Detection and Tracking of Flying Small Bats under Complex Backgrounds

Kakeru Sugimoto[1], Kazusa Usio[2], Ryota Sugimori[2], Emyo Fujioka[3], Hiroaki Kawashima[4], Shizuko Hiryu[5], and Hitoshi Habe[6,7]

[1] Graduate School of Science and Engineering, Kindai University, Japan
[2] Graduate School of Life and Medical Sciences, Doshisa University, Japan
[3] Organization for Research Initiatives and Development, Doshisha University, Japan
[4] School of Social Information Science, University of Hyogo, Jpapan
[5] Faculty of Life and Medical Sciences, Doshisha University, Japan
[6] Department of Informatics, Faculty of Informatics, Kindai University, Japan
[7] Cyber Informatics Research Institute, Kindai University, Japan

Abstract. Bats emit ultrasonic waves during their flight and listen to the echoes to understand their surroundings. To understand the unique ecology of bats, various efforts have been made. Among them, computer-aided automatic detection and tracking would play an important role. This enables us to analyze the movement precisely. Bats are nocturnal, small in size, and move at high speeds, which are pretty unfavorable conditions for detection and tracking. However, the state of the art of multiple object tracking (MOT) methods yields better performance in object detection and tracking. In this paper, we use YOLOv7, a new version of YOLO, for object detection and OC-SORT, a kind of MOT method, for object tracking and compare the accuracy of each method with other existing methods. The video images used in this study have complex backgrounds, and the accuracy of detection could be low because bats are assimilated into the background. Additionally, the shadows of bats are miss-detected as bats. To cope with such difficult situations, we first calculate the inter-frame differences to extract moving objects clearly and then detect the shadows of bats as another object class to avoid the miss-detection of bat shadows as bats. We finally compare the difference in performance with the actual video of flying bats.

Keywords: Object Detection· Moving Object Tracking · YOLO · SORT · Small Object · Complex Background

1 Introduction

Bats are the only mammals that can fly. Bats, which are not very visible, emit ultrasonic waves themselves and listen to their echoes to understand their surroundings and targets. This feature is called echolocation, and efforts are underway to elucidate its mechanism[1].

However, because bats are nocturnal, the images are often dark. Additionally, the background is often complex, and because bats are small and move at high speed, they may blend in with the background or become blurred. Therefore, it is difficult and time-consuming for the human eyes to distinguish between bat images.

In addition, when capturing video outside, we need a light source to lit bats under dark conditions. This often results in the shadow of the bat, making it difficult to distinguish it from the actual bat. The sample of the actual video used in this study is shown in Fig. 1.

The goal of this study is to improve the accuracy of bat detection and tracking. Our method is based on YOLOv7[2] developed by Chien-Yao Wang et al. for detection and OC-SORAT[3] developed by Jinkun Cao et al. for tracking. To achieve the goal, we first extract bat regions under complex backgrounds using inter-frame differences. Then, we apply two-class detection, not only detecting actual bats but also their shadows as another class of objects. This enables us to reduce the false detection of shadows as bats. We think that these methods can be applied to any data if they can correctly detect and track complex images that are small, dark, and fast, such as the images used in this study.

(a) Original Video (b) Bats and Shadow

Fig. 1. Example of Actual Video

2 Related Work

2.1 Measurement with Microphone Arrays

Fujioka et al. [4] simultaneously measured the flight routes of wild bats and the direction of ultrasonic pulse emission using microphone arrays to elucidate the mechanism of echolocation, but the microphone arrays have limitations in their observable environment, such as ultrasonic interference when multiple bats are flying and limited space for observation.

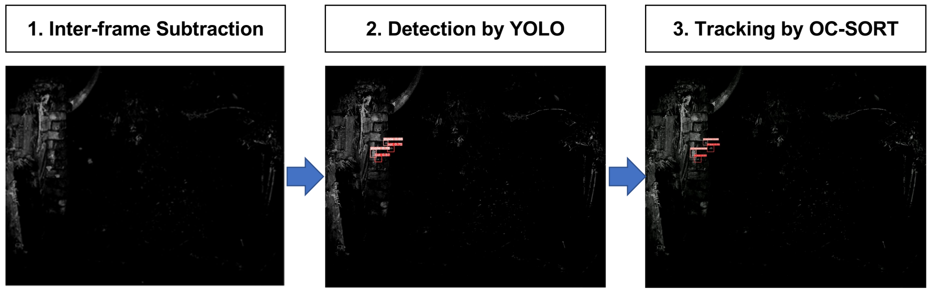

Fig. 2. Overview of the proposed method

2.2 Object detection by Background Subtraction

Background subtraction[5–7] can be used for separating the background and foreground portions. Additionally, background subtraction[8][9][10] using a mixed Gaussian model can incorporate complex background and adaptively updates the model.

In our previous study, we extracted and detected moving objects by background subtraction for bat videos and then tracked them using nearest neighbor search and motion model[11]. However, this method was not sufficiently accurate in some areas and required tuning for each application environment.

3 Proposed Method

This section describes our proposed method.

3.1 Overview

As described earlier, our method is based on YOLO and OC-SORT. In addition to those methods, we apply inter-frame subtraction to extract small moving objects, and we perform two-class detection: actual bats and their shadows. This would decrease the false detection of shadows as actual bats. The flow of this method is as follows and shown in Fig. 2:

1. Inter-frame subtraction
2. Two-class bat detection by YOLO
3. Bat tracking and ID association by OC-SORT

Each process will be described in the subsequent sections.

3.2 Inter-frame Subtraction

Inter-frame subtraction is used to extract the position of flying bats, to improve the detection accuracy of bats. More specifically, at time t, the difference between the images at t and $t-1$ and the difference between the images at $t+1$ and t are computed from three consecutive images. Then, the disjunction of the two subtracted images is calculated to obtain an image in which the moving object is emphasized. Since the resulting image generates salt and pepper noise, the salt and pepper noise is removed by shrinking and expanding the image. Finally, the brightness of the original image is increased by 30 at pixels where moving objects are detected and decreased by 100 at other pixels. This image is used in the subsequent detection process.

3.3 Two-class Bat Detection by YOLO

YOLOv7 is a detection method in YOLO, developed in 2022 by the same group as YOLOv4, Scaled-YOLOv4, and YOLOR. This method is reported to outperform existing methods on the MS COCO dataset significantly[2]. We expect it also yields good results in this study. We also use YOLOv5 for detecting bats. YOLOv5 is an older method but is widely used for detection.

As mentioned earlier, we perform two-class detection: actual bats and their shadows. As shown in Fig. 1(b), flying bats and their shadows are quite similar in their appearance. If we train the object detection model so that only the bat region will be detected, may false detection and miss detection would happen. To cope with this issue, we train the model to detect the actual bats and shadows as different classes. This makes the detection model able to distinguish between the two objects and suppress false detections.

3.4 Bat tracking and ID association by OC-SORT

Observation-Centric SORT (OC-SORT), created by Cao et al. in 2022, is an object tracking method that is robust to occlusion and nonlinear motion while maintaining the simple, real-time, online approach characteristic of traditional SORT [3]. OC-SORT solves the problems of traditional SORT[12] and achieves state-of-the-art performance in the latest MOT benchmarks. OC-SORT created a robust model for occlusion and nonlinear motion with the following three techniques: Observation-centric Online Smoothing, Observation-Centric Momentum, and Observation-Centric Recovery.

4 Experiments

We conduct three kinds of experiments to evaluate and compare each component of the proposed method. The details and results of each experiment will be described in the followings.

The training data in all experiments contains 300 images of 10 seconds at the same scene. The number of epochs in training was 300.

4.1 Exp1: Inter-Frame Subtraction

In the first experiment, we examine the effect of inter-frame subtraction. To this end, we compare the accuracy of bat detection with inter-frame subtraction and without it. The object detection method is YOLOv5[13].

Fig.3 shows an example of bat detection results by YOLOv5, with and without inter-frame subtraction. Table 1 summarizes the evaluation results. As shown in Table 1, the inter-frame difference improved both recall and mAP by about 16 points. However, the recall value is still low at 56%. The main reason for this is that the value of the inter-frame difference was smaller for bats with small movements and bats moving in the depth direction.

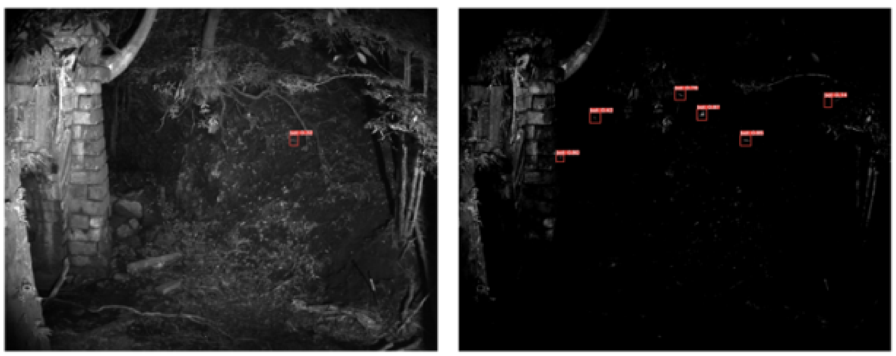

Fig. 3. Exp1: Detection results

Table 1. Exp1: Performance comparison of inter-frame subtraction(YOLOv5)

	Precision	Recall	mAP50
w/o inter-frame subtraction	88.0%	39.9%	45.5%
w/ inter-frame subtraction	83.1%	**55.9%**	**61.4%**

4.2 Exp2: Two-Class Bat Detection

In the second experiment, we see the effect of the two-class bat detection, i.e., bats and shadows are treated as different classes. We will compare the performance of bat detection using the two-class setting and the standard single-class setting, which aim to detect bats only.

Also, we compare two object detection models: YOLOv5 and YOLOv7. Although YOLOv7 is newer model than YOLOv5, it is worth evaluating which is better for our application domain.

Table 2 summarizes the results. The results show that the two-class bat detection strategy is effective, as both models give higher performance. Especially for YOLOv7, all evaluations were more than 10 points higher than in the one-class setting. However, this may not be true for all results since only the recall value dropped in YOLOv5.

Also, when comparing the two detection models: YOLOv5 and YOLOv7, contrary to expectations, YOLOv5 gives better results. This implies that YOLOv7 shows better results for standard data, such as people, dogs, and cars, but not for all data.

Table 2. Exp2: Effect of two-class bat detection

Model	Detection Method	Precision	Recall	mAP50
YOLOv5	One-Class	83.1%	55.9%	61.4%
	Two-Class	**83.8%**	55.5%	**65.6%**
YOLOv7	One-Class	55.8%	47.1%	45.1%
	Two-Class	79.1%	**57.0%**	61.4%

4.3 Exp3: Bat Tracking

In the third experiment, we will evaluate the results of tracking bats using OC-SORT. For OC-SORT we use the YOLOv5 model which gives the best detection performance among all the detection results we have tested. The comparison is made between the ByteTrack[14] and OC-SORT.

The observations from the experimental results are (1) OC-SORT reduced the number of false track detections of overlapping bats, and is more accurate than ByteTrack, (2) OC-SORT is highly dependent on detection same as ByteTrack. Hence, if there is a missing detection in the middle of a trajectory, it may be regarded as a different individual, or terminate the tracking. One example of the individuals tracked by OC-SORT is shown in Fig.4. The coordinate axes of the graph are aligned with the actual image size.

5 Conclusions

To detect and track flying bats under complex backgrounds, we apply YOLO and OC-SORT. In addition to the standard method, we perform the inter-frame subtraction to extract flying bats even when the bats are small and fast and the backgrounds are complex. Also, we conduct two-class bat detection: bats

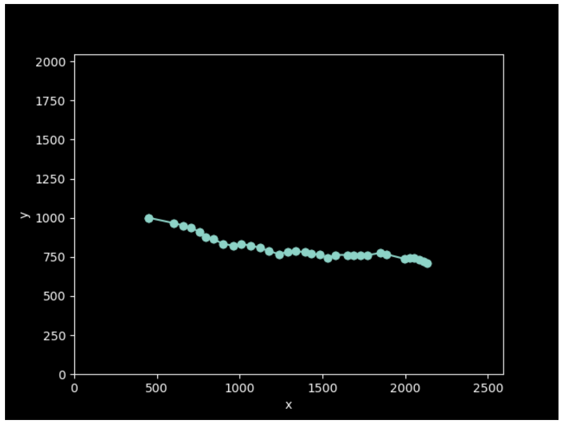

Fig. 4. An example of tracking results

and their shadows are differently detected. Experimental results demonstrate the proposed framework works well for the actual video containing flying bats.

Future work includes improving the detection so that bats with small movements. At the tracking stage, it is necessary to cope with missing detection in the middle of a trajectory.

This work was partly supported by JSPS KAKENHI JP21H05302.

References

1. K.Hase et al., Bats enhance their call identities to solve the cocktail party problem, Communications Biology, volume1, Article number: 39 (2018)
2. Chien-Yao Wang, Alexey Bochkovskiy, Hong-Yuan Mark Liao "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors" arXiv:2207.02696 (2022).
3. Cao, Jinkun, Weng, Xinshuo, Khirodkar, Rawal, Pang, Jiangmiao, Kitani, Kris. "Observation-centric sort: Rethinking sort for robust multi-object tracking" arXiv preprint arXiv:2203.14360 (2022).
4. E. Fujioka, I. Aihara, M. Sumiya, K. Aihara, and S. Hiryu. Echolocating bats use future-target information for optimal foraging. In PNAS, 4 (2016).
5. A.Elgammal, D.Harwood, and L.S.Davis, "Non-parametric background model for background subtraction" In Proceedings6th ECCV, (2000).
6. C.Stauffer, W.E.L.Grimson "Adaptive background mixture models for real-time tracking" Proceedings 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, (1999).
7. O.Barnich, M.V.Droogenbroeck "ViBe: A Universal Background Subtraction Algorithm for Video Sequences" IEEE Transactions on Image Processing, Vol. 20, Issue .6, June (2011)
8. P. KaewTraKulPong, R. Bowden "An Improved Adaptive Background Mixture Model for Realtime Tracking with Shadow Detection" In: P.RemagninoGraeme, A.JonesNikos, ParagiosCarlo S. Regazzoni(eds.) Video-Based Surveillance Systems, pp.135-144

9. Z.Zivkovic "Improved adaptive Gaussian mixture model for background subtraction", Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004.
10. Z.Zivkovic, F.Heijden "Efficient adaptive density estimation per image pixel for the task of background subtraction" In: T.K.Ho, Murray Hill, G. Sanniti et.al, Pattern Recognition, Letters Volume 27, Issue 7, May 2006, Pages.773-780
11. E. Fujioka, M. Fukushiro, K.Ushio, K. Kohyama, H. Habe, S. Hiryu, "Three-Dimensional Trajectory Construction and Observation of Group Behavior of Wild Bats during Cave Emergence." Journal of Robotics and Mechatronics 33 (3): pp. 556–63, 2021.
12. Bewley, Alex, Ge, Zongyuan, Ott, Lionel, Ramos, Fabio, Upcroft, Ben. "simple online and realtime tracking". In 2016 IEEE International Conference on Image Processing (ICIP), pp. 3464–3468, (2016).
13. Ultralytics YOLOv5: https://github.com/ultralytics/yolov5. Last accessed 30 Nov 2022
14. Yifu Zhang, Peize Sun, Yi Jiang, Dongdong Yu, Fucheng Weng, Zehuan Yuan, Ping Luo, Wenyu Liu, Xinggang Wang. "ByteTrack: Multi-Object Tracking by Associating Every Detection Box" arXiv 2110.06864 (2022)