arXiv:2504.04148v1 [cs.RO] 5 Apr 2025

A General Peg-in-Hole Assembly Policy Based on Domain Randomized Reinforcement Learning

Xinyu Liu[1], Aljaz Kramberger[1], Leon Bodenhagen[1]

[1]The Maersk Mc-Kinney Moller Institute, University of Southern Denmark, Odense, Denmark.

Contributing authors: xinl@mmmi.sdu.dk;

Abstract

Generalization is important for peg-in-hole assembly, a fundamental industrial operation, to adapt to dynamic industrial scenarios and enhance manufacturing efficiency. While prior work has enhanced generalization ability for pose variations, spatial generalization to six degrees of freedom (6-DOF) is less researched, limiting application in real-world scenarios. This paper addresses this limitation by developing a general policy GenPiH using Proximal Policy Optimization(PPO) and dynamic simulation with domain randomization. The policy learning experiment demonstrates the policy's generalization ability with nearly 100% success insertion across over eight thousand unique hole poses in parallel environments, and sim-to-real validation on a UR10e robot confirms the policy's performance through direct trajectory execution without task-specific tuning.

Keywords: Deep reinforcement learning, Sim-to-real, Peg-in-hole assembly

1 Introduction

With rapid advancements in robotics and artificial intelligence (AI), robots are increasingly deployed in various industrial applications, particularly for automating mass production to enhance production efficiency. Peg-in-hole assembly, which is the fundamental operation of robotic assembly, becomes important and attracts research attention [1, 2]. Recently, the integration of DRL has enabled notable improvements in policy's generalization ability in this task [3, 4]. Typically, DRL-based policy is trained to process observations, such as target poses, and output corresponding actions, like joint positions or end-effector movements, to guide the robot in task execution [5].

These policies can adapt to changing working scenarios, making them highly effective in addressing environmental uncertainties. This adaptability is particularly beneficial for flexible and customized manufacturing. Previous studies have successfully applied DRL to peg-in-hole assembly, achieving strong performance in generalizing across varying object poses[6, 7]. However, most research focuses on various planar positions, and the peg is already roughly aligned with the hole. Generalization to various spatial pose with 6 degrees of freedom (DOF) has been less explored.

This study addresses the gap by employing a DRL-based learning method to implement peg-in-hole assembly with variations in the hole poses. A simulation environment including Universal Robot UR10e robot and Cranfield benchmark [8] models is constructed using NVIDIA's Isaac Sim and Isaac Lab [9] as the extension for training the assembly policies. The PPO algorithm is used for policy learning, as it is known for its stability and capability to process continuous data. The trained policy is then deployed in a real-world setup for experimental validation. In this paper, we outline the following contributions:

- A general-purpose simulation environment for training assembly policies.
- An assembly policy with generalization ability to various spatial hole pose.

2 Related Work

2.1 Deep Reinforcement Learning-Based Robotic Control

DRL-based methods have recently become increasingly popular in robotics because of their capacity to process high-dimensional and continuous data. It combines deep learning and reinforcement learning methods, using neural networks as the policy to process high-dimensional observation and output continuous actions. It has been applied for complex real-world robotic manipulation tasks where the observation space stretches over multiple dimensions, and action space requires continuous values[10]. Generalization of robot tasks is a popular research topic, which gained traction in the past with research of statistical methods [11], whereas today, novel simulation-based DRL methods are taking the forefront. Unlike traditional one-off motion planning or control methods designed for single robotics tasks, learned policies—when appropriately trained—can generalize across multiple tasks in unstructured environments and unknown scenarios. Actor-critic (AC) algorithms are one of the most popular DRL algorithm types, including Proximal policy optimization (PPO) [12] and Soft Actor-Critic (SAC) [13]. PPO is more stable with generalized advantage estimator [14] while SAC is more efficient in complex tasks due to entropy regularization. In this paper, the PPO algorithm is used to learn the PiH task.

2.2 Peg-in-Hole Assembly

Peg-in-hole assembly is a fundamental industrial operation that has been researched for decades. These studies mainly focus on assembly performance, including assembly precision and generalization ability to various working scenarios [5]. For peg-in-hole and similar tasks, researchers apply the Deep Deterministic Policy Gradient (DDPG) [15] algorithms to train the control policy for precise timber assembly [16]. In terms