The 11th Asian-Australasian Conference on Precision Agriculture (ACPA 11)
October 14-16, 2025, Chiayi, Taiwan

DEEP REINFORCEMENT LEARNING BASED ROBOTIC ARM CONTROL FOR AUTONOMOUS HARVESTING

Shao-Yang Chang[1], Al Bashir[2], Azlan Zahid[2], Chen-Kang Huang[1, 3*]

[1] Department of Biomechatronics Engineering, National Taiwan University, Taiwan.
[2] Department of Biological and Agricultural Engineering, Texas A&M University, USA.
[3] Center for Intelligent Agriculture Education and Research, National Taiwan University, Taiwan.

*Corresponding Author: ckhuang94530@ntu.edu.tw

ABSTRACT

Inverse Kinematics (IK) is a traditional method used for robotic arm manipulation, relying heavily on precise calibration and huge computational demands for arms with higher Degrees of Freedom (DoF). In contrast, Deep Reinforcement Learning (DRL) is an innovative approach to manipulation that exhibits greater tolerance for calibration inaccuracies. It trains using noise added to joint angles, allowing it to learn how to compute accurate trajectories even with inaccuracies in the joint angles. In this study, the Proximal Policy Optimization (PPO) algorithm was employed as DRL to train the UR10e robotic arm to guide its end-effector to reach the 3D coordinates with proper orientations of the target object. To obtain the coordinates of the on-plant strawberry fruits as the target object, the YOLOv11 algorithm was trained as the vision model, where the dataset was collected from greenhouse-grown strawberries using an Intel RealSense Depth D435i camera. NVIDIA IsaacSim was used as the simulated environment to train the PPO algorithm. A 3D model of the robotic arm was imported to the environment, and its physical properties, such as rigidity, damping, torque limits, and velocity limits, were defined. This setup enables the robotic arm to learn optimal policies through a trial-and-error approach for reaching its target strawberry for harvesting operation, where the real-time 3D coordinates were sent from the vision module. For the validation experiment on strawberry harvesting in an actual greenhouse, a ROS2 environment was set up to enable communication between system nodes. Experimental results demonstrated that the proposed DRL framework and YOLOv11 vision model successfully achieved precise robotic arm control and adaptability in complex agricultural harvesting tasks.

Keywords: Deep Reinforcement Learning, Robot arm control, Strawberry segmentation, 3D coordinates, IsaacSim Simulation.

INTRODUCTION

Robotic arms are increasingly applied in modern agriculture for precision harvesting and related tasks (Williams and Polydoros, 2025). Traditional control methods, such as inverse kinematics (IK), require precise calibration and involve heavy computation for high degrees of freedom, making them less suitable for dynamic agricultural environments where uncertainties are common. In contrast, deep reinforcement learning (DRL) provides greater robustness by training policies under noisy conditions, enabling the robotic arm to adapt to calibration errors and operate with lower runtime computational demands.

This study applies the proximal policy optimization (PPO) algorithm developed by Schulman *et al.* (2017) within IsaacSim to train a robotic arm for agricultural harvesting. A module based on vision is integrated to detect the cutting point of crops, which are then used as targets for the trained policy to guide the robotic arm's end-effector. The objective is to develop and evaluate a DRL framework that achieves accurate, adaptable, and efficient robotic arm control in agricultural environments.

MATERIAL AND METHODS

The training and evaluation of the robotic arm was conducted in IsaacSim, where a 3D model of UR10e was imported, and its physical properties, like joint rigidity, damping, torque limits, and velocity limits, were specified to replicate realistic motion dynamics. A greenhouse environment was created with varying target positions and orientations to simulate agricultural harvesting conditions. The control framework was based on the PPO algorithm, with the state space defined by joint angles and target coordinates, while the action space consisted of continuous joint angle increments mapped to trajectory commands. The reward function combined distance and orientation accuracy with penalties for excessive velocity. Noise was introduced during training to improve robustness against calibration errors.

A vision module was integrated to detect the cutting point of crops using artificial intelligence (AI) image recognition techniques, and the identified 3D coordinates were transmitted to the PPO policy as target inputs. To align the camera coordinate system with the robot base, a hand-eye calibration procedure was performed, ensuring that detected points could be transformed into reachable poses for the manipulator. The trained policy was interfaced with the robotic arm through ROS2, where action messages were published to the trajectory controller via the UR driver, enabling communication with the physical robot. This setup provides a pipeline from visual perception to motion execution, allowing the system to operate reliably in harvesting tasks.

The DRL policy was compared against the cuRobo proposed by Sundaralingam *et al.* (2023), a planner that computes trajectories through IK and trajectory optimizations. Both methods were evaluated in the same IsaacSim environment with an identical UR10e robotic arm model. A set of five hundred distinct target poses was generated across the workspace, and both DRL and cuRobo were tested on the same targets set.

RESULTS & DISCUSSION

cuRobo computes a complete trajectory in advance through IK and optimization, whereas DRL outputs an action at every step during execution. cuRobo required a mean planning time of 56.81 milliseconds per target, while DRL needed 0.14 milliseconds to complete a step on average, resulting in 4.20 milliseconds of compute per second at 30 steps per second. Despite the different definitions, these results show that DRL achieves lower overall computational demand while maintaining real-time control.

The DRL policy generated joint trajectories with a mean length of 4.38 rad, whereas cuRobo produced longer trajectories with a mean length of 5.81 rad, representing a reduction of 25%. The shorter path of DRL is attributed to its incremental computation, which promotes direct movements, while the single computation optimization of cuRobo may result in detours or branch switching when facing orientation constraints, which leads to longer trajectories. The mean acceleration under DRL reached 8.42 rad/s^2, which is substantially higher than the value of 0.97 rad/s^2 obtained by cuRobo. This outcome reflects the nature of DRL's incremental control strategy, where frequent micro adjustments accumulate into larger accelerations. DRL achieves faster and shorter paths, but at the cost of higher accelerations that may increase mechanical stress.

Table 1 Performance comparison between DRL policy and cuRobo planner.

Planner	Joint trajectories length (rad)	Joint acceleration (rad/s^2)
DRL	4.38	8.42
cuRobo	5.81	0.97

CONCLUSIONS

This study demonstrates the potential of DRL for robotic arm control in agricultural harvesting. Compared to traditional IK planners such as cuRobo, the DRL framework delivers faster computation, reduced trajectory length, and improved adaptability under calibration uncertainties. The integration of a vision module enables from perception to action execution, ensuring accurate targeting of crop cutting points. But the DRL approach results in higher joint accelerations that may increase mechanical stress. Future work will focus on enhancing the smoothness of the DRL policy to reduce abrupt accelerations and further improve motion quality in harvesting tasks.

ACKNOWLEDGEMENTS

This research was supported by the “Global Agricultural Innovation and Sustainability Alliance: Integrating Plant Science, Environmental Solutions, Veterinary Practices, and Smart Technologies” under the University Academic Alliance in Taiwan (UAAT), funded by the Ministry of Education (MoE), Taiwan. I also gratefully acknowledge the Texas A&M AgriLife Research and Extension Center at Dallas for providing the greenhouse facilities and experimental equipment, and valuable support in the development of computer vision models, as well as for hosting an exchange student during this study.

REFERENCES

Schulman, J., F. Wolski, P. Dhariwal, A. Radford, and O. Klimov. 2017. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*.

Sundaralingam, B., S. K. S. Hari, A. Fishman, C. Garrett, K. Van Wyk, V. Blukis, A. Millane, H. Oleynikova, A. Handa, and F. Ramos. 2023. Curobo: Parallelized collision-free robot motion generation. *2023 IEEE International Conference on Robotics and Automation (ICRA)*.

Williams, E., and A. Polydoros. 2025. Zero-Shot Sim-to-Real Reinforcement Learning for Fruit Harvesting. *arXiv preprint arXiv:2505.08458*.