Figure-8 Trajectory Tracking for Franka Emika using Reinforcement Learning

Natcha Jengjirapas

ETH Zurich, Switzerland

System Architecture

- **Robot Hardware:** Franka Emika Panda
- **Physics Engine:** MuJoCo
- **Framework:** Gymnasium
- **RL Algorithm:** PPO
- **CPU:** AMD Ryzen 9 7950x
- **OS:** Window 11 with WSL2 (Ubuntu 22.04/24.04)

RL: State & Observation Space

$$[q, \dot{q}, \Delta p, \Delta R, v_{des}, w_{des}, sin\phi, cos\phi, a, b]$$

- _get_obs function
- Total: 30-dim float32 vector

RL: State & Observation Space

$$[q, \dot{q}, \Delta p, \Delta R, v_{des}, w_{des}, sin\phi, cos\phi, a, b]$$

Body Schema

- _get_obs function
- Total: 30-dim float32 vector

RL: State & Observation Space

$$[q, \dot{q}, \underline{\Delta p, \Delta R}, v_{des}, w_{des}, sin\phi, cos\phi, a, b]$$

Tracking Errors

- _get_obs function
- Total: 30-dim float32 vector

RL: State & Observation Space

$$[q, \dot{q}, \Delta p, \Delta R, v_{des}, w_{des}, sin\phi, cos\phi, a, b]$$

Reference Dynamics

- _get_obs function
- Total: 30-dim float32 vector

RL: State & Observation Space

$$[q, \dot{q}, \Delta p, \Delta R, v_{des}, w_{des}, sin\phi, cos\phi, a, b]$$

Phase Embedding

- _get_obs function
- Total: 30-dim float32 vector

RL: State & Observation Space

$$[q, \dot{q}, \Delta p, \Delta R, v_{des}, w_{des}, sin\phi, cos\phi, a, b]$$

Task Parameters

- _get_obs function
- Total: 30-dim float32 vector

RL: Action Space

$$q_{cmd}^{(t+1)} = \text{clip}\left(q_{act}^{(t)} + (a^{(t)} \cdot \lambda),\, q_{min},\, q_{max}\right)$$

- Action: Normalized in [-1 , 1]
- Scale: 0.02 - 0.08 rad/step
- Frequency: 40 Hz control loop

RL: Action Space

$$q_{cmd}^{(t+1)} = \text{clip}\left(q_{act}^{(t)} + (a^{(t)} \cdot \lambda),\ q_{min},\ q_{max}\right)$$

Action Vector

- Action: Normalized in [-1 , 1]
- Scale: 0.02 - 0.08 rad/step
- Frequency: 40 Hz control loop

RL: Action Space

$$q_{cmd}^{(t+1)} = \text{clip}\left(q_{act}^{(t)} + (a^{(t)} \cdot \lambda),\ q_{min},\ q_{max}\right)$$

Scaling Factor

- Action: Normalized in [-1 , 1]
- Scale: 0.02 - 0.08 rad/step
- Frequency: 40 Hz control loop

RL: Action Space

$$q_{cmd}^{(t+1)} = \text{clip}\left(q_{act}^{(t)} + (a^{(t)} \cdot \lambda),\ q_{min},\ q_{max}\right)$$

Delta Integration

- Action: Normalized in [-1 , 1]
- Scale: 0.02 - 0.08 rad/step
- Frequency: 40 Hz control loop

RL: Action Space

$$q_{cmd}^{(t+1)} = \underbrace{\text{clip}\left(q_{act}^{(t)} + (a^{(t)} \cdot \lambda),\ q_{min},\ q_{max}\right)}_{\text{Safety Guard}}$$

- Action: Normalized in [-1 , 1]
- Scale: 0.02 - 0.08 rad/step
- Frequency: 40 Hz control loop

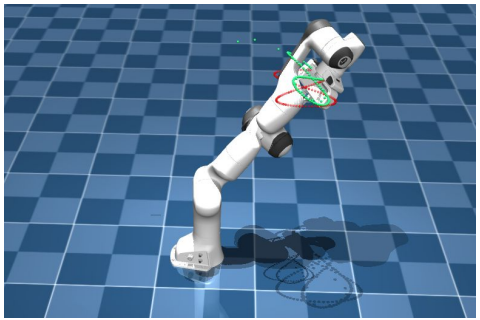
RL: Reward Function

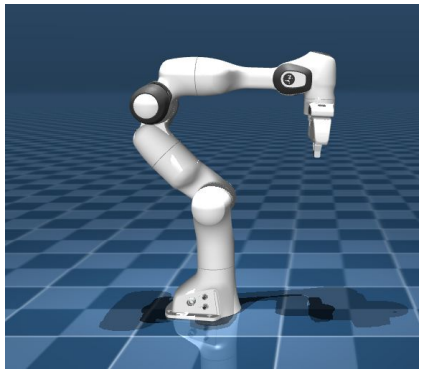
$$R = -(w_p \|\Delta p\|^2 + w_R \|\Delta R\|^2 + w_u \|a\|^2)$$

- Position: tracking accuracy
- Rotation: orientation alignment
- Action Penalty: encourage smooth torque profiles

RL: Environment Initialization & Constraints

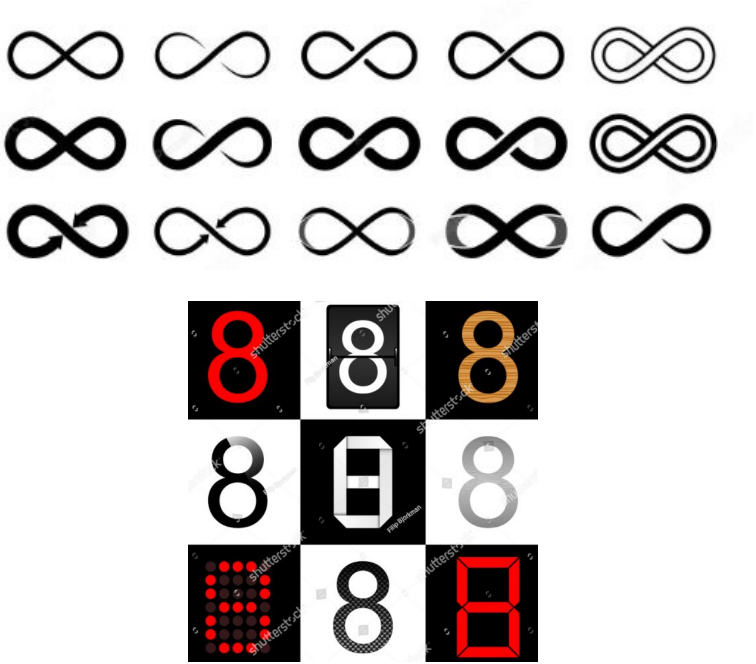
- Athletic Home Pose: $\mathbf{q}_{home} = [0, -0.785, 0, -2.356, 0, 1.571, 0.785]$
- Termination Criteria:
  - Terminate_pos_err_m = 1.2
  - Termination_grace_steps = 25





RL: Reset

- Randomization:
  - Width (a): [0.10, 0.17] m
  - Height (b): [0.08, 0.12] m
  - X-Axis: [0.25, 0.40] m
  - Y-Axis: [0.00, 0.10] m
  - Phase: [0.00, 2π] m



Training & Refinement

- PPO:
  - MlpPolicy
  - n step: 2048
  - batch: 256
  - ent_coef: 0.005
  - Total Time Step: 500,000
- Env:
  - Position Weight : 25 -> 100 -> 1000
  - Orientation Weight: 0.5
  - Action Weight: 0.02
  - action_scale = 0.02 -> 0.05 -> 0.08
  - F_hz = 0.75 -> 1 -> 1.3

Performance & Results

- Training Seed: 0
- Testing Seed: 24
- Success: < 5cm

Table 1: Frequency 1.0Hz Results Across 10 Randomized Episodes ($N = 10$)

Test Case	**Size** ($a \times b$)	**Pos. Error (cm)**	**Ori. Error (rad)**	**Success Rate**
In-Distribution	0.14×0.10	2.62	0.3418	100%
Small Scale	0.10×0.06	2.17	0.3435	100%
Kinematic Limit*	0.22×0.15	4.99	0.3472	**40%**

Table 2: Frequency 1.3Hz Results Across 10 Randomized Episodes ($N = 10$)

Test Case	**Size** ($a \times b$)	**Pos. Error (cm)**	**Ori. Error (rad)**	**Success Rate**
In-Distribution	0.14×0.10	2.86	0.2277	100%
Small Scale	0.10×0.06	2.11	0.1965	100%
Kinematic Limit*	0.22×0.15	6.8	0.2493	**0%**

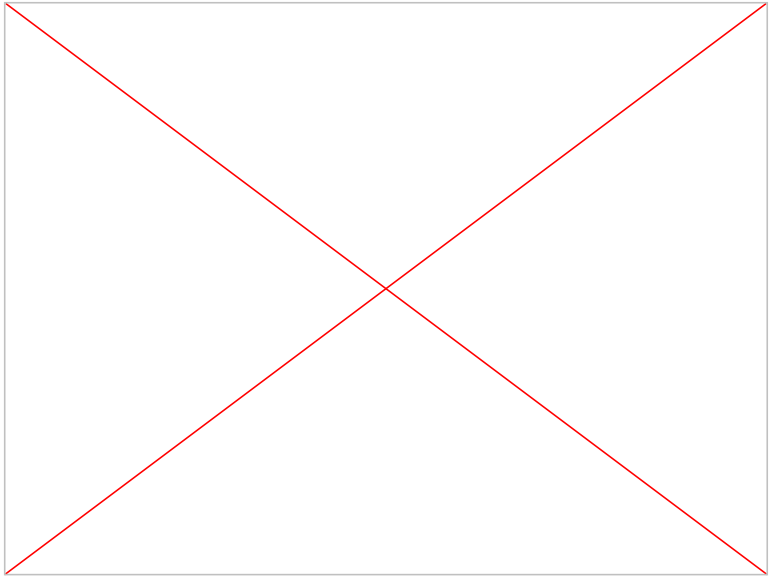
Performance & Results

Table 3: In-Distribution Validation Stability (Specialized Task)

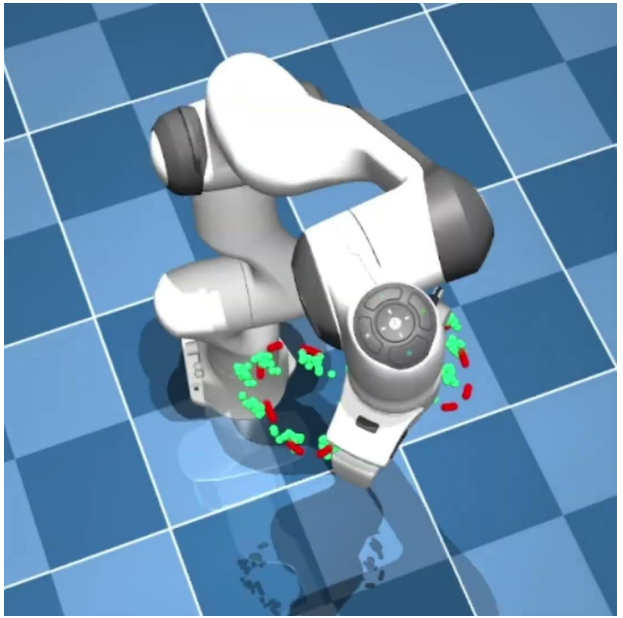
Metric	**Unit**	**10-Run Validation**	**50-Run Validation**
Avg. Position Error	cm	1.79	1.84
Avg. Orientation Error	rad (°)	0.1280 (7.33°)	0.1291 (7.40°)
Success Rate	%	**100%**	**100%**

Small Out of Distribution

1.0 Hz

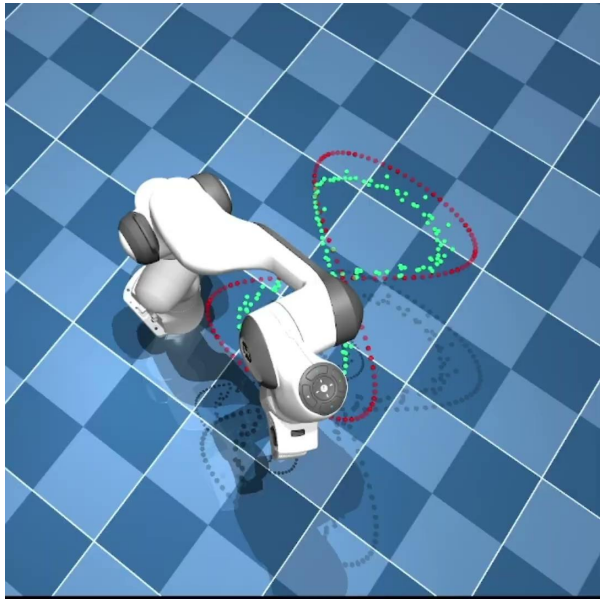


1.3 Hz

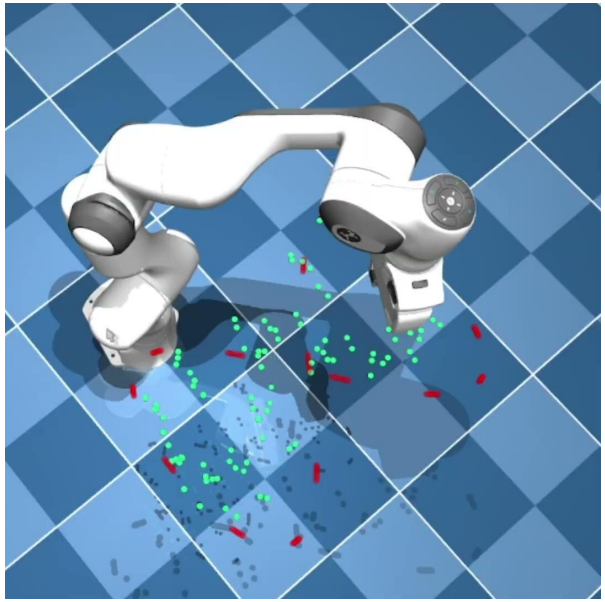


Large Out of Distribution

1.0 Hz

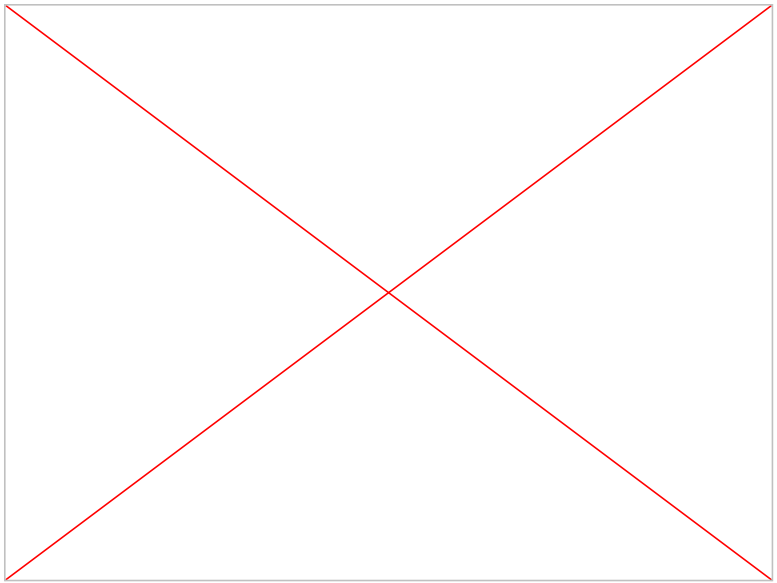


1.3 Hz

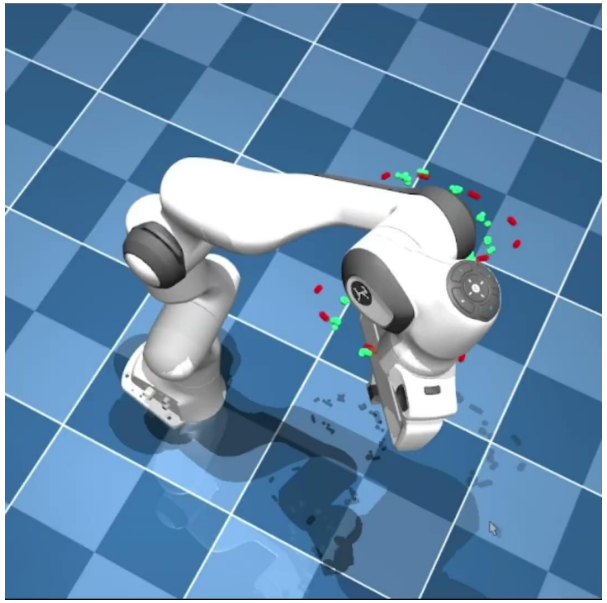


In Distribution

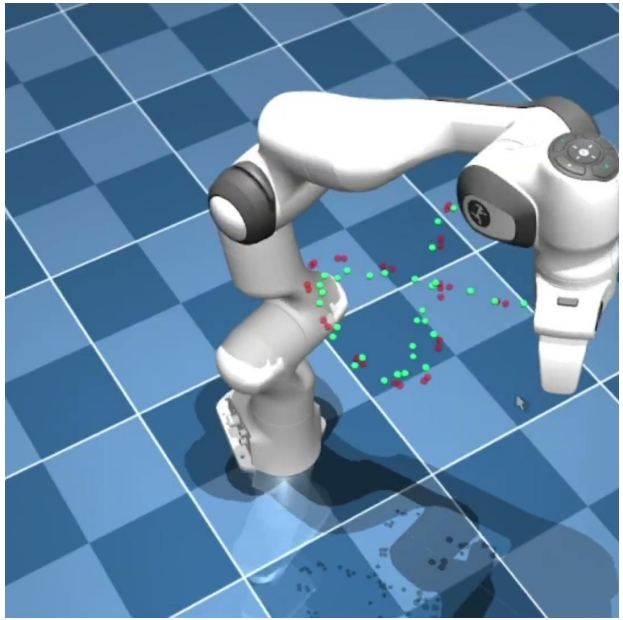
1.0 Hz



1.3 Hz



Specialized Version



Conclusion

- Achieved a consistent 2.62 cm avg tracking error
- 100% success rate on specialized tasks
- Generalized policy adapted to small OOD
- Identified the physical limit of the policy at a = .22 m, where Joint 4 and Joint 5 reaching its mechanical stop.

Joint Limitation

Joint 0: -2.90 to 2.90 radians

Joint 1: -1.76 to 1.76 radians

Joint 2: -2.90 to 2.90 radians

Joint 3: -3.07 to -0.07 radians

Joint 4: -2.90 to 2.90 radians

Joint 5: -0.02 to 3.75 radians

Joint 6: -2.90 to 2.90 radians