

# LocoMan: Advancing Versatile Quadrupedal Dexterity with Lightweight Loco-Manipulators

Changyi Lin<sup>1</sup>, Xingyu Liu<sup>1</sup>, Yuxiang Yang<sup>2</sup>, Yaru Niu<sup>1</sup>,  
Wenhai Yu<sup>3</sup>, Tingnan Zhang<sup>3</sup>, Jie Tan<sup>3</sup>, Byron Boots<sup>2</sup>, Ding Zhao<sup>1</sup>  
<https://linchangyi1.github.io/LocoMan>

**Abstract**—Quadrupedal robots have emerged as versatile agents capable of locomoting and manipulating in complex environments. Traditional designs typically rely on the robot’s inherent body parts or incorporate top-mounted arms for manipulation tasks. However, these configurations may limit the robot’s operational dexterity, efficiency, and adaptability, particularly in cluttered or constrained spaces. In this work, we present LocoMan, a dexterous quadrupedal robot with a novel morphology to perform versatile manipulation in diverse constrained environments. By equipping a Unitree Go1 robot with two low-cost and lightweight modular 3-DoF loco-manipulators on its front calves, LocoMan leverages the combined mobility and functionality of the legs and grippers for complex manipulation tasks that require precise 6D positioning of the end effector in a wide workspace. To harness the loco-manipulation capabilities of LocoMan, we introduce a unified control framework that extends the whole-body controller (WBC) to integrate the dynamics of loco-manipulators. Through experiments, we validate that the proposed whole-body controller can accurately and stably follow desired 6D trajectories of the end effector and torso, which, when combined with the large workspace from our design, facilitates a diverse set of challenging dexterous loco-manipulation tasks in confined spaces, such as opening doors, plugging into sockets, picking objects in narrow and low-lying spaces, and bimanual manipulation.

## I. INTRODUCTION

Recent advances in the capability of quadrupedal robots have enabled them to traverse highly complex terrains [1, 2, 3, 4, 5, 6], perform acrobatic skills [7, 8, 9, 10, 11], and interact with humans [12]. While most of these works focus on improving the *mobility* of quadrupedal robots, the *manipulation* capability of quadrupedal robots still remains limited and often requires special designs to achieve tasks such as pressing buttons[13] and kicking balls[14, 15]. To be effectively deployed in daily life, quadrupedal robots are required to possess versatile manipulation skills with enhanced dexterity, in addition to their inherent locomotion capability.

Integrating manipulation into quadrupedal robots proves to be a difficult task. As illustrated in Table. I, without external hardware modifications, quadrupedal robots have to utilize their legs, head, or torso to move or transport objects [13, 15, 16, 17], which limits their ability to control the 6D poses of the object and perform high-precision tasks. A popular solution to address this challenge is to use a top-mounted robot arm [18]. However, this solution often comes



Figure 1: Equipped with loco-manipulators, LocoMan is proficient in handling versatile manipulation tasks. (a) With a single loco-manipulator, LocoMan not only can perform manipulation tasks that require precision and stability, but also excels at operating in extremely narrow space in its compact form. (b) With both loco-manipulators installed on the two front legs, LocoMan is able to perform bimanual manipulation tasks when standing upright. (c) LocoMan is also capable of loco-manipulation, e.g. carrying objects with its loco-manipulators while walking.

at the cost of increased payload and decreased agility, and prevents the robot from reaching narrow spaces or performing more dexterous tasks such as bimanual manipulation. While foot-mounted grippers offer a lightweight alternative to full robot arms, existing work adopts 1-DoF grippers with limited functionality [19]. Therefore, it can be challenging to achieve versatile dexterous loco-manipulation using existing manipulator solutions for legged robots.

In this work, we present LocoMan, a comprehensive solution for versatile and skilled manipulation with legged robots. To reduce the payload and increase dexterity, LocoMan mounts two custom-designed lightweight 3-DoF manipulators on the front *calves* of the robot, and uses existing leg joints in addition to the manipulators for manipulation, as illustrated in Fig. 1. We refer to our designed manipulator as “loco-manipulator”, which draws conceptual inspiration from the front limbs of animals such as great apes and bears, which are adept at both locomotion and skilled manipulation.

For precise and robust control of LocoMan to perform loco-manipulation tasks, we design a unified whole-body control framework, which tracks the command kinematically and dynamically through joint impedance commands. The integration of the loco-manipulator’s lightweight design with

<sup>1</sup> Carnegie Mellon University

<sup>2</sup> University of Washington

<sup>3</sup> Google Deepmind

Table I: Comparison of the capabilities between robots with different morphology.

Methods	[13, 15, 20]	[18, 21]	[19]	Ours
Additional Manipulator	None	6DoF Top-mounted	1DoF Leg-mounted	3DoF Leg-mounted
6D Operational Space	✗	✓	✗	✓
Narrow Space Manipulation	✗	✗	✓	✓
Bi-Manual Manipulation	✗	✗	✓	✓
Loco-Manipulation	✓	✓	✗	✓

precise whole-body control enables LocoMan to operate in diverse modes, from reaching an object from narrow-space with single-arm, bimanual manipulation while sitting upright, to carrying an object while walking. We further design a Finite State Machine (FSM) to manage the transitions of these operation modes.

We evaluate the design of loco-manipulator on a Unitree Go1 robot and find that the added loco-manipulators can significantly increase the robot’s workspace (by 99.01% for single-arm manipulation and 118.28% for bimanual manipulation) with negligible weight increase (<2.5%). In addition, the whole-body controller enables precise trajectory tracking with a mean pose error of 1.89 mm and 0.047 rad. We further validate the capabilities of LocoMan in a number of real-life manipulation tasks under user teleoperation. In addition to standard tasks such as cabinet opening and drink pouring, LocoMan can reach narrow spaces with limited vertical clearance (0.25 m for locomotion and 0.09 m for manipulation). Thanks to the lightweight design of loco-manipulator, Loco-Man can stand on its rear legs in an upright pose and perform bimanual manipulation tasks such as grasping a pair of socks simultaneously and hoisting a basket with a rope.

In summary, the contribution of this paper is as follows:

- 1) We present LocoMan, leg-mounted with two lightweight loco-manipulators to improve manipulation capability of quadrupedal robots.
- 2) We design a unified whole-body controller for precise, simultaneous 6D pose tracking for both the end effectors and the torso.
- 3) We evaluate that LocoMan is capable of precise trajectory tracking in a large workspace.
- 4) We demonstrate that LocoMan can perform versatile real-life manipulation tasks, including cabinet opening, charger inserting, drink pouring, grasping in a narrow space, and basket hoisting.

## II. RELATED WORK

### A. Quadrupedal Loco-Manipulation

**Without Additional Manipulators** Researchers have a long history exploring the manipulation capability of quadrupedal robots. Without an additional manipulator, researchers have designed the robot to use its head [17, 22, 23, 24], torso [16], or foot [13, 14, 15, 25, 26, 27, 28] to interact with objects in the environment, where [22, 23, 24] rely on multiple robots to push larger objects. However, without an

additional mechanism such as a gripper, these approaches are limited to relatively simple tasks that do not require fetching objects, such as kicking balls [14], pressing buttons [13], pushing boxes [24], opening doors [27], and probing [26].

**With Additional Manipulators** Many previous works mount a dedicated arm on top of the quadrupedal robot to perform more complicated manipulation tasks such as pick-and-place [29, 30], pulling doors [31, 21, 32], tablecloth spreading [33], turning wheels [34], wiping a whiteboard [18], and collaboratively carrying large objects [35, 36]. On the other hand, equipping a robot with a dedicated arm typically adds to the mechanical complexity and power requirements, leading to an increase in both weight and expense. Moreover, such an addition makes it difficult to operate in a narrow space and reduces the quadrupedal robot’s agility. Some other works alleviate these problems by installing more lightweight manipulators or grippers on the mouth [37] or foot [19, 20] of the quadrupedal robot. Tsvetkov et al. [19] propose a novel design of a small-scale quadrupedal robot with manipulators built into the legs, each requiring three additional actuators, where two are used for grasping. It facilitates single-arm and two-arm manipulation, but it only supports specifying target positions in joint space. Arm et al. [20] propose a learning-based controller that can track quadrupedal foot position targets through whole-body motions, and a 2-DoF gripper is attached to the foot to perform tasks such as collecting rock samples. However, how to efficiently reach a specified 6D pose of the end effector on the leg to perform more complex manipulation tasks is still underexplored. While some hexapod robots [38, 39, 40, 41, 42] use legs as manipulators, where integrated grippers can fetch objects with high degrees of freedom, prehensile manipulation with dexterity could be challenging for quadrupedal robots.

### B. Controllers for Leg-Manipulation

To accomplish a wide range of manipulation tasks, the controller of the loco-manipulation robot need to support a variety of operation modes, including reaching a narrow space with a single arm, walking while carrying an object, and bi-manual manipulation with upright pose. While prior works have achieved similar tasks such as grasping [21, 31, 18], object pushing and kicking [17, 14, 15], and object carrying [41] using optimization-based [21, 31, 29, 17, 41] or learning-based [16, 14, 15, 18, 13] methods, most of these controllers are designed for specific tasks, and additional effort is required for multi-task support. In contrast, LocoMan adopts a unified control framework, which uses the same low-level whole-body controller [43] for all five operation modes. By tracking desired trajectories at both the kinematics and dynamics level, our unified framework requires little task-specific tuning and performs all manipulation tasks with high accuracy.

## III. DESIGN OF LOCO-MANIPULATOR

To accomplish versatile manipulation tasks, the robot must be able to reach desired 6D poses in its workspace. While

Table II: Specifications of a loco-manipulator.

DoFs	Dimension [mm <sup>3</sup> ]	Weight [g]	Cost[\$]	4 Servos	2 Bearings [d × D × h]	3D Printer
3 (joints) + 1 (gripper)	179 × 61 × 42	147	370	XC330-T288-T	5 mm × 8 mm × 2.5 mm	Formlabs® Form 3+

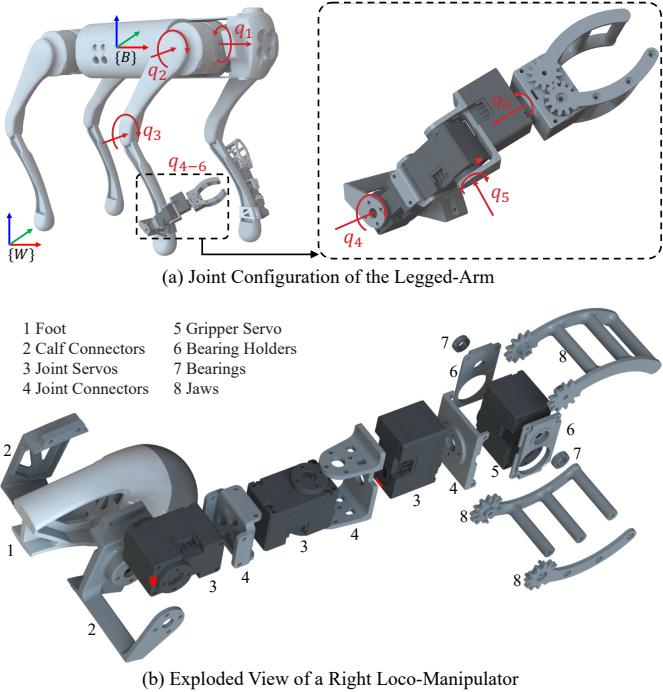


Figure 2: Design of our loco-manipulator. (a) The joint configuration of the legged-arm. The legged-arm has six joints including three from the leg and three from loco-manipulator. (b) The components of a right loco-manipulator shown in an exploded view.

quadrupedal robots can utilize torso movements to aid manipulation [20, 19], enabling 6D pose reaching directly from the manipulator can significantly improve the precision and range of manipulation tasks. Combining the loco-manipulator DoFs with the existing joint DoFs, each front limb of LocoMan has 6 DoFs in total, and can reach arbitrary 6D space poses *without* body movement.

The design of LocoMan is inspired by the anatomy of human arms. As illustrated in Fig. 2(a), we attach custom-designed loco-manipulators to the calf of the front legs of a quadrupedal robot, and utilize existing leg joints ( $q_{1-3}$ ) together with the gripper joints ( $q_{4-6}$ ) for 6DoF space manipulation. The original leg joints,  $q_{1-3}$ , similar to human shoulders and elbows, are used mainly for movement and positional tracking. The added joints of loco-manipulator, ( $q_{4-6}$ ), similar to the human wrist, are primarily used for orientation tracking. Integrated together, each front limb of LocoMan can reach a wide variety of space poses effectively.

We carefully orient the joints of loco-manipulator to seamlessly integrate with the existing structure of the original robot. Specifically, we design the first joint ( $q_4$ ) to be aligned with the calf joint of the front leg, design the last joint ( $q_6$ ) to point towards the gripper, and design the second joint ( $q_5$ ) to be perpendicular to the other two joints. In this way, the

gripper can reach out to far spaces during manipulation (front-right foot of Fig. 2(a)), while staying tucked to the calf during standing and walking (front-left foot of Fig. 2(a)) with minimal interference with locomotion capabilities.

The components of loco-manipulator are lightweight, low cost, and easy to fabricate. As illustrated in Fig. 2(b), for the actuators of a loco-manipulator, we select four Dynamixel servos, which are compact, highly dynamic, and capable of providing position, velocity, and torque feedback. loco-manipulator is mounted on the robot’s calf via two calf connectors, which are meticulously designed so that their internal structures seamlessly interlock with the skeletal framework of the robot’s calf. To make loco-manipulator more compact, we designed the gripper as two pairs of rotating jaws with gear engagement, enabling symmetric opening and closing of the gripper with a single servo. One pair of jaws is connected to the servo horns, while the other pair is mounted with two bearings. Except for the servos (label 3 and label 5 in Fig. 2(b)) and the bearings (label 7), the other components of loco-manipulator, including the calf connectors, joint connectors, bearing holders, and jaws, can be 3D printed using a standard 3D printer. Integrated together, each loco-manipulator weighs 147g (1.23% of the robot’s weight), and can be easily fabricated with a material cost of \$370 (Table II).

#### IV. A UNIFIED FRAMEWORK FOR WHOLE-BODY LOCO-MANIPULATION

##### A. Notation

We denote the state of LocoMan as  $\mathbf{q} = (\mathbf{x}^{\text{torso}}, \mathbf{q}_j) \in \mathbb{R}^{24}$ , where  $\mathbf{x}^{\text{torso}} \in \mathbb{R}^6$  denotes the position and orientation of the floating torso base, and  $\mathbf{q}_j \in \mathbb{R}^{18}$  denotes the joint angles of LocoMan, including the legs’ (12 DoF) and the loco-manipulators’ (6 DoF). We denote  $\mathbf{x}^{\text{eef}} = (\mathbf{x}_{\text{left}}^{\text{eef}}, \mathbf{x}_{\text{right}}^{\text{eef}}) \in \mathbb{R}^{12}$  and  $\mathbf{x}^{\text{foot}} \in \mathbb{R}^{24}$  as the cartesian positions and orientations of two grippers and four feet, which can be computed from the state vector using forward kinematics. We use the bar notation ( $\bar{\cdot}$ ) to denote all desired states.

##### B. Overview

Effective loco-manipulation requires the robot to complete a large variety of tasks accurately and robustly, from in-place picking to locomotion while carrying objects. In light of this requirement, we design a unified control framework for LocoMan (Fig. 3), where a *robot commander* specifies the desired state of LocoMan based on the action command from user, and a *unified whole-body controller* (WBC) tracks the desired state based on LocoMan’s kinematics and dynamics model. The robot commander supports a variety of operation modes (Table III), such as locomotion and in-place manipulation, where the mode switch is governed by a finite state

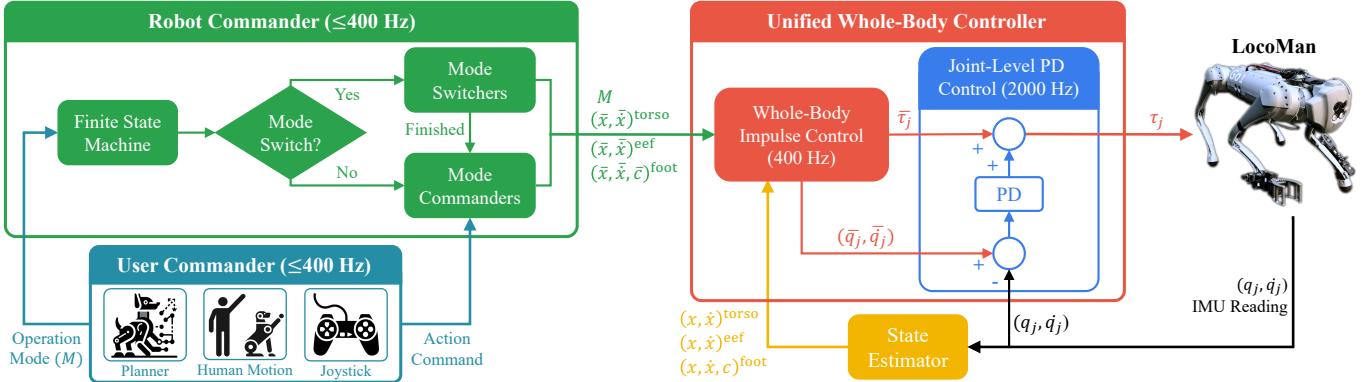


Figure 3: The unified framework for whole-body loco-manipulation. The robot commander converts the desired action command from user into the standardized robot command based on a specific operation mode ( $M$ ). The unified whole-body controller, including whole-body impulse control and joint-level PD control, computes torques for each joint of LocoMan to track the desired robot command,

Table III: Hierarchical tracking objectives and state estimators for the five operation modes. The subscripts  $\{m\}$ ,  $\{s\}$ ,  $\{l, r\}$  indicate the end effector in manipulation ( $m$ ), the feet in swing ( $s$ ), and the left ( $l$ ) and right ( $r$ ) grippers respectively.

Operation Mode	Hierarchical Tracking Objectives in WBC				State Estimator
	Objective 1	Objective 2	Objective 3	Objective 4	
Single Foot Manipulation	Torso Position	Torso Orientation	Foot <sub>m</sub> Position	-	Kinematics-Based
Single Gripper Manipulation	Torso Position	Torso Orientation	Gripper <sub>m</sub> Position	Gripper <sub>m</sub> Orientation	Kinematics-Based
Bimanual Manipulation	Gripper <sub>l,r</sub> Positions	Gripper <sub>l,r</sub> Orientations	-	-	Kinematics-Based
Locomotion	Torso Velocity	Torso Orientation	Foot <sub>s</sub> Positions	-	Kalman Filter-Based
Loco-Manipulation	Torso Velocity	Torso Orientation	Foot <sub>s</sub> Positions	Gripper <sub>l,r</sub> Orientations	Kalman Filter-Based

machine (FSM). Within each mode, the FSM launches a mode-specific *mode commander* to convert user specified target (e.g. walking speed or gripper pose) into desired states of the torso  $(\bar{x}, \bar{\dot{x}})^{\text{torso}}$ , end effector  $(\bar{x}, \bar{\dot{x}})^{\text{eef}}$ , and foot  $(\bar{x}, \bar{\dot{x}}, \bar{c})^{\text{foot}}$ . We further design a set of *mode switchers* to smoothly transition between different operation modes. For robust operation and accurate tracking of the desired state, we extend a *whole-body controller* to solve the desired position  $\bar{q}_j$ , velocity  $\bar{\dot{q}}_j$  and torque  $\bar{\tau}_j$  of each joint based on the complete kinematics and dynamics model of the robot. Then a joint-level PD controller computes the final torque  $\tau_j$  for each joint at a higher frequency. In addition, we implement two state estimators to accurately estimate the robot state based on sensor feedback.

### C. Robot Commander

The *robot commander* supports five different operation modes (Table III), namely locomotion, locomanipulation, single-foot manipulation, single-gripper manipulation, and bimanual manipulation. Each *mode commander* converts different user commands from planners, human motions, or joysticks to desired states of LocoMan.

For single-arm manipulation, the user specifies the desired torso 6D pose, and the desired gripper 6D pose or foot 3D position. For bimanual manipulation, the user commands the desired 6D poses for each gripper. The stance feet remain fixed to the ground during the three manipulation modes. For locomotion and loco-manipulation, the user specifies the desired velocity, roll, pitch, and height of the robot torso, together with the desired gripper orientation for the later mode.

To ensure a smooth transition between operation modes, the FSM launches a specific *mode switcher* during mode

transitions. For example, from locomotion mode to single-arm manipulation mode, LocoMan is commanded to zero velocity until all feet are on the ground, and then moves the torso to the rear opposite the end effector being manipulated.

### D. Unified Whole-Body Controller

Given the desired states of the torso  $(\bar{x}, \bar{\dot{x}})^{\text{torso}}$ , end effector  $(\bar{x}, \bar{\dot{x}})^{\text{eef}}$ , and foot  $(\bar{x}, \bar{\dot{x}})^{\text{foot}}$  and foot contact force  $\bar{c}^{\text{foot}}$ , the WBC computes the desired position  $\bar{q}_j$ , velocity  $\bar{\dot{q}}_j$  and torque  $\bar{\tau}_j$  of each joint.

Our WBC implementation is based on the two-step implementation from Kim et al. [43] with additional modeling of the loco-manipulators. In the first step, WBC tracks the desired states *kinematically* by computing the desired position  $\bar{q}_j$  and velocity  $\bar{\dot{q}}_j$  of each joint using Inverse Kinematics (IK). Since the robot may not have sufficient degrees of freedom to track the entire desired state, we assign a *objective priority* to each component of the desired states (Table III), and solve objectives in descending priority by projecting low-priority objectives into the null-space of high-priority ones. In the second step, WBC tracks the desired state *dynamically* by optimizing joint torque commands  $\bar{\tau}_j$ , where the objective is to track the desired torso acceleration, and the constraints include rigid body dynamics and friction cone limits. Please refer to the original work [43] for additional details.

## V. EXPERIMENTS

We design a set of experiments to validate the capability of LocoMan in versatile manipulation tasks. Specifically, we aim to answer the following questions: (1) How large is the workspace of LocoMan and how does it compare with the

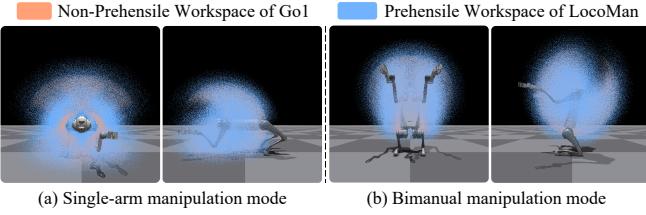


Figure 4: Our loco-manipulator turns the original non-prehensile (in orange) workspace of Go1 to a prehensile workspace (in blue) and expand the reachable area by more than 80%.

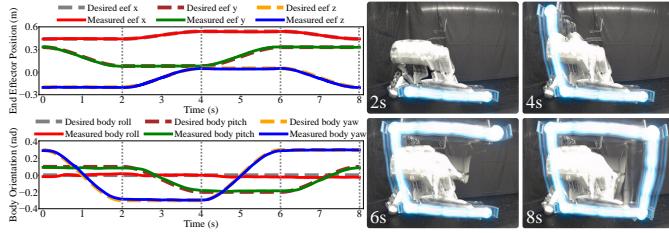


Figure 5: The figure on the left shows trajectory tracking of the end effector position and the torso orientation when drawing a spatial square. The figure on the right depicts the corresponding 3D trajectory tracked by an LED light attached to the gripper.

workspace of the original robot? (2) How well can LocoMan track a desired trajectory in its workspace? (3) What real-life manipulation tasks can LocoMan perform? and (4) What is the benefit of the leg-mounted manipulator design of LocoMan, compared to the top-mounted arms [18, 21]?

#### A. Experiment Setup

As illustrated in Fig. 1, we install two loco-manipulators on the front legs of Unitree Go1, a quadrupedal robot. The servos of the manipulators are powered by the robot via an additional voltage converter (from 24V to 12V). The algorithms of the motion command and control framework run on a desktop, and the control signals for the quadrupedal robot’s motors and the loco-manipulators’ servos are sent via cables.

#### B. Workspace Analysis

We compare the workspace of LocoMan with the workspace of the unmodified Unitree Go1 robot as illustrated in Fig. 4. The workspace is approximated by uniformly sampling multiple angles for each joint within its joint limit, recording the joint configurations without self-collision, and plotting the end effector positions (center of the gripper for LocoMan and foot for the original robot). With additional loco-manipulators, LocoMan extends the workspace of the end effector much further from the original foot toes to the prehensile gripper. Quantitatively, the inclusion of the loco-manipulator expands the workspace volume by at least 80%: from  $0.34m^3$  to  $0.68m^3$  in the single-arm manipulation mode, from  $0.38m^3$  to  $0.83m^3$  in the bimodal mode, and from  $0.63m^3$  to  $1.14m^3$  in the combined volume.

#### C. Trajectory Tracking

We design a set of experiments to evaluate the trajectory tracking performance of LocoMan. First, we manually design

Table IV: Translational and rotational MAEs of end effector and torso in trajectory tracking. Unit is mm for lengths and rad for angles.

Trajectories	$e_{\text{eff}}^t$	$e_{\text{eff}}^R$	$e_{\text{torso}}^t$	$e_{\text{torso}}^R$
Gripper-M w/o Torso Motion	1.67	0.045	1.92	0.0075
Gripper-M w Torso Motion	1.89	0.047	2.38	0.016
Foot-M w/o Torso Motion	0.93	-	2.02	0.0078
Foot-M w Torso Motion	1.10	-	2.26	0.017
Standing w Torso Motion	-	-	1.05	0.014

trajectories of the gripper and torso, where the gripper follows a desired rectangular trace with a spatial bounding box of  $0.25m \times 0.1m \times 0.25m$ , and the body moves towards the direction of the gripper. We evaluate the quality of the tracking using the Mean Absolute Error (MAE) of the commanded desired poses and the estimated poses from the kinematics-based state estimator for the manipulator and the torso. As illustrated in Fig. 5, the whole-body controller can enable the robot to accurately track the desired trajectory with tracking MAE of  $1.06mm$  in the end effector position and  $0.023rad$  in the torso orientation.

To further evaluate the tracking performance of the whole-body controller, we define five tasks using the end effector with and without torso movement while standing. For each task, we define four different intermediate poses for the end effector and the torso. We then generate an 8-second trajectory with total 3200 target 6D poses using cubic Bezier interpolation from each start pose to its end pose. As illustrated in Table IV, the translational and rotational MAEs of the gripper are only  $1.89mm$  and  $0.047rad$  respectively even with torso movement, which are much more accurate than learning-based method for foot manipulation [20] with translational error of  $57mm$ . On the other hand, we notice that the translational errors of the torso during manipulation are much larger than those during standing, which may be due to the uneven deformation of the three stance feet. However, the translational errors of the manipulator are smaller than the torso translational errors, benefiting from the task hierarchy of the whole-body controller.

#### D. Low-cost Vision-based Teleoperation Platform

In LocoMan, the robot torso and end effectors need to be commanded simultaneously, which would be difficult to teleoperate using a single joystick, keyboard or sensor glove. Therefore, we develop a low-cost vision-based teleoperation platform to teleoperate LocoMan using motions of human hands and torso. As illustrated in Fig. 6 (a), the core of our teleoperation platform is a pair of 3D-printed rigid shells worn by palms and thumbs of both hands, as well as the human torso. The shells are densely covered with AprilTags [44] whose corners and IDs can be robustly detected. Using a method similar to [45, 46], we set up calibrated high-resolution ZED cameras at multiple views and the 6D pose of the rigid shells can be obtained by minimizing the re-projection errors of AprilTag corners in the multiple-view images. The 6D pose of the torso and palm shells are directly used to command the 6D pose of the torso and end effector of quadrupedal

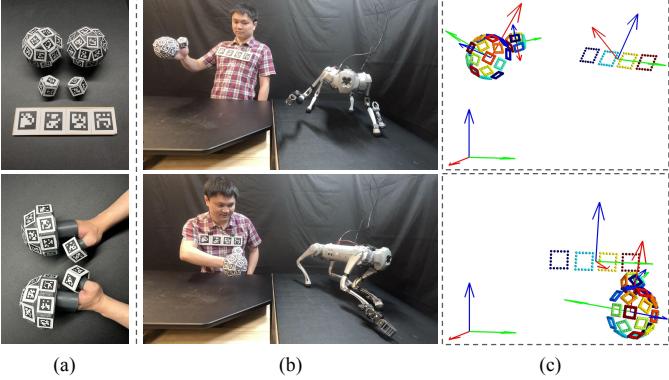


Figure 6: Low-cost teleoperation platform. (a) The rigid shells worn by palms, thumbs and torso created by 3D printing. (b) Simultaneous control of the 6D pose of the torso and the end effector of the quadrupedal robot. (c) Visualization of the estimated corresponding 6D poses of the user’s right palm, right thumb and the torso using 3D point clouds of AprilTag edges.

robot, while the relative angles between the 6D poses of the thumb and palm shells are used to command the opening and closing of the robot gripper. In this way, the teleoperation users can move their torso and both of their hands and fingers to naturally control all motion of the quadrupedal robot. With a few cameras and mounts, our teleoperation platform is especially low-cost in hardware compared to previous works, while allowing the users to simultaneously command torso pose, two end effector poses, and two gripper openings of the quadrupedal robot.

#### E. Versatile Manipulation in Real-World Tasks

Using our teleoperation platform, we validate LocoMan’s capabilities listed in Table I by commanding LocoMan to complete the following real-world robotic tasks using teleoperation:

*1) Manipulation with 6DoF End Effector:* To validate that LocoMan is capable of performing manipulations in the 6D operational space, we evaluate LocoMan on two challenging tasks: opening a sliding drawer and the swing door of a cabinet. These two tasks demand precise control over the robot end effector pose throughout the interaction.

As illustrated in Fig. 7 (a), when opening the sliding drawer, LocoMan successfully controls the end effector to move linearly along the sliding direction of the drawer while maintaining a stable orientation axis parallel to the movement path. Fig. 7(b) demonstrates LocoMan’s capability in opening a cabinet swing door, where it maneuvers the end effector in an arc trajectory and adjusts the gripper orientation to stay perpendicular to the door plane to avoid collision with the door handle.

Furthermore, we teleoperate LocoMan to perform two additional tasks that require precise motion control: inserting a charger into a power socket shown in Fig. 8 (a), LocoMan can accurately align the charger with the socket, demonstrating its ability to precisely adjust its end effector position. In Fig. 8 (b), LocoMan pours liquid from one cup to another, which highlights its ability to stably modulate the end effector orientation.



Figure 7: LocoMan shows its capability of manipulation in 6D operational space in the tasks of opening two typical types of cabinets under vision-based human motion teleoperation. (a) LocoMan skillfully opens a sliding drawer. (b) LocoMan adeptly opens a swing door cabinet.

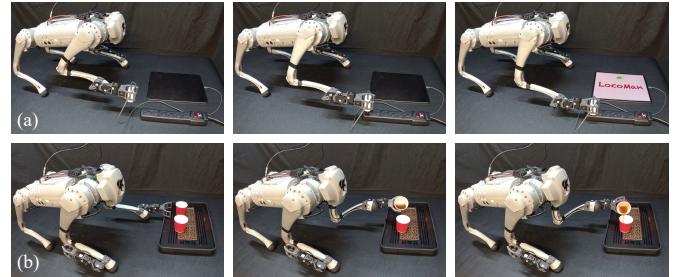


Figure 8: LocoMan demonstrates its precision and stability in completing two tasks that demand fine control under joystick teleoperation. (a) Insert a charger with meticulous alignment. (b) Pour liquid from a cup to another, carefully modulating the pour angle and rate.

*2) Manipulation in Narrow Space:* Fetching objects from an extremely narrow space is a challenge for quadrupedal robots due to the limited space for adjusting both the leg and manipulator joint angles.

As illustrated in Fig. 9 (a), the task requires LocoMan to walk below a bed to grasp an object located under a cabinet and then carry it out. This confined environment only offers a 0.25 m clearance below the bed and an even more constricted 0.09 m height under the cabinet. As illustrated in Fig. 9 (b), LocoMan can walk stably below the vertically restricted space, avoiding any contact with the bed. This is only possible without an additional robot arm installed on the top, such as those in [18, 21]. Thanks to the compact design of loco-manipulator, LocoMan can deftly maneuver its end effector to extend under the cabinet and adjust its pose to successfully grasp the object. LocoMan then retrieves its loco-manipulator while holding the object in the gripper and exits the narrow space. The successful completion of this challenging task demonstrates LocoMan’s competency in loco-manipulation within narrow and complex environments. This showcases its potential for deployment in scenarios such as search and rescue missions, which often necessitate both agile locomotion and manipulation in cramped spaces.

*3) Bimanual Manipulation:* Enabling quadrupedal robots to execute bimanual manipulation significantly broadens the spectrum of tasks they can undertake. With dual lightweight loco-manipulators installed on LocoMan’s two front legs, it becomes practical for LocoMan to stand upright and engage in complex bimanual tasks.

Hoisting objects with a rope is a task that requires precise bi-

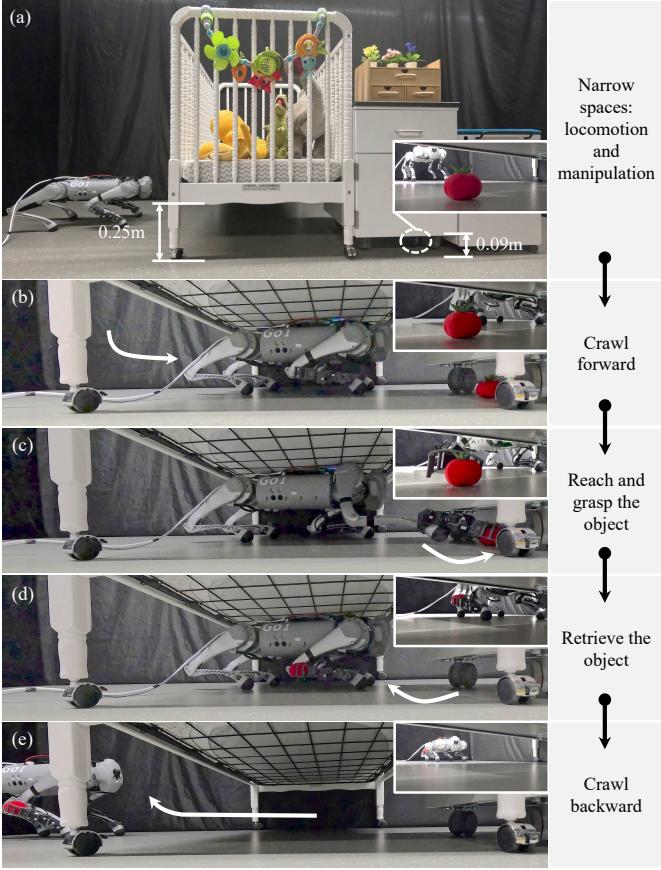


Figure 9: LocoMan demonstrates its locomotion and manipulation capabilities in constrained environments under joystick teleoperation. (a) The task requires the robot to walk beneath a crib with a clearance of 0.25 m and grasp the object under the cabinet with a clearance of 0.09 m. (b) LocoMan crawls forward to approach the targeted object. (c) LocoMan extends its loco-manipulator to reach and grasp the object. (d) LocoMan retrieves the object. (e) LocoMan crawls backward and exits the crib.

manual coordination to ensure the rope does not slip back. As illustrated in Fig. 10 (a), while maintaining an upright stance, LocoMan adeptly adjusts the 6D pose of an end effector to securely grasp the rope and applies a downward pull. Upon reaching the limits of its workspace, the pulling end effector maintains its grip on the rope, steadfastly holding it in place until the alternate end effector takes over to continue pulling the rope. The seamless transition and successful execution of the hoisting operation vividly illustrate LocoMan’s advanced bimanual coordination capabilities, which showcase the practical versatility and enhanced operational scope of quadrupedal robots equipped with our proposed loco-manipulators.

Furthermore, as Fig. 10 (b) shows, LocoMan can grasp a pair of socks hanged about 1 m high from the ground, which is relatively high compared to the height of the robot torso during four legs stance.

*4) Loco-Manipulation:* As illustrated in Fig. 1 (c), we instruct LocoMan to manipulate a flashlight aimed at a predetermined orientation trajectory within the world frame, while trotting forward at a velocity of  $0.15m/s$  and shifting left at  $0.05m/s$ . This necessitates the coordination of its locomotion and manipulation capabilities. Throughout the execution of

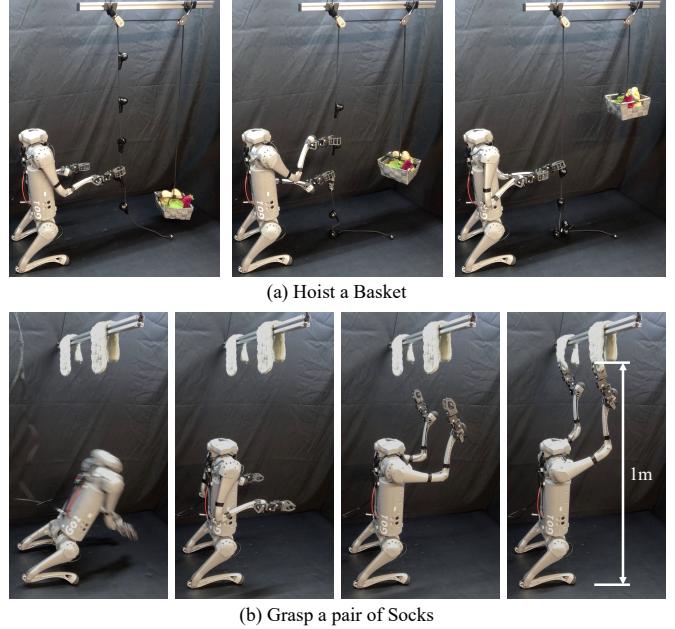


Figure 10: LocoMan is capable of performing challenging bimanual manipulation tasks with its dual lightweight loco-manipulators. (a) LocoMan collaboratively operates its two grippers to hoists a basket under vision-based human motion teleoperation. (b) LocoMan can reach 1m high to grasp a pair of socks under joystick teleoperation.

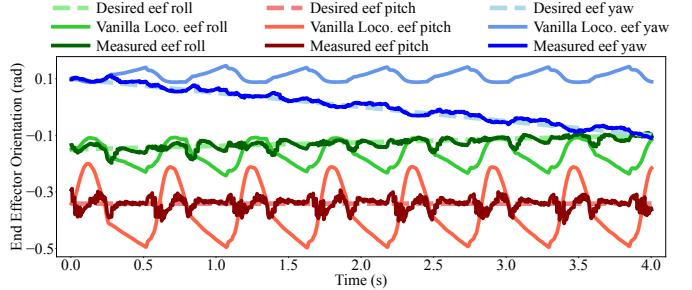


Figure 11: Vanilla Loco. denotes carrying the flashlight using the vanilla locomotion mode. The curves indicate that the loco-manipulation mode of LocoMan can consistently track the desired orientation of the flashlight in the loco-manipulation mode.

this task, we record the orientation of the end effector and also the orientation of its conjunct foot for calculating the vanilla end effector orientation based on their initial relative orientation. As illustrated in Fig. 11, in the loco-manipulation mode, LocoMan can track the desired orientation of the flashlight significantly better than relying solely on locomotion.

## VI. CONCLUSION

In this paper, we present LocoMan, a novel approach that enhances the manipulation dexterity of quadrupedal robots through the integration of two lightweight loco-manipulators, expanding their operational workspace and enabling precise control over complex 6D manipulation tasks. Our design effectively combines the mobility of quadrupedal robots with the functionality of manipulators, without compromising agility or requiring extensive payload capacity. The developed unified control framework ensures accurate and stable movement

across a spectrum of tasks, illustrating LocoMan’s versatility in environments ranging from confined spaces to tasks requiring intricate dual-arm coordination.

## VII. ACKNOWLEDGEMENTS

The work is partially supported by Google Deepmind with an unrestricted grant. The authors also acknowledge support from the National Science Foundation under grants CNS-2047454.

## REFERENCES

- [1] F. Jenelten, J. He, F. Farshidian, and M. Hutter, “Dtc: Deep tracking control,” *Science Robotics*, vol. 9, no. 86, p. eadh5401, 2024.
- [2] S. Choi, G. Ji, J. Park, H. Kim, J. Mun, J. H. Lee, and J. Hwangbo, “Learning quadrupedal locomotion on deformable terrain,” *Science Robotics*, vol. 8, no. 74, p. eade2256, 2023.
- [3] J. Lee, J. Hwangbo, L. Wellhausen, V. Koltun, and M. Hutter, “Learning quadrupedal locomotion over challenging terrain,” *Science robotics*, vol. 5, no. 47, p. eabc5986, 2020.
- [4] R. Yang, G. Yang, and X. Wang, “Neural volumetric memory for visual locomotion control,” in *CVPR 2023*, 2023.
- [5] A. Kumar, Z. Fu, D. Pathak, and J. Malik, “Rma: Rapid motor adaptation for legged robots,” *arXiv preprint arXiv:2107.04034*, 2021.
- [6] B. Lindqvist, S. Karlsson, A. Koval, I. Tevetzidis, J. Haluška, C. Kanelakis, A.-a. Agha-mohammadi, and G. Nikolakopoulos, “Multimodality robotic systems: Integrated combined legged-aerial mobility for subterranean search-and-rescue,” *Robotics and Autonomous Systems*, 2022.
- [7] Y. Yang, G. Shi, X. Meng, W. Yu, T. Zhang, J. Tan, and B. Boots, “Cajun: Continuous adaptive jumping using a learned centroidal controller,” *arXiv preprint arXiv:2306.09557*, 2023.
- [8] C. Li, M. Vlastelica, S. Blaes, J. Frey, F. Grimmerger, and G. Martius, “Learning agile skills via adversarial imitation of rough partial demonstrations,” in *Conference on Robot Learning*. PMLR, 2023.
- [9] Z. Zhuang, Z. Fu, J. Wang, C. G. Atkeson, S. Schwertfeger, C. Finn, and H. Zhao, “Robot parkour learning,” in *CoRL 2023*, 2023.
- [10] X. Cheng, K. Shi, A. Agarwal, and D. Pathak, “Extreme parkour with legged robots,” *arXiv preprint arXiv:2309.14341*, 2023.
- [11] D. Hoeller, N. Rudin, D. Sako, and M. Hutter, “Anymal parkour: Learning agile navigation for quadrupedal robots,” *arXiv preprint arXiv:2306.14874*, 2023.
- [12] A. Xiao, W. Tong, L. Yang, J. Zeng, Z. Li, and K. Sreenath, “Robotic guide dog: Leading a human with leash-guided hybrid physical interaction,” in *ICRA 2021*, 2021.
- [13] X. Cheng, A. Kumar, and D. Pathak, “Legs as manipulator: Pushing quadrupedal agility beyond locomotion,” *arXiv preprint arXiv:2303.11330*, 2023.
- [14] Y. Ji, Z. Li, Y. Sun, X. B. Peng, S. Levine, G. Berseth, and K. Sreenath, “Hierarchical reinforcement learning for precise soccer shooting skills using a quadrupedal robot,” in *IROS 2022*, 2022.
- [15] Y. Ji, G. B. Margolis, and P. Agrawal, “Dribblebot: Dynamic legged manipulation in the wild,” *arXiv preprint arXiv:2304.01159*, 2023.
- [16] S. Jeon, M. Jung, S. Choi, B. Kim, and J. Hwangbo, “Learning whole-body manipulation for quadrupedal robot,” *IEEE RA-L*, 2023.
- [17] M. Sombolostan and Q. Nguyen, “Hierarchical adaptive locomanipulation control for quadruped robots,” in *ICRA 2023*, 2023.
- [18] Z. Fu, X. Cheng, and D. Pathak, “Deep whole-body control: learning a unified policy for manipulation and locomotion,” in *CoRL 2023*, 2023.
- [19] Y. Tsvetkov and S. Ramamoorthy, “A novel design and evaluation of a dactylus-equipped quadruped robot for mobile manipulation,” in *IROS 2022*, 2022.
- [20] P. Arm, M. Mittal, H. Kolvenbach, and M. Hutter, “Pedipulate: Enabling manipulation skills using a quadruped robot’s leg,” *arXiv preprint arXiv:2402.10837*, 2024.
- [21] J.-P. Sleiman, F. Farshidian, M. V. Minniti, and M. Hutter, “A unified mpc framework for whole-body dynamic locomotion and manipulation,” *IEEE RA-L*, 2021.
- [22] M. Sombolostan and Q. Nguyen, “Hierarchical adaptive control for collaborative manipulation of a rigid object by quadrupedal robots,” *arXiv preprint arXiv:2303.06741*, 2023.
- [23] O. Nachum, M. Ahn, H. Ponte, S. Gu, and V. Kumar, “Multi-agent manipulation via locomotion using hierarchical sim2real,” *arXiv preprint arXiv:1908.05224*, 2019.
- [24] M. J. Mataric, M. Nilsson, and K. T. Simsarin, “Cooperative multi-robot box-pushing,” in *Proceedings 1995 IEEE/RSJ International Conference on Intelligent Robots and Systems. Human Robot Interaction and Cooperative Robots*, vol. 3. IEEE, 1995, pp. 556–561.
- [25] F. Shi, T. Homberger, J. Lee, T. Miki, M. Zhao, F. Farshidian, K. Okada, M. Inaba, and M. Hutter, “Circus anymal: A quadruped learning dexterous manipulation with its limbs,” in *ICRA 2021*, 2021.
- [26] H. Kolvenbach, C. Bärtschi, L. Wellhausen, R. Grandia, and M. Hutter, “Haptic inspection of planetary soils with legged robots,” *IEEE RA-L*, 2019.
- [27] T. Topping, G. Kenneally, and D. E. Koditschek, “Quasi-static and dynamic mismatch for door opening and stair climbing with a legged robot,” in *ICRA 2017*, 2017.
- [28] S. Kim, M. Sorokin, J. Lee, and S. Ha, “Human Motion Control of Quadrupedal Robots using Deep Reinforcement Learning,” in *Proceedings of Robotics: Science and Systems*, New York City, NY, USA, 2022.
- [29] N. Yokoyama, A. W. Clegg, E. Undersander, S. Ha, D. Batra, and A. Rai, “Adaptive skill coordination for robotic mobile manipulation,” *arXiv preprint arXiv:2304.00410*, 2023.
- [30] J.-R. Chiu, J.-P. Sleiman, M. Mittal, F. Farshidian, and M. Hutter, “A collision-free mpc for whole-body dynamic locomotion and manipulation,” in *ICRA 2022*, 2022.
- [31] J.-P. Sleiman, F. Farshidian, and M. Hutter, “Versatile multicontact planning and control for legged loco-manipulation,” *Science Robotics*, vol. 8, no. 81, p. eadg5014, 2023.
- [32] M. Mittal, D. Hoeller, F. Farshidian, M. Hutter, and A. Garg, “Articulated object interaction in unknown scenes with whole-body mobile manipulation,” in *IROS 2022*, 2022.
- [33] X. Chu, S. Wang, M. Feng, J. Zheng, Y. Zhao, J. Huang, and K. S. Au, “Model-free large-scale cloth spreading with mobile manipulation: Initial feasibility study,” in *2023 IEEE 19th International Conference on Automation Science and Engineering (CASE)*. IEEE, 2023, pp. 1–6.
- [34] H. Ferrolho, V. Ivan, W. Merkt, I. Havoutis, and S. Vijayakumar, “Roloma: Robust loco-manipulation for quadruped robots with arms,” *Autonomous Robots*, vol. 47, no. 8, pp. 1463–1481, 2023.
- [35] F. De Vincenti and S. Coros, “Centralized model predictive control for collaborative loco-manipulation,” *RSS 2023*, 2023.
- [36] J. Kim and K. Akbari Hamed, “Cooperative locomotion via supervisory predictive control and distributed nonlinear controllers,” *Journal of Dynamic Systems, Measurement, and Control*, 2022.
- [37] A. Lykov, M. Litvinov, M. Konenkov, R. Prochii, N. Burtsev, A. A. Abdulkarim, A. Bazhenov, V. Berman, and D. Tsatsas, “Cognitivedog: Large multimodal model based system to translate vision and language into action of quadruped robot,” *arXiv preprint arXiv:2401.09388*, 2024.
- [38] A. Roennau, G. Heppner, M. Nowicki, and R. Dillmann, “Lauron v: A versatile six-legged walking robot with advanced maneuverability,” in *2014 IEEE/ASME International Conference on Advanced Intelligent Mechatronics*. IEEE, 2014, pp. 82–87.
- [39] G. Heppner, T. Buettner, A. Roennau, and R. Dillmann, “Versatile-high power gripper for a six legged walking robot,” in *Mobile Service Robotics*. World Scientific, 2014, pp. 461–468.
- [40] G. Heppner, A. Roennau, J. Oberländer, S. Klemm, and R. Dillmann, “Laurope-six legged walking robot for planetary exploration participating in the spacebot cup,” *WS on Advanced Space Technologies for Robotics and Automation*, vol. 2, no. 13, pp. 69–76, 2015.
- [41] J. Whitman, S. Su, S. Coros, A. Ansari, and H. Choset, “Generating gaits for simultaneous locomotion and manipulation,” in *IROS 2017*, 2017.
- [42] W. Brinkmann, A. Dettmann, L. C. Danter, C. Schulz, T. Stark, and A. Brandt, “Enhancement of the six-legged robot mantis for assembly and construction tasks in lunar mission scenarios within a multi-robot team,” in *Proceedings: International Symposium on Artificial Intelligence, Robotics and Automation in Space*, 2020.
- [43] D. Kim, J. Di Carlo, B. Katz, G. Bledt, and S. Kim, “Highly dynamic quadruped locomotion via whole-body impulse control and model predictive control,” *arXiv preprint arXiv:1909.06586*, 2019.
- [44] E. Olson, “Apriltag: A robust and flexible visual fiducial system,” in *ICRA 2011*, 2011.
- [45] X. Liu, R. Jonschkowski, A. Angelova, and K. Konolige, “Keypose: Multi-view 3d labeling and keypoint estimation for transparent objects,” in *CVPR 2020*, 2020.
- [46] X. Liu, S. Iwase, and K. M. Kitani, “Stereobj-1m: Large-scale stereo image dataset for 6d object pose estimation,” in *ICCV 2021*, 2021.

## APPENDIX

### A. Joint Limits of Loco-Manipulator

The range of motion for loco-manipulator's joints is illustrated in Fig. 12, with angles of 280°, 210°, and 360° for the first through the final joint, respectively.

The motion limitation of the first joint arises from a collision between the first servo and the support structure that is part of the outer calf connector (Fig. 2). This support structure is designed primarily to connect to the opposite horn of the first servo, playing a vital role in stabilizing the manipulator. The design of the support structure aims to minimize its impact on the essential range of motion of the first servo. This critical range is necessary for extending the gripper during transitions from locomotion to manipulation and for positioning the gripper forwards and towards the ground, representing the primary workspace. The limitations in the motion of the second joint are due to self-collision. For the third joint, it does not face stringent motion restrictions, except for those imposed by the cable's length.

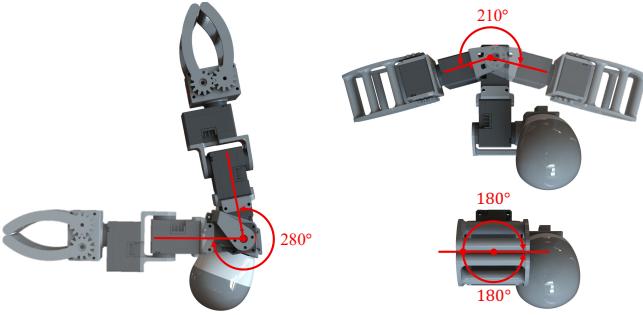


Figure 12: The range of motion for loco-manipulator's joints. From the first joint to the final one, the corresponding angular movements are 280°, 210°, and 360° respectively.

### B. Whole-Body Impulse Control

**Notation** from IV: robot state  $\mathbf{q} = (\mathbf{x}^{\text{torso}}, \mathbf{q}_j) \in \mathbb{R}^{24}$ , where  $\mathbf{x}^{\text{torso}} \in \mathbb{R}^6$  denotes the position and orientation of the floating torso, and  $\mathbf{q}_j \in \mathbb{R}^{18}$  denotes the joint angles of LocoMan. The bar notation ( $\bar{\cdot}$ ) denotes all desired states.

1) **Hierarchical Objective Tracking:** As depicted in Table III, there are  $n$  tracking objectives in different operation modes, comprising  $n$  tracking tasks to be executed. The null space projection technique is utilized to track each objective in hierarchical order (from 1-st to  $n$ -th).

In the  $i$ -th ( $i = 1, \dots, n$ ) task, the tracking items consist of the desired position  $\bar{\mathbf{x}}_i$ , velocity  $\dot{\bar{\mathbf{x}}}_i$ , and acceleration  $\ddot{\bar{\mathbf{x}}}_i$  of the  $i$ -th tracking objective, which are tracked by computing the desired position  $\bar{\mathbf{q}}_i$ , velocity  $\dot{\bar{\mathbf{q}}}_i$ , and acceleration  $\ddot{\bar{\mathbf{q}}}_i$  for both the floating torso and the joints. The desired states ( $\bar{\mathbf{q}}_n$ ,  $\dot{\bar{\mathbf{q}}}_n$ , and  $\ddot{\bar{\mathbf{q}}}_n$ ) for the last tracking objective would be the final desired states ( $\bar{\mathbf{q}}$ ,  $\dot{\bar{\mathbf{q}}}$ , and  $\ddot{\bar{\mathbf{q}}}$ ) of LocoMan.

$$\begin{aligned}\bar{\mathbf{q}}_i &= \bar{\mathbf{q}}_{i-1} + \mathbf{J}_{i|\text{pre}}^\dagger (\bar{\mathbf{x}}_i - \mathbf{x}_i - \mathbf{J}_i (\bar{\mathbf{q}}_{i-1} - \mathbf{q})) \\ \dot{\bar{\mathbf{q}}}_i &= \dot{\bar{\mathbf{q}}}_{i-1} + \mathbf{J}_{i|\text{pre}}^\dagger (\dot{\bar{\mathbf{x}}}_i - \mathbf{J}_i \dot{\bar{\mathbf{q}}}_{i-1}) \\ \ddot{\bar{\mathbf{q}}}_i &= \mathbf{J}_{i|\text{pre}}^{\text{dyn} \dagger} (\ddot{\bar{\mathbf{x}}}_i^{\text{cmd}} - \mathbf{J}_i \dot{\mathbf{q}} - \mathbf{J}_i \dot{\bar{\mathbf{q}}}_{i-1})\end{aligned}\quad (1)$$

The acceleration command  $\ddot{\mathbf{x}}_i^{\text{cmd}}$  is computed by:

$$\ddot{\mathbf{x}}_i^{\text{cmd}} = \bar{\mathbf{x}}_i + \mathbf{K}_p^{\text{acc}} (\bar{\mathbf{x}}_i - \mathbf{x}_i) + \mathbf{K}_d^{\text{acc}} (\dot{\bar{\mathbf{x}}}_i - \dot{\mathbf{x}}_i) \quad (2)$$

where  $\mathbf{K}_p^{\text{acc}}$  and  $\mathbf{K}_d^{\text{acc}}$  denote position and velocity feedback gains.

To compute the desired states of LocoMan from the tracking objective, we need to use the pseudo-inverse of the Jacobian. In Equation (1), we track the desired position and velocity kinematically, and the desired acceleration dynamically. Therefore, two types of pseudo-inverses are utilized. One is the SVD-based pseudo-inverse denoted by  $\{\cdot\}^\dagger$ , which is defined as:

$$\begin{aligned}\mathbf{J} &= \mathbf{U} \Sigma \mathbf{V}^T \\ \mathbf{J}^\dagger &= \mathbf{V} \Sigma^\dagger \mathbf{U}^T\end{aligned}\quad (3)$$

where  $\mathbf{J}$  is the Jacobian and  $\mathbf{U}$  and  $\mathbf{V}$  are the orthogonal matrices. Besides,  $\Sigma$  is the diagonal matrix, and  $\Sigma^\dagger$  is formed by taking the reciprocal of each non-zero element on the diagonal of  $\Sigma$ , leaving the zeros in place, and then transposing the matrix.

The other is the dynamically consistent pseudo-inverse denoted by  $\{\cdot\}^\ddagger$  with the definition of:

$$\mathbf{J}^\ddagger = \mathbf{A}^{-1} \mathbf{J}^\top (\mathbf{J} \mathbf{A}^{-1} \mathbf{J}^\top)^{-1} \quad (4)$$

where  $\mathbf{A}$  is the mass matrix.

Furthermore,  $\mathbf{J}_i$  is the Jacobian of the  $i$ -th tracking objective, while  $\mathbf{J}_{i|\text{pre}}$  is the projection of it into the null space of all previous tasks. And the symbol  $\{\cdot\}^{\text{dyn}}$  denotes the dynamically consistent constraints.  $\mathbf{J}_{i|\text{pre}}$  and  $\mathbf{J}_{i|\text{pre}}^{\text{dyn}}$  are computed with the following steps:

$$\begin{aligned}\mathbf{N}_{i-1} &= \mathbf{N}_0 \mathbf{N}_{1|0} \cdots \mathbf{N}_{i-1|i-2} \\ \mathbf{J}_{i|\text{pre}} &= \mathbf{J}_i \mathbf{N}_{i-1} \\ \mathbf{J}_{i|i-1} &= \mathbf{J}_i (\mathbf{I} - \mathbf{J}_{i-1}^\dagger \mathbf{J}_{i-1}) \\ \mathbf{N}_{i|i-1} &= \mathbf{I} - \mathbf{J}_{i|i-1}^\dagger \mathbf{J}_{i|i-1} \\ \mathbf{N}_{i-1}^{\text{dyn}} &= \mathbf{N}_0^{\text{dyn}} \mathbf{N}_{1|0}^{\text{dyn}} \cdots \mathbf{N}_{i-1|i-2}^{\text{dyn}} \\ \mathbf{J}_{i|\text{pre}}^{\text{dyn}} &= \mathbf{J}_i \mathbf{N}_{i-1}^{\text{dyn}} \\ \mathbf{J}_{i|i-1}^{\text{dyn}} &= \mathbf{J}_i (\mathbf{I} - \mathbf{J}_{i-1}^{\text{dyn} \dagger} \mathbf{J}_{i-1}) \\ \mathbf{N}_{i|i-1}^{\text{dyn}} &= \mathbf{I} - \mathbf{J}_{i|i-1}^{\text{dyn} \dagger} \mathbf{J}_{i|i-1}^{\text{dyn}}\end{aligned}\quad (5)$$

where  $\mathbf{N}_{i-1}$ ,  $\mathbf{J}_{i|i-1}$ , and  $\mathbf{N}_{i|i-1}$  represent the null space of the previous  $(i-1)$  tasks, the projection of the Jacobian  $\mathbf{J}_i$  into the null space of the  $(i-1)$ -th task, and the independent null space introduced by the  $i$ -th task.

Here,  $i$  is within the range of  $1, \dots, n$ . And the initial items are defined as:

$$\begin{aligned}\mathbf{N}_0 &= \mathbf{I} - \mathbf{J}_c^\dagger \mathbf{J}_c \\ \mathbf{N}_0^{\text{dyn}} &= \mathbf{I} - \mathbf{J}_c^{\text{dyn} \dagger} \mathbf{J}_c \\ \mathbf{J}_0 &= \mathbf{J}_c \\ \bar{\mathbf{q}}_0 &= \mathbf{q} \\ \dot{\bar{\mathbf{q}}}_0 &= \mathbf{0} \\ \ddot{\bar{\mathbf{q}}}_0 &= \mathbf{J}_c^\dagger (-\mathbf{J}_c \dot{\mathbf{q}})\end{aligned}\quad (6)$$

where  $\mathbf{J}_c$  is the Jacobian of the feet in contact.

2) **Quadratic Programming:** With the desired acceleration  $\bar{\ddot{q}}$  computed in the previous step, we solve a Quadratic Program (QP) to optimize for ground reaction forces. The QP problem is formulated as:

$$\min_{\mathbf{f}_r, \delta_t} \mathbf{f}_r^\top \mathbf{Q}_1 \mathbf{f}_r + \delta_t^\top \mathbf{Q}_2 \delta_t \quad (7)$$

$$\text{s.t. } \mathbf{S}_f(\mathbf{A}\ddot{\mathbf{q}}^{\text{cmd}} + \mathbf{b} + \mathbf{g}) = \mathbf{S}_f \mathbf{J}_c^\top \mathbf{f}_r \quad (\text{floating torso dyn.})$$

$$\ddot{\mathbf{q}}^{\text{cmd}} = \bar{\ddot{\mathbf{q}}} + \begin{bmatrix} \delta_t \\ \mathbf{0}_j \end{bmatrix} \quad (\text{floating torso acceleration})$$

$$\mathbf{W}\mathbf{f}_r \geq \mathbf{0} \quad (\text{contact force constraints})$$

where the optimization objective is to minimize the reaction force  $\mathbf{f}_r$  and the relaxation variable  $\delta_t$  for the floating torso acceleration with the weights  $\mathbf{Q}_1$  and  $\mathbf{Q}_2$ .  $\mathbf{S}_f$ ,  $\mathbf{A}$ ,  $\mathbf{b}$ ,  $\mathbf{g}$ , and  $\mathbf{W}$  are the floating torso selection matrix, the mass matrix, the Coriolis force, the gravitation force, and the force constraint matrix for foot contact, respectively.

With the optimized reaction force  $\mathbf{f}_r$  and relaxation variable  $\delta_t$  for the floating torso acceleration, we compute the joint torque command with the multi-body dynamics formulation:

$$\begin{bmatrix} \bar{\tau}_t \\ \bar{\tau}_j \end{bmatrix} = \mathbf{A}\ddot{\mathbf{q}}^{\text{cmd}} + \mathbf{b} + \mathbf{g} - \mathbf{J}_c^\top \mathbf{f}_r \quad (8)$$

where  $\tau_j$  is the torque command for LocoMan's joints.

3) **Joint-Level PD Control:** We compute the desired position  $\bar{q}_j$ , velocity  $\dot{\bar{q}}_j$ , and torque  $\bar{\tau}_j$  for LocoMan's joints in the previous steps, which runs in 400Hz. The final torque command  $\tau_j$  applied to each joint is computed by a joint-level PD controller that runs in 2000Hz:

$$\tau_j = \bar{\tau}_j + \mathbf{K}_p^{\text{tau}} (\bar{q}_j - q_i) + \mathbf{K}_d^{\text{tau}} (\dot{\bar{q}}_j - \dot{q}_i) \quad (9)$$

The high frequency joint-level PD feedback enables the robot to track desired poses smoothly and accurately.

4) **PD Gains for Real Robot:** In the experiments described in V, we tune the PD gains of the tracking objectives for the whole-body impulse controller (Table V) and the PD gains of the hip, thigh, and calf joints for the joint-level PD controller (Table VI). We use the SDK from Dynamixel® and set the position gain values to  $[8, 6, 6]^T \times 128$  and the velocity gain values to  $[65, 40, 40]^T \times 16$  for the three joints of locomanipulator in all operation modes, where 128 and 16 are the default scales in the SDK.

Table V: PD gains of the tracking objectives for computing the acceleration errors in whole-body impulse control.

Tracking Objective	$\mathbf{K}_p^{\text{acc}}$	$\mathbf{K}_d^{\text{acc}}$
Torso Position (Locomotion)	$[100, 100, 100]^T$	$[10, 10, 10]^T$
Torso Orientation (Locomotion)	$[100, 100, 100]^T$	$[10, 10, 10]^T$
Torso Position (Manipulation)	$[100, 100, 100]^T$	$[1, 1, 1]^T$
Torso Orientation (Manipulation)	$[100, 100, 100]^T$	$[1, 1, 1]^T$
Foot Position	$[100, 100, 100]^T$	$[10, 10, 10]^T$
Foot Orientation	$[100, 100, 100]^T$	$[10, 10, 10]^T$
Gripper Position	$[100, 100, 100]^T$	$[10, 10, 10]^T$
Gripper Orientation	$[100, 100, 100]^T$	$[10, 10, 10]^T$

Table VI: PD gains of the hip, thigh, and calf joints for computing the applied torque in joint-level PD control.

Leg Mode	$\mathbf{K}_p^{\text{tau}}$	$\mathbf{K}_d^{\text{tau}}$
Stance Legs in Locomotion	$[30, 30, 30]^T$	$[1, 1, 1]^T$
Swing Legs in Locomotion	$[30, 30, 30]^T$	$[1, 1, 1]^T$
Stance Legs in Loco-Manipulation	$[30, 30, 30]^T$	$[1, 1, 1]^T$
Swing Legs in Loco-Manipulation	$[20, 20, 20]^T$	$[0.8, 0.8, 0.8]^T$
Stance Legs in Single-Arm Manipulation	$[60, 60, 60]^T$	$[2, 2, 2]^T$
Swing Legs in Single-Arm Manipulation	$[60, 60, 60]^T$	$[2, 2, 2]^T$
Stance Legs in Bimanual Manipulation	$[100, 100, 100]^T$	$[2.5, 2.5, 2.5]^T$
Swing Legs in Bimanual Manipulation	$[30, 30, 30]^T$	$[1, 1, 1]^T$

### C. State Estimation

As illustrated in Table III, there are two state estimators in our system, including the kinematics-based estimator used in the manipulation modes and the Kalman filter-based estimator used in the locomotion mode and the loco-manipulation mode.

As depicted in Fig. 13 (a), at the moment when LocoMan finishes the transition process from other operation modes to manipulation modes, the kinematics-based estimator sets the base frame of LocoMan as the world frame. Furthermore, we also record the positions of the contact foot as  $\mathbf{P}_{\text{init}}^W$ , which can be computed using forward kinematics and have the same values in the world frame or the base frame at the reset moment. During manipulation, the world frame remains fixed, and we compute the current positions of the contact foot in the base frame as  $\mathbf{P}_{\text{cur}}^B$ . Assuming that there is no slip occurring between the contact foot and the ground, we employ the Procrustes analysis method to solve the transform matrix from the world frame to the base frame  $\mathbf{T}_W^B$ .

As for locomotion, we implement the Kalman filter-based estimator which fuses the IMU information. As Fig. 13 (b) shows, the world frame is always updated to be the zero-yaw world frame.

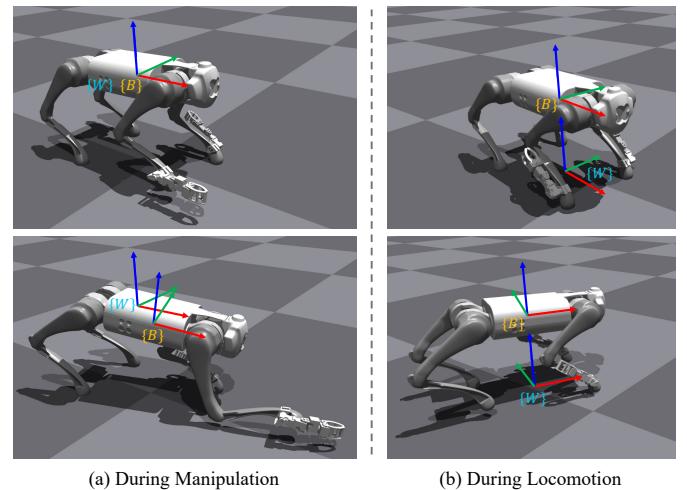


Figure 13: The definition of the world frame in different state estimators. (a) The kinematics-based estimator set the initial base frame of LocoMan as the world frame. (b) The Kalman filter-based estimator always has the zero-yaw world frame as the world frame.