

Learning Quadrupedal Locomotion for a Heavy Hydraulic Robot Using an Actuator Model

Minho Lee¹, Hyeonseok Kim², Jin Tak Kim³, Sangshin Park³, Jeong Hyun Lee¹, Jungsan Cho³, and Jemin Hwangbo*,¹

Abstract—The simulation-to-reality (sim-to-real) transfer of large-scale hydraulic robots presents a significant challenge in robotics because of the inherent slow control response and complex fluid dynamics. The complex dynamics result from the multiple interconnected cylinder structure and the difference in fluid rates of the cylinders. These characteristics complicate detailed simulation for all joints, making it unsuitable for reinforcement learning (RL) applications. In this work, we propose an analytical actuator model driven by hydraulic dynamics to represent the complicated actuators. The model predicts joint torques for all 12 actuators in under 1 microsecond, allowing rapid processing in RL environments. We compare our model with neural network-based actuator models and demonstrate the advantages of our model in data-limited scenarios. The locomotion policy trained in RL with our model is deployed on a hydraulic quadruped robot, which is over 300 kg. This work is the first demonstration of a successful transfer of stable and robust command-tracking locomotion with RL on a heavy hydraulic quadruped robot, demonstrating advanced sim-to-real transferability.

Index Terms—Hydraulic/Pneumatic Actuators, Legged Robots, Reinforcement Learning

I. INTRODUCTION

Quadruped robots have been widely used for navigating through various obstacles and complex terrains [1]–[7]. They have demonstrated stability [4] and terrain adaptability [1], [6]–[8], showing their potential for various applications.

Many hydraulic quadruped robots have been operated with model-based control [4], [5], [9]–[11]. Model-based control is a control strategy that relies on a mathematical model of the system’s dynamics. The model predicts the robot’s future behavior and optimizes the control inputs accordingly, enabling effective decision-making and real-time adaptation. For example, the hydraulic quadruped robot HyQ [10] utilized model-based control to achieve stable locomotion, demonstrating the effectiveness of model-based control in

This work was supported by the Korea Research Institute for Defense Technology planning and advancement(KRIT) grant funded by the Korean government(DAPA(Defense Acquisition Program Administration)). (No. 20-107-C00-007-02(KRIT-CT-22-001), Development of Walking-Driving Hybrid Locomotion Control for Multi-Legged Robot Platform, 2024)

* corresponding author

¹ Robotics and Artificial Intelligence Lab, KAIST, Daejeon, South Korea

² Robotics Team, HYUNDAI Rotem, Uiwang, Gyeonggi-do, South Korea

³ Hydraulic Robot Laboratory, Human-Centric Robotics R&D Department, Korea Institute of Industrial Technology, Ansan, Gyeonggi-do, South Korea

myno1126@kaist.ac.kr,
hyeonseok0427@hyundai-rotem.co.kr,
jintagi@kitech.re.kr, pss@kitech.re.kr,
josualee@kaist.ac.kr, chojs@kitech.re.kr,
jhwangbo@kaist.ac.kr

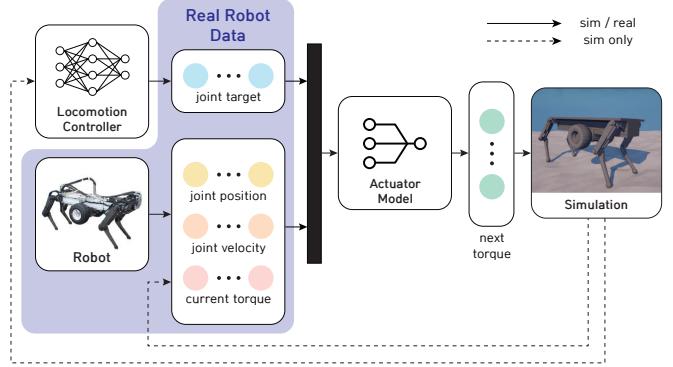


Fig. 1. Actuator model framework

such systems [12]. Model-based control provides accurate and reliable control using an accurate system dynamics model. However, it can be computationally intensive and less adaptable to dynamic or unpredictable environments.

Recently, the control of quadruped robots has increasingly relied on reinforcement learning (RL) approaches [7], [13]–[15]. RL-based control offers greater flexibility, adapting more effectively to changing conditions and learning near-optimal policies without an explicit model, making it more robust in unpredictable scenarios [8], [16]. Despite the advantages of RL-based control, model-based control has been widely used in hydraulic robots [11] due to its ability to manage the complex dynamics of hydraulic systems. The low simulation-to-reality (sim-to-real) transferability in hydraulic systems makes it challenging to deploy RL-based controllers on real-world hardware platforms.

Sim-to-real transfer is considered more difficult in hydraulic systems than in electric motor systems because of the inherent slow response and the complex fluid dynamics of hydraulic systems [4], [9]–[11]. Hydraulic robots comprise multiple interconnected cylinders, making accurate simulation difficult due to fluid compressibility and variations in pressure delivery speed in real systems. Furthermore, the large size and heavy weight of the robot can cause undesirable bending and compression of mechanical parts, further complicating the sim-to-real transfer. Moreover, collecting data from such a system without a reliable controller is challenging, as accidents can result in serious damage to the robot.

While previous approaches have informed our work, our method differs in two key ways. First, we propose a simple analytical actuator model derived from hydraulic actuator dynamics, which facilitates sim-to-real transfer. This model

accurately captures the behavior of hydraulic systems with negligible computational cost($< 1\mu s$ for all 12 actuators simultaneously). Recent studies often rely on neural network-based actuator models [13], [17], [18]. While these models can provide accurate predictions within the domains they are trained on, they tend to generalize poorly to out-of-distribution scenarios. This poses a serious risk in large-scale hydraulic systems, where untested conditions during data collection can lead to hazardous failures. Moreover, collecting isolated actuator data is impractical for legged robots, as joint loads vary depending on configuration and motion history. These dependencies make it difficult to replicate realistic operating conditions outside the full robotic system. We capture the real-world dynamics of hydraulic actuators—including external loads and fluid effects—by collecting locomotion data directly from a quadruped robot. Building on this, we develop an analytical model capable of handling a broad range of operating conditions to support robust reinforcement learning training.

Also, we validate our approach on a fully hydraulic quadruped robot. Unlike prior studies that focus on single-actuator hardware demonstrations [17], [19], our work showcases the model’s effectiveness in a complete, real-world multi-actuator system. In this study, we make the following key contributions:

- Proposal of an analytical actuator model with accurate torque prediction, showing high sim-to-real transferability.
- Development of an RL-based locomotion controller for a hydraulic quadrupedal robot weighing over 300kg.
- Demonstration of hydraulic-powered RL locomotion at 1m/s, including adaptability to dynamic conditions.

To the best of our knowledge, this is the first demonstration of locomotion at 1 m/s on a hydraulic quadruped robot weighing over 300 kg using an RL-based locomotion controller.

II. RELATED WORKS

RL-based control approaches have recently gained widespread adoption in the quadruped robot community. Previous work [13] demonstrated the successful use of RL for learning agile and dynamic motor skills in legged robots, achieving high performance in complex tasks such as walking, running, and recovering from falls. More recent developments [20] adopted the RL-trained policies to traverse various terrains, enhancing the sim-to-real transferability and improving robustness in quadruped locomotion. However, it is challenging to apply RL directly to hydraulic quadruped robots due to the difficulty in sim-to-real transferability in hydraulic systems.

Several attempts have been made to improve the sim-to-real transferability between theoretical estimations and actual hydraulic systems [9], [11], [17]–[19], [21]. One line of work [19] proposed a nonlinear adaptive robust control strategy for a hydraulic actuator to estimate system parameters and handle nonlinearities. However, this approach is heavily dependent on the accuracy of the system model and is



Fig. 2. A hybrid-locomotion platform combined with wheeled & quadruped system

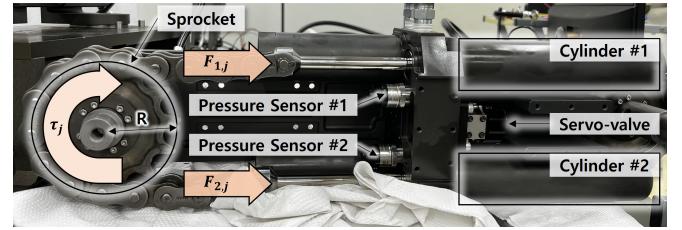


Fig. 3. Structure of the hydraulic actuator, where two hydraulic cylinders function as a single unit to generate torque on a revolute joint. The illustration details how chains transfer force from the hydraulic cylinders to the joint. Each leg has three such actuators (roll, pitch, and knee), all following the same structural design.

computationally demanding. Another work [17] presented a neural network inverse model of a hydraulic actuator to track a force trajectory accurately. However, their method was limited to a single actuator system and was verified only in a disturbance-free environment. Recent advancements [18] introduced adaptive online neural predictive control for hydraulic actuators, demonstrating low tracking error and robustness to model uncertainties. However, the use of neural networks requires a large amount of high-quality training data, and real-time application remains challenging due to their high computational cost.

III. METHODS

A. Hydraulic quadruped robot

Our robot is a hybrid locomotion platform used to validate the proposed actuator model and the locomotion performance of a policy trained with the actuator model. Our robot weighs over 300kg and has a length exceeding 1.8 meters. It is a ground vehicle developed to navigate complex terrains while carrying heavy supplies. It features a four-legged design and a two-wheel differential drive system, as shown in Fig. 2. The robot can handle various environments, driving at high speeds on flat ground and transitioning to walking when driving is impractical. The leg mechanism comprises a 3-degree-of-freedom hydraulically actuated joint. Each rotary joint integrates a sprocket and chain mechanism, as shown in Fig. 3, which addresses torque nonlinearity issues while having a wide range of motion. Each leg has three revolute actuators—roll, pitch, and knee joint—all of which share the

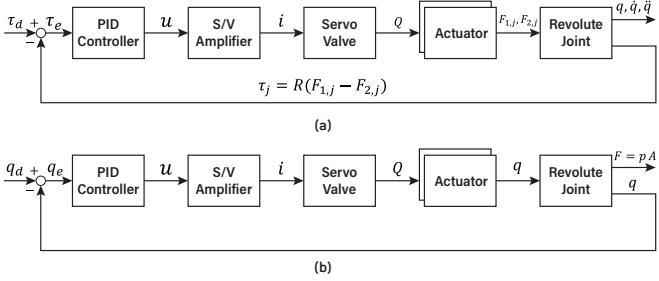


Fig. 4. Control diagram of the joint with (a) torque PID control and (b) position PID control

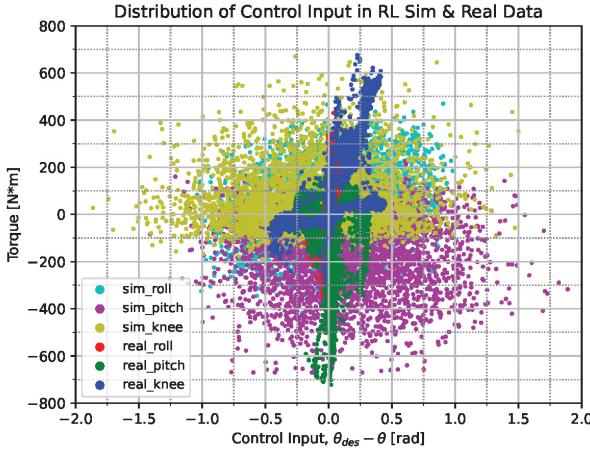


Fig. 5. Data distribution of simulation and real data

same structural design. The torque for each joint is controlled by adjusting the pressures and displacements within two hydraulic cylinders arranged in a bidirectional pulling structure, which together form a single revolute actuator. A total of 24 hydraulic cylinders control the movement, all of which are connected to a single pump that supplies the fluid from outside the robot. In this study, we focus on the quadrupedal locomotion of the robot. All experiments are conducted with the driving system deactivated.

B. Hydraulic actuator control

Each joint operates under two distinct low-level control modes: torque and position PID control, each suited for different tasks. The block diagrams for both controls are shown in Fig. 4, where u is the servo valve input, i is the electric current, Q is the flow rate, and q is the current position of the joint.

Torque PID control adjusts hydraulic pressure to achieve a target torque, making it computationally efficient and suitable for simulations where torque output is assumed to match the target. However, real-world discrepancies arise due to measurement noise and delays, reducing reliability in dynamic tasks like locomotion. Position PID control regulates actuator position for greater stability and precision in real-world applications, as it relies on accurate encoders. While less sensitive to measurement errors, it requires a detailed actuator model for effective simulation, making implementation more complex.

We address these challenges by proposing an actuator model that simplifies hydraulic actuator dynamics, enabling efficient simulation of position control with minimal computational overhead. Our actuator model predicts torque output based on its physical state, allowing rapid and realistic simulation of position control. This solution bridges the gap between simulation and real-world performance, enabling effective RL-based training without complex actuator modeling.

C. Actuator model

Previous work [17] proposed a first-order differential equation for a single hydraulic cylinder as

$$\dot{f} = g(x)\dot{x} + h(f, x)x_s \quad (1)$$

$$\dot{x}_s = \theta(x_s) + \psi(x_s)u \quad (2)$$

where,

$$g(x) = -A^2\beta \left(\frac{1}{V_{0A} + Ax} + \frac{1}{V_{0B} + A(L - x)} \right)$$

$$h(f, x) = C_d w \beta A \sqrt{\frac{P_S - P_T}{\rho}} - \text{sgn}(x_s) \frac{f}{\rho A}$$

$$\cdot \left(\frac{1}{V_{0A} + Ax} + \frac{1}{V_{0B} + A(L - x)} \right).$$

Here, f is the force, x is the position of the cylinder, x_s is the valve opening, u is the control input (i.e., the reference position), $\theta(x_s)$, and $\psi(x_s)$ are nonlinear functions related to the valve's model and controller. The function $\text{sgn}(x_s)$ is the sign function, L is the cylinder stroke, A is the cylinder area, w is the area gradient, β is the bulk modulus of the hydraulic fluid, ρ is the oil density, V_{0A} and V_{0B} are the dead volume of each cylinder channel, C_d is the discharge coefficient of the valve, P_S is the supply pressure, and P_T is the return pressure. By rewriting (1) into delta notations, the following equation can be obtained:

$$\Delta f = g(x)\Delta x + h(f, x)x_s\Delta t. \quad (3)$$

However, (3) is formulated for a single cylinder system, which does not account for the dynamic interactions present in a multi-cylinder environment. Direct application of (3) to systems with multiple interconnected cylinders, as in our robot, can lead to significant errors due to hydraulic coupling effects. In our case, multiple actuators share a common hydraulic circuit, where fluid compressibility and varying pressure delivery speeds cause discrepancies in the return pressures between the cylinders. Nonlinearity and the number of state variables also require high computational costs, which makes it challenging to apply in real-time simulations with 400 parallel environments. Additionally, in environments where forces are applied in a consistent direction, resistive forces and large impacts could be neglected. In contrast, the direction of force application continuously changes during the locomotion of our robot, making resistive forces and dynamic impacts non-negligible factors that require additional consideration.

To address these limitations, we extend the actuator model beyond the single-cylinder formulation (3) by introducing

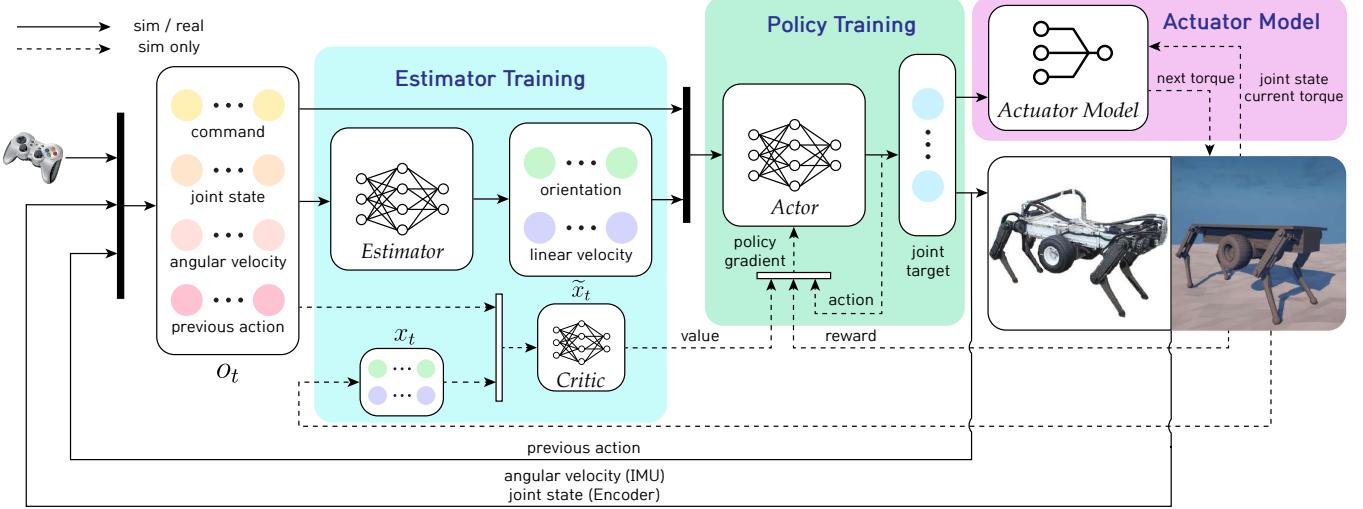


Fig. 6. Overall concurrent training framework

additional terms, including k_3 to model resistive forces and k_4 to account for corrective forces during sudden impacts. Furthermore, we simplify the original formulation while preserving its ability to capture essential hydraulic dynamics, thereby enhancing its practicality for real-world applications. Specifically, we propose the following model:

$$\begin{aligned} \Delta f = & k_1 \Delta x - k_2 f - k_3 \dot{x} \\ & + k_4 \Delta x \cdot \max(-f \cdot \text{sgn}(\Delta x), 0). \end{aligned} \quad (4)$$

In (3), $g(x)$ represents the position-dependent mechanical leverage effect as a function of x . However, its variation remains minimal within the robot's operating range, which can be approximated as a constant k_1 . $h(f, x)x_s dt$ represents the force-dependent pressure drop across the hydraulic system. f is scaled 1000 times larger than x . The applied force primarily influences this pressure drop, while other factors remain constant or vary negligibly. Thus, it can be approximated as a negative linear function of force, with k_2 capturing the system-dependent hydraulic characteristics. This simplification eliminates the need to solve nonlinear differential equations, substantially reducing computational cost while maintaining sufficient accuracy for RL training.

The term $k_3 \dot{x}$ is introduced to model the velocity-dependent resistive force, which was negligible in (1) when rapid changes in the direction of force application were minimal. However, in our experiments, the actuator's movement direction changes continuously, making fluid resistance a significant factor. Therefore, incorporating this resistive term is essential to accurately capture the actuator's real-world behavior.

We introduce the corrective term k_4 to account for sudden impact scenarios. The control input is conflicted by the physical dynamics of the actuator when the direction of motion is opposite to the current force. This happens when the actuator is attempting to resist an excessive external load from an impact, yet it is forced to move in a direction opposite to the force. The system's control objective is misaligned with its actual physical response, resulting in

discrepancies between the intended control and the resulting motion in the simulation without appropriate correction. Our robot experiences high external loads and reaction forces, and these effects are not adequately accounted for in the previous model (1). To address this issue, we introduce a corrective force that adjusts the control input in these scenarios. This term prevents excessive reductions in control force, ensuring that the model's responses remain consistent with the system's actual dynamics, even in the presence of large external disturbances or impacts.

The derivation of the corrective term begins with a reconsideration of the fundamental hydraulic pressure dynamics governing the actuator. The actuator force f can be represented as $f = p \cdot A$ with the pressure p and the effective actuator area A . The hydraulic pressure dynamics within the actuator chamber, subject to fluid compressibility and instantaneous displacement demands, is given by:

$$\frac{dp}{dt} = \frac{B}{V} (Q_{in} - A\dot{x}). \quad (5)$$

B denotes the fluid bulk modulus, V is the actuator chamber volume, Q_{in} is the volumetric inflow. In sudden impact scenarios, a rapid displacement command Δx leads to an instantaneous and significant inflow demand, described by $Q_{in} \approx A\Delta x/\Delta t$.

We hypothesize that under impact conditions, the commanded displacement rate $\Delta x/\Delta t$ can be considered significantly larger than the actual actuator velocity \dot{x} . The actuator piston motion is constrained by flow capacity and oil viscosity, which limit its acceleration under sudden commands. Consequently, while the commanded velocity reflects the instantaneous displacement demand, the actual piston velocity remains much smaller at the initial moment of impact. Hence, the second term can be regarded as negligible, leading to the approximation $\Delta p \approx (BA/V)\Delta x$.

Furthermore, we assume that the pressure variation caused by an impact should scale with the applied impact force. To capture this effect, we introduce a dimensionless normalization with respect to the maximum actuator force,

thereby hypothesizing that the corrective term can represent the nonlinear and transient nature of hydraulic responses under sudden impact conditions. The resulting equation is expressed as:

$$\Delta f_{\text{imp}} = \frac{BA^2}{Vf_{\max}} f \Delta x. \quad (6)$$

The final adjusted term, designed to activate under actual impact conditions (i.e., when $f \cdot \Delta x < 0$), is given by $k_4 \Delta x \cdot \max(-f \cdot \text{sgn}(\Delta x), 0)$, as shown in (4). This corrective force plays a critical role in obtaining a stable RL policy. During RL exploration in simulation, the robot frequently encounters situations involving sudden impacts. The RL actor undergoes various attempts and experiences numerous falls throughout the process of learning locomotion policies, resulting in frequent and substantial impacts. This is shown in Fig. 5, where simulation data exhibits a broader distribution compared to real data, particularly when the torque and control input are in opposite directions. If the actuator model is constructed without properly accounting for impacts, it will predict entirely incorrect values when encountering these situations, thereby hindering the RL training process. Additionally, our analytical model can estimate k_4 even with a small amount of data, whereas learning-based methods rely on large and diverse input data. Since curve fitting requires significantly fewer data points than neural network training, our model remains reliable despite lower occurrences of impact cases.

By applying $\Delta f = f_{\text{next}} - f$ and the relationships $\tau = Rf$, $\Delta x = x_{\text{des}} - x$, $\dot{x} = R\dot{q}$, and $x = Rq$, where R is the radius of the sprocket, the final expression of the actuator model is obtained from (4) as:

$$\begin{aligned} \tau_{\text{next}} = & k_1 R^2 (q_{\text{des}} - q) + (1 - k_2)\tau - k_3 R^2 \dot{q} \\ & + k_4 R(q_{\text{des}} - q) \cdot \max(-\tau \cdot \text{sgn}(q_{\text{des}} - q), 0). \end{aligned} \quad (7)$$

Therefore, (7) shows that our model can predict τ_{next} corresponding to the current torque τ , current joint state q and \dot{q} , and the target position q_{des} .

The coefficients of the actuator model from (7) are obtained by fitting the data from real-world robot operations. We collected data from operating the robot for 20 seconds with a locomotion policy trained without an actuator model. The locomotion policy was operated at a frequency of 100Hz, while joint positions, target positions, joint velocities and torques were recorded at a higher frequency of 1000Hz. The high-frequency data collection allowed us to capture detailed and precise information about the actuator dynamics in real-robot operations.

We implemented three baseline neural network architectures—MLP, LSTM, and GRU—under the same dataset and training conditions as the proposed actuator model to ensure a fair comparison. Each network was configured with size (48, 64, 12), and hyperparameters such as hidden layer size were tuned to achieve the best accuracy. Training was performed for 1,000 iterations until convergence, providing sufficient optimization.

TABLE I
GLOBAL REWARDS

Reward	Expression
Command	$r_v = k_{\text{cmd}} \{ \exp(-\ \text{cmd}_{xy} - v_{xy}\ ^2) \cdot (1 + \exp(-0.5\ \text{cmd}_{xy} - v_{xy}\)) + \exp(-1.5(\text{cmd}_z - \omega_z)^2) \}$
Yaw Command Error	$r_{\text{yaw}} = \begin{cases} 10c_f k_{\text{yaw}} (\text{cmd}_z - \omega_z)^2, & \text{if } \omega_z = 0, \\ c_f k_{\text{yaw}} (\text{cmd}_z - \omega_z)^2, & \text{else.} \end{cases}$
Base Height	$r_h = k_h \exp(-40 h_0 - h)$
Base Motion	$r_m = k_m (v_z^2 + 0.02(\omega_x + \omega_y))$
Torque	$r_{\tau} = c_f k_{\tau} \ \tau\ ^2$
Torque Clip	$r_{\tau_{\text{clip}}} = c_f k_{\tau_{\text{clip}}} \ \tau_{\text{clip}}\ ^2$
Nominal Position	$r_q = \begin{cases} 10c_f k_q \ q_t - q_{\text{nom}}\ , & \text{if } \text{cmd} = 0, \\ c_f k_q \ q_t - q_{\text{nom}}\ , & \text{else.} \end{cases}$
Joint Velocity	$r_{\dot{q}} = c_f k_{\dot{q}} \ \dot{q}_t\ ^2$
Joint Acceleration	$r_{\ddot{q}} = c_f k_{\ddot{q}} \ \ddot{q}_t - \ddot{q}_{t-1}\ ^2$
Action smoothness	$r_s = c_f k_s (0.5\ q_t^{\text{des}} - 2q_{t-1}^{\text{des}} + q_{t-2}^{\text{des}}\ ^2 + \ q_t^{\text{des}} - q_{t-1}^{\text{des}}\ ^2)$

TABLE II
LOCAL REWARDS

Reward	Expression
Flight Phase	$r_{\text{fl}} = c_f k_{\text{fl}} \text{ (if } \forall i, \text{contact}_i = \text{False)}$
Airtime	$r_{\text{air}, i} = \begin{cases} k_a \min(\max(T_{s,i} - T_{a,i}, -0.25), 0.25) & \text{if } \ \text{cmd}\ = 0, \\ k_a \min(T_{a,i}, 0.2) & \text{if } T_{a,i} < 0.25, \\ k_a \min(T_{s,i}, 0.2) & \text{if } T_{s,i} < 0.25, \\ 0 & \text{else.} \end{cases}$
Foot Slip	$r_{\text{slip}, i} = c_f k_{\text{slip}} \ v_{\text{foot}, i, xy}\ ^2 \text{ (if contact}_i = \text{True)}$
Foot Clearance 1	$r_{\text{cl1}, i} = k_{\text{cl1}} \ v_{\text{foot}, i}\ (h_i - h_{\text{tar}})^2 \text{ (if contact}_i = \text{False \& } \ \text{cmd}\ > 0)$
Foot Clearance 2	$r_{\text{cl2}, i} = k_{\text{cl2}} r_{\text{cl1}, t-1, i} \text{ (if contact}_i = \text{True)}$
GRF Smoothness	$r_{\text{grf}, i} = c_f k_{\text{grf}} \{ 0.5(f_{t,i}^{\text{grf}} - 2f_{t-1,i}^{\text{grf}} + f_{t-2,i}^{\text{grf}})^2 + (f_{t,i}^{\text{grf}} - f_{t-1,i}^{\text{grf}})^2 \}$
Action Clip	$r_{\text{act}, j} = c_f k_{\text{act}} q_{t,j}^{\text{des, clip}} $
Joint Limit	$r_{l,j} = c_f k_l \text{ (if } q_{t,j} > \text{limit)}$

D. Estimator and policy training with actuator model

The locomotion policy of the quadruped robot is trained with RL in 400 parallel environments with flat terrain. The initial state is assigned randomly for each environment, and curriculum learning is employed with factor c_f to facilitate smooth state transitions. Concurrent training [15] is introduced to train a robust locomotion policy and the state estimator for our robot, as the overall training framework is shown in Fig. 6.

Reward functions and coefficients are applied as shown in Table I and II. Global rewards are calculated for the whole robot, and local rewards are calculated for each joint or foot. The total reward is calculated by summing all the global and local rewards. The simulation terminates during training when the robot makes contact with the ground, except for its feet and shanks. Observation and kinematics randomization are applied to be robust against disturbances from real-world experiments.

TABLE III
RMSE AND MAPE LOSS FOR ACTUATOR MODEL AND MLP, LSTM, GRU

Models	Command 0.4 m/s			Command 1.0 m/s (1)			Command 1.0 m/s (2)			Command 1.0 m/s (3)			Command 1.0 m/s (4)		
	RMSE	τ_j > 50		RMSE	τ_j > 50		RMSE	τ_j > 50		RMSE	τ_j > 50		RMSE	τ_j > 50	
		RMSE	MAPE [%]		RMSE	MAPE [%]		RMSE	MAPE [%]		RMSE	MAPE [%]		RMSE	MAPE [%]
Actuator model (Ours)	7.45	8.06	4.22	8.44	9.15	4.77	7.90	8.67	4.48	10.6	11.4	5.65	10.8	11.7	5.72
MLP	36.0	41.7	20.9	47.1	53.5	27.3	38.9	44.8	22.8	57.5	64.4	33.2	57.7	63.8	32.6
LSTM	33.3	37.2	19.8	41.1	46.5	23.8	35.3	39.7	20.6	50.1	55.8	28.2	51.2	56.9	28.7
GRU	27.4	30.7	15.6	34.4	38.3	19.9	28.9	32.2	16.4	41.8	46.1	24.0	42.1	45.9	24.0

The RL training network consists of three parts: actor, critic, and estimator. All three networks are designed using MLP. The estimator estimates the current robot state x_t from the observation o_t , and the actor receives the estimated robot state \tilde{x}_t with o_t and outputs an action a_t , which is the desired joint position. The observation state o_t consists of

$$o_t = (\omega, q_t, \dot{q}_t, q_{t-1}^{des}, cmd) \quad (8)$$

where ω stands for the base angular velocity, q_t and \dot{q}_t are the joint states of time t , q_{t-1}^{des} is the previous action which is the desired joint position from the previous time step, and cmd is the velocity command. The critic outputs a value from o_t and x_t , and the policy is trained with PPO. The estimator is trained by supervised learning with x_t and \tilde{x}_t . The locomotion policy and the estimator are operated at 100Hz simultaneously, while the simulation and the position controller in the real robot are operated at 1000Hz. The training is conducted in Raisim [22] using an AMD Ryzen Threadripper 3990X and NVidia RTX 3090, with 20,000 iterations taking around 10 hours.

IV. RESULTS

A. Evaluation of actuator model

We compared the proposed actuator model's prediction results with the baseline neural networks (MLP, LSTM, and GRU) under the conditions described in Section III-C. In total, five real-robot locomotion experiments were conducted. One trial was performed with a forward command of 0.4 m/s, while the remaining four were conducted at a forward command of 1.0 m/s. During these experiments, the locomotion policy generated target positions at 100Hz, and o_t were recorded at 1000Hz. We compared the predicted torques with the measured torques at each timestep using the collected data. RMS and Mean Absolute Percentage (MAP) errors with a threshold of 50 N·m for the actual torque values were calculated. This threshold is introduced because the error becomes disproportionately more prominent in the MAPE metric when the actual torque values are small. Table III summarizes the results for RMS and MAP errors, whereas Fig. 7 also shows that our model has the highest accuracy in predicting the torque values. Our model outperforms the trained neural networks in terms of torque prediction accuracy across all experimental data, demonstrating its performance in sim-to-real transfer. This accuracy enables the RL-based controller to operate reliably across a broader range of conditions, ensuring stable and robust locomotion. Unlike neural networks, which struggle with out-of-distribution data,

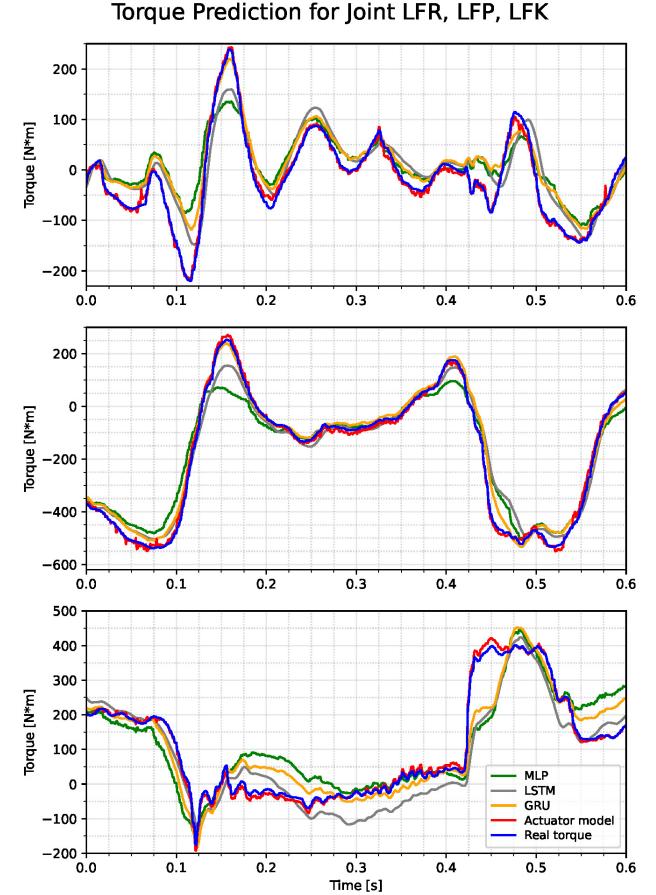


Fig. 7. Torque prediction of the actuator model of joint LFR, LFP, LFK

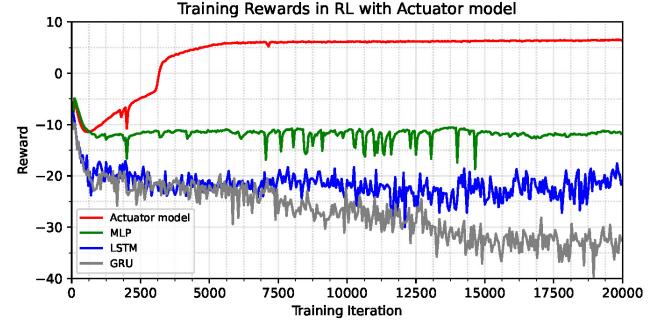


Fig. 8. Reward shown according to the training iteration.

our actuator model provides consistent torque predictions even in dynamic environments, allowing the RL policy to adapt to disturbances and variations in terrain.

In addition, we implemented the proposed actuator model

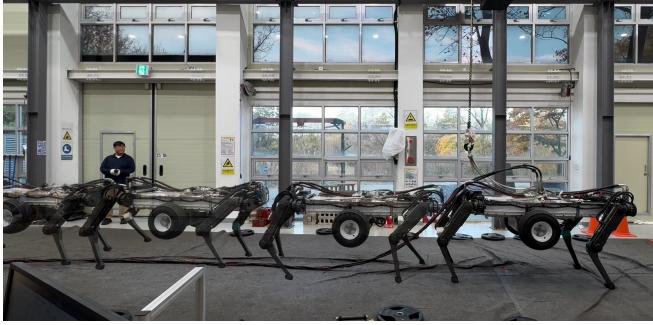


Fig. 9. Locomotion results with 1.0 m/s command, trained with our actuator model.

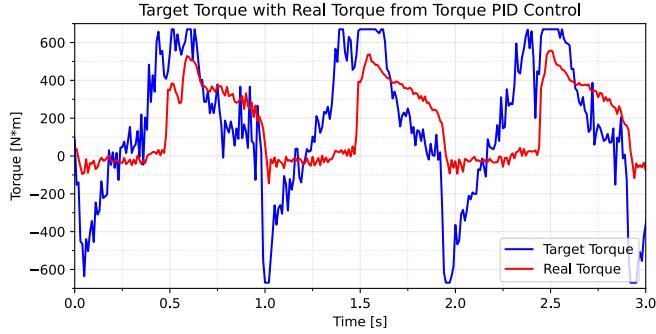


Fig. 10. Torque control results, actual vs target torque with torque control and baseline models(MLP, LSTM, GRU) in the simulation and trained locomotion policies with RL. The reward values recorded during training are shown in Fig. 8. Notably, our actuator model outperformed the exploration process in RL, as the latter failed to achieve appropriate reward values and did not converge to a stable result.

The proposed model's performance arises from its ability to generalize beyond the narrow operational range of the training data. Although the dataset for the actuator model was limited, our model successfully predicted the corresponding torques in unstable situations. The baseline models were also trained with the same data but faced difficulties, often failing to generate reasonable outputs in unstable situations and consequently breaking down during RL training. To learn dynamic locomotion successfully, the actuator model must handle exceptional cases such as falls, where large impact forces frequently occur. Unlike the neural networks trained solely on standard walking data, the proposed actuator model incorporates the k_4 term, which applies corrective forces under impact conditions based on the actuator's physical state. This mechanism enabled the model to remain accurate in out-of-distribution scenarios and effectively respond to dynamic disturbances, thereby ensuring robustness during training and real-world applications. Our model's performance demonstrated its ability to generalize beyond the training data and handle complex, non-ideal situations, which is critical for robust quadruped robot locomotion.

We also compared the computational cost of training. The proposed actuator model required 10 hours to train the locomotion policy with RL, whereas the baseline models required between 36 and 48 hours under the same training

conditions. This result highlights that our model achieves superior accuracy while being more than three times faster to train, thereby offering advantages from both performance and efficiency perspectives.

B. Evaluation of the locomotion performance

We compared the robot's locomotion performance with the trained policies using three different control approaches. These included torque PID control trained in RL without an actuator model, position PID control trained in RL with stiff PD estimation instead of an actuator model, and position PID control trained in RL with the proposed actuator model. All experiments were conducted indoors on a flat floor, where the robot walked about 30 meters in a straight trajectory to provide a controlled evaluation environment for the trained policies. The robot's walking performance was evaluated in two main areas: command tracking and locomotion stability. The supplementary material includes detailed experimental videos that illustrate these approaches.

1) Command Tracking: The robot was tested with tracking commands of 0.4 m/s and 1.0 m/s. The torque control demonstrated satisfactory performance at a speed of 0.4 m/s, with the command being tracked effectively. However, as the command speed increased, torque control struggled to keep the torque output in sync with the target torque, as shown in Fig. 10, causing the robot to lose balance and tip forward. The robot struggled to follow the target torque in torque control, resulting in failures during dynamic locomotion. In contrast, the position controls followed the commands up to 1.0 m/s, demonstrating improved robustness to higher speeds. Notably, the locomotion controller trained with the actuator model showed sufficient capability, as in Fig. 9, to increase the command speed beyond 1.0 m/s. This highlights the advantage of using an accurate actuator model, which not only enhances locomotion stability but also enables smoother control transitions, even in untrained conditions. However, this could not be thoroughly tested due to spatial constraints in the experimental setup and limitations related to the crane's speed, which was used to prevent the robot from tipping over. Nevertheless, it is expected that the locomotion controller trained with the actuator model would have been able to handle even higher command speeds under more favorable conditions.

2) Locomotion stability: The torque control exhibited significant issues with locomotion stability as the command speed increased, resulting in the robot failing to maintain stability. The actuator delays shown in Fig. 10 caused the robot to walk with pronounced impacts, resulting in noticeable vibrations and sounds. The robot frequently lost stability and fell when its feet encountered obstacles or the ground. The position control without the actuator model showed stability during locomotion but faced challenges due to its stiff nature of high-gain PD estimation. This stiffness restricted the control's ability to lift the legs sufficiently, especially when encountering obstacles such as uneven surfaces. This limitation resulted in instability and compromised locomotion performance in more complex environments, as

holding the foot slightly above the ground. Moreover, the stiff nature of the controller resulted in a less natural walking pattern, with sudden vibrations to the robot's body. However, the position control trained with the proposed actuator model demonstrated superior locomotion stability. The locomotion was natural and smooth, with the robot's feet reaching sufficient height for stable movements, as exactly shown inside the simulation environment. The learned policy demonstrated the ability to perform turning and lateral walking on flat terrain without requiring retraining, further supporting its robustness under the current training conditions. Overall locomotion was stable, and the robot exhibited more fluid and adaptive movements, even in dynamic conditions with sudden disturbances. This adaptivity was facilitated by the RL-based controller, which leveraged the accurate torque estimation of the actuator model to develop more robust control policies.

These results demonstrate that the proposed actuator model not only accurately simulates the behavior of hydraulic actuators but also enables the seamless transfer of simulation-based policies to real-world robotic systems. The stable and robust locomotion confirms the effectiveness of our actuator model under a wide range of conditions, both in simulation and real-world experiments.

V. CONCLUSION

This study demonstrated the effectiveness of the proposed actuator model by showing both accurate torque prediction and improvement of real-world locomotion. Our actuator model enabled the development of a robust RL-locomotion controller, demonstrating successful sim-to-real transfer in heavy hydraulic robots. The RL policy trained with the actuator model outperformed traditional methods, delivering smooth and stable locomotion both in simulation and on the real robot, even at higher speeds and under dynamic conditions.

However, the model's performances were not tested in scenarios that deviate significantly from the training data, such as abrupt directional changes or extreme external forces. Additionally, the long-term durability of the actuator model remains unassessed, and the experimental setup limited testing beyond 1.0 m/s.

Future work will focus on expanding real-world testing to include a broader range of conditions, such as higher speeds and diverse terrains, and refining the actuator model by incorporating factors such as actuator wear and hydraulic variations to ensure long-term reliability. We also aim to evaluate the model's adaptability to different robotic platforms and tasks, increasing its generalizability beyond the quadruped robot used in this study. In the long term, we aim to develop controllers that combine the robustness of RL with the interpretability of model-based approaches, contributing to more reliable deployment in real-world systems.

REFERENCES

- [1] A. W. Winkler, C. Mastalli, I. Havoutis, M. Focchi, D. G. Caldwell, and C. Semini, "Planning and execution of dynamic whole-body locomotion for a hydraulic quadruped on challenging terrain," in *2015 IEEE International Conference on Robotics and Automation (ICRA)*, 2015, pp. 5148–5154.
- [2] X. Rong, Y. Li, J. Ruan, and B. Li, "Design and simulation for a hydraulic actuated quadruped robot," *Journal of Mechanical Science and Technology*, vol. 26, no. 4, pp. 1171–1177, 2012.
- [3] S. Jeon, M. Jung, S. Choi, B. Kim, and J. Hwangbo, "Learning whole-body manipulation for quadrupedal robot," *IEEE Robotics and Automation Letters*, vol. 9, no. 1, pp. 699–706, 2024.
- [4] H. Khan, S. Kitano, M. Frigerio, M. Camurri, V. Barasol, R. Featherstone, D. G. Caldwell, and C. Semini, "Development of the lightweight hydraulic quadruped robot — minihyq," in *2015 IEEE International Conference on Technologies for Practical Robot Applications (TePRA)*, 2015, pp. 1–6.
- [5] S. Kuindersma, R. Deits, M. Fallon, A. Valenzuela, H. Dai, F. Permenter, T. Koolen, P. Marion, and R. Tedrake, "Optimization-based locomotion planning, estimation, and control design for the atlas humanoid robot," *Autonomous Robots*, vol. 40, no. 3, pp. 429–455, Mar 2016. [Online]. Available: <https://doi.org/10.1007/s10514-015-9479-3>
- [6] Y. Kim, J. H. Lee, C. Lee, J. Mun, D. Youm, J. Park, and J. Hwangbo, "Learning semantic traversability with egocentric video and automated annotation strategy," *IEEE Robotics and Automation Letters*, vol. 9, no. 11, pp. 10423–10430, 2024.
- [7] H. Kim, H. Oh, J. Park, Y. Kim, D. Youm, M. Jung, M. Lee, and J. Hwangbo, "High-speed control and navigation for quadrupedal robots on complex and discrete terrain," *Science Robotics*, vol. 10, no. 102, p. eads6192, 2025. [Online]. Available: <https://www.science.org/doi/abs/10.1126/scirobotics.ads6192>
- [8] 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. [Online]. Available: <https://www.science.org/doi/abs/10.1126/scirobotics.abc5986>
- [9] M. Sirospour and S. Salcudean, "Nonlinear control of hydraulic robots," *IEEE Transactions on Robotics and Automation*, vol. 17, no. 2, pp. 173–182, 2001.
- [10] C. Semini, N. G. Tsagarakis, B. Vanderborght, Y. Yang, and D. G. Caldwell, "Hyq - hydraulically actuated quadruped robot: Hopping leg prototype," in *2008 2nd IEEE RAS & EMBS International Conference on Biomedical Robotics and Biomechatronics*, 2008, pp. 593–599.
- [11] B. Cho, S.-W. Kim, S. Shin, J.-H. Oh, H.-S. Park, and H.-W. Park, "Energy-efficient hydraulic pump control for legged robots using model predictive control," *IEEE/ASME Transactions on Mechatronics*, vol. 28, no. 1, pp. 3–14, 2023.
- [12] B. Ugurlu, I. Havoutis, C. Semini, and D. G. Caldwell, "Dynamic trot-walking with the hydraulic quadruped robot—hyq: Analytical trajectory generation and active compliance control," in *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2013, pp. 6044–6051.
- [13] J. Hwangbo, J. Lee, A. Dosovitskiy, D. Bellicoso, V. Tsounis, V. Koltun, and M. Hutter, "Learning agile and dynamic motor skills for legged robots," *Science Robotics*, vol. 4, no. 26, p. eaau5872, 2019. [Online]. Available: <https://www.science.org/doi/abs/10.1126/scirobotics.aau5872>
- [14] V. Tsounis, M. Alge, J. Lee, F. Farshidian, and M. Hutter, "Deepgait: Planning and control of quadrupedal gaits using deep reinforcement learning," *IEEE Robotics and Automation Letters*, vol. 5, no. 2, pp. 3699–3706, 2020.
- [15] G. Ji, J. Mun, H. Kim, and J. Hwangbo, "Concurrent training of a control policy and a state estimator for dynamic and robust legged locomotion," *IEEE Robotics and Automation Letters*, vol. 7, no. 2, pp. 4630–4637, 2022.
- [16] 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. [Online]. Available: <https://www.science.org/doi/abs/10.1126/scirobotics.adc2256>
- [17] S.-W. Kim, B. Cho, S. Shin, J.-H. Oh, J. Hwangbo, and H.-W. Park, "Force control of a hydraulic actuator with a neural network inverse model," *IEEE Robotics and Automation Letters*, vol. 6, no. 2, pp. 2814–2821, 2021.
- [18] Y. Wang, C. Zhao, and Z. Shi, "Adaptive online neural predictive control of hydraulic actuator using fpga acceleration," *IEEE Transactions on Industrial Electronics*, vol. 71, no. 4, pp. 3924–3934, 2024.
- [19] C. Guan and S. Pan, "Nonlinear adaptive robust control of single-rod electro-hydraulic actuator with unknown nonlinear parameters," *IEEE*

Transactions on Control Systems Technology, vol. 16, no. 3, pp. 434–445, 2008.

- [20] A. Kumar, Z. Fu, D. Pathak, and J. Malik, “Rma: Rapid motor adaptation for legged robots,” 2021.
- [21] K. B. Tcheumchoua, S. Nam, and W. K. Chung, “Torque control of hydraulic pressure servo valve driven actuator with deep neural network,” in *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2022, pp. 12 512–12 519.
- [22] J. Hwangbo, J. Lee, and M. Hutter, “Per-contact iteration method for solving contact dynamics,” *IEEE Robotics and Automation Letters*, vol. 3, no. 2, pp. 895–902, 2018.