ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) | 978-1-7281-6327-7/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICASSP49357.2023.10095516

HIGH-SPEED DRONE DETECTION BASED ON YOLO-V8

Jun-Hwa Kim, Namho Kim, Chee Sun Won
Department of Electronics Electrical Engineering, Dongguk University, Seoul, Korea

ABSTRACT

Detecting drones in a video is a challenging problem due to their dynamic movements and varying range of scales. Moreover, since drone detection is often required for security, it should be as fast as possible. In this paper, we modify the state-of-the-art YOLO-V8 to achieve fast and reliable drone detection. Specifically, we add Multi-Scale Image Fusion and P2 Layer to the medium-size model (M-model) of YOLO-V8. Our model was evaluated in the 6th WOSDETC challenge.

Index Terms— Drone Detection, Small-object detection, YOLO-V8

1. INTRODUCTION

The 6th Drone-vs-Bird Detection Challenge in 2023 is a grand challenge jointly organized by the International Workshop on Small-Drone Surveillance, Detection and Counteraction Techniques (WOSDETC) and the International Conference on Acoustics, Speech and Signal Processing (ICASSP). This challenge aims to build a surveillance system that can respond to various environmental and safety issues related to drones. The dataset for the challenge has been expanded to include more realistic situations such as drones with birds. In particular, drones captured at far distances are often very small, making them challenging to detect and distinguish from birds. Therefore, improving the performance of small drone detection is crucial [1, 2]. Additionally, since real-time drone detection is essential for security and safety purposes, detection should be as fast as possible.

To improve both small drone detection performance and detection speed, we have adopted the state-of-the-art YOLO-V8 model [3]. Specifically, we use the medium-size (M-model) YOLO-V8 for high-speed detection, and we have also adopted the Multi-Scale Image Fusion (MSIF) technique [4]. Additionally, we have added a P2 layer in the YOLO-V8 architecture to deal with small-sized objects more effectively. In training, we have intensively used a modified copy & paste technique as an augmentation method to increase the number of drone appearances in the training images.

2. METHODS

Our drone detection strategy focuses on improving detection performance, particularly for small drones. To achieve this, we employ YOLO-V8 [3], which is an updated version of YOLO [5]. YOLO-V8 achieves state-of-the-art performance through model structure optimization, optimization of the anchor box or anchor-free scheme, and various data augmentation techniques. Our deep-learning architecture is based on the computationally less expensive medium-size YOLO-V8 (i.e., M-model YOLO-V8).

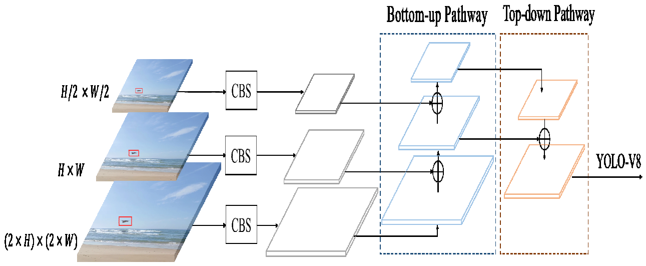

Fig. 1. The structure of MSIF. H and W are the height and width of the input image, respectively. CBS is a series of convolution layer, batch normalization layer, and SiLU layer. The output of the feature map is fed to the backbone of YOLO-V8.

Note that instead of resizing the original image, the image space is divided into multiple tiles, and each image tile can be treated sequentially to detect objects in a large-resolution image [6]. Although this tile-based method can be useful for detecting very small drones, the sequential access of tile-based images can be a bottleneck for real-time execution. In this paper, instead of the multi-tile approach, we adopt Multi-Scale Image Fusion (MSIF) [4] to deal with various sizes of drones. As shown in Fig 1, MSIF extracts the pixel-level features for three different scales of the input image, which are fused into one feature map through bottom-up and top-down structures. Then, the combined feature map is transferred to the input of YOLO-V8. Additionally, since the high-resolution feature map at the bottom of the feature pyramid has weak semantic information but strong spatial local features [7], we added a P2 layer to the feature pyramid of YOLO-V8.

Since drones usually occupy only a small portion of the whole image space, it is challenging to provide images with drones in various positions and scales to train deep networks. Therefore, to diversify the background and scale of the drones in the training images, we intensively utilize the copy & paste scheme [8]. That is, the cropped and scaled drones are pasted in various image spaces for training. The location for the paste is selected such that it does not overlap with the already drone-occupied area. An example image of the copy & paste scheme is shown in Fig. 2.

3. EXPERIMENTS RESULT

3.1. Implementation details and Dataset

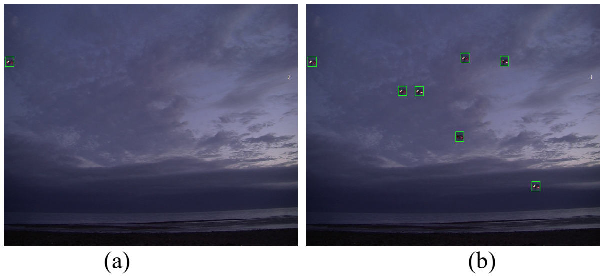
(a) (b)

Fig. 2. Example of copy & paste in Drone-vs-Bird dataset: (a) original image, (b) pasted image. The objects for copy & paste are from the Drone-vs-Bird dataset.

We evaluated the modified YOLO-V8-M model using the Drone-vs-Bird dataset [9], which consists of 77 training and 30 testing videos. Not all frames from the videos were included in the training dataset, but every fifth frame was used. YOLO-V8 [3] provides five different models {N, S, M, L, X} based on channel depth and number of filters. We adopted the M-model for our backbone architecture as it provides a good trade-off between detection accuracy and speed. We trained YOLO-V8-M for 93 epochs with a batch size of 16 using SGD optimization. To maintain the aspect ratio of the original image, we resized the images to 640 pixels along the long axis during training and 1280 pixels during testing. During the testing phase, we applied horizontal flip and multi-scale augmentation with scales of 1.0, 0.83, and 0.67. For non-maximum suppression (NMS), we set the confidence threshold and IoU (Intersection-over-Union) threshold to 0.25 and 0.1, respectively. All experiments were conducted using NVIDIA GTX 1080Ti GPU and Pytorch 1.12.1 environment.

3.2. Results

Table 1 shows our results for the Drone-vs-Bird Detection Challenge [9]. The results are Average Precision (AP) for each testing video. A detection is counted as correct when its IoU with a ground truth box is above 0.5. The fps (frames per second) of the P2 layer and MSIF [4] with YOLO-V8-M model is 45.7 fps and 17.6 fps respectively at image sizes of 640 and 1280. The fps is the result of measurement including all execution times pre-processing, model inference, and NMS stage.

Table 1. Detection performance for Drone-vs-Bird test dataset.

Sequence	AP	Sequence	AP
GOPR5867_001	0.521	dji_phantom_mountain	0
GH010037_solo_split02	0.361	GOPR5843_004	0.867
GH010039_matrice_split02	0	GOPR5847_001	0.564
GH010040_inspire_split03	0	GOPR5853_002	0.266
GH010045_phantom_split01	0.623	GOPR5856_001	0.539
VID_20220306_170118_01	0.077	GOPR5862_001	0.620
VID_20220306_170541_01	0.057	GOPR5868_001	0.483
VID_20220311_122209_01	0	VID_20210606_141851_01	0
VID_20210417_143217_01	0	VID_20210606_143947_04	0
VID_20210606_141511_01	0.041	VID_20211010_143610_01	0.733
GOPR5852_001	0.280	VID_20211010_143610_01	0.770
GOPR5861_001	0.606	4k_2020-06-22_C0006_split_01_01	0.001
VID_20211012_081448_01	0	4k_2020-07-29_C0020_01	0.914
2019_10_16_C0003_52_30_mavic	0.241	4k_2020-07-29_C0021_01	0.803
dji_mavick_mountain_cross	0	VID_20210417_143930_02	0.011
Overall AP	0.189		

4. CONCLUSION

To achieve fast and reliable drone detection, we modified the latest version of YOLO-V8. We applied a multi-scale image fusion to the M-model of YOLO-V8, enabling us to effectively train on multi-scale drone images. Additionally, we added a P2 layer to YOLO-V8 and extensively used copy & paste augmentation to improve detection performance, particularly for small objects. \

5. ACKNOWLEDGMENT

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (NRF-2023R1A2C1003588).

6. REFERENCES

[1] H. Sun, Y. Jian, S. Jiaquan, L. Dong, N. Z. Liu and Z. Huiyu, “TIB-NET: Drone detection network with tiny iterative backbone,” *Ieee Access*, pp.130697-130707, 2020.

[2] T. Sangam, R. D. Ishan, S. Wapas and S. Mubarak, “Tranvisdrone: Spatio-temporal transformer for vision-based drone-to-drone detection in aerial videos,” *arXiv preprint arXiv:2210.08423,* 2022.

[3] Ultralytics/ Ultralytics. Available online: https://github.com/ultralytics/ultralytics.

[4] N. Kim, J. H. Kim and C. S. Won, “FAFD: Fast and Accurate Face Detector,” *Electronics*, vol. 11(6), pp. 875, 2022.

[5] J. Redmon, S. Divvala, R. Girshick and A. Farhadi, “You only look once: Unified, real-time object detection,” *In Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 779-788, 2016.

[6] 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)*, pp. 1-8, 11 2021.

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

[8] M. Kisantal, Z. Wojna, J. Murawski, J. Naruniec and K. Cho, “Augmentation for small object detection,” *arXiv preprint arXiv:1902.07296*, 2019.

[9] Drone-vs-Bird Detection Challenge, WOSDETC:2023. Available online: https://wosdetc2023.wordpress.com