arXiv:2603.02443v1 [cs.RO] 2 Mar 2026

通过结合模型与学习型控制实现安全全身运动操作

Alexander Schperberg[†*]、Yeping Wang[‡] 和 Stefano Di Cairano[†]

摘要——运动与操作的同步进行，使机器人能够突破固定基座的限制，与环境交互。然而，在兼顾接触交互过程中的安全性与柔顺性的同时，协调腿式移动与机械臂操作仍然颇具挑战。为此，我们提出一种全身控制器，将机械臂的基于模型的导纳控制与用于腿式移动的强化学习（RL）策略相结合。导纳控制器将外部力和力矩（例如人类在物理交互过程中施加的力和力矩）映射为期望的末端执行器速度，从而实现柔顺行为。手臂与腿部控制器协同跟踪这些速度，实现统一的六自由度力响应。基于模型的设计可实现精确的力控制，并通过参考调节器（RG）提供安全保证；此外，利用神经网络增强的卡尔曼滤波器进行可靠的基座速度估计，进一步提升了系统的鲁棒性。我们使用配备六自由度机械臂和腕部安装的六自由度力/扭矩传感器的 Unitree Go2 四足机器人，在仿真和硬件平台上验证了该方法。结果表明，该方法能够准确跟踪交互驱动速度，呈现柔顺行为，并在动态场景下实现安全、可靠的性能。

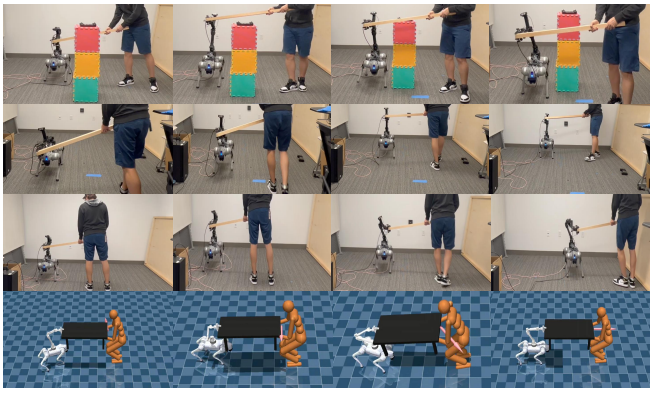


图1. 我们提出一种基于模型与学习的组合式全身控制器，可在仿真与硬件中实现人类与配备机械臂的四足机器人之间的安全协作。

I. 引言

配备机械臂的腿式机器人不仅有望克服不平坦地形或复杂的人类环境，还能扩大机械臂的整体工作空间，从而拓展操作范围，并胜任从搜救到人机协作的任务。对此类复杂系统进行建模本身就极具挑战性，现有方法通常要么求解一个联合处理运动与操作（运动操作）的大规模优化问题 [1],，要么依赖基于学习的方法 [2], ；随着可控关节数量增加，这些方法要么缺乏正式的安全性保证，要么需要海量数据。由于系统复杂，即使是基于学习的方法通常也会采用分层架构 [3], ，将手臂与腿部策略分开；但这会限制性能，因为两种策略无法协同配合以完成预期任务。尽管一些最先进算法开始采用更统一的策略，只需单次训练[4], ，但如何充分控制接触力仍是运动操作领域一项活跃的研究课题。这一点尤为突出，因为许多方法主要关注末端执行器位姿跟踪。此外，尽管强化学习（RL）已在富接触任务中取得成功，但通常只是隐式地处理接触力，而不对其进行显式调节。然而，调节力的能力至关重要，它能实现柔顺行为，而这对于动觉示教、安全的人机交互、避免机器人或他人在接触过程中受到损伤，以及通过优化身体姿态来改善力可操作度椭球、进而产生更大的任务对齐力都不可或缺。

我们的工作属于全身移动操作中的力控制领域。尽管有些研究考虑了接触力 [5], ，但它们通常不会使用力/力矩（F/T）传感器直接测量接触力，而是依靠简单的弹簧阻尼模型进行估算。尤其是，鉴于仿真器难以模拟如此复杂的接触力，我们认为，采用力/力矩传感器的方法对于安全性和准确性至关重要。更重要的是，许多现有方法忽视了安全考量，或在控制器设计中缺乏形式化保证——有些方法甚至只采用基于学习的策略 [5], [6],——因此我们力求采用不同的方法。我们将用于机械臂的基于模型的导纳控制器与用于腿式移动的基于学习的强化学习策略相结合。选择基于模型的机械臂设计，不仅是因为机械臂会直接接触人类或物体，对安全性和准确性有更高要求，也因为这种设计使我们能够实施并保证安全性。在我们的方案中，这是通过参考调节器（RG）实现的 [7]。尽管手臂与腿部控制器采用模块化设计，我们的方法仍通过在强化学习训练过程中纳入导纳控制器，处理移动操作中的耦合动力学。

[†]Alexander Schperberg和Stefano Di Cairano就职于三菱电机研究实验室（MERL），剑桥，马萨诸塞州，02139，美国。{schperberg, dicairano}@merl.com

[‡]Yeping Wang在本文部分工作期间就职于三菱电机研究实验室（MERL），剑桥，马萨诸塞州，02139，美国。yeping@cs.wisc.edu

*通讯作者

该控制器跟踪末端执行器的力和力矩，并通过雅可比矩阵将其映射为组合臂-基座系统期望的线速度和角速度。最后，由于导纳控制器和参考调节器（RG）均为状态反馈方法，状态估计会直接影响控制器性能 [8]。因此，我们设计了一种结合神经网络的卡尔曼滤波器，以提高线速度和角速度的估计精度。在训练过程中，我们还利用卡尔曼滤波器的状态协方差矩阵为强化学习观测提供信息，从而缩小仿真到现实差距（见第 III-C 节）。

总体而言，我们提出以下**贡献**：
1) 我们提出了一种用于腿式移动操作的基于模型与学习的组合式全身控制器，该控制器兼顾腿部和机械臂，通过力/力矩传感器显式跟踪力和力矩，同时保证安全性。2) 我们通过参考调节器，在机械臂的控制器设计中实施安全保证。3) 我们构建了一种结合神经网络的卡尔曼滤波器，用于对复杂系统动力学建模，以便估计基座的线速度和角速度，从而直接提升组合式控制器的稳定性和准确性。4) 我们通过仿真和硬件实验验证了所提出的方法。

II. 相关工作

移动操作领域可根据共同采用的方法分为以下几类：第 II-.1 节：仅修改操作侧的规划与控制架构，同时采用标准的模型预测控制或基于强化学习的运动控制器（即不使用全身控制）；第 II-.2 节：集成基于模型的全身控制器，以考虑运动与操作之间的耦合；第 II-.3 节：将基于学习的方法应用于全身控制，这可能涉及两步过程：先训练低层策略，再使用感知和控制训练高层策略；也可以采用一步训练的统一策略；第 II-.4 节：采用结合基于模型的操作控制与基于学习的运动控制器的方法，或反之。最后，第 II-.5 节重点讨论移动操作中的基于力的控制器，以实现柔顺行为。

1）不考虑全身运动的控制：通过机器人 API 调用默认的基于模型的控制器来实现移动操作，相关演示见 [9]–[11]。例如，Zhang等人 [9]专注于在线抓取位姿融合，使机器人在运动过程中也能抓取物体。Yokoyama 等人 [10] 利用贝叶斯基函数收集多种操作任务的数据，包括开门，而运动则由现有控制器负责。Arcari 等人 [11] 研究长时域抓取与放置任务，通过 SPOT API 协调高层命令，引导机器人导航至物体处、拾取物体并将其运送至目标位置。这些工作的局限在于，它们将操作与运动模块彼此分离，且未考虑腿部与机械臂同时运动所产生的复杂动力学，从而降低了性能或延长了任务完成时间。

2）基于模型的全身控制： [1], [12]–[14]展示了通过纯基于模型的方法实现机械臂与腿部的耦合。 [1], [12], [13],中使用全身逆动力学模型，确保系统在通过轨迹优化（TO）计算并由模型预测控制器（MPC）跟踪的参考轨迹附近保持稳定。Murooka 等人 [14]使用基于接触的点规划与控制来操作物体，同时考虑人形机器人全身动力学。这些工作倾向于采用简化的机器人基座/机械臂动力学，以降低在线基于模型控制的计算成本，但这可能导致性能受限。

3）基于强化学习的全身控制：端到端基于学习的运动操作全身控制器也取得了成功成果 [2]–[4]。Liu 等人 [2] 提出了视觉全身控制器（VBC）。VBC采用低层级目标到达策略，并在第二阶段训练中利用该策略学习高层级任务规划策略。RoboDuet [3] 是另一项工作，它训练了两个策略：一个用于运动，另一个用于操作。研究展示了对六自由度末端执行器位姿以及机器人基座速度的成功跟踪。与此同时，Fu 等人[4] 提出统一策略，避免了训练两个不同策略及后续奖励调优的复杂性。四足机器人的关节与机械臂同时接受训练，以利用领域自适应技术跟踪期望的末端执行器位姿和速度。尽管基于学习的方法因其泛化能力以及对复杂动力学的建模能力而颇具吸引力，但其输出不可预测，因此难以量化安全性。

4）结合模型与基于学习的全身控制：为应对基于学习的方法的不可预测性，也有研究采用模型与基于学习的方法相结合的方式 [15], [16]。例如，ALMA [15] 是一种在线运动规划框架，它将用于机械臂、跟踪期望末端执行器位姿的基于模型的模型预测控制控制器，与用于维持平衡的基于学习的运动策略相结合。然而，腿部与机械臂控制器独立运行，该框架并未展现出动态协调的运动。类似地，Zimmermann 等人 [16] 同样结合了模型与基于学习的组件，但效果有限，因为该方法缺乏机械臂与腿部之间的紧密协调。总体而言，这种模块化分离可能导致子系统之间的误差传播，从而产生不平滑且可能不安全的整体行为。

5）全身控制中的隐式力跟踪：尽管大多数运动操作研究侧重于跟踪参考末端执行器位姿和基座速度，但它们并未考虑接触力之间的相互作用，以实现柔顺行为。这种柔顺性有利于动觉示教、防止机器人或其环境受损，并促进人机协作。仅有少数工作探索了这一领域 [5], [6], [17]–[20]。例如，

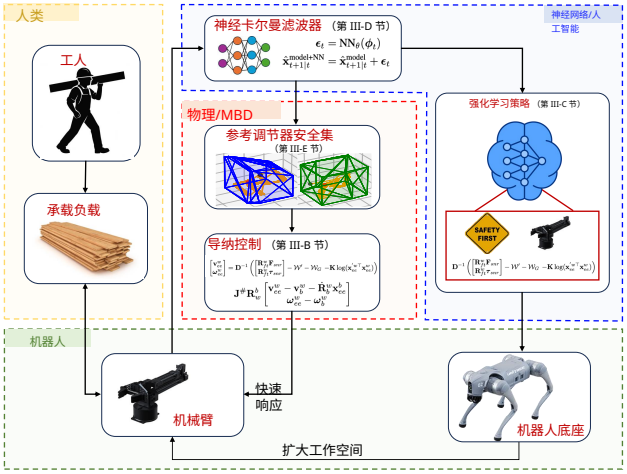


图2. 我们的框架架构将神经网络/人工智能方法与基于模型的设计（MBD）相结合，用于安全的人机协作。

Portela 等人 [5] 训练了一种以直接目标力作为输入的强化学习策略，使远程操作者能够调节机械臂的柔顺性。该方法使用阻抗控制器，要求机器人硬件支持直接力矩指令，且无需力/力矩（F/T）传感器。也有一些工作侧重于通过力交互学习下游任务，例如利用安装式机械臂进行跌倒恢复 [17]，或利用机器人足部运球 [18], [19]。尽管这些工作通常为机械臂和腿部分别采用独立策略，以实现基于力的运动操作控制，Zhi 等人 [6] 展示了一种统一策略，可在不需要力/力矩传感器的情况下集成力与位置控制。该策略采用力估计器，根据目标位置与当前位置之间的偏差来预测外力（即基于弹簧的模型）。尽管这些工作对运动操作采用了以力为中心的方法，但其力跟踪能力仍然有限，这主要是由于缺少力/力矩传感器，因而难以胜任需要精确、准确力控制的任务。重要的是，它们在控制设计中并未提供安全保证。

III. 方法

为明确问题，考虑这样一个应用示例：人类与配备机械臂的四足机器人共同搬运负载。机械臂抓取物体，并必须对施加于末端执行器的力旋量作出顺应性响应，使其沿着人类引导的运动方向移动。随后利用这种运动来操控机器人底座，使机器人跟随与其共同搬运负载的人类。如图 2 所示，我们提出一种结合模型与学习的全身控制器，以实现安全的人机协作。

A. 问题建模

我们将机械臂建模为一种快速、低惯量、短程的执行器—传感器，用于执行快速的局部修正；而机器人底座则作为一个较慢、高惯量、长程的身体，通过重新定位来维持运动并扩大可达工作空间。在操作中，机械臂执行即时的修正响应，随后基座调整整个身体，以保持平滑、连续的运动。

具体而言，末端执行器应在外部力旋量 $\mathcal{W} \in \mathbb{R}^6$的作用下，跟随期望的线速度$\mathbf{v}_{ee}^{w} \in \mathbb{R}^3$ 和角速度 $\boldsymbol{\omega}_{ee}^{w} \in \mathbb{R}^3$，同时保持从第 III-E 节的参考调节器获得的参考位姿$\mathbf{x}_{ee}^{'w} \in SE(3)$和参考力旋量 $\mathcal{W}' \in \mathbb{R}^6$ 。我们在世界坐标系中将此行为建模为：

$$\mathcal{W} - \mathcal{W}' = \mathbf{K} \log(\mathbf{x}_{ee}^{'w\top} \mathbf{x}_{ee}^{w}) + \mathbf{D} \begin{bmatrix} \mathbf{v}_{ee}^{w} \\ \boldsymbol{\omega}_{ee}^{w} \end{bmatrix} \tag{1}$$

其中$\mathbf{K}$ 和$\mathbf{D}$ 分别表示刚度和阻尼系数，它们是对角矩阵，维度为 $\mathbb{R}^{6\times6}$ ；$\mathbf{x}_{ee}^{w} \in SE(3)$ 是末端执行器的当前位姿。术语 $\log(\mathbf{x}_{ee}^{'w\top}\mathbf{x}_{ee}^{w}) \in \mathbb{R}^6$ 表示期望$\mathbf{x}_{ee}^{'w}$ 与当前$\mathbf{x}_{ee}^{w}$ 末端执行器位姿之间的线性与角度差。

在一个人类与机器人协作移动桌子的示例中（图1底部），假设 z 轴指向上方，则末端执行器在 xy 平面内应具有柔顺性，而在 z 方向和旋转方面应具有更高的刚度。

B. 全身导纳控制

我们考虑安装在腕部、位于末端执行器与机械臂之间的力/力矩传感器，如图6所示。传感器数据[$\mathbf{F}_{snr} \in \mathbb{R}^3, \boldsymbol{\tau}_{snr} \in \mathbb{R}^3$] 以力/扭矩传感器坐标系表示。利用公式1，我们得到末端执行器的目标线速度和角速度：

$$\begin{bmatrix} \mathbf{v}_{ee}^{w} \\ \boldsymbol{\omega}_{ee}^{w} \end{bmatrix} = \mathbf{D}^{-1} \left(\begin{bmatrix} \mathbf{R}_{ft}^{w} \mathbf{F}_{snr} \\ \mathbf{R}_{ft}^{w} \boldsymbol{\tau}_{snr} \end{bmatrix} - \mathcal{W}' - \mathcal{W}_G - \mathbf{K} \log(\mathbf{x}_{ee}^{'w\top} \mathbf{x}_{ee}^{w}) \right) \tag{2}$$

其中， $\mathcal{W}_G \in \mathbb{R}^6$ 是末端执行器重力所产生的力和力矩，$\mathbf{R}_{ft}^{w} \in \mathbb{R}^{3\times3}$ 是从世界坐标系到力/力矩传感器坐标系的旋转矩阵，用于定义式（1）中的 $\mathcal{W}$ 。

世界坐标系中的末端执行器速度 $[\mathbf{v}_{ee}^{w}, \boldsymbol{\omega}_{ee}^{w}]$ 由基座速度 $[\mathbf{v}_{b}^{w}, \boldsymbol{\omega}_{b}^{w}]$与末端执行器相对于基座的相对速度 $[\mathbf{v}_{ee}^{b}, \boldsymbol{\omega}_{ee}^{b}]$共同构成：

$$\begin{aligned} \mathbf{v}_{ee}^{w} &= \mathbf{v}_{b}^{w} + \dot{\mathbf{R}}_{b}^{w} \mathbf{x}_{ee}^{b} + \mathbf{R}_{b}^{w} \mathbf{v}_{ee}^{b} \\ \boldsymbol{\omega}_{ee}^{w} &= \boldsymbol{\omega}_{b}^{w} + \mathbf{R}_{b}^{w} \boldsymbol{\omega}_{ee}^{b} \end{aligned} \tag{3}$$

其中 $\mathbf{R}_{b}^{w}$ 是从世界坐标系到机器人基座坐标系的旋转矩阵。我们指出，公式3是通过对从世界坐标系到末端执行器坐标系的末端执行器位置和姿态求导得出的，其中包括机械臂本身和基座的速度贡献。随后，我们将从基座坐标系到末端执行器坐标系的期望速度 $[\mathbf{v}_{ee}^{b}, \boldsymbol{\omega}_{ee}^{b}]$ 映射为关节

空间：

$$\dot{\mathbf{q}}_{arm} = \mathbf{J}^{\#} \begin{bmatrix} \mathbf{v}_{ee}^{b} \\ \boldsymbol{\omega}_{ee}^{b} \end{bmatrix} \tag{4}$$

其中 $\mathbf{J}^{\#} = \mathbf{J}^{\top}\left(\mathbf{J}\mathbf{J}^{\top} + \lambda^{2}\mathbf{I}\right)^{-1}$ 是机械臂雅可比矩阵 $\mathbf{J}$ 的阻尼伪逆，λ 是阻尼因子。将公式3代入公式4，得到期望的机械臂关节速度：

$$\dot{\mathbf{q}}_{arm} = \mathbf{J}^{\#}\mathbf{R}_{w}^{b} \begin{bmatrix} \mathbf{v}_{ee}^{w} - \mathbf{v}_{b}^{w} - \dot{\mathbf{R}}_{b}^{w}\mathbf{x}_{ee}^{b} \\ \boldsymbol{\omega}_{ee}^{w} - \boldsymbol{\omega}_{b}^{w} \end{bmatrix} \tag{5}$$

其中，$[\mathbf{v}_{ee}^{w},\ \boldsymbol{\omega}_{ee}^{w}]$ 根据公式 2，$[\mathbf{v}_{b}^{w},\ \boldsymbol{\omega}_{b}^{w}]$ 利用第 III-D 节所述的卡尔曼滤波器进行估计，$[\mathbf{R}_{b}^{w},\ \dot{\mathbf{R}}_{b}^{w}]$ 由安装在机器人基座上的陀螺仪测得，而$\mathbf{x}_{ee}^{b}$ 则通过机械臂的正向运动学计算得出。

C. 用于基座运动的强化学习策略

我们训练强化学习策略，以控制机器人的基座，并使其与第 III-B 节所述的机械臂导纳控制器协同工作。在最初为使用 Go1 足式机器人进行纯运动而开发的 MuJoCo Playground 近端策略优化设置 [21], 基础上，我们将其改编为配备机械臂的 Go2机器人，并修改目标，使其跟踪机械臂末端执行器速度，而非基座速度。超参数、底层网络和训练步数与 [21]大致相同。图1底部展示了在仿真中评估我们策略的示例：机器人与人类协同搬运一张桌子。观测空间包括：1）投影到机体坐标系中的重力；2）腿部关节位置和速度；3）基座的线速度和角速度；4）前一个动作；以及5）从基于模型的导纳控制器接收的期望末端执行器速度 $[\mathbf{v}_{ee}^{w},\ \boldsymbol{\omega}_{ee}^{w}]$，该控制器在强化学习训练过程的每个时间步长运行。按照 MuJoCo Playground 的设置 [21],，动作空间由腿部关节位置指令$\mathbf{q}_{\text{leg}} \in \mathbb{R}^{12}$ 构成，这些指令经过缩放后与默认关节位置$\mathbf{q}_{\text{default}} \in \mathbb{R}^{12}$相加：

$$\mathbf{q}_{\text{leg}} = \mathbf{q}_{\text{default}} + k_{a}\boldsymbol{a} \tag{6}$$

其中，k_a 是网络 $\boldsymbol{\alpha}$原始输出的动作缩放系数。奖励是多个项的加权和，详见表I。主要奖励旨在鼓励机器人跟踪期望的末端执行器速度。除运动跟踪外，奖励还包含促进能效的项，例如对关节扭矩和关节速度的惩罚。最后，我们还使用若干塑形奖励和惩罚项来引导运动行为的特定方面，例如鼓励正确的摆腿并减少足部滑移。这些组成部分有助于策略生成目标明确、高效且符合物理规律的动作。需要注意的是，在训练和评估期间，我们都不会向策略提供 MuJoCo 中的基座真实速度。相反，我们通过将模拟器中的理想值替换为根据第 III-D 节所述状态估计流程得到的含噪估计值，对实际的状态估计不确定性进行建模，从而使状态估计更接近机器人可能接收到的结果。具体而言，在每个时间步长 t，我们取 MuJoCo 中的基座原始速度值，并添加零均值高斯噪声 $\boldsymbol{\eta}_t \sim \mathcal{N}(\mathbf{0},\ \mathbf{P}_t)$，其中$\mathbf{P}_t$是状态估计器在更新步骤中得到的状态协方差矩阵。该采样流程保留了估计状态的方差和交叉相关性，确保学习到的策略经历真实的速度估计噪声，而非依赖模拟器提供的完美信息。

表I 奖励函数

奖励	**表达式**
速度跟踪	
末端执行器线速度跟踪	$\phi(\mathbf{v}_{ee}^{\prime w} - \mathbf{v}_{ee}^{w})$
末端执行器角速度跟踪	$\phi(\boldsymbol{\omega}_{ee}^{\prime w} - \boldsymbol{\omega}_{ee}^{w})$
节能	
关节力矩（惩罚项）	$\lVert\boldsymbol{u}\rVert + \lvert\boldsymbol{u}\rvert$
动作变化率（惩罚项）	$\lVert\mathbf{a}_t - \mathbf{a}_{t-1}\rVert^2$
能耗（惩罚项）	$\left\lvert\dot{\boldsymbol{q}}_{\text{leg}}\right\rvert^{\top}\lvert\boldsymbol{u}\rvert$
行为	
姿态偏差（惩罚）	$\phi(\boldsymbol{q}_{\text{leg}} - \boldsymbol{q}_{\text{default}})$
静止站立（惩罚）	$\left\lvert\boldsymbol{q}_{\text{leg}} - \boldsymbol{q}_{\text{default}}\right\rvert \cdot \mathbf{1}_{\lVert\mathbf{v}_{ee}^{w}\rVert<0.01}$
Z轴线速度（惩罚）	$v_{b,z}^2$
XY平面角速度（惩罚）	$\omega_{b,x}^2 + \omega_{b,y}^2$
足部腾空时间	$(\mathbf{t}_{\text{air}} - 0.1)^{\top}\mathbf{c}^{\text{first}} \cdot \mathbf{1}_{\lVert\mathbf{v}_{ee}^{w}\rVert>0.01}$
摆动间隙（惩罚）	$\sum_i \lvert h_i - h_{\max}\rvert \cdot \lVert\mathbf{v}_i\rVert_{xy}^2 \cdot c_i^{\text{first}} \cdot \mathbf{1}_{\lVert\mathbf{v}_{ee}^{w}\rVert>0.01}$
摆动高度（惩罚）	$\sum_i \left(\frac{h_i^{\text{峰值}} - h^{\text{最大值}}}{\quad}\right)$
足部滑移（惩罚）	$\sum_i c_i \lVert\mathbf{v}_{i,xy}\rVert^2 \cdot \mathbf{1}_{\lVert\mathbf{v}_{ee}^{w}\rVert>0.01}$
终止（惩罚）	$1_{\mathbf{z}_b \cdot \mathbf{z}_w < 0}$

其中，$\phi(x) = \exp(-\lVert x\rVert^2/\sigma)$，$\sigma$ 是一个可调参数，用于控制奖励对跟踪误差的敏感程度。$\boldsymbol{q}_{\text{leg}}$、$\dot{\boldsymbol{q}}_{\text{leg}}$、$\boldsymbol{u} \in \mathbb{R}^{12}$ 分别表示腿部关节的位置、速度和力矩。v_b、z、ω_b、x、ω_b、y 分别表示基座沿 z 轴的线速度、沿 x 轴的角速度以及沿 y 轴的角速度。$\mathbf{v}_i$、xy 表示地面 xy 平面内第 i 个足部的速度。$\mathbf{a}_t \in \mathbb{R}^{12}$ 表示强化学习策略在时间步长 t输出的动作向量，对应于缩放后的腿部关节位置指令或 $\mathbf{a}_t = k_a\boldsymbol{\alpha}_t$。$\mathbf{z}_b$和 $\mathbf{z}_w$ 分别表示基座和世界坐标系的 z 轴。$\mathbf{t}_{\text{air}} \in \mathbb{R}^4$ 包含四个足部的腾空阶段持续时间。$c_i \in \{0,1\}$ 表示第 i 个足部是否与地面接触。$c_i^{\text{first}} \in \{0,1\}$ 表示第 i 个足部此前是否处于腾空状态并在当前帧与地面接触，即 $\mathbf{c}^{\text{first}}$ 仅在足部触地时激活相关项。h_i 表示第 i 个足部距地面的距离，$h_{\max}$ 是指定最大距离的超参数，而 h_i^{peak} 表示第 i 个足部在当前摆动阶段达到的最大高度。为明确符号含义，在强化学习训练的语境中，$\mathbf{v}_{ee}^{\prime w}$、$\boldsymbol{\omega}_{ee}^{\prime w}$表示期望末端执行器速度，即导纳控制器（或式 2）的输出；而 $\mathbf{v}_{ee}^{w}$、$\boldsymbol{\omega}_{ee}^{w}$ 表示当前速度。

D. 速度估计与不确定性

足式机器人运动会产生不必要的振荡噪声，尤其会影响机器人基座线速度和角速度的估计，表示为 $\hat{\mathbf{x}} = [\mathbf{v}_{b}^{w},\ \boldsymbol{\omega}_{b}^{w}]^{\top} \in \mathbb{R}^6$（不要与末端执行器位置的表示法$\mathbf{x}_{ee}^{w}$混淆）。由于计算末端执行器的线速度和角速度 $[\mathbf{v}_{ee}^{w}, \boldsymbol{\omega}_{ee}^{w}]$需要用到这些量，因此基座速度估计中的误差会直接影响导纳控制器的性能。鉴于

系统的复杂性，我们利用神经网络根据动作捕捉数据学习系统动力学，并将真实值作为参照，再将所得模型整合到卡尔曼滤波器中，以改善状态估计。具体而言，神经网络 $\mathrm{NN}_\theta(\phi_t)$ 经参数化后，用于估计在时间步长 t处动力学模型与真实值之间的差异，其中误差定义为 $\boldsymbol{\epsilon}_t$。在这方面， $\boldsymbol{\epsilon}_t$ 具有明确的物理数学含义，即 $\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t^\top = \mathbf{Q}_t$，也就是过程噪声的协方差。因此，我们将卡尔曼滤波器的预测步骤写为：

$$\hat{\mathbf{x}}_{t+1}^{\text{model}} = \text{Dyn}(\hat{\mathbf{x}}_t, \mathbf{f}_t) + \mathbf{w}_t \tag{7}$$

$$\mathbf{y}_t = h(\hat{\mathbf{x}}_t) + \mathbf{v}_t \tag{8}$$

$$\boldsymbol{\epsilon}_t = \text{NN}_\theta(\boldsymbol{\phi}_t) \tag{9}$$

$$\hat{\mathbf{x}}_{t+1|t}^{\text{model+NN}} = \hat{\mathbf{x}}_{t+1|t}^{\text{model}} + \boldsymbol{\epsilon}_t \tag{10}$$

$$\mathbf{Q}_t = \boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t^\top \tag{11}$$

$$\mathbf{P}_{t+1|t} = \mathbf{A}_d\mathbf{P}_{t|t}\mathbf{A}_d^\top + \mathbf{Q}_t \tag{12}$$

以及测量更新步骤：

$$\mathbf{S}_t = \mathbf{C}_t\mathbf{P}_{t+1|t}\mathbf{C}_t^\top + \mathbf{R}_t \tag{13}$$

$$\mathbf{K}_t = \mathbf{P}_{t+1|t}\mathbf{C}_t^\top\mathbf{S}_t^{-1} \tag{14}$$

$$\tilde{\mathbf{y}}_t = \mathbf{y}_t - \mathbf{C}_t\hat{\mathbf{x}}_{t+1|t}^{\text{model+NN}} \tag{15}$$

$$\hat{\mathbf{x}}_{t+1|t+1} = \hat{\mathbf{x}}_{t+1|t} + \mathbf{K}_t\tilde{\mathbf{y}}_t \tag{16}$$

$$\mathbf{P}_{t+1|t+1} = \mathbf{P}_{t+1|t} - \mathbf{K}_t\mathbf{S}_t\mathbf{K}_t^\top \tag{17}$$

其中，$\hat{\mathbf{x}}_{t+1|t}^{\text{model}}$ 是系统模型 Dyn(·) 给出的状态；$\hat{\mathbf{x}}_{t+1|t}^{\text{model+NN}}$ 是神经网络 $\mathrm{NN}_\theta(\phi_t)$ 进行模型校正后的状态；$\mathbf{C}_t$ 是观测矩阵；$\mathbf{S}_t$ 是新息协方差；$\mathbf{K}_t$ 是卡尔曼增益；$\mathbf{P}_t$ 是状态估计协方差；$\mathbf{y}_t$ 是测量更新值。$\mathbf{A}_d$ 是根据我们的系统动力学 Dyn(·) 推导出的离散时间线性化：

$$\text{Dyn}(\mathbf{x}_k, \mathbf{f}_k) = \begin{bmatrix} \boldsymbol{\omega}_k + \Delta T\left(\sum_{i=1}^n \hat{\mathbf{I}}^{-1}\left(\mathbf{p}_k^i \times \mathbf{f}_k^i\right)\right) \\ \mathbf{v}_k + \Delta T\left(\sum_{i=1}^n \frac{1}{m}\mathbf{f}_k^i + \mathbf{g}\right) \end{bmatrix} \tag{18}$$

其中，$\mathbf{f}_k^i \in \mathbb{R}^3$、$\mathbf{p}_k^i \in \mathbb{R}^3$ 分别表示世界坐标系中第 i 条腿的地面反作用力和落足位置，$\hat{\mathbf{I}} \in \mathbb{R}^{3\times3}$ 是惯性张量的斜对称矩阵，ΔT 是离散化的时间步长，$\mathbf{g} \in \mathbb{R}^3$ 是常量重力向量。

对于测量模型 $h(\hat{\mathbf{x}}_t)$，我们使用根据惯性测量单元测量值得到的机器人基座角速度。我们还根据与地面接触的足部估计机器人基座线速度 $\mathbf{v}_{\text{base}} \in \mathbb{R}^3$ ：

$$\mathbf{v}_{\text{base}} = -\frac{1}{n_c}\sum_{i=1}^{n}\mathbf{R}_b^w\left(\dot{\mathbf{p}}^i + \boldsymbol{\omega}_{\text{imu}} \times \mathbf{p}^i\right) \tag{19}$$

其中，$\boldsymbol{\omega}_{\text{imu}} = [\omega_x, \omega_y, \omega_z]^\top$ 是惯性测量单元角速度向量，n_c 是与地面接触的足部数量，而 $\dot{\mathbf{p}}^i$ 是世界坐标系中的足部速度。

用于预测模型误差的网络采用简单的双层前馈架构，每层包含 256 个隐藏单元，并使用 ReLU 激活函数。训练目标是最小化预测修正值与

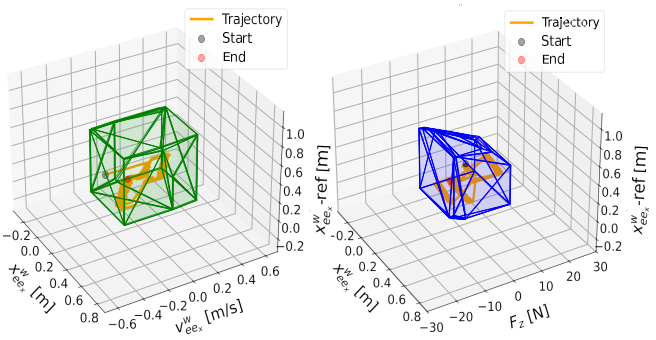


图 3. 图 5 所示示例轨迹的最大输出可容许集（MOAS）。

$\boldsymbol{\epsilon}_i^{\text{predicted}}$ 真实模型误差 $\boldsymbol{\epsilon}_i^{\text{true}}$ 之间的均方误差，损失函数如下：

$$\mathcal{L}(\theta) = \frac{1}{N}\sum_{i=1}^{N}\left\|\boldsymbol{\epsilon}_i^{\text{predicted}} - \boldsymbol{\epsilon}_i^{\text{true}}\right\|^2 \tag{20}$$

尽管模型误差网络仅使用真实硬件数据进行训练，但其预测的修正量会在仿真中的强化学习策略训练期间纳入其中。具体而言，我们使用卡尔曼滤波器的协方差（如第 III-D 节所述）向观测值注入真实噪声，从而鼓励策略学习对真实世界噪声的鲁棒性。

最后，为了提高平滑性，我们对卡尔曼滤波器更新后的状态应用平滑系数为 λ 的低通滤波器：

$$\hat{\mathbf{x}}_t^{\text{filtered}} = \lambda\hat{\mathbf{x}}_t + (1-\lambda)\hat{\mathbf{x}}_{t-1}^{\text{filtered}} \tag{21}$$

E. 使用参考调节器提供安全保证

为了确保式（1）的安全性，我们采用参考调节器（RG），通过渐近调整参考信号，确保系统在用户定义的约束内运行。如 [7], 中详细所述，该方法需要计算最大输出可容许集（MOAS），$\mathcal{O}_\infty$，该集合定义了系统能够在不违反约束的情况下运行的所有初始状态和参考信号：

$$\mathcal{O}_\infty = \{(x_0, v) \mid y(t) \in \mathcal{Y} \text{ for all } t \geq 0\} \tag{22}$$

其中 x_0 是系统的初始状态，v 是参考输入，$y(t)$是系统输出，而 $\mathcal{Y}$ 表示满足代表安全性和性能要求的约束条件的允许输出集合。实际上，MOAS 的计算是一个迭代过程，用于验证给定状态和参考输入在足够长的时间范围内是否始终处于可容许边界内，从而确保收敛至平衡点邻域。该过程首先初始化位置、速度、力和力矩、运动学约束、控制增益以及离散化参数。在位置、速度、参考位置、力和力矩以及参考力和力矩上生成网格，以对轨迹进行采样。对于每种网格组合，算法使用控制速度，在离散时间步上模拟系统

$[\mathbf{v}_{ee}^{w}$、 $\boldsymbol{\omega}_{ee}^{w}]$ ，通过欧拉积分更新位置。如果轨迹过程中违反了任何位置、速度、力和力矩或运动学约束，则从集合中剔除该网格点。否则，将相应的值添加到 $\mathcal{O}_\infty$ 中。最终的 $\mathcal{O}_\infty$ 由所有满足全部约束且可受控的初始条件组成。在我们的案例中，$\mathcal{O}_\infty$ 沿位置、速度和力的 x、y和 z 轴分别独立计算，并包含绕夹爪基座法向轴的力矩。

一旦获得 $\mathcal{O}_\infty$ ，即可用它调整当前状态下的输入参考设定点，从而确保存在满足无限时域内输出约束的可行控制序列，即$\mathbf{x}_{ee}^{'w}, \mathcal{W}^{'} = \mathcal{O}_\infty($ o $o = [\ _{ee}^{w}\ \ _{ee}^{'w}\ , \ _{ee}^{w}\ \mathcal{W}\ \mathcal{W}^{'}\ ]$ $_{ee}^{'w}$ $W^{'}$),

其中$\mathbf{x}$,$\mathbf{x}$,ref$\mathbf{v}$,,ref和$\mathbf{x}$,refref 分别为期望的末端执行器参考位置和力和力矩。请注意，如果根据 $\mathcal{O}_\infty$， o 不会导致约束违反，那么$\mathbf{x}_{ee}^{'w}, \mathcal{W}^{'} = \mathbf{x}_{ee,\text{ref}}^{'w}\ \mathcal{W}^{'}$ ref，即参考信号保持不变。为实现高效的实时可行性检查，我们将 $\mathcal{O}_\infty$ 的离散样本存储在*KD*树中，以便快速执行最近邻查询。给定查询状态 o_q，我们判断 $o_q \in \mathcal{O}_\infty$ ，或找到最近的可行状态$o^* \in \mathcal{O}_\infty$:

$$o^* = \arg\min_{o \in \mathcal{O}_\infty} \|o - o_q\| \tag{23}$$

如果由于模型误差或显著扰动导致 $o \notin \mathcal{O}_\infty$ 或式 23 不可行，则保持$\mathbf{x}_{ee}^{w}$,$\mathbf{v}_{ee}^{w}, \mathcal{W}$ 不变，并求解最接近的$\mathbf{x}_{ee}^{*w}$ 和 $\mathcal{W}^*$，使系统回到可行集，此时$\mathbf{x}_{ee}^{'w}, \mathcal{W}^{'} = \mathbf{x}_{ee}^{*w}, \mathcal{W}^*$。

表 II 结合神经网络的卡尔曼滤波器：估计误差统计

指标	$\boldsymbol{\omega_x}$ (弧度/秒)	$\boldsymbol{\omega_y}$ (弧度/秒)	$\boldsymbol{\omega_z}$ (弧度/秒)
MAE	0.0031	0.0013	0.0009
RMSE	0.0035	0.0017	0.0013
STD	0.0018	0.0015	0.0013
指标	$\boldsymbol{v_x}$ (米/秒)	$\boldsymbol{v_y}$ (米/秒)	$\boldsymbol{v_z}$ (米/秒)
MAE	0.0385	0.0235	0.0202
RMSE	0.0564	0.0297	0.0273
STD	0.0511	0.0297	0.0272

IV. 实验验证

我们在一台 Unitree Go2 四足机器人上验证了该框架。该机器人配备 D1 六自由度伺服机械臂，且机械臂腕部安装了 BOTA[22]力/力矩传感器（图6）。运动强化学习控制器以 50 Hz 运行，机械臂的导纳控制器以 40 Hz 运行（D1 机械臂电机的收发频率受限于 40 Hz），状态估计器则以 500 Hz 运行。由于参考调节器（RG）的计算（即最大控制不变集的计算）已离线完成，在线部署速度很快：我们只需通过*KD*树 查询系统是否处于安全集内，并据此调整参考信号。我们展示了力控制器在人机协作任务中响应外部力和力矩的能力：机器人与一名人类共同扶着一块木板，朝不同方向移动，包括越过障碍物，如图所示图1。在该任务中，机器人仅根据人类施加在末端执行器上的推拉力旋量，生成适当的线速度和角速度，以跟随人类的动作。通过将参考调节器输入状态中的参考力旋量设为零，我们实现了这类人机协作任务。请注意，我们不对末端执行器在 x、 y和 z 位置施加位置参考，但会对姿态参考施加约束：夹爪只允许绕夹爪基座法向轴旋转。

在超过 25 秒的运动操作过程中，图 4 展示了末端执行器在 x 和 y 轴上的指令速度或交互驱动线速度跟踪能力（左图）、在 z 轴上的线速度和角速度跟踪能力（中图），以及在 x 和 y 轴上的角速度跟踪能力（右图）。指令速度是根据腕部力/力矩传感器测得的力旋量，通过式（2）中的力控制器计算得出的。总体而言，在所有 x、 y和 z 轴上，线速度分量的均方误差为 $\leq$ 0.005 米^2/秒^2，角速度分量的均方误差为 $\leq$ 0.029 弧度^2/秒^2 。结果表明，系统在全部 6 个自由度上均能实现准确的力跟踪。图 4 还显示，在 15.0 秒至 20.0 秒之间， z 角速度的误差较大，且在跟踪 x 和 y 轴上的角速度时存在额外的延迟和噪声。需要指出的是，这些误差可能来自多种因素，包括执行器和传感器误差、外部扰动，以及同时跟踪全部 6 个自由度的难度。这些误差可能直接受到导纳控制器增益调节的影响，也可能是由于必须优先考虑某些方向分量而非其他分量所致。

我们通过参考调节器（RG）确保力控制器的稳定性和安全性。必要时，参考调节器会将期望参考信号修改为新的参考信号，以确保约束满足和控制稳定性。图 3 展示了 MOAS：左侧为位置、速度和参考位置，右侧为位置、力和参考位置。通过一个示例轨迹（详见图 5），我们可以看到，轨迹的起点和终点均处于 MOAS 范围内，表明在所有时间步长上都满足约束且具有递归可行性。在图 5 中，我们还比较了在有无参考调节器的情况下，跟踪期望轨迹 $x_{ee_x}^{w}$ 并采用期望的零参考力 F_z （以实现柔顺性）的效果。对于位置跟踪（上方子图），如果不修改规定的参考信号，在 2.5 s 至 5.0 s 以及 5.0 s 至 7.5 s 这两个时段内，控制器可能会超出位置约束限值。这显然是由于参考位置在 5.0 s 和 7.5 s 时快速变化，导致这些时间步长上的速度出现尖峰，从而违反约束（子图中部）。我们还展示了（子图下方），控制器能够成功地将力降回零，并通过适当的速度指令纠正力扰动。

在表 II 中，我们使用平均绝对误差（MAE）、均方根误差（RMSE）和标准差（STD）作为指标，验证了第 III-D 节中的状态估计器对线速度和角速度的估计，并将估计状态与动作捕捉系统提供的真实值进行比较。

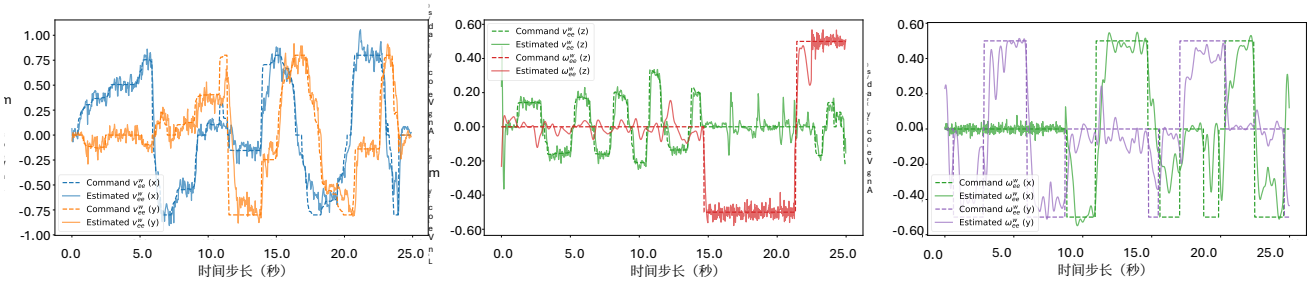


图 4. 人机协作任务期间的速度跟踪结果。左图显示命令末端执行器在 x/y 方向上的线速度跟踪结果，中图显示在 z 方向上的线速度和角速度跟踪结果，右图显示在 x/y 方向上的角速度跟踪结果。命令线速度和角速度由输出导纳控制器直接生成，该控制器以 6 自由度力/力矩信息作为输入。

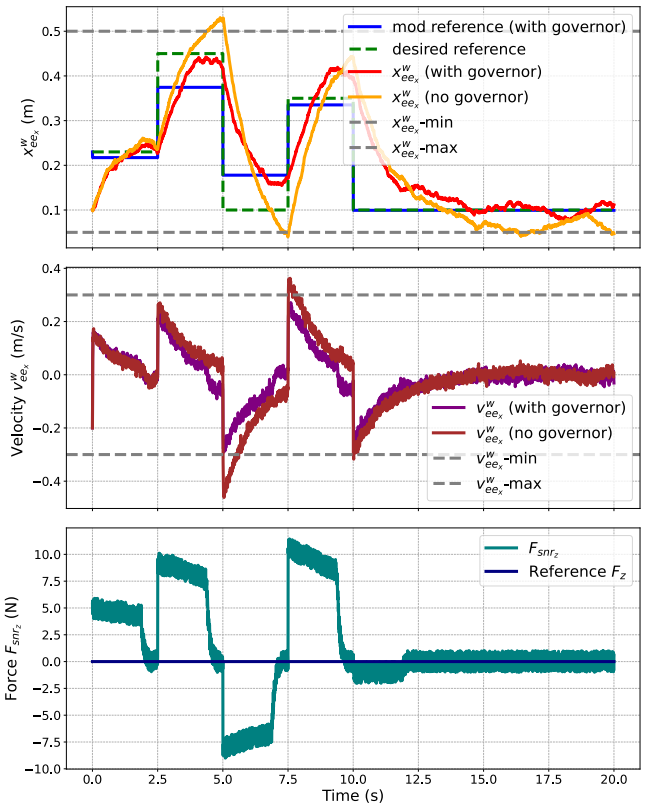


图 5. 展示参考调节器对安全性和约束满足影响的示例轨迹。

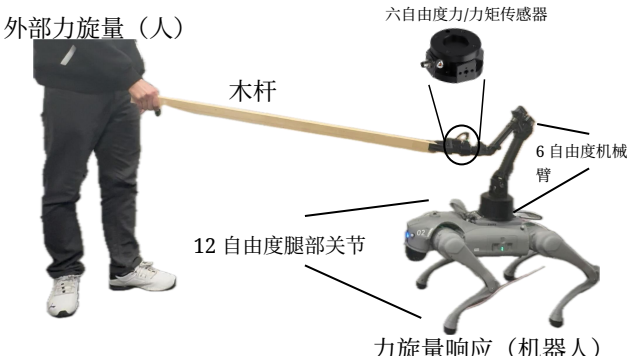


图6. 硬件验证设置。

V. 结论

总而言之，我们成功演示了通过基于模型和学习的框架实现力跟踪移动操作。该框架能够接受用户施加的力旋量，并作出适当的线速度和角速度响应；线速度分量和角速度分量的均方误差（MSE）分别为 ≤ 0.005 $\mathrm{m}^2/\mathrm{s}^2$ 和≤ 0.029 $\mathrm{rad}^2/\mathrm{s}^2$ 。通过参考调节器保证安全性，展现了我们针对机械臂的基于模型的方法的优势；参考调节器确保输出的末端执行器位置和速度始终处于可行集内，并满足位置、速度和运动学约束。最后，为了补偿复杂动力系统状态估计中的误差，对于腿式机械臂机器人，我们在卡尔曼滤波器中使用了神经网络，因为速度估计的改善有助于提升力控制器的性能。未来的工作将致力于同时改善所有自由度的跟踪，尤其是力矩跟踪；使用机器人抓取各种不同物体的数据训练我们的神经卡尔曼滤波器，以考虑重量分布的变化；并在人机交互任务中加入视觉信息。

参考文献

[1] L. Sentis 和 O. Khatib，“通过对行为基元进行分层控制来综合全身行为”，《国际人形机器人杂志》，第 02 卷，第 4 期，第 505–518 页，2005 年。[2] M. Liu、Z. Chen、X. Cheng、Y. Ji、R. Qiu、R. Yang 和 X. Wang，“腿式运动操作的视觉全身控制”，第八届机器人学习会议，2024 年。[3] G. Pan、Q. Ben、Z. Yuan、G. Jiang、Y. Ji、S. Li、J. Pang、H. Liu 和 H. Xu，“Roboduet：通过跨本体部署实现全身腿式运动操作”，*arXiv* 预印本 *arXiv:2403.17367*，2024 年。[4] Z. Fu、X. Cheng 和 D. Pathak，“深度全身控制：学习用于操作和运动的统一策略”，载于机器人学习会议（*CoRL*），2022 年。[5] T. Portela、G. B. Margolis、Y. Ji 和 P. Agrawal，“学习腿式操作的力控制”，载于*2024*年*IEEE*机器人与自动化国际会议（*ICRA*），第 15366–15372 页，2024 年。[6] P. Zhi 、 P. Li 、 J. Yin 、 B. Jia 和 S. Huang，“学习腿式运动操作的统一力与位置控制”，*arXiv* 预印本 *arXiv:2505.20829*，2025 年。[7] E. Garone、S. Di Cairano 和 I. Kolmanovsky，“具有约束的系统的参考与指令调节器：理论与应用综述”，《自动化》，第 75 卷，第 306–328 页，2017 年。[8] A. Schperberg、S. Tsuei、S. Soatto 和 D. Hong，“Saber：用于异构机器人自主导航的数据驱动运动规划器”，*IEEE*机器人与自动化快报，第 6 卷，8093，2021 4，8086 no. pp. –。

[9] J. Zhang、N. Gireesh、J. Wang、X. Fang、C. Xu、W. Chen、L. Dai 和 H. Wang，“Gamma：基于在线抓取位姿融合的抓取能力感知移动操作策略学习”，载于*2024*年*IEEE*机器人与自动化国际会议（*ICRA*），第 1399–1405 页，2024 年。[10] N. Yokoyama、A. Clegg、J. Truong、E. Undersander、T.-Y. Yang、S. Arnaud、S. Ha、D. Batra 和 A. Rai，“Asc：面向机器人移动操作的自适应技能协调”，*IEEE*机器人与自动化快报，第 9 卷，第 1 期，第 779–786 页，2024 年。[11] E. Arcari、M. V. Minniti、A. Scampicchio、A. Carron、F. Farshidian、M. Hutter 和 M. N. Zeilinger，“用于机器人移动操作的贝叶斯多任务学习模型预测控制”，*IEEE*机器人与自动化快报，第 8 卷，第 6 期，第 3222–3229 页，2023 年。[12] M. Posa、C. Cantu 和 R. Tedrake，“一种通过接触对刚体轨迹进行优化的直接方法”，《国际机器人研究杂志》，第 33 卷，第 1 期，第 69–81 页，2014 年。[13] M. P. Murphy、B. Stephens、Y. Abe 和 A. A. Rizzi，“高自由度动态操作”，载于《无人系统技术 *XIV*》（R. E. Karlsen、D. W. Gage、C. M. Shoemaker 和 G. R. Gerhart 编），第 8387 卷，第 83870V 页，国际光学与光子学会，SPIE，2012 年。[14] M. Murooka、S. Nozawa、Y. Kakiuchi、K. Okada 和 M. Inaba，“面向人形机器人的大型重物体全身推动操作及接触姿态规划”，载于*2015*年*IEEE*国际机器人与自动化会议（*ICRA*），第 5682–5689 页，2015 年。[15] C. D. Bellicoso、K. Kr¨amer、M. St¨auble、D. Sako、F. Jenelten、M. Bjelonic 和 M. Hutter，“ALMA——面向可控扭矩机器人的关节式运动与操作”，载于*2019*年机器人与自动化国际会议 *(ICRA)*，第 8477–8483 页，2019 年。[16] S. Zimmermann、R. Poranne 和 S. Coros，“Go fetch!——使用波士顿动力Spot与外部机械臂进行动态抓取”，载于*2021*年*IEEE*机器人与自动化国际会议 *(ICRA)*，第 4488–4494 页，2021 年。[17] Y. Ji、G. B. Margolis 和 P. Agrawal，“Dribblebot：野外动态足式操作”，载于*2023*年*IEEE*机器人与自动化国际会议 *(ICRA)*，第 5155–5162 页，2023 年。[18] Y. Ma、F. Farshidian 和 M. Hutter，“学习手臂辅助的跌倒损伤减轻与恢复，以实现足式移动操作机器人”，载于*2023*年*IEEE*机器人与自动化国际会议 *(ICRA)*，第 12149–12155 页，2023 年。[19] T. Haarnoja 等，“利用深度强化学习为双足机器人学习敏捷足球技能”，《style id='13'>《科学·机器人学》，第 9 卷，第 89 期，文章编号 eadi8022，2024 年。[20] A. Schperberg、Y. Shirai、X. Lin、Y. Tanaka 和 D. Hong，“使用无迹卡尔曼滤波器的富接触机器人系统自适应力控制器”，载于*IEEE-RAS*第*22*届人形机器人国际会议 *(Humanoids)*，第 1–8 页，IEEE，2023 年。[21] K. Zakka、B. Tabanpour、Q. Liao、M. Haiderbhai、S. Holt、J. Y. Luo、A. Allshire、E. Frey、K. Sreenath、L. A. Kahrs、等，“MuJoCo Playground”，*arXiv*预印本 *arXiv:2502.08844*，2025 年。[22] Bota Systems，“力/扭矩传感器。” https://www.botasys. com/，2025 年。访问日期：2025年7月21日。