IEEE Sensors Letters VOL. 7, NO. 2, FEBRUARY 2023

IEEE Sensors Council

Sensor applications

A Novel Pet Trajectory Prediction Method for Intelligent Plant Cultivation Robot

Zihan Jiang[1*], Bo Jin[2*], and Yichen Song[3]

[1] School of Advanced Technology, Xi'an Jiaotong-Liverpool University, Suzhou 215000, China
[2] Institute of Systems and Robotics, Department of Electrical and Computer Engineering, University of Coimbra, 3030290 Coimbra, Portugal
[3] College of Control Science and Engineering, Zhejiang University, Hangzhou 310027, China
[*] Member, IEEE

Manuscript received 20 November 2022; revised 1 December 2022; accepted 16 January 2023. Date of publication 20 January 2023; date of current version 3 February 2023.

Abstract—Intelligent plant cultivation robots will have certain working obstacles due to various moving objects, such as pets. This letter studies how to predict the trajectory of pets to avoid collisions. In previous intelligent plant cultivation robot systems, target-tracking technology was used to track pets and avoid collisions. However, collisions often occur even after tracking the trajectory of the pet because the trajectory of the pet's movement is difficult to determine. Therefore, the Informer algorithm is used to predict the trajectory of pets in this letter. It is presented that collisions can be avoided based on the predicted trajectory. Experimental results on the data set show that this algorithm has better accuracy than most common algorithms, so trajectory prediction based on this algorithm can make the robot avoid obstacles better. Finally, by verifying two data sets, the public data set and the self-made data set, it is determined that the performance of the proposed method is optimized by 3%–50% compared with other methods.

Index Terms—Sensor applications, deep learning, distance estimation, Informer, robot, trajectory prediction.

I. INTRODUCTION

In this period of rapid development of robots, there are different kinds of robots in all kinds of industries, such as medical robots [1], robotic arms [2], underwater robots [3], and robots in other industries. There is no doubt that robots are a scorching topic nowadays. Likewise, intelligent plant cultivation robot system [4], [5], which performs better than the existing plant cultivation systems as it costs less, saves more time, and achieves better quality, also emerged in the field of plant cultivation. The system performs particularly prominent when it is used indoors.

Intelligent plant cultivation robots can complete the work of a static environment, but the problem of dealing with pet distractions remains to be solved. Trajectory prediction has been widely applied in various fields at present, including ship trajectory [6], typhoon trajectory [7], and vehicle trajectory [8]. From the perspective of trajectory prediction, this letter predicts the route of pets in advance so that the intelligent plant robot can avoid the pets.

In the current field of trajectory prediction research, the research on trajectory prediction can be divided into three areas: 1) physics-based approaches; 2) classical machine learning approaches; 3) deep learning approaches. In the physics approach, trajectory prediction is performed by modeling some of the intuitive features of a pedestrian and then generalizing the past state to the future based on different physical models [9], [10], [11]; in the classical machine learning approach, the trajectory prediction problem is transformed from model-driven to data-driven, and various classical machine learning models are heavily used in this approach [12], [13], [14]; In deep learning methods, trajectory prediction is viewed as a sequence generation problem, where deep learning models learn valuable information from pedestrian movement trajectories to perform pedestrian trajectory prediction [15], [16], [17].

The method adopted in this letter is supervised learning, which is a deep learning method, and it requires a lot of historical data. In order to reduce the workload of the manual label, the Yolov5+Deep Sort is adopted for automatic labeling. Through target recognition and track following, the track data of pet moving is automatically recognized, and the purpose of reducing manual labels can be finally achieved.

Informer [18] is applied in the field of trajectory prediction for the first time in this letter, and it is compared with other trajectory prediction algorithms. Experiments show that Informer has a better performance than other algorithms in the field of trajectory prediction.

Furthermore, 3-D pet trajectory prediction based on a two-layer Informer is carried out in this letter to realize the full function of collision avoidance.

In general, the main contributions of this letter are as follows.

1) Applying Informer in track prediction for the first time.
2) Using a two-layer Informer for 3-D trajectory prediction.
3) Employing Yolov5+Deep Sort for automatic labeling reduces 95% of manual labeling.

II. INTELLIGENT PLANT CULTIVATION ROBOT SYSTEM

As shown in Fig. 1, the experimental object used in this letter is an intelligent plant cultivation robot composed of robot hardware facilities and a robot development platform.

The robot mainly consists of five modules: 1) the plant cultivation module; 2) the optical module; 3) the mapping module; 4) the path planning module; and 5) the speech processing module.

Corresponding author: Zihan Jiang (e-mail: 1095773538@qq.com).
Zihan Jiang and Bo Jin contributed equally to this work.
Associate Editor: L. Azpilicueta.
Digital Object Identifier 10.1109/LSENS.2023.3238468

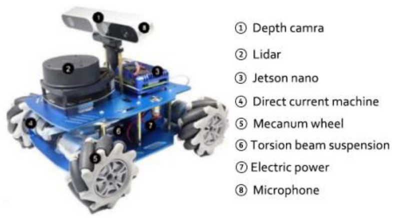

Fig. 1. Intelligent plant cultivation robot system.

TABLE 1. Improvement of Informer

Disadvantages of Transformer	Improvements of Informer
Complicated point multiplication of Self-Attention	Using ProbSparse Self-Attention
Multi-layer network overlay, large memory footprint	Distillation to reduce dimensions and network parameters
Slow step-by-step decoding prediction generation speed	Getting all results in one step by Generative Style Decoder

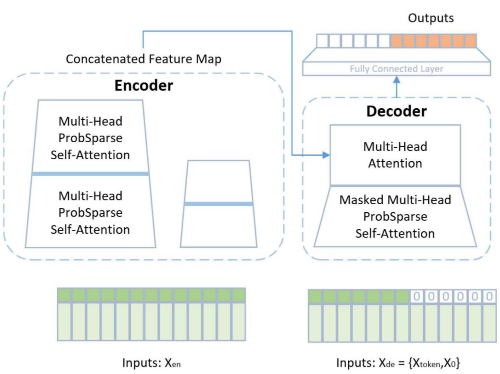

Fig. 2. Informer network structure.

III. TRAJECTORY PREDICTION

The trajectory prediction algorithm used in this letter is the Informer algorithm improved by Transformer [19]. The principle of Informer will be briefly introduced as follows.

Informer is proposed to overcome the shortcomings of the Transformer and consider the characteristics of long-term time series prediction, as shown in Table 1.

The overall Informer model architecture, similar to the Transformer, is divided into Encoder and Decoder, as shown in Fig. 2.

IV. DATA SETS

Data is essential in the field of deep learning. The self-made pet-tracking data set in this letter includes 400 000 marked targets, and the total time of the data set reaches 30 min, with 30 frames per second. There is no doubt that the labeling task of this dataset is challenging because there are too many needed tags. Consequently, manual and automatic labeling are combined to complete our experiment, which significantly reduces the workload required for labeling.

A. Manual Labeling of Object Images

In this letter, the target tracking method of YOLOv5X+Deep Sort is used to reduce the workload required in manual marking. First,

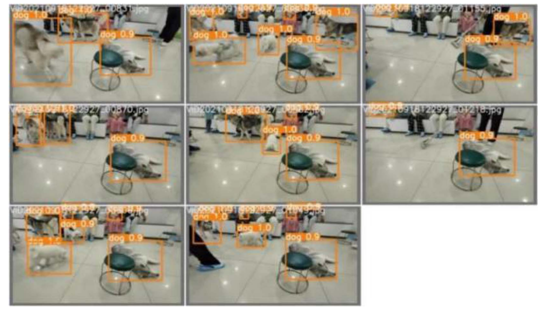

Fig. 3. Actual testing results of YOLOv5x.

Fig. 4. Deep sort actual testing results.

YOLOv5X needs to be trained for pet target recognition. As the object recognition method in this letter belongs to supervised learning, object category information and object border position information need to be obtained. Then, Labeling is used to label the collected training samples in the experimental part. Altogether, 1000 images are manually selected from the self-made data set for annotation.

After the image annotation, the data set is divided into the training set, verification set, and test set, with the training set accounting for 70%, verification set 20%, and test set for 10%, respectively.

B. Automatic Labeling of Object Images

After the training of the YOLOv5X model is completed, the user object information can be gathered by automatic image marking. It is roughly calculated that automatic labeling reduces more than 95% of the workload compared with manual labeling. Next, the two algorithms of automatic marking are briefly introduced in the following section.

1) YOLOv5x: YOLOv5 is a single-stage target detection algorithm. Compared with YOLOv4 [20], YOLOv5 has smaller mean weight files [21], [22], shorter training time, and faster inference speed based on less reduction in average detection accuracy. YOLOv5 network structure is divided into four parts: 1) Input, 2) Backbone, 3) Neck, and 4) Head.

In this letter, the YOLOv5X model is finally decided as the object detection model in this letter by considering both accuracy and speed. The final experimental results are shown in Fig. 3.

It can be seen from Fig. 3 that YOLOv5x can better detect the position of pets in the image and replace manual marking to some extent.

2) Deep Sort [23]: Deep Sort adopts recursive Kalman filtering and hypothesis tracking methods of frame-by-frame data association. The final effect diagram of the Deep Sort implementation is shown in Fig. 4.

As seen from Fig. 4, Deep Sort, having a good effect of tracking objects, can better mark the moving trajectory data of pets.

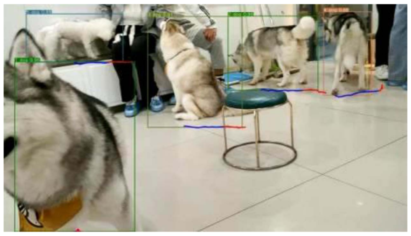

Fig. 5. Schematic diagram of pet trajectory prediction.

TABLE 2. Experimental Results of Pet Data Set

	ADE	FDE
Vanilla LSTM [26]	7.8	15.4
Social LSTM [26]	7.9	15.3
SGAN [27]	7.1	12.1
Sophie [28]	6.5	11.8
PTF [29]	6.6	12.4
GAT [30]	6.3	10.7
Social GAT [30]	6.3	10.6
Social STGCNN [31]	6.1	10.2
RSGB [32]	6.7	11.3
Ours	5.9	10.4

V. EXPERIMENTAL RESULTS OF TRAJECTORY PREDICTION

Two data sets are mainly adopted in the experiment of this letter. The first is the open ETH data set, which verifies the extensive application of the trajectory prediction method. The second is a self-made pet trajectory movement data set, through which the trajectory prediction algorithm is applied to practical applications.

A. Evaluation Principle

This letter uses two indicators to evaluate the model's performance: ADE [24] and FDE [25]. ADE represents the mean square error of all predicted positions and actual positions during the prediction period. FDE demonstrates the distance between the final predicted position and the corresponding absolute position at the end of the predicted trajectory. The calculation method is shown in

$$\text{ADE} = \frac{\sum_{n\in N}\sum_{t\in T_r}||\hat{p}_t^n - p_t^n||_2}{N * T_p} \tag{1}$$

$$\text{FDE} = \frac{\sum_{n\in N}||\hat{p}_t^n - p_t^n||_2}{N}, t = T_{\text{final}}. \tag{2}$$

B. Self-Made Data Set (Pet Data Set)

The pet information, which forms the self-made pet data set in this letter, was gathered and labeled from a long-time shooting of a constantly changing scene using the perspective of a stationary intelligent plant cultivation robot. The schematic diagram of the prediction is shown in Fig. 5.

As seen from Table 2, the algorithm proposed in this letter performs better than other typical trajectory prediction algorithms in the pet data set. It can forecast the trajectory of pets more accurately.

TABLE 3. Experimental Results of ETH Data Set

	ADE	FDE
Vanilla LSTM [26]	1.09	2.41
Social LSTM [26]	1.09	2.35
SGAN [27]	0.87	1.62
Sophie [28]	0.7	1.43
PTF [29]	0.73	1.65
GAT [30]	0.68	1.29
Social GAT [30]	0.69	1.29
Social STGCNN [31]	0.64	1.11
RSGB [32]	0.8	1.53
Ours	0.61	1.21

C. Open Data Set (ETH Data Set)

ETH is a binocular vision-based pedestrian database that can be used for a variety of studies on pedestrian movement.

As can be seen from Table 3, the algorithm proposed in this letter performs better than other typical trajectory prediction algorithms in ETH data sets. It can accurately predict the trajectory of pedestrians, which has a specific promotion value.

VI. EXTENDED APPLICATION (DOUBLE LAYER INFORMER FOR OBSTACLE AVOIDANCE)

In this section, two-layer Informer are used to predict the 3-D trajectory of pet movement. The function of each layer is, to: 1. Predict 2-D trajectory of pet movement; 2. Predict the distance between the pet and the robot. The first layer of Informer has been fully described in the previous parts, so this section puts emphasis on the second layer.

The vision module of the intelligent plant cultivation robot system proposed in this letter is a binocular camera based on the depth camera, so simple triangulation can be used to learn the distance between the pet and the robot visually. When the pet appears in the visible range of the camera for the first time, i.e., the distance between the pet and the robot is measured, and the distance is denoted as s_1.

The information we can get from the target detection frame includes (x, y, w, h) and the target frame can be drawn by calculating the four points (x, y), $(x + w, y)$, $(x, y + h)$, and $(x + w, y + h)$.

On the premise that s_1 is known, this letter deduces s_n at each time point by calculating the area ratio of target detection frames between each time point. The specific calculation formula is shown in

$$s_n = s_1 \frac{w_n h_n}{w_1 h_1}. \tag{3}$$

Therefore, this letter predicts $s$at different time points by predicting w, h. The evaluation indexes of the prediction results are still ADE and FDE, with the value of ADE and FDE being 5.641 and 9.551, respectively. This shows that the algorithm proposed in this letter has good prediction results. Combined with the pet movement trajectory predicted by the first-layer Informer, the schematic diagram of the final realization is shown in Fig. 6.

In Fig. 6, the blue ball represents the robot, the green ball the pet, s_n the predicted distance between the pet and the robot, the dotted line is the predicted distance, and the solid line is the predicted direction of movement.

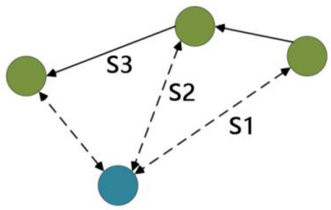

Fig. 6. Pet track prediction scheme.

VII. CONCLUSION

This letter effectively solves the problem of too many collisions between intelligent plant cultivation robots and pets. First, the original trajectory-tracking method is improved to the trajectory prediction method. Second, the Informer is used in the field of track prediction for the first time and has obtained excellent results. At the same time, two data sets have been verified in this letter, and the experiment shows that this letter has a 3%–50% improvement compared with other methods in the comparison method. Finally, the combination of Informer employed in pet track prediction on images and Informer based on distance prediction realizes the track prediction on the 3-D level, achieving a rather remarkable effect.

REFERENCES

[1] T. A. Lawrie et al., "Robot-assisted surgery in gynaecology," *Cochrane Database Systematic Rev.*, no. 4, 2019, Art. no. CD011422, doi: 10.1002/14651858.CD011422.pub2.
[2] J. Shi, J. Xu, J. Sun, and Y. Yang, "Iterative learning control for time-varying systems subject to variable pass lengths: Application to robot manipulators," *IEEE Trans. Ind. Electron.*, vol. 67, no. 10, pp. 8629–8637, Oct. 2020.
[3] X. Zheng, W. Wang, M. Xiong, and G. Xie, "Online state estimation of a fin-actuated underwater robot using artificial lateral line system," *IEEE Trans. Robot.*, vol. 36, no. 2, pp. 472–487, Apr. 2020.
[4] J. Zihan, L. Lizhuang, J. Kaiyang, S. Bingyu, and L. Mingyu, "Preliminary design of a new intelligent plant cultivation robot," in *Proc. IEEE Asia-Pacific Conf. Image Process., Electron. Comput.*, 2021, pp. 810–814, doi: 10.1109/IPEC51340.2021.9421192.
[5] Z. Jiang, Y. Guo, K. Jiang, M. Hu, and Z. Zhu, "Optimization of intelligent plant cultivation robot system in object detection," *IEEE Sensors J.*, vol. 21, no. 17, pp. 19279–19288, Sep. 2021, doi: 10.1109/JSEN.2021.3077272.
[6] H. Xue and T. Chai, "Vessel track prediction based on fractional gradient recurrent neural network with maneuvering behavior identification," *Sci. Program.*, vol. 2021, 2021, pp. 1–11.
[7] D. Xu, F. Shen, J. Min, and A. Shu, "Assimilation of GPM microwave imager radiance for track prediction of typhoon cases with the WRF hybrid En3DVAR system," *Adv. Atmospheric Sci.*, vol. 38, no. 6, pp. 983–993, 2021.
[8] Y. Xing, C. Lv, and D. Cao, "Personalized vehicle trajectory prediction based on joint time series modeling for connected vehicles," *IEEE Trans. Veh. Technol.*, vol. 69, no. 2, pp. 1341–1352, Feb. 2020.
[9] F. Nashashibi and S. Ammoun, "Real time trajectory prediction for collision risk estimation between vehicles," in *Proc. IEEE 5th Int. Conf. Intell. Comput. Commun. Process.*, 2009, pp. 417–422, doi: 10.1109/ICCP.2009.5284727.
[10] C. G. Keller and D. M. Gavrila, "Will the pedestrian cross? A study on pedestrian path prediction," *IEEE Trans. Intell. Transp. Syst.*, vol. 15, no. 2, pp. 494–506, Apr. 2014, doi: 10.1109/TITS.2013.2280766.
[11] A. S. Willsky, E. B. Sudderth, E. Fox, and M. I. Jordan, "Bayesian nonparametric inference of switching dynamic linear models," *IEEE Trans. Signal Process.*, vol. 59, no. 4, pp. 1569–1585, Apr. 2011, doi: 10.1109/TSP.2010.2102756.
[12] F. Solera, R. Cucchiara, and S. Calderara, "Socially constrained structural learning for groups detection in crowd," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 5, pp. 995–1008, May 2016, doi: 10.1109/TPAMI.2015.2470658.
[13] A. Rangesh, M. M. Trivedi, and N. Deo, "How would surround vehicles move? A unified framework for maneuver classification and motion prediction," *IEEE Trans. Intell. Veh.*, vol. 3, no. 2, pp. 129–140, Jun. 2018, doi: 10.1109/TIV.2018.2804159.
[14] J. Wang, K. Li, Y. Li, and X.-Y. Lu, "Pedestrian trajectory prediction combining probabilistic reasoning and sequence learning," *IEEE Trans. Intell. Veh.*, vol. 5, no. 3, pp. 461–474, Sep. 2020, doi: 10.1109/TIV.2020.2966117.
[15] D. Q. Huynh, H. Xue, and M. Reynolds, "PoPPL: Pedestrian trajectory prediction by LSTM with automatic route class clustering," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 1, pp. 77–90, Jan. 2021, doi: 10.1109/TNNLS.2020.2975837.
[16] L. Zhu, R. Quan, Y. Wu, and Y. Yang, "Holistic LSTM for pedestrian trajectory prediction," *IEEE Trans. Image Process.*, vol. 30, pp. 3229–3239, 2021, doi: 10.1109/TIP.2021.3058599.
[17] A. Hasan, K. Driggs-Campbell, K. Shin, R. Li, and Z. Huang, "Long-term pedestrian trajectory prediction using mutable intention filter and warp LSTM," *IEEE Robot. Automat. Lett.*, vol. 6, no. 2, pp. 542–549, Apr. 2021, doi: 10.1109/LRA.2020.3047731.
[18] H. Zhou et al., "Informer: Beyond efficient transformer for long sequence time-series forecasting," in *Proc. AAAI Conf. Artif. Intell.*, 2021, pp. 11106–11115.
[19] A. Vaswani et al., "Attention is all you need," *Adv. Neural Inf. Process. Syst.*, vol. 30, pp. 5998–6008, 2017.
[20] C.-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao, "Scaled-YOLOv4: Scaling cross stage partial network," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Nashville, TN, USA, 2021, pp. 13024–13033, doi: 10.1109/CVPR46437.2021.01283.
[21] C. Y. Wang, "CSPNet: A new backbone that can enhance learning capability of CNN," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops*, 2020, pp. 390–391.
[22] T. Y. Lin et al., "Feature pyramid networks for object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 2117–2125.
[23] N. Wojke, A. Bewley, and D. Paulus, "Simple online and realtime tracking with a deep association metric," in *Proc. IEEE Int. Conf. Image Process.*, 2017, pp. 3645–3649.
[24] H. Bai, S. Cai, N. Ye, D. Hsu, and W. S. Lee, "Intention-aware online POMDP planning for autonomous driving in a crowd," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2015, pp. 454–460.
[25] P. Raksincharoensak, T. Hasegawa, and M. Nagai, "Motion planning and control of autonomous driving intelligence system based on risk potential optimization framework," *Int. J. Automot. Eng.*, vol. 7, no. 1, pp. 53–60, 2016.
[26] A. Alahi et al., "Social LSTM: Human trajectory prediction in crowded spaces," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 961–971.
[27] A. Gupta et al., "Social GAN: Socially acceptable trajectories with generative adversarial networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 2255–2264.
[28] A. Sadeghian et al., "SoPhie: An attentive GAN for predicting paths compliant to social and physical constraints," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 1349–1358.
[29] J. Liang et al., "Peeking into the future: Predicting future person activities and locations in videos," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 5725–5734.
[30] V. Kosaraju et al., "Social-BiGAT: Multimodal trajectory forecasting using bicycle-GAN and graph attention networks," in *Proc. 33rd Int. Conf. Neural Inf. Process. Syst.*, Red Hook, NY, USA: Curran Associates Inc., 2019, p. 1.
[31] A. Mohamed et al., "Social-STGCNN: A social spatio-temporal graph convolutional neural network for human trajectory prediction," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 14424–14432.
[32] J. Sun, Q. Jiang, and C. Lu, "Recursive social behavior graph for trajectory prediction," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 660–669.