10th Student Symposium on Mechanical and Manufacturing Engineering, 2022

Learning Robot Skill Sequences with Reinforcement Learning for Off-Planet Assembly

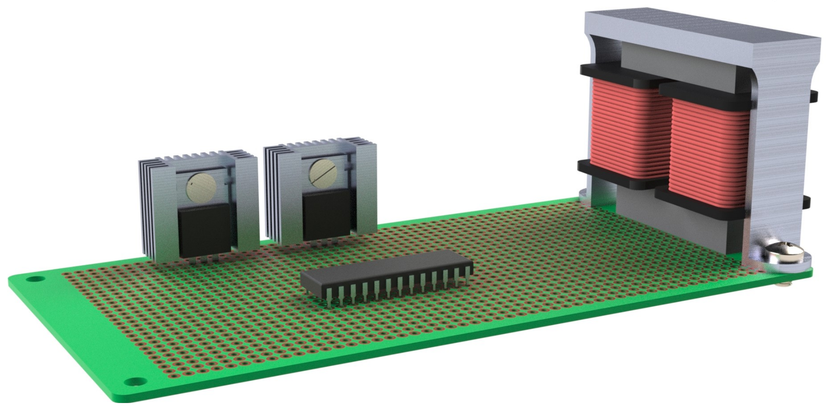
P. L. Møller, V. Druskinis, F. J. Christensen, M. L. Debijadi
Department of Materials and Production, Aalborg University, DK

1. Introduction

Through recent years there has been an increased focus on developing technologies that allow for In-Space-Assembly (ISA) due to the increased flexibility of operations it provides.
Any colonization of other planetary bodies, such as Mars, will rely on establishing In-Situ Resource Utilization (ISRU) to become self-sufficient, which has a high level of synergy with assembly-from-parts. Autonomous robotic systems are therefore highly relevant as this type of low-level assembly will require complex and flexible technology.
Finding an optimal assembly sequence is an NP-hard problem which can lead to extensive computation time for an exhaustive search for the single optimal solution. By utilizing RL, an agent can be used to reduce this search time by building up experience through trial and error.

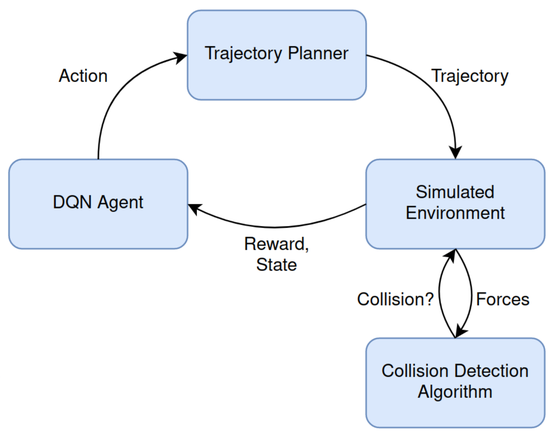
2. Assembly Product

Various products are relevant to colonization of other planetary bodies, however, power generation and infrastructure is a general predecessor to most other necessary technologies and therefore a simplified inverter will be used as a standardized assembly board

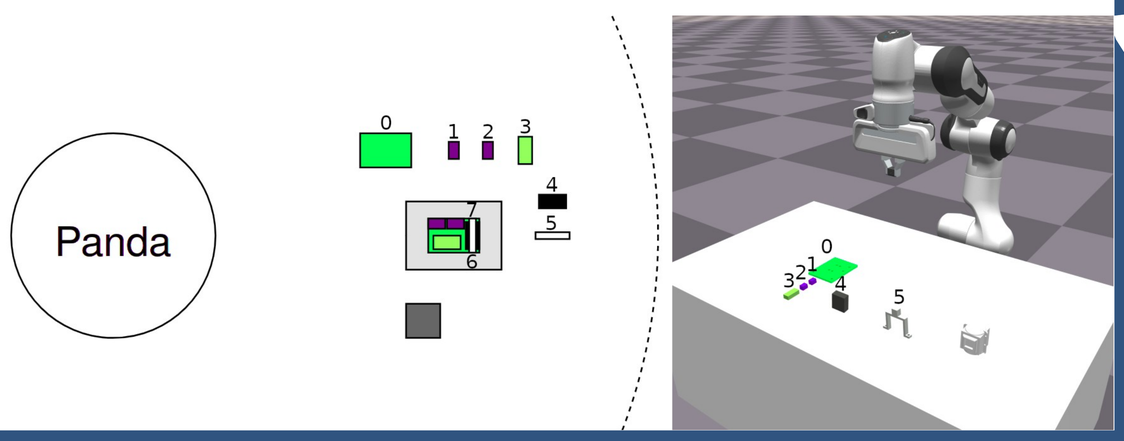


3. Experimental Setup

A DQN agent will be trained to output an action, which will be interpreted by the trajectory planner. Throughout the traj-Ectory, forces will be meas-ured and fed into a collision detection algorithm which will determine feasibility of the action. If no collisions are det-ected, a new state and a rewar will be computed and another iteration will be carried out.



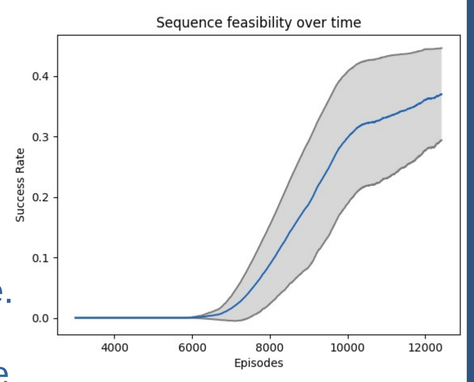
The environment will be simulated through Nvidia Isaac Gym, as visualized below.



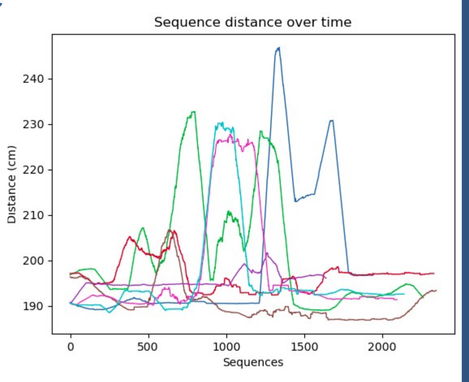
4. Experiments

The DQN was trained over ten sessions for 20000 steps, after which the training itself was evaluated to establish whether the agents were learning to act as intended, followed by an assertion of optimality for the best output sequences.

From the feasibility of output sequences through the training it could be seen that the agents would learn to generate com-plete sequences after approximately 6000 steps and tended to converge towards approximately 40% success rate.



From the minimum distance convergence test there were no clear signs of con-verging towards the real optimal solution after sequences could be completed, however, it could be seen that the exp-loration of the agents would occasionally cause them to diverge, after which they could converge back towards sequences of lower distances.



Finally, the optimality of the best sequences output by each training session were evaluated by exhaustively searching for the real optimal solution. Comparisons can be seen in the table below.

Test	Output Sequence	Distance	Deviation
1	[0,1,4,5,2,3,8,9,10,11,6,7,13,14,15]	1.880	0.70%
2	N/A	N/A	N/A
3	[0,1,2,3,4,5,8,9,10,11,6,7,13,14,15]	1.867	0.00%
4	[0,1,4,5,2,3,8,9,10,11,6,7,13,14,15]	1.880	0.70%
5	[0,1,2,3,4,5,6,7,8,9,10,11,13,14,15]	1.885	0.96%
6	[0,1,2,3,4,5,8,9,10,11,6,7,13,14,15]	1.867	0.00%
7	[0,1,2,3,4,5,8,9,10,11,6,7,13,14,15]	1.867	0.00%
8	[0,1,6,7,2,3,8,9,4,5,10,11,13,14,15]	1.939	3.86%
9	N/A	N/A	N/A
10	[0,1,2,3,4,5,8,9,10,11,6,7,13,14,15]	1.867	0.00%

4. Conclusions

A reinforcement learning algorithm can be designed to generate feasible assembly sequences based on a fixed discrete action space, while collision detection algorithms can be used to determine feasibility the actions. Reward schemes based on distances can be used to incentivize the generation of feasible and optimal assembly sequences. A critical point was found for the agents, which was learning to apply a tool change and continue the assembly afterwards, where either they would succeed and complete the sequence, or fail and become stuck in a local minima. Experiments indicated that agents showed signs of learning as intended. Success rates of output sequences increase over time and sequences converge near the real optimal assembly sequence.

Acknowledgement

The authors of this work gratefully acknowledge Grundfos for sponsoring the 10th MechMan Symposium

AALBORG UNIVERSITY
DENMARK
Department of Materials and Production
www.mp.aau.dk

GRUNDFOS