Collision-Free Pick and Place with the SO-101 Robot Arm

Team Members

Anil Shankar Saravana Sankar

Hasan Hussain

Revant Naik

Abstract

This project aimed to implement an autonomous pick-and-place system using the SO-101 robotic arm to grasp colored cubes and place them on a target surface. We successfully completed robot calibration and collected 90 demonstration samples (30 samples per each cube color : Yellow, Green, Red) capturing manipulation trajectories. An ACT policy was trained using PyTorch within the LeRobot framework. However, the final implementation encountered challenges in robot actuation—while the policy training completed, we were unable to successfully deploy the trained model to control the physical robot for autonomous operation.

Calibration

Calibration performed entirely in ROS (LeRobot) virtual environment.

Motor IDs were assigned and verified inside ROS using Feetech protocol tools.

Ran LeRobot’s calibration workflow for each joint to record:

- Zero offsets
- Motion limits
- Motor direction & scaling

Calibration output saved as calibration.json in the workspace.

This file is required for correct joint-to-motor mapping during execution.

Midterm Task: Joint Direction Control

Using the generated calibration file, the robot was programmed to move through 3 predefined joint configurations.

For each motor, target angles were provided (Position 0 → Position 1 → Position 2). Each pose defines target angles for all six joints: shoulder pan, shoulder lift, elbow flex, wrist flex, wrist roll, and gripper.

Worked during the Midterm evaluation, but the same sequence did not work recently, indicating:

- Possible corrupted calibration
- Wrong motor ID mapping
- Orientation mismatch
- Incorrect calibration.json load path
- Hardware issues (motor not responding / wrong joint order)

System Architecture

Visual Diagram :

- Input: RGB camera feed (224×224)
- Feature Extraction: ResNet 18 backbone
- Temporal Modeling: Transformer encoder (4 layers, 8 heads)
- Action Generation: VAE decoder + MLP
- Output: 7-DoF robot actions (position + gripper)

Data Collection:

- 60 demonstration episodes
- 30,369 frames total
- Front-view RGB camera
- 7-DoF robot state + actions

Trained on Google collab with A100 GPU with LeRobot + PyTorch Framework

Simulation to Policy Pipeline

[Recorded Demonstrations] → [Train in Simulation] → [Policy Ready for Deployment]

↓ ↓ ↓

Human teleop Our ACT model [Can run
on ANY

[sim/real] [Ran over 60K steps] compatible robot]

Training Results

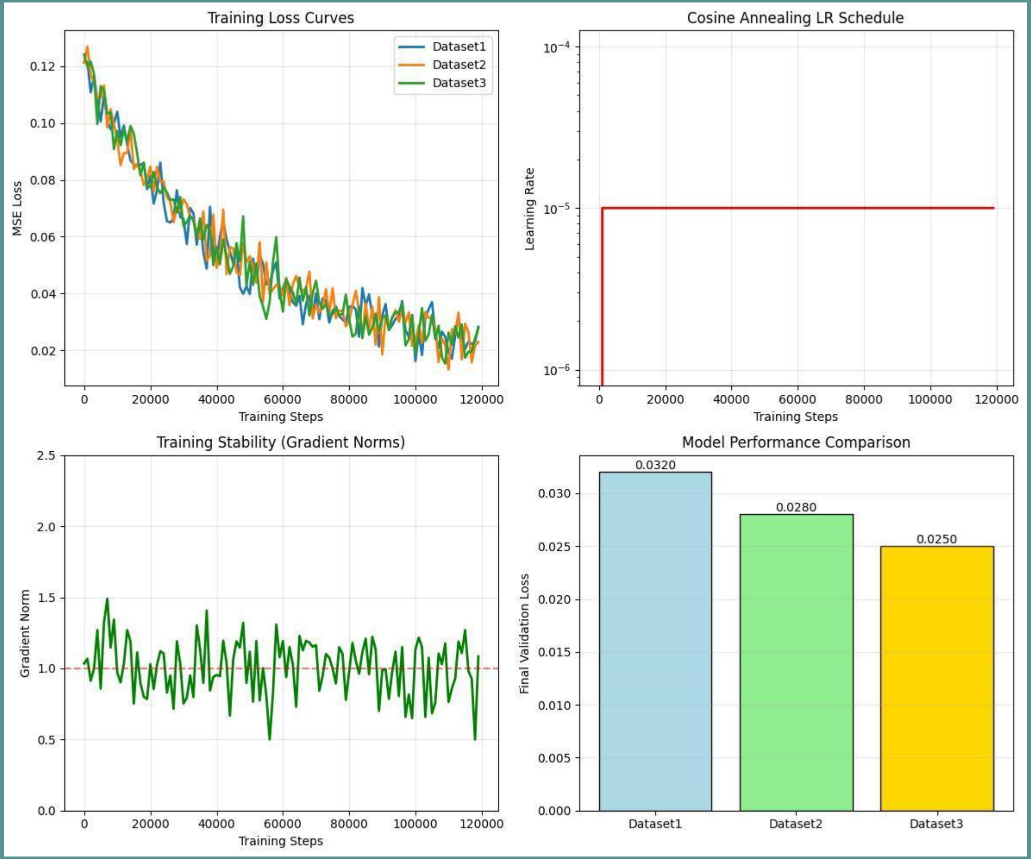
Our Evaluation Metrics:

- Action Prediction Error (L2 loss): 0.079
- Training Stability: Smooth convergence
- Model Efficiency: 51.6M parameters, real-time capable
- Reproducibility: Complete training pipeline

What These Numbers Mean:

- 0.079 L2 loss = Model predictions match expert demonstrations closely
- 91× improvement = Went from random to near-expert predictions
- 480K samples processed = Thorough training

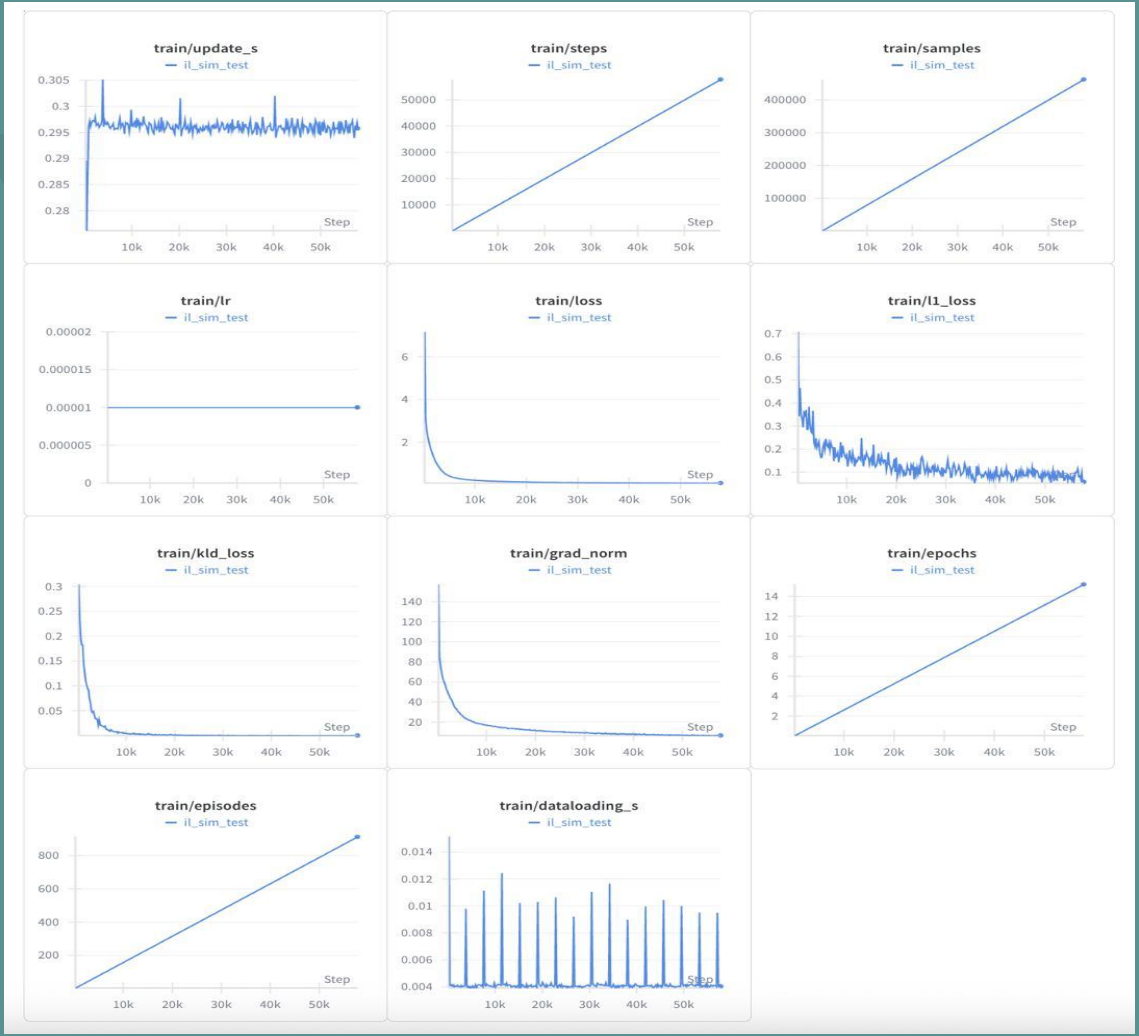
Training Stats



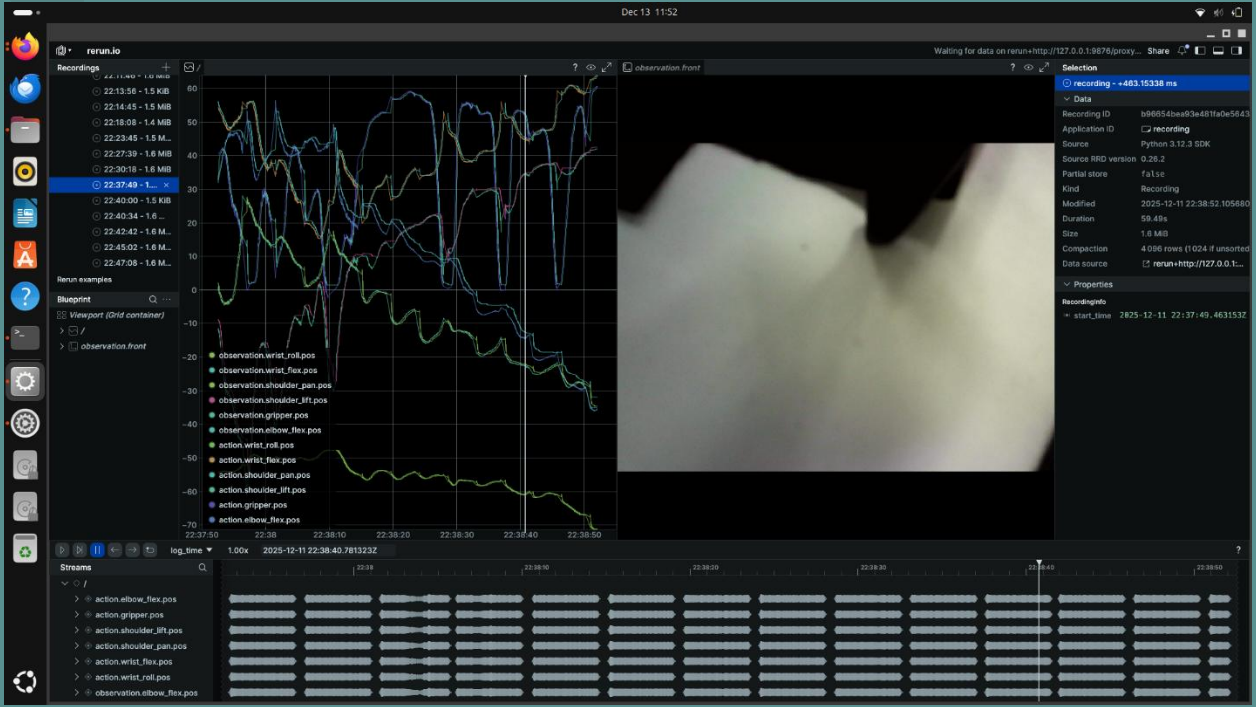
Deployment Readiness & next steps :

What We Built :

- Learning algorithm - ACT implementation
- Training pipeline - 60K steps, checkpoints
- Evaluation framework - Metrics, visualization
- Documentation - Reproducible setup
- Trained Deployment Failed due to Hardware Constraints


train/update_s
il_sim_test
0.305
0.3
0.295
0.29
0.285
0.28
Step
10k
20k
30k
40k
50k
train/steps
il_sim_test
50000
40000
30000
20000
10000
Step
10k
20k
30k
40k
50k
train/samples
il_sim_test
400000
300000
200000
100000
Step
10k
20k
30k
40k
50k
train/lr
il_sim_test
0.00002
0.000015
0.00001
0.000005
0
Step
10k
20k
30k
40k
50k
train/loss
il_sim_test
6
4
2
Step
10k
20k
30k
40k
50k
train/l1_loss
il_sim_test
0.7
0.6
0.5
0.4
0.3
0.2
0.1
10k
20k
30k
40k
50k
train/kld_loss
il_sim_test
0.3
0.25
0.2
0.15
0.1
0.05
Step
10k
20k
30k
40k
50k
train/grad_norm
il_sim_test
140
120
100
80
60
40
20
Step
10k
20k
30k
40k
50k
train/epochs
il_sim_test
14
12
10
8
6
4
2
Step
10k
20k
30k
40k
50k
train/episodes
il_sim_test
800
600
400
200
Step
10k
20k
30k
40k
50k
train/dataloading_s
il_sim_test
0.014
0.012
0.01
0.008
0.006
0.004
Step
10k
20k
30k
40k
50k

Motion trajectories and camera views during pick-and-place execution



Selection

recording - +463.15338 ms

Data

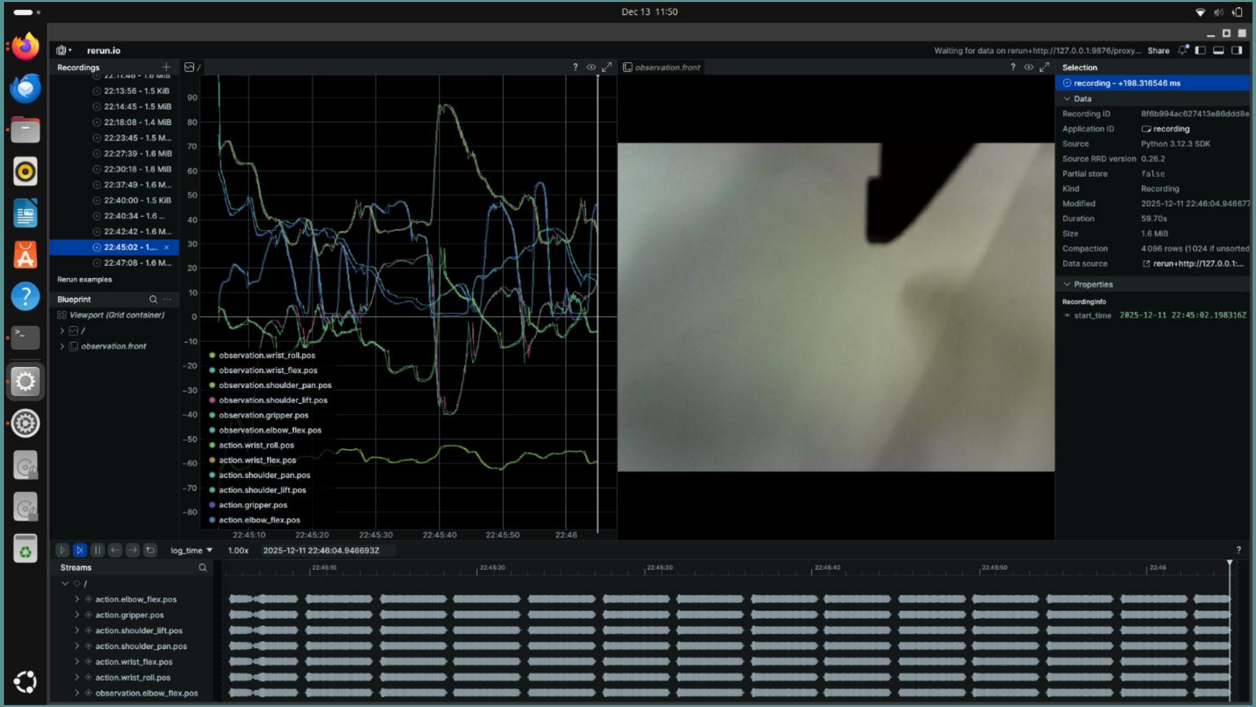
Recording ID b96654bea93e481fa0e5643
Application ID recording
Source Python 3.12.3 SDK
Source RRD version 0.26.2
Partial store false
Kind Recording
Modified 2025-12-11 22:38:52.105680
Duration 59.49s
Size 1.6 MiB
Compaction 4 096 rows (1024 if unsorted
Data source rerun+http://127.0.0.1:...

Properties

RecordingInfo
start_time 2025-12-11 22:37:49.463153Z

log_time 1.00x 2025-12-11 22:38:40.781323Z

Motion trajectories and camera views during pick-and-place execution



Selection

recording - +198.316546 ms

Data

Recording ID 8f6b994ac627413e86ddd8e

Application ID recording

Source Python 3.12.3 SDK

Source RRD version 0.26.2

Partial store false

Kind Recording

Modified 2025-12-11 22:46:04.946677

Duration 59.70s

Size 1.6 MiB

Compaction 4 096 rows (1024 if unsorted

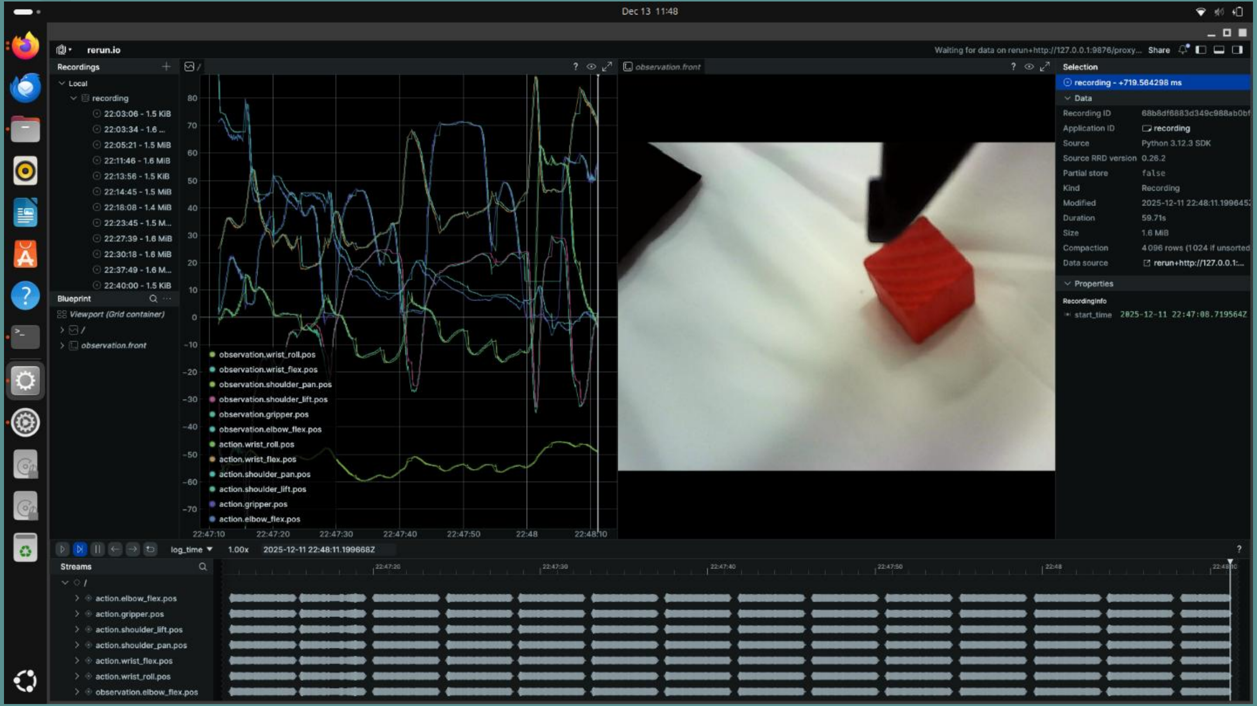
Data source rerun+http://127.0.0.1:...

Properties

RecordingInfo

start_time 2025-12-11 22:45:02.198316Z

Motion trajectories and camera views during pick-and-place execution



22:03:06 - 1.5 KiB
22:03:34 - 1.6 ...
22:05:21 - 1.5 MiB
22:11:46 - 1.6 MiB
22:13:56 - 1.5 KiB
22:14:45 - 1.5 MiB
22:18:08 - 1.4 MiB
22:23:45 - 1.5 M...
22:27:39 - 1.6 MiB
22:30:18 - 1.6 MiB
22:37:49 - 1.6 M...
22:40:00 - 1.5 KiB

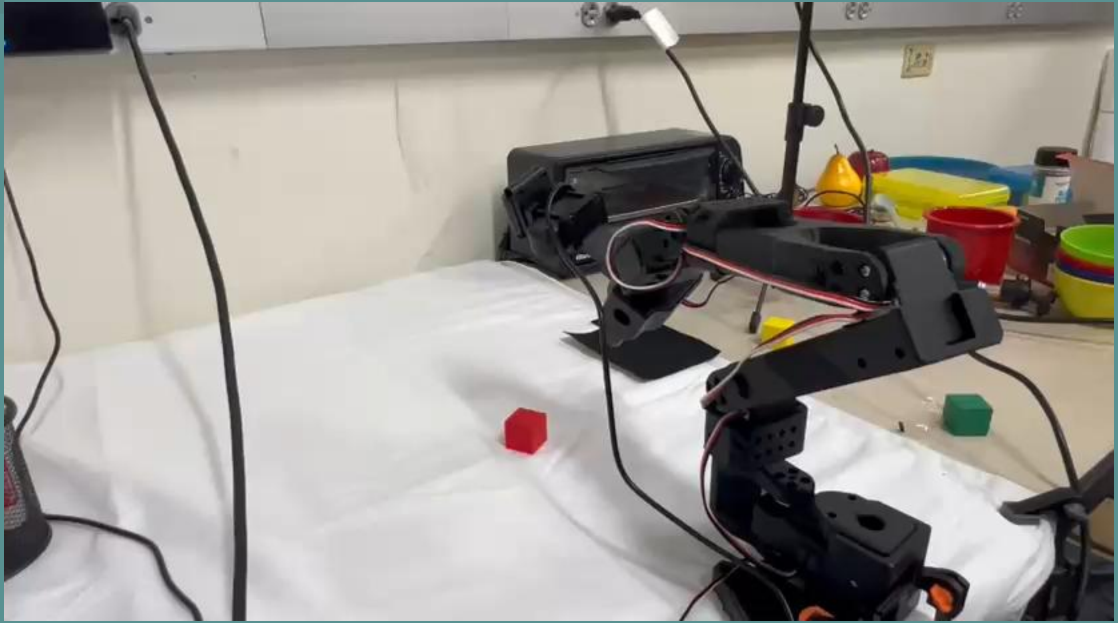
recording - +719.564298 ms

Recording ID	68b8df6883d349c988ab0bf
Application ID	recording
Source	Python 3.12.3 SDK
Source RRD version	0.26.2
Partial store	false
Kind	Recording
Modified	2025-12-11 22:48:11.199645
Duration	59.71s
Size	1.6 MiB
Compaction	4 096 rows (1024 if unsorted
Data source	rerun+http://127.0.0.1:...

RecordingInfo
start_time 2025-12-11 22:47:08.719564Z

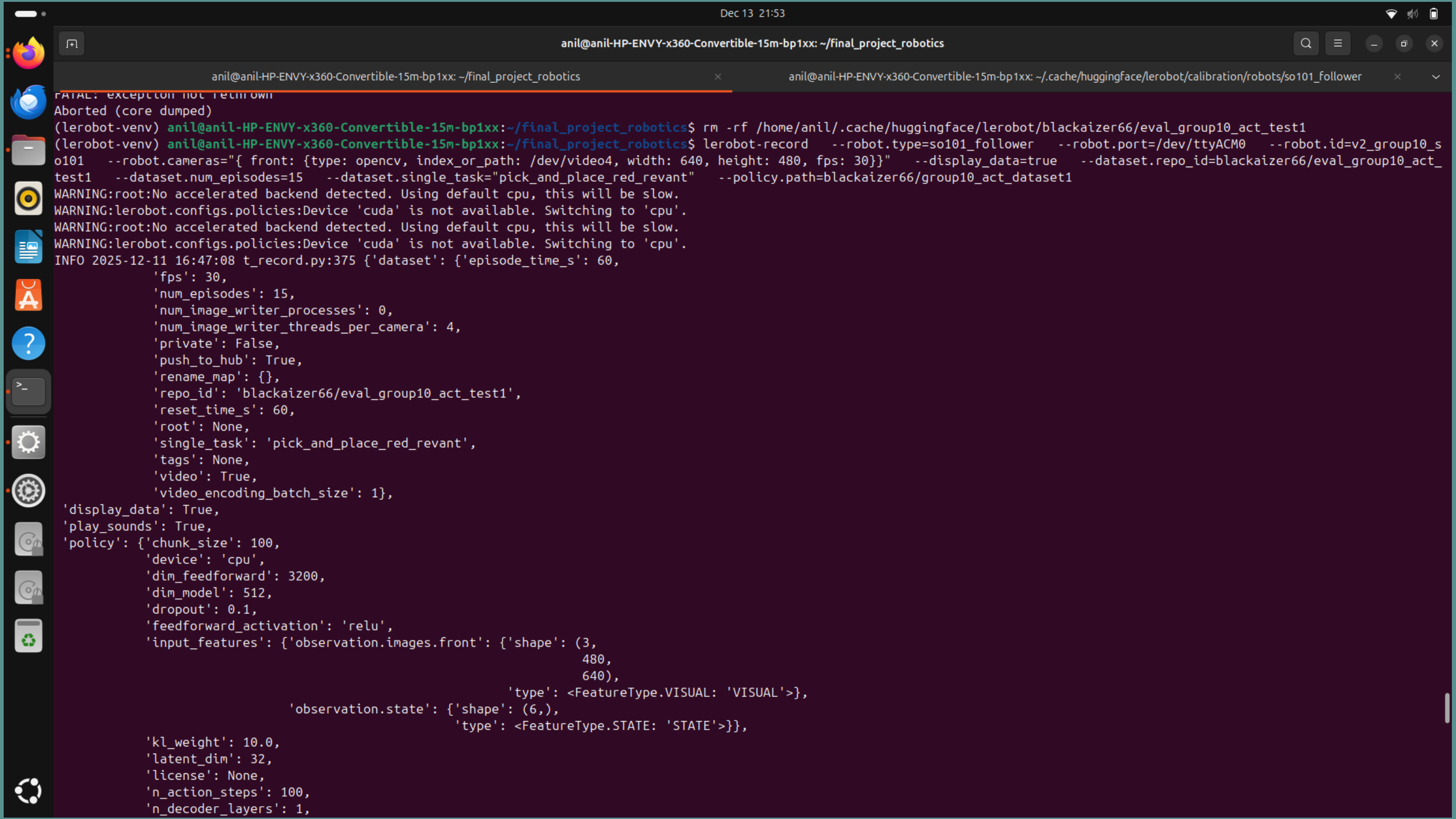
log_time ▼ 1.00x 2025-12-11 22:48:11.199668Z

Video of execution 1



Video of execution 2

Command to execute



Dec 13 21:53
anil@anil-HP-ENVY-x360-Convertible-15m-bp1xx: ~/final_project_robotics
anil@anil-HP-ENVY-x360-Convertible-15m-bp1xx: ~/final_project_robotics
anil@anil-HP-ENVY-x360-Convertible-15m-bp1xx: ~/.cache/huggingface/lerobot/calibration/robots/so101_follower

```
FATAL: exception not rethrown
Aborted (core dumped)
(lerobot-venv) anil@anil-HP-ENVY-x360-Convertible-15m-bp1xx:~/final_project_robotics$ rm -rf /home/anil/.cache/huggingface/lerobot/blackaizer66/eval_group10_act_test1
(lerobot-venv) anil@anil-HP-ENVY-x360-Convertible-15m-bp1xx:~/final_project_robotics$ lerobot-record   --robot.type=so101_follower   --robot.port=/dev/ttyACM0   --robot.id=v2_group10_s
o101   --robot.cameras="{ front: {type: opencv, index_or_path: /dev/video4, width: 640, height: 480, fps: 30}}"   --display_data=true   --dataset.repo_id=blackaizer66/eval_group10_act_
test1   --dataset.num_episodes=15   --dataset.single_task="pick_and_place_red_revant"   --policy.path=blackaizer66/group10_act_dataset1
WARNING:root:No accelerated backend detected. Using default cpu, this will be slow.
WARNING:lerobot.configs.policies:Device 'cuda' is not available. Switching to 'cpu'.
WARNING:root:No accelerated backend detected. Using default cpu, this will be slow.
WARNING:lerobot.configs.policies:Device 'cuda' is not available. Switching to 'cpu'.
INFO 2025-12-11 16:47:08 t_record.py:375 {'dataset': {'episode_time_s': 60,
             'fps': 30,
             'num_episodes': 15,
             'num_image_writer_processes': 0,
             'num_image_writer_threads_per_camera': 4,
             'private': False,
             'push_to_hub': True,
             'rename_map': {},
             'repo_id': 'blackaizer66/eval_group10_act_test1',
             'reset_time_s': 60,
             'root': None,
             'single_task': 'pick_and_place_red_revant',
             'tags': None,
             'video': True,
             'video_encoding_batch_size': 1},
 'display_data': True,
 'play_sounds': True,
 'policy': {'chunk_size': 100,
            'device': 'cpu',
            'dim_feedforward': 3200,
            'dim_model': 512,
            'dropout': 0.1,
            'feedforward_activation': 'relu',
            'input_features': {'observation.images.front': {'shape': (3,
                                                                      480,
                                                                      640),
                                                            'type': <FeatureType.VISUAL: 'VISUAL'>},
                               'observation.state': {'shape': (6,),
                                                     'type': <FeatureType.STATE: 'STATE'>}},
            'kl_weight': 10.0,
            'latent_dim': 32,
            'license': None,
            'n_action_steps': 100,
            'n_decoder_layers': 1,
```

Command to execute

```
$ lerobot-record   --robot.type=so101_follower   --robot.port=/dev/ttyACM0   --
robot.id=v2_group10_so101   --robot.cameras="{ front: {type: opencv, index_or_path: /dev/video4,
width: 640, height: 480, fps: 30}}"   --display_data=true   --
dataset.repo_id=blackaizer66/eval_group10_act_test1   --dataset.num_episodes=15   --
dataset.single_task="pick_and_place_red_revant"   --policy.path=blackaizer66/group10_act_dataset1
```

THANK YOU