Drone-vs-Bird Detection Challenge at IEEE AVSS2019

Angelo Coluccia, Alessio Fascista
University of Salento, Department of Engineering
via Monteroni, 73100 Lecce, Italy
angelo.coluccia@unisalento.it

Arne Schumann[†], Lars Sommer[‡,†]
[†]Fraunhofer IOSB
Fraunhoferstrasse 1, 76131 Karlsruhe, Germany
[‡]Vision and Fusion Lab, Karlsruhe Institute of Technology KIT
Adenauerring 4, 76131 Karlsruhe, Germany
firstname.lastname@iosb.fraunhofer.de

Marian Ghenescu
UTI Grup, Romania
107A Oltenitei Avenue, 041393 Bucharest 4, Romania
marian.ghenescu@uti.eu.com

Tomas Piatrik
School of Electronic Engineering and Computer Science, Queen Mary University of London
Mile End Road, London, E1 4NS, United Kingdom
tomas.piatrik@qmul.ac.uk

Geert De Cubber
Royal Military Academy of Belgium, Department of Mechanics, Unmanned Vehicle Centre
30, Av. De La Renaissance, 1000 Brussels, Belgium
geert.decubber@rma.ac.be

Mrunalini Nalamati, Ankit Kapoor, Muhammad Saqib, Nabin Sharma, Michael Blumenstein
University of Technology Sydney
Broadway, Ultimo, NSW 2007, Australia
nalamati@student.uts.edu.au

Vasileios Magoulianitis, Dimitrios Ataloglou, Anastasios Dimou,
Dimitrios Zarpalas, Petros Daras
Information Technologies Institute, Centre for Research and Technology Hellas (CERTH)
6th km Harilaou - Thermi, 57001, Thessaloniki, Greece
magoulianitis@iti.gr

Celine Craye, Salem Ardjoune
CerbAir Research Lab
Boulogne-Billancourt, France
celine.craye@cerbair.com

David de la Iglesia, Miguel Méndez, Raquel Dosil, Iago González
Gradiant - Galician Research and Development Center In Advanced Telecommunications
Fonte das Abelleiras, s/n - CITEXVI - 36310 Vigo, Spain
mmendez@gradiant.org

Abstract

This paper presents the second edition of the "drone-vs-bird" detection challenge, launched within the activities of the 16-th IEEE International Conference on Advanced Video and Signal-based Surveillance (AVSS). The challenge's goal is to detect one or more drones appearing at some point in video sequences where birds may be also present, together with motion in background or foreground. Submitted algorithms should raise an alarm and provide a position estimate only when a drone is present, while not issuing alarms on birds, nor being confused by the rest of the scene. This paper reports on the challenge results on the 2019 dataset, which extends the first edition dataset provided by the SafeShore project with additional footage under different conditions.

1. Introduction

Adopting effective detection and countermeasure techniques to face the rising threat of small drones — whose payload capability can be exploited for terrorism attacks using explosives or chemical weapons, as well as for smuggling of drugs and other illegal activities — is a great concern of security agencies today. Both in the context of critical infrastructure and border protection, but also for general persecution of illegal activities and anti-terrorism activities, surveillance and detection technologies based on different modalities are under investigation, with different trade-offs in complexity, range, and capabilities [1]. Given their characteristics, drones can be easily confused with birds, particularly at long distance, which makes the surveillance tasks even more challenging. The use of video analytics can solve the issue, but effective algorithms are needed, which are able to operate under unfavorable conditions, namely weak contrast, long range, low visibility, etc.

In 2017 the first edition of the *International Workshop on Small-Drone Surveillance, Detection and Counteraction Techniques* (WOSDETC) [2] was organized in Lecce, Italy, as part of the 14th edition of the *IEEE International Conference on Advanced Video and Signal based Surveillance* (AVSS). Supported by the SafeShore Consortium[1], one of the initiatives was to launch the *drone-vs-bird detection challenge*, whose aim was to address one of the main issues arising in the described context [3, 4]. A second edition of the challenge has been launched in 2019, and this paper reports on its definition and results. The challenge was again part of WOSDETC, co-located with the 16th edition of AVSS and held in Taipei, Taiwan. The challenge aimed at attracting research efforts to identify novel solutions to the problem of discrimination between birds and drones, by providing an annotated video dataset recorded at shore areas in different conditions. The 2019 challenge dataset has been doubled by courtesy of the Fraunhofer IOSB institute, Karlsruhe, Germany, which provided 6 additional videos with different cameras and background so as to cover also land scenarios (not only maritime ones, as for the first edition based solely on the SafeShore footage).

[1]The project "SafeShore" has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 700643.

Figure 1. Map of the research groups participating in the challenge.

The challenge's goal is to detect drones appearing at some time in a short video sequence while not generating detections on birds and other distracting scene elements. All the participants of the challenge were asked to submit score files with their detection results and a companion paper describing the applied methodology. The worldwide distribution of the research institutions that have been interested in the 2019 challenge is shown in Fig. 1. About 15 different research groups requested access to the dataset for participation in the competition.

2. Challenge Dataset and Evaluation Metric

For the 2019 challenge a new dataset has been made available. The training data consists of a collection of 11 MPEG4-coded static-camera videos where a drone enters the scene at some point. Annotation is provided in separate files in terms of frame number and bounding box of the target given as $[top\_x\ top\_y\ width\ height]$ for those frames where drones are present. Birds or other scene elements are not annotated. Several examples of frames extracted from the training videos are shown in Fig. 2. The first four images depict maritime sequences from first challenge edition while the remaining four depict the newly added scenes. Compared to the first challenge edition, the difficulty of the task is increased by the need to cope with very diverse background and illumination conditions, as well as with different scales (zoom), viewpoints, low contrast, and presence of birds.

Two days before the challenge deadline, 3 new static-camera video sequences have been provided to participants

Figure 2. Sample frames extracted from the training videos.

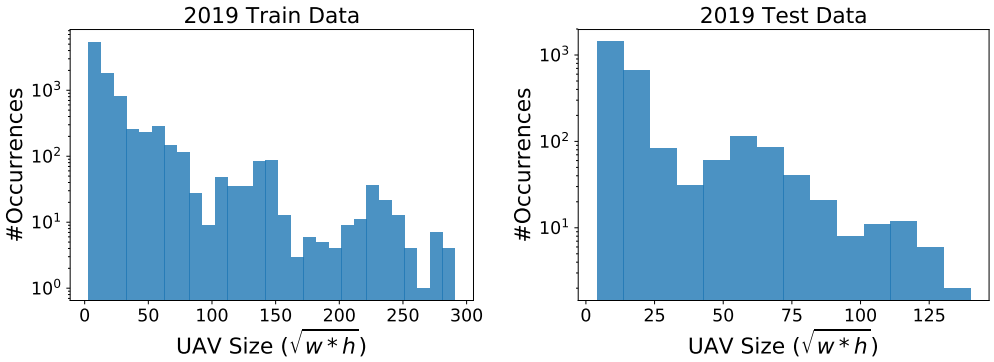

Figure 3. Distribution of UAV sizes across the ground truth annotations in train and test data. Note the large number of very small objects that had to be detected (log scale axis).

for testing their methods. While the first installment of the challenge relied on a single test video, this greater number of very diverse test videos helps to better assess the developed algorithms under various conditions and avoids saturation effects on the performance metrics. Participants were asked to submit one file per test sequence providing the frame number and estimated bounding boxes for the frames where their algorithm detects the presence of drones. For frames not reported, no detection was assumed. Only one configuration of the algorithm parameters has been allowed.

A particular challenge in the 2019 data was the large number of very distant drones in train and test data that had to be detected. Fig. 3 shows the number of ground truth annotations for various drone sizes (i.e., distances). The majority of cases required detection of objects with an average box edge width of 20 pixels or less. In the training data, the smallest annotations were of size 3 pixels and in the test data 4 pixels. The plots also illustrate the large range of detection sizes that the developed models had to be able to handle, i.e., from 3-4 pixels up to almost 300 pixels.

The submitted algorithms were evaluated by matching ground truth to detection via the well established Intersection over Union (IoU) measure, which is determined as

$$IoU = \frac{\text{detection} \cap \text{ground truth}}{\text{detection} \cup \text{ground truth}}.$$

If a detection achieves an IoU with a ground truth annotation higher than a threshold (usually 0.5), it is counted as a true positive detection (TP). Otherwise a detection is counted as a false positive (FP). Ground truth annotations which were not assigned a TP, are counted as false negatives (FN), i.e., missed detections. Based on these measures, precision and recall of the detection algorithm can be computed and aggregated into the F_1-score:

$$F_1 = 2\frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}}.$$

The final ranking of submitted methods is obtained based on the F_1-score, ranging from the best (maximum) possible value of 1.0 to the worst (minimum) value of 0.0.

3. Participation and Best Proposed Algorithms

We briefly summarize the methodology of the four best performing approaches. All approaches rely on convolutional neural networks and some of the most prominent object detection meta-architectures from literature. However, the approaches greatly differ in the way these models are used and adapted.

Nalamati et al. [5] from University of Technology, Sydney, evaluate deep learning based object detection architectures for the task of drone detection. The single-stage SSD [6] and two-stage Faster R-CNN [7] meta-architectures are applied. As backbone convolutional neural networks, the Inception-v2 [8] and ResNet-101 [9] models are chosen. The conducted evaluation finds that among the investigated architectures Faster R-CNN with ResNet-101 performs best while the single-stage SSD generally performs worse than the two-stage Faster R-CNN.

Magoulianitis et al. [10] from CERTH, Greece, propose to use super-resolution jointly with object detection in order to locate very small (i.e., distant) drones better. The DCSCN model [11] is chosen as the super-resolution component of the proposed approach and upscales the image by a factor of 2. For object detection, the Faster R-CNN detector is applied. Both models are trained jointly so that the object detector can maximally benefit from the super-resolution result.

Craye and Ardjoune [12] from CerbAir Research Lab, France, have proposed a detection approach based on two separate convolutional neural networks. First, U-Net [13] — a semantic segmentation network originally designed for medical imaging — is used to identify areas of interest within the larger image. Instead of using single RGB images as input, N grayscaled images of the video sequence with a pre-defined step T between consecutive images are

fed into the network to integrate a temporal aspect. Then, areas of interest are obtained by applying a blob finder on the background-foreground mask generated by the U-Net. Finally, a classification network, ResNet, is used to determine whether those areas contain a drone or not.

De la Iglesia et al. [14] from GRADIANT, Spain, propose a detection algorithm based on RetinaNet, which employs the Feature Pyramid Network (FPN) [15] architecture to perform predictions at different scales. While lower pyramidal levels contain fine-grained details which are useful to detect small objects, the upper levels aggregate dense spatial information that are more adequate for larger objects. As the third pyramid level (P5) already provides enough context to detect most of the larger drones, the two last feature pyramid levels (P6, P7) are discarded. Furthermore, a lower feature pyramid level is additionally considered for prediction to account for very small drones. K-means clustering is performed in order to determine appropriate anchor boxes used for bounding box regression.

The four approaches vary strongly in how they address the drone detection problem. [5] and [14] rely purely on existing object detection CNN meta-architectures. While [5] carries out an evaluation of different architecture combinations, [14] goes a step further and adapts a state-of-the-art object detection framework specifically to the task of drone detection. In contrast to this, the other two teams rely on additional components in their detection approach. In both cases, these components pre-process the image for the detection task. However, the motivation is a very different one. While one approach relies on super-resolution to improve detection of very distant drones, the other performs a pre-selection of image regions based on motion information so that the subsequent classifier can solve a more limited and potentially easier problem. [12] is the only one of the four approaches to consider temporal information. Both, approaches [12] and [10] also rely on extended training data from additional sources other than the provided challenge training set.

4. Results

For evaluation and ranking of the submitted approaches the F_1 score was computed across all three test sequences for an IoU threshold of 0.5. The results are given in Table 1. The scores of the different teams vary strongly with the two teams that use additional components for super-resolution or motion segmentation achieving the best result. The overall top score was achieved by team 3, which relied on motion information. The Faster R-CNN based approach of team 1 achieves a comparatively low score, while the specifically adapted feature pyramid network of team 4 performs quite a bit better.

A more detailed analysis of the types of errors that lead to the resulting scores is depicted in Fig. 4. The figure shows the number of true positives, misses, and false positives for various object sizes. It can be observed that teams 2 and 3 perform very accurately for objects of size 32 pixels or above. The other approaches have difficulty detecting some of these larger objects, likely due to an overly strong focus on small objects during training. Most errors for all teams occur in the very small object range below 32 pixels. Teams 2 and 3 can detect more than half of these very small objects and simultaneously produce a comparatively small number of false positive detections.

Table 1. Final scores of the top-4 algorithms on the test videos.

Approach	F_1 Score
Nalamati et al. (Team 1) [5]	0.12
Magoulianitis et al. (Team 2) [10]	0.68
Craye and Ardjoune (Team 3) [12]	**0.73**
de la Iglesia et al. (Team 4) [14]	0.41

Due to the main errors occurring on very small object sizes, we additionally evaluated with less restrictive IoU thresholds. For very small box sizes, a single pixel in deviation from the correct location can have a huge impact on the IoU measure, while the detection is still essentially correct. Figure 5 shows precision, recall, and F_1 score for all teams across different IoU thresholds. Most importantly, it can be observed that regardless of IoU value, the relative ranking of all teams remains the same. However, for lower IoU thresholds particularly the results of team 3 show that very small objects can still be detected quite accurately. This is indicated by the continuing rise in F_1 score with lower IoU. For an IoU of 0.3 recall lies above 90% while precision is close to 100%. For IoU below 0.3 saturation is observed for most teams, indicating that there are no further detections located near ground truth boxes.

Qualitative analysis of the submitted results shows that teams 1 and 4, which rely on standard object detection architectures, generate a larger amount of false positives. Particularly fixed scene elements that bare some resemblance to very distant drones cause re-occurring false positive detections. A temporal aspect to the approaches might well be able to filter these false positives. Conversely, team 3 benefits strongly from the motion segmentation element, which seems to reduce the number of false positives on background elements greatly. However, in one of the test sequences, which contains a motorway in the background, several false positives are caused by moving cars. This might indicate that the subsequent classifier after the motion segmentation could still be improved. Overall, the inclusion of additional training data, as well as motion information appears to be a very promising approach to drone detection at long distance. This is in line with observations made at the previous challenge, in which the winning entry

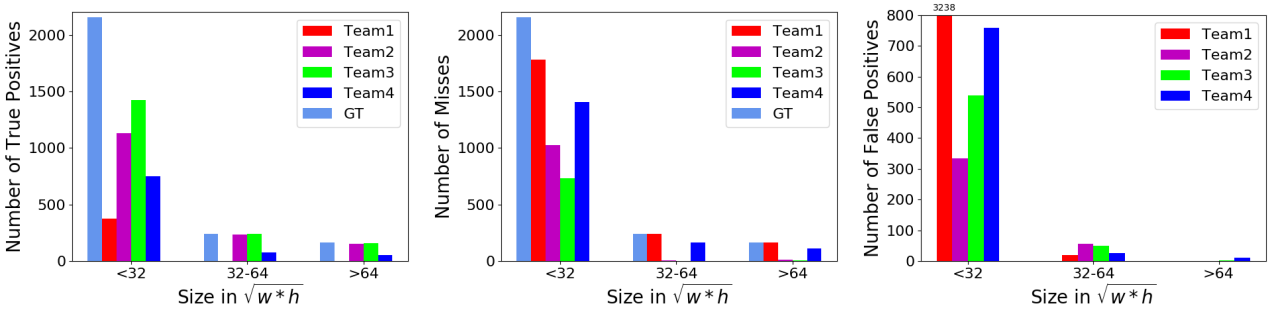

Figure 4. Numbers of true positive detections, misses, and false positives for various drone size ranges. Values were computed for IoU=0.5.

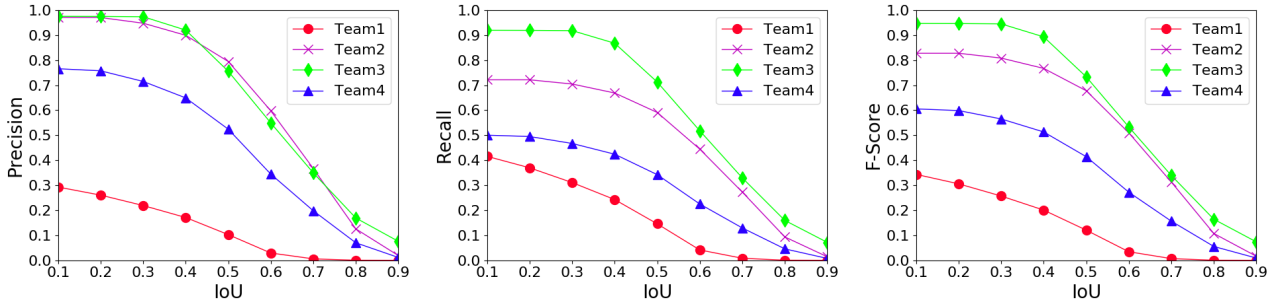

Figure 5. Precision, recall, and F_1-score curves for various IoU thresholds of the top 4 teams.

also relied on added training data and temporal information for filtering errors [16].

5. Conclusions

This work reported on the results of the 2019 drone-vs-bird detection challenge, held with the 16th IEEE International Conference on Advanced Video and Signal based Surveillance (AVSS). Compared to the previous installment, a new and more challenging drone dataset was provided and a variety of approaches were submitted that differed strongly in the methods chosen to address the problem. Ultimately, a combination of added training data and use of motion segmentation achieved the best result. However, strong results were also achieved by combining drone detection with super-resolution and reliance on recent state-of-the-art object detection architectures. Since these approaches address different aspects of the detection problem, a combination of their key properties might be an interest avenue for future research.

References

[1] Arne Schumann, Lars Sommer, Thomas Müller, and Sascha Voth. An image processing pipeline for long range uav detection. In *Emerging Imaging and Sensing Technologies for Security and Defence III; and Unmanned Sensors, Systems, and Countermeasures*, volume 10799, page 107990T. International Society for Optics and Photonics, 2018. 2

[2] Angelo Coluccia, Marian Ghenescu, Tomas Piatrik, Geert De Cubber, Arne Schumann, Lars Sommer, Johannes Klatte, Tobias Schuchert, Juergen Beyerer, Mohammad Farhadi, et al. Drone-vs-bird detection challenge at ieee avss2017. In *14th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, pages 1–6. IEEE, 2017. 2

[3] Horizon2020. The SafeShore project. In *http://safeshore.eu*, funded by the European Commission under the "Horizon 2020" program, grant agreement No 700643. 2

[4] G. De Cubber, R. Shalom, A. Coluccia, O. Borcan, R. Chamrád, T. Radulescu, E. Izquierdo, and Z. Gagov. The SafeShore system for the detection of threat agents in a maritime border environment. In *IARP Workshop on Risky Interventions and Environmental Surveillance*, Les Bons Villers, Belgium, May 2017. 2

[5] Mrunalini Nalamati, Ankit Kapoor, Muhammed Saqib, Nabin Sharma, and Michael Blumenstein. Drone detection in long-range surveillance videos. In *16th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS) Workshop*. IEEE, 2019. 4, 5

[6] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander C Berg. Ssd: Single shot multibox detector. In *European conference on computer vision*, pages 21–37. Springer, 2016. 4

[7] S. Ren, K. He, R. Girshick, and J. Sun. Faster r-cnn:towards real-time object detection with region proposal networks: Towards real-time object detection with region proposal networks. In *Advances in neural information processing systems*, 2015. 4

[8] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2818–2826, 2016. 4

[9] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016. 4

[10] Vasileios Magoulianitis, Dimitrios Ataloglou, Anastasios Dimou, Dimitrios Zarpalas, and Petros Daras. Does deep super-resolution enhance uav detection? In *16th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS) Workshop*. IEEE, 2019. 4, 5

[11] Jin Yamanaka, Shigesumi Kuwashima, and Takio Kurita. Fast and accurate image super resolution by deep cnn with skip connection and network in network. In *International Conference on Neural Information Processing*, pages 217–225. Springer, 2017. 4

[12] Celine Craye and Salem Ardjoune. Spatio-temporal semantic segmentation for drone detection. In *16th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS) Workshop*. IEEE, 2019. 4, 5

[13] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015. 4

[14] David de la Iglesia, Miguel Mendez, Raquel Dosil, and Iago Gonzalez. Drone detection cnn for close and long range surveillance in mobile applications. In *16th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS) Workshop*. IEEE, 2019. 5

[15] Tsung-Yi Lin, Piotr Dollár, Ross Girshick, Kaiming He, Bharath Hariharan, and Serge Belongie. Feature pyramid networks for object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2117–2125, 2017. 5

[16] Arne Schumann, Lars Sommer, Johannes Klatte, Tobias Schuchert, and Jürgen Beyerer. Deep cross-domain flying object classification for robust uav detection. In *14th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS) Workshop*. IEEE, 2017. 6