**Digital Twin-Driven Deep Reinforcement Learning**

**for Adaptive Robot Control in Robotic Construction**

1. **Introduction**
   * Background
     + Construction is known as one of the most stressful occupations due to its physically and psychologically demanding tasks performed in a hazardous and difficult work environment. In addition, traditional labor-intensive production methods are becoming less productive due to lack of skilled labors and an aging workforce.
     + Recent advancements in robotics technology is changing the nature of work. Autonomous robots have been applied to help human workers and/or replace their repetitive works to address the productivity issues.
   * Problem Statement
     + To improve the performance of the autonomous robots in construction, an adaptive robot control algorithm is needed because the construction site is dynamic (i.e., unstructured and evolving).
   * Knowledge gap
     + In recent years, several researchers have demonstrated the potential of deep reinforcement learning (DRL), which enables robot agent to learn an optimal control policy through repetitive interactions with a simulation environment, for a more adaptative robot control method. Once the agent learned adaptive control policy in the simulation environment, its trained policy can be deployed on the real robot control.
     + However, it has not been known thoroughly whether a DRL agent can find an adaptive control policy in dynamic construction environment.
   * Research Objective
     + The authors develop and test a digital twin-driven DRL framework to investigate DRL’s potential for adaptive robot control in dynamic construction site.
     + The digital twin can capture dynamic site conditions in near real-time and creates a virtual simulation environment where DRL agent can interact with.
     + The authors test whether DRL agent can learn optimal policy stably in the dynamic digital twin environment and whether such policy shows adaptive control performance in response to dynamic site environment.
2. **Potential of DRL for robot control in construction** 
   * Autonomous robot in construction
     + Current method and limitation
   * Potential of DRL for robot control in construction
     + Active interaction with simulation environment
     + Adaptive and autonomous control
   * Knowledge gap
     + DRL model with well-designed rewards can adapt well to a certain stationary learning environment (e.g., Atari games, Go). However, when the environment changes, even a well-trained DRL model may not be able to adapt to a changing environment; thus, producing reliable results will be challenging because such model is typically only adapted to a specific environment (Won et al 2020). For these reasons, it has not been known thoroughly whether a DRL agent can find an adaptive control policy stably, and whether such a policy gives us reliable control performance in response to dynamic construction site environment.
     + To train DRL agent adaptive control policy, realistic construction simulation environment is also needed.
3. **Digital twin-driven deep reinforcement learning for adaptive robot control**
   * Research objective
     + To develop and test a digital twin-driven DRL framework to investigate DRL’s potential for adaptive robot control in a dynamic construction site.
     + A digital twin is a real-time virtual replica of physical assets, which represents the project’s performance, the geometry of assets, and resource status. With the aid of empowered IoT sensors, the digital twin can automatically update ‘as-built’ BIM based on ‘as-designed’ BIM, thus it can show all dynamic site conditions near real-time. In addition, the digital twin can be used to simulate “what-if” scenarios to provide high-fidelity and rich learning environment for the DRL agent.
     + The authors test whether a DRL agent can learn optimal policy in a dynamic site environment stably and whether such policy shows better adaptive control performance than a rule-based behavior in three different site environment profiles.
   * Framework overview

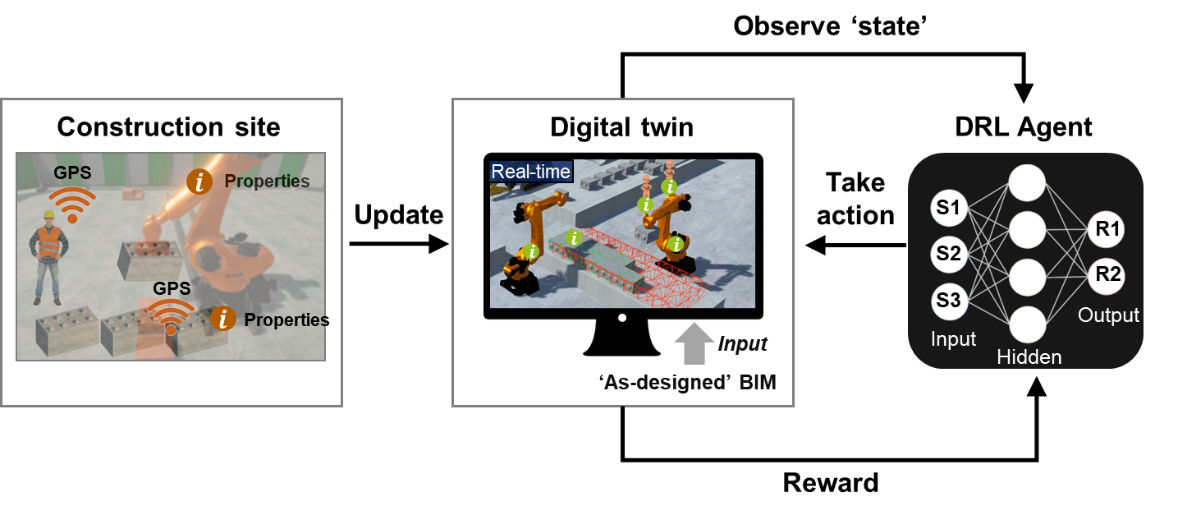


Fig 1. Digital twin-driven DRL framework

* + - Real-time environment updates (based on BIM).
    - A DRL agent can interact with the dynamic digital twin environment and test their actions in different ‘what-if’ scenarios. Specifically, the DRL agent earns positive or negative rewards for its robot control actions given that observation states and updates the parameters of the policy network. As a result, the agent can find an adaptive robot control policy that maximizes the project performance by maximizing cumulative rewards in response to dynamic site environment.

****

Fig 2. DRL learning in digital twin

* + Deep reinforcement learning via proximal policy optimization (PPO)
    - Markov decision processes
    - Policy gradients
    - A proximal policy optimization (PPO) algorithm, one of the most efficient policy gradient methods for continuous action space, is applied for the DRL algorithm (Schulman et al. 2017). PPO enables multiple updates per minibatch sample to promote sample efficiency and guarantees the stability of policy optimization by limiting the update amplitude of policy (Zhang et al. 2019).
    - Deep neural network architecture for PPO
    - State, Action, and Reward Function

1. **Experiments and Results**
   * + An experiment is designed to explore the capability of the suggested framework. In this experiment, the authors created a virtual robotic construction scenario where real robots are synchronized with virtual robots in the digital twin via robot operating system (ROS) which is a set of software libraries and tools that can be used to control motion of real robots (Quigley et al. 2009). Thus, PPO agent can interact with more realistic robot actions in the digital twin.
     + In this scenario, the authors tested the stability of policy training by checking learning curve of three indicators (cumulative reward, building success rate, and construction time). Then, we tested whether such trained policy shows more adaptive control performance than the rule-based model’s behavior in three different dynamic site profiles (dynamic delivery status, dynamic brick transport status, dynamic building speed status).
   * PPO model training in simulation scenarios
     + The authors trained a PPO agent with a simulation scenario. The scenario is that one loading robot load prefab interlocking bricks to a loader truck, and it delivers them from the stockyard to assembly location (bridge construction site), and another two assembly robots assemble them to construct a small-scale mock-up bridge (Fig.3). The bridge consists of a total of 49 bricks with 4 different types.

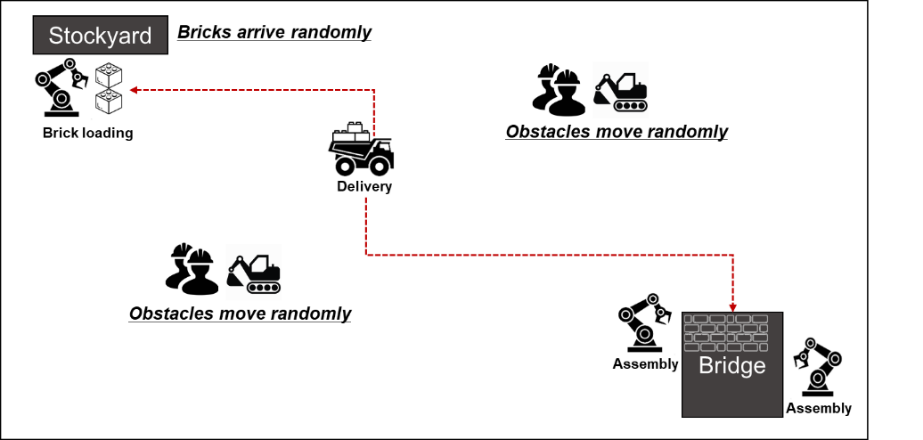


Fig 3. Simulation scenario overview

* + - In this simulation scenario, the PPO agent take actions both for loading robot and assembly robots for given site states (brick status, truck status, and assembly status) to train its policy network. For each simulation tick, the agent takes one action per robot. For the loading robot, it decides which bricks (type 1 to 4) and how many (1 to 4) to load on the loader truck. For assembly robots, it decides in what order to assemble the bricks delivered via the loader truck and whether to send the truck back or wait. The agent gets feedbacks for their actions according to the robots’ performances (productivity and idle time) and overall project performance (bridge construction time), and the feedbacks update the policy network (Fig.4).

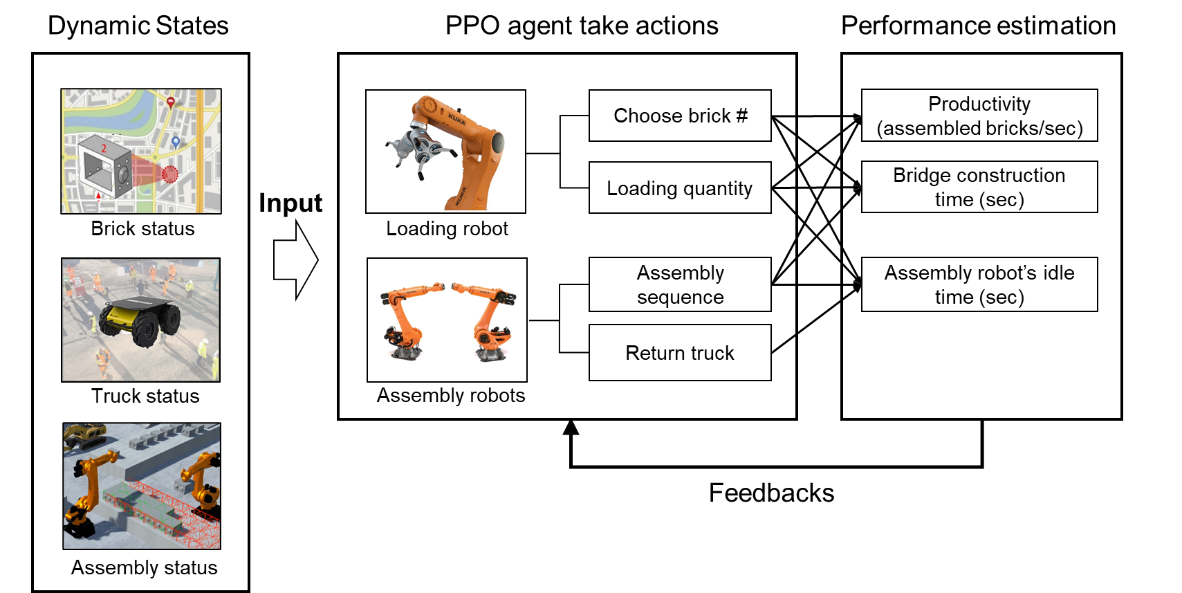


Fig. 4 PPO training scenarios (to be revised)

* + - Hyper parameter setting (Training Epoch, hours, 45 parallel learning, gamma, theta, alpha (learning rate), batch size)
  + Results of PPO training for optimal control policy

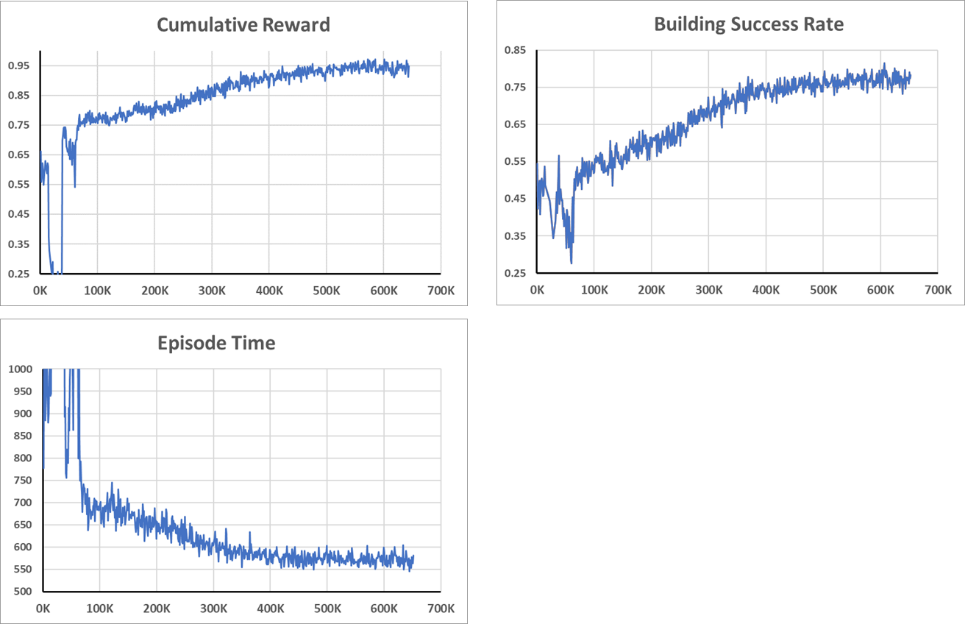


Fig. 5 Learning curve for (a) cumulative reward (b) building success rate (c) episode time (to be revised)

* + PPO policy’s adaptive control performance in the dynamic site environment
    - The authors test the trained policy’s adaptive control performance in three different dynamic environment profiles and compared it with rule-based model’s behavior.
    - There are three environmental parameters to demonstrate dynamic site conditions: 1) brick delivery ratio for changing availability of bricks in the stockyards; 2) loader truck’s speed for changing brick transport efficiency on site; 3) assembly speed of assembly robots for changing bridge assembly efficiency.
    - The average performance of the trained model

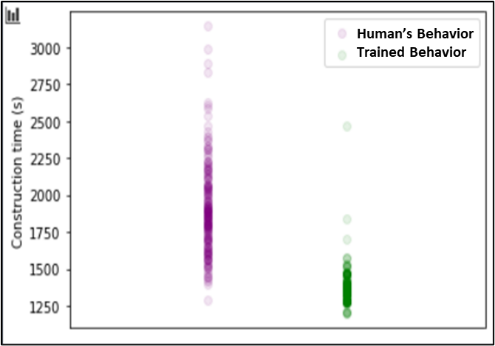
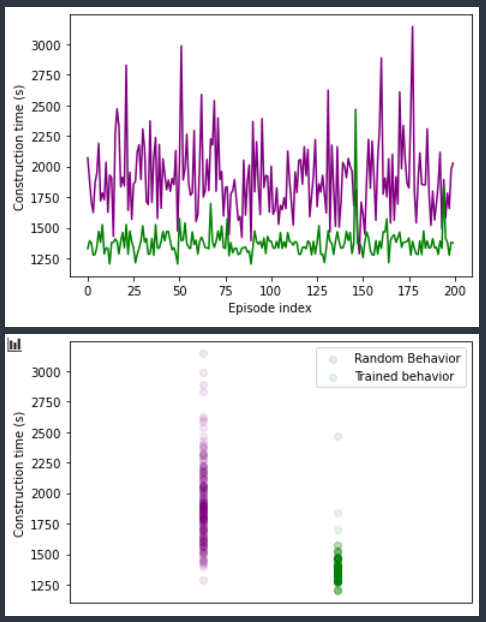


Fig. 6 Average project performance (bridge construction time) between (a) trained model and (b) rule-based model (to be revised)

* + - Adaptive control in three dynamic environment profiles and comparison between trained behavior and rule-based behavior

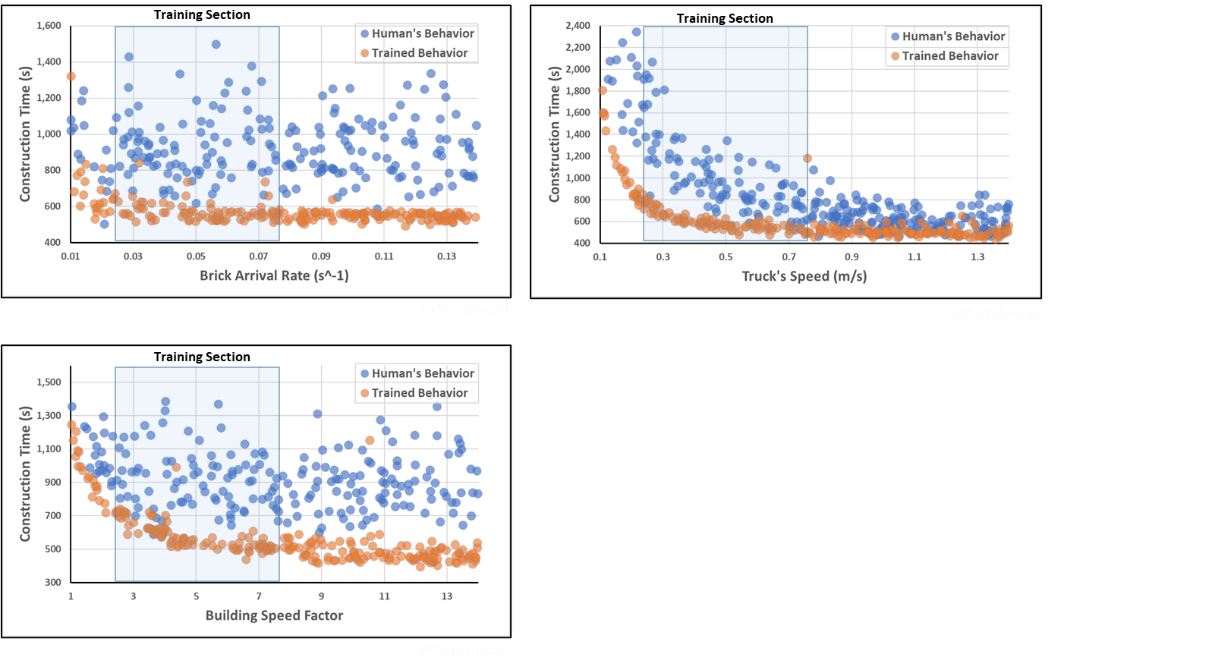


Fig. 7 Adaptive control performance in dynamic environment profiles (a) brick arrival (b) truck’s speed, (c) building speed factor(to be revised)

1. **Discussion**
   * Importance of DRL-based adaptive control policy in construction
     + Multiple robot controls
     + Generalizability to other equipment
     + Adaptive decision-making to unseen or unpredictable site conditions
   * Potential of digital twin for future DRL application.
     + Real-time site observations and virtualize to reduce local optima while DRL learning and real-world application.
     + Possibility of online learning-based DRL relearning in construction
   * Limitation and future research
     + When construction project is not fully observable and if multiple robots are not fully connected each other, single agent based DRL is not realistic.
2. **Conclusion**
3. **Acknowledgments**

**References**