2021 17th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS) | 978-1-6654-3396-9/21/$31.00 ©2021 IEEE | DOI: 10.1109/AVSS52988.2021.9663844

Drone-vs-Bird Detection Challenge at IEEE AVSS2021

Angelo Coluccia, Alessio Fascista
University of Salento, Department of Innovation Engineering
via Monteroni, 73100 Lecce, Italy
name.surname@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
name.surname@iosb.fraunhofer.de

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

Fatih Cagatay Akyon, Ogulcan Eryuksel, Kamil Anil Ozfuttu, Sinan Onur Altinuc
OBSS AI
OBSS Technology, Universiteler Mah.
1606. Cad. No:4/1/307 Cyberpark Cyberplaza C blok 3.kat
Cankaya/Ankara
{fatih.akyon, ogulcan.eryuksel, anil.ozfuttu, sinan.altinuc}@obss.com.tr

Fardad Dadboud, Vaibhav Patel, Varun Mehta, Miodrag Bolic
Computational Analysis and Acceleration Research Group (CARG)
School of Electrical Engineering and Computer Science (SEECS), University of Ottawa
800 King Edward, Ottawa, On., Canada
{fardad.dadboud, vpate029, vmehta2, miodrag.bolic}@uottawa.ca

Iraj Mantegh
National Research Council Canada (NRC)
Montreal, Qc., Canada
iraj.mantegh@cnrc-nrc.gc.ca

Abstract

This paper presents the 4-th edition of the "drone-vs-bird" detection challenge, launched in conjunction with the the 17-th IEEE International Conference on Advanced Video and Signal-based Surveillance (AVSS). The objective of the challenge is to tackle the problem of detecting the presence of one or more drones in video scenes where birds may suddenly appear, taking into account some important effects such as the background and foreground motion. The proposed solutions should identify and localize drones in the scene only when they are actually present, without being confused by the presence of birds and the dynamic nature of the captured scenes. The paper illustrates the results of the challenge on the 2021 dataset, which has been further

extended compared to the previous edition run in 2020.

1. Introduction

The ever-increasing widespread of unmanned aerial vehicle (UAV) technologies in the modern society is opening the door to a number of unprecedented opportunities, but poses at the same time considerable practical risks. This calls for the need of advanced solutions to counteract the possible misuse of UAVs for illegal activities. Indeed, due to their accessibility and ease of use, UAVs can be also employed for smuggling (illegal transportation at borders or in restricted areas), illegal surveillance, privacy violation, interference with aircraft operations and terrorist attacks.

To combat these raising threats, advanced surveillance and detection systems based on different technologies are currently under investigation in the literature [1, 2]. A good trade-off between cost and detection range is achieved by video analytics-based UAV techniques that process sequences of images acquired by cameras in a given surveilled area. To be effective in detecting UAVs, such techniques need to correctly handle their visual similarity with other small flying objects such as birds, which make the surveillance task a challenging problem, especially when the targets are spread across several pixels in the image and the acquisitions are made under unfavorable conditions such as reduced visibility, long distances and weak contrast. The task becomes even more difficult due to the peculiar characteristics of small UAVs, which can reach speeds greater than 100 km/h and perform rapid maneuvers, including stop&go motions [3].

Aiming at advancing the state-of-the-art on UAV detection based on video-analytics, in 2017 the first edition of the *International Workshop on Small-Drone Surveillance, Detection and Counteraction Techniques* (WOSDETC) [4] was organized as part of the 14th edition of the *IEEE International Conference on Advanced Video and Signal based Surveillance* (AVSS), held in Lecce, Italy. In conjunction with the workshop, a grand challenge called *drone-vs-bird detection challenge* was launched thanks to the support of the SafeShore Consortium[1]. As the name suggest, the aim of the challenge is to tackle one of the main issues of the considered scenarios [5, 6]. In 2019, we run a second edition of the challenge, again as part of WOSDETC and co-located with the 16th edition of AVSS and held in Taipei, Taiwan [7]. 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 [8]. Since its first edition, the challenge has attracted more than 120 research groups working on the design of novel techniques to tackle the problem of discriminating drones from birds, from all over the world. The main goal is to correctly identify the presence of drones suddenly appearing in a video sequence, without being confused by birds or additional disturbing objects in the scene. The 2021 challenge dataset comprises different video sequences covering both maritime and land scenarios, acquired under different cameras and background conditions. All the participants submitted a set of score files containing the detection results, together with a companion paper describing the proposed methodology. More than 70 different research groups requested access to the dataset for participation in this edition of the challenge.

Figure 1. Examples of drone types present in the training set, i.e., Parrot Disco, 2 custom fix wing drones, DJI Inspire, DJI Phantom, DJI Mavic, DJI Matrice and 3DR Solo Robotics.

2. Challenge Dataset and Evaluation Protocol

For the Drone-vs-Bird Detection Challenge 2021, 77 different video sequences have been made available as training data. These video sequences originate from the previous installment of the challenge and were collected using MPEG4-coded static cameras by the SafeShore project, by the Fraunhofer IOSB research institute and by the ALADDIN[2] project. On average, the video sequences consist of 1,384 frames, while each frame contains 1.12 annotated drones. The video sequences are recorded with both static cameras and moving cameras and the resolution varies between 720×576 and 3840×2160 pixels. In total, 8 different types of drones exist in the dataset (see Fig. 1), *i.e.* 3 with fixed wings and 5 rotary ones. For each video, a separate annotation file is provided, which contains the frame number and the bounding box (expressed as [top$_x$ top$_y$ width height]) for the frames in which drones enter the scenes.

The video sequences provided for training exhibit a high variability in difficulty, including sequences with sky or vegetation as background, different weather types (cloudy, sunny), direct sun glare and variation in camera characteristics (see Fig. 2). The large variations in drone sizes due

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

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

Figure 2. Sample frames extracted from the training videos exhibit the large variability of the dataset.

to the strongly varying distance of the drones from the camera and the large number of small-sized drones make the detection task very challenging [8]. Birds that are not individually annotated occur in several of the videos and are the main disturbing object type (see Fig. 3).

The challenge test set consists of further 21 video sequences for which no annotations are provided. Fourteen of these video sequences originate from the previous installment of the challenge and exhibit similar characteristics to the training set, as most locations displayed in the test set are also contained in the training set, but not all. In this 2021 edition, seven further video sequences have been added to the test set, which comprise novel backgrounds, two novel rotary drone types, disturbing objects (*e.g.*, planes) and drones in front of structured background (see Fig. 4). The test set contains five sequences recorded with moving camera. Note that sequences exceeding 1 minute are shortened in order to avoid individual videos from dominating the resulting evaluation scores. The dataset is freely available for download upon signing a Data Usage Agreement (DUA). The annotations are available at https://github.com/wosdetc/challenge.

In the Drone-vs-Bird Detection Challenge 2021, each submission had to include a result file for each video sequence, providing the frame numbers and the predicted bounding boxes ([top_x, top_y, width, height]) and corresponding confidence scores for each frame. For frames not reported in the files, no detection was assumed. Average Precision (AP) was employed as evaluation metric to rank the submitted results. For this, detections are counted as true positive detections, if the Intersection over Union (IoU) criterion with a ground truth annotation exceeds 0.5. Detections with an IoU below 0.5 are counted as false positive detections, while ground truth annotations without an assigned detection are counted as false negative detections. Note that the test sequences have been made publicly available three days before the submission deadline.

Figure 3. Examples with birds that are the main disturbing objects.

Figure 4. Sample frames extracted from the test videos exhibiting differences to the training set.

3. Participation and Best Proposed Algorithms

We briefly summarize the methodology of the two best performing approaches, and also of the winner of the 2020 edition, for comparison purposes.

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.

Akyon et al. from OBSS AI, Ankara, used YOLOv5 [9], a single-shot detection model that provides good performance and high inference speed. Synthetic data generation and object tracking based improvements were applied to improve performance. Authors used real data mixed with synthetically generated data selectively to fine tune the YOLOv5 model. Synthetic drone images were generated with different backgrounds with various generation parameters. Generation parameters were optimized and a subset of synthetic data was selected with a data-centric approach. To further improve the performance, authors used a Kalman-based object tracking [10] to extract temporal information about the detected drones. Tracker and temporal information were used to delete very short-lived predictions and boost the confidence of a detection in a particular track.

CARG-UOTTAWA team from Computational Analysis and Acceleration Research Group (CARG) at the University of Ottawa and National Research Council Canada (NRC) proposed a single-stage YOLOV5-based [9] object detector for small UAV detection and classification. The team utilized YOLOV5x architecture which leverages Cross Stage Partial (CSP) [11] backbone with a faster inference speed. The YOLOV5 model applied Path Aggregation Network

(PANet) [12] that is similar to the Feature Pyramid Network (FPN). The PANet has another bottom-up pathway beyond the two FPN's pathways, bottom-up and up-down paths. The YOLOV5 model also applied scaling, color space adjustments, and mosaic training data augmentation. Besides, authors combined the challenge dataset with one of the publicly available UAV air to air dataset, that is called "Det-Fly" [13], to add more complex background and lighting conditions in the training dataset.

4. Results

The results for each team in the Drone-vs-Bird Detection Challenge 2021 are given in Table 1. The AP is given for every sequence of the test set, while the AP computed over all sequences is given at the bottom. The results are further compared to the winning entry of the Drone-vs-Bird Detection Challenge 2020 and two baseline detection methods, *i.e.* Faster R-CNN [14] with Feature Pyramid Network (FPN) [15] and RetinaNet [16]. The winner of the Drone-vs-Bird Detection Challenge 2020 adapted Cascade R-CNN [17] for the task of drone detection by modifying the anchor scales. Furthermore, the training data was extended by an additional drone dataset. To train the two baseline methods, we employed the object detection toolbox MMDetection [18]. For each model, we use ResNet-50 [19] as backbone architecture, weights pre-trained on MS COCO [20] for initialization and the standard configurations. During training and evaluation, all input images are scaled to 1920×1080 pixels.

OBSS AI obtained the overall best AP, ranking first in the Drone-vs-Bird Detection Challenge 2021. For almost each sequence, OBSS AI achieves higher detection scores compared to CARG-UOTTAWA. Compared to the results of the winning entry of the previous edition, OBSS AI exhibits similar results on the sequences from the last challenge. AP values close to 1 are achieved for some sequences (*e.g.* dji_phantom_close), as the occurring drones are comparatively large and clearly visible. Poor detection results are obtained in case of sequences recorded at twilight (*e.g.* 2019_11_14_C0001_11_23_inspire_1m) or sequences that contain segments with very distant drones (*e.g.* GOPR5847_001). While the results of both baseline methods are similar to OBSS AI for most sequences, their performance clearly drops in case of fix wing drones (*e.g.* two_parrot_2).

OBSS AI outperforms CARG-UOTTAWA on most sequences that are added in 2021 to the test set. In video sequence *4k_2020-07-29_C0020_01*, CARG-UOTTAWA fails to detect the two distant drones, as the predicted bounding boxes are too large and thus, the IoU criterion is not satisfied. While OBSS AI achieves superior results on scenes with novel drone types and/or novel backgrounds (*e.g.* VID_20210606_143947_04), CARG-UOTTAWA is more robust towards disturbing objects such as planes (*e.g.* GOPR5868_001). In case of a distant drone in front of structured objects, *i.e.* street lamp, both submitted approaches are not able to detect the drone (*i.e.* 4k_2020-06-22_C0006_split_01_01). Compared to the baseline methods, OBSS AI exhibits higher AP values for sequences containing novel drone types and unknown background, while the results are almost similar for the other sequences.

Table 2 reports the recall rates for each team in the Drone-vs-Bird Detection Challenge 2021, the winning entry of the Drone-vs-Bird Detection Challenge 2020 and the two baselines. While OBSS AI achieves higher recall rates on most sequences compared to CARG-UOTTAWA, RetinaNet exhibits the highest overall recall. However, the overall AP achieved for RetinaNet is lower due to a higher number of false alarms and detections with low confidence scores. The recall rates for CARG-UOTTAWA are comparatively low: reason for this are inaccurate bounding box predictions, yielding IoU values below 0.5 to ground truth boxes. Decreasing the IoU threshold value to accept detections as true positives clearly improves the overall AP for CARG-UOTTAWA (see Table 3).

Qualitative detection results for both teams and the corresponding ground truth are given in Figures 5 and 6. Both methods achieve good detection results for different drone types and a large range of drone sizes in different environments (see Figure 5). Negative detection results given in Figure 6 indicate remaining challenges such as distant drones, drones in front of structured backgrounds and disturbing objects.

5. Conclusions

This work reported on the results of the 2021 drone-vs-bird detection challenge, held with the 17th IEEE International Conference on Advanced Video and Signal based Surveillance (AVSS). The challenge attracted remarkable interest, given the timeliness and significance of the topic. The solutions proposed by the two best teams have been described and compared, also with respect to the winner of the 2020 edition. As in previous installments, the prominent ingredients of the solutions are convolutional neural networks coupled with additional processing blocks typical of moving object detection, possibly in conjunction with augmentation of the dataset via synthetic generation or addition from other existing repositories.

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,*

Sequence	AP				
	OBSS AI	CARG-UOTTAWA	Winner 2020	Faster R-CNN	RetinaNet
GOPR5843_004	0.953	0.628	0.977	**1.000**	0.989
GOPR5845_002	0.992	0.874	0.987	**0.994**	0.988
GOPR5847_001	0.663	0.624	0.657	0.693	**0.736**
dji_matrice_210_midrange_cruise	0.985	0.915	**0.999**	0.997	0.988
dji_mavick_mountain_cross	0.877	0.579	0.773	0.892	**0.903**
dji_phantom_mountain	0.666	0.354	0.631	0.640	**0.731**
dji_phantom_close	**0.990**	0.902	0.988	0.837	0.698
two_parrot_2	0.927	0.867	**0.935**	0.302	0.436
parrot_disco_long_session_2_1m	**0.546**	0.210	0.423	0.133	0.086
2019_08_19_C0001_27_46_solo_1m	0.810	0.463	**0.900**	0.855	0.737
2019_08_19_C0001_57_00_inspire_1m	0.873	0.736	0.912	**0.898**	0.737
2019_09_02_C0002_11_18_solo	0.928	0.925	0.946	**0.954**	0.943
2019_10_16_C0003_52_30_mavic	0.519	0.523	0.548	0.697	**0.699**
2019_11_14_C0001_11_23_inspire_1m	0.100	0.088	0.389	0.779	**0.788**
4k_2020-06-22_C0006_split_01_01	0.000	0.000	-	0.014	**0.080**
4k_2020-07-29_C0020_01	0.541	0.000	-	0.522	**0.812**
4k_2020-07-29_C0021_01	0.998	0.312	-	**1.000**	**1.000**
GOPR5853_002	**0.618**	0.286	-	0.606	0.470
GOPR5868_001	0.823	**1.000**	-	0.989	0.919
VID_20210606_141851_01	**0.117**	0.000	-	0.054	0.039
VID_20210606_143947_04	**0.212**	0.000	-	0.168	0.000
Overall	**0.677**	0.500	-	0.628	0.620

Table 1. Detailed comparison for each team in the Drone-vs-Bird Detection Challenge 2021, the winning entry of the Drone-vs-Bird Detection Challenge 2020 and two baselines. The AP is given for every sequence of the test set. While the first block gives the results for sequences originating from the previous installment of the challenge, the second block reports the results for the novel sequences. The overall averaged result is given at the bottom.

Figure 5. Image crops showing qualitative detection results for OBSS AI (red), detection results for CARG-UOTTAWA (blue) and the corresponding GT (green). Both methods achieve good detection results for different drone types and a large range of drone sizes in different environments.

Sequence	Recall				
	OBSS AI	CARG-UOTTAWA	Winner 2020		
GOPR5843_004	0.968	0.646	0.982	**1.000**	0.995
GOPR5845_002	0.993	0.917	**0.998**	0.994	0.993
GOPR5847_001	0.667	0.632	0.660	0.695	**0.748**
dji_matrice_210_midrange_cruise	0.988	0.942	0.999	**0.998**	0.992
dji_mavick_mountain_cross	0.896	0.676	0.818	0.914	**0.931**
dji_phantom_mountain	0.725	0.384	0.695	0.669	**0.779**
dji_phantom_close	**0.991**	0.942	0.989	0.889	0.816
two_parrot_2	**0.952**	0.902	0.951	0.305	0.462
parrot_disco_long_session_2_1m	**0.897**	0.248	0.687	0.242	0.266
2019_08_19_C0001_27_46_solo_1m	0.843	0.536	**0.929**	0.889	0.873
2019_08_19_C0001_57_00_inspire_1m	0.884	0.769	**0.931**	0.902	0.896
2019_09_02_C0002_11_18_solo	0.949	0.936	0.954	0.958	**0.960**
2019_10_16_C0003_52_30_mavic	0.645	0.566	0.688	0.783	**0.813**
2019_11_14_C0001_11_23_inspire_1m	0.111	0.095	0.568	0.779	**0.792**
4k_2020-06-22_C0006_split_01_01	0.000	0.000	-	0.248	**0.531**
4k_2020-07-29_C0020_01	0.733	0.007	-	0.874	**0.914**
4k_2020-07-29_C0021_01	0.998	0.556	-	**1.000**	**1.000**
GOPR5853_002	**0.677**	0.286	-	0.606	0.470
GOPR5868_001	0.998	**1.000**	-	**1.000**	**1.000**
VID_20210606_141851_01	0.188	0.011	-	0.181	**0.256**
VID_20210606_143947_04	**0.673**	0.000	-	0.394	0.029
Overall	0.745	0.549	-	0.713	**0.747**

Table 2. Resulting recall for each team in the Drone-vs-Bird Detection Challenge 2021, the winning entry of the Drone-vs-Bird Detection Challenge 2020 and two baselines. The recall is given for every sequence of the test set and the overall recall is given at the bottom.

Figure 6. Image crops showing qualitative detection results for OBSS AI (red), detection results for CARG-UOTTAWA (blue) and the corresponding GT (green). Remaining challenges are distant drones, small drones in front of structured backgrounds and disturbing objects.

IoU	0.1	0.2	0.3	0.4	0.5	0.6
OBSS AI	0.744	0.741	0.737	0.723	0.677	0.575
CARG-UOTTAWA	0.687	0.642	0.634	0.588	0.500	0.348

Table 3. Average Precision for various IoU threshold values used to accept detections as true positives.

and Countermeasures, volume 10799, page 107990T. International Society for Optics and Photonics, 2018. 2

[2] Angelo Coluccia, Gianluca Parisi, and Alessio Fascista. Detection and classification of multirotor drones in radar sensor networks: A review. *Sensors*, 20(15), 2020. 2

[3] Angelo Coluccia, Alessio Fascista, and Giuseppe Ricci. Online estimation and smoothing of a target trajectory in mixed stationary/moving conditions. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 4445–4449, 2019. 2

[4] 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

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

[6] 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

[7] Angelo Coluccia, Alessio Fascista, Arne Schumann, Lars Sommer, Marian Ghenescu, Tomas Piatrik, Geert De Cubber, Mrunalini Nalamati, Ankit Kapoor, Muhammad Saqib, Nabin Sharma, Michael Blumenstein, Vasileios Magoulianitis, Dimitrios Ataloglou, Anastasios Dimou, Dimitrios Zarpalas, Petros Daras, Celine Craye, Salem Ardjoune, David De la Iglesia, Miguel Mndez, Raquel Dosil, and Iago Gonzlez. Drone-vs-bird detection challenge at ieee avss2019. In *2019 16th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, pages 1–7, 2019. 2

[8] Angelo Coluccia, Alessio Fascista, Arne Schumann, Lars Sommer, Anastasios Dimou, Dimitrios Zarpalas, Miguel Méndez, David De la Iglesia, Iago González, Jean-Philippe Mercier, et al. Drone vs. bird detection: Deep learning algorithms and results from a grand challenge. *Sensors*, 21(8):2824, 2021. 2, 3

[9] Glenn Jocher, Alex Stoken, Jirka Borovec, NanoCode012, ChristopherSTAN, Liu Changyu, Laughing, tkianai, Adam Hogan, lorenzomammana, yxNONG, AlexWang1900, Laurentiu Diaconu, Marc, wanghaoyang0106, ml5ah, Doug, Francisco Ingham, Frederik, Guilhen, Hatovix, Jake Poznanski, Jiacong Fang, Lijun Yu, changyu98, Mingyu Wang, Naman Gupta, Osama Akhtar, PetrDvoracek, and Prashant Rai. ultralytics/yolov5: v3.1 - bug fixes and performance improvements. `https://github.com/ultralytics/yolov5`, October 2020. 4

[10] Joaqun Alori, Alan Descoins, Braulio Ros, and Agustn Castro. tryolabs/norfair: v0.3.1. `https://github.com/tryolabs/norfair`, 2021. 4

[11] CY Wang, HYM Liao, IH Yeh, YH Wu, PY Chen, and JW Hsieh. Cspnet: A new backbone that can enhance learning capability of cnn. arxiv 2019. *arXiv preprint arXiv:1911.11929*. 4

[12] Shu Liu, Lu Qi, Haifang Qin, Jianping Shi, and Jiaya Jia. Path aggregation network for instance segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 8759–8768, 2018. 5

[13] Ye Zheng, Zhang Chen, Dailin Lv, Zhixing Li, Zhenzhong Lan, and Shiyu Zhao. Air-to-air visual detection of micro-uavs: An experimental evaluation of deep learning. *IEEE Robotics and Automation Letters*, 6(2):1020–1027, 2021. 5

[14] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. *Advances in neural information processing systems*, 28:91–99, 2015. 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] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. Focal loss for dense object detection. In *Proceedings of the IEEE international conference on computer vision*, pages 2980–2988, 2017. 5

[17] Zhaowei Cai and Nuno Vasconcelos. Cascade r-cnn: Delving into high quality object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 6154–6162, 2018. 5

[18] Kai Chen, Jiaqi Wang, Jiangmiao Pang, Yuhang Cao, Yu Xiong, Xiaoxiao Li, Shuyang Sun, Wansen Feng, Ziwei Liu, Jiarui Xu, Zheng Zhang, Dazhi Cheng, Chenchen Zhu, Tianheng Cheng, Qijie Zhao, Buyu Li, Xin Lu, Rui Zhu, Yue Wu, Jifeng Dai, Jingdong Wang, Jianping Shi, Wanli Ouyang, Chen Change Loy, and Dahua Lin. MMDetection: Open mmlab detection toolbox and benchmark. *arXiv preprint arXiv:1906.07155*, 2019. 5

[19] 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 (CVPR)*, June 2016. 5

[20] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft coco: Common objects in context. In *European conference on computer vision*, pages 740–755. Springer, 2014. 5