UNIVERSITY OF TORONTO
FACULTY OF APPLIED SCIENCE & ENGINEERING
SYNCERE AI

CLARA: Compliant policy for Low-cost Assistive Robotic Arms

Eric Xie, Haitong Wang, Matthew Lisondra, Angus Fung, Aaron Hao Tan

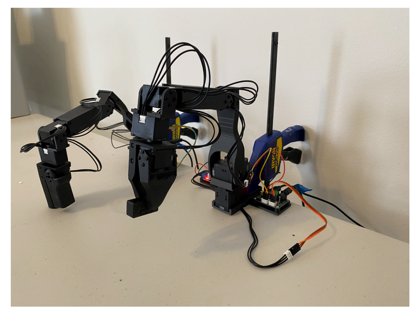
GitHub

Syncere AI

Motivation

How can we train a robot manipulator to be compliant and safe?

- Robots must react in real time to disturbances
- External force sensors are expensive (Franka arm costs $10k+). Koch LeRobot is an open source 5-DOF robotic arm for around $200

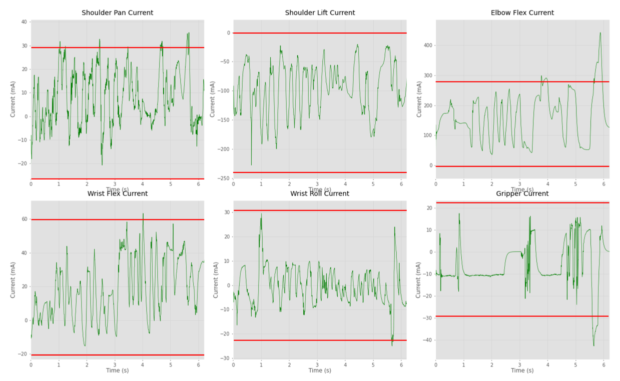


Koch arm V1.1

Koch playing chess

Contributions

1) CLARA, an end-to-end compliant control pipeline, combining SOTA U-Net diffusion model and impedance controller trained in RL.
2) Bilateral teleoperation to provide operator force feedback by reading follower arm's current to detect contact.

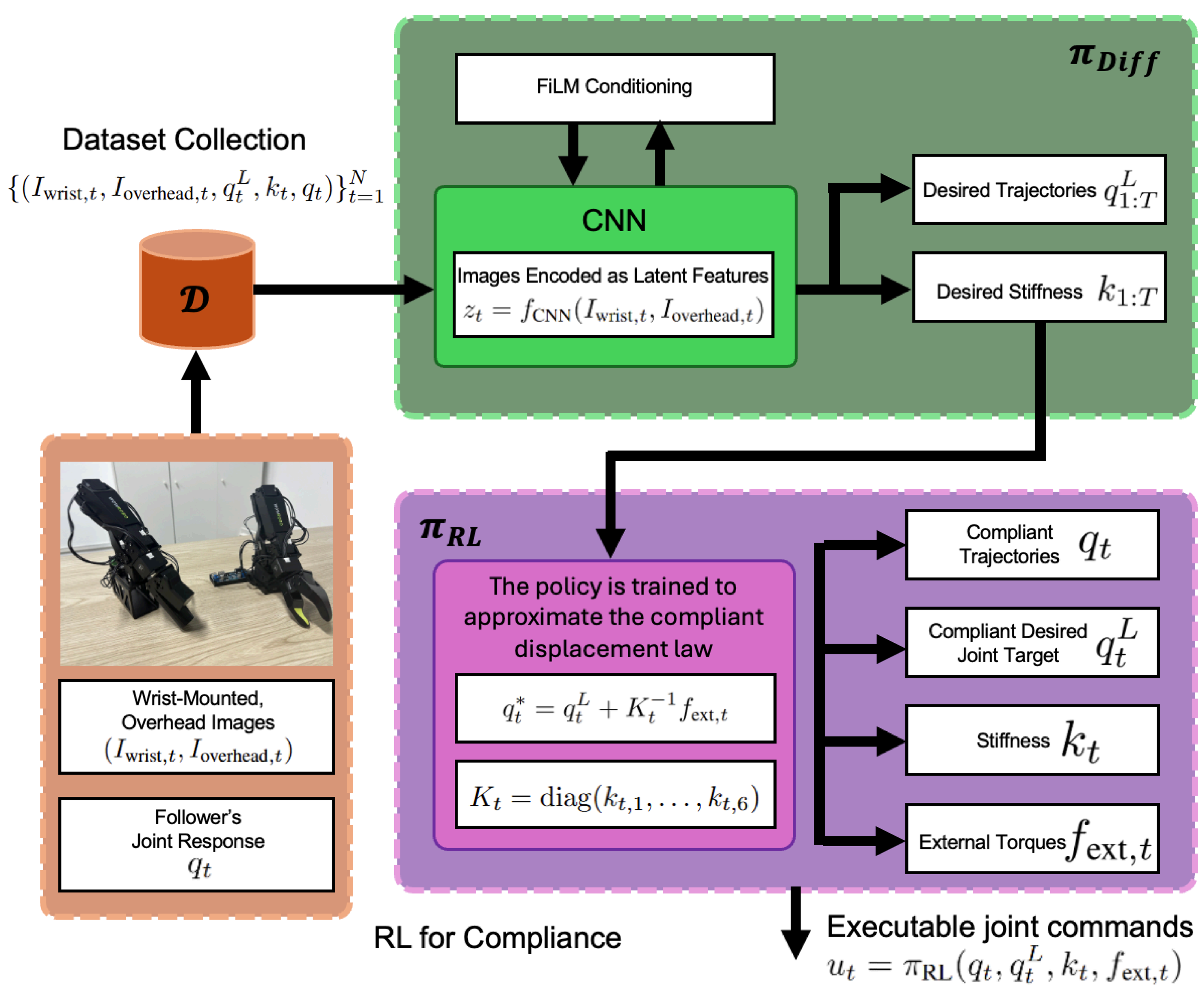


: Free movement max threshold

: Real time current reading

Pipeline

Diffusion-Based Planning



RL for Compliance

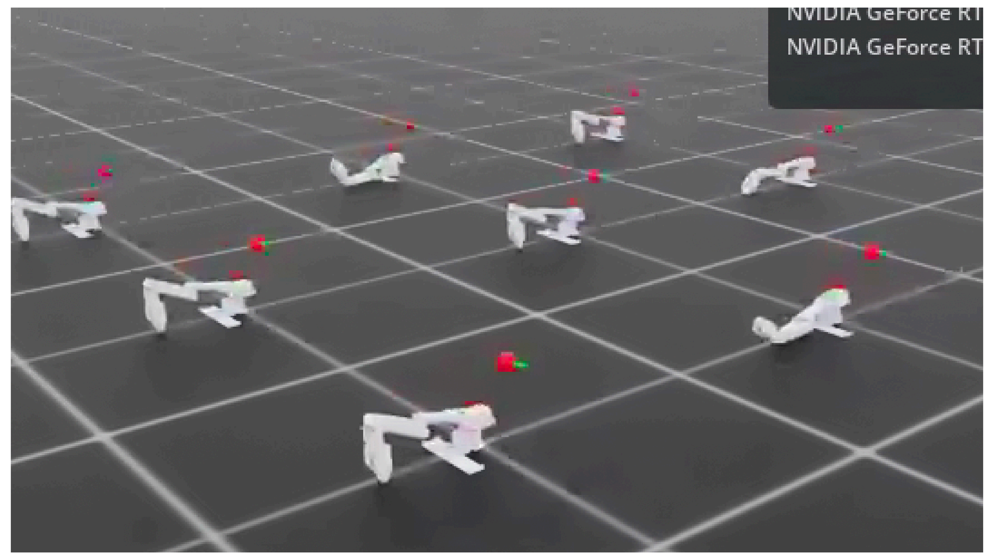
Executable joint commands

$u_t = \pi_{\text{RL}}(q_t, q^L_t, k_t, f_{\text{ext},t})$

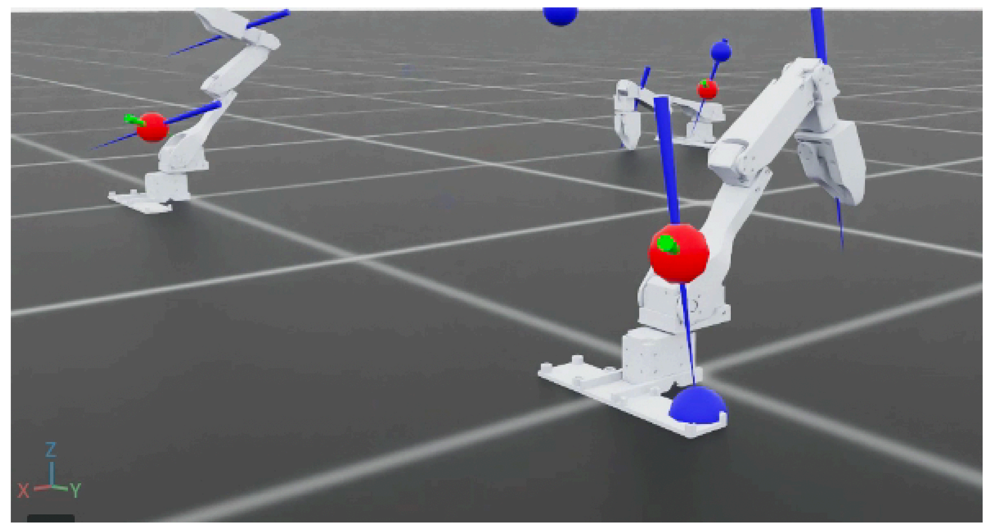
RL Training

Training the impedance controller with curriculum learning in NVIDIA's Isaac Lab

Stage 1: Position only



Stage 2: Sample external force



Reward function:

$$Reward^{(t)} = -w_1 \|\mathbf{x}^{(t)}_{\text{curr}} - \mathbf{x}^{(t)}_{\text{target}}\|_2 + w_2 \left[1 - \tanh\left(\frac{\|\mathbf{x}^{(t)}_{\text{curr}} - \mathbf{x}^{(t)}_{\text{target}}\|_2}{\sigma}\right)\right] - w_3 \cdot 2\arccos\left(\left|\langle \mathbf{q}^{(t)}_{\text{curr}}, \mathbf{q}^{(t)}_{\text{target}} \rangle\right|\right) - w_4 \|\mathbf{u}^{(t)} - \mathbf{u}^{(t)}\|_2$$

- If a force is sampled, the equilibrium distance computed by Hooke's law is added to target position

Important considerations

- Constrained orientation sampling due to Koch's limited range of movement
- Domain randomization, observation & action noise, random action delays to overcome sim-to-real gap
- Using Weight and Biases (wandb) to visualize training results

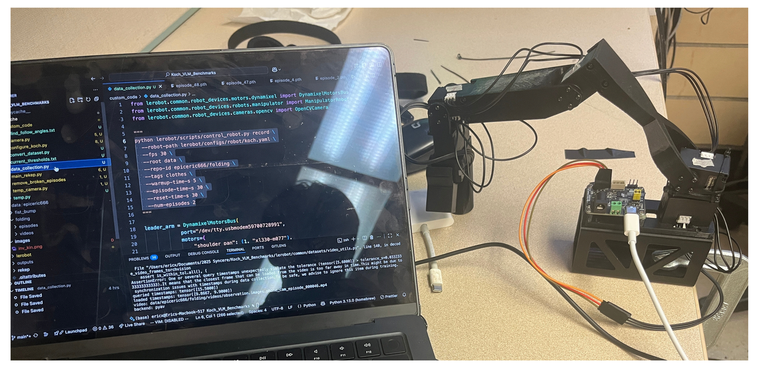
Training Procedure

- Proximal Policy Optimization (PPO) with RSL-RL
- 3000 episodes on 2 NVIDIA 3070 GPUs

Data Collection

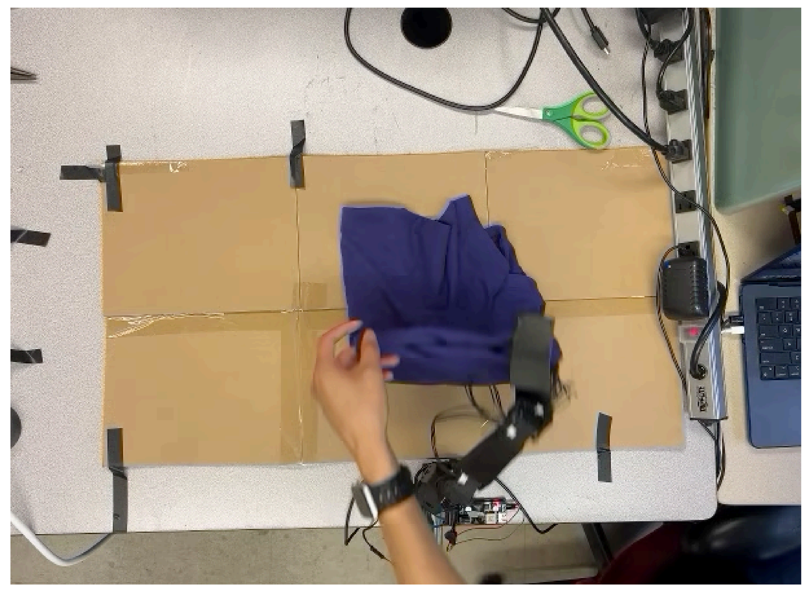
Setup

Follower arm performing task

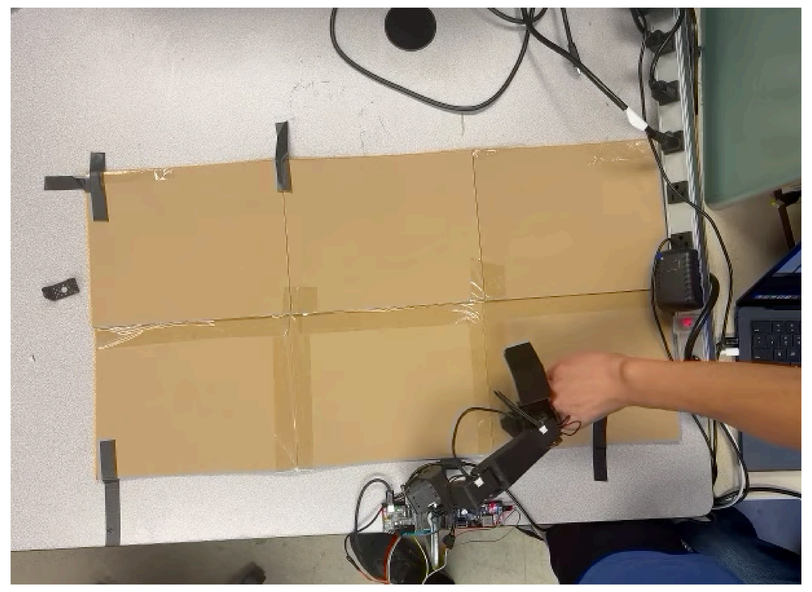


Leader arm & computer providing input

Two main tasks



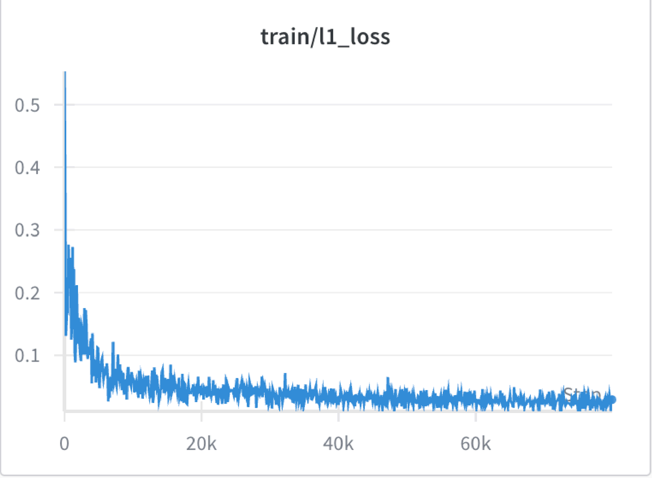
Shirt folding



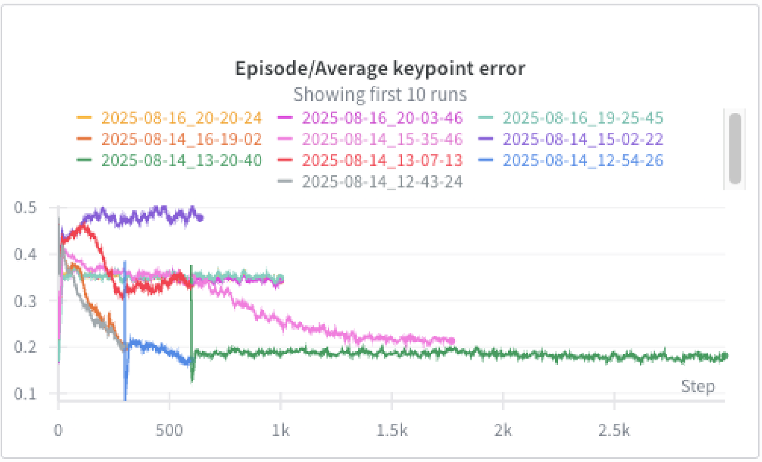
Fist bump

Results

1) CLARA



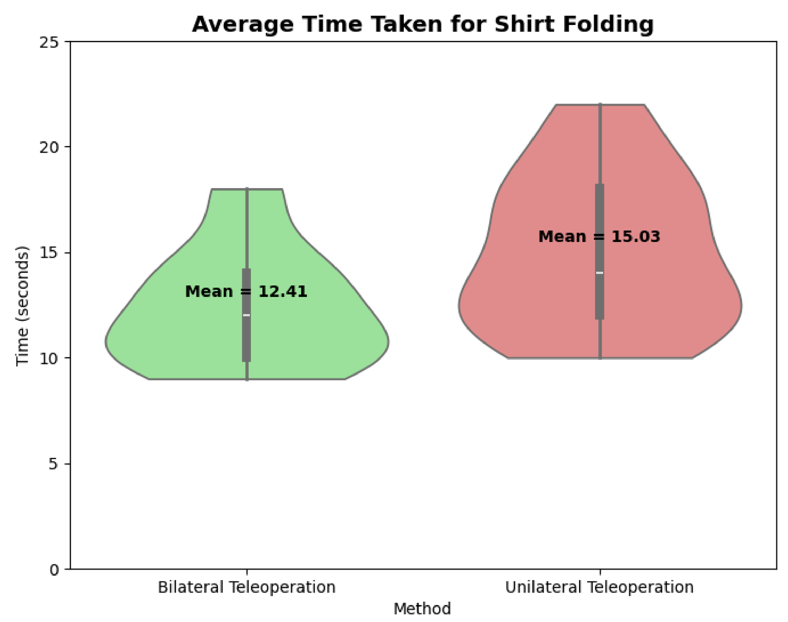
Diffusion loss plot



RL target error plot across runs

- Ideal diffusion loss curve
- RL decreases slowly and plateaus at around 10cm

2) Bilateral teleoperation



Comparison violin plot

- Data split across 50 episodes of training
- Decreased teleoperation time by 17%
- Saved time compounds over large training process
- User rates safer and more intuitive