Outdoor Wild Bird Detection based on YOLO algorithm

Bo-Cheng Zhu[1], and Tzung-Han Lin[1], Yao-Chuan Tsai[2], Kuang-Wen Hsieh[2], Fuh-Min Fan[2], Perng-Kwei Lei[2]

[1]National Taiwan University of Science and Technology, Taipei 10607, Taiwan
National Chung-Hsing University, Taichung 402, Taiwan
Keywords: Small Object, Object detection, Deep Learning

ABSTRACT

This study focuses on outdoor bird detection in video surveillance to reduce the risk of avian influenza (AI) in poultry farms. Once a bird is detected, our system will trigger another action to drive away the bird. We utilized YOLO algorithm for object detection, and add diversity samples to the dataset to train the model, then recognition accuracy is improved.

1. INTRODUCTION

In the recent years, object detection has been widely used in various fields. For instance, pedestrian detection can be used in autonomous vehicles to assist vehicles avoiding pedestrians on the road [1]; Flying objects detection can be used in drones, to avoid drones colliding in the air, or to avoid flying creatures [2]. As a result it improves the safety.

Because several regions in the world are suffering from severe bird flu disasters, it needs a solution to keep wild birds away from poultry. For this purpose, we try to use the object detection technology of YOLO for small wild bird identification. "You only look once" (YOLO) is one state of the art algorithm for objection. If the wild bird can be detected, we intend to combine one additional laser ray to scare it away. Therefore, it would dramatically reduce the risk of avian influenza.

In this paper, we proposed a practical method to improve the accuracy of small object detection. We refer to the copy-paste method proposed by Kisantal [3] to manually attach small objects where they appear. For example, if there is only one frisbee in the park, his method will make the frisbee appear in many regions on the image. However, we modify the method to randomly put several wild birds on image to simulate the natural situation of wild birds. In addition, we also refer to the concept of image flipping proposed by Laroca [4]. In his method, the region of interested is flipped either horizontally or vertically to increase the diversity of objects' poses in order to improve the recognition accuracy. In addition, with combination of these two methods, the accuracy can be significantly improved comparing to each.

Based on the fore-mentioned method, the accuracy of detection of our proposed method has been greatly improved. For example, increasing the diversity of dataset can effectively improve the recognition of wild birds in images.

2. EXPERIMENT

The flowchart of our implementation is shown as **Fig. 1**. In this section, we will introduce how to collect images, modify images, and why we use YOLO algorithm [5].

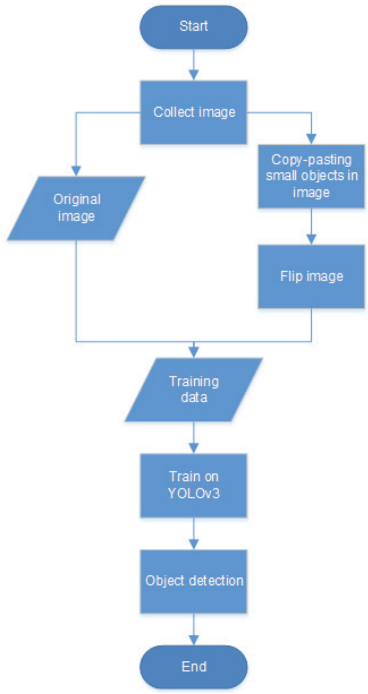


Fig.1 System flowchart

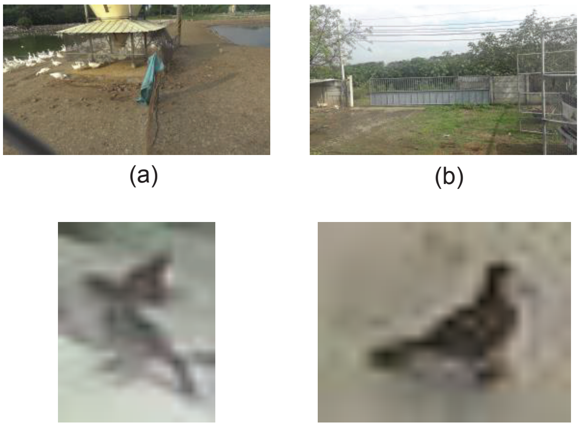
2.1 The dataset

Our dataset contains 900 images taken from different poultry farms. These images were taken with the same type of camera and they are stored in size of 1920 x 1080 pixels. The model of camera we used is SONY SNC-VM772. To cover a wide region in the poultry farm, the camera is allocated on the roof to shot from top-view. Each camera takes thousands of images. Among these images, we manually selected those images with wild birds as candidate images. **Fig. 2** shows different poultry farm and detected objects in the dataset. **Fig. 2** (a) and (b) represent different poultry farms, and **Fig. 2** (c) and (d) represent different kinds of wild birds.

In our experiments, the main purpose was to detect these birds, which are smaller than 32 x 32 pixels comparing to 1920 x 1080 pixels. These are defined as small objects in MS COCO [6], as shown in Table 1.

Table 1 Small, medium, large objects defined in MS COCO [6]

Object	Min rectangle area	Max rectangle area
Small	0 x 0	32 x 32
Medium	32 x 32	96 x 96
Large	96 x 96	∞ x ∞



(a) (b) (c) (d)

Fig. 2 Sample images of the dataset

Table 2 The architecture of Darknet-53

	Type	Filters	Size	Output
	Convolutional	32	3 × 3	256 × 256
	Convolutional	64	3 × 3 / 2	128 × 128
1×	Convolutional	32	1 × 1	
	Convolutional	64	3 × 3	
	Residual			128 × 128
	Convolutional	128	3 × 3 / 2	64 × 64
2×	Convolutional	64	1 × 1	
	Convolutional	128	3 × 3	
	Residual			64 × 64
	Convolutional	256	3 × 3 / 2	32 × 32
8×	Convolutional	128	1 × 1	
	Convolutional	256	3 × 3	
	Residual			32 × 32
	Convolutional	512	3 × 3 / 2	16 × 16
8×	Convolutional	256	1 × 1	
	Convolutional	512	3 × 3	
	Residual			16 × 16
	Convolutional	1024	3 × 3 / 2	8 × 8
4×	Convolutional	512	1 × 1	
	Convolutional	1024	3 × 3	
	Residual			8 × 8
	Avgpool		Global	
	Connected		1000	
	Softmax			

2.2 YOLOv3

In our experiment, we used the YOLO version 3 with Darknet-53 [5], and tuned it to train the model based on our dataset. Darknet-53 is a new feature extraction network for YOLO that combines darknet-19 and residual network. **Table 2** shows the architecture of Darknet-53. YOLO version 3 uses a similar feature pyramid network [7] method and adds more anchor boxes for multi-scale prediction, which improves the problem of small object detection by YOLO version 2 [8].

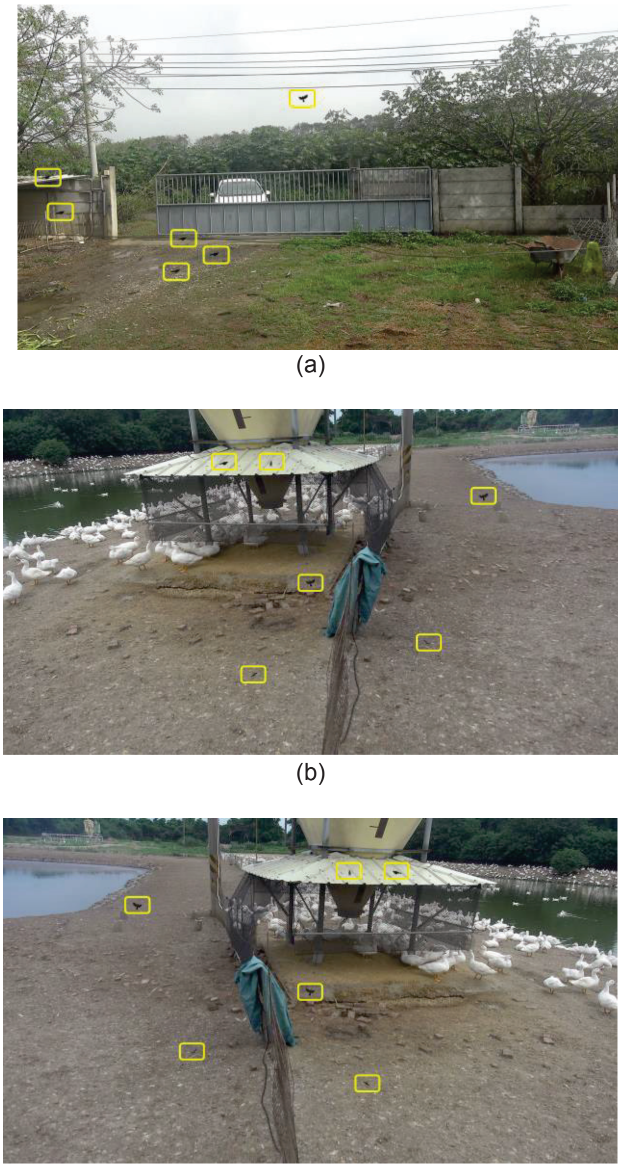
2.3 Dataset modification

In order to solve the problem of insufficient samples and diversity of small objects, we use copy-paste method to add small objects. First of all, we manually selected a small object from the image, cropped it out and remove background. Second, we selected other images with the same target but different backgrounds. Finally, we used the copy-paste method to paste the small objects into multiple and random locations on the image to generate a new image. After completing above procedure, we use flip-operation to modify the image, and then modify the copy-paste image to further generate a new sample image. **Fig. 3** shows an example. **Fig. 3** (a) and (b) show the result of using the copy-paste method multiple times at the same image by different small objects. **Fig. 3** (c) is the result of flipped (b). Yellow boxes represent small objects that are pasted on the image. The dataset split ratio is up to 80% for training, 10% for validation, 10% for testing.

3. RESULTS AND DISCUSSION

By slightly tuning YOLO, combining the two methods to generate images, and adding the original dataset to generate a new dataset, the training results are better than that with only original dataset.

In the case of detection only for small objects, it is recommended to lower the IOU threshold, because the ground truth of small objects is only 32 x 32 pixels or smaller, and a slight gap with the predicted bounding box will cause the identified object to become a false negative which will lead to decrease in accuracy. According to the experimental result of **Fig. 4**, it is suggested that the IOU threshold can be set to 0.25, which is the best value to detect only small wild birds.



(a)

(b)

(c)

Fig. 3 Examples of images in the dataset.

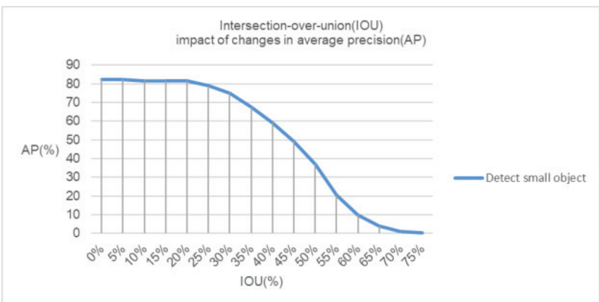


Fig. 4 Experimental result

4. CONCLUSION

In the case of original dataset, it is speculated that the reason for poor detection of small objects may be caused by a small sample of objects and insufficient diversity. We proposed a practical method to solve these problems by using copy-paste small objects or flipping images, or combining the two methods. This method can solve the problem of insufficient data and diversity, and effectively improve the accuracy of detecting small objects like wild birds. In future work, it may be possible to use different backgrounds combined with detection objects in the original dataset to improve data diversity.

ACKNOWLEDGEMENTS

The research project is supported by Council of Agriculture of Taiwan.

REFERENCES

[1] P. Dollar, C. Wojek, B. Schiele, and P. Perona, “Pedestrian Detection: An Evaluation of the State of the Art,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, No. 4, pp. 743–761 (2012).

[2] A. Rozantsev, V. Lepetit, and P. Fua, “Flying objects detection from a single moving camera,” *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, (2015).

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

[4] R. Laroca, E. Severo, L. A. Zanlorensi, L. S. Oliveira, G. R. Goncalves, W. R. Schwartz, and D. Menotti, “A Robust Real-Time Automatic License Plate Recognition Based on the YOLO Detector,” 2018 International Joint Conference on Neural Networks (IJCNN), (2018).

[5] J. Redmon and A. Farhadi, “Yolov3: An incremental improvement,” arXiv preprint arXiv:1804.02767, (2018).

[6] T. Y. Lin, M. Maire, S. Belongie, L Bourdev, R Girshick, J. Hays, P. Perona, D. Ramanan, C. L. Zitnick, P Dollar, “Microsoft coco: Common objects in context,” In European conference on computer vision, pp. 740-755 (2014, September). Springer, Cham.

[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] J. Redmon and A. Farhadi, “YOLO9000: better, faster, stronger,” In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 7263-7271 (2017).