India's Armed Forces," no. 370, p. 16, 2020.
- [5] S. Park, H. T. Kim, S. Lee, H. Joo, and H. Kim, "Survey on Anti-Drone Systems: Components, Designs, and Challenges," *IEEE Access*, vol. 9, pp. 42635–42659, 2021, doi: 10.1109/ACCESS.2021.3065926.
- [6] H. Bnouachir, M. Chergui, and H. Medromi, "Smart Truck Distribution In An Open-Pit Mine," vol. 15, no. 1, 2023.
- [7] V. Jain, A. Jain, V. Garg, A. Jain, M. Demirci, and M. C. Taplamacioglu, "Siamese Neural Networks for Pandemic Detection Using Chest Radiographs," vol. 14, no. 2, 2022.
- [8] I. Guvenc, F. Koohifar, S. Singh, M. L. Sichitiu, and D. Matolak, "Detection, Tracking, and Interdiction for Amateur Drones," *IEEE Commun. Mag.*, vol. 56, no. 4, pp. 75–81, Apr. 2018, doi: 10.1109/MCOM.2018.1700455.
- [9] X. Chang, C. Yang, J. Wu, X. Shi, and Z. Shi, "A Surveillance System for Drone Localization and Tracking Using Acoustic Arrays," in *2018 IEEE 10th Sensor Array and Multichannel Signal Processing Workshop (SAM)*, Sheffield: IEEE, Jul. 2018, pp. 573–577. doi: 10.1109/SAM.2018.8448409.
- [10] J. Busset et al., "Detection and tracking of drones using advanced acoustic cameras," in *Unmanned/Unattended Sensors and Sensor Networks XI; and Advanced Free-Space Optical Communication Techniques and Applications*, SPIE, Oct. 2015, pp. 53–60. doi: 10.1117/12.2194309.
- [11] M. Jian, Z. Lu, and V. C. Chen, "Drone detection and tracking based on phase-interferometric Doppler radar," in *2018 IEEE Radar Conference (RadarConf18)*, Apr. 2018, pp. 1146–1149. doi: 10.1109/RADAR.2018.8378723.
- [12] J.-M. Shin, Y.-S. Kim, T.-W. Ban, S. Choi, K.-M. Kang, and J.-Y. Ryu, "Position Tracking Techniques Using Multiple Receivers for Anti-Drone Systems," *Sensors*, vol. 21, no. 1, Art. no. 1, Jan. 2021, doi: 10.3390/s21010035.
- [13] S. Kapania, D. Saini, S. Goyal, N. Thakur, R. Jain, and P. Nagrath, "Multi Object Tracking with UAVs using Deep SORT and YOLOv3 RetinaNet Detection Framework," in *Proceedings of the 1st ACM Workshop on Autonomous and Intelligent Mobile Systems*, Bangalore India: ACM, Jan. 2020, pp. 1–6. doi: 10.1145/3377283.3377284.
- [14] R. Jadhav, R. Patil, A. Diwan, S. M. Rathod, and M. Inamdar, "Aerial Object Detection and Tracking using YOLOv4 and DeepSORT," in *2022 International Conference on Industry 4.0 Technology (I4Tech)*, Sep. 2022, pp. 1–6. doi: 10.1109/I4Tech55392.2022.9952705.
- [15] "Roboflow Universe: Open Source Computer Vision Community," *Roboflow*. Accessed: Nov. 01, 2022. [Online]. Available: <https://universe.roboflow.com/>
- [16] "Caltech-UCSD Birds-200-2011." Accessed: Mar. 07, 2022. [Online]. Available: <http://www.vision.caltech.edu/visipedia/CUB-200-2011.html>
- [17] M. L. Pawelczyk and M. Wojtyra, "Real World Object Detection Dataset for Quadcopter Unmanned Aerial Vehicle Detection," *IEEE Access*, vol. 8, pp. 174394–174409, 2020, doi: 10.1109/ACCESS.2020.3026192.
- [18] A. Coluccia et al., "Drone vs. Bird Detection: Deep Learning Algorithms and Results from a Grand Challenge," *Sensors*, vol. 21, no. 8, p. 2824, Apr. 2021, doi: 10.3390/s21082824.
- [19] N. Wojke, A. Bewley, and D. Paulus, "Simple Online and Realtime Tracking with a Deep Association Metric," *arXiv.org*. Accessed: Jan. 10, 2024. [Online]. Available: <https://arxiv.org/abs/1703.07402v1>.
- [20] G. Yasmine, G. Maha, and M. Hicham, "Overview of single-stage object detection models: from Yolov1 to Yolov7," in *2023 International Wireless Communications and Mobile Computing (IWCMC)*, Jun. 2023, pp. 1579–1584. doi: 10.1109/IWCMC58020.2023.10182423.
- [21] N. S. Punn, S. K. Sonbhadra, S. Agarwal, and G. Rai, "Monitoring COVID-19 social distancing with person detection and tracking via fine-tuned YOLO v3 and Deepsort techniques." *arXiv*, Apr. 27, 2021. Accessed: Jan. 11, 2024. [Online]. Available: <http://arxiv.org/abs/2005.01385>.