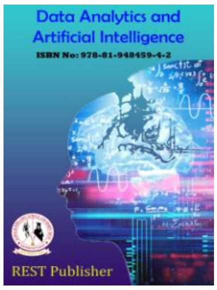
Data Analytics and Artificial Intelligence

Vol: 3(2), 2023

REST Publisher; ISBN: 978-81-948459-4-2

Website: http://restpublisher.com/book-series/daai/

DOI: https://doi.org/10.46632/daai/3/2/20



Mobile Robotic Arm for Opening Doors Using Proximal Policy Optimization

***M. Kokila, G. Amalredge**

St. joseph's college of arts and science for women, Hosur, Tamil Nādu, India

*Corresponding Author Email: kokila2259@gmail.com

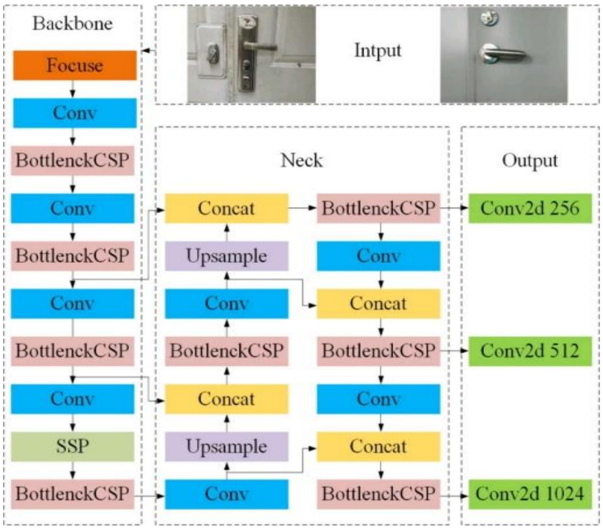
Abstract: *The traditional robotic arm control method has strong dependence on the application scenario. To improve the reliability of the mobile robotic arm control when the scene is disturbed, this paper proposes a control method based on an improved proximal policy optimization algorithm. This study researches mobile robotic arms for opening doors. At first, the door handle position is obtained through an image-recognition method based on YOLOv5. Second, the simulation platform CoppeliaSim is used to realize the interaction between the robotic arm and the environment. Third, a control strategy based on a reward function is designed to train the robotic arm and applied to the opening-door task in the real environment. In this paper PPO algorithm is used to solve the result. The experimental results show that the proposed method can accelerate the convergence of the training process. Besides, our method can effectively reduce the jitter of the robotic arm and improve the stability of control.*

keywords: *yolov5; CoppeliaSim; mobile robotic arm; control strategy; PPO algorithm; deep reinforcement learning; proximal strategy optimization.*

1. INTRODUCTION

In the field of modern industrial technology, robotic arms are used to perform dangerous and repetitive tasks. To expand the working range of the robotic arms, the researchers combined the robotic arm with the mobile platform to form a mobile robotic arm. The mobile robotic arm has the moving function of the mobile platform and the flexible operation function of the robotic arm. At present, the mobile robotic arm is widely used in the fields of life services, industrial manufacturing, and space exploration, and its research has important theoretical value and broad application prospects [1]. The motion planning of mobile robotic arms is usually achieved by demonstration in traditional industrial applications. For application scenarios with changing working environments, traditional control methods are becoming increasingly challenging to meet the needs of users to control robotic arms quickly. In this paper, the recognition method of YOLOv5 [19] is used for visual guidance to quickly and accurately identify the door handle target. For the application of the deep reinforcement learning method in the control of the mobile robotic arm, the training convergence speed is slow and the stability is poor because of the sparse reward value of the algorithm. To reduce training time and help the algorithm converge, this paper proposes an improved PPO algorithm. It can learn complex behaviors, handle continuous action spaces, and find good policies in environments with high-dimensional state spaces.
The contributions of this work are as follows:

- ✓ The method of YOLOV5 is applied to door-handle recognition, and the relationship between distance and pixel ratio is experimentally fitted for door-handle localization.
- ✓ The state space, action space and reward function are designed to realize the interaction between the robotic arm and the environment.
- ✓ An improved PPO algorithm is presented by expanding the experience pool to reduce the exploration of the useless environment.
- ✓ The comparison experiments show that the improved PPO algorithm is faster and more stable than the TRPO and PPO algorithms.



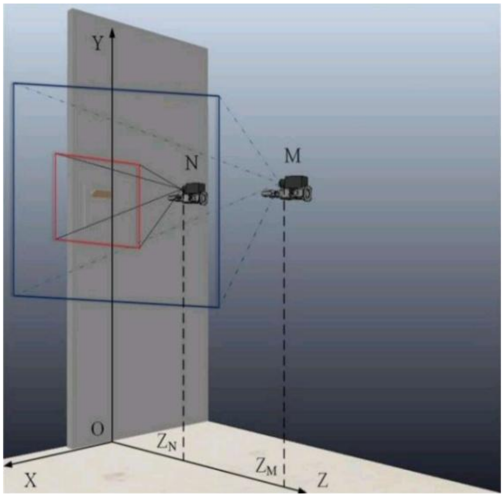
2. HANDLE RECOGNITION AND POSITIONING METHOD DESCRIPTION:

In this section, the proposed identification and positioning methods are described. With the increasing demand for target detection accuracy and speed in practical applications, the current target detection algorithms mainly use detection methods based on deep learning. Target detection algorithms based on deep learning include Faster R- CNN [20], SSD [21], YOLOv3 [22], YOLOv5, etc. Compared with traditional detection methods, YOLOv5 has the characteristics of good robustness, strong generalization, and high precision.

Yolov5 Network Model: The YOLOv5 network structure consists of four parts: Input, Backbone, Neck, and Output layer.

The network structure: The input terminal Input adopts a data-enhancement method to perform an adaptive scaling operation on the input image. The focus structure at the front end of Backbone effectively improves the quality of image extraction through convolution operations. The center and scale prediction (CSP) structure enhances the network's ability to fuse features. The spatial pyramid pooling (SPP) module first uses kernels of varied sizes to perform max-pooling operations. Neck aggregates the feature information with the output features of the CSP module and fully integrates the image features of different layers. The output layer Output adopts the generalized intersection over union (GIOU) loss method to increase the measure of the intersection scale, which solves the situation that the two boxes do not intersect.

Handle Positioning: Move the robotic arm in front of the door and adjust the pose so that the camera is parallel to the door. This experiment aims to measure the vertical distance between the camera and the door plane and calculate the ratio of the recognition frame's length to the picture's length when the camera is constrained by a parallel plane. The relationship between the two is described by the fitting curve, and the fitting curve is substituted according to the length ratio.



The camera position varies from N to M; corresponding to the projection on the Z-axis ZN respectively ZM represents the vertical distance from the camera to the door plane—the trail of opening the door.

Door Opening Trajectory of the Mobile Robotic Arm: The door-opening trajectory of the mobile robotic arm is limited to the rotation angle of the door axis and the door handle. The following Figure 5 shows the process of the door rotating from A to B around the door axis O. The rotation angle of the door handle around the door axis is the same as the door rotation angle

p. The gripping position of the door handle is set as the initial point of the trajectory, and the force analysis is conducted using the lever principle. Less force is required to turn the door handle as the force is applied away from the door handle pivot axis. Here, the grasping position is set to 3/4 of the length of the door handle, and the robotic arm can reduce the grasping force as much as possible to complete the grasping operation without falling off. Under these constraints, the trajectory of the robotic arm's end is determined.

Reinforcement Learning PPO Algorithm and Improved PPO Method: For the traditional control algorithm of the robotic arm, there are multiple solutions of target position, which makes the exploration efficiency of the end of the robotic arm inefficient with the increase of dimension. The reinforcement learning method is used to improve the control efficiency and stability of the robotic arm. During the interaction between the robotic arm and the environment, the reward feedback optimization strategy of the environment for the action in the current state is used to realize a better control strategy under specific performance indicators.

Principle of PPO Algorithm: The PPO (proximal policy optimization) algorithm can deal with large-scale and continuous space control, which is suitable for continuous actions such as robotic arm grasping. The key idea of the PPO algorithm is to limit the update range of the new strategy by adjusting the proportion of the new strategy and the old strategy to solve the problem that it is difficult to determine the value of the learning rate. This restriction prevents the strategy from converging when the learning rate is too large, and the training time is long when the learning rate is too low. This restriction can effectively improve the control efficiency of the robotic arm and improve the stability of the control.PPO is improved based on TRPO (trust region policy optimization) algorithm. In view of the low sampling efficiency of the TRPO algorithm, the PPO algorithm adopts the method of importance sampling to solve the problem. The algorithm can improve the efficiency and reliability of TRPO data when only first-order optimization is used.

Improvement of PPO Algorithm: In the reinforcement learning PPO method, the agent continuously obtains experience and rewards in action during the training process and trains the model through these data. Full data coverage cannot be achieved in the real environment, so the agent will think that the reward problem will prolong the training process and reduce the training efficiency. To reduce the appearance of sparse rewards, this paper is an improved method of reward shaping for PPO. This method randomly selects a certain number of states in the sequence as the new target when the sequence length reaches or exceeds the target during the reinforcement learning training process.

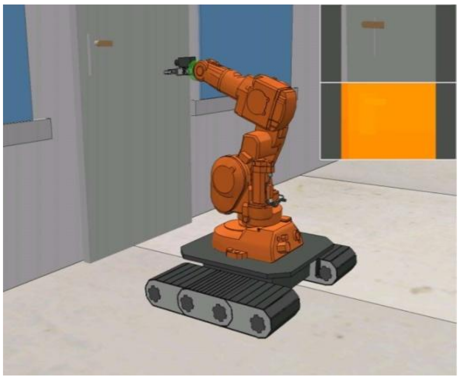
Definition of State Space and Action Space: The state space mainly includes the base of the robotic arm, the relative position between the end of the robotic arm and the target, the rotation coordinates of each joint of the robotic arm under limited conditions, and the state of the end of the robotic arm. Rd is set as the reward function of the state space, w represents the reward weight, and the reward function equation of the relative distance between the gripper and the target T is established. The action space design lies in the rotation angle of each joint of the base and the robotic arm and includes the rotation angle when the end clamping claw holds the target. The base, boom, jib, and wrist of the robotic arm rotate around the axis respectively and set the rotation angle around the axis and the corresponding value range.

Design of Reward Function**:** To solve the problem of sparse grasping rewards for robotic arms, formal rewards were used to improve learning efficiency. The state action is updated and trained according to the distance reward function network. The end gripper of the robotic arm can get the reward value of the target distance after each action. The training process of reinforcement learning tends to increase in the direction of reward, and the reward value increases gradually with the decrease of distance.

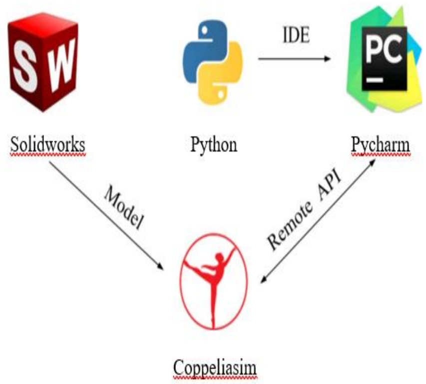
Result of Simulation Experiment: To assess the reliability of the method, a simulation environment and a realistic experimental scenario are experimentally set up. For the control task of opening doors on a mobile robotic arm platform, reinforcement learning algorithms such as TRPO, PPO and improved PPO are tested in the simulation environment. The reward values corresponding to the algorithms are used as key metrics to compare the merits. The models with good training results are applied to the recognition grasping task of the mobile robotic arm in a realistic scenario.

3. CONSTRUCTION OF THE SIMULATION ENVIRONMENT

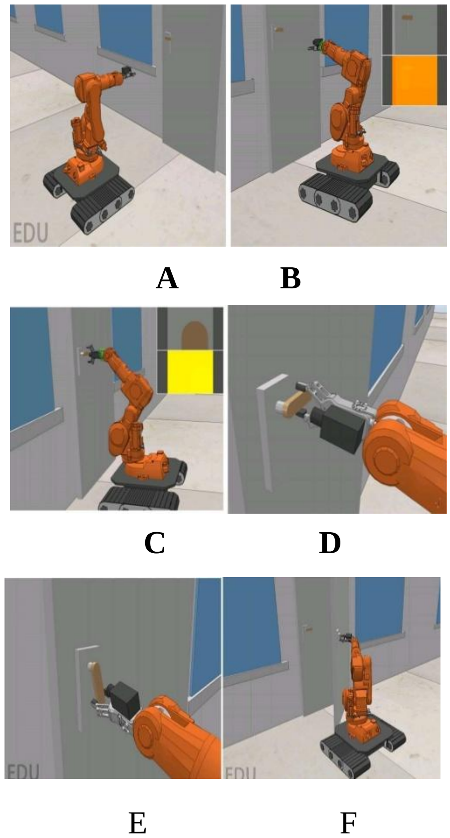
CoppeliaSim is a robotics simulation platform with an integrated development environment. The platform has a rich physics engine for building cross-platform simulation environments.



The main body of the mobile robotic arm is built from a combination of the arm, track, camera, and jaws in CoppeliaSim, and the mobile arm is sized to match the actual structure and dimensions. Python controls the movements of the mobile arm platform in CoppeliaSim through the API interface, providing a good experimental environment to verify whether the mobile arm can accurately follow the control procedure to open the door.



Join Simulation Experiments: The experimental simulation platform adopts the method of joint simulation, and Python controls CoppeliaSim through an API interface to realize the control of the motion simulation of the mobile robotic arm.



The effect of the simulation experiment. (a) The platform moving to the operation area. (b) Adjusting the position. (c) The platform captures the image. (d) The grasping operation. (e) Revolving door handle. (f) The door-opening operation. The experiment uses the proposed improved PPO algorithm control strategy in the experiment. a. shows the mobile robotic arm platform moving to the operation area and adjusting the position. b—d show the grasping operation process of the robotic arm reaching the grasping position according to the relative position information. e,f show the mobile robotic arm completing the door-opening operation.

Results and Discussion of Physical Experiment: The mobile robotic arm platform mainly includes the robotic arm, crawler car, camera, gripper, etc. First, after the mobile robotic arm platform reaches the working area, it adjusts to a state parallel to the door through its position and attitude information. The camera and the robotic arm form a hand—eye system. The mobile robotic arm platform determines the target of the door handle and calculates the relative position through the method of target detection. Finally, the model trained in the simulation environment is applied to the actual crawling experiment. In addition, the crawling success rates of TRPO, PPO, and improved PPO algorithms are compared to verify the effectiveness of the control strategy of improved PPO algorithms.

Object Detection and Positioning: Based on the PyTorch framework, the operating environment is set up, the dataset is trained using the YOLOv5 model, the number of training iterations of the dataset is set to 300 times, and the main data change as the number of iterations increases during the training process. The door-handle recognition and detection models extract large amounts of feature information from images.

The identification results show the frame with a better recognition frame selects the border of the door handle. The recognition rate is as high as 90%, and the experimental results show that the door-handle detection model based on the YOLOv5 model has the characteristics of high precision, good robustness, and fast speed.

Experiments on Opening Doors: The mobile robotic arm was tested in a real-life scenario. Figure 16a shows that the moving robotic arm platform moves to the door and adjusts the posture. Information through the camera to complete the recognition and positioning of the door handle.

The effect of the simulation experiment. (a) The platform moving to the operation area. (b) Adjusting the position. (c) Image recognition. (d) The grasping operation. (e) Revolving door handle. (f) The door-opening operation.

4. CONCLUSIONS AND FUTURE WORKS

This paper proposes an improved method of reward shaping for PPO, which increases the storage of experience values during training, thus reducing the exploration of useless environments by the robotic arm. This method is applied to door opening with a mobile robotic arm. The robotic arm identifies and locates the door handle through the YOLOV5 method and applies the model trained in the simulation environment to the real grasping task. To measure the performance of the proposed method, this paper has also compared PPO with TRPO and improved PPO by comparing the solution quality using a reward indicator and actual error. In future work, the application scenarios of mobile robotic arms will be expanded by adding various types of door handles.

ACKNOWLEDGEMENT

I am delighted to express my heartfelt appreciation to our departments head and staff, as well as family and friends. This paper is possible by their encouragement assistance and support.

REFERENCE

[1]. Liu, B. Recent Advancements in Autonomous Robots and Their Technical Analysis. Math. Probl. Eng. 2021, 2021, 6634773.[CrossRef]

[2]. Cardenas, A.; Boodwine, B.; Skaar, S.; Seelinger, M. Vision-Based Control of a Mobile Base and On-Board Arm. Int. J. Robot. Res.2003, 22, 677-698. [CrossRef]

[3]. Huang, S.-J.; Huang, J.-C. Vision guided dual arms robotic system with DSP and FPGA integrated system structure. J. Mech. Sci. Technol. 2011, 25, 2067-2076. [CrossRef]

[4]. Sharma, A.; Noel, M.M. Design of alow-costfive-finger anthropomorphic robotic arm with nine degrees of freedom.Robot. Comput.lntegr.Manuf.2012,28,551—558[CrossRef]

[5]. Ali, M.M.; Hui, L.; Stoll, R.; Thurow, K. Arm grasping for mobile robot transportation using Kinect sensor and kinematic analysis. In Proceedings of the 2015 IEEE International Instrumentation and Measurement Technology Conference (12MTC), Pisa, Italy, 11-14 May 2015; pp. 516-521.

[6]. Renjith, R.; Reshma, R.; Arun, K.V. Design and implementation of traffic sign and obstacle detection in a self-driving car using SURF detector and Brute force matcher. In Proceedings of the 2017 IEEE International Conference on Power, Control, Signals and Instrumentation Engineering (ICPCSI), Chennai, India, 21—22 September 2017.

[7]. Belda, K.; Rovny, O. Predictive control of 5 DOF robot arm of autonomous mobile robotic system motion control employing mathematical model of the robot arm dynamics. In Proceedings of the 2017 21st International Conference on Process Control (PC), Strbske Pleso, Slovakia, 6—9 June 2017.

[8]. Chung, W.; Rhee, C.; Shim, Y.; Lee, H.; Park, S. Door-Opening Control of a Service Robot Using the Multifingered Robot Hand. IEEE Trans. Ind. Electron. 2009, 56, 3975-3984. [CrossRef]

[9]. Arduengo, M.; Torras, C.; Sentis, L. A Versatile Framework for Robust and Adaptive Door Operation with a Mobile Manipulator Robot. arXiv 2019, arXiv:1902.09051.

[10]. Finn, C.; Levine, S. Deep Visual Foresight for Planning Robot Motion. In Proceedings of the 2017 IEEE International Conference on Robotics and Automation (ICRA), Singapore, 29 May—3 June 2016.