Imitation Learning for Construction Robotics: A Case Study on Wood Joining Bimanual Manipulation

Yuezhen Gao[1], Huyue Li[1], Hridey Jain[1], Ali Golabchi[1] and Qipei Mei[1]
[1] Department of Civil and Environmental Engineering, University of Alberta, Edmonton, Canada.
Email: yuezhen@ualberta.ca

INTRODUCTION

The construction sector faces persistent challenges including safety hazards, labour shortages, inefficiency, and material waste. Robotic platforms have improved automation in bricklaying, drilling, and plywood wall assembly. However, most existing systems rely on pre-programming that integrates perception, planning, and control modules and lack adaptive inference ability, resulting in limited flexibility for complex construction tasks.

To advance robotic solutions for construction automation, we present a case study on applying imitation learning to a wood-joining task. This woodworking operation involves grasping the wooden components with both hands, aligning them carefully, and interlocking them. Our results demonstrate the ability of a bimanual robotic platform to autonomously assemble wooden joints, providing practical insights into deploying and fine-tuning imitation learning models for construction robotics.

METHODS

Dual-arm setups are popular in robotic learning research due to their flexibility, transferability, and similarity to human operators [1]. In this study, we use the mobile ALOHA platform as our dual-arm setup for data collection. The platform integrates joint-space mapping between the leader and follower robot arms [2]. To set up the environment, tongue-and-groove joints were prepared, and a black tablecloth with four markers was used as the background to enhance visual contrast and help object localization (Fig 1).

Imitation learning is a machine learning method in which agents learn tasks from human demonstrations. We deployed the Action Chunking with Transformers (ACT), an imitation learning model that trains a transformer policy as the decoder of a conditional variational autoencoder to model a generative distribution over action sequences [2]. ACT predicts short sequences of low-level actions from images and robot states, mitigating compounding errors compared to single-step behaviour cloning (BC), and showing robust execution in longer-horizon tasks.

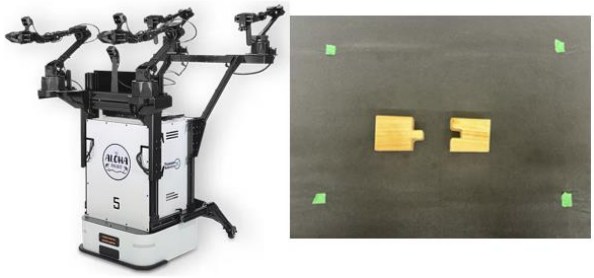

Fig 1 Data collection platform and experiment setup.

RESULTS AND DISCUSSION

We fine-tuned the hyperparameters of the ACT policy with 100 demonstrations and evaluated multiple models. To improve performance, we co-trained our task with the ALOHA dataset [3] and made a left-right tasks-switch dataset. The best model was obtained with a batch size set to 32, chunk size of 75, fps set to 50, and co-trained for 200k steps. Selected test results are shown below (Table 1). During testing, we observed two factors that affected inference results: (a) FPS. Running evaluation at higher rates is beneficial only if the system can process every frame in real time. Due to hardware constraints, evaluation was conducted at 15 FPS. b) max_relative_target. It defines the maximum angular change permitted for a motor in a single step. While it improves motion smoothness, it also introduces compounding error, resulting in less success rate.

CONCLUSIONS

In this case study, we fine-tuned ACT model with wood-joining demonstrations and evaluated it on the dual-arm mobile ALOHA platform, showing the bimanual manipulation ability of automated wooden joint assembly. We also observed the influence of certain parameters and will collect more demonstrations for diverse woodworking tasks to further enhance construction automation.

REFERENCES

[1] Smith C. Rob Auton Syst 60: 1340-53, 2012.
[2] Zhao TZ et al. arXiv 2304.13705: 1-10, 2023.
[3] Fu Z et al. arXiv 2401.02117: 1-10, 2024.

Table 1: Success rate for the wood joining task.

Chunk size	Co-train	Left Pick	Right Pick	Insertion
50	No	35%	15%	0%
50	Yes	35%	20%	10%
75	Yes	81%	94%	71%