Received XX Month, XXXX; revised XX Month, XXXX; accepted XX Month, XXXX; Date of publication XX Month, XXXX; date of current version XX Month, XXXX.

Digital Object Identifier 10.1109/XXXX.2022.1234567

The Drone-vs-Bird Detection Grand Challenge at ICASSP 2023: A Review of Methods and Results

Angelo Coluccia[1], Senior Member, IEEE, Alessio Fascista[1], Member, IEEE, Lars Sommer[2], Arne Schumann[2], Anastasios Dimou[3], and Dimitrios Zarpalas[3]

[1]Department of Innovation Engineering, University of Salento, 73100 Lecce, Italy
[2]Fraunhofer Center for Machine Learning, Fraunhofer IOSB, 76131 Karlsruhe, Germany
[3]Centre for Research and Technology Hellas, The Visual Computing Lab, Information Technologies Institute, 57001 Thessaloniki, Greece

Corresponding author: Angelo Coluccia (email: angelo.coluccia@unisalento.it).

ABSTRACT This paper presents the 6th edition of the Drone-vs-Bird detection challenge, jointly organized with the WOSDETC workshop within the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) 2023. The main objective of the challenge is to advance the current state-of-the-art in detecting the presence of one or more Unmanned Aerial Vehicles (UAVs) in real video scenes, while facing challenging conditions such as moving cameras, disturbing environmental factors, and the presence of birds flying in the foreground. For this purpose, a video dataset was provided for training the proposed solutions, and a separate test dataset was released a few days before the challenge deadline to assess their performance. The dataset has continually expanded over consecutive installments of the Drone-vs-Bird challenge and remains openly available to the research community, for non-commercial purposes. The challenge attracted novel signal processing solutions, mainly based on deep learning algorithms. The paper illustrates the results achieved by the teams that successfully participated in the 2023 challenge, offering a concise overview of the state-of-the-art in the field of drone detection using video signal processing. Additionally, the paper provides valuable insights into potential directions for future research, building upon the main pros and limitations of the solutions presented by the participating teams.

INDEX TERMS deep learning, drone detection, image and video signal processing, unmanned aerial vehicles (UAV)

I. INTRODUCTION

UNMANNED Aerial Vehicles (UAVs), commonly known as drones, have gained immense popularity and found diverse applications in recent years, encompassing areas such as monitoring, environmental protection, support to communication systems [1]–[3]. While their versatility and capabilities offer numerous benefits, there is a growing need for effective *UAV detection systems* due to the concerns raised on various aspects, including security, safety, and privacy [4]. The 2023 annual report of the Federal Aviation Administration (FAA) of US reported multiple incidents over the recent years [5], mainly caused by malicious or suspicious usage, or inadvertent misuse of UAVs, proving the severity and timely importance of the problem.

Detecting and identifying *unauthorized drones* can help preventing potential threats, safeguarding critical infrastructures, and protecting individual privacy. The increasing utilization of UAVs highlights the need not only for implementing and regulating rules regarding drone flights, but also for establishing effective UAV detection systems to accurately localize intruders or unauthorized drones. For instance, certain sensitive areas, such as airports, military bases, government facilities, or nuclear power plants, require heightened security measures. Similarly, critical infrastructure such as power plants, oil refineries, and telecommunications networks are crucial to the functioning of societies [6].

UAV detection systems play a crucial role also in ensuring *air traffic safety* as the risk of collisions with manned aircrafts also rises with the ever increasing number of UAVs in the airspace [7]. By identifying and tracking UAVs, authorities can implement appropriate measures to prevent collisions, maintain the integrity of flight paths, and reduce the potential for accidents or disruptions.

This article has been accepted for publication in IEEE Open Journal of Signal Processing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/OJSP.2024.3379073

UAVs have been misused also for various illegal activities, including smuggling drugs, contraband, or weapons across borders [8]. In this respect, UAV detection systems can aid law enforcement agencies in identifying and apprehending individuals involved in such activities. Rapid detection and response enhance the effectiveness of border control operations, mitigating the risks associated with illegal trafficking.

The increasing need for high performing counter-drone systems stresses the need for innovative approaches providing reliable automatic detection and identification of drones, in a variety of environments and scenarios. In terms of operational capacity, an effective detection system must be able to detect threats from *diverse drone types*, including custom-made and entirely new designs. Another challenge faced by current systems is the adaptation to a new or evolving environment (i.e. weather, sunlight, vegetation, surrounding infrastructures, presence of birds) which can make systems ineffective. Modern Counter-UAV systems build upon a number of detection technologies (e.g. visual, RF, radar, acoustic) [9], [10], to overcome specific challenges and limitations of each individual modality.

While there is a lot of ongoing research on the subject from many different communities, e.g. [11]–[16], the availability of datasets that can be exploited for training UAV detection systems is rather limited. Some organizations and companies may have proprietary ones. These datasets may include real-world scenarios, proprietary detection algorithms, and sensor data. However, access to such datasets is typically restricted and may require partnerships or agreements with the data owners.

In-house data collection provides the advantage of tailoring the dataset to specific needs and capturing data in specific operational contexts, but it is an expensive and time-consuming task in terms of equipment, regulations, privacy preservation, and most importantly annotation. As a matter of fact, capturing data utilizing diverse UAV types, under different environmental conditions, including birds that may interfere with the detection capabilities, is not possible for most research teams.

Collaboration and data-sharing initiatives among industry stakeholders, government agencies, and research communities can help increase data availability for UAV detection system training. In this paper, we present the *Drone-vs-Bird Detection Grand Challenge*, an initiative that combined data capturing campaigns from European projects, in order to offer to the research community a comprehensive dataset on visual capturings of UAVs, manually annotated, aiming to promote research on the domain. The Challenge focuses on providing the means to advance the drone detection state of the art (SoA), by seeking for innovative signal processing solutions for video data sequences. Since the launching of Drone-vs-Bird Detection Grand Challenge, numerous academic groups or companies have received the dataset, in order to train or evaluate their own drone detection methods.

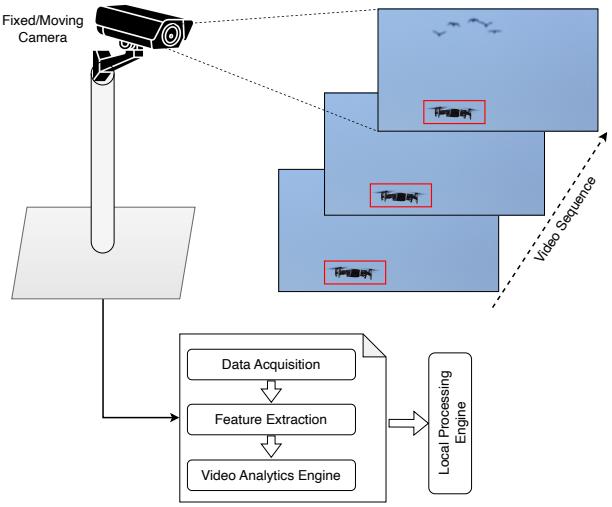

FIGURE 1. General operational scenario of the Drone-vs-Bird Detection Challenge.

The first edition of the *International Workshop on Small-Drone Surveillance, Detection and Counteraction Techniques* (WOSDETC) [17] was organized in 2017 as part of *IEEE International Conference on Advanced Video and Signal based Surveillance* (AVSS), held in Lecce, Italy. In conjunction with the workshop, the grand challenge called *Drone-vs-Bird Detection Challenge* was launched. In 2019, a second edition of the challenge was organized, again as part of WOSDETC and co-located with the 16th edition of AVSS held in Taipei, Taiwan [18]. A third edition of the Drone-vs-Bird challenge was organized in 2020, initially planned as part of the 17th edition of AVSS in Washington DC, USA, but then run as virtual event due to the COVID-19 pandemic [19]. The fourth edition of the challenge was organized in conjunction with the 17th AVSS in 2021 as a virtual event [20]. The fifth edition of the challenge was held in conjunction with ICIAP 2021 (May 2022) [21] in Lecce. The present work extends the short (two-page) paper in [22] and provides a more deep overview of the methodologies and outcomes of the 6th edition of the Grand Challenge, which was held as part of ICASSP 2023 on Rhodes Island, Greece.

A very high number of access requests to the dataset have been filed since the challenge inception, by numerous teams from various countries. This reflects the global reach and popularity of the challenge, showcasing the widespread interest and involvement of research communities from around the world. However, it is worth noting that only a handful of teams manage to submit valid results, confirming the general difficulty of the detection task and the need of further research and advancements in the topic. To offer a quantitative summary, in the 2023 edition, we received approximately 100 requests for the dataset, had about 20 registrations for the grand challenge, and received 8 successful submissions by the deadline from 4 distinct teams.

FIGURE 2. Examples of drone types present in the training set, i.e., Parrot Disco, 2 custom fixed-wing drones, DJI Inspire, DJI Phantom, DJI Mavic, DJI Matrice and 3DR Solo Robotics.

II. The Drone-vs-Bird Detection Challenge Dataset

In this section, we give a comprehensive description of the dataset utilized for the Drone vs. Bird Detection Grand Challenge at ICASSP 2023. We begin by outlining its primary attributes, including the diverse range of video sequences encompassed, the various models of drones and disturbing objects encountered, as well as the variability observed within the considered scenes. Furthermore, we provide a concise summary of other drone datasets freely accessible in existing literature to offer a holistic perspective. Lastly, we present a brief overview of the participation history, spanning from the challenge's first edition to the latest 2023 edition.

A. The Challenge Training Dataset

The Drone-vs-Bird Detection Challenge dataset encompasses a diverse collection of 77 video sequences, serving as training data for all participating teams. This dataset has undergone progressive evolution over the editions of the challenge. Initially, a portion of the videos was obtained through experimental campaigns conducted within the SafeShore project[1], utilizing MPEG4-coded static cameras. These recordings were subsequently augmented by additional sequences contributed by the Fraunhofer IOSB research institute, sourced from various locations across Germany. In 2020, the ALADDIN project[2] introduced 45 more videos, incorporating the use of moving cameras for acquisition. Overall, the training dataset to date comprises a combination of sequences captured with both static and moving cameras, featuring diverse resolutions ranging from 720×576 to 3840×2160 pixels. Note that static cameras allow a straightforward application of motion detection methods as initial detector or in addition to appearance based detection methods, while sequences recorded by moving cameras require a camera motion compensation.

Each sequence contains an average of approximately 1,384 frames, with an average of 1.12 annotated drones per frame. As illustrated in Figure 2, the dataset encompasses eight distinct types of commercial drones, including Parrot Disco, DJI Inspire, DJI Phantom, DJI Mavic, DJI Matrice, 3DR Solo Robotics, and two custom fixed-wing drones. Among these, three types possess fixed wings, while the remaining five exhibit rotary wings.

The training dataset is comprised of sequences provided by different research institutes, which recorded their data at different locations under varying conditions, thus offering a large variety of scenes and backgrounds. It features the presence of both static and moving camera sequences, of different lengths, with frame characteristics changing also within a same sequence (e.g., the camera may first point to the sky but then follow the drone on the land, with trees background or maritime scene or others). More specifically, scenes include urban areas, woodlands, agricultural areas, urban areas and rivers in Central Europe, maritime areas as well as Mediterranean landscapes and cities, resulting in varying levels of difficulty for the detection algorithms. A diverse range of backgrounds is observed, including sky, buildings, water surfaces, and different kinds of vegetation, i.e., trees, grassland, bushes, and rocks. The dataset further incorporates different weather conditions such as cloudy and sunny, and different recording times such as daytime, dawn and nighttime. Moreover, it encompasses challenges such as direct sun glare and variations in camera characteristics, as depicted in Figure 3. While drones are annotated in the dataset, birds, often appearing as main disturbing objects, specifically in more than one-third of the sequences, are not annotated (further discussion on this point will be provided in Sec. VI).

The distance between the drones and the camera exhibits significant variability across and within the videos, leading to considerable variations in drone sizes, as showcased in Figure 4. The drone sizes range from as small as 15 pixels to over 1,000,000 pixels. The majority of annotated drones have sizes less than 16^2 pixels or fall within the range of 16^2 to 32^2 pixels. The presence of small-sized drones poses a particularly challenging detection task. To facilitate the training process, each video sequence is accompanied by a separate annotation file, available on GitHub (at https://github.com/wosdetc/challenge). This file contains information on the frames in which drones enter the scenes, along with their precise locations expressed as bounding boxes in the form of $[\mathrm{top}_x\ \mathrm{top}_y\ w\ h]$. In this notation, $(\mathrm{top}_x, \mathrm{top}_y)$ represents the coordinates of the top right corner, while w and h indicate the width and height of the bounding box, respectively. While drones are annotated in the dataset, birds, often appearing as main disturbing objects, i.e. in more than one third of the sequences, are not annotated.

B. The Challenge Test Dataset

The challenge test set encompasses an additional 30 video sequences, for which no annotations are provided. Among these, 16 video sequences are inherited from initial editions of the challenge. Most of the locations depicted in these

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

[2]The project "ALADDIN" has been granted funding from the European Union's Horizon 2020 research and innovation programme, with grant agreement No. 740859.

This article has been accepted for publication in IEEE Open Journal of Signal Processing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/OJSP.2024.3379073

FIGURE 3. Sample frames extracted from the training videos showing the large variability of the dataset.

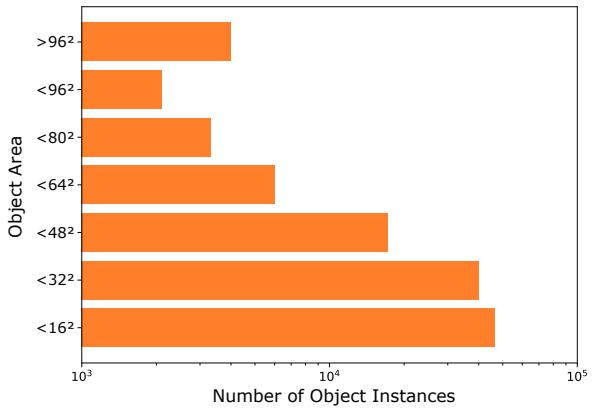

FIGURE 4. Distribution of drone sizes across the ground truth annotations in the training dataset.

sequences are also present in the training set and exhibit similar characteristics. To increase the difficulty, the test set has been enriched with new video sequences that introduce novel backgrounds and two distinct types of rotary drones. Furthermore, the test set features the presence of additional disturbing objects, such as planes, and includes scenarios where drones are located against structured backgrounds, as shown in Figure 5. To ensure a fair evaluation, sequences exceeding 30 seconds in duration have been shortened to prevent a few individual videos from dominating the whole evaluation process.

As concerns the level of overlap between the training and test sets in the Drone-vs-Bird dataset, it is worth highlighting that most sequences in the test set are recorded at completely unseen locations, whereas the remaining sequences that share similar (but not identical) backgrounds with the training set have been acquired using different perspectives and recording times. The latter is indeed a common strategy

FIGURE 5. Sample frames extracted from the test videos showcase notable differences compared to the training set.

to minimize the overlap in the datasets while balancing the efforts necessary to conduct the acquisition campaign (in some practical cases, in fact, changing to completely different scenes may be not even possible due to constraints on the movement of hardware equipment). Thus, the Drone-vs-Bird dataset allows to assess the effectiveness of the proposed detection methods (including, for instance, aspects such as the generalization ability) with nearly-zero overlap between training and test sets.

The dataset, including both the training set and the test set, is openly available for download. However, to access the dataset, interested individuals must first sign a Data Usage Agreement (DUA) to comply with the terms and conditions regarding the use and handling of the data. For convenience, the annotations for the dataset can be accessed at the following URL: https://github.com/wosdetc/challenge.

C. Other Drone Datasets

For the sake of completeness, we now review other publicly available drone detection datasets, in comparison with the Drone-vs-Bird dataset. It is important to note that datasets based on other sensor modalities are not considered in this overview, although some of the datasets may also include EO (Electro-Optical) or IR (Infrared) imagery in addition to (visible-light) video sequences.

The first dataset we discuss is the *Drone Dataset: Amateur Unmanned Air Vehicle Detection*, released in 2019 [23]. This dataset includes over 4000 images featuring DJI Phantom drones. Images have a resolution between 300×168 pixels and 4k, and the dataset also comprises images with non-drone objects.

The *Small Target Detection database (USC-GRAD-STDdb)* [24] was built using 115 video segments downloaded from YouTube. The frames have a resolution of 1280×720 pixels, with specific annotations available for about 25,000 frames. They include more than 56,000 small objects, categorized as drones, birds, boats, vehicles, and people. Out of the 115 video segments, 57 contain either drones or birds, while the Drone-vs-Bird dataset specifically considers the simultaneous presence of both drones and birds in the scene.

TABLE 1. Attributes covered by the overviewed UAV datasets.

	Several Types of UAV Models	Variety of Back-grounds	Different Weather	Possible Moving Camera	Different Pixel Sizes	Presence of Birds or Other objects
Amateur UAV Detection Dataset [23]	✗	✓	✗	✗	✓	✗
USC-GRAD-STDdb [24]	✓	✓	✗	✓	✓	✓
Purdue UAV Dataset [25]	✗	✗	✗	✓	✓	✗
Flying Object Detection [26]	✓	✗	✗	✓	✓	✓
Real-World Object Detection Dataset [27]	✓	✓	✓	✓	✓	✗
Anti-UAV Challenge Dataset [28]	✓	✓	✓	✓	✓	✗
Multi-view Drone Tracking Datasets [29]	✓	✗	✓	✗	✗	✗
VisioDECT [30]	✓	✓	✓	✗	✓	✗
DUT Anti-UAV [31]	✓	✓	✓	✓	✓	✗
Halmstad Drone Dataset [32]	✗	✓	✗	✓	✓	✓
USC Drone Dataset [33]	✗	✓	✓	✗	✗	✗
Drone-vs-Bird	✓	✓	✓	✓	✓	✓

The *Purdue UAV dataset* [25] is a smaller dataset comprising only five video sequences, for a total of 1829 frames. These video sequences were recorded using a custom airframe with a camera and have a frame rate of 30 frames per second. Images have a resolution that is either 1920×1080 or 1280×960 pixels. Moreover, the annotations for the ground truth are openly available for download.

Another dataset worth mentioning is the *Flying Object Detection from a Single Moving Camera* dataset [26]. The dataset consists of 20 video sequences, with each image having a resolution of 752×480 pixels and containing, on average, two similar objects that challenge the detection task. The video sequences were acquired with a commercial UAV mounting a standard camera, resulting in varying drone appearances caused by changing orientations, lighting conditions, and other factors. Furthermore, this dataset includes 20 video sequences featuring aircraft sourced from YouTube, exhibiting image resolutions ranging from 640×480 to 1280×720 pixels.

A more recent dataset is the *Real World Object Detection Dataset for Quadcopter Unmanned Aerial Vehicle Detection* [27]. This dataset encompasses an extensive collection of 51446 training images and an additional 5375 images specifically allocated for testing purposes. The images themselves were procured through a combination of internet downloads and author-captured content, all adjusted to adhere to a uniform resolution of 640×480 pixels. Within the training set, about 52,676 different instances of drones can be found. Conversely, the test set is composed of about 2863 drone instances, alongside 2750 images void of any drone presence. To expedite the annotation procedure, an innovative semi-automated labeling pipeline was effectively implemented. Notably, within the training set, approximately 40.8% of the drones are confined to dimensions smaller than 32×32 pixels, while about 23.4% exceed the threshold of 96×96 pixels. In the test set, similar proportions reveal that about 36.3% of the drones are of smaller dimensions than 32×32 pixels, while a noteworthy 28.3% surpass the dimension of 96×96 pixels.

The *Anti-UAV Challenge* dataset [28] was released in 2020 as part of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Unlike the Drone vs. Bird Detection Challenge, the Anti-UAV Challenge focuses on the task of tracking a single object. The dataset consists of a total of 160 video sequences, including IR and EO imagery. The IR images have a resolution of 640×512 pixels, while the EO images have a resolution of 1920×1080 pixels. About 100 video sequences have been annotated, serving as training data for tracking algorithms. The videos were recorded through a rotating platform equipped with a static camera. Consequently, the acquisition campaign is limited only to a few selected scenarios. Additionally, this dataset focuses on four specific drone types: DJI Inspire, DJI Phantom, DJI MavicAir, and DJI MavicPRO.

On the other hand, the *Multi-view drone tracking datasets* [29] were proposed to deal with the problem of reconstructing 3D flight trajectories, using as acquisition system an ad-hoc network of cameras. These datasets consist of five separate datasets. The first four datasets accounts for the presence of an hexacopter captured with different cameras, while the fifth dataset involves three different types of drones. The datasets consider different acquisition setups, with a number of cameras changing from 4 to 7. Moreover, the flight duration varies between 2 and 10 minutes. In terms of annotations, they are provided in the form of a single point for the first four datasets. Compared to the Drone vs.

Bird Detection Challenge dataset, these Multi-view drone tracking datasets are smaller in size and lack the inclusion of diverse environmental settings. Not least, they are tailored for different goals, such as tracking of single objects or reconstruction of 3D trajectories.

The *VISIODECT dataset* [30], released in 2022, comprises 20,924 sample images and associated annotations, encompassing six drone models operating in three distinct scenarios (cloudy, sunny, and evening), at various altitudes and distances ranging from 30 m to 100 m. The data is available in three different file formats (txt, xml, csv) and was generated over 1 year and 8 months at 12 different locations. Each video sequence was converted into JPEG image frames with dimensions of 852×480 pixels. These frames were organized and stored in repositories, each representing a specific model class and scenario sub-class. To enhance data quality, a teams of professionals cleaned each repository by manually selecting image frames that did not feature drones in the background. Data annotation was conducted by manually delineating bounding boxes around each image file, resulting in the creation of corresponding label files. To maintain a consistent naming convention and minimize errors, label files for each scenario sub-class were named to align with their respective image files and stored in repositories accordingly. Differently from the Drone-vs-Bird dataset, the VISIODECT does not include sequences acquired from moving cameras and does not account for the presence of birds or similar objects.

Another very recent dataset is the *Anti-UAV Detection and Tracking* from Dalian University of Technology (DUT) [31]. The whole dataset divides in two separated subsets: one for detection and the other for tracking. The detection dataset, which has a similar scope as the Drone-vs-Bird, accounts for 10,000 images in total, in which the training, testing, and validation sets have 5200, 2200 and 2600 images, respectively. All frames and images have been manually annotated. Image resolution spans from 160×240 to 3744×5616, offering a large variability in the UAV sizes across different sequences. There are more than 35 different UAV models appearing in the detection dataset, flying in outdoor environments including sky, dark clouds, jungles, high-rise buildings, residential buildings, farmland, and playgrounds. Compared to our Drone-vs-Bird dataset, the DUT dataset mainly lacks the presence of birds or other disturbing flying objects.

The *Halmstad dataset* represents another valuable source of video sequences meant for UAV detection [32]. Data have been captured at three airports in Sweden (Halmstad, Gothenburg, and Malm) and comprise 650 video sequences, including also some non-copyrighted material from the YouTube channel "Virtual Airfield operated by SK678387" used to enrich the target categories (mainly airplanes and helicopters). The dataset features only 3 different types of UAVs and all the videos have a resolution of 640×512 pixels, and a total duration of 10 seconds each. The maximum distance at which UAVs are captured from the fixed cameras is about 200 m. Given the limited scenario considered for the construction of the dataset (airports only), it does not provide high variability in terms of scenes and weather conditions.

The *USC Drone Dataset* represents another freely-avilable dataset specifically constructed for video-based object detection and tracking [33]. It contains only 30 sequences, all recorded at the USC campus. The sequences include the presence of a single drone model but span a variety of different backgrounds, different angles of acquisition and variable weather conditions. The dataset has the objective of capturing real UAV attributes such as fast maneuvering, occlusions, and high illumination, just to mention a few. All video sequences have a fixed resolution of 1920×1080, with each individual video lasting approximately one minute. To partially compensate for the limited variability in terms of scenes and drone models, the dataset also uses model-based data augmentation techniques that synthesize training images and annotate location of each drone within frames automatically.

In Table 1 we compare the major attributes of our dataset against different datasets identified in the literature. From experience in the field, a number of challenges can be identified, which need to be overcome for detecting drones "in the wild". The latter, in fact, are found i) of differing types, shapes, sizes and models (varying from tiny to large ones), ii) within a variety of backgrounds that may cause false positives, iii) within varying environmental, weather and time (day, dawn, sunshine, cloudy, dark) conditions, iv) using stable footage or not, v) in videos of differing resolution and drone sizes in pixels, vi) with the simultaneous presence of other flying objects (e.g., birds) that might cause false positives. Table 1 shows that indeed our proposed dataset is the one that meets all the expected criteria.

III. Description of Tasks and Evaluation Metrics

A. Detection Task

The detection task of the Drone-vs-Bird Detection Challenge 2023 requires that participating teams submit a set of result files. These files should encompass each video sequence, with explicit indications of the frame numbers in which drones were detected. Alongside the frame numbers, the predicted position of the drones within the frame must be provided in the same format of annotations, namely bounding boxes denoted by $[\mathrm{top}_x\ \mathrm{top}_y\ w\ h]$. Additionally, result files should include confidence scores for each frame, aiding in the assessment of the algorithm's uncertainty on its predictions. In cases where a frame does not contain any reported detections, it will be assumed that no drones were detected in that particular frame.

While the use of additional training data is permitted, teams must provide detailed descriptions regarding the quantity and nature of the supplementary data employed. It is essential for teams relying on additional data to submit an additional result of their method, indicating the performance

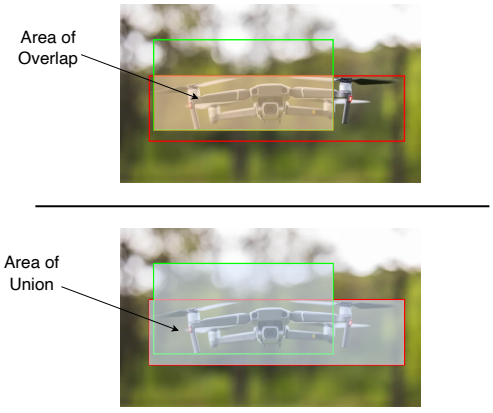

FIGURE 6. Intersection over Union (IoU) metric that captures the goodness in predicting a drone position in a frame.

achieved solely using the provided training data. However, the ultimate evaluation and ranking for the challenge will be based on the overall best score achieved, irrespective of the utilization of additional data. The ultimate goal of the algorithms should be to achieve precise and accurate localization of drones, ensuring that the estimated bounding boxes closely correspond to the actual UAVs positions.

B. Evaluation Metrics

The evaluation process for the Drone-vs-Bird Detection Challenge employs the widely-adopted Average Precision (AP) metric, which is commonly utilized in object detection tasks such as the COCO object detection challenge. The AP metric is based on the Intersection over Union (IoU) criterion, which measures the overlap between the estimated bounding box and the ground truth bounding box surrounding the UAV in the scene. The IoU is calculated as the ratio of the area of overlap between the two boxes to the total area of their union, as shown Figure 6.

To determine the accuracy of detections, a threshold (typically 0.5) is applied to the IoU. If the IoU between a detected UAV and a ground truth annotation exceeds the threshold, it is considered a true positive detection. Conversely, detections with an IoU below the threshold are counted as false positives. Any ground truth annotations that are not assigned to a detection are regarded as false negatives, representing missed detections. By calculating the area under the precision-recall curve, the AP metric provides a comprehensive evaluation of a detector's performance, capturing the trade-off between precision and recall. This single metric thus effectively summarizes the overall precision-recall characteristics of each proposed algorithm. It is important to note that the test sequences are made available to participants one week prior to the submission deadline; teams are requested to run their algorithms on these test data and submit the results. Eventually, teams that realize the performance of their algorithms are qualitatively inadequate, typically withdraw from the challenge and refrain from submitting any results.

IV. Drone Detection Algorithms

In the following the drone detection algorithms used for submissions are briefly discussed.

OBSS AI (OBSS Teknoloji Ankara, Turkey) proposed a drone detection framework that comprises an initial deep learning based drone detector, a sequence classifier and template matching. Based on their previous approach [21], YOLOv5m6 [34] was employed as drone detection model. The model was trained on four different drone detection datasets to increase the models generalization ability. In addition, a new synthetic drone detection dataset, which consists of random background images and randomly placed drone objects, was employed to improve the detection performance in case of complex and unseen backgrounds. The model was trained for 10 epochs using the default YOLOv5 training configuration, while the image scale was set to 1344 pixels. While an image based drone classification model was utilized in their previous work to improve the detection accuracy, OBSS AI modified this classifier to classify image sequences [35]. For this purpose, an object tracker generated tracks for detected objects. Then, eight instances of a track were fed to a sequence classifier model, computing a drone probability. This drone probability was combined with the object detectors' drone probability using geometric mean. To train the classifier, a dataset was semi-automatically created by using their detector and tracker to generate object tracks from training video sequences. These tracks were exported and manually labeled as drone, bird, and other. To overcome missing detections in case of complex backgrounds, OBSS utilized a template matching approach. Therefore, historical data were stored for each tracked object. If the object detector failed, a template matching algorithm was applied near the last object location in a small search region, i.e. image width / 10 $\times$ image height / 10. Predicted bounding boxes were then fed to the sequence classifier model to calculate the drone probability of the object.

IIT (Indian Institute of Technology Jammu) proposed a detection scheme comprised of three stages. Initially, YOLOv7 [36] is applied as drone detection model, which was trained on 60 videos from the Drone-vs-Bird challenge train set. To reduce the number of false positive detections, detections were filtered based on the confidence score in the subsequent stage. For this purpose, IIT estimated the number of drones n for each sequence and only considered the corresponding n bounding boxes with the highest confidence scores. The number of drones per sequence was derived from the number of detections for each image throughout the entire sequence. In the last stage, a CSRT tracker [37] was employed to account for missed detections in complex environments. As the tracker is less reliable than YOLOv7 in detecting accurate bounding boxes, IIT proposed a scheme to fuse bounding boxes estimated by both YOLOv7 and the tracker. If YOLOv7 did not detect a drone, detections from the previous frame were used to initialize trackers. Then bounding boxes predicted by the tracker were used until detections become

available again. To identify false positive detections, the IoU between detections and the trackers bounding boxes was used. If the IoU was less than 0.3, detections were considered as false positive detections. Additional details can be found in the ICASSP paper [38].

DU (Dongguk University, South Korea) adapted the medium-size YOLOv8 [39] model for drone detection. To account for various drone sizes, a multi-scale image fusion (MSIF) [40] approach was employed. MSIF extracts features for three different scales of the input image, which are fused into one feature map through bottom-up and top-down structures. The combined feature map was then used as input of the YOLOv8 model. To improve the detection accuracy in case of small drones, the P2 layer of the backbone was added to the feature pyramid of YOLOv8 due to strong spatial local features. DU further applied data augmentation to increase the number of drone appearances in the training images. For this, a copy and paste scheme [41] was employed. Cropped and scaled drones were randomly located under the condition that the new location did not overlap with an already drone-occupied area. For training, every fifth frame of the Drone-vs-Bird challenge train set was used and the image scale was set to 640 pixels. The YOLOv8 model was trained for 93 epochs with a batch size of 16. For inference, the image scale was set to 1280 pixels. Furthermore, DU applied horizontal flipping and multi-scale augmentation. For more details, we refer the reader to the ICASSP paper [42].

Note that all approaches are based on detection methods applied on single images. Hence, no approach considers motion based detection methods to identify possible drone locations. However, OBSS employs temporal information by using an additional tracker, which can be useful in case of distant drones or complex backgrounds.

SNU (Shandong Normal University, China) adapted Single Shot Multi-Box Detector (SSD) [43] as detection model. To account for small drones, SNU added a shallow feature pyramid network and attention module. For training, SNU used images from the Drone-vs-Bird challenge train set and set the image scale to 300 pixels. The interested reader is referred to the ICASSP paper [44] for more details.

Team	Average Precision
OBSS AI Submission 2	**0.852**
OBSS AI Submission 1	0.841
OBSS AI Submission 3	0.811
IIT Submission 1	0.450
IIT Submission 2	0.367
IIT Submission 3	0.357
Dongguk University	0.189
Shandong Normal University	0.121

TABLE 2. Results of Drone-vs-Bird Detection Challenge 2023.

V. Performance Assessment

A. Analysis of the results

The final ranking of the Drone-vs-Bird challenge 2023 is reported in Table 2, showing the AP for each submission. The winning entry was submitted by OBSS AI, which substantially outperformed the other participating teams.

The AP values for each sequence are given in Table 3. For this, we only considered the best submission from each team. OBSS achieved the best AP on all sequences, exhibiting good detection results on most sequences. However, poor AP values are obtained in case of scenes with weak contrast between drone and structured background, e.g. VID_20210606_143947_04 and VID_20210606_141511_01. The detection results for IIT clearly differ for the different sequences. While high detection accuracies are achieved on several sequences, all drones are missed in other sequences. DU and SNU exhibit poor AP values on most sequences, while good detection results are only achieved for scenes with large UAVs and simple (non-structured) background.

The number of submitted detections and overall recall are given in Table 4. The test sequences comprise about 18000 annotated drones. While the number of detections submitted by OBSS and DU exceed the number of annotations, IIT and SNU submitted clearly less detections. Thus, OBSS exhibits a high recall rate, whereas IIT and SNU show poor recall rates. Though the high number of submitted detections, DU achieved a poor recall rate, which indicates that most detections are false positive detections.

The recall rate for each sequence are listed in Table 5. OBSS exhibits good recall rates except for some sequences, which comprise scenes with weak contrast between drone and background. The recall rates for IIT clearly differ for the different sequences. For some sequences, all drones are correctly detected, while all drones are missed in other sequences. DU and in particular SNU show numerous scenes without any or only few detections.

Notice that static cameras allow a straightforward application of motion detection methods as initial detector or in addition to appearance-based detection methods, while sequences recorded by moving cameras require a camera motion compensation. Moreover, all approaches are based on detection methods applied on single images; hence, none of them considers motion-based detection methods to identify possible drone locations. However, OBSS partially exploits temporal information by using an additional tracker, which can be useful in case of distant drones or complex backgrounds.

To further analyze the detection results, we computed the AP values and recall rates for different drone sizes (see Table 6 and Table 7, respectively). OBSS achieved the best AP values and highest recall rates for all drone sizes. Though the recall rate and AP increases with larger drones, OBSS shows a high recall rate and good AP even for small drones whose size is less than 16^2 pixels. The APs and recall rates obtained

Sequence	AP			
	OBSS	IIT	DU	SNU
GOPR5867_001	**0.997**	0.629	0.521	0.821
GH010037_solo_split02	**1.000**	0.546	0.361	0.078
GH010039_matrice_split02	**0.964**	0.000	0.000	0.000
GH010040_inspire_split03	**0.581**	0.000	0.000	0.000
GH010045_phantom_split01	**0.942**	0.561	0.623	0.000
VID_20220306_170118_01	**0.734**	0.643	0.077	0.000
VID_20220306_170541_01	**0.994**	0.932	0.057	0.000
VID_20220311_122209_01	**1.000**	**1.000**	0.000	0.000
VID_20210417_143217_01	**0.939**	0.018	0.000	0.004
VID_20210606_141511_01	**0.477**	0.032	0.041	0.000
GOPR5852_001	**1.000**	0.060	0.280	0.000
GOPR5861_001	**0.998**	0.613	0.606	0.436
VID_20211012_081448_01	**0.945**	0.759	0.000	0.000
2019_10_16_C0003_52_30_mavic	**0.781**	0.206	0.241	0.315
dji_mavick_mountain_cross	**0.891**	0.496	0.000	0.000
dji_phantom_mountain	**0.651**	0.327	0.000	0.000
GOPR5843_004	**0.923**	0.675	0.867	0.000
GOPR5847_001	**0.681**	0.595	0.564	0.000
GOPR5853_002	**0.659**	0.145	0.266	0.000
GOPR5856_001	**0.995**	0.465	0.539	0.169
GOPR5862_001	**0.996**	0.455	0.620	0.744
GOPR5868_001	**0.995**	0.985	0.483	0.781
VID_20210606_141851_01	**0.641**	0.001	0.000	0.000
VID_20210606_143947_04	**0.236**	0.000	0.000	0.000
VID_20211010_143610_01	**0.997**	0.994	0.733	0.117
VID_20211012_175158_02	**0.938**	0.638	0.770	0.000
4k_2020-06-22_C0006_split_01_01	**0.698**	0.000	0.001	0.000
4k_2020-07-29_C0020_01	**0.981**	0.490	0.914	0.000
4k_2020-07-29_C0021_01	**1.000**	**1.000**	0.803	0.000
VID_20210417_143930_02	**0.861**	0.390	0.011	0.105

TABLE 3. Detailed comparison for each team in the Drone-vs-Bird Detection Challenge 2023. The AP is given for every sequence of the test set.

Team	#Detections	Recall
OBSS AI Submission 2	38175	**0.906**
IIT Submission 1	11854	0.486
Dongguk University	30471	0.392
Shandong Normal University	2912	0.133

TABLE 4. Number of detections and recall.

by IIT are in the same range for different UAV sizes, yielding the best results for drone sizes in the range between 16^2 and 32^2 pixels. While the recall rates for DU are similar for different drone sizes, the AP values are worse for smaller drones. This indicates that more false positive detections are caused in case of small drone sizes. The results for SNU show that only large drones are detected, whereas all small drones are missed. One reason for this is the used image scale, which results in clearly down-scaled input images, so that small drones only comprise a few pixels.

Sequence	Recall			
	OBSS	IIT	DU	SNU
GOPR5867_001	**1.000**	0.683	0.522	0.834
GH010037_solo_split02	**1.000**	0.547	0.529	0.108
GH010039_matrice_split02	**0.970**	0.000	0.000	0.000
GH010040_inspire_split03	**0.696**	0.000	0.000	0.000
GH010045_phantom_split01	**0.973**	0.566	0.730	0.000
VID_20220306_170118_01	**0.776**	0.643	0.287	0.000
VID_20220306_170541_01	**0.998**	0.936	0.255	0.000
VID_20220311_122209_01	**1.000**	**1.000**	0.000	0.000
VID_20210417_143217_01	**0.970**	0.044	0.000	0.007
VID_20210606_141511_01	**0.574**	0.056	0.202	0.000
GOPR5852_001	**1.000**	0.195	0.368	0.005
GOPR5861_001	**0.998**	0.613	0.624	0.472
VID_20211012_081448_01	**0.970**	0.774	0.000	0.000
2019_10_16_C0003_52_30_mavic	**0.872**	0.266	0.280	0.368
dji_mavick_mountain_cross	**0.923**	0.541	0.000	0.000
dji_phantom_mountain	**0.752**	0.358	0.000	0.000
GOPR5843_004	**0.945**	0.686	0.908	0.000
GOPR5847_001	**0.690**	0.595	0.582	0.000
GOPR5853_002	**0.754**	0.145	0.280	0.000
GOPR5856_001	**1.000**	0.474	0.567	0.253
GOPR5862_001	**1.000**	0.545	0.649	0.755
GOPR5868_001	**1.000**	0.987	0.487	0.828
VID_20210606_141851_01	**0.700**	0.025	0.011	0.000
VID_20210606_143947_04	**0.759**	0.011	0.000	0.000
VID_20211010_143610_01	**1.000**	0.995	0.799	0.117
VID_20211012_175158_02	**0.968**	0.667	0.876	0.000
4k_2020-06-22_C0006_split_01_01	**0.814**	0.002	0.058	0.000
4k_2020-07-29_C0020_01	**0.988**	0.676	0.958	0.000
4k_2020-07-29_C0021_01	**1.000**	**1.000**	**1.000**	0.000
VID_20210417_143930_02	**0.868**	0.391	0.167	0.110

TABLE 5. Detailed comparison for each team in the Drone-vs-Bird Detection Challenge 2023. The Recall is given for every sequence of the test set.

Team	AP			
	$< 16^2$	$> 16^2 \& < 32^2$	$> 32^2 \& < 64^2$	$> 64^2$
OBSS	**0.756**	**0.760**	**0.834**	**0.908**
IIT	0.485	0.544	0.444	0.473
DU	0.114	0.130	0.290	0.216
SNU	0.000	0.000	0.003	0.289

TABLE 6. AP for different drone sizes.

Team	Recall			
	$< 16^2$	$> 16^2 \& < 32^2$	$> 32^2 \& < 64^2$	$> 64^2$
OBSS	**0.843**	**0.909**	**0.883**	**0.925**
IIT	0.549	0.630	0.464	0.387
DU	0.398	0.402	0.418	0.370
SNU	0.000	0.000	0.014	0.310

TABLE 7. Recall for different drone sizes.

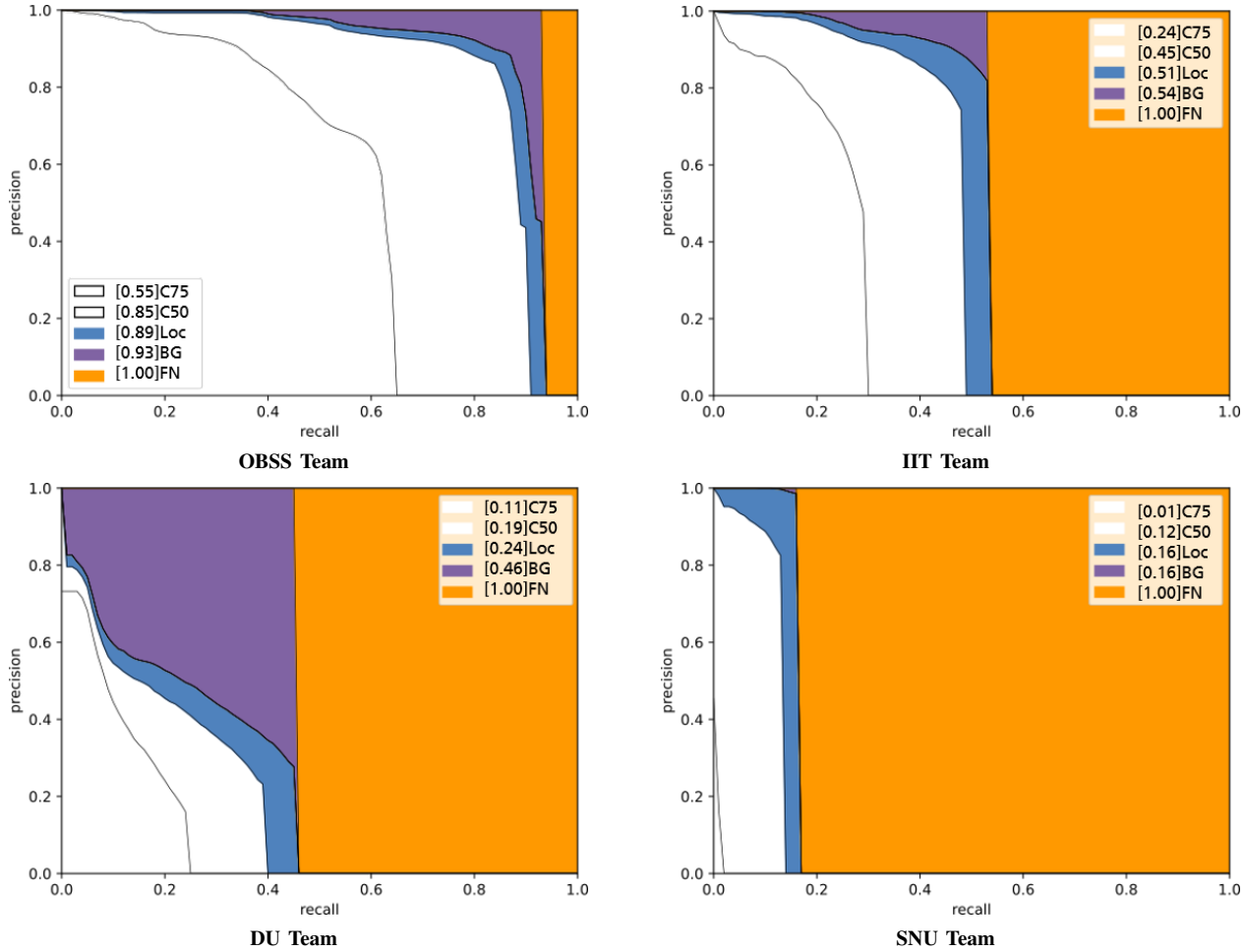

FIGURE 7. Error analysis using the COCO evaluation toolbox [45]. For each team, a series of precision recall curves (PRCs) is given. C75 is the PRC for an IoU of 0.75 to accept detections as true positives, while C50 is the PRC for an IoU of 0.5 as used within this challenge. Loc depicts localization errors. For this, the IoU criterion is set to 0.1. BG shows false positive detections caused by the background and FN illustrates the remaining false negatives.

A detailed analysis of the occurring errors is given in Figure 7. A series of precision recall curves (PRCs) is given for each team. C75, C50 and Loc are the PRCs for IoU thresholds of 0.75, 0.5 and 0.1, respectively. Due to the less strict IoU threshold, Loc depicts inaccurately localized detections. BG points out false positive detections caused by the background, while FN shows remaining false negative detections. For OBSS, the remaining errors are caused by inaccurate localization, false positive detections due to background clutter and missed detections, while no error source clearly dominates. The main error source for IIT is missed detections. One reason for this could be the applied filtering scheme, as several drone detections might be filtered out. In addition to the high number of missed detections, DU exhibits numerous false positive detections caused by the background. This indicates that the applied model is not able to accurately distinguish between drones and clutter objects. For SNU, the errors are mainly due to false negative detections. As already discussed one reason for this is the inappropriate down-scaling of the input images. The high numbers of missed detections for IIT, DU and SNU indicate the poor generalization ability of the applied models. In contrast to OBSS, these teams considered only the Drone-vs-Bird challenge train set and no additional datasets for training. However, the test set comprises multiple scenes with partially complex backgrounds, which are unseen during training and thus, may cause missed detections due to unexpected appearances of drones. This indicates that most approaches are robust only in case of sequences comparable with those in the training set, e.g. drones with the sky as background, while performance may significantly change in case of variations in the UAV sizes and complexity of the background, as structured background often yields weak contrast to drones. Considering that the best performance have been obtained by using additional datasets for training, it is apparent that the diversity in the training data plays an important role for both adaptability and generalizability.

Examples of qualitative detection results for all teams are given in Figure 8 and Figure 9. Note that only detections with a confidence score above 0.5 are considered. In case of large drones and unstructured background, all approaches achieved good detection results (see Figure 8). However, small drones

as well as drones in front of background are only detected by OBSS and IIT or only by OBSS. In case of more complex backgrounds or weak contrast between drone and background, all approaches have issues to correctly detect drones. Besides more diverse training data or novel data augmentation techniques, the usage of temporal information could be beneficial for such scenarios.

B. Discussion

The 6th edition of the Drone-vs-Bird challenge involved the participation of four distinct research teams. The analyses and results reported in Section V-A clearly demonstrated that the algorithm proposed by the OBSS team significantly outperformed the approaches proposed by the other participating teams over all the sequences provided in the test set. From a more technical point of view, the superiority of the detection framework proposed by OBSS can be ascribed to its ability to mitigate effects introduced by mobile cameras and to detect distant drones. Although all the teams applied only appearance-based detectors on single frames without considering more extended motion information, OBSS inserted a tracking approach in the processing loop that helped the proposed method to identify with high probability the presence of drones even in case of small and blurry appearance. Moreover, the drone detection model used by OBSS was trained on four different drone detection datasets to increase the model generalization ability and was further augmented with a synthetic drone dataset that lead to improved the detection performance in case of complex and unseen backgrounds. On the other hand, the methods proposed by IIT, DU and SNU teams could be considered effective only for sequences with simple (non-structured) backgrounds and with large drones appearing at same instant. When facing scenes with more complex backgrounds or smaller drones, all the methods from IIT, DU, and SNU tend to exhibit a too high number of missed detections, while the method proposed by OBSS is able to limit the number of miss detections or false alarms, though at the price of a reduced AP.

Overall, for the case of large drones and unstructured background, all approaches achieved satisfactory detection results. However, small drones as well as drones in front of backgrounds are detected by OBSS and, only in part, by IIT. All approaches suffered in case of more complex backgrounds or weak contrast between drone and background. One of the primary difficulty arises from managing mobile cameras and detecting distant drones. Another important aspect to highlight is that most of the algorithms do not explicitly incorporate birds in a supervised manner during the design phase due to the absence of annotated bird data. Consequently, instances where multiple birds are present in test sequences (including scenes with entire flocks) tend to result in increased false alarms across all methods, as birds share small visual characteristics with small (distant) UAVs. Not least, the majority of the models adopted by the participating teams were trained on a few different real and synthetic datasets, thus exhibiting a rather poor generalization capability.

VI. Conclusion

This paper presented an overview of the outcomes from the 6th edition of the Drone vs Birds Detection Grand Challenge at ICASSP 2023. The four methods proposed by the participating teams exhibit distinct design elements, leading to a complementary set of interesting aspects. Notably, the primary difficulties arise from managing mobile cameras and detecting distant drones. Another important aspect to highlight is that most of the algorithms do not explicitly incorporate birds in a supervised manner during the design phase due to the absence of annotated bird data. Consequently, instances where multiple birds are present in test sequences (including scenes with entire flocks) tend to result in increased false alarms across all methods, as birds share small visual characteristics with small (distant) UAVs. Incorporating bird targets into the training dataset has been proved a challenging and labour-intensive task. In drone tracking footage, birds appear as small and blurry flying objects, often not easy to be identified in single images as bird without utilising motion information within a sequence. Furthermore, in case of flock of birds, annotation would be a very time-consuming task, that cannot promise a satisfactory accuracy. Addressing this issue necessitates devising strategies to integrate bird data, particularly given the visual similarity between distant fixed-wing UAVs and birds. It should be also kept in mind that the goal is drone detection, not classification of the rest of the scene. Designing a method for more general object detection and classification would lead to a different approach for future extension of the challenge, incorporating an additional class representing birds at the design stage, as well as other classes for similar flying objects (e.g., airplanes). To this aim, a first step could be to annotate birds only in videos where their appearance is evident enough both for the sake of annotation and useful training of the detection method, which also makes it possible to consider the adoption of semi-automatic annotation tools. Additionally, videos solely of birds (available on the Internet) could be used to train a method with bird appearance features. Another possibility would be to generate videos with drones and birds, by augmenting a drone video with synthetically generated flying bird(s). This would alleviate the hassle of bird annotation, but requires to construct suitable methods for realistic bird flights generation. All such aspects warrant further exploration and will be a focal point in upcoming editions of the Drone vs. Bird Detection Challenge.

More generally, understanding the main factors that contribute to the evident performance variations exhibited by each algorithm across different sequences is an important direction of further research, in particular for what concerns the ability to cope with arbitrarily-complex backgrounds. Future

FIGURE 8. Qualitative examples for OBSS AI (green), IIT (red), Dongguk (blue) and Shandong (yellow) showing good detection results in case of large drones and unstructured background.

editions of the challenge could also incorporate additional assessments: besides the mentioned multi-class extension, other performance aspects such as computational efficiency (including real-time capabilities) could be investigated. The use of a shared Docker container installed on a remote machine (e.g., using one of the cloud facilities) could be a viable solution to compare the runtime of the proposed algorithms on the same hardware platform and assess whether they are suitable for real-time implementation. The latter is expected to evolve into a crucial requirement in the future editions of the challenge, given the increasing importance of promptly detecting drones as a fundamental prerequisite for modern UAV detection systems.

In conclusion, all the inquiries above aim to unravel the intricate trade-offs inherent in the multitude of approaches and methodological combinations adopted for drone detection based on video signal processing, contributing to a deeper understanding of their underlying mechanisms and highlighting, at the same time, the most promising research directions.

REFERENCES

[1] H. V. Nguyen, H. Rezatofighi, B.-N. Vo, and D. C. Ranasinghe, "Online UAV Path Planning for Joint Detection and Tracking of Multiple Radio-Tagged Objects," *IEEE Transactions on Signal Processing*, vol. 67, no. 20, pp. 5365–5379, 2019.

[2] K. K. Nguyen, A. Masaracchia, V. Sharma, H. V. Poor, and T. Q. Duong, "RIS-Assisted UAV Communications for IoT With Wireless Power Transfer Using Deep Reinforcement Learning," *IEEE Journal of Selected Topics in Signal Processing*, vol. 16, no. 5, pp. 1086–1096, 2022.

[3] A. Fascista, "Toward Integrated Large-Scale Environmental Monitoring Using WSN/UAV/Crowdsensing: A Review of Applications, Signal Processing, and Future Perspectives," *Sensors*, vol. 22, no. 5, 2022. [Online]. Available: https://www.mdpi.com/1424-8220/22/5/1824

[4] V. Chamola, P. Kotesh, A. Agarwal, Naren, N. Gupta, and M. Guizani, "A Comprehensive Review of Unmanned Aerial Vehicle Attacks and Neutralization Techniques," *Ad Hoc Networks*, vol. 111, p. 102324, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1570870520306788

[5] (2023) UAS Sightings Report. [Online]. Available: https://www.faa.gov/uas/resources/public_records/uas_sightings_report

[6] (2021) A Drone Tried to Disrupt the Power Grid. It Won't Be the Last. [Online]. Available: https://www.wired.com/story/drone-attack-power-substation-threat/

[7] (2022) Drone operator above Tesla Giga Berlin spoils routine descent for passenger plane. [Online]. Available: https://www.teslarati.com/tesla-giga-berlin-drone-operator-berlin-brandenburg-airport-plane/

[8] (2022) Security forces foil narco-terrorism bid. [Online]. Available: https://tinyurl.com/5hy2h6v7

[9] I. Guvenc, F. Koohifar, S. Singh, M. L. Sichitiu, and D. Matolak, "Detection, Tracking, and Interdiction for Amateur Drones," *IEEE Communications Magazine*, vol. 56, no. 4, pp. 75–81, 2018.

[10] A. Coluccia, G. Parisi, and A. Fascista, "Detection and Classification of Multirotor Drones in Radar Sensor Networks: A Review," *Sensors*, vol. 20, no. 15, 2020. [Online]. Available: https://www.mdpi.com/1424-8220/20/15/4172

[11] A. G. Haddad, M. Ahmed Humais, N. Werghi, and A. Shoufan, "Long-Range Visual UAV Detection and Tracking System with Threat Level Assessment," in *IECON 2020 The 46th Annual Conference of the IEEE Industrial Electronics Society*, 2020, pp. 638–643.

[12] B. K. S. Isaac-Medina, M. Poyser, D. Organisciak, C. G. Willcocks, T. P. Breckon, and H. P. H. Shum, "Unmanned Aerial Vehicle Visual Detection and Tracking Using Deep Neural Networks: A Performance Benchmark," in *Proceedings of the IEEE/CVF International Confer-*

FIGURE 9. Qualitative examples for OBSS AI (green), IIT (red), Dongguk (blue) and Shandong (yellow) showing reasons for missed detections.

ence on Computer Vision (ICCV) Workshops, October 2021, pp. 1223–1232.

[13] J. Zhao, J. Zhang, D. Li, and D. Wang, "Vision-Based Anti-UAV Detection and Tracking," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 12, pp. 25 323–25 334, 2022.

[14] J. Li, D. H. Ye, M. Kolsch, J. P. Wachs, and C. A. Bouman, "Fast and robust UAV to UAV detection and tracking from video," *IEEE Transactions on Emerging Topics in Computing*, vol. 10, no. 3, pp. 1519–1531, 2021.

[15] S. Samaras, E. Diamantidou, D. Ataloglou, N. Sakellariou, A. Vafeiadis, V. Magoulianitis, A. Lalas, A. Dimou, D. Zarpalas, K. Votis, P. Daras, and D. Tzovaras, "Deep Learning on Multi Sensor Data for Counter UAV ApplicationsA Systematic Review," *Sensors*, vol. 19, no. 22, 2019. [Online]. Available: https://www.mdpi.com/1424-8220/19/22/4837

[16] T. Müller, H. Widak, M. Kollmann, A. Buller, L. W. Sommer, R. Spraul, A. Kröker, I. Kaufmann, A. Zube, F. Segor *et al.*, "Drone detection, recognition, and assistance system for counter-uav with vis, radar, and radio sensors," in *Automatic Target Recognition XXXII*, vol. 12096. SPIE, 2022, pp. 94–108.

[17] A. Coluccia, M. Ghenescu, T. Piatrik, G. De Cubber, A. Schumann, L. Sommer, J. Klatte, T. Schuchert, J. Beyerer, M. Farhadi *et al.*, "Drone-vs-bird detection challenge at IEEE AVSS2017," in *14th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*. IEEE, 2017, pp. 1–6.

[18] A. Coluccia, A. Fascista, A. Schumann, L. Sommer, M. Ghenescu, T. Piatrik, G. De Cubber, M. Nalamati, A. Kapoor, M. Saqib, N. Sharma, M. Blumenstein, V. Magoulianitis, D. Ataloglou, A. Dimou, D. Zarpalas, P. Daras, C. Craye, S. Ardjoune, D. De la Iglesia, M. Mndez, R. Dosil, and I. Gonzlez, "Drone-vs-Bird Detection Challenge at IEEE AVSS 2019," in *2019 16th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, 2019, pp. 1–7.

[19] A. Coluccia, A. Fascista, A. Schumann, L. Sommer, A. Dimou, 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*, vol. 21, no. 8, p. 2824, 2021.

[20] A. Coluccia, A. Fascista, A. Schumann, L. Sommer, A. Dimou, D. Zarpalas, F. C. Akyon, O. Eryuksel, K. A. Ozfuttu, S. O. Altinuc, F. Dadboud, V. Patel, V. Mehta, M. Bolic, and I. Mantegh, "Drone-vs-Bird Detection Challenge at IEEE AVSS2021," in *2021 17th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, 2021, pp. 1–8.

[21] A. Coluccia *et al.*, "Drone-vs-bird detection challenge at ICIAP 2021," in *Image Analysis and Processing. ICIAP 2022 Workshops*. Cham: Springer International Publishing, 2022, pp. 410–421.

[22] A. Coluccia, A. Fascista, A. Schumann, L. Sommer, A. Dimou, D. Zarpalas, and N. Sharma, "Drone-vs-Bird Detection Grand Challenge at ICASSP2023," in *2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2023, pp. 1–2.

[23] M. C. Aksoy, A. S. Orak, H. M. zkan, and B. Selimoglu, "Drone Dataset: Amateur Unmanned Air Vehicle Detection," 2019.

[24] B. Bosquet, M. Mucientes, and V. Brea, "STDnet: A ConvNet for Small Target Detection," in *Proceedings of the 29th British Machine Vision Conference*, Newcastle (UK), 2018.

[25] J. Li, D. H. Ye, T. Chung, M. Kolsch, J. Wachs, and C. Bouman, "Multi-target detection and tracking from a single camera in Unmanned Aerial Vehicles (UAVs)," in *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2016, pp. 4992–4997.

[26] A. Rozantsev, V. Lepetit, and P. Fua, "Flying objects detection from a single moving camera," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 4128–4136.

[27] M. Ł. Pawełczyk and M. Wojtyra, "Real World Object Detection Dataset for Quadcopter Unmanned Aerial Vehicle Detection," *IEEE Access*, vol. 8, pp. 174 394–174 409, 2020.

[28] "Anti-UAV Challenge," https://anti-uav.github.io/.

[29] J. Li, J. Murray, D. Ismaili, K. Schindler, and C. Albl, "Reconstruction of 3D flight trajectories from ad-hoc camera networks," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2020, pp. 1621–1628.

[30] S. O. Ajakwe, V. U. Ihekoronye, D.-S. Kim, and J. M. Lee, "Dronet: Multi-tasking framework for real-time industrial facility aerial surveillance and safety," *Drones*, vol. 6, no. 2, 2022.

[31] J. Zhao, J. Zhang, D. Li, and D. Wang, "Vision-based anti-uav detection and tracking," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 12, pp. 25 323–25 334, 2022.
[32] F. Svanstrm, C. Englund, and F. Alonso-Fernandez, "Real-time drone detection and tracking with visible, thermal and acoustic sensors," in *2020 25th International Conference on Pattern Recognition (ICPR)*, 2021, pp. 7265–7272.
[33] Y. Chen, P. Aggarwal, J. Choi, and C.-C. J. Kuo, "A deep learning approach to drone monitoring," in *Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, 2017, pp. 686–691.
[34] G. Jocher *et al.*, "ultralytics/yolov5: v3.1 - Bug Fixes and Performance Improvements," https://github.com/ultralytics/yolov5, Oct. 2020.
[35] F. C. Akyon, E. Akagunduz, S. O. Altinuc, and A. Temizel, "Sequence Models for Drone vs Bird Classification," *arXiv preprint arXiv:2207.10409*, 2022.
[36] C.-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao, "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 7464–7475.
[37] L. Alan, T. Vojíř, L. Čehovin, J. Matas, and M. Kristan, "Discriminative correlation filter tracker with channel and spatial reliability," *International Journal of Computer Vision*, vol. 126, no. 7, pp. 671–688, 2018.
[38] S. K. Mistry, S. Chatterjee, A. K. Verma, V. Jakhetiya, B. N. Subudhi, and S. Jaiswal, "Drone-vs-Bird: Drone Detection Using YOLOv7 with CSRT Tracker," in *ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2023, pp. 1–2.
[39] G. Jocher *et al.*, "Ultralytics YOLOv8," https://github.com/ultralytics/ultralytics.
[40] N. Kim, J.-H. Kim, and C. S. Won, "FAFD: Fast and Accurate Face Detector," *Electronics*, vol. 11, no. 6, p. 875, 2022.
[41] M. Kisantal, Z. Wojna, J. Murawski, J. Naruniec, and K. Cho, "Augmentation for small object detection," *arXiv preprint arXiv:1902.07296*, 2019.
[42] J.-H. Kim, N. Kim, and C. S. Won, "High-Speed Drone Detection Based On Yolo-V8," in *ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2023, pp. 1–2.
[43] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "SSD: Single shot multibox detector," in *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14*. Springer, 2016, pp. 21–37.
[44] P. Dong, C. Wang, Z. Lu, K. Zhang, W. Wan, and J. Sun, "S-Feature Pyramid Network and Attention Model for Drone Detection," in *ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2023, pp. 1–2.
[45] "COCO evaluation toolbox ," https://cocodataset.org/#detection-eval.

Angelo Coluccia(M'13 - SM'16) received the Ph.D. degree in Information Engineering in 2011 and is currently an Associate Professor of Telecommunications at the Department of Engineering, University of Salento (Lecce, Italy). He has been a research fellow at Forschungszentrum Telekommunikation Wien (Vienna, Austria), and has held a visiting position at the Department of Electronics, Optronics, and Signals of the Institut Supérieur de l'Aéronautique et de l'Espace (ISAE-Supaero, Toulouse, France). His research interests are in the area of multi-channel, multi-sensor, and multi-agent statistical signal processing for detection, estimation, localization, and learning problems. Relevant application fields are radar, wireless networks (including 5G and beyond), and emerging network contexts (including intelligent cyber-physical systems, smart devices, and social networks). He is Member of the Sensor Array and Multichannel Technical Committee and of the Data Science Initiative for the IEEE Signal Processing Society.

Alessio Fascista(M'19) received the Ph.D. degree in Engineering of Complex Systems from the University of Salento (Lecce, Italy) in 2019. He is currently an Assistant Professor of Telecommunications at the Department of Innovation Engineering, University of Salento. He has held a visiting position at the Department of Telecommunications and Systems Engineering of the Universitat Autonoma de Barcelona (UAB, Spain) in 2018, and at the Department of Electrical Engineering of the Chalmers University of Technology (Gothenburg, Sweden) in 2022. His main research interests are in the field of telecommunications with focus on statistical signal processing for detection, estimation, and localization in terrestrial wireless systems. He is Member of IEEE and Member of the Technical Area Committee in Signal Processing for Multisensor Systems of EURASIP. He serves as an Associate Editor for the IEEE Open Journal of the Communications Society (OJ-COMS).

Lars Sommeris working as research scientist in the Video Exploitation Systems Department, Fraunhofer IOSB. He received the B.S. and M.S. degrees in electrical engineering and information technology from the Karlsruhe Institute of Technology (KIT), in 2011 and 2014, respectively. In 2020, he received the Ph.D. degree from the Karlsruhe Institute of Technology (KIT). His research mainly focuses on image analysis using machine learning, especially deep learning based classification, detection and segmentation and explainable AI.

Arne Schumannreceived his diploma in computer science in 2011 from Karlsruhe Institute of Technology (KIT). He has since worked as researcher on several computer vision subjects at KIT, the Fraunhofer Institute of Optronics, System Technologies and Image Exploitation (IOSB) and Queen Mary University of London. He received his PhD in 2019 from KIT. He is now a senior scientist in the Video Exploitation Systems Department at Fraunhofer IOSB and his work primarily focuses on deep learning methods for image exploitation, including object detection, classification, few-shot learning and data-centric AI methods. He has authored or co-authored over 45 scientific publications, several of which focus on the subject of UAV detection, tracking and classification.

Anastasios Dimouis a Researcher at the Information Technologies Institute (ITI) of the Centre for Research and Technology Hellas (CERTH). He received the Diploma in Electrical and Computer Engineer from Aristotle University of Thessaloniki (AUTH), Greece, the Professional Doctorate in Engineering (PDEng) in Information and Communication Technology from the Technical University of Eindhoven (TU/e), the Netherlands, and the PhD degree on image processing for surveillance applications from the Universidad Politcnica de Madrid (UPM), Spain. His work has led to the co-authoring of more than 40 publications in refereed international journals and conferences, including 8 journals, 32 conferences and 2 book chapters. He regularly acts as a reviewer for multiple conferences and journals in his domain.

Dimitrios Zarpalas is a Principal Researcher (grade B) at the Information Technologies Institute (ITI) of the Centre for Research and Technology Hellas (CERTH). He holds the diploma of Electrical and Computer Engineer from Aristotle University of Thessaloniki, A.U.Th, an MSc in computer vision from The Pennsylvania State University, and a PhD in medical informatics (School of Medicine, A.U.Th). His main research interest are on 3D/4D computer vision and machine learning, such as tele-immersion applications: volumetric video, 4D reconstruction of moving humans, their hologram compression and transmission in real-time; 3D motion capturing, analysis and evaluation; 3D object recognition and 3D shape descriptor extraction; 3D medical image processing, shape analysis of anatomical structures; while in the past has also worked in indexing, search and retrieval and classification of 3D objects, proteins and 3D model watermarking. He has (co-)authored more than 75 papers in peer reviewed international journals, conference proceedings, and books (including one IEEE Distinguished paper and one IEEE conference best paper award).