Drone-vs-Bird: Drone Detection Using YOLOv7 with CSRT Tracker

Conference Paper · June 2023

DOI: 10.1109/ICASSP49357.2023.10095146

6 authors, including:

Ajeet Kumar Verma
Indian Institute of Technology Jammu

Vinit Jakhetiya
IIT JAMMU

Sunil Jaiswal
K|Lens GmbH

DRONE-VS-BIRD: DRONE DETECTION USING YOLOV7 WITH CSRT TRACKER

Sahaj K. Mistry[1*], *Shreyas Chatterjee* [1*], *Ajeet K. Verma*[1], *Vinit Jakhetiya*[1], *Badri N. Subudhi*[2], *Sunil Jaiswal*[3]

[1]Department of Computer Science and Engineering, IIT Jammu, India
[2]Department of Electrical Engineering, IIT Jammu, India, [3]K|Lens GmbH, Germany

ABSTRACT

Drone detection has become a critical component of surveillance, but detecting small drones remains a challenge. In this work, we introduce a new drone detection scheme that uses a customized YOLOv7 combined with a tracker. Using simple YOLOv7 has some limitations, including false positives and difficulty detecting drones in complex environments. To address these issues, we propose two solutions. First, we use a simple method based on the probability or confidence level of detected objects to reduce the number of false positives. Second, we integrate YOLOv7 with object tracking to detect drones in complex environments. Our proposed method was successful in achieving a top-5 ranking in the 6^{th} edition of the Drone-vs-Bird challenge, which was organized by the ICASSP Signal Processing Grand Challenge in 2023.

1. INTRODUCTION

The increasing risk posed by small drones is a significant concern for security agencies and governments worldwide, highlighting the need for effective detection and countermeasure techniques [1]. Unfortunately, current drone detection algorithms face several limitations. For example, these algorithms struggle to differentiate between drones and birds when images or videos are captured from a distance. Additionally, they are often ineffective when the background is similar to that of the drones, leading to inefficient detection. Even the latest version of YOLOv7 fails to address these issues adequately. With this view, it is required to have new and better algorithms which can accurately detect drones.

2. PROPOSED METHOD

The proposed algorithm comprises 3 steps: 1. Initial drone detection using the YOLOv7 [2], 2. Reducing the number of false positives based on the probability/confidence score, 3. Applying the CSRT tracker [3] to detect the drones in complex environments/backgrounds based upon the Intersection over union (IoU) between YOLOv7 bounding boxes and tracker bounding boxes. The details of each step are given below:

*Authors have equal contributions.

1. ***Drone detection using the YOLOv7***: The YOLOv7 is not explicitly designed to detect drones. We have fine-tuned the YOLOv7 on the drone vs bird dataset with this view. Out of 77 videos available in this dataset, 60 videos have been used for training and the remaining for validation. The mean average precision (mAP) achieved is 0.885 for the validation dataset. The YOLOv7 is able to detect drones, but it also gives false positives (suggesting any object to be a drone, but in reality, it's not a drone). With this view, reducing the number of false positives is essential.

(a). With False Positives (before applying threshold).

(b). Without False Positives (after applying threshold).

Fig. 1: YOLOv7 Results after removing the false positives.

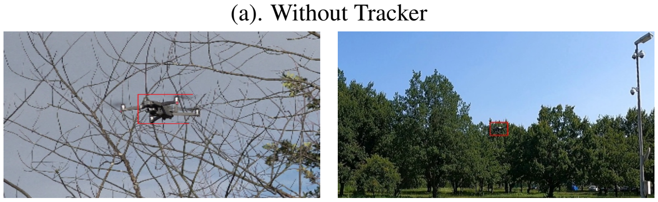
2. ***Reducing the number of false positives***: When YOLOv7 detects any object other than the drone, the associated probability/confidence score will be low. We find the number of drones in the video using a threshold and consider only that many bounding boxes throughout the video. We take the bounding boxes with the highest probability/confidence scores. For example, for some videos with more than one drone, if YOLOv7 suggests that more than 80% of the frames have more than one drone, we conclude that the video has n number of drones (where n is the number of drones in the maximum number of frames). The corresponding n-best bounding boxes with the highest probability/confidence score are used to reduce the number of false positives in near real time. This simple method is able to significantly reduce the number of false positives (as shown in Fig. 1(a) and (b)).

3. ***Tracker-based drone detection in a complex environment***: As previously discussed, YOLOv7 struggles to de-

ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) | 978-1-7281-6327-7/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICASSP49357.2023.10095146

tect drones in complex environments consistently throughout video frames. However, the algorithm can detect drones in some frames, and this information can be combined with a tracker to identify unidentified drones. While the tracker is less reliable than YOLOv7 in detecting accurate bounding boxes, we propose a method that fuses the bounding boxes estimated by both YOLOv7 and the tracker. In cases where YOLOv7 does not detect a drone, we suggest using the previous frame's YOLOv7 bounding boxes to initialize trackers. We use the bounding boxes identified by the tracker until YOLOv7 bounding boxes become available again. If YOLOv7 detects a false positive, we propose using IoU between YOLOv7 and the tracker's bounding boxes to identify the false positive. If the IoU is greater than 0.3, we consider YOLOv7's output a true positive since the AP of the bounding box would be higher. Otherwise, we consider the tracker's bounding boxes more reliable than YOLOv7's. This approach of combining a tracker with YOLOv7 significantly improves the accuracy of drone detection in complex environments, as demonstrated in Fig. 2 (a) and (b).

(a). Without Tracker



(b). With Tracker

Fig. 2: Detection of drones in a complex environment when YOLOv7 is combined with Tracker.

Test Video Name	**ScoreA**	**ScoreB**	**ScoreC**
GOPR5867_001	0.62871	0.49929	0.32863
GH010037_solo_split02	0.54624	0.99249	0.99834
GH010039_matrice_split02	0.00000	0.00000	0.00000
GH010040_inspire_split03	0.00000	0.00000	0.00000
GH010045_phantom_split01	0.56078	0.47320	0.47320
VID_20220306_170118_01	0.64274	0.64793	0.64793
VID_20220306_170541_01	0.93163	0.94461	0.94461
VID_20220311_122209_01	1.00000	1.00000	1.00000
VID_20210417_143217_01	0.01819	0.01709	0.01710
VID_20210606_141511_01	0.03189	0.03376	0.01199
GOPR5852_001	0.06006	0.05688	0.05688
GOPR5861_001	0.61269	0.56567	0.55634
VID_20211012_081448_01	0.75892	0.75893	0.75893
2019_10_16_C0003_52_30_mavic	0.20594	0.24077	0.24191
dji_mavick_mountain_cross	0.49635	0.37647	0.37510
dji_phantom_mountain	0.32743	0.15145	0.15145
GOPR5843_004	0.67512	0.62285	0.62285
GOPR5847_001	0.59494	0.29450	0.26575
GOPR5853_002	0.14476	0.15503	0.15503
GOPR5856_001	0.46520	0.44719	0.41339
GOPR5862_001	0.45517	0.28003	0.27590
GOPR5868_001	0.98531	0.98531	0.98531
VID_20210606_141851_01	0.00074	0.00026	0.00009
VID_20210606_143947_04	0.00043	0.00035	0.00035
VID_20211010_143610_01	0.99360	0.99360	0.99360
VID_20211012_175158_02	0.63755	0.03698	0.03698
4k_2020-06-22_C0006_split_01_01	0.00001	0.00107	0.00107
4k_2020-07-29_C0020_01	0.49007	0.49007	0.49007
4k_2020-07-29_C0021_01	1.00000	1.00000	1.00000
VID_20210417_143930_02	0.38964	0.11792	0.11792
Overall	**0.45047**	**0.36735**	**0.35667**

A: Results with YOLOv7 + custom filter, **B**: Results with YOLOv7 + custom filter + tracker (IOU threshold of 0.3), **C**: Results with YOLOv7 + custom filter + tracker (IOU threshold of 0.4)

Table 1: Results of the proposed algorithm on Drone-vs-Bird testing dataset (in terms of AP).

3. RESULTS AND CONCLUSION

The proposed algorithm is applied to the Drone-vs-Bird testing dataset, which consists of 30 videos with varying backgrounds. The experimental results are shown in table 1, consisting of the following: using the retrained all the parameters YOLOv7 with a custom filter and the same combined with a tracker with different IoU thresholds. It was observed that the proposed method of reduction of false positives significantly improves the performance of fine-tuned YOLOv7. At the same time, including the tracker in methods B and C with the YOLOv7 helps the proposed algorithm detect the drone in a significantly large amount of frames.

As seen in Table 1, the mAP of methods B and C is degrading. One of the possible reasons is that the tracker is detecting the drone but is not able to accurately create the bounding box around the drones (as shown in Fig. 2(a) and (b)), which results in poor mAP. In the future, better incorporation or modification of the tracker parameters may be proposed to detect drones accurately with higher mAP.

4. REFERENCES

[1] A. Coluccia et al., "Drone vs. bird detection: Deep learning algorithms and results from a grand challenge," *Sensors*, 2021.

[2] Chien-Yao Wang, Alexey Bochkovskiy, and Hong-Yuan Mark Liao, "Yolov7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," *arXiv preprint arXiv:2207.02696*, 2022.

[3] Alan Lukežič, Tom'aš Voj'iř, Luka Čehovin Zajc, Jiř'i Matas, and Matej Kristan, "Discriminative correlation filter tracker with channel and spatial reliability," *International Journal of Computer Vision*, 2018.