Hindawi
Mathematical Problems in Engineering
Volume 2021, Article ID 1392362, 15 pages
https://doi.org/10.1155/2021/1392362

Research Article

Key Points Tracking and Grooming Behavior Recognition of *Bactrocera minax* (Diptera: Trypetidae) via DeepLabCut

Wei Zhan,[1] **Yafeng Zou**,[1] **Zhangzhang He**,[2] **and Zhiliang Zhang**[1]

[1]*School of Computer Science, Yangtze University, Jingzhou 434023, China*
[2]*Insect Ecology Laboratory, College of Agriculture, Yangtze University, Jingzhou 434025, China*

Correspondence should be addressed to Wei Zhan; zhanwei814@yangtzeu.edu.cn

Received 28 June 2021; Revised 16 July 2021; Accepted 22 July 2021; Published 3 August 2021

Academic Editor: Yunchao Tang

Statistical analysis of *Bactrocera* grooming behavior is important for pest control and human health. Based on DeepLabCut, this study proposes a noninvasive and effective method to track the key points of *Bactrocera minax* and to detect and analyze its grooming behavior. The results are analyzed and calculated automatically by a computer program. Traditional movement tracking methods are invasive; for instance, the use of artificial pheromone may affect the behavior of *Bactrocera minax*, thus directly affecting the accuracy and reliability of experimental results. Traditional research studies mainly rely on manual work for behavior analysis and statistics. Researchers need to play the video frame by frame and record the time interval of each grooming behavior manually, which is time-consuming, laborious, and inaccurate. So the advantages of automated analysis are obvious. Using the method proposed in this paper, the image data of 94538 frames from 5 adult *Bactrocera* were analyzed and 14 key points were tracked. The overall tracking accuracy was as high as 96.7%. In the behavior analysis and statistics, the average accuracy rate of the five grooming behavior was all above 96%, and the accuracy rate of the remaining two grooming behavior was over 87%. The experimental results show that the automatic noninvasive method designed in this paper can track many key points of *Bactrocera minax* with high accuracy and ensure the accuracy of insect behavior recognition and analysis, which greatly reduces the manual observation time and provides a new method for key points tracking and behavior recognition of related insects.

1. Introduction

Bactrocera minax is one of the most serious pests in citrus [1]. Since the 1960s, citrus diseases and insect pests have been increasing year by year. The economic losses caused by *Bactrocera minax* are as high as 300 million yuan per year, and its resulting rotten of citrus endangers human health [2]. It has been confirmed that inhibition of grooming behavior increases mortality in insect-pathogen bioassay [3, 4]. Grooming behavior provides effective solutions for *Bactrocera minax* control.

First of all, let us introduce what is grooming behavior. Grooming is a broad definition, including all forms of body surface care. For animals, grooming is a very important activity for healthy survival [5] and is also a very common behavior [6]. Grooming has multiple functions: removing foreign dust particles from the epidermis and the surface of sensory organs [7], removing secretions and epidermal lipids from the soiled body surface [8, 9], collecting pollen particles as food [10], and removing external parasites or pathogens [11–16]. The grooming behavior of insects is a very important part of their defense mechanism [17]. The role of insect grooming and hygienic activities is gaining recognition in the field of insect pathology [18]. Hygienic behavior has been shown to play a key role in disease prevention in insects [19, 20].

The researchers needed to do a lot of experiments in Petri dishes in the exact same environment. They used special drugs to inhibit the grooming behavior of the insects. At this point, they need to perform a large number of statistics on the effects of different drugs on grooming behavior. At present, the key point tracking technology in the field is to add markers on the body of insects, which may interfere with the action of *Bactrocera minax* and affect the experimental

results. In biomechanics, genetics, behavior, and neuroscience, extracting animal behavior without using markers is usually the key to measure behavior effect [21]. In traditional *Bactrocera minax* behavior studies, researchers mainly rely on manual work to obtain the behavior time parameters and observe, analyze, and record the times of each behavior by playing videos frame by frame [22], which is a very time-consuming, labor-consuming, boring process, and the relativity is extremely low. On the contrary, the progress of computer vision constantly inspires data analysis methods to reduce manual labor [23–27]. Computer vision and deep learning have been widely used in different areas, such as medical diagnosis [28], detection of COVID-19 [29], defect detection in the industry [30], classification of crop pests in agriculture [31], face recognition in life [32], and automatic driving of car application [33]. Technologies bring us a lot of convenience. Domestic and international research on animal behavior classification technology and target tracking technology has been strengthened, and certain progress has been made [34–37]. We have tried some deep learning algorithms, such as ABRS, which can only find the time point of different behavior switching in a complex environment but cannot track target or recognize the accurate grooming behavior [38]. Another deep learning algorithm is DeepLabCut, which can track the target key points but cannot solve the behavior classification problem [32]. Thus, it is not suitable for our environment. So we want to integrate the characteristics of these two algorithms to invent a new algorithm, which can not only accurately track key points but also identify and classify insect behavior.

The main purpose of this paper is to develop an algorithm that can automatically track the trajectory of *Bactrocera minax* key points and accurately identify its grooming behavior, so as to get rid of the invasive and artificial analysis process of *Bactrocera minax* grooming behavior. We used the DeepLabCut open-source toolkit and optimized it further for our specific needs. We extended the training data by data enhancement technology, chose the residual network (ResNet-50) as the backbone network, identify the key points of *Bactrocera minax* using optimized DeepLabCut algorithm, and filter the abnormal jitter points. Then we detected and classified the grooming behavior through the relative position relationship of the key points. Finally, 14 key points tracking, grooming behavior recognition, and time interval statistics of *Bactrocera minax* were realized.

2. Materials and Methods

2.1. Experimental Equipment and Environment. The computer hardware device GPU used in this experiment is an NVIDIA 1660Ti, CPU is Intel Core i5-9300h, 16 GB ram; software development environment is Python 3.6, using TensorFlow GPU 1.14.0, using the open-source toolbox DeepLabCut 2.1.5, mainly using libraries such as NumPy 1.17.3, OpenCV 3.4.5, and Matplotlib 3.1.1.

The grooming behavior videos were recorded by Sony camera (Sony, hxr-mc58c) of 5 adult *Bactrocera*. The shooting time was from May to August in 2019. It was a captive laboratory colony. The detailed culture conditions are illustrated in this paper: feeding behavior of *Bactrocera minax* (Diptera: Trypetidae) on male inflorescence of *Castanea mollissima* (Fagales: Fagaceae) [22]. The shooting resolution is 1920×1080, and the frame rate is 25 frames/s. The size of the culture dish (35×20 mm) allows the *Bactrocera minax* to move freely in the dish.

2.2. Experimental Methods

2.2.1. Data Acquisition and Processing. High-definition camera was used to record the adult video of *Bactrocera minax* (Figure 1(a)). We manually capture the key frames in the original video and increased the proportion of pixels occupied by the target. It meant that the screenshots were taken with as little background as possible and then use it as training data of neural network. We can also automatically capture the key frames through the DeepLabCut open-source toolbox to save time and improve efficiency [21], but the effect of model training will not be so good. Our data preprocessing was divided into four processes, as shown in Figure 1(b). In the process of image data filtering, we removed the samples with fuzzy, obviously noisy image data, and saved samples that contain most of the key points. In order to make the training model more accurate, the methods of flipping, changing scale, and changing contrast were used to expand the original image to generate similar image data in the process of image data expansion [39]. In the last step of data annotation, we marked the key points to be detected (Figure 1(c)), including head, left antennae, right antennae, left side of the body, the right side of the body, left forefoot, right forefoot, left middle foot, right middle foot, left hindfoot, right hindfoot, left wing, right wing, and head.

2.2.2. Training Model. To meet the needs of the insect behavior researcher, we choose the suitable backbone network, adjust model hyperparameters, optimize DeepLabCut, and design behavior recognition algorithms.

Bactrocera minax's key points are difficult to identify because of their small joints and fast movements. In order to improve the recognition accuracy, we deepen the number of neural network layers to extract more feature information. However, as the number of network layers is too much, the analysis rate was reduced. The backbone of the neural network we choose is ResNet-50 [21, 40] (Figure 2). The analysis speed of neural network in the same depth is greatly accelerated, and the performance degradation caused by network deepening was effectively solved [40].

At this time, we divided the labeled images into a training set and a validation set at a ratio of 8 to 2, and then through the neural network, the batch size is 8. The training adopts batch processing, and multiple images were processed at a time, so the size of the pictures should be consistent. If the size of the input image does not meet the required size, we will shrink the shorter side randomly to between 256 and 480 pixels, and the long side and the short side were scaled in equal proportion to ensure the length-to-width ratio of the image remains unchanged, and then, it was

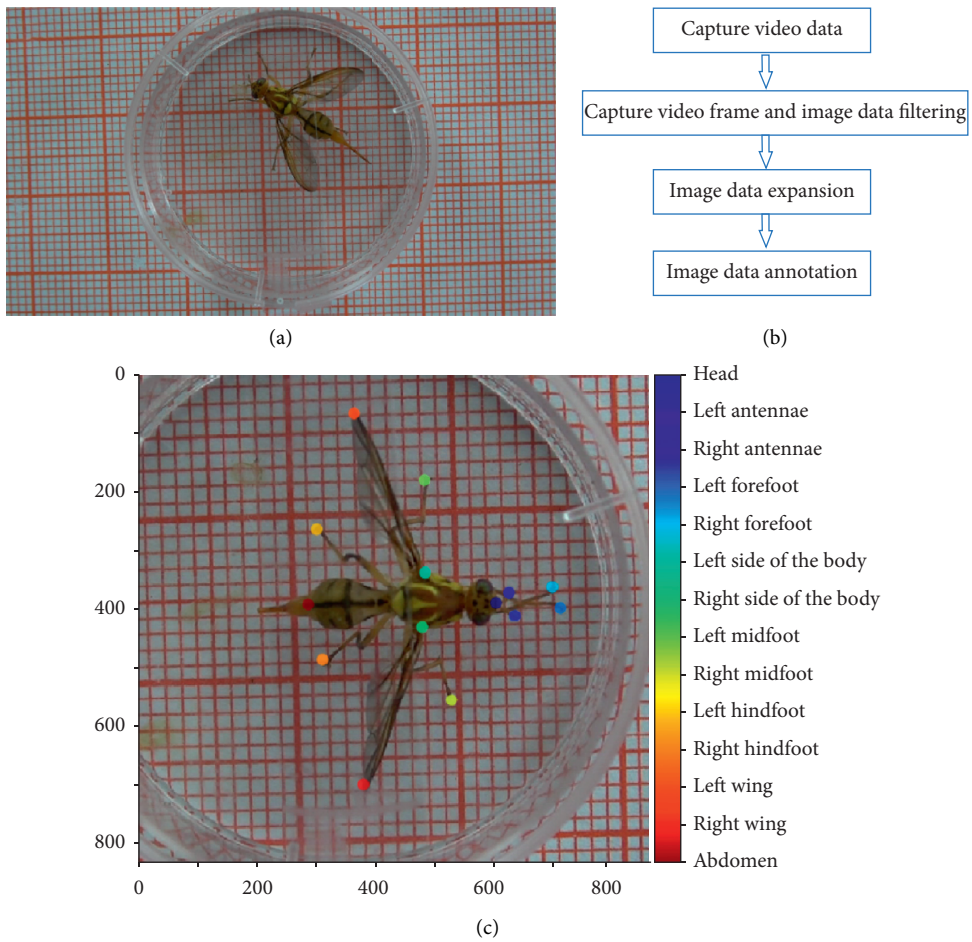


Figure 1: Acquisition and processing of data of adult *Bactrocera minax*. (a) Experimental recording environment of *Bactrocera minax*, grid size = 1 mm. (b) Data preprocessing process. (c) We mark the key points of the intercepted images, and the key points are selected from the experimental parts.

randomly cut into multiple $224 * 224$ pictures. The loss function is Huber loss:

$$\text{Huber loss} = \begin{cases} 0.5 * \sum_{1}^{n} X^2, & \text{当} |X| \le k, \\ k * \sum_{1}^{n} |X| - 0.5 * k^2, & \text{当} |X| > k. \end{cases} \quad (1)$$

The loss value is calculated by forwarding propagation, and the weight and bias are updated by backpropagation to achieve the purpose of training parameters, where k is set to 1.

Using the method of migration learning [41–43], the parameters of the ImageNet dataset pretraining model are taken as initialization parameters. ImageNet dataset includes more than 20,000 class targets, and more than 14 million image URLs are manually annotated by ImageNet. Using the ImageNet dataset as input and training weight as initialization parameter can get better weight parameters faster. Figure 3 shows the process of model training and video analysis.

Then, the annotation dataset in Section 2.2.1 is used as input to the neural network. If the training result always fails to achieve the specified accuracy, it is necessary to expand the dataset and adjust the training hyperparameters. After a large number of experiments, we set the learning rate of the hyperparameters as a ladder, the first 10,000 times of iterative learning rate is 0.005, the 10,000 to 400,000 times learning rate is set to 0.002, and the latter learning rate is set to 0.001.

We evaluate the training model (the results of the model evaluation are shown in Figure 4(a)) and optimize the model. First, the model is used to analyze the video, extract the wrong prediction frames, correct the key points of the error detection, manually move to the correct position, and train again. Then we can get a better and more accurate model by repeating the above training process. After repeating iterative training and parameter adjustment, we finally get a better model. We input the video into this trained model to obtain the coordinates and probability (confidence level) each key point and save this information in the CSV table, which is used to the key points tracking, grooming behavior recognition, and statistics of *Bactrocera minax* in Section 2.2.3. If the *Bactrocera minax* move quickly, the camera will not be able to accurately capture the location of the key points, and the image will be blurred,

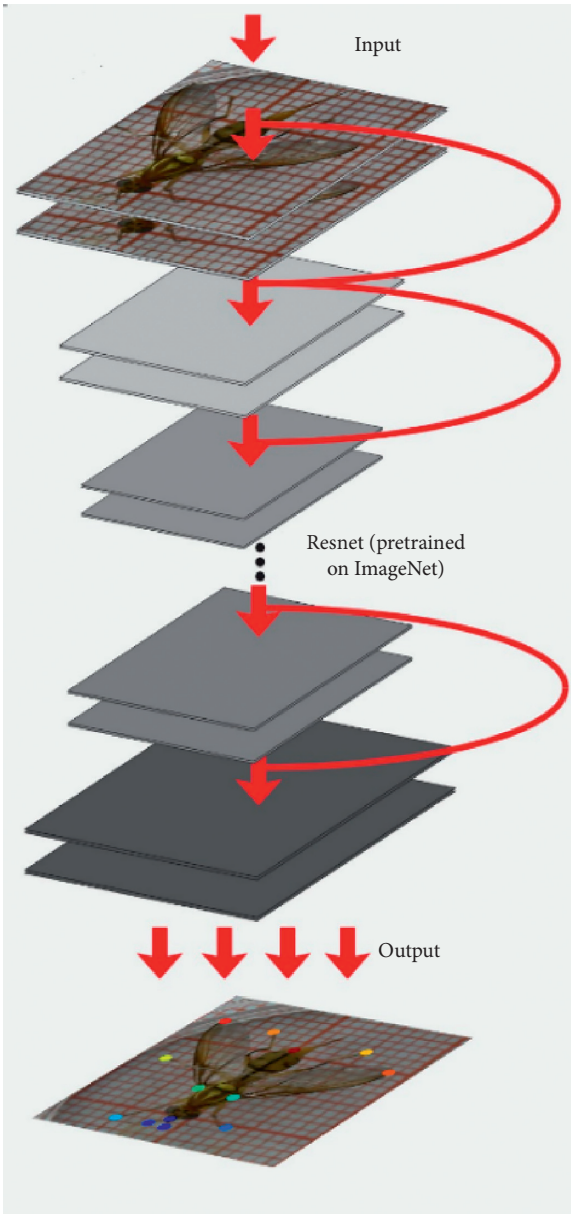


Figure 2: Train a deep neural network (DNN) architecture to predict the body part locations on the basis of the corresponding image.

such as the wing in Figure 4(b). Another possibility is that the key points are hidden behind the body, such as the left hindfoot in Figure 4(c).

2.2.3. Keypoints Tracking, Grooming Posture Recognition, and Statistics of Bactrocera minax. On the premise of tracking key points with high accuracy, the method used in this paper classifies grooming behavior by the location relationship of key points. At this time, we found out the relationship between behavior and key points through long-term observation of giant *Bactrocera minax* behavior. Because the culture dish is placed vertically, the citrus fly cannot groom its body with its forefoot and hindfoot at the same time; otherwise, it cannot be adsorbed on the culture

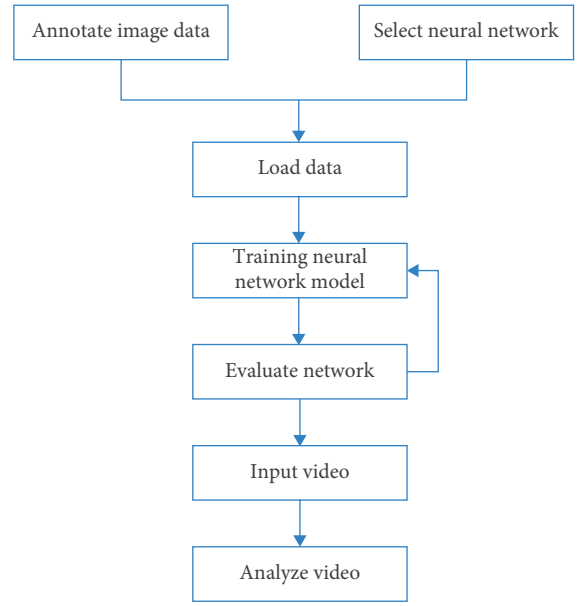


Figure 3: Process of model training and video analysis.

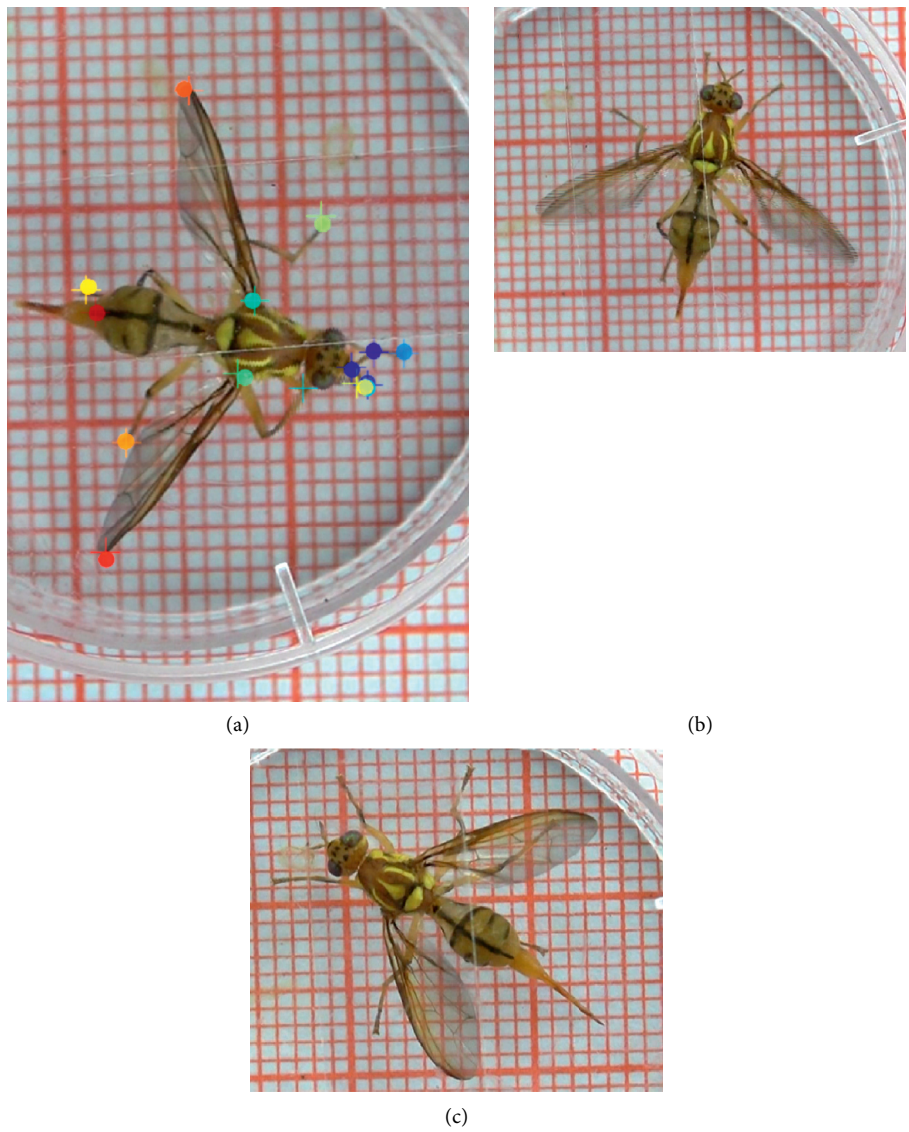
dish. Traversing the coordinates of each key point, if the pixel of the forefoot coordinate changes and the hindfoot remains unchanged, the behavior is divided into the first category (the forefoot grooming, midfoot with the forefoot grooming, and the antennae grooming). If the coordinates of the hindfoot change continuously and the forefoot remains unchanged, the behaviors are divided into the second category (the wings grooming, hindfoot grooming, the abdomen grooming, and midfoot with the hindfoot grooming). If the coordinates of forefoot, midfoot, and hindfoot are constantly changing, it is the moving process of *Bactrocera minax*, and the behavior detection is excluded. Then, the grooming behavior of each part is represented by the coordinate relationship of the key points. Figure 5 shows the states of seven grooming behaviors of machine analysis.

Figure 5 shows still images of the seven different grooming behaviors. The seven behaviors in Figure 6 correspond to those in Figure 5 one by one, reflecting more clearly the coordinate relationship and movement law of each carding behavior.

After a lot of experiments and comparison with the original video, we set the confidence level of 0.8 as the detection threshold. When it is lower than 0.8, the coordinates are not accurate. But we still want to find the slightly correct coordinates of these key points. We use median function to eliminate noise:

$$X_5 = \frac{X_0 + X_1 + X_2 + X_3 + X_4 + X_6 + X_7 + X_7 + X_8 + X_9}{10}. \tag{2}$$

Although the coordinates of these key points are not accurate, the motion is continuous, and blur or occlusion of a single image does not affect the judgment. So we can predict the action in this frame as the same as that in the previous frame. We analyze the behavior of each frame and then save the frame number in the list, and each different



(a) (b) (c)

Figure 4: The results of model evaluation and two cases of inaccurate positioning of key points. (a) The cross is the result of manual marking and the circle is the result of machine analysis. (b) The rapid movement of the *Bactrocera minax* causes its wings to blur. (c) The left hindfoot of the *Bactrocera minax* is covered by the body.

behavior is stored in its own list. If it is detected that the coordinate relationship of each key point in a frame does not meet any of the above classification rules, the system cannot predict the behavior at this time. It will automatically play the first 10 frames and the last 10 frames of the current frame and help the system predict the frame through the staff confirmation, so as to ensure the correct classification of the behavior.

Figure 7 shows a simple illustration of the general process of our experiments.

The *Bactrocera minax* video recorded by Sony camera (Sony, hxr-mc58c) is transferred to the server, and the video data are loaded into the neural network model. We analyze the video through the trained model to get the coordinates of each key point. All frame numbers of the video are stored in different behavior lists through coordinate relationship. Read these lists, there may be some similar grooming behaviors leading to classification errors. So the list may miss some frames or store the frame number of other behaviors. In order to further reduce the error, we need to eliminate some error detection, so we use the following methods to reduce the error of statistical time (Figure 8).

The black square is the frame number stored in the list (correct frame detection), and white squares are the frame number that are not stored in this list. Let us assume that the video has 27 frames, among them, grooming behavior X occurred in 1–26 frames. So correct frame detection is 1, 2, 4, 6, 8, 11, 12, 14, 16, 18, 20, 22, 24, and 26, and missed detection is 3, 5, 7, 9, 10, 13, 15, 17, 19, 21, 23, and 25. There is no behavior at frame 27. We need to recover frames that were not detected correctly. So we create a window with a size of 25. Starting from the first frame, we can see that 25 frames are selected in the red window (1–25 frames, 25 frames is 1 second, which depends on the number of frames

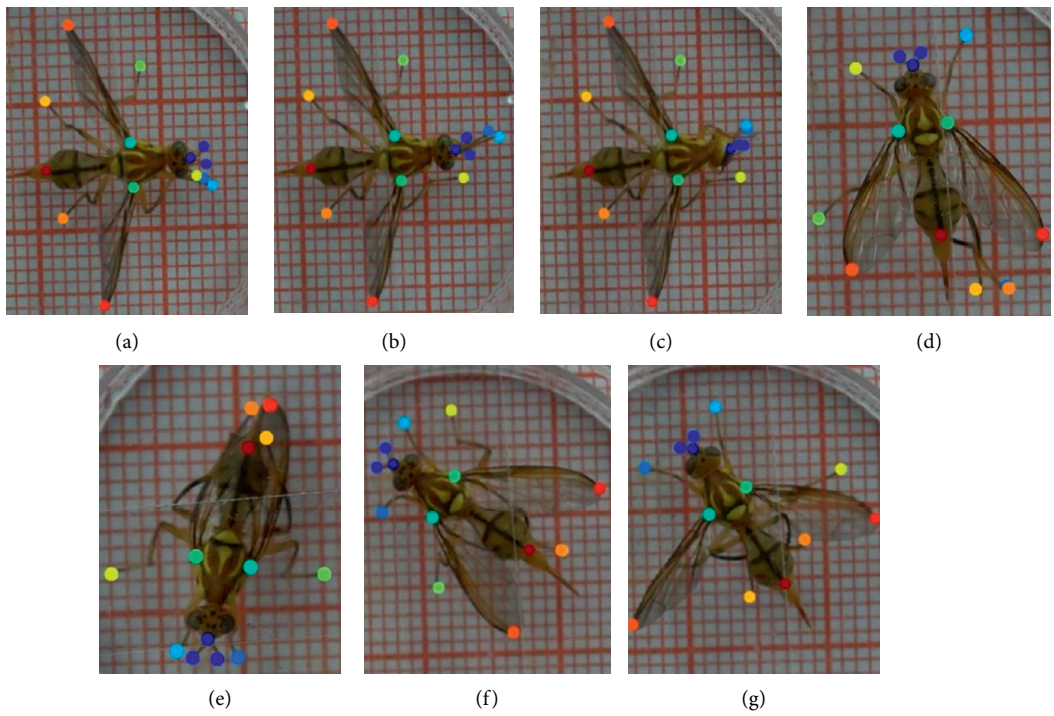


Figure 5: Seven grooming posture of *Bactrocera minax*. (a) Forefoot and midfoot reciprocal grooming. (b) Forefoot grooming. (c) Antennae grooming. (d) Midfoot and hindfoot reciprocal grooming. (e) Wing grooming. (f) Hindfoot grooming. (g) Abdomen grooming.

of video recording). We assume that more than half of these 25 frames are classified as the behavior X. In other words, there are more than 13 black squares in the red window. We judge that the behavior X occurred in 1–25 frames and the detection is correct; otherwise, the detection is wrong. When the detection is correct, we record the first of 25 frames as the beginning time of the behavior. Next, we need to find the end time of the behavior. To avoid missing any frame, we use the sliding window method (stride = 1) to analyze the next 25 frames (2–26) from the second frame and then the next 25 frames (3–27) from the third frame and so on. At this time, we have saved start time. If 2–26 frames are also detected correctly, the start time is still the first recorded frame, and no change is made until it is detected incorrectly. If the detection is not correct, in other words, there are less than 13 black squares in the red window. Obviously, we can see that on the third detection (3–27), there are only 12 black squares in the red window. The condition of correct detection is not satisfied. So the last frame of these 25 frames is not grooming behavior X. Now we will use the last frame of behavior X (the last black square, frame 26) as the end time of the behavior. In this way, we find the start and the end time of behavior X, which is the exact time period when behavior X first appears in the video. Similarly, if the list stores hundreds of frame numbers instead of just these 27 frames, then we start from the end frame (frame 27), repeat the above operation, and find the start frame and end frame again. In this way, we can find the specific time of all grooming behaviors X in the video. Similarly, read the other list, classify all behaviors, and record the time. We also invite entomologists to help us manually count the duration of each grooming behavior in the video. They observe, analyze, and record the times of each behavior by playing videos frame by frame. Finally, they count the duration of each grooming behavior, and the time unit is seconds. Then we calculate the accuracy (equation (3)) and difference (equation (4)) between the two methods:

$$\text{accuracy} = \frac{\text{time of method statistics in this paper}}{\text{time of manual statistics}}, \tag{3}$$

$$\text{difference} = \frac{\text{time of manual statistics} - \text{time of method statistics in this paper}}{\text{time of manual statistics}}. \tag{4}$$

Figure 9 shows the video analysis process.

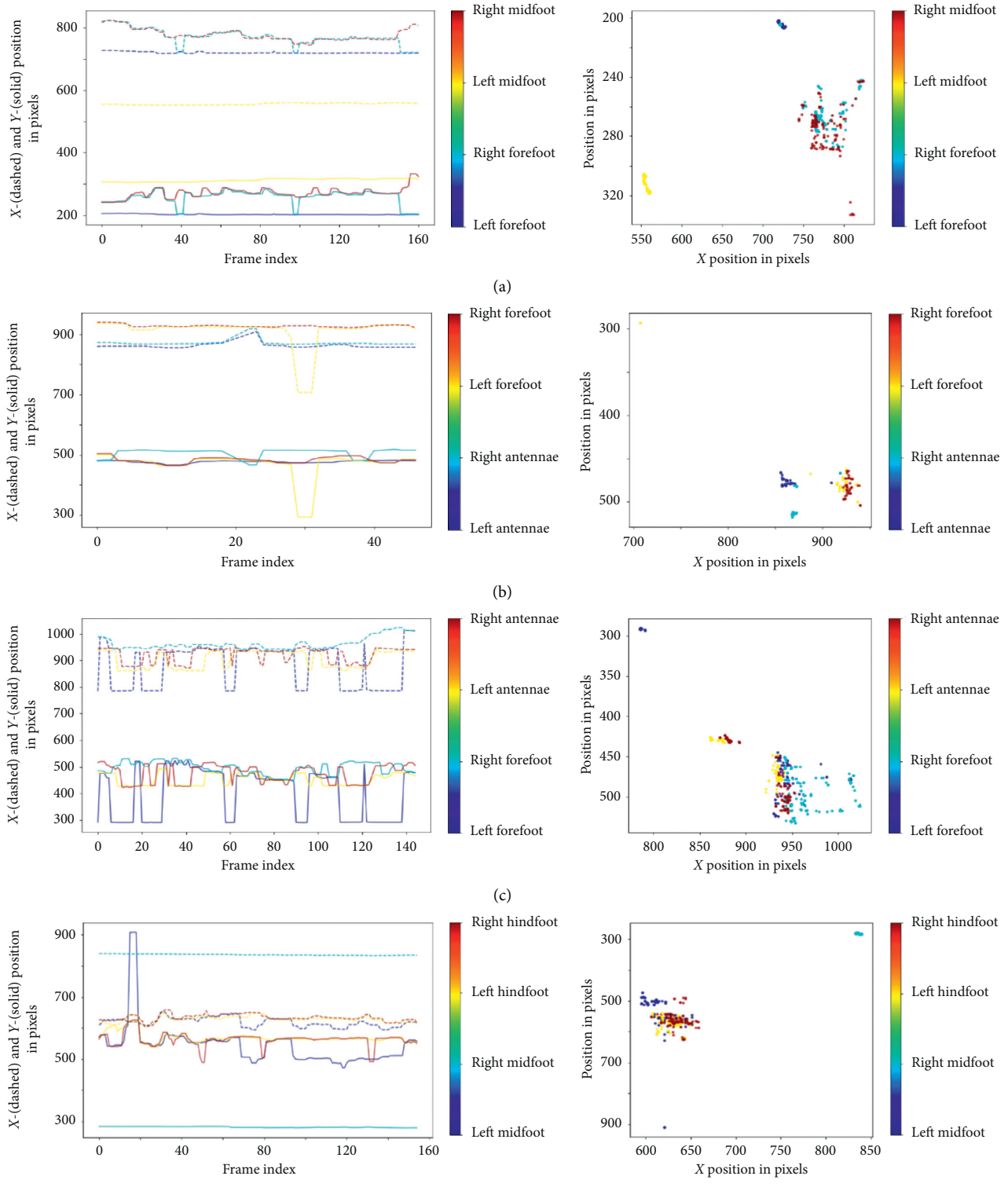


Figure 6: Continued.

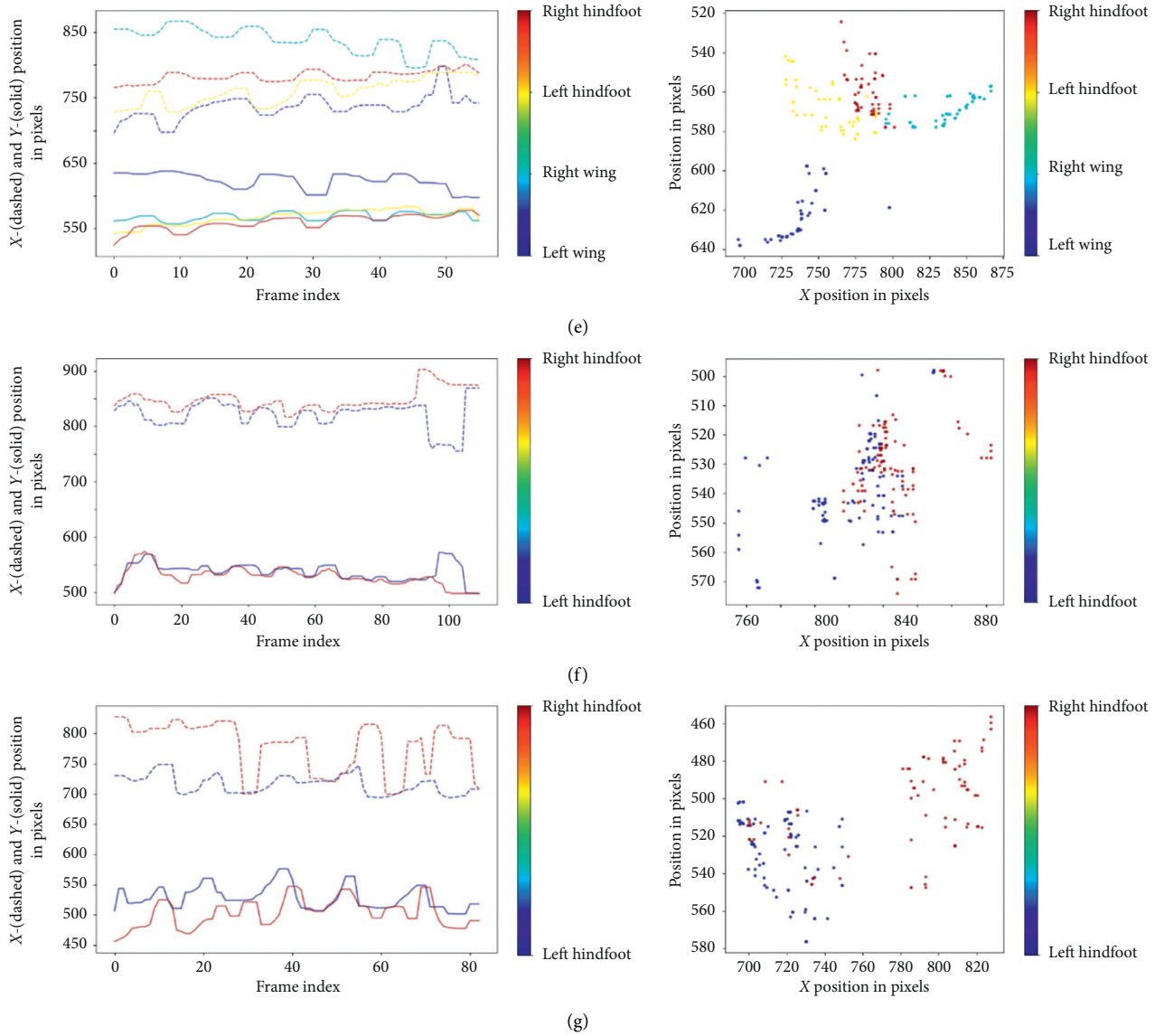


Figure 6: Movement law and coordinate relationship of seven grooming behaviors. (a) In the process of the forefoot and midfoot reciprocal grooming: one midfoot is close to one or more forefeet. (b) In the process of grooming the forefeet: the forefeet are close together. (c) In the process of the antennae grooming: the forefoot and antennae move at the same time and the coordinates change constantly. (d) In the process of the hindfoot and midfoot reciprocal grooming: one midfoot is close to one or more hindfoot. (e) In the process of wing grooming: the coordinates of hindfoot and wing change constantly. (f) In the process of hindfeet grooming: the hindfeet are close together. (g) In the process of the abdomen grooming: hindfeet grooming is similar to abdomen grooming, only hindfeet movement, but in the process of abdomen grooming, hindfeet are not close together.

3. Experimental Results

The video of *Bactrocera minax* collected in the laboratory of oriental *Bactrocera minax* in Agricultural College of Yangtze University in May 2019 was selected as the evaluation sample. The collected videos were 25 frames, with a resolution of $1920 * 1080$. 224 images intercepted from ten videos were marked with "left antennae," "right antennae," "left side of the body," "right side of the body," "left forefoot," "right forefoot," "left midfoot," "right midfoot," "left hindfoot," "right hindfoot," "left wing," "right wing," "head," and "abdomen" with a total of 3136 data. The training results of ResNet-50 are shown in Figure 10. The validation loss value of the first 100,000 cycle training of the network decreases rapidly, and the loss value reduces to 0.00211. Finally, the loss has converged to the specified range of 0.001 passing through 412200 iterations. After 441600 repeated iterations, the loss value tends to be stable, and the final validation loss is 0.00079, which has reached the training goal [21]. We also tried to use another neural network resnet_101 to train. At first, it got better results than ResNet-50. After a lot of iterations, their loss value was almost the same. But it takes more time to detect [37], so ResNet-50 is still a better choice.

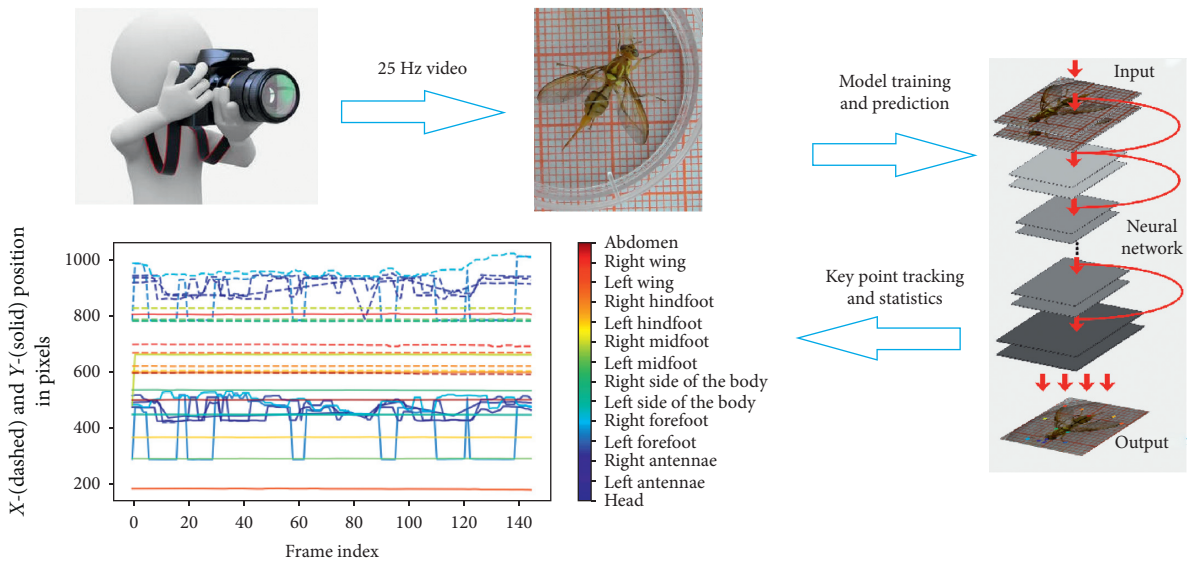


Figure 7: Simple flowchart of key point tracking.

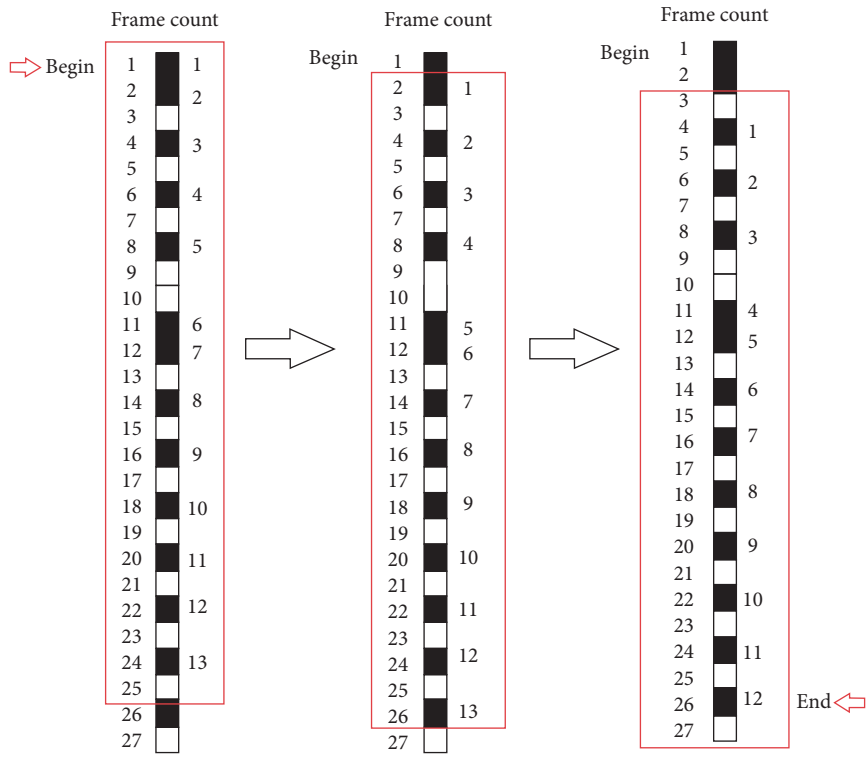


Figure 8: Elimination of false detection and recovery of missed detection.

We used a single 1660ti GPU to analyze the video of *Bactrocera minax* with an average duration of 12 minutes. The running time experiment results are shown in Figure 11, with the unit of minutes.

It can be seen from Figure 11 that, for a video with 1920×1080 pixels and a duration of about 12 minutes, the analysis time of 6G 1660ti GPU is about 45 minutes, and the processing speed is about 7 Hz. When we use hardware with good performance to handle video with different resolutions, for instance, one can process the 682×540 pixel frames at around 30 Hz on an NVIDIA 1080Ti GPU, and low-resolution videos with 204×162 pixel frames are analyzed at around 85 Hz [44]. This means that when we deal with small-resolution video with better hardware, we can process data in real time. For humans, we cannot do the same endless counting as machines because the process is time-consuming and tedious. Therefore, it is difficult to calculate their exact time. If you exclude breaks, a 12-minute video usually takes about 5 hours, which is 6-7 times longer than a machine.

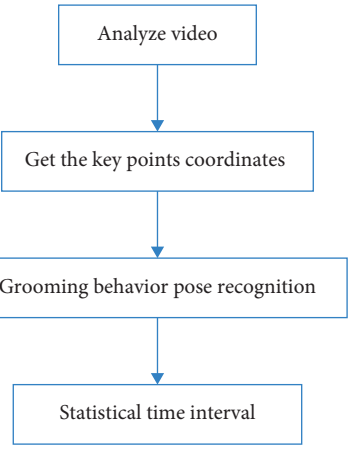


Figure 9: Grooming posture pose recognition and analysis.

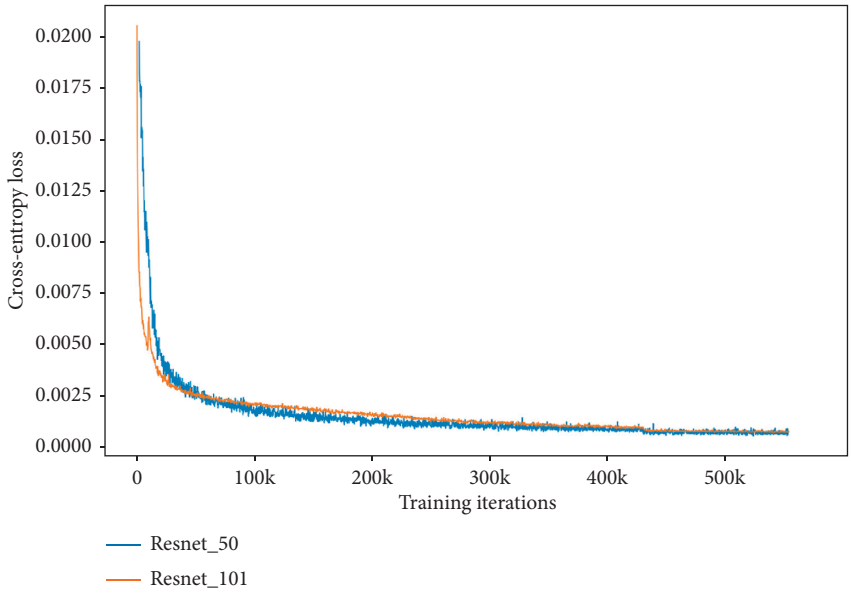


Figure 10: Training results of neural network.

Finally, after analyzing the video with the program, we get the coordinate information of our 14 key points, and the program stores them in the table file. In this experiment, we analyze the data of 94538 frames from 5 videos of no. 01–05 and summarize the confidence level of machine detection of each key point (Table 1).

From Table 1, we can select the coordinates of key points greater than a certain confidence level according to our accuracy requirements and select reliable and satisfactory data. In this experiment, we need to count the specific time interval of the grooming behavior of *Bactrocera minax* and have high requirements for the accuracy of the key points. We can see that the confidence level only decreases by 0.261% when it is increased from 0.9 to 0.95, while the accuracy decreases by 0.855% when it is increased from 0.9 to 0.99, which is a larger decrease. This means that we will discard more key point information, of which perhaps most of the key point detection is accurate, and 0.95 confidence is already a very high value. Therefore, we choose the experimental threshold with a confidence level greater than 0.95. The recognition rate of the key points whose confidence level is greater than 0.99 has an overall target tracking accuracy of more than 96.720%. Figure 12 shows the detailed video detection results corresponding to each adult fly whose confidence level is greater than 0.99.

In similar papers on machine learning and deep learning algorithms, we use the same video data for testing, and their accuracy rates of tracking key points are shown in Table 2. We have achieved 95.2% accuracy with the original DeepLabCut algorithm [44]. In fast animal pose recognition using deep neural networks, we use the same training data and it results in 94.6% of peak performance [45].

There is no suitable algorithm to complete similar tasks in the classification and statistics of the grooming behavior of *Bactrocera minax*. Therefore, we compare the results of machine prediction in our manuscript with the results of expert manual statistics. Table 3 records the statistical results of the above two methods.

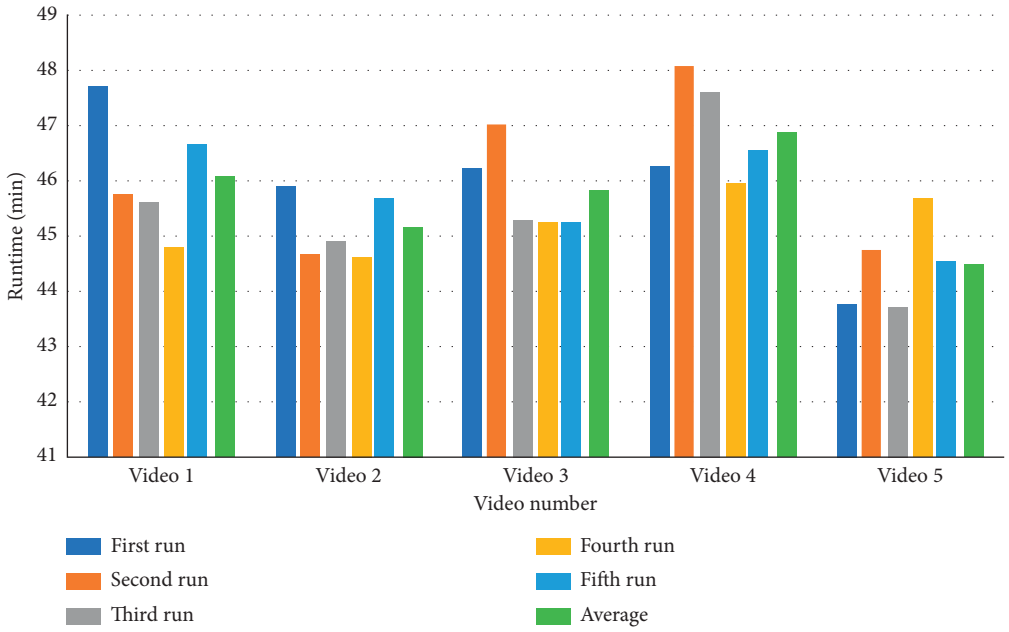


Figure 11: Statistics of computer analysis video time consumption.

Table 1: Confidence level of key points.

Key points	The confidence level is lower than 0.8 (%)	The confidence level is greater than 0.9 (%)	The confidence level is greater than 0.95 (%)	The confidence level is greater than 0.99 (%)
Head	0.181	99.730	99.666	99.490
Left antenna	1.355	98.374	98.129	97.618
Right antenna	0.983	98.750	98.542	97.888
Left forefoot	3.243	96.365	96.024	95.287
Right forefoot	2.062	97.588	97.264	96.442
Left side of body	0.109	99.850	99.828	99.774
Right side of body	0.186	99.741	99.686	99.540
Left middle foot	4.405	95.109	94.651	93.408
Right middle foot	3.628	95.906	95.689	95.305
Left hindfoot	2.938	96.198	95.327	93.606
Right hindfoot	6.990	92.414	91.825	90.630
Left wing	2.207	97.803	97.697	97.433
Right wing	1.431	98.404	98.268	97.954
Abdomen	0.116	99.829	99.800	99.709
Average value	2.131	97.57	97.314	96.720

4. Discussion

In terms of computer vision, a lot of machine learning image processing models and deep learning models have been developed to accurately classify and identify crop pests [46, 47] and a key points tracking and grooming behavior recognition system of *Bactrocera minax* has been developed inspired by human behavior estimation [48]. The accuracy of traditional machine vision classification mainly depends on the input features. The extracted features are mainly shape, color, texture, and so on, which are usually made by hand, and the original data are converted into feature vectors [49–51]. Some researchers have classified some behaviors of mice, chickens, and other animals [52–54], and they have a good effect on large and obvious targets. There is no good solution for small targets such as *Bactrocera minax* because its body color is similar and the characteristics are not obvious. Therefore, like many researchers, we improve the accuracy rate by applying the deep learning model [55], which is a method to automatically extract abstract features [56] without the need of manual operation. However, in the actual environment, due to the small target and poor video quality of the *Bactrocera minax*, many open-source models cannot meet the needs of the scene. For example, YOLOv4 [57], which was recently published this year, has a high accuracy rate in target detection and is widely used, but it still cannot accurately detect, identify, and track the small key points of *Bactrocera minax*. DeepLabCut [44] can

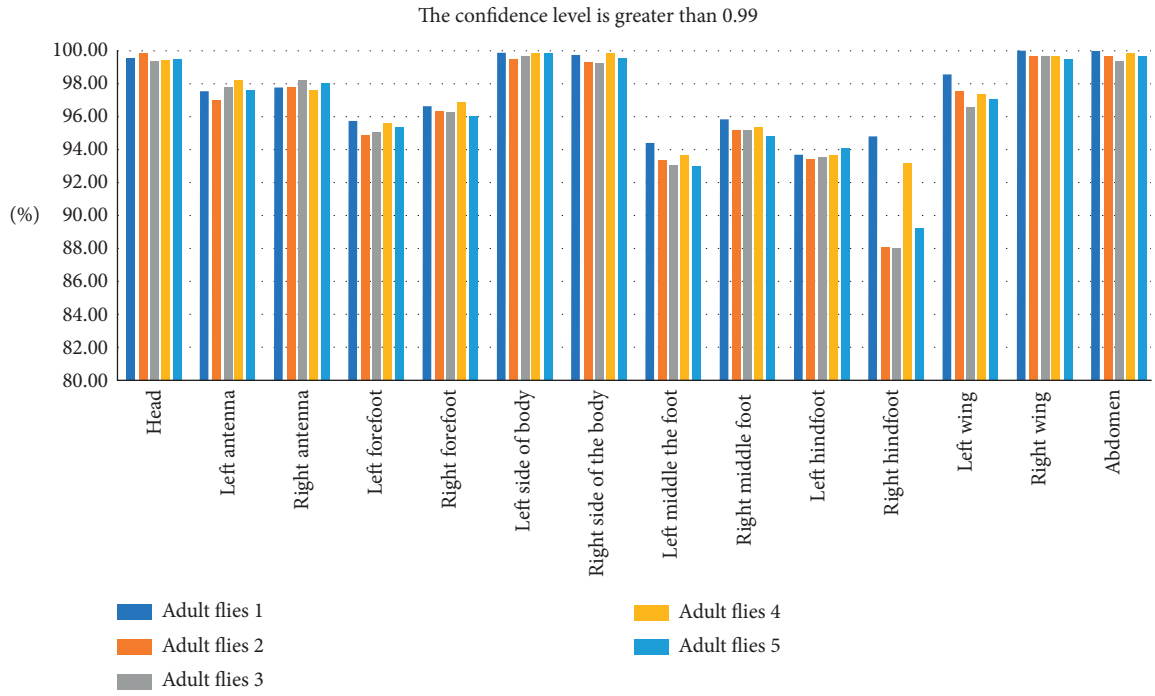


Figure 12: The detailed video detection results corresponding to each adult fly.

Table 2: Comparison of several detection methods.

Methods	Test set accuracy (%)
Original DeepLabCut algorithm	95.2
LEAP algorithm	94.6
Methods of this paper	96.7

manually select the key points needed for the experiment to mark and then track the movement of the target key points, but it is unable to classify and record the behavior changes of *Bactrocera minax*. In view of their shortcomings, the method adopted in this paper classifies the grooming behavior through the location of key points on the premise of tracking key points with high accuracy.

This experiment is a noninvasive key point tracking of *Bactrocera minax*. At present, the key point tracking technology in the field is to add markers on the body of insects, which may interfere with the action of *Bactrocera minax* and affect the experimental results. In biomechanics, genetics, behavior, and neuroscience, extracting animal behavior without using markers is usually the key to measure behavior effect [21]. The above experiments and the experimental results show that we have successfully and accurately tracked the movement track of *Bactrocera minax*, and the key points of its body parts using the computer with an overall accuracy rate are of over 96.7% (the confidence level is greater than 0.99). Compared with the original DeepLabCut algorithm [44] with target tracking accuracy rate of 95.2% and deep learning algorithm with LEAP [45] accuracy rate of 94.6%, the accuracy of the method used in this paper has been improved. In the statistical analysis of grooming behavior, the average accuracy rates of forefoot and midfoot reciprocal grooming, midfoot and hindfoot reciprocal grooming, hindfoot grooming, wing grooming, and abdomen grooming are all above 96%, and the accuracy rates of antennae and forefoot grooming are above 87%. The accuracy of classification decreased significantly for the activities of antennae grooming and forefoot grooming. The possible reasons are as follows: (1) the grooming behaviors of antennae and forefoot are all in a small area of the head, and these key points of Bactrocera are too close and crowded, leading to the improvement of error detection. (2) There is no obvious distinction between grooming behavior itself. (3) It is possible that the duration of grooming behavior is so short and the switching frequency is so high that these behaviors only last for a few frames, which leads to an increase of misjudgment. (4) The definition and frame rate of the video recorded by video equipment are low, which reduces the analysis performance of the computer.

In the future, we will continue to optimize our method by improving the real-time performance of the algorithm, shortening the analysis time, and reducing the requirements of the algorithm for the hardware equipment through the model quantization [58, 59] and the use of MobileNet lightweight neural network [60]; by improving the accuracy of deep neural network model and grooming behavior classification within the head range, and reducing the false detection rate of behavior classification through expanding data, collecting more data of *Bactrocera minax* in different environments; by reducing the number of blocked key points through multidimensional video recording of *Bactrocera minax*; by overcoming self-occlusions and sensor noise problems [61, 62]; and further improving the accuracy of key points recognition through leverages multiview RGB-D data and self-supervised, data-driven learning algorithms [63].

Table 3: Classification accuracy of individual behaviors.

Grooming behavior	Machine prediction (s)	Expert statistics (s)	Accuracy (%)	Difference degree
Antennae grooming	78.24	89.24	87.67	0.1232
Forefoot grooming	132.96	151.64	87.68	0.1231
Forefoot and midfoot reciprocal grooming	573.28	592.42	96.77	0.0323
Hindfoot grooming	228.80	235.84	97.01	0.0298
Wing grooming	468.44	474.48	98.73	0.0127
Abdomen grooming	304.40	316.64	96.01	0.0386
Midfoot and hindfoot reciprocal grooming	243.24	252.40	96.37	0.0284

5. Conclusion

In this paper, data enhancement technology is used to expand the training data, and the key points of *Bactrocera minax* are identified using the DeepLabCut toolbox and deep residual network. The grooming behavior is judged by the relative position of the key points. The tracking of key points and behavior recognition and statistics of grooming behavior are realized. The experiment shows that this method can effectively track the key points of *Bactrocera minax*, and the accuracy rate is more than 96.7%. Compared with other methods, the accuracy rate has been further improved. This method can recognize the grooming behavior of *Bactrocera minax* and count its duration. The average accuracy is more than 94%, which greatly reduces the statistical time of manual observation within the allowable error range. This paper also puts forward several specific methods to further improve the accuracy and efficiency of system detection.

Data Availability

The data used to support the findings of this study are available at https://github.com/ZZl0897/BehaviorDetection.

Conflicts of Interest

The authors declare no conflicts of interest.

Authors' Contributions

W.Z. and Y.Z. conceptualized the study and were involved in project administration; Z.H. and Y.Z. performed data curation and were responsible for methodology; Y.Z. and Z.Z. were responsible for software; W.Z. and Z.H. contributed resources; Y.Z., W.Z., and Z.Z. performed formal analysis; W.Z., Z.H., and Y.Z. supervised the study; Z.H. and Y.Z. validated the data; ; Y.Z. prepared the original draft; W.Z., Y.Z., Z.H., and Z.Z reviewed and edited the manuscript.

Acknowledgments

The authors are thankful to Zhangzhang He (Insect Ecology Laboratory, College of Agriculture, Yangtze University, Jingzhou, China) for commenting on earlier version of the manuscript. This research was funded by the China University Industry-University-Research Innovation Fund "New Generation Information Technology Innovation Project" (2019ITA03004) and the National Natural Science Foundation of China (31772206 and 31972274).

References

[1] W. Jin-Tao, D. Yong-Cheng, L. I. Zong-Kai et al., "Overview of the use of the sterile insect technique to control the Chinese citrus fruit fly," *Chinese Journal of Applied Entomology*, vol. 50, 2013.

[2] Y. Ke-Xi, Z. Qiong, and J. Qi, "Feeding, mating and oviposition behaviours of the adults of Bactrocera minax enderlein," *Journal of Natural ence of Hunan Normal University*, vol. 35, 2012.

[3] A. Yanagawa and S. Shimizu, "Defense strategy of the termite, Coptotermes formosanus Shiraki to entomopathogenic fungi," *Japanese Journal of Environmental Entomology and Zoology*, vol. 16, pp. 17–22, 2005.

[4] J. P. Galvanho, M. P. Carrera, D. D. O. Moreira, M. Erthal Jr., C. P. Silva, and R. I. Samuels, "Imidacloprid inhibits behavioral defences of the leaf-cutting ant acromyrmex subterraneus subterraneus (Hymenoptera: Formicidae)," *Journal of Insect Behavior*, vol. 26, no. 1, pp. 1–13, 2013.

[5] M. S. Mooring, D. T. Blumstein, and C. J. Stoner, "The evolution of parasite-defence grooming in ungulates," *Biological Journal of the Linnean Society*, vol. 81, no. 1, pp. 17–37, 2004.

[6] K. Böröczky, A. Wada-Katsumata, D. Batchelor, M. Zhukovskaya, and C. Schal, "Insects groom their antennae to enhance olfactory acuity," *Proceedings of the National Academy of Sciences*, vol. 110, no. 9, pp. 3615–3620, 2013.

[7] R. W. Phillis, A. T. Bramlage, C. Wotus et al., "Isolation of mutations affecting neural circuitry required for grooming behavior in Drosophila melanogaster," *Genetics*, vol. 133, no. 3, pp. 581–592, 1993.

[8] N. F. Carlin, B. Holldobler, and D. S. Gladstein, "The kin recognition system of carpenter ants (Camponotus spp.)," *Behavioral Ecology and Sociobiology*, vol. 20, pp. 219–227, 1986.

[9] M. Ozaki, A. Wada-Katsumata, K. Fujikawa et al., "Ant nestmate and non-nestmate discrimination by a chemosensory sensillum," *Science*, vol. 309, no. 5732, pp. 311–314, 2005.

[10] W. Rath, "Co-adaptation of Apis cerana fabr. and varroa jacobsoni oud," *Apidologie*, vol. 30, no. 2-3, pp. 97–110, 1999.

[11] D. G. Boucias and J. C. Pendland, *Principles of Insect Pathology*, Kluwer Academic Publisher, Boston, MA, USA, 1998.

[12] D. Kovac and U. Maschwitz, "Secretion-Grooming in aquatic beetles (Hydradephaga): a chemical protection against contamination of the hydrofuge respiratory region," *Chemoecology*, vol. 1, no. 3-4, pp. 131–138, 1990.

[13] A. Yanagawa and S. Shimizu, "Resistance of the termite, Coptotermes formosanus Shiraki to Metarhizium anisopliae due to grooming," *BioControl*, vol. 52, no. 1, pp. 75–85, 2007.

[14] I. Lusebrink, K. Dettner, and K. Seifert, "Stenusine, an antimicrobial agent in the rove beetle genus stenus (coleoptera, staphylinidae)," *Naturwissenschaften*, vol. 95, no. 8, pp. 751–755, 2008.

[15] Y.-S. Peng, Y. Fang, S. Xu, L. Ge, and M. E. Nasr, "The resistance mechanism of the Asian honey bee, Apis cerana Fabr., to an ectoparasitic mite, Varroa jacobsoni Oudemans," *Journal of Invertebrate Pathology*, vol. 49, no. 1, pp. 54–60, 1987.

[16] C. M. Vincent and S. M. Bertram, "Crickets groom to avoid lethal parasitoids," *Animal Behaviour*, vol. 79, no. 1, pp. 51–56, 2010.

[17] M. Zhukovskaya, A. Yanagawa, and B. Forschler, "Grooming behavior as a mechanism of insect disease defense," *Insects*, vol. 4, no. 4, pp. 609–630, 2013.

[18] F. E. Vega and H. K. Kaya, *Insect Pathology*, Academic Press, San Diego, CA, USA, 2nd edition, 2012.

[19] R. R. James and J. Xu, "Mechanisms by which pesticides affect insect immunity," *Journal of Invertebrate Pathology*, vol. 109, no. 2, pp. 175–182, 2012.

[20] J. A. I. Swanson, B. Torto, S. A. Kells, K. A. Mesce, J. H. Tumlinson, and M. Spivak, "Odorants that induce hygienic behavior in honeybees: identification of volatile compounds in chalkbrood-infected honeybee larvae," *Journal of Chemical Ecology*, vol. 35, no. 9, pp. 1108–1116, 2009.

[21] T. Nath, A. Mathis, and A. C. Chen, "Using DeepLabCut for 3D markerless pose estimation across species and behaviors," *Nature Protocols*, vol. 14, no. 7, p. 2152, 2019.

[22] H. E. Zhang-Zhang, H. Deng-Ke, D. U. Tian-Hua et al., "Feeding behaviour of Bactrocera minax (Diptera: Trypetidae) on male inflorescence of Castanea mollissima (fagales: fagaceae)," *Acta Entomologica Sinica*, vol. 61, no. 4, pp. 458–467, 2018.

[23] A. I. Dell, "Automated image-based tracking and its application in ecology," *Trends in Ecology & Evolution*, vol. 29, pp. 417–428, 2014.

[24] A. Gomez-Marin, J. J. Paton, A. R. Kampff, R. M. Costa, and Z. F. Mainen, "Big behavioral data: psychology, ethology and the foundations of neuroscience," *Nature Neuroscience*, vol. 17, pp. 1455–1462, 2014.

[25] D. J. Anderson and P. Perona, "Toward a science of computational ethology," *Neuron*, vol. 84, pp. 18–31, 2014.

[26] H. Wang, L. Dong, H. Zhou, L. Luo, J. Wu, and Y. Tang, "YOLOv3-Litchi detection method of densely distributed litchi in large vision scenes," *Mathematical Problems in Engineering*, vol. 2021, Article ID 8883015, 11 pages, 2021.

[27] Y. Tang, M. Chen, C. Wang, L. Luo, J. Li, and X. Zou, "Recognition and localization methods for vision-based fruit picking robots: a review," *Frontiers in Plant Science*, vol. 11, p. 510, 2020.

[28] M. Bakator and D. Radosav, "Deep learning and medical diagnosis: a review of literature," *Multimodal Technologies & Interaction*, vol. 2, no. 3, 2018.

[29] T. B. Alakus and I. Turkoglu, "Comparison of deep learning approaches to predict COVID-19 infection," *Chaos Solitons & Fractals*, vol. 140, Article ID 110120, 2020.

[30] X. Yin, Y. Chen, A. Bouferguene et al., "A deep learning-based framework for an automated defect detection system for sewer pipes," *Automation in Construction*, vol. 109, no. 2020, Article ID 102967, 2019.

[31] K. Thenmozhi and U. S. Reddy, "Crop pest classification based on deep convolutional neural network and transfer learning," *Computers and Electronics in Agriculture*, vol. 164, Article ID 104906, 2019.

[32] W. Cui, W. Zhan, J. Yu, C. Sun, and Y. Zhang, "Face recognition via convolutional neural networks and siamese neural networks," in *Proceedings of the 2019 International Conference on Intelligent Computing, Automation and Systems (ICICAS)*, pp. 746–750, Chongqing, China, 2019.

[33] A. Miglani and N. Kumar, "Deep learning models for traffic flow prediction in autonomous vehicles: a review, solutions, and challenges," *Vehicular Communications*, vol. 20, 2019.

[34] H. B. Zhang, Y. X. Zhang, B. Zhong et al., "A comprehensive survey of vision-based human action recognition methods," *Sensors*, vol. 19, no. 5, 2019.

[35] S. You, H. Zhu, M. Li et al., "A review of visual trackers and analysis of its application to mobile robot," 2019, https://arxiv.org/abs/1910.09761.

[36] Z. Zhang, W. Zhan, Z. He, and Y. Zou, "Application of spatio-temporal context and convolution neural network (CNN) in grooming behavior of Bactrocera minax (Diptera: Trypetidae) detection and statistics," *Insects*, vol. 11, p. 565, 2020.

[37] C. Sun, W. Zhan, J. She, and Y. Zhang, "Object detection from the video taken by drone via convolutional neural networks," *Mathematical Problems in Engineering*, vol. 2020, Article ID 4806359, 10 pages, 2020.

[38] P. Ravbar, K. Branson, and J. H. Simpson, "An automatic behavior recognition system classifies animal behaviors using movements and their temporal context," *Journal of Neuroscience Methods*, vol. 326, Article ID 108352, 2019.

[39] M. D. Bloice, "Biomedical image augmentation using augmentor," *Bioinformatics*, vol. 35, no. 21, pp. 4522–4524, 2019.

[40] K. He, "Deep residual learning for image recognition," in *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, Las Vegas, NV, USA, 2016.

[41] J. Donahue, "DeCaf: a deep convolutional activation feature for generic visual recognition," in *International Conference on Machine Learning*, vol. 647–655, PMLR, Beijing, China, 2014.

[42] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, "How transferable are features in deep neural networks?" in *Advances in Neural Information Processing Systems*, vol. 3320–3328, Curran Associates, Red Hook, NY, USA, 2014.

[43] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*, vol. 1, MIT Press, Cambridge, MA, USA, 2016.

[44] A. Mathis, "DeepLabCut: markerless pose estimation of user-defined body parts with deep learning," *Nature Neuroscience*, vol. 21, no. 9, pp. 1281–1289, 2018.

[45] T. D. Pereira, D. E. Aldarondo, L. Willmore et al., "Fast animal pose estimation using deep neural networks," *Nature Methods*, vol. 15, 2019.

[46] M. Martineau, D. Conte, R. Raveaux, I. Arnault, D. Munier, and G. Venturini, "A survey on image-based insect classification," *Pattern Recognition*, vol. 65, pp. 273–284, 2017.

[47] S. Ye, S. Lu, X. Bai, and J. Gu, "ResNet-Locust-BN network-based automatic identification of east Asian migratory locust species and instars from RGB images," *Insects*, vol. 11, p. 458, 2020.

[48] Z. Cao, T. Simon, S. E. Wei et al., "Realtime multi-person 2D pose estimation using part affinity fields," 2016, https://arxiv.org/abs/1611.08050.

[49] Q. Cai and D. J. He, "Identification of vegetable leaf-eating pests based on image analysis," *Journal of Computer Applications*, vol. 30, pp. 1870–1872, 2010, in Chinese.

[50] F. Q. Zhang, *Machine Vision and Wavelet Analysis-Based Farmland Pest Identification System*, Zhengzhou University, Henan, China, 2003, in Chinese with English Abstract.
[51] S. H. Lu and S. J. Ye, "Using an image segmentation and support vector machine method for identifying two locust species and instars," *Journal of Integrative Agriculture*, vol. 19, pp. 1301–1313, 2020.
[52] J. B. I. Rousseau, P. B. A. V. Lochem, W. H. Gispen et al., "Classification of rat behavior with an image-processing method and a neural network," *Behavior Research Methods Instruments & Computers*, vol. 32, no. 1, pp. 63–71, 2000.
[53] Z. Min and Z. Heng-Yi, "Automatic recognition of rat's postures based on naive bayes classifier," *Space Medicine & Medical Engineering*, vol. 2, 2005.
[54] F. Lao, G. Teng, J. Li et al., "Behavior recognition method for individual laying hen based on computer vision," *Transactions of the Chinese Society of Agricultural Engineering*, vol. 28, no. 24, pp. 157–163, 2012.
[55] C. Wen, D. Wu, H. Hu, and W. Pan, "Pose estimation-dependent identification method for field moth images using deep learning architecture," *Biosystems Engineering*, vol. 136, pp. 117–128, 2015.
[56] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, 2015.
[57] A. Bochkovskiy, C. Y. Wang, and H. Y. M. Liao, "YOLOv4: optimal speed and accuracy of object detection," 2020, https://arxiv.org/abs/2004.10934.
[58] Z. Cai, X. He, J. Sun et al., "Deep learning with low precision by half-wave Gaussian quantization," in *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, Honolulu, HI, USA, 2017.
[59] J. Wu, C. Leng, Y. Wang et al., "Quantized convolutional neural networks for mobile devices," 2015, https://arxiv.org/abs/1512.06473.
[60] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L. Chen, "MobileNetV2: inverted residuals and linear bottlenecks," in *Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 4510–4520, Salt Lake City, UT, USA, 2018.
[61] A. Zeng, K. T. Yu, S. Song et al., "Multi-view self-supervised deep learning for 6D pose estimation in the amazon picking challenge," in *Proceedings of the 2017 IEEE International Conference on Robotics and Automation (ICRA)*, IEEE, Singapore, 2017.
[62] M. Chen, Y. Tang, X. Zou, K. Huang, L. Li, and Y. He, "High-accuracy multi-camera reconstruction enhanced by adaptive point cloud correction algorithm," *Optics and Lasers in Engineering*, vol. 122, pp. 170–183, 2019.
[63] M. Chen, Y. Tang, X. Zou, Z. Huang, H. Zhou, and S. Chen, "3D global mapping of large-scale unstructured orchard integrating eye-in-hand stereo vision and SLAM," *Computers and Electronics in Agriculture*, vol. 187, Article ID 106237, 2021.