International Journal of Computer Vision and Signal Processing, 15(1), 74-81(2025) ORIGINAL ARTICLE

A Tracking Method of Multiple Animals Using YOLOv5

Toshifumi Kimura

School of Human Science and Environment, University of Hyogo, Japan

Hidetoshi Ikeno

Faculty of Informatics, The University of Fukuchiyama, Japan

Mizue Ohashi

School of Human Science and Environment, University of Hyogo, Japan

Ryuichi Okada

Department of Biology, Graduate School of Science, Kobe University, Japan

Mamiko Ozaki

Department of Chemical Science and Engineering, Kobe University, Japan. KYOUSEI Science Center for life and Nature, Nara Women's University, Japan

Hiroyuki Ai

Faculty of Science, Department of Earth System Science, Fukuoka University, Japan

Shunya Habe

Semco Co., Ltd., Japan

Teijiro Isokawa

Graduate School of Engineering, University of Hyogo, Japan

Abstract

In behavioral experiments on social animals in ethology, it is important to understand not only the detailed location of each of the target animals, but also their swarm behavior emerging from their interactions. Recently, many systems have been developed to support animal behavior analysis. However, these systems require videos under good conditions for easy discrimination of targets from background, leading to their limited applications. In this paper, a tracking system that is robust to different experimental conditions is proposed. The proposed system adopts YOLOv5, a deep neural network based system, as an object detector from video images and incorporates the existing K-Track system for tracking the detected objects. The performance of the proposed system is evaluated using actual videos obtained from behavioral experiments, and robust detection of target animals and their tracking is possible.

Contribution of the Paper: Accurate animal tracking is achieved through the combination of object detection and bi-directional associations for detected objects.

Keywords: behavioral experiments, object detection, tracking objects, deep neural networks

IJCVSP
International Journal of Computer Vision and Signal Processing

ISSN: 2186-1390 (Online)
http://cennser.org/IJCVSP

Article History:
Received: 10/4/2025
Revised: 14/7/2025
Accepted: 24/11/2025
Published Online: 25/11/2025

1. INTRODUCTION

Social animals have been used in many ethological studies to understand the mechanisms of their behavior. It is necessary to analyze not only the behavior of each individual but the behavior of individuals with their interactions. Among many kinds of social animals, bees and ants are popular animals (insects). This is because they can be easily and quickly bred and various levels of analysis can be performed, such as measuring neural activity in their brains and secretion of chemical substances, and observing their behavior under various environmental conditions. Our interests are on the analysis of how these insects interact with each other using image sequences taken by digital cameras. The use of images can reduce the influence on the behavior of target insects.

Recently, many types of digital video cameras with high performance and low cost have become available. Thus many researchers have used them to record the behaviors of animals in their experiments; these cameras can record the behaviors of long duration with high resolution as a video. On the other hand, analyzing the behaviors from the video, such as extracting and tracking individuals from a set of images, is still done manually by the researchers themselves. In addition to being time-consuming and laborious, manual operations are prone to errors, such as false-positive detection of individuals and undetected individuals from the image frames. Therefore, the development of automatic or semi-automatic analysis methods from recorded videos is an important challenge.

Several computer programs for automatic tracking and analysis of animals have been presented and are available to researchers. *id-Tracker* [1] is a software for mice and insects (vinegar flies and ants) where several individuals can be tracked at the same time. The tracking can be done by using the texture of a target individual's back as a cue to identify each of the individuals in the image frame. Using texture is advantageous for achieving non-invasive to target individuals, such as putting marks on individuals, but images with high resolution are required to discriminate textures among individuals, leading to high computational cost for image processing and large capacity storage for images. It is also necessary to cope with texture changes against light condition due to the movements of individuals.

Branson et al. have developed an automatic tracking software called *Ctrax* [2], where the fruit fly, *Drosophila melanogaster*, as a target animal, but it is applicable to other animals such as cockroaches and mice. This software is also capable of identifying and tracking individuals without the use of tags. The main goal of this software is to identify behaviors for target individuals from the short duration of their trajectories, so the locations of individuals in the image are not always accurate. Feldman et al. have developed a scheme for analyzing the behavior of honeybees [3]. This scheme is intended to extract certain behaviors of honeybees on their nest with non-dense condition, such as waggle dance of honeybees, so it is not good at various types of behaviors.

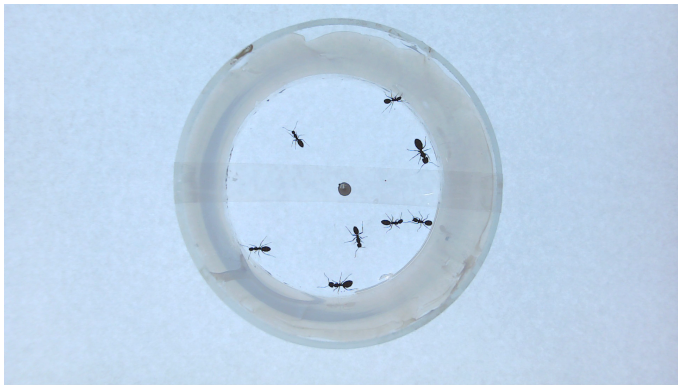
Kimura et al. have developed two types of software, called *K-Track* and *K-Track-kai* [4, 5, 6], for automatic tracking of multiple honeybees on their nest from low resolution video frames. These software achieves simultaneous extraction for the locations, velocities, accelerations of target individuals, as well as the waking distance from a given position and changes in distances between individuals. This turns out to be applicable to track other types of animals, such as ants, antmimicking spiders, spiders and praying mantis under the flat area. Although various types of videos can be used by these software, they tend to fail to track target individuals walking at the corner and boundary in an experiment arena.

All of the software mentioned above requires image frames from which targets and background can be easily distinguished. This can be done by preparing a configuration for experiments with a uniform lighting condition and a uniform background. There are still many experimental conditions and videos obtained by experiments with lower illumination and complex backgrounds to avoid the behavior of individuals from their environment. Therefore, it is necessary to develop a tracking method that is robust to the environmental conditions around the target individuals.

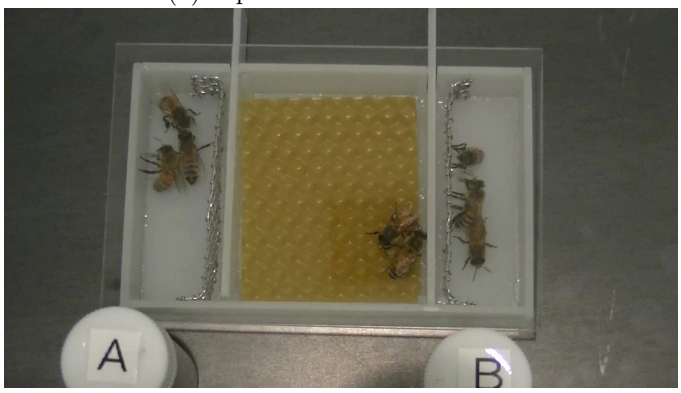
In this paper, we propose an automatic tracking method for social animals and evaluate this method using video images obtained by experiments. Our method is based on *K-Track-kai* for tracking target individuals [6] and uses YOLOv5 [7], a version of YOLO systems [8] consisting of deep neural networks, as an object detector from video images. We chose YOLOv5 as the object detector because it is easy to install on various systems (Windows/macOS), and it provides well-developed tools for creating training samples for target objects. The original *K-Track-kai* adopts the so-called background subtraction method to distinguish target individuals from the background. However, this approach is not effective in extracting individuals under low-illumination conditions. YOLOv5, on the other hand, can detect target individuals in various environments by using a well-prepared set of training samples, but it does not include functionality for tracking the detected individuals. Therefore, combining these two methods is expected to result in a more robust individual tracking system. It is shown that our method can extract the target individuals under non-uniform environments and can track them automatically.

This paper is organized as follows. In section 2, video images in our experiments and the our method are described. Experimental results and their discussions are shown respectively in sections 3 and 4. Section 5 finishes this paper with conclusion.

Email address: kimura@shse.u-hyogo.ac.jp (Toshifumi Kimura)



(a) experimental field for ants



(b) experimental field for honeybees

Figure 1: Experimental configurations for analyzing animals' behaviors.

2. MATERIALS AND METHODS

In this section, examples of video images for the target insects are shown, then our presented method is described. Our method consists of two stages of processing, i.e., objects (animals) detection from the video image and object tracking by connecting objects between successive images.

2.1. Target animals and experimental conditions

Two types of animals and their experimental conditions are used to evaluate the proposed method. An experimental configuration of ants is shown in Fig. 1(a). There are several ants that can walk around in a round area of a glass dish (12cm diameter). In this experiment, sufficient illumination is obtained in the whole of arena to extract each of the ants can be easily extracted [9]. Figure 1(b) shows a configuration of an experiment using honeybees. In this experiment, there are three compartments in the experimental area; honeybees in each compartment can move around in their compartment but it is not possible to go through other compartments [10]. A piece of honeybee nest is placed in the middle compartment, so the background of the area is not uniform but has a complex structure. For this experiment, low illumination is set; this is intended not to affect the behavior of honeybees, and this configuration is common in honeybee nests.

2.2. Detecting objects from images

In *K-Track* and its improved *K-Track-kai* systems, walking animals are first extracted. This is done using a method called background subtraction, where the moving individuals are extracted by subtracting successive frames with a threshold. This is effective if the conditions of background (non-animals) are adequately configured, e.g. uniform area and sufficient illumination, in other words, this is not robust if these conditions cannot be prepared.

In order to extract target animals under different environmental conditions, we adopt YOLOv5 [7], which is a deep neural network-based system, as an object detector. A set of detected objects can be extracted from each image frame in a video. Each detected object is represented by a two-dimensional coordinate of bounding box around the object with its confidence (probability) for detection.

YOLO (or YOLOv5) can perform detection and identification (classification) tasks simultaneously using a single neural network. The system consists of three components, called Backbone, Neck, and Head. The Backbone extracts image features from the input image using a convolutional neural network, then these features are integrated in the Neck component, and the Head makes bounding boxes of the detected objects from the integrated features as well as classifications for each of the image regions with their bounding boxes. Several pre-trained models (sets of connection weights in the neural network) that can detect and identify various objects are attached to the distribution of YOLOv5, and popular objects such as person and car can be detected by using these models. Target animals used in the ethological experiments, such as honeybees and ants, are hardly detected by these models, so additional training is required for YOLOv5. To train the network, sets of image regions for target animals are prepared with their annotations. To create annotations for image regions, the software LabelImg [11] is used.

2.3. Tracking objects over image frames

YOLOv5 does not deal with temporal contexts for target objects; this system detects objects for each frame in a video but does not perform binding identical objects across successive frames. In this section, we present an object binding scheme for tracking target animals in video. The proposed scheme requires a set of image frames for the target insects before its processing; real-time tracking is not for our purpose.

Our scheme uses the following information obtained from YOLOv5, denoted by:

- $O_i(t)$ as i-th object detected in the image frame at time t,
- $N(t)$ as the number of objects detected in the image frame at time t,
- $x_i^{LU}(t), y_i^{LU}(t)$ as the leftmost and uppermost coordinate of bounding box on $O_i(t)$,

- $x_i^{RD}(t), y_i^{RD}(t)$ as the rightmost and lowermost coordinate of bounding box on $O_i(t)$,
- $x_i^C(t), y_i^C(t)$ as the center coordinate of bounding box on $O_i(t)$.

Several definitions are provided using these information. Euclidean distance between the objects $O_i(t)$ and $O_j(t)$, $d\left(O_i(t), O_j(t)\right)$ as

$$d\left(O_i(t), O_j(t)\right) = \sqrt{\left(x_i^C(t) - x_j^C(t)\right)^2 + \left(y_i^C(t) - y_j^C(t)\right)^2}. \quad (1)$$

Overlap between two objects is defined by whether the bounding boxes of two objects intersect. Overlap for two objects in the different time frame $O_i(t)$ and $O_j(t')$, denoted by $p(O_i(t), O_j(t'))$, is defined as a Boolean value: $p(O_i(t), O_j(t')) = 1$ if $x_j^{LU}(t') - x_i^{LU}(t) < x_i^{RD}(t) - x_i^{LU}(t)$ and $y_i^{LU}(t) - y_j^{LU}(t') < y_i^{LU}(t) - x_i^{RD}(t)$ hold, otherwise $p(O_i(t), O_j(t')) = 0$, where $x_j^{LU}(t') > x_i^{LU}(t)$ and $y_i^{LU}(t) > y_j^{LU}(t')$ are assumed. A relation between objects is defined by the symbol $\rightarrow$. $O_i(t) \rightarrow O_j(t+1)$ defines that a relation is created from $O_i(t)$ to $O_j(t+1)$.

Then the proposed scheme for binding objects over image frames is presented. This scheme consists of the following steps:

1. (forward search) Make connections from the objects in the image at time t to the objects in the image at time $(t+1)$ according to the minimum distance between objects with their overlap: for $i = 1, \cdots, N(t)$,

$$j = \underset{N(t+1)}{k=1,\cdots,} \frac{1}{d(O_i(t), O_k(t+1))} \cdot p(O_i(t), O_k(t+1)),$$
$$O_i(t) \rightarrow O_j(t+1). \quad (2)$$

   This process is applied from the image at frame $t = 1$ to the frame at time $t = (T-1)$ where T is the number of frames in a video.

2. (backward search) Make connections from the objects in the image at time $(t+1)$ to the objects in the image at time t according to the minimum distance between objects with their overlap: for $i = 1, \cdots, N(t+1)$,

$$j = \underset{N(t)}{k=1,\cdots,} \frac{1}{d(O_i(t+1), O_k(t))} \cdot p(O_i(t+1), O_k(t)),$$
$$O_i(t+1) \rightarrow O_j(t). \quad (3)$$

   This process is applied from the image at the frame $t = (T-1)$ to the frame at the time $t = 1$.

3. (binding objects over frames) The connection relations between the image frames at the time t and $(t+1)$ are examined. If $O_i(t) \rightarrow O_j(t+1)$ and $O_j(t+1) \rightarrow O_i(t)$ holds for pairs of i and j, then it is assumed that the object $O_i(t)$ is the the same object as $O_j(t+1)$, and the connection between these objects is fixed. This process is performed for all detected objects in all frames.

4. (binding by selecting connections from candidates) Remaining connections after the previous step are mainly caused by the number of detected objects differing between image frames, due to undetected objects from the image frame. First, for each object, connections to this object between two consecutive images are removed if it already had its fixed connection between these frames.

5. (removing redundant connections) Then, for each of objects, a connection from/to this object between two consecutive images is selected where the distance of this connection is minimum without considering the overlap of two objects, and the connections other than the selected connection are removed.

6. (creating trajectories for each object) Final step is to collect the connections between the image frames to create each trajectory for the detected objects. A trajectory of an object contains a time series of its center coordinates detected in the image frames.

A schematic example of making connections across image frames is shown in Fig. 2, where six objects are detected in the frames at the time $(t-1)$ and the time $(t+1)$ and five objects are detected at the time t. The connections from the steps 1 and 2 are shown in Fig. 2(a). In this figure, the created connections are represented by dotted lines with arrows, where the directions of the arrows, i.e. down and up, represent the created connections by forward and backward search, respectively. Objects O_1, O_2, and O_3 can be considered identical over time $(t-1)$, t, and $(t+1)$ because the forward and backward connections are the same. These connections are fixed by the step 3. On the other hand, the objects O_4, O_5, and O_6 have inconsistent links between forward and backward searches. These connections are selected and removed by steps 4 and 5. The final configuration of the connections between the objects is shown in Fig. 2(b).

3. EXPERIMENTAL RESULTS

The proposed method is evaluated through tracking target individuals from videos obtained in ethological experiments. Two types of videos are used to evaluate the detection performance. One video, called video-A, contains eight ants in a flat arena (Fig. 1(a)), where the frame rate per second in the video is 30, the number of image frames is 9560, and each frame has 1920 pixels width and 1080 pixels height. The other video, called video-B, contains eight honeybees in the compartments (Fig. 1(b)), where the frame rate per second in the video is 30, the number of image frames is 5085, and each frame has 1920 pixels width and 1080 pixels height. To evaluate the tracking

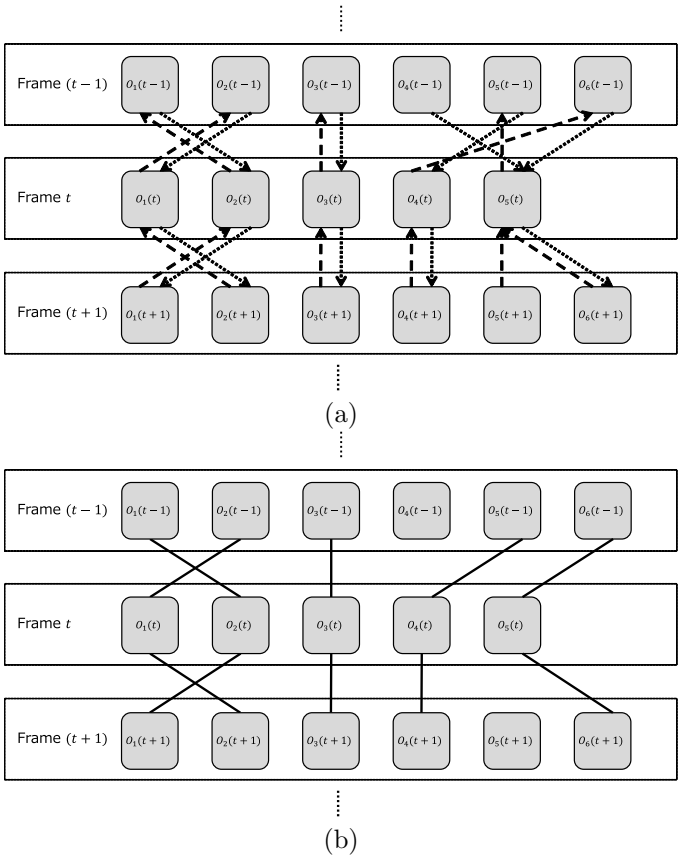


Figure 2: Connecting objects between successive image frames. (a) Connecting detected objects by forward and backward directions in image sequence. First, for the frames at $\cdots(t-1), t, (t+1), \cdots$, detected objects between frames are associated by step 1 in the proposed algorithm. The associated connections are denoted by the dotted lines between objects. Then, for the frames at $\cdots, (t+1), t, (t-1)\cdots$, detected objects are associated by step 2 in the proposed algorithm. The connections are denoted by the broken lines. (b) Connections are fixed for the objects if consistent connections are present between frames (e.g. $O_3(t-1)$ and $O_3(t)$), and connections are chosen from candidates if non-consistent connections occur (e.g. $O_5(t)$ and $O_5(t+1)$, $O_6(t+1)$). Resultant connections are denoted by the solid lines.

performance, video-A is used as the input for the proposed method.

3.1. Experimental setup

In order to detect target animals from videos, it is necessary to train YOLOv5 with some image frames with their annotations. In this experiment, we randomly selected 20 frames out of 9560 frames from video-A, and also 20 frames out of 5085 frames from video-B. We then annotated the locations of the targets in these frames. The total number of target animals with their annotations is 320 (two sets of 20 image frames with eight targets in each frame). These images were used to train YOLOv5 with the pre-training model yolov5s as the base model. The number of training epoch was 1000. We used 0.5 of confidence threshold is used for detecting objects from the image frame by the trained YOLOv5.

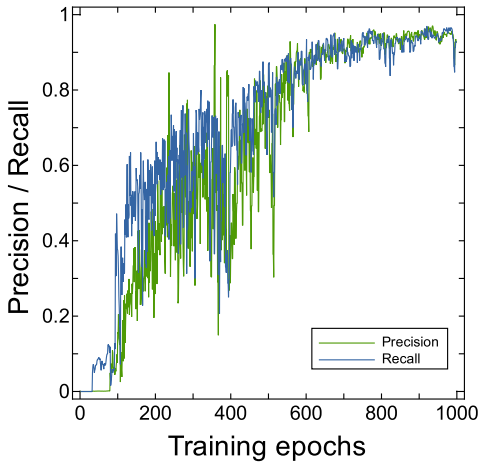


Figure 3: Changes of precision and recall on detection in training YOLOv5 with a set of training images (the number of epochs for training is 1000).

.

3.2. Detection by YOLOv5

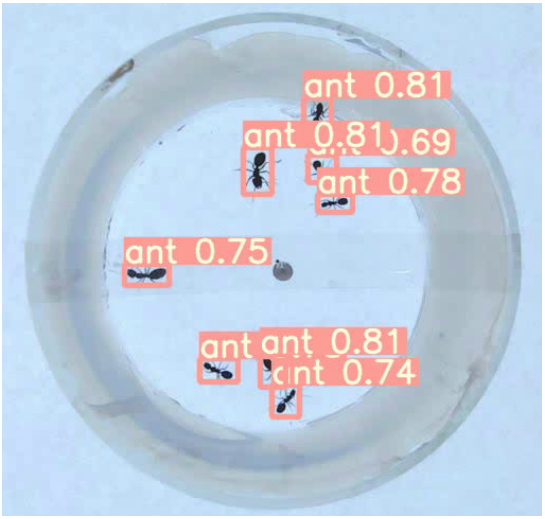
First, we show the changes of precision and recall values for object detection during the training YOLOv5, in Fig. 3. These values are calculated from the training dataset (not from the validation dataset). It is shown that the training in YOLOv5 is successfully conducted.

Detection was performed for all frames in video-A and video-B. Examples of detection in these video are shown in Fig. 4. Correct detection for the targets in the images could be performed in these images. The detection success rates for video-A and video-B were calculated as 95.4% (72944 of all 76480 individuals) for the video-A and 94.8% (38572 of all 40680 individuals) for the video-B. Detection was performed on another video, which was not the video-A but had similar experimental conditions, and in this case, 96.8% (73827 of all 76264 individuals) of detection success rate was achieved. Despite the very limited number of samples are used for training (1.67% in the video-A and 3.15% in the video-B) and the presence of non-uniform background in the images (especially in video-B), high detection accuracy could be achieved by using YOLOv5.

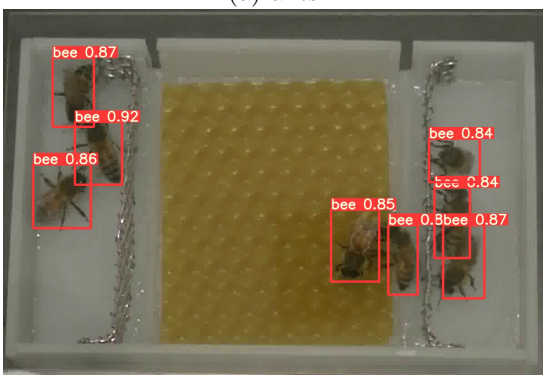
3.3. Tracking target individuals

In this section, we focused on the tracking of target individuals by the proposed scheme from video-A. Figure 5 shows the trajectories of eight ants for the first 900 frames in video-A, where the red and blue lines represent the trajectories of ants by the proposed scheme and those by a researcher, respectively. The trajectories by manual tracking are obtained from the reference [9], where manual tracking is efficiently performed with the help of *K-Track-kai*. From this figure, the trajectories by the proposed scheme are quite similar to the manually tracked ones.

We then present an example of individual tracking by comparing our proposed method with YOLO11 [12], a state-of-the-art object tracking algorithm. Figure 6 illustrates an



(a) ants



(b) honeybees

Figure 4: Examples of the detected objects by YOLOv5. Region of each detected object is displayed by its bounding box attached around it, the word and number at the upper-left of the bounding box shows the name of the detected object and confidence for detection of the object, respectively. In each case of experimental conditions (a) and (b), each individual is successfully detected.

example of ant tracking, where two ants collide and then move apart. The region containing these two ants is highlighted in Fig. 6(a), where the ants are approaching each other near the edge of the arena. The tracking results using YOLO11 are shown in Fig. 6(b) at frames 1855, 1869, and 1885. As shown in this figure, a different identification number (ID #109) is assigned to the ant originally labeled as ID #4 at frame 1885. This indicates that YOLO11 failed to maintain the original ID after the ant collided with another one (ID #6), resulting in a reassignment of a new ID. In contrast, our proposed method consistently maintains the correct ID assignments for both ants throughout the collision, as demonstrated in Fig. 6(c).

To quantitatively evaluate the tracking performance, we measured the Euclidean distance between the center point of the bounding box for the target detected by the proposed scheme and the point of the detected by manually tracking for each of the image frames. Figure 7 shows the errors averaged over the number of frames (measured in

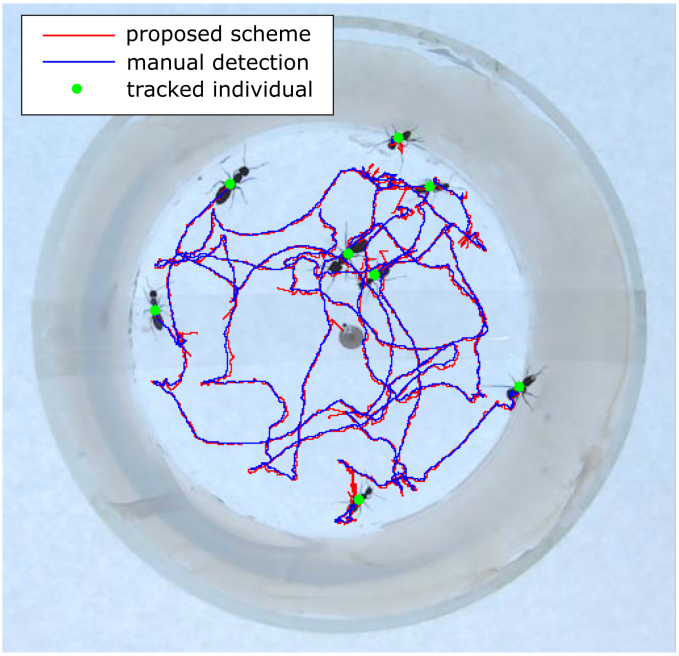


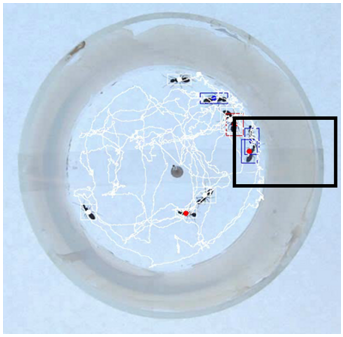
Figure 5: An example of tracking ants by the proposed scheme and by manual tracking as baseline. Green points indicate the tracked individuals, and red and blue lines represent trajectories of ants by the proposed scheme and manual tracking, respectively. Comparing the tracked trajectories with respect to the trajectories by the manual tracking, the proposed scheme can obtain almost same trajectories by the manually tracked ones.

pixels) for the eight target individuals as a boxplot. From this figure, it can be seen that most of detection can be made with a maximum error of 8 pixels, which corresponds to 0.155 cm in the real world (0.019 cm per pixel). Adequate tracking could be performed by the proposed scheme from this result. There are 26 events for failure of binding of targets on tracking by the proposed scheme, due to undetected targets from the image frames.

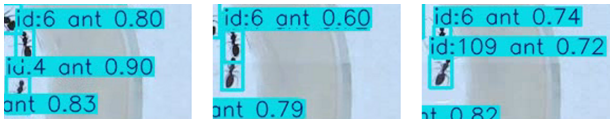
4. DISCUSSION

The proposed tracking scheme relies on detection objects in the image frames, so it is applicable to various experimental configurations where the backgrounds against the objects in the images have complex texture or low illumination. On the other hand, false detections or undetected objects could have adverse effects on the tracking.

The proposed scheme in the current version has no procedures for binding objects without their connections, e.g. $O_4(t-1)$ and $O_5(t+1)$ in Fig. 2(b). Thus, misidentification of trajectories can occur if these unbounded objects are present by a target animal being getting on the top of another animal. This can at least be detected by checking the number of detected objects per frame; usually the number of animals is not changed in laboratory experiments. The occurrence of undetected objects is mainly due to the untrained image patterns for the targets, so it is useful



(a) Snapshot at frame 1885



frame 1855 frame 1869 frame 1885

(b) Tracking by YOLOv11 object tracker

frame 1855 frame 1869 frame 1885

(c) Tracking by the proposed scheme

Figure 6: Examples of tracking results by YOLO11 [12] object tracker and by the proposed scheme, under a situation in collision of two ants. (a) Two ants make collision at the region shown by the black rectangle, (b) tracking by YOLO11 fails after the collision, i.e., new identification number 109 is assigned to the ant with identification number 4 at frame 1855, and (c) tracking by the proposed scheme is successful under the same situation.

to make additional training of the detector by these images. It is effective to include the mechanisms for dealing with short-range trajectories before/after a connection of objects is lost, as implemented in *K-Track-kai*.

The proposed scheme achieves accurate target trajectories compared to those obtained by manual tracking, but some estimation differences between them still remain. This is caused by the ways of calculating the center point of the detected object. In our proposed scheme, the center point of the object is calculated as the center point of the bounding box attached to the object. Manual tracking and *K-Track* variants calculate the center point of the detected object as the center of gravity in the detected object. This could be a reason for the difference (or error) in the location of a detected object. Also, the locations of misdetected objects can make a big difference by the proposed scheme. When several targets come together, the locations of detected objects tend to be in the center of these targets. This can be improved by considering short trajectories before making a cluster of targets.

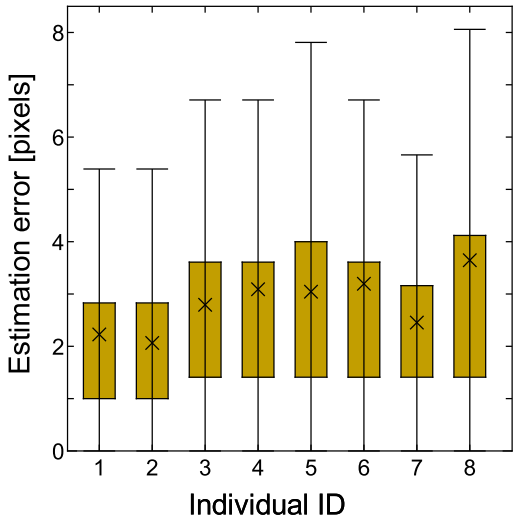


Figure 7: Estimation errors between the proposed scheme and manual tracking for eight individuals. These are calculated from the video-A with 5400 image frames.

5. CONCLUSION

In this paper, we have proposed a method for tracking multiple animals and demonstrated its performance using actual videos. Our proposed method consists of a deep neural network called YOLOv5 as an object detector from the image frames and a scheme of binding objects detected over image frames by *K-Track-kai*. Experimental results showed that (1) only a limited number of training samples was sufficient for detecting target animals, and (2) binding of objects by using forward and backward search in the video sequence was effective for tracking objects across frames. Regarding point (1), the proposed method can be used on low-resource computers, such as laptops without a GPU (Graphics Processing Unit). In such environments, training samples for YOLOv5 can still be prepared and processed on these computers.

The proposed method is expected to be used for videos obtained under different conditions in ethological experiments, due to its easy configuration and robustness. But it is necessary to improve the object binding the mechanisms to achieve more robust tracking by incorporating the mechanisms used in *K-Track-kai*. It is also possible to incorporate the techniques used in *DeepSort* [13], a deep-learning based system for tracking objects. These remain for our future work.

References

[1] A. Pérez-Escudero, J. Vicente-Page, R. C. Hinz, S. Arganda, G. G. de Polavieja, idtracker: tracking individuals in a group by automatic identification of unmarked animals, Nature methods 11 (7) (2014) 743–748.

[2] K. Branson, A. A. Robie, J. Bender, P. Perona, M. H. Dickinson, High-throughput ethomics in large groups of Drosophila, Nature methods 6 (6) (2009) 451–457.

[3] A. Feldman, T. Balch, Automatic Identification of Bee Movement, in: Proc. 2nd Int. Workshop on the Mathematics and algorithms of social insects, 2003, pp. 53–59.
[4] T. Kimura, M. Ohashi, R. Okada, H. Ikeno, A new approach for the simultaneous tracking of multiple honeybees for analysis of hive behavior, Apidologie 42 (2011) 607–617.
[5] T. Kimura, M. Ohashi, K. Crailsheim, T. Schmickl, R. Okada, G. Radspieler, H. Ikeno, Development of a new method to track multiple honey bees with complex behaviors on a flat laboratory arena, PLoS ONE 9 (1) (2014) e84656.
[6] T. Kimura, M. Ohashi, K. Crailsheim, T. Schmickl, R. Okada, G. Radspieler, T. Isokawa, H. Ikeno, A Heuristic Trajectory Decision Method to Enhance the Tracking Performance of Multiple Honeybees on a Flat Laboratory Arena, Transactions of the Institute of Systems, Control and Information Engineers (ISCIE) 32 (3) (2019) 113–122.
[7] G. Jocher, Ultralytics yolov5 (2020). doi:10.5281/zenodo.3908559.
URL `https://github.com/ultralytics/yolov5`
[8] J. Redmon, S. Divvala, R. Girshick, A. Farhadi, You Only Look Once: Unified, Real-time Object Detection, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 779–788.
[9] H. Mizutani, K. Tagai, S. Habe, Y. Takaku, T. Uebi, T. Kimura, T. Hariyama, M. Ozaki, Antenna cleaning is essential for precise behavioral response to alarm pheromone and nestmate–nonnestmate discrimination in japanese carpenter ants (camponotus japonicus), Insects 12 (9) (2021) 773(1–13).
[10] S. E. Hewlett, D. M. Wareham, A. B. Barron, Honey bee (apis mellifera) sociability and nestmate affiliation are dependent on the social environment experienced post-eclosion, Journal of Experimental Biology 221 (Pt3) (2017) jeb.173054.
[11] Labelimg, `https://github.com/HumanSignal/labelImg`, accessed 2025/3/1.
[12] G. Jocher, J. Qiu, Ultralytics yolo11 (2024).
URL `https://github.com/ultralytics/ultralytics`
[13] D. P. N. Wojke, A. Bewley, Simple Online and Realtime Tracking with a Deep Association Metric, CoRR abs/1703.07402.