

## Drone-vs-bird detection challenge at IEEE AVSS2017

Angelo Coluccia

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

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

Arne Schumann<sup>†</sup>, Lars Sommer<sup>‡,†</sup>, Johannes Klatt<sup>†</sup>, Tobias Schuchert<sup>†</sup>, Juergen Beyerer<sup>†,‡</sup>  
<sup>†</sup>Fraunhofer IOSB

Fraunhoferstrasse 1, 76131 Karlsruhe, Germany  
<sup>‡</sup>Vision and Fusion Lab, Karlsruhe Institute of Technology KIT  
Adenauerring 4, 76131 Karlsruhe, Germany  
firstname.lastname@iosb.fraunhofer.de

Mohammad Farhadi, Ruhallah Amandi  
ArkaInvent

No. 3, Jahan St., Taleqani, Tehran, Iran  
{farhadi, amandi}@arkainvent.com

Cemal Aker, Sinan Kalkan

KOVAN Research Lab., Computer Engineering, Middle East Technical University  
Dumlupinar Bulvari, 06800, Ankara, Turkey  
{cemal, skalkan}@ceng.metu.edu.tr

Muhammad Saqib<sup>\*</sup>, Nabin Sharma<sup>\*</sup>, Sultan Daud Khan Makkah<sup>‡</sup>, Michael Blumenstein<sup>\*</sup>

<sup>\*</sup>University of Technology Sydney Broadway  
Ultimo, NSW 2007, Australia

<sup>‡</sup>Makkah Technology Valley  
Kingdom of Saudi Arabia

muhammad.saqib@student.uts.edu.au

## Abstract

*Small drones are a rising threat due to their possible misuse for illegal activities, in particular smuggling and terrorism. The project SafeShore, funded by the European Commission under the Horizon 2020 program, has launched the “drone-vs-bird detection challenge” to address one of the many technical issues arising in this context. The goal is to detect a drone appearing at some point in a video where birds may be also present: the algorithm should raise an alarm and provide a position estimate only when a drone is present, while not issuing alarms on birds. This paper reports on the challenge proposal, evaluation, and results<sup>1</sup>.*

## 1. Introduction

Small drones are a rising threat due to their possible misuse for illegal activities such as smuggling of drugs as well as for terrorism attacks using explosives or chemical weapons. Several surveillance and detection technologies are under investigation at the moment, with different trade-offs in complexity, range, and capabilities.

The project SafeShore, funded by the European Commission under the “Horizon 2020” program, grant agreement No 700643, is addressing this ambitious goal within a general framework of border protection [1, 2]. One of the initiatives of the SafeShore Consortium has been the organization of the *International Workshop on Small-Drone Surveillance, Detection and Counteraction Techniques* (WOSDETC) as part of the 14th edition of the *IEEE International Conference on Advanced Video and Signal based Surveillance* (AVSS). In conjunction with this event, the *drone-vs-bird detection challenge* has been launched to address one of the main issues arising in the described context. Indeed, given their characteristics, drones can be easily confused with birds, which makes the surveillance tasks even more challenging especially in maritime areas where bird populations may be massive. The use of video analytics can solve the issue, but effective algorithms are needed able to operate also under unfavorable conditions, namely weak contrast, long range, low visibility, etc.

The challenge was aimed at attracting research efforts to identify novel solutions to the problem outlined above, i.e., discrimination between birds and drones, by providing an annotated video dataset recorded at shore areas in different conditions. The challenge goal is to detect a drone appearing at some time in a short video sequence where birds are also present: the algorithm should raise an alarm and provide a position estimate only when a drone is present, while

<sup>1</sup>The project “SafeShore” leading to this work has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 700643.

not issuing alarms on birds. All the participants to the challenge were asked to submit score files with their results and a companion paper describing the applied methodology.

## 2. Dataset and evaluation metric

For the challenge the following dataset has been made available: a collection of 5 MPEG4-coded 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, i.e.

$[top\_x \ top\_y \ width \ height]$

only when the drone is present.

A few examples of frames extracted from the videos released to train the algorithms are shown in Fig. 1. It is apparent the difficulty of coping with very diverse background and illumination conditions, as well as with different scales (zoom), viewpoints, low contrast, and presence of birds.

A few days before the challenge deadline, a different video sequence has been provided for testing. Authors then submitted one file providing the frame number and estimated bounding box (always in the format  $[top\_x \ top\_y \ width \ height]$ ) only for the frames where the algorithm detects the presence of the drone. For frames not reported, no detection is assumed. Fig. 2 shows the central part of this sequence, with clouds differently illuminated that create a significant clutter, and the presence of a bird moving closer and closer to the drone — the rightmost figure is a zoom where the two moving objects are about to cross the same point on the projected plane of the image.

A penalty is computed frame-by-frame as the area (in pixels) of the smallest box that includes both the true and estimated bounding boxes, normalized by the area of the targets bounding box in order to be meaningfully averaged over all frames. Two examples are reported in Fig. 3.

For frames with no target a bounding box  $[0 \ 0 \ 1 \ 1]$  is used, i.e., located at the origin with 1 pixel area. A synthetic performance indicator is obtained as some average score of the penalties, with the best (smallest) possible score being equal to 1. To take into account the non-uniform distribution of the penalties, the root mean square value of the error is taken as final score (more details in Sec. IV).

## 3. Participation and best proposed algorithms

The challenge has attracted remarkable interest, with about 20 different research groups requesting access to the dataset for participation to the competition. The worldwide distribution of the research institutions that have been interested in the challenge is shown in Fig. 4. It is also worth noticing that none of the participants is a member of the SafeShore consortium, nor a research partner/collaborator.



Figure 1. Sample frames extracted from the videos released to train the algorithms.



Figure 2. Sample frames extracted from the video released to test the algorithms.

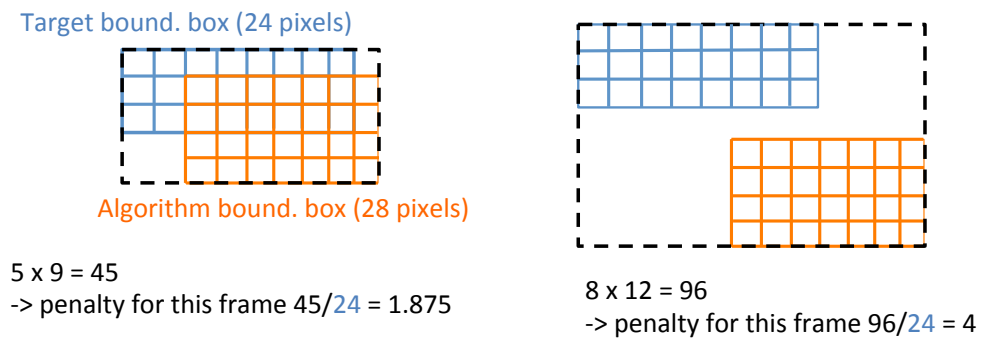


Figure 3. Example of calculation of the performance metric.

At a glance, the prominent ingredient of the solutions that have been proposed is the use of neural networks and deep learning approaches, coupled with additional processing blocks and ideas. As basic building block, convolutional neural networks (CNNs) have been used. These are a class of deep, feed-forward artificial neural network that use a variation of multilayer perceptrons to significantly reduce the pre-processing. The first layers that receive an input signal are convolution filters that basically try to label it by “mixing” (convolving) the input signal with the current filter information. The resulting signal is passed on to the next layer; each layer, in a sense, represents a feature of interest to be learned. Since convolution is translation-invariant, the output signal is not dependent on where the features are located, but simply whether the features are present, which is a powerful property for image recognition applications. Then, signals from the convolution layer are processed to reduce the impact of noise and variations (“subsampling”), e.g. by averaging, resizing, or contrast reduction. Neurons in the last layers are fully connected, to mimic high-level reasoning where all possible paths are considered. In the following the most successful algorithms developed for the drone-vs-bird challenge are briefly described.

Aker and Kalkan from KOVAN Research Lab., Computer Engineering, Middle East Technical University Ankara, Turkey [3], have used an end-to-end object detection method based on CNNs to predict the location of the drone in the video frames. In order to be able to train the network, the authors created an artificial dataset by combining real drone and bird images with different background videos. The results show that the variance and the scale of the dataset make it possible to perform well on drone detection problem.

Saqib et al. from University of Technology Sydney, Australia, and Makkah Technology Valley, Kingdom of Saudi Arabia [4], have considered Faster R-CNN [5] with Caffe deep learning library. The Caffe-based pre-trained models are publicly available for most of the object detectors. There are too few images in the dataset to learn a deep model from scratch. Therefore, to take full advantage of network architectures, the authors have used transfer learning from ImageNet to fine-tune the models. The fine-tuning process helps the system to converge faster and perform better. Various network architectures have been tested such as ZF [6], VGG16 and VGG M 1024 [7] to train the system (see details in the paper) and evaluate the performance on the test dataset. ZF is a 8 layered architecture containing 5 convolutional layers and 3 fully-connected layers. Similarly, VGG16 is a 16 layered architecture that has 13 convolutional layers and 3 fully connected layers.

Schumann et al. from Fraunhofer IOSB, Karlsruhe, Germany, and Vision and Fusion Lab, Karlsruhe Institute of Technology KIT, [8], have proposed a detection framework

composed of two core modules: the first module detects regions which are likely to contain a UAV followed by a classification module to distinguish each hypothesis into UAV or distractor classes, such as birds. To detect regions which are likely to contain an UAV, two complementary detection techniques are considered which exhibit promising results on video sequences containing UAVs at different distances. Depending on whether the video images are recorded by static cameras or moving cameras, median background subtraction or a deep learning based method are applied, respectively. To reduce the high number of false alarms, a CNN classifier is also used.

In general, to classify UAVs in real world data is a challenging task due to varying object dimensions (in the range of less than ten to hundreds of pixels), large variety of existing UAVs, and often lack of training data. Furthermore, the classification is impeded by varying illumination conditions, differing backgrounds, and localization errors of the detector. To address the various object dimensions, in [8] it is proposed to use a small network that is optimized to handle low resolution objects such as UAVs at large distances. A proprietary dataset is used to train the CNN classifier. The dataset is composed of crawled and self-acquired UAV images, bird images of a publicly available dataset and crawled background images to account for the large variety of existing UAVs, other distracting flying objects, and varying illumination conditions and backgrounds.

Finally, Faster RCNN with the VGG16 model has been also used in the approach proposed by Amandi and Farhadi from ArkaInvent Research, Tehran, Iran, [9]. Therein, moving object detection is combined with single deep neural network object detector; along with finding of moving objects, object detection step applies on each frame using three classes: drone, bird, other. If the detection accuracy is higher than a threshold and related to the previous step, the algorithm accepts it but if the detection is out of the predicted bound the result of the object detection is rejected. The algorithm finds moving objects combined with the history of previous detections and temporarily object detection results are accepted if there is no accurate detection.

## 4. Results

The computation of the per-frame penalty has been performed based on the metric described in Sec. 2. Multiple bounding boxes are counted as additional penalties for the same frame. A final score is calculated to obtain a ranking for the average behavior on the whole test video; in particular it is computed as square root of the mean squared penalty across frames.

Results are listed in Table 1, only for the algorithms for which result data have been provided by the deadline; up to three different settings for the same algorithm were allowed, so as to test different solutions under the same approach.



Figure 4. Map of the research groups participating to the challenge.

Interestingly, the results spontaneously grouped, irrespective of the specific setting; thus, for this specific outcome, it seems that the algorithm is more important than the fine-tuning of its parameters.

Table 1. Final score of the algorithms on the test video

algorithm [8] (setting 1)	1.0
algorithm [8] (setting 2)	1.2963
algorithm [8] (setting 3)	2.6347
algorithm [9] (unique setting)	3.1896
algorithm [3] (setting 3)	7.7556
algorithm [3] (setting 2)	13.3270
algorithm [3] (setting 1)	18.8449
algorithm [4] (setting 2)	110.7539
algorithm [4] (setting 3)	124.1251
algorithm [4] (setting 1)	176.8701
all other teams did not provide results	--

Generally speaking, all algorithms were able to detect the drone; they differed in the ability to cope with clutter, change of illumination conditions, and presence of birds. Moreover, some of the approaches have a better localization ability than others. By looking at the per-frame penalty over the whole video sequence, which is 660 frames long, different behaviors can be observed. For some of the algorithm, the penalty is more flat over the whole video; for others, the errors are very unevenly distributed. In particular, a few larger penalties can dominate the final performance due to wrong bounding box quite far apart the ground truth; moreover, sometimes larger penalties may arise due to sensitivity to clutter; finally, missing detections have an impact too.

Two representative examples of very different penalty distributions are shown in Fig. 5. This motivated the need of a synthetic final score able to take into account both the average performance but also the presence/absence of very different penalty values; the ultimate choice has been the simplest of such “higher-order” metrics, i.e., the root mean square, but it is an interesting direction for future research to design more sophisticated metrics for a sharper assessment.

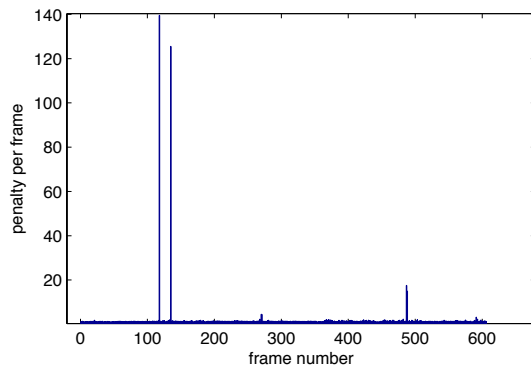
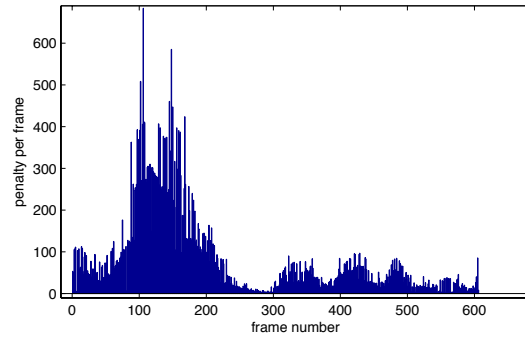


Figure 5. Example of per-frame penalty distribution across the whole test video.

At the end of the evaluation process, the most successful algorithm has been the one proposed in [8]: in its best setting, the value of the penalty reached its absolute minimum (1.0). This is a great achievement, although performance can be of course different on other test videos. It is part of the SafeShore Consortium’s future plans to elaborate a more advanced version of this challenge for the next year, based on the experience of this first, yet satisfying, edition. The winner has been awarded by a Nvidia TX2 platform.

## 5. Conclusions

The paper reported on the “drone-vs-bird detection challenge” launched by the SafeShore Consortium within the International Workshop on Small-Drone Surveillance, Detection and Counteraction Techniques (WOSDETC), co-located with the 14th IEEE International Conference on Advanced Video and Signal based Surveillance (AVSS) held in Lecce, Italy. The challenge has attracted remarkable interest, with about 20 different research groups participating from all over the world. The prominent ingredient of the solutions that have been proposed is the use of neural networks and deep learning approaches, coupled with additional processing blocks typical of moving object detection, but with innovative ideas to cope with the peculiarities of the challenge. A more advanced edition is planned for the next year, based on the lessons learned from this edition.

## References

- [1] Horizon2020. The SafeShore project. In <http://safeshore.eu>, funded by the European Commission under the “Horizon 2020” program, grant agreement No 700643.
- [2] 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.
- [3] Cemal Aker and Sinan Kalkan. Using deep networks for drone detection. In *IEEE International Workshop on Small-Drone Surveillance, Detection and Counteraction Techniques*, Lecce, Italy, Aug. 2017.
- [4] Muhammad Saqib, Nabin Sharma, Sultan Daud Khan Makkah, and Michael Blumenstein. A study on detecting drones using deep convolutional neural networks. In *IEEE International Workshop on Small-Drone Surveillance, Detection and Counteraction Techniques*, Lecce, Italy, Aug. 2017.
- [5] 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.
- [6] M. D. Zeiler and R. Fergus. Visualizing and understanding convolutional networks. In *European conference on computer vision*, 2014.
- [7] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. Technical report, arXiv preprint arXiv:1409.1556, 2014.
- [8] Arne Schumann, Lars Sommer, Johannes Klatte, Tobias Schuchert, and Jurgen Beyerer. Deep cross-domain flying object classification for robust UAV detection. In *IEEE International Workshop on Small-Drone Surveillance, Detection and Counteraction Techniques*, Lecce, Italy, Aug. 2017.
- [9] Mohammad Farhadi and Ruhallah Amandi. Drone detection using combined motion and shape features. In *IEEE International Workshop on Small-Drone Surveillance, Detection and Counteraction Techniques*, Lecce, Italy, Aug. 2017.