[1]Zhiye Yao

[2*]Mengxing Huang

Machine Vision-based Trajectory Tracking of Honey Bee Crawling Motion

Journal of Electrical Systems

Abstract: - With the maturity of machine vision technology in artificial intelligence and the rise of the concept of "smart agriculture", machine vision-based methods are widely interested in studying biological behavior, but the technology for acquiring biological behavior information is still unsound. In this paper, we propose a machine vision-based bee crawling trajectory tracking method, which automatically records the bee crawling trajectory in the video information and obtains the description of bee crawling behavior. The method establishes the description of bee crawling motion through bee target detection, bee crawling motion tracking and bee crawling motion prediction as a method to study bee crawling behavior through artificial intelligence techniques. The method has the advantages of high timeliness and low labor cost and reduces the problems of traditional research through manual observation records and bioassay methods with high influence of human subjective judgment factors and high application costs.

Keywords: Machine Vision, Trajectory Tracking, Honey Bee, SSD, SORT.

1 Introduction

In recent years, with the iteration and development of machine vision methods in the field of artificial intelligence, the accuracy of target detection, target classification, and target tracking has increased, the cost of research and development and application has decreased, and the target motion track tracking methods based on machine vision have been deployed to a variety of fields. At the same time, the concept of "smart agriculture" in the field of agriculture has triggered the cooperation between artificial intelligence and agriculture to deal with and analyze the problems in the field of agriculture from the perspective of artificial intelligence methods. Machine vision-based target tracking is of wide interest in biological behavior research.

It has been found through research that honey bees, which are biological indicators of environmental change with high economic and environmental value, will show sublethal manifestations when exposed to pesticide environments, especially nerve agent pesticides, such as neonicotinoids. This is characterized by behavioral abnormalities such as the inability to fly, body convulsions, incoherent crawling movements, and other poisoning behaviors. The current methods for studying and determining honey bee poisoning are cumbersome and labour-intensive. For example, in honey bee research, the study of honey bee poisoning usually involves observing the mortality rate for 48 hours or using instruments to test various physiological tests of honey bees to determine the condition and degree of bee poisoning, which requires a lot of labour cost investment and a large human factor in the observation or experimental process; in agricultural honey bee farming, farmers can judge the condition and degree of bee poisoning based on their experience, which also leads to the fact that only professionals can judge. In this paper, we propose a machine vision-based bee crawling trajectory tracking method to obtain a description of the bee crawling motion by recording the bee crawling motion in the video information. Thus, it provides key information for the subsequent determination of whether the bees are poisoned and the degree of poisoning by the change in crawling behavior. This method, compared to human observation and recording, can effectively reduce the influence of human subjective judgment on experimental results and reduce the human cost of judging honey bee behavior, which also has wide application value.

1 College of Information and Communication Engineering, Hainan University, Haikou, P.R. 570228, China, yaozhiye@hainanu.edu.cn
2* College of Information and Communication Engineering, Hainan University, Haikou, P.R. 570228, China, huangmx09@163.com

Major Science and Technology Project of Hainan Province ZDKJ202017-2-2

yaozhiye@hainanu.edu.cn 126@YZYym

hnhkhmx@126.com HMX@19742024

At present, artificial intelligence as a tool to solve agricultural problems is still in the development stage. This paper proposes a machine vision-based tracking of honey bee crawling movement as a method to study honey bee behavior through artificial intelligence technology and provides data support for further analysis of the relationship between honey bee behavior and poisoning behavior subsequently.

2 Related Work

Machine vision technology in the field of artificial intelligence has gradually matured. Methods to study the behavior or movement of bees through artificial intelligence techniques have been emerging. In terms of technical methods, they are mainly divided into target detection, target tracking, and target position prediction. Target detection is used to obtain the position of the target in the image, tracking algorithms are used to track the target based on the obtained position, and the position of the target is predicted for the next period based on the obtained position and the predicted position is corrected to reduce the computational cost of prediction.

Target detection techniques aim at finding the target of interest in the image and determining its classification and position. Therefore, it is divided into One Stage and Two Stages depending on whether the target classification and target position are acquired simultaneously. The famous algorithms for one-stage target detection are the YOLO series and SSD, among which the YOLO series algorithm uses the CNN model to achieve end-to-end target detection, using the whole image as the input of the network, and directly outputs the class of targets contained in the image and the position of each target in the output layer. The SSD algorithm, also known as "Single Shot Detection", uses VGG16 as the base model and adds a convolutional layer to obtain feature maps at different network scales to achieve detection of targets at different scales in the image. Xiuming Guo et al. [2], proposed an adaptive small target recognition algorithm by using YOLOv3 and SSD to compare the effect of small target recognition for bees in complex agricultural environments, and the new algorithm improved the accuracy by 2.6% and 1.8%, respectively, compared to the two models. The feasibility of using YOLO series and SSD algorithms as honey bee target detection methods is likewise demonstrated. The well-known algorithms for second-stage target detection are RCNN, Fast-RCNN, and Faster-RCNN, which are all based on deep learning and use convolutionalnetworks as a method for recognizing targets, and the main idea is to obtain a feature map by convolutional featureextraction of the image, generate candidate regions according to the feature map is a region generation network RPN, use ROI-pooling to map the candidate Finally, according to the ROI output feature vector, the candidate region category is calculated and the final position of the detection frame is obtained by regression. However, since the two-stage method requires more computational resources and consumes more time than the one-stage method, and there is no significant difference in recognition accuracy, it is more worthwhile to use the one-stagetarget detection technique.

The target tracking technique, which can be viewed as a multivariate estimation, for example, given a target state s, s^i_t denotes the i target state at frame t. Five key steps exist in this technique are: state initialization, state processing model, motion model, tracking method, and evaluation metrics. The initialization methods are divided into detection-based automatic tracking (DBT) and artificial non-automatic tracking (DFT). DBT is known as the mainstream of current applications because it has the advantage of being able to associate any target to the tracking trajectory due to the fact that it does not need to set the state of the new target artificially compared to DFT, and it has the feature of detection before tracking. Based on the DBT method， Yang Qingyu et al. [3] proposed an adaptive infrared target size change detection tracking method with size adaptive feature and 10% tracking efficiency compared to the original DBT. The processing modes are divided into online and offline processing, and the main difference is whether to consider the impact of subsequent frame information on the current frame information, so for real-time detection tend to use online processing, and for video detection prefer to use offline processing. To simplify the tracking difficulty, motion models are introduced in target tracking, which analyze the motion behavior from the current target and thus estimate the potential motion position in subsequent frames, thus achieving the effect of reducing the position search space and reducing the tracking difficulty. The motion model is often matched with the tracking method. Currently, the Kalman filter method is commonly used for motion estimation, and the actual detected target position is matched with the predicted target position using Hungarian matching to adjust the prediction effect. The common algorithms are SPRT and SORT. The final indicators to evaluate the tracking effect are MOTA, which is the accuracy of tracking, reacting to the missed detection and false alarm of the target; MOTP, which is the accuracy of tracking, reacting to the degree of matching between the actual and predicted, usually expressed by the intersection and ratio of the detection frame to the prediction frame.

In summary, this paper compares the target detection results of first-order target detection based on SSD algorithm for honey bee; DBT is selected to initialize the target motion state, and SORT target tracking algorithm framework is selected to track the crawling motion trajectory of honey bee; Kalman filter is used as the motion model in SORT framework to predict and update the position of the target, and the Hungarian algorithm is used to calculate the intersection ratio of tracking frame and prediction frame. The Kalman filter is updated according to the matching result. Thus, a description of the bee's crawling motion is obtained based on the video information.

3 Research Methodology

3.1 Dataset

Datasets were collected from collected bees in captive-bred Chinese honey bee hives in Haikou in 2023, and filmed from April to May. The collected collecting bees were placed in crawling observation boxes and video shooting was conducted from 10:00 a.m. to 5:00 p.m. A total of 3071 valid images were obtained for making the training set for target detection, and 10 sets of valid bee crawling videos were used as target tracking processing videos. Bee labeling was performed manually on the bee images. The camera pixel separation rate was 1920×1080 pixels.

The generated original image dataset contains 3071 image information, and the target position in each image is represented as (x, y, w, h), (x, y) is the centre coordinate point, w and h are the width and height of the target frame, respectively. For the original image and then separately after the normalization process to obtain the input image as the input image data set of SSD. In order to improve the recognition accuracy of small targets, this study creates further sub-image data for each image on the normalized input image dataset, i.e., each image is segmentedto obtain a sub-image of the normalized image, which constitutes a sub-dataset of the input image.

The generated original video data set contains 10 video data of 10 seconds duration, each of which has 2 bees moving in a 10 cm diameter circular area. For the original videos, the input images are then normalized separately to obtain the input video data set as SSD. The data image sample is shown below (see Fig.1.).

Fig. 1. Input image example

3.2 Object Detection

Using SSD as a target detector, this detector is built based on VGG-16 network, borrowing the idea of first-order target detection pioneered by YOLO, using the same detection network to output the class of the target and the location of the target, and combining the RPN in Faster-RCNN network for feature fusion of multi-scale convolution results, so as to improve the accuracy of detection. The structure is mainly divided into backbone, extra, localization and classification. Backbone and extra are connected in series, backbone is used to extract the feature maps in the images and extra is used to perform multi-scale feature extraction on the feature maps. Set 6 groups of output feature maps, there is a convolutional progression relationship between the feature maps, as the feature maps go through more convolutional layers, the smaller the size of the feature maps, the more abstract the features represented by the feature maps. The kth feature map of different sizes satisfies the formula:

$$s_k = s_{min} + \frac{s_{max} - s_{min}}{m-1}(k-1) \tag{1}$$

This step draws on the YOLO idea. Since the number of targets in this research method is small in tracking the bee crawling trajectory, thus the target overlap is not considered.

The model training process goes through positive and negative sample tagging to determine which a priori frame generated from the feature map matches the GT in the training image, and the matching method uses IOU with the threshold set to 0.5, i.e., the intersection ratio between the GT and the a priori frame is greater than 0.5 which is considered as the corresponding match. For the position loss function L_{loc} and classification loss L_{conf} of the SSD algorithm, it is set as the joint loss function:

$$L(x, c, l, g) = \frac{1}{N}(L_{conf}(x, c) + \alpha L_{loc}(x, l, g)) \tag{2}$$

Where N is the number of positive samples of the prior, c is the category confidence prediction, l is the predicted value of the position of the corresponding boundary of the prior frame, g is the position parameter of GT(ground truth), and α is a factor of 1. The position loss function uses smoothL1 loss, and the classification loss function uses cross-entropy loss.

The processing of the predicted negative samples, i.e., the samples with wrong predictions, is hard sample mining. The processing method is to arrange the input predictions in descending order according to the category confidence, take out k negative samples, and then add the k negative samples to the negative samples of the next iteration to train the network. For each prediction frame, the image background is judged according to the predicted category and the category confidence, and the prediction frame indicating the background is filtered. The prediction frames with lower confidence are then filtered based on the confidence threshold. Since the output of the network is the corrected parameters of the prior frame, the prediction frame needs to be decoded to obtain the true parameters of the prior frame. The true prior frames are then sorted in descending order of confidence, and the top k predictor frames are retained. The prediction frames that overlap with each other are filtered out using the extreme value suppression algorithm (NMS) to obtain the final prediction results.

3.3 Object Tracking

SORT algorithm core for Kalman filter and Hungarian matching algorithm, its process Kalman filter prediction, Hungarian algorithm prediction to match the real frame with the tracking frame, Kalman filter update The linear velocity model is introduced in the SORT algorithm with Kalman filter for position prediction, and the Hungarian algorithm uses the IOU distance as a weight, and is not considered the same target when the IOU is less than a threshold (0.3).

The Kalman filter performs prediction and update as a motion model for position prediction. Among them, the principle of Kalman filter is that the filter estimates the state of the current moment based on the state of the previous moment to get the a priori estimate of the state of the current moment, and then uses the true measurement of the current moment and then corrects the a priori estimate to get the estimate of the current moment.

X $(x, y, s, r, x_v, y_v, s_v)$ state descriptive variables, P is the state covariance matrix of the descriptive variables, F the state transfer variance matrix, determined by the motion model Kalman filter, H the observation matrix, andZ(x, y, s, r) the observation variables.

The current state is predicted using the state of the previous moment, applying the state transfer matrix F scope state description vector x to obtain the next moment state variable$X' =$ FX， $P' = FPF^T$, and the current moment's result is updated for the predicted result, satisfying the formula:

$$x = Fx + Bu \quad (3)$$

$$P = \alpha Fx F^T + Q \quad (4)$$

Where u is the accelerated B is the transfer matrix of u. The updating process satisfies the following equations:

$$K = PH^T(HPH^T + R)^{-1} \quad (5)$$

$$x = x + (K(z - Hx)) \quad (6)$$

$$P = (I - KH)P \quad (7)$$

The Hungarian algorithm is a combinatorial optimization algorithm for solving the task assignment problem in polynomial time to associate the data association of multi-target tracking, i.e., to solve the maximum matching problem of the bipartite graph, i.e., to determine whether the tracked target is the same target in the previous stage. In this paper, the Hungarian algorithm is implemented to match the detection frame and the tracking frame, and the whole process is to traverse the detection frame and the target frame to match each other, and if the over-match is successful, it will be retained, and if it is not successful, it will be deleted.

4 Experimental Process and Results

The individual bees were detected by SSD and then tracked using SORT according to the location parameters returned from the detection, and the flow chart of SSD training method is as follows(see Fig.2.):

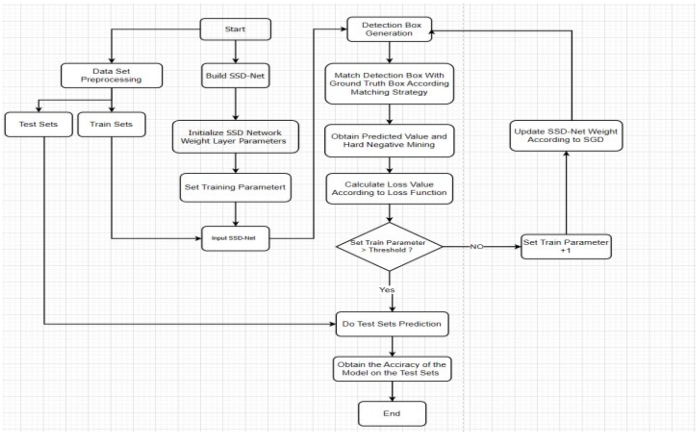

Fig. 2. The flow chart of SSD training method

In the experiment, the dataset containing 3071 images was segmented according to 70% training set and 30% test set, and the average recall during SSD training was 98.3% and the prediction accuracy converged to 88.6%. The accuracy f the training set is shown in the Fig.3.

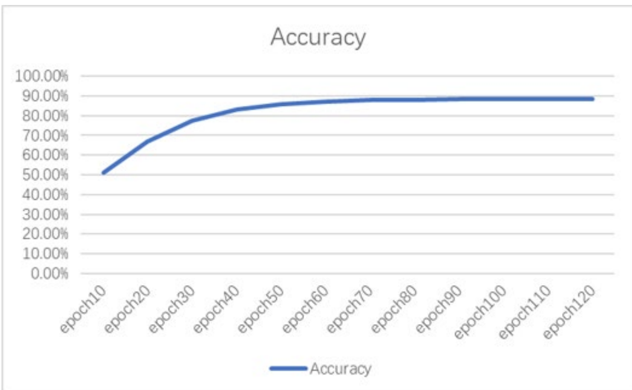

Fig. 3. The accuracy of the training set

SSD prediction of bees in 10 groups of videos was obtained for a total of 20 bees in 10 groups in obtaining a recognition rate of 100% for each bee in 10 seconds.

The overall trajectory tracking flow chart is Fig.4.

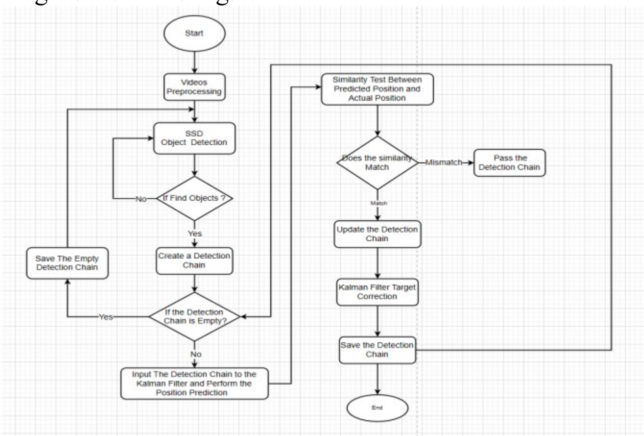

Fig. 4. The whole trajectory tracking flow chart

State transition matrix:

$$F = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \quad (8)$$

Measurement matrix:

$$H = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix} \quad (9)$$

Covariance matrix:

$$P = diag([10,10,10,10,1e4,1e4,1e4]^T) \quad (10)$$

$$Q = diag([1,1,1,1,0.01,0.01,1e-4]^T) \quad (11)$$

$$R = diag([1,1,10,10]^\wedge T) \quad (12)$$

Finally, the tracking results of the target tracking algorithm for 10 sets of videos are listed in the table 1.

Table 1. The tracking result

No.	Detection tracking rate /%	Undetected rate /%	False alarm rate /%
1	88.1	11.9	3
2	81.7	18.3	5
3	84.6	1.4	2
4	93.2	6.8	7
5	84.3	15.7	5
6	86.6	13.4	2
7	85.3	14.7	3
8	84.1	15.9	4
9	85.2	14.8	2
10	85.6	14.4	3

According to the above experimental results, it is found that even though the detection accuracy of target detection using SSD target detection model with a prediction accuracy of about 88.6% for frame-by-frame in 10 groups of videos reaches 100%, the detection rate in trajectory tracking using SORT is generally not higher than 90%.

5 Conclusion

In this paper, we use SSD target detection with the SORT target tracking method to transform the observation of bee crawling behavior description only manually to the automatic acquisition of bee crawling behavior description using a computer. By quantifying the bee crawling behavior by representing the three variables of bee position, speed, and direction within a time series, the influence of artificial subjective judgment factors on the results of traditional research through observation and physiological detection methods is greatly reduced, and data support is provided for further research on the midway characteristics of honey bees by analyzing their health behavior and honey bee poisoning behavior.

Acknowledgements:

The funding for this project was provided by supplementary funds from the Major Science and Technology Project of Hainan Province, under grant number ZDKJ202017-2-2.

References

[1] K. Bozek, L. Hebert, A. S. Mikheyev and G. J. Stephens, "Towards Dense Object Tracking in a 2D Honeybee Hive," 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 4185-4193(2016)

[2] GUO Xiuming, ZHU Yeping, LI Shijuan, ZHANG Jie, LYU Chunyang, LIU Shengping. Scale Adaptive Small Objects Detection Method in Complex Agricultural Environment: Taking Bees as Research Object. Smart Agriculture, 2022, 4(1): 140-149.

[3] YANG Qingyu, WANG Yongrang, LI Hao, TANG Shanjun, LI Fang, ZHANG Guohua. Adaptive Detection and Tracking Algorithm for Infrared Target Size Variation. Infrared Technology , 2022, 44(11): 1176-1185.

[4] Quigley TP, Amdam GV, Harwood GH. Honey bees as bioindicators of changing global agricultural landscapes. Curr Opin Insect Sci. 2019;35:132-137..

[5] Yuan Hong, Huang Jun, Wang Ruigang & Wang Shuangli. (2022). Analysis of bee flow automatic monitoring technology. Bee Journal (08), 5-11

[6] R. Girshick, "Fast R-CNN," 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 2015, pp. 1440-1448, doi: 10.1109/ICCV.2015.169.

[7] Zeng, X., Ouyang, W., Yang, B., Yan, J., Wang, X. (2016). Gated Bi-directional CNN for Object Detection. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds) Computer Vision – ECCV 2016. ECCV 2016. Lecture Notes in Computer Science(), vol 9911. Springer, Cham.

[8] L. Jiao et al., "A Survey of Deep Learning-Based Object Detection," in IEEE Access, vol. 7, pp. 128837-128868, 2019,

[9] He, K., Zhang, X., Ren, S., & Sun, J. (2015). Spatial pyramid pooling in deep convolutional networks for visual recognition. IEEE transactions on pattern analysis and machine intelligence, 37(9), 1904-1916.

[10] H. Rezatofighi, N. Tsoi, J. Gwak, A. Sadeghian, I. Reid and S. Savarese, "Generalized Intersection Over Union: A Metric and a Loss for Bounding Box Regression," 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 2019, 658-666,

[11] E. Neuburger and V. Krebs, "Einführung in Die Theorie des Linearen Optimalfilters (Kalman-Filter) (Introduction to Linear Optimal Filtering Theory (Kalman Filter))," in IEEE Transactions on Systems, Man, and Cybernetics, vol. SMC-6, no. 11, pp. 796-796,

[12] V. -T. Luu, V. -H. Tran, E. Poliakov and C. -C. Huang, "D2D: Divide to Detect, A Scale-Aware Framework for On-Road Object Detection Using IR Camera," 2023 IEEE International Conference on Consumer Electronics (ICCE), Las Vegas, NV, USA, 2023, pp. 1-5

[13] M. Huang, B. Wang, J. Wan and C. Zhou, "Improved blood cell detection method based on YOLOv5 algorithm," 2023 IEEE 6th Information Technology,Networking,Electronic and Automation Control Conference (ITNEC), Chongqing, China, 2023, pp. 992-996

[14] T. Hu and J. Zhang, "Highly overlapping strawberry leaf detection based on YOLOv5s-MBLS deep learning method," 2023 3rd International Conference on Neural Networks, Information and Communication Engineering (NNICE), Guangzhou, China, 2023, pp. 145-148

[15] Law, H., Deng, J. (2018). CornerNet: Detecting Objects as Paired Keypoints. In: Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y. (eds) Computer Vision – ECCV 2018. ECCV 2018. Lecture Notes in Computer Science(), vol 11218. Springer, Cham