

Article

Open Access

# MonkeyTrail: A scalable video-based method for tracking macaque movement trajectory in daily living cages

Meng-Shi Liu<sup>1,2,3</sup>, Jin-Quan Gao<sup>4,5</sup>, Gu-Yue Hu<sup>1,2,3,6</sup>, Guang-Fu Hao<sup>1,2,3</sup>, Tian-Zi Jiang<sup>1,2,3,6</sup>, Chen Zhang<sup>7,\*</sup>, Shan Yu<sup>1,2,3,\*</sup>

<sup>1</sup> Brainnetome Center, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China

<sup>2</sup> National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China

<sup>3</sup> School of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing 101408, China

<sup>4</sup> Technology Management Center, SAFE Pharmaceutical Technology Co., Ltd., Beijing 100176, China

<sup>5</sup> Model R&D Center, Beijing Life Biosciences Co., Ltd., Beijing 100176, China

<sup>6</sup> Key Laboratory for NeuroInformation of Ministry of Education, School of Life Science and Technology, University of Electronic Science and Technology of China, Chengdu, Sichuan 611731, China

<sup>7</sup> Department of Neurobiology, School of Basic Medical Sciences, Beijing Key Laboratory of Neural Regeneration and Repair, and Advanced Innovation Center for Human Brain Protection, Capital Medical University, Beijing 100069, China

## ABSTRACT

Behavioral analysis of macaques provides important experimental evidence in the field of neuroscience. In recent years, video-based automatic animal behavior analysis has received widespread attention. However, methods capable of extracting and analyzing daily movement trajectories of macaques in their daily living cages remain underdeveloped, with previous approaches usually requiring specific environments to reduce interference from occlusion or environmental change. Here, we introduce a novel method, called MonkeyTrail, which satisfies the above requirements by frequently generating virtual empty backgrounds and using background subtraction to accurately obtain the foreground of moving animals. The empty background is generated by combining the frame difference method (FDM) and deep learning-based model (YOLOv5). The

entire setup can be operated with low-cost hardware and can be applied to the daily living environments of individually caged macaques. To test MonkeyTrail performance, we labeled a dataset containing >8 000 video frames with the bounding boxes of macaques under various conditions as ground-truth. Results showed that the tracking accuracy and stability of MonkeyTrail exceeded that of two deep learning-based methods (YOLOv5 and Single-Shot MultiBox Detector), traditional frame difference method, and naïve background subtraction method. Using MonkeyTrail to analyze long-term surveillance video

Received: 19 January 2022; Accepted: 17 March 2022; Online: 17 March 2022

Foundation items: This work was supported by the National Key Research and Development Program of China (2017YFA0105203, 2017YFA0105201), National Science Foundation of China (31771076, 81925011), Strategic Priority Research Program of the Chinese Academy of Sciences (CAS) (XDB32040201), Beijing Academy of Artificial Intelligence, and Key-Area Research and Development Program of Guangdong Province (2019B030335001)

\*Corresponding authors, E-mail: [czhang@ccmu.edu.cn](mailto:czhang@ccmu.edu.cn); [shan.yu@nlpr.ia.ac.cn](mailto:shan.yu@nlpr.ia.ac.cn)

&Current address: School of Computing, National University of Singapore, Singapore 119077, Singapore

This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Copyright ©2022 Editorial Office of Zoological Research, Kunming Institute of Zoology, Chinese Academy of Sciences

recordings, we successfully assessed changes in animal behavior in terms of movement amount and spatial preference. Thus, these findings demonstrate that MonkeyTrail enables low-cost, large-scale daily behavioral analysis of macaques.

**Keywords:** Movement trajectory tracking; Video-based behavioral analyses; Background subtraction; Virtual empty background; Occlusion

## INTRODUCTION

Due to their similarity to humans in terms of genetics, physiology, behaviors, and structural/functional characteristics of the brain (Gibbs et al., 2007), macaques are widely used as an effective animal model to study human brain disorders, such as Parkinson's disease (Bezard et al., 2001), Alzheimer's disease (Beckman & Morrison, 2021), autism spectrum disorders (Liu et al., 2016b), and Rett syndrome (Chen et al., 2017). In these studies, behavioral analyses provide important information for validating models and developing effective treatments (Krakauer et al., 2017; Lehner, 1987; Nice, 1954). However, traditional behavioral analyses conducted by human experimenters are both time-consuming and susceptible to subjective biases (Bateson & Martin, 2021). More importantly, behavioral parameters that can be analyzed manually are very limited. Thus, automatic methods for behavioral analysis have attracted increasing attention in recent years. Among them, video-based methods built on rapid technological developments in computer vision have become promising approaches (Mathis et al., 2020).

Video-based automatic analysis of macaque behavior usually involves measurement of movement amount (quantifying overall movement intensity) (Caiola et al., 2019; Hashimoto et al., 1999; Togasaki et al., 2005; Yabumoto et al., 2019), movement trajectory (measuring trajectory of body during movements) (Bala et al., 2020; Francisco et al., 2019; Graving et al., 2019; Lind et al., 2005; Mathis et al., 2018; Ueno et al., 2019; Walton et al., 2006; Yabumoto et al., 2019; Yao et al., 2019), and behavioral categorization (categorizing different types of activities) (Ballesta et al., 2014; Hu et al., 2020a, 2020b; Wiltischko et al., 2015). Movement trajectory can provide vital information for various purposes. Movement trajectory length not only reflects overall activity level, but also records important spatial information about movement (Yabumoto et al., 2019), which can be used to categorize different behaviors for quantifying movement characteristics specific to Parkinson's disease, drunkenness, and aging (Caiola et al., 2019; Liu et al., 2016b; Pandya et al., 2015; Togasaki et al., 2005; Walton et al., 2006; Yabumoto et al., 2019). Due to its wide applications, video-based trajectory tracking has been actively developed in recent years and can be divided into methods based on traditional image processing (Lind et al., 2005; Walton et al., 2006; Yabumoto et al., 2019) and on deep learning (Bala et al., 2020; Francisco et al., 2019; Graving et al., 2019; Mathis et al., 2018; Ueno et al., 2019; Yao et al., 2019).

Background subtraction is an essential technique used in

traditional video-based trajectory tracking (Lind et al., 2005; Walton et al., 2006; Yabumoto et al., 2019), which considers the animal as the foreground and assumes the background is stable. Thus, subtracting the background from individual frames can highlight the animal. If the contrast between the selected background and animal is high, background subtraction can achieve accurate tracking (Walton et al., 2006). Compared with deep learning-based methods, background subtraction does not need to learn features of the animal, and thus needs no model training (Bala et al., 2020; Yao et al., 2019). However, background subtraction is very sensitive to environmental changes in the video. Many background subtraction-based tracking methods require specific environments (Walton et al., 2006; Yabumoto et al., 2019) to provide a clean and stable background, e.g., using specialized cages made of transparent material (Ueno et al., 2019). These special requirements have prevented the wide use of automatic tracking of animals in daily living environments.

Deep learning-based trajectory tracking methods have capitalized on the rapid development in deep neural networks. Powerful tracking applications (Graving et al., 2019; Mathis et al., 2018) have been introduced in the field of animal motion tracking in recent years. Most tracking key points (Bala et al., 2020; Graving et al., 2019; Mathis et al., 2018; Yao et al., 2019) are based on pose estimation or track the whole animal (Francisco et al., 2019; Ueno et al., 2019) using object detection models. High-dimensional motion information can be obtained by recording the trajectory of key points (Johansson, 1973). However, pose estimation is sensitive to occlusion. Although object detection models, such as YOLOv5 (Jocher, 2021; Redmon et al., 2016) and Single-Shot MultiBox Detector (SSD) (Liu et al., 2016a), can alleviate the problem of occlusion, they fail under severely occluded conditions, e.g., when macaques are behind dense mesh. The daily living cages of macaques contain bars, mesh, and other objects (e.g., water bottle, food box), which creates many occlusions for cameras outside the cage. To solve this issue, specific environments with less occlusion are also needed for deep learning-based methods (Bala et al., 2020; Caiola et al., 2019; Graving et al., 2019; Yao et al., 2019), thus creating the same problems as traditional methods. In addition, deep learning-based methods require a large amount of labeled data to train the model (Mathis et al., 2020), which is difficult to obtain for individual applications (Mathis et al., 2018), especially for animals with complex gestures such as macaques (Bala et al., 2020).

Despite significant progress in video-based animal trajectory tracking, important challenges still exist. A low-cost tracking method that can be efficiently applied to normal living environments is highly desirable. Here, we developed an effective method, MonkeyTrail, for tracking macaque body trajectory. The system only requires normal surveillance cameras mounted outside the cage. Furthermore, it needs no special environment and only requires a small amount of data for training. This method provides a low-cost, scalable solution to accurately track body movements of macaques in their daily living cages.

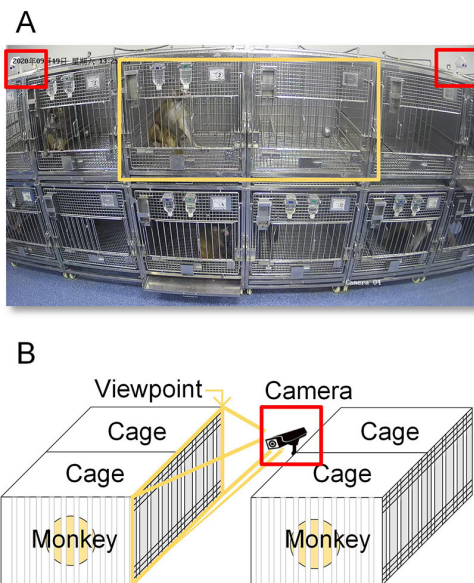
## MATERIALS AND METHODS

### Animals and recording environment

This study was conducted according to the international standards of non-human primate care and use and was approved by the Institutional Review Board/Ethics Committee of Capital Medical University (AEEI-2019-077). The monkeys were housed in animal rooms under a temperature of 18–26 °C and humidity of 40%–70%. Video data of three adult macaques in their daily living cages were used to develop and validate the tracking method. The macaques were provided with water, certified primate biscuits, vegetables and fruit daily. The animals were individually housed in cages adjacent to each other. The room was maintained on two light/dark cycles: 12 h light/dark or 11.5 h light/12.5 h dark. A regular high-definition (HD) surveillance camera (1 920×1 080 pixels, 25 Hz refresh rate, 4 mm focal length) was mounted on the other side of the room above the height of the cages. This setup ensured minimal disturbance to the normal living environment. However, it introduced technical challenges for automatic video tracking, including occlusion due to the metal bars and mesh of the cage and variable background due to movable objects, e.g., toys and pull-out plates, which needed to be overcome in this study. The overall recording environment and camera setup are shown in Figure 1.

### Overall workflow of MonkeyTrail

We devised a Python-based program called MonkeyTrail, which uses background subtraction supplemented with YOLOv5 to record the trajectory of individually caged



**Figure 1 Overall recording environment and camera setup**

A: One frame of recorded video, showing arrangement of monkey cages. For each recording, two cages in upper and middle positions with better visibility (marked by yellow box) were analyzed by proposed method. Position of camera in A is marked by red box. B: Diagram showing setup of recording cameras mounted on the other side of the room above cage height. Yellow and red boxes in B correspond to A.

macaques in their daily living environment. The major steps involved in the procedure are briefly described below.

In MonkeyTrail, video preprocessing first establishes parameters for the algorithm, automatically crops the video size, and sets the frame rate for video data. Next, a virtual empty background is generated by combining the frame difference method (FDM) and YOLOv5. The bounding boxes are then located using background subtraction with the generated empty backgrounds and simple image processing techniques. Finally, the centers of the bounding boxes in individual frames are connected to each other, producing a moving trajectory. The length of the trajectory is used to calculate the total amount of movements. The MonkeyTrail workflow is shown in Figure 2.

The source code of MonkeyTrail is available at <https://github.com/Xingheliu/MonkeyTrail>.

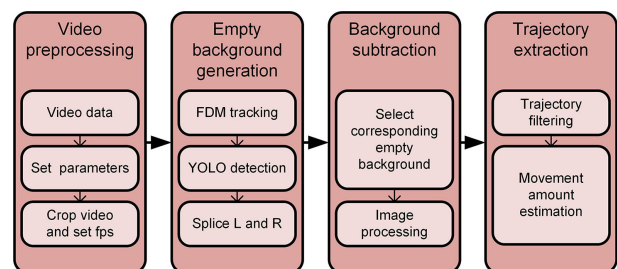
### Video preprocessing

First, the region of interest in the first video frame is manually selected, which includes a single cage housing a macaque, and then the frames per second (fps) are set for further analyses. After all parameters are obtained, MonkeyTrail will automatically crop the video size and set the fps. This step eliminates irrelevant pixels from the videos to facilitate downstream processing and avoid interference from other animals. According to the first frame of each video, the initial tracking frame of the macaque and the center point of the cage are manually selected. The following steps are automatic.

### Empty background generation

A key step in MonkeyTrail is to automatically update the empty background so that the background subtraction method (BSM) can be used for the cages. The generation of an empty background can be divided into two steps: (1) screening frames with high movements in the video sequence into two sets (L and R, representing frames with animal located in the left and right half of the cage, respectively); and (2) selecting the closest pair of L and R frames in a time sequence, and stitching together each empty area without macaques to form a complete empty background.

In the first step, FDM is used to initially obtain the L and R sets, with YOLOv5 then used to obtain reliable L and R sets. Specifically, there are three sub-steps: (1) The frame pixel difference is obtained using FDM. The amount of pixel difference can provide rough estimations of animal movement. At the same time, the position of pixel differences can provide the approximate position of the bounding box for the animal.



**Figure 2 MonkeyTrail workflow**

(2) As the tracking position obtained by FDM will be more reliable when the animal is actively moving, a high-movement threshold (i.e., pixel difference threshold) is used to select frames in the video sequence and classify the frames into L and R sets based on the position of the bounding box. (3) YOLOv5 is then used to detect the L and R sets obtained in the previous step, and only preserve frames with a high-confidence bounding box.

In the second step, two suitable frames in the L and R sets are used to generate an empty background. This process involves two sub-steps: (1) To improve empty background quality, two frames are selected in the L and R sets with a high degree of location discrimination between tracking boxes. The latter criterion ensures that in these frames, the animal is on one side of the cage and the other side represents a half-empty background. (2) The L and R half-empty backgrounds are then spliced to provide a complete empty background. After successfully splicing multiple pairs of L and R frames, a series of automatically generated empty backgrounds are obtained.

### Background subtraction in daily living cages

With frequently updated empty backgrounds, background subtraction combined with proper image processing will provide accurate bounding boxes of the macaques for tracking. Specifically, the original RGB-color frames are converted into gray scale, with the foreground then obtained by background subtraction. The extracted foreground is then subjected to standard image processing, including spatial median filtering, binarizing, eroding, and dilating (Gonzalez & Woods, 2002). The bounding box is formed by finding the minimal rectangular area that covers the foreground.

The parameters used for image processing in our program were optimized according to the experimental environment used. If applied to other environments, they can be adjusted appropriately to obtain the best results. In addition, to avoid over-processing the image under severe occlusion, which can result in excessive elimination of foreground results, previous FDM results are used to judge the size of the tracking frame. For example, if the foreground is too small and in a high-density occlusion area, and if the FDM result is also in this area, then a bounding box with the same size as the immediate previous frame will be tracked.

### Trajectory extraction and movement amount estimation

MonkeyTrail extracts trajectory by connecting the center point of the bounding boxes in consecutive frames, which can be used to reflect the center location of the macaque's body. To reduce disturbance of the tracking point caused by subtle movements of the limbs or head, movements smaller than a preset threshold are discarded. In other words, when the Euclidean distance of the tracking point compared to the previous time point does not exceed this threshold, the change in the tracking point is filtered out. The distance of the trajectory represents the amount of movement.

### Manually annotated test dataset

We created a test dataset to quantify the performance of MonkeyTrail in comparison to other methods. To test its reliability in various macaque movement states, the selected

data included continuous movement, continuous stillness, and transition between movement and stillness. To test MonkeyTrail with different levels of occlusion, we selected data that included cases of macaques occluded by metal bars, mesh grids, and parts of their own body. Furthermore, the three monkeys used in the test dataset have different appearances, including size and fur coloring. None of the animals in the test dataset were present in the training set. Finally, to test MonkeyTrail performance across different lighting conditions, the dataset included both day and night video data.

The entire test dataset contained 55 min of video, composed of 13 consecutive clips of varying lengths to cover all conditions mentioned above. To reduce the workload of manual labeling, the frame rate was decreased to 2 or 5 fps, resulting in a total frame number of 8130. We manually selected the bounding box containing the entire animal as ground-truth using the labeling platform LabelImg (Tzutalin, 2015).

## RESULTS

### Updating empty background to detect monkeys in cages using background subtraction

Many methods using background subtraction to locate macaques in videos require a physically created empty background obtained by removing the animal from the monitoring environment. These empty backgrounds cannot be frequently updated. However, due to changes in the environment over time (e.g., cage movement, changes in lighting conditions, and introduction of new objects in the frame), the empty background obtained in advance will not match the true background over long periods of recording.

Typical changes in the background of a randomly chosen daily video recording (~2 h) are shown in Figure 3. The three sequentially generated (~1 h intervals) empty backgrounds of the same cage (Figure 3C–E) (see methods for details) showed that the background changed in brightness and detailed appearance. To highlight these changes, we used a video frame with the monkey in cage (Figure 3B) to perform background subtraction with frames showing in Figure 3C–E, in which Figure 3E is most close to the moment of Figure 3B. The background subtraction results are illustrated by the white pixels in Figure 3F–H. As seen in Figure 3H, background subtraction successfully highlighted the macaque, but using empty backgrounds obtained 1 or 2 h earlier resulted in low-quality outcomes (Figure 3F, G). Thus, the longer the time elapsed since obtaining the empty background, the poorer the subtraction results, making it more difficult to locate the foreground using image processing techniques. The above example illustrates the importance of frequently updating empty backgrounds to detect caged monkeys using background subtraction.

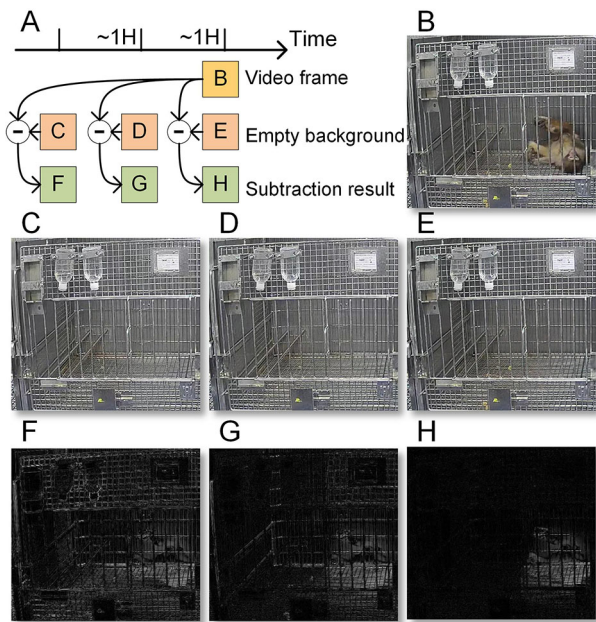
### Application of YOLOv5 to generate empty background and visualize tracking results

To achieve better background subtraction, we introduced a method in MonkeyTrail to frequently update the empty background with minimal animal disturbance. Simply, we



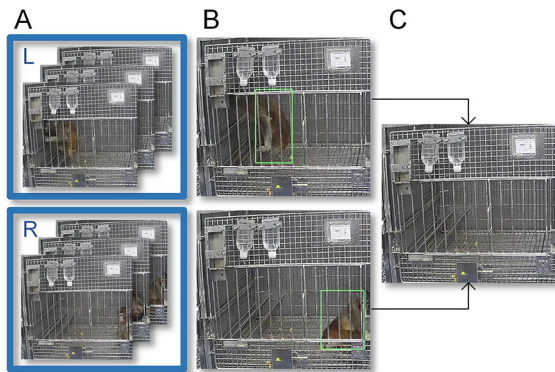
detected the frames in which the animal was on one side of the cage, thus providing a half-empty background, and then spliced the left and right half-empty backgrounds adjacent to each other in time to generate a virtual empty background.

As seen in Figure 4, the L and R frame sets (Figure 4A) showed animals in the left and right halves of the cage, respectively. The L and R frames were first detected using FDM to roughly locate the animal in the frame, followed by the deep learning-based model of YOLOv5 to select frames with a high-confidence bounding box. Two frames (Figure 4B) from the L and R set close in time were then paired. Finally, the areas where there were no macaques in the two frames were stitched together to generate a virtual empty background



**Figure 3 Influence of environmental changes on efficacy of background subtraction**

A: Relationships among B–H. (C, F), (D, G), (E, H) are empty backgrounds at certain times and corresponding background subtraction results. Empty backgrounds of C, D, and E were obtained at 1 h intervals. Real-time frame B subtracted from C–E is a video frame near time of E.



**Figure 4 Method to frequently update empty background**

(Figure 4C). For more details, see section "Empty background generation" in Materials and Methods.

With the frequently updated empty background, background subtraction showed good performance in highlighting foreground objects (cf. Figure 5A–C) using simple image processing techniques (see methods for details).

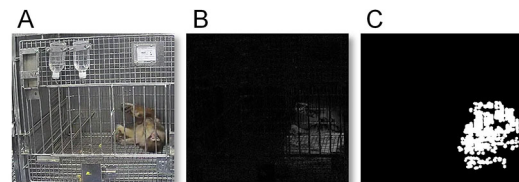
Thus, based on the extracted foreground containing the animal, the bounding box and movement trajectory can be easily obtained. Furthermore, MonkeyTrail can provide bounding boxes reasonably fit to the trunk of the monkey under various occlusion and lighting conditions (day and night) (Figure 6; Supplementary Videos S1, S2).

#### Accuracy of MonkeyTrail in generating bounding boxes

MonkeyTrail capitalizes on the combined strengths of BSM, FDM, and YOLOv5. Thus, we next quantitatively analyzed the improvements in MonkeyTrail over these three methods and the deep learning model SSD. We prepared a manually annotated dataset to obtain the ground-truth (see Materials and Methods for details), and then compared the MonkeyTrail, BSM, FDM, YOLOv5, and SSD results. To provide a comprehensive comparison, the testing dataset contained 13 video clips from three animals, covering various movement modes, occlusions, and lighting conditions (day and night). To facilitate fair comparison, the FDM, BSM, YOLOv5, and SSD processes were the same as the corresponding functions in MonkeyTrail. For example, image processing in BSM without an updated background is the same as MonkeyTrail, and the YOLOv5 model for object detection is the same as that used by MonkeyTrail. Tracking accuracy was determined by IoU, which measures the degree of overlap between the bounding box generated by individual methods ( $r_t$ ) and that of the ground-truth ( $r_o$ ), and is defined as:

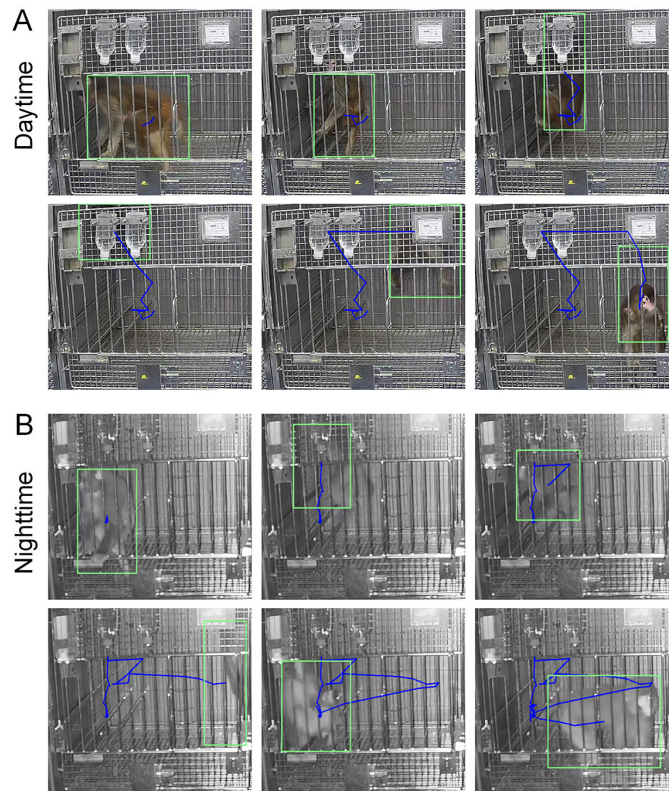
$$IoU = \frac{|r_t \cap r_o|}{|r_t \cup r_o|} \quad (1)$$

To visualize the tracking results of different methods, we displayed the IoU value for each frame in the test dataset with a continuous time axis, as shown in Figure 7A–E. To better understand how animal activity and occlusions can affect the different methods, we plotted the estimated amount of total activity, as well as the period in which the animals were severely occluded, as shown in Figure 7F.



**Figure 5 Background subtraction process with generated empty background**

A: One video frame showing typical situation in daily living cage. B: Background subtraction between A and virtual empty background generated temporally close to A, thus highlighting foreground containing animal. C: Image processing result of B, after spatial median filtering, binarizing, eroding, and dilating. A and B are redrawn from Figure 3B and H, respectively.



**Figure 6 Representative tracking results for macaque during daytime and nighttime**

Green box and blue line represent bounding box and trajectory, respectively. Sequence of frames is from left to right, then top to bottom. Time interval between each frame is >10 s. These examples include different motions and various levels of occlusion.

Among the methods, MonkeyTrail provided the most accurate and stable tracking results, reflected by overall larger IoU values and smaller variability (Figure 7A). In comparison, both SSD and YOLOv5 showed highly fluctuating performance (Figure 7B, C), with reasonable tracking results interrupted by periods of zero IoU, indicating failure to detect the animal at all (Figure 7B, C). Importantly, these failures usually occurred when the animals were occluded (cf. Figure 7F), suggesting that SSD and YOLOv5 are highly sensitive to occlusion. Without frequently updating the empty background, traditional BSM exhibited noisy results. Thus, it was difficult to obtain accurate macaque foreground areas by standard image processing, leading to inaccurate tracking results (Figure 7D). In addition, FDM provided relatively accurate tracking results, but only when the animals were actively moving (Figure 7E, cf. Figure 7E). These analyses provide insights into the advantages and disadvantages of different methods and illustrate how MonkeyTrail can combine their strengths to achieve accurate and stable tracking.

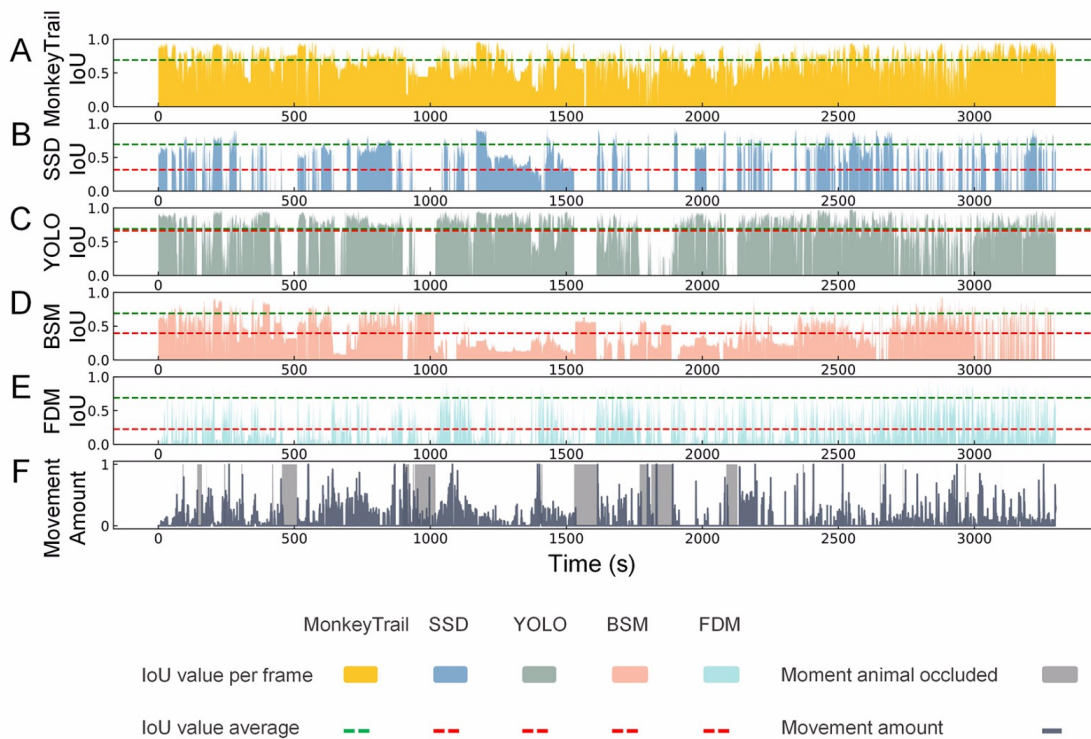
To provide a more comprehensive and quantitative comparison, we compared tracking success rates with a systematically varying overlap threshold, which is often used for gauging the performance of object tracking methods (Wu et al., 2013). The results are shown in Figure 8. When the “overlap threshold” of a certain frame was greater than the set threshold, the frame was regarded as a success, and the percentage of total successful frames in all frames was defined as the “success rate”. Compared with SSD, BSM, and

FDM, the success rate of MonkeyTrail was consistently higher over the entire overlap threshold range. Although YOLOv5 seemed to be more accurate in a small proportion of frames, it failed to detect monkeys completely in ~15% of frames; in contrast, MonkeyTrail exhibited more stable tracking results, yielding more favorable overall performance.

#### Application of MonkeyTrail to extract movement patterns in daily living cages

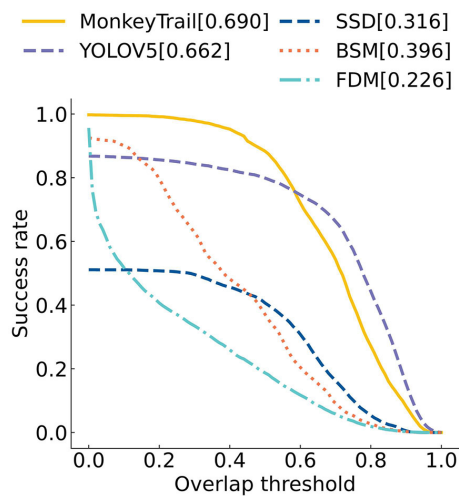
To demonstrate the practical value of our method in analyzing the behaviors of macaques, we used the trajectories recorded by MonkeyTrail to calculate the amount of movement and spatial preference of two macaques over two time periods separated by one year, each containing recordings of five consecutive days. Movement amount and spatial preference are useful indicators of behavioral changes induced by factors such as drug injection, surgery, and changes in external conditions (Caiola et al., 2019; Togasaki et al., 2005; Yabumoto et al., 2019). Daily monitoring of these parameters can reveal their acute and chronic effects on behavior.

Total activity is shown in Figure 9. In addition to the obvious results that sleep-awake cycles affect activity patterns, monkeys in the 2019 recordings showed a bimodal pattern of active movement during the daytime in the cages, which may be the combined result of physiological (i.e., napping) and housekeeping (i.e., feeding, lighting) factors. In 2020, both the lighting schedule and housekeeping activity pattern changed. As a result, total amount of activity of animals changed to a



**Figure 7 Visualization of performance in generating bounding boxes by different methods**

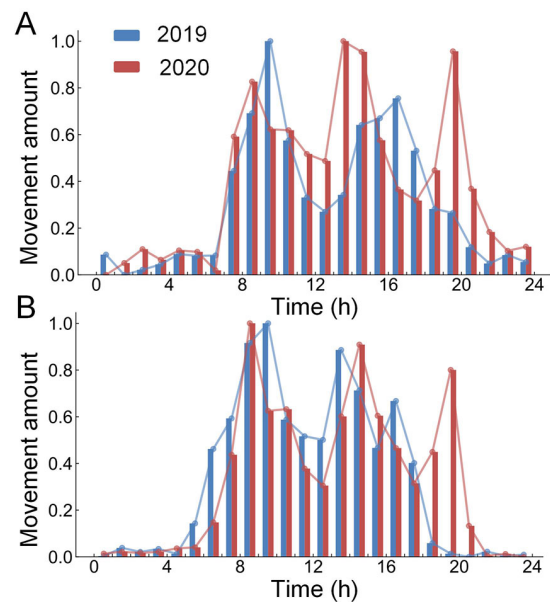
Results of several trajectory tracking methods were compared with results of manual annotation to calculate accuracy. IoU, which measures accuracy of bounding box, was plotted for individual frames concatenated in time. A–E: Results of MonkeyTrail, SSD, YOLOv5, BSM, and FDM, with IoU shown in different colors. Green dashed line indicates mean value of IoU for MonkeyTrail, and red dashed lines represent mean values of IoU for corresponding methods. F: Amount of motion (calculated by length of trajectory movement) was plotted with the same time frame as in A–E. Gray box represents time when macaque is occluded by parts of cage. Data were from three monkeys, including 8 130 frames.



**Figure 8 Tracking success rates with systematically varying overlap thresholds for different tracking methods**

Success rate is percentage of total number of frames with IoU values greater than predefined threshold. Number in square brackets indicates average IoU value.

trimodal pattern. These results illustrate the value of tracking total activity amount to capture behavioral changes in monkeys in daily living cages. Such changes not only reflect



**Figure 9 Daily total activity patterns of two monkeys (A and B) captured with MonkeyTrail**

Blue and red columns represent results obtained in 2019 and 2020, respectively. Average activity counts in each hourly time segment were obtained from 5-day recordings.



the influence of unintentional external factors, as above, but also provide valuable information for validating intentional treatments, such as drug administration or surgical operations.

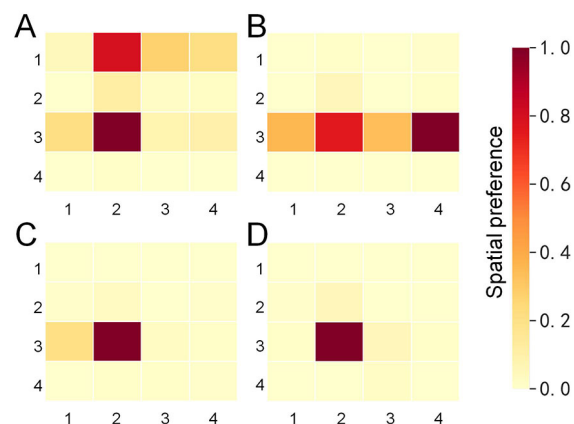
The spatial preferences are shown in Figure 10. We first divided the two-dimensional (2D) projection of the cage into 16 areas. We then counted the number of times the macaque's trajectory passed through each area (normalized by the maximum number found in one area) to determine its spatial preferences. Similar to total activity amount, we analyzed the average spatial preferences of the macaques from 1100h to 1200h for five consecutive days in both 2019 and 2020. In 2019, monkey A showed a preference for hanging at the top of the cage (Figure 10A), while monkey B preferred to sit at the bottom of the cage (Figure 10C). A year later, monkey B still showed a preference for sitting on the base of the cage (Figure 10D), but monkey A had changed its behavioral preference, more often sitting than hanging (Figure 10B). As shown in Figures 9, 10, tracking animal behavior in their daily living cages can provide useful information regarding movement patterns in both temporal and spatial domains.

## DISCUSSION

In this study, we introduced a simple method named MonkeyTrail to track the movement trajectories of macaques. MonkeyTrail is based on frequently updating the empty background to apply background subtraction effectively and accurately for analyzing videos recorded outside animal cages. To the best of our knowledge, due to its minimal recording environment and hardware requirements, MonkeyTrail is the first method that can be deployed in parallel to monitor many monkeys in their daily living cages.

Several limitations are worth noting. Firstly, this method does not include steps to eliminate image distortion due to different viewing angles, therefore we recommend that the camera be mounted directly opposite the cage of interest. Secondly, if an animal remains at one side of the cage for a prolonged period, it can be difficult to generate a virtual empty background, which may impact tracking accuracy. However, our empirical data showed that, on average, updating the background once every 40 min showed reasonably good results, thus providing an ample time window to generate multiple virtual empty backgrounds. Thirdly, our method is primarily designed for individually caged animals. Although it can be used to track multiple macaques in the same cage as long as they are spatially separated, the method cannot be readily extended to situations in which multiple animals interact with each other.

Movement trajectories extracted by MonkeyTrail can be used to analyze spatial preference and movement amount. Compared with motion recorded by pixel differences (Caiola et al., 2019; Hashimoto et al., 1999), motion calculated by the distance of trajectory, although not suited to detect subtle movements of small body parts, can provide more accurate results at the whole-body movement level, especially under conditions with severe occlusion. Trajectory also provides spatial information, which is missing in pixel differences. Although posture estimation can be used to analyze more detailed movement patterns, overall movement trajectory of animals can still provide important information. For example, Pandya et al. (2015) verified the relationship between



**Figure 10 Spatial preference of macaques extracted by MonkeyTrail**

A, B/C, D, results of monkeys A/B obtained in 2019 and 2020, respectively. Horizontal and vertical axes of heat map represent X and Y coordinates of cage, respectively. Each heat map region represents number of times macaque's trajectory passed through this space, normalized by maximum number found in one region (color-coded). Each heat map was obtained by averaging trajectory data of five days.

mitochondrial aging and age-dependent motor function loss by analyzing changes in distance and speed of movement of different-aged macaques, which directly supported the mitochondrial aging theory. Yabumoto et al. (2019) analyzed changes in spatial patterns of movement trajectory in macaques before and after alcohol injection, thus revealing the effects of alcohol in movement control. Liu et al. (2016b) recorded the locomotion trajectory of monkeys in cages and found anxiety-associated behaviors, e.g., repetitive, circular locomotion, in animals with MeCP2 overexpression, mimicking autism spectrum disorders in humans.

In addition to extracting movement trajectories, MonkeyTrail can also provide outlines of animals for individual frames, with either bounding boxes or body masks, which may be instrumental for future algorithms aimed at pose detection. In addition, the bounding boxes and their contents provided by MonkeyTrail can serve as training samples to train or fine-tune other deep learning-based methods for more sophisticated detection and recognition.

MonkeyTrail uses long-term video recordings of macaques in their daily living cages using ordinary HD cameras mounted outside the cage. This low-cost setup is scalable to automatically track many animals, thus allowing large-scale applications. In addition to behavioral tracking in future experiments, our method can also be used to analyze stored data retrospectively for animal rooms equipped with video recording devices. Furthermore, this method can be readily extended to three-dimensional (3D) tracking with depth cameras (Yabumoto et al., 2019), thus providing more comprehensive information regarding the movement patterns of animals.

## SUPPLEMENTARY DATA

Supplementary data to this article can be found online (Supplementary Video S1, S2).



## COMPETING INTERESTS

The Institute of Automation, Chinese Academy of Sciences has submitted patent application on the methods described in this paper (application No. 2021116673622; invented by M.S.L and S.Y.; pending).

## AUTHORS' CONTRIBUTIONS

M.S.L., C.Z., J.Q.G., T.Z.J., and S.Y. designed the experiments. C.Z. and J.Q.G. provided the original video data. M.S.L. developed the algorithm. M.S.L., G.Y.H., G.F.H., M.S.L., and S.Y. wrote the manuscript. All authors read and approved the final version of the manuscript.

## REFERENCES

- Bala PC, Eisenreich BR, Yoo SBM, Hayden BY, Park HS, Zimmermann J. 2020. Automated markerless pose estimation in freely moving macaques with OpenMonkeyStudio. *Nature Communications*, **11**(1): 4560.
- Ballesta S, Reymond G, Pozzobon M, Duhamel JR. 2014. A real-time 3D video tracking system for monitoring primate groups. *Journal of Neuroscience Methods*, **234**: 147–152.
- Bateson M, Martin PR. 2021. *Measuring Behaviour: an Introductory Guide*. 4<sup>th</sup> ed. Cambridge: Cambridge University Press.
- Beckman D, Morrison JH. 2021. Towards developing a rhesus monkey model of early Alzheimer's disease focusing on women's health. *American Journal of Primatology*, **83**(11): e23289.
- Bezard E, Dovero S, Prunier C, Ravenscroft P, Chalon S, Guilloteau D, et al. 2001. Relationship between the appearance of symptoms and the level of nigrostriatal degeneration in a progressive 1-methyl-4-phenyl-1, 2, 3, 6-tetrahydropyridine-lesioned macaque model of Parkinson's disease. *Journal of Neuroscience*, **21**(17): 6853–6861.
- Caiola M, Pittard D, Wichmann T, Galvan A. 2019. Quantification of movement in normal and parkinsonian macaques using video analysis. *Journal of Neuroscience Methods*, **322**: 96–102.
- Chen YC, Yu JH, Niu YY, Qin DD, Liu HL, Li G, et al. 2017. Modeling rett syndrome using TALEN-edited *MECP2* mutant cynomolgus monkeys. *Cell*, **169**(5): 945–955.E10.
- Francisco FA, Nührenberg P, Jordan AL. 2019. A low-cost, open-source framework for tracking and behavioural analysis of animals in aquatic ecosystems. *BioRxiv*: 571232.
- Gibbs RA, Rogers J, Katze MG, Bumgarner R, Weinstock GM, Mardis ER, et al. 2007. Evolutionary and biomedical insights from the rhesus macaque genome. *Science*, **316**(5822): 222–234.
- Gonzalez RC, Woods RE. 2002. *Digital Image Processing*. 2<sup>nd</sup> ed. Prentice Hall: Upper Saddle River.
- Graving JM, Chae D, Naik H, Li L, Koger B, Costelloe BR, et al. 2019. DeepPoseKit, a software toolkit for fast and robust animal pose estimation using deep learning. *Elife*, **8**: e47994.
- Hashimoto T, Izawa Y, Yokoyama H, Kato T, Moriizumi T. 1999. A new video/computer method to measure the amount of overall movement in experimental animals (two-dimensional object-difference method). *Journal of Neuroscience Methods*, **91**(1-2): 115–122.
- Hu GY, Cui B, Yu S. 2020a. Joint learning in the spatio-temporal and frequency domains for skeleton-based action recognition. *IEEE Transactions on Multimedia*, **22**(9): 2207–2220.
- Hu GY, Cui B, He Y, Yu S. 2020b. Progressive relation learning for group activity recognition. *In: Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Seattle: IEEE, 977–986.
- Jocher G. 2021. YOLOv5. <https://github.com/ultralytics/yolov5>.
- Johansson G. 1973. Visual perception of biological motion and a model for its analysis. *Perception & Psychophysics*, **14**(2): 201–211.
- Krakauer JW, Ghazanfar AA, Gomez-Marín A, Maciver MA, Poeppel D. 2017. Neuroscience needs behavior: correcting a reductionist bias. *Neuron*, **93**(3): 480–490.
- Lehner PN. 1987. Design and execution of animal behavior research: an overview. *Journal of Animal Science*, **65**(5): 1213–1219.
- Lind NM, Vinther M, Hemmingsen RP, Hansen AK. 2005. Validation of a digital video tracking system for recording pig locomotor behaviour. *Journal of Neuroscience Methods*, **143**(2): 123–132.
- Liu W, Anguelov D, Erhan D, Szegedy C, Reed S, Fu CY, et al. 2016a. SSD: single shot multibox detector. *In: Proceedings of the 14th European Conference on Computer Vision*. Amsterdam: Springer, 21–37.
- Liu Z, Li X, Zhang JT, Cai YJ, Cheng TL, Cheng C, et al. 2016b. Autism-like behaviours and germline transmission in transgenic monkeys overexpressing MeCP2. *Nature*, **530**(7588): 98–102.
- Mathis A, Mamidanna P, Cury KM, Abe T, Murthy VN, Mathis MW, et al. 2018. DeepLabCut: markerless pose estimation of user-defined body parts with deep learning. *Nature Neuroscience*, **21**(9): 1281–1289.
- Mathis A, Schneider S, Lauer J, Mathis MW. 2020. A primer on motion capture with deep learning: principles, pitfalls, and perspectives. *Neuron*, **108**(1): 44–65.
- Nice M M. 1954. Reviewed work: *The Herring Gull's World*. A study of the social behaviour of birds by Niko Tinbergen. *Bird-Banding*, **25**(2): 81–82.
- Pandya JD, Grondin R, Yonutas HM, Haghazadeh H, Gash DM, Zhang ZM, et al. 2015. Decreased mitochondrial bioenergetics and calcium buffering capacity in the basal ganglia correlates with motor deficits in a nonhuman primate model of aging. *Neurobiology of Aging*, **36**(5): 1903–1913.
- Redmon J, Divvala S, Girshick R, Farhadi A. 2016. You only look once: Unified, real-time object detection. *In: Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas: IEEE, 779–788.
- Togasaki DM, Hsu A, Samant M, Farzan B, DeLanney LE, Langston JW, et al. 2005. The Webcam system: a simple, automated, computer-based video system for quantitative measurement of movement in nonhuman primates. *Journal of Neuroscience Methods*, **145**(1-2): 159–166.
- Tzutalin. 2015. LabelImg. <https://github.com/tzutalin/labelImg>.
- Ueno M, Hayashi H, Kabata R, Terada K, Yamada K. 2019. Automatically detecting and tracking free-ranging Japanese macaques in video recordings with deep learning and particle filters. *Ethology*, **125**(5): 332–340.
- Walton A, Branham A, Gash DM, Grondin R. 2006. Automated video analysis of age-related motor deficits in monkeys using EthoVision. *Neurobiology of Aging*, **27**(10): 1477–1483.
- Wiltschko AB, Johnson MJ, Iurilli G, Peterson RE, Katon JM, Pashkovski SL, et al. 2015. Mapping sub-second structure in mouse behavior. *Neuron*, **88**(6): 1121–1135.
- Wu Y, Lim J, Yang MH. 2013. Online object tracking: A benchmark. *In: Proceedings of 2013 IEEE Conference on Computer Vision and Pattern Recognition*. Portland, OR, USA: IEEE, 2411–2418.
- Yabumoto T, Yoshida F, Miyauchi H, Baba K, Tsuda H, Ikenaka K, et al. 2019. MarmotDetector: a novel 3D automated system for the quantitative assessment of marmoset behavior. *Journal of Neuroscience Methods*, **322**: 23–33.
- Yao Y, Jafarian Y, Park HS. 2019. MONET: multiview semi-supervised keypoint detection via epipolar divergence. *In: Proceedings of 2019 IEEE/CVF International Conference on Computer Vision*. Seoul: IEEE, 753–762.