Workers Demo, Robots Learn: Inverse Reinforcement Learning with Cross-Embodiment Videos for Construction Robots

Lei Huang, Tong Ren, and Zhengbo Zou

Department of Civil Engineering and Engineering Mechanics, Columbia University

IROS 2026 PITTSBURGH

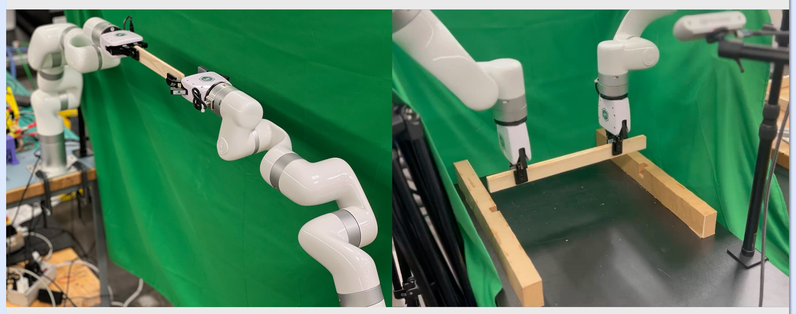
5th Workshop on Future of Construction

Motivation

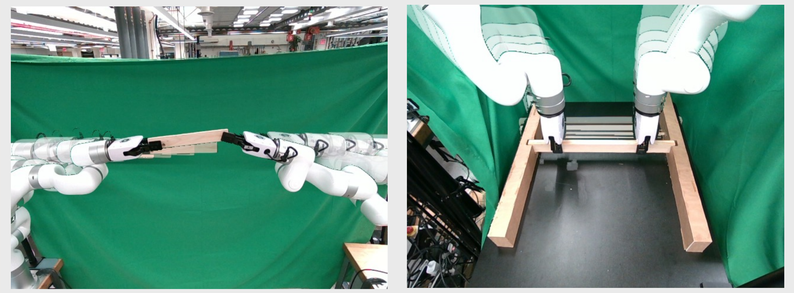
Construction robots obtain higher autonomy by learning policies via **Reinforcement Learning (RL)**, but

1) Each policy requires a proper *dense reward function*
2) Design of such reward function requires *iterative tuning*
3) *Handcrafting reward is not scalable* to the large repertoire of construction tasks

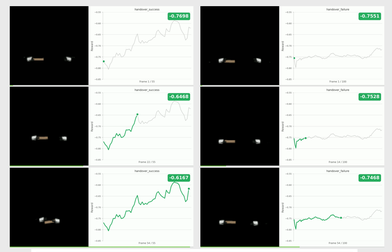
Tested Tasks

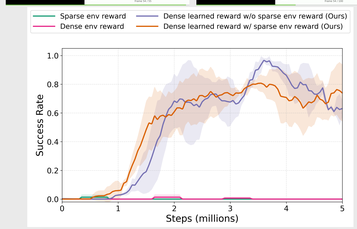


Experiments



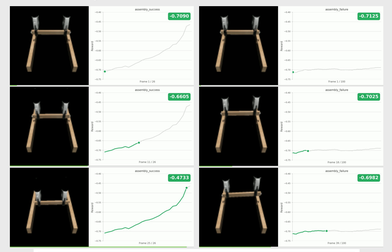
Handover Reward: Success vs Failure

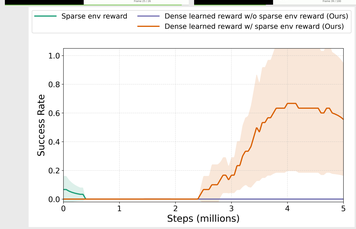




Reward Setting	Success Rate (%)
Sparse Env. Reward	0
Dense Env. Reward	0
Learned Reward	68
Learned Reward w/ Sparse Env. Reward	72

Assembly Reward: Success vs Failure

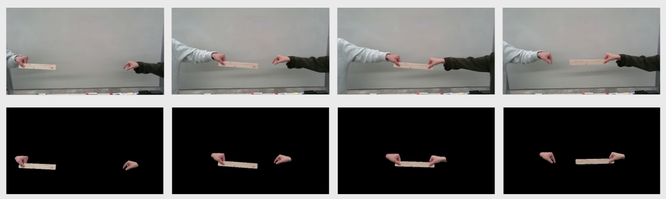




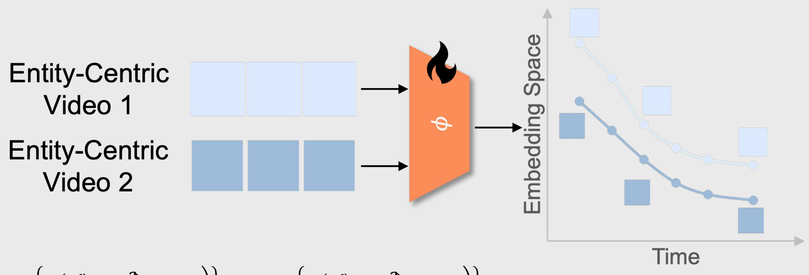
Reward Setting	Success Rate (%)
Sparse Env. Reward	0
Learned Reward	0
Learned Reward w/ Sparse Env. Reward	64

Inverse Reinforcement Learning with Human Videos

1. Object-Centric Segmentation of Human Videos



2. Reward Learning from Human Videos



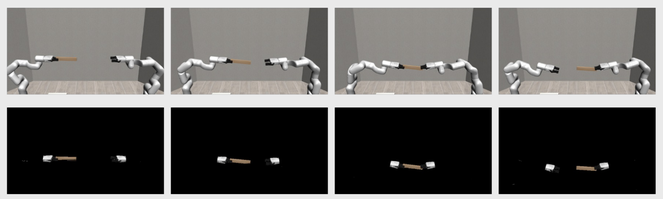
Temporal Alignment

$V_i = \left\{\phi\left(I_i^1\right), \phi(I_i^2) \ldots \phi(I_i^{L_i})\right\}$ and $V_i = \left\{\phi\left(I_i^1\right), \phi(I_i^2) \ldots \phi(I_i^{L_i})\right\}$

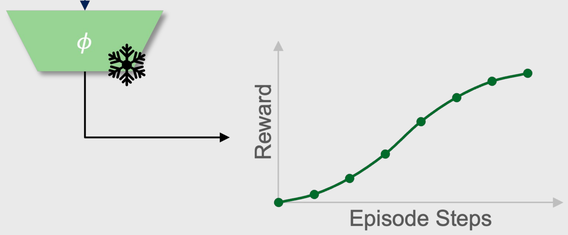
$\widetilde{V_{ij}^t} = \sum_k^{L_j} \alpha_k V_j^k$ where $\alpha_k = \frac{e^{-\left\|v_i^t - v_j^k\right\|^2}}{\sum_k^{L_j} e^{-\left\|v_i^t - v_j^k\right\|^2}}$ $\mu_{ij}^t = \sum_k^{L_i} \beta_{ijt}^k k$ where $\beta_{ijt}^k = \frac{e^{-\left\|\widetilde{v_{ij}^t} - v_i^k\right\|^2}}{\sum_k^{L_i} e^{-\left\|\widetilde{v_{ij}^t} - v_i^k\right\|^2}}$

Loss function: $L_{ij}^t = \left(\mu_{ij}^t - t\right)^2$

3. Reinforcement Learning with Learned Reward



Segmented Episode Rollout



Goal Embedding: $g = \sum_{i=1}^N \phi(I_i^{L_i})/N$

Learned Reward Function: $r(s) = 1/k \left\|\phi(s) - g\right\|_2^2$

where $k = 1/N \sum_{i=1}^N \left\|\phi\left(I_i^1\right) - g\right\|_2$

Conclusions

- Construction workers' operation videos contain implicit reward signaling task progress
- Temporal encoder learns a latent space where distances to goal embedding reflects the progress
- The latent distances are effective reward for training bimanual construction robots via RL