Track Boosting and Synthetic Data Aided Drone Detection

Fatih Cagatay Akyon, Ogulcan Eryuksel, Kamil Anil Ozfuttu, Sinan Onur Altinuc
OBSS AI
OBSS Technology, Ankara, Turkey
{fatih.akyon, ogulcan.eryuksel, anil.ozfuttu, sinan.altinuc}@obss.com.tr

Abstract

This is the paper for the first place winning solution of the Drone vs. Bird Challenge, organized by AVSS 2021. As the usage of drones increases with lowered costs and improved drone technology, drone detection emerges as a vital object detection task. However, detecting distant drones under unfavorable conditions, namely weak contrast, long-range, low visibility, requires effective algorithms. Our method approaches the drone detection problem by fine-tuning a YOLOv5 model with real and synthetically generated data using a Kalman-based object tracker to boost detection confidence. Our results indicate that augmenting the real data with an optimal subset of synthetic data can increase the performance. Moreover, temporal information gathered by object tracking methods can increase performance further.

1. Introduction

Initially being used for military applications, the use of drones has been extended to multiple application fields, including traffic and weather monitoring [5], smart agriculture monitoring [17], and many more [15]. Furthermore, with the COVID-19 pandemic, there has been a radical increase in the use of drones not only for autonomous delivery of essential grocery and medical supplies but also to enforce social distancing. Nowadays, small quadcopters can be easily purchased on the Internet at low prices, which brings unprecedented opportunities but also poses several threats in terms of safety, privacy, and security [7].

The Drone vs. Bird Detection Challenge was launched in 2017, during the first edition of the International Workshop on Small-Drone Surveillance, Detection and Counteraction Techniques (WOSDETC) [4] as part of the 14th edition of the IEEE International Conference on Advanced Video and Signal based Surveillance (AVSS). This challenge aims to address the technical issues of discriminating between drones and birds [4]. Given their characteristics, in fact, drones can be easily confused with birds, particularly at long distances, which makes the surveillance task even more challenging. The use of video analytics can solve the issue, but effective algorithms are needed that can operate under unfavorable conditions, namely weak contrast, long-range, low visibility, etc.

To overcome these issues, firstly, we use synthetic data selectively to enrich the dataset. Secondly, we make use of the Kalman filter [18] based object tracking method to track objects across time to eliminate false positives and enhance detection performance. Lastly, we propose a track boosting method for boosting the confidence scores of detections based on track statistics.

2. Related Work

In recent years, the application of deep learning-based detection methods has led to excellent results for a wide range of applications, including drone detection. However, due to the absence of large amounts of drone detection datasets, a two-staged detection strategy has been proposed in [16]. First, the authors examined the suitability of different flying object detection techniques, i.e., frame differencing and background subtraction techniques, locally adaptive change detection, and object proposal techniques [11], to extract region candidates in video data from static and moving cameras. In the second stage, a small CNN classification network is applied to distinguish each candidate region into drone and clutter categories.

In [3], Gagn and Mercier (referred to as Alexis team) proposed a drone detection approach based on YOLOv3 [13] and taking a single RGB frame as input. By integrating an image tiling strategy, this approach is able to detect small drones in high-resolution images successfully. Alexis Team leveraged the public PyTorch implementation of YOLOv3 with Spatial Pyramid Pooling (YOLOv3-SPP) made available by Ultralytics [8]. Spatial Pyramid Pooling [6] is a simple technique for which the input features are processed by pooling layers of different sizes in parallel and then concatenated to generate fixed-length feature vectors. More-
over, EagleDrone Team proposed a YOLOv5 based drone
detection modality with a linear sampling-based data sub-
sampling method. They propose to set the sampling proba-
bilities using calculated loss per image. In addition to that,
they utilize an ESRGAN based super-resolution technique
to detect small and low-resolution drones.

3. Proposed Technique

3.1. Detection Model

The proposed technique focuses on a combination of
two methods to improve the accuracy of drone detection
performance using YOLOv5 [9] object detection method.
YOLOv5 is selected because of its speed and performance
on object detection tasks. In addition, it supports anchor
optimization, which is proven to improve performance [19]
and feature pyramids [10] that handle objects at different
scales.

3.2. Synthetic Data

![Figure 1. Samples from synthetically generated drones images.](image)

The use of synthetic data in deep learning appears help-
ful in scenarios where data is scarce or unavailable. Al-
though synthetic data alone cannot show the same perfor-
mance as real data, it has been seen that it increases perfor-
mance when used alongside real data [19]. Since there is no
general method for creating synthetic data, each problem
requires a unique approach. For the drone tracking prob-
lem, a method for creating labeled, randomized composi-
tions by positioning 3D drone objects in front of 2D back-
grounds were designed. This method was chosen because it
is challenging to create a 3D randomized environment for
the drone detection problem, and a location-independent
object such as a drone can be used appropriately with 2D
backgrounds. To generate the dataset, 3D drone models
were rendered with various conditions such as position, ro-
tation and lighting, and post-process effects on the random-
ized background images. Some samples from the syntheti-
cally generated dataset can be seen in Fig. 1.

The most complex challenge encountered in the method
is developing a solution to bridge the so-called "domain
gap". To achieve this, experiments were conducted on the
properties of the generated synthetic data, and the properties
that make it similar to real data were investigated. In line
with these studies, we created datasets with varying sam-
ple sizes that were optimized based on the features discov-
ered. Finally, a series of experiments were conducted; first,
a dataset was used with mixing synthetic and real data; sec-
ond, a model trained on a synthetic dataset was used as a
backbone for real data training.

3.3. Object Tracking And Tracker Based Confi-
dence Boosting

Object tracking algorithms are used to provide contin-
uity of object detections over time. While tracking the ob-
jects is not directly required, it provides temporal informa-
tion about the objects in the video that can further improve
performance.

3.3.1 Tracker Method

A simple Kalman-based tracking method is applied [2] over
the predictions of the object detection network. Kalman-
based tracking uses a position and velocity-based pro-
cess, object detection methods as measurements, and a hit
counter mechanism. The tracking parameters are optimized
for drone tracking with possibly moving cameras.

One benefit of using a tracker system is that for a track to
be formed successfully, the object detection model needs to
provide a consistent stream of predictions close to the ob-
ject’s predicted location. Therefore false positives occurring
at random positions usually fail to build up the necessary hit
count to form a track, as shown in Fig. 2. This has a positive
impact on the mAP score.

![Figure 2. Tracker results and false positive detections on a frame.](image)

The tracks provide a spatiotemporal dimension and continu-
ity over the predictions. This information is used in various
ways to improve performance.

3.3.2 Track Boosting Method

The tracks provide a spatiotemporal dimension and continu-
ity over the predictions. This information is used in various
However, tracking can provide object predictions where the object detection method provides no predictions; tracker-only predictions are not very useful compared to detection predictions due to their low IoU rate and test set not having annotations for occluded objects. Furthermore, in conducted experiments, including the tracker predictions had a negative impact on the mAP score. Therefore only detections from the object detection method are used in the Tracker Boosting Method.

Also, object detection model confidence may vary significantly with moving objects as the object moves or changes orientation over time. With the track information provided by the tracking algorithm, we increased the confidence of the predictions within a track by averaging the max confidence score in the track with the confidence provided by the object detection algorithm as shown in Eq. (1) where \( S'_{i,j} \) is the score for a prediction with \( i \) as track number and \( j \) as the position in the track and \( s_i \) is the vector of scores for track \( i \).

\[
S'_{i,j} = \frac{S_{i,j} + \max(s_i)}{2}
\]  

4. Experiments

4.1. Used Datasets

Before conducting the experiments, we randomly selected 15 of the videos from drone-vs-bird set [19] as the validation set. Then extracted the frames from training and validation videos. We observed that using a subset of the training frames instead of the whole set prevents over-fitting resulting in improved accuracy. Therefore, uniformly sampled \( 10^4 \) frames from the drone vs. bird training is used as training data in experiments. This dataset will be referred as "Real Dataset".

Synthetic datasets with different features were generated using the method mentioned in sec. 3.2. These datasets with special features are:

- **Original** Original rendered image without specific features
- **Noise** Image rendered with film grain noise post processing effect
- **Optimal drone sizes** Drones sized according to normal distribution
- **Blur** Rendered image with Gaussian Blur optimized for backgrounds

The dataset generated with \( 10^4 \) images by optimizing these features is called "Synthetic Dataset".

To perform experiments in which synthetic data will be used alongside real data, we created a combined dataset of \( 1.05 \times 10^4 \) images which will be referred as "Combined Dataset". This dataset includes all \( 10^4 \) samples from "Real Dataset" and an optimal sub-sample of 500 images from 'Synthetic Dataset'. The reason to include just a small subset of the synthetic images is to avoid the domain gap mentioned in sec. 3.2.
Moreover, we have also trained models with a combined dataset containing images from mav-vid [14] and realworld-uav [12] drone detection datasets. However, since their distributions differ from the drone-vs-bird dataset, our drone-vs-bird validation accuracy dropped, so we do not mention combined dataset training results in this paper.

4.2. Results

Firstly, a series of experiments were performed on various synthetic datasets to find the optimal synthetic features. All experiments were performed with COCO pretrained YOLOv5m6 model of 1333 input image size, 8 batch size, and 10 epochs setup and with datasets described in 4.1. For data sampling, inference and evaluation, our open source vision package SAHI [1] is utilized. As seen in Fig. 4, experiments revealed that the most important property is the amount of blur applied to the image. In addition, it was understood that the distribution of drone sizes and noise effect similar to film grain significantly affect the performance. The results of the synthetic data experiments are used to create the optimized synthetic dataset known as ‘Synthetic Dataset’. The training results obtained with the ‘Synthetic Dataset’ show that with the correct features, a training set consisting of only synthetic images gives us acceptable results considering no real images were required in the process.

As seen in Table 1, real data gives better results than synthetic data; however, augmenting real data with synthetically generated data improves the validation results by up to 4.2 AP in all scenarios. Moreover, by applying a Kalman filter-based tracker, base results can be improved by up to 1 AP. More importantly, applying the track boosting method on top of a tracker provides us with an additional 1.5 and 0.6 AP improvement in real and combined dataset experiments, respectively.

5. Conclusion

Our results show that a YOLOv5 model fine-tuned only on synthetically generated images can achieve acceptable performance on drone detection tasks. Moreover, mixing an optimal subset of synthetic data yielded much better results than using real and synthetic images by themselves. Usage of the tracker improves upon the object detection performance in all cases. This improvement may be a result of filling out missing frames and eliminating the false positives by the tracker’s internal mechanism. These results can be further improved by adjusting the frame predictions using the track information. Using the maximum confidence value in a track as a reference value, the overall mAP score increased. In cases where both the tracking algorithm and the object detection provide a prediction using the prediction from the object detection model results in improved accuracy as well.

References

[1] F. C. Akyon, S. O. Altinuc, and A. Temizel. Slicing aided hyper inference and fine-tuning for small object detection. arXiv preprint arXiv:2202.06934, 2022.

[2] J. Alori, A. Descoins, B. Ros, and A. Castro. tryolabs/norfair: v0.3.1. https://github.com/tryolabs/norfair, 2021.

[3] A. Coluccia, A. Fascista, A. Schumann, L. Sommer, A. Dibou, D. Zarpalas, M. Méndez, D. De la Iglesia, I. González, J.-P. Mercier, et al. Drone vs. bird detection: Deep learning algorithms and results from a grand challenge. Sensors, 21(8):2824, 2021.

[4] A. Coluccia, A. Fascista, A. Schumann, L. Sommer, M. Gheneuscu, T. Piatrik, G. De Cubber, M. Nalamati, A. Kapoor, M. Saqib, et al. Drone-vs-bird detection challenge at ieee avss2019. In 2019 16th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), pages 1–7. IEEE, 2019.

[5] M. Elloumi, R. Dhaou, B. Escrig, H. Idoudi, and L. A. Saidane. Monitoring road traffic with a uav-based system. In 2018 IEEE Wireless Communications and Networking Conference (WCNC), pages 1–6. IEEE, 2018.

[6] K. He, X. Zhang, S. Ren, and J. Sun. Spatial pyramid pooling in deep convolutional networks for visual recognition. IEEE transactions on pattern analysis and machine intelligence, 37(9):1904–1916, 2015.
[7] T. Humphreys. Statement on the security threat posed by unmanned aerial systems and possible countermeasures. *Over-sight and Management Efficiency Subcommittee, Homeland Security Committee, Washington, DC, US House*, 2015. 1

[8] G. Jocher, Y. Kwon, J. Veitch-Michaelis, Marc, G. Bianconi, F. Baltac, D. Suess, W. Xinyu, T. M. Shead, T. Havlik, P. Skalski, J. Hu, F. Reveriano, Falak, and D. Kendall. ultralytics/yolov3: 43.1map@0.5:0.95 on coco2014. https://github.com/ultralytics/yolov3, 2020. 1

[9] G. Jocher, A. Stoken, J. Borovec, NanoCode012, ChristopherSTAN, L. Changyu, Laughing, tkianai, A. Hogan, Lorenzomammana, yxNONG, AlexWang1900, L. Diaconu, Marc, wanghaoyang0106, ml5ah, Doug, F. Ingham, Frederik, Guilhen, Hatovix, J. Poznanski, J. Fang, L. Yu, changyu98, M. Wang, N. Gupta, O. Akhtar, PetrDvoracek, and P. Rai. ultralytics/yolov5: v3.1 - bug fixes and performance improvements. https://github.com/ultralytics/yolov5, Oct. 2020. 2

[10] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie. Feature pyramid networks for object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2117–2125, 2017. 2

[11] T. Müller. Robust drone detection for day/night counter-uav with static vis and swir cameras. In *Ground/Air Multisensor Interoperability, Integration, and Networking for Persistent ISR VIII*, volume 10190, page 1019018. International Society for Optics and Photonics, 2017. 1

[12] M. Paweczyk and M. Wojtyra. Real world object detection dataset for quadcopter unmanned aerial vehicle detection. *IEEE Access*, 8:174394–174409, 2020. 4

[13] J. Redmon and A. Farhadi. Yolov3: An incremental improvement. *arXiv preprint arXiv:1804.02767*, 2018. 1

[14] A. Rodriguez-Ramos, J. Rodriguez-Vazquez, C. Sampedro, and P. Campoy. Adaptive inattentional framework for video object detection with reward-conditional training. *IEEE Access*, 8:124451–124466, 2020. 4

[15] H. Shakhatreh, A. H. Sawalmeh, A. Al-Fuqaha, Z. Dou, E. almaita, I. Khalil, N. S. Othman, A. Khreishah, and M. Guizani. Unmanned aerial vehicles (uavs): A survey on civil applications and key research challenges. *IEEE Access*, 7:48572–48634, 2019. 1

[16] L. Sommer, A. Schumann, T. Müller, T. Schuchert, and J. Beyerer. Flying object detection for automatic uav recognition. In *2017 14th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, pages 1–6. IEEE, 2017. 1

[17] P. Tokekar, J. Vander Hook, D. Mulla, and V. Isler. Sensor planning for a symbiotic uav and ugv system for precision agriculture. *IEEE Transactions on Robotics*, 32(6):1498–1511, 2016. 1

[18] G. Welch, G. Bishop, et al. An introduction to the kalman filter. 1995. 1

[19] J. Wojtanowski, M. Zygmunt, T. Drozd, M. Jakubaszek, M. Życzkowski, and M. Muzal. Distinguishing drones from birds in a uav searching laser scanner based on echo depolarization measurement. *Sensors*, 21(16):5597, 2021. 2, 3