ATRos: Learning Energy-Efficient Agile Locomotion for Wheeled-legged Robots

Jingyuan Sun[1*], Hongyu Ji[2*†], Zihan Qu[3†], Chaoran Wang[4†], Mingyu Zhang[5†]

Abstract— Hybrid locomotion of wheeled-legged robots has recently attracted increasing attention due to their advantages of combining the agility of legged locomotion and the efficiency of wheeled motion. But along with expanded performance, the whole-body control of wheeled-legged robots remains challenging for hybrid locomotion. In this paper, we present *ATRos*, a reinforcement learning (RL)-based hybrid locomotion framework to achieve hybrid walking-driving motions on the wheeled-legged robot. Without giving predefined gait patterns, our planner aims to intelligently coordinate simultaneous wheel and leg movements, thereby achieving improved terrain adaptability and improved energy efficiency. Based on RL techniques, our approach constructs a prediction policy network that could estimate external environmental states from proprioceptive sensory information, and the outputs are then fed into an actor-critic network to produce optimal joint commands. The feasibility of the proposed framework is validated through both simulations and real-world experiments across diverse terrains, including flat ground, stairs, and grassy surfaces. The hybrid locomotion framework shows robust performance over various unseen terrains, highlighting its generalization capability. The details about real-world experiments can check (`https://baoziweiyuebing.github.io/ATRos-Wheeled_legged-robot-hybrid-locomotion/`)

I. INTRODUCTION

Robots capable of agile and robust locomotion are becoming increasingly important in real-world scenarios such as industrial inspection, warehouse automation, urban delivery, and disaster response. While legged robots have demonstrated remarkable progress in traversing unstructured terrains, they often suffer from relatively high energy consumption and limited efficiency during long-range navigation. Wheeled robots, on the other hand, offer high-speed and energy-efficient mobility on structured surfaces but lack the versatility to overcome obstacles and operate in irregular environments. Wheeled-legged robots have recently emerged as a promising solution, combining the best of both paradigms by enabling rolling on flat terrain and stepping when required to overcome obstacles. This hybrid locomotion capability not only increases operational versatility but also introduces new opportunities for energy-efficient, long-duration missions.

Despite recent progress, hybrid locomotion for wheeled–legged robots remains highly challenging. Compared with purely legged systems, wheeled–legged platforms must cope with additional dynamics introduced by continuous wheel rolling, nonholonomic constraints, and frequent transitions between rolling and stepping modes. These robots also present a distinct energy–agility trade-off: wheeled locomotion on flat or moderately uneven terrain is typically more energy-efficient, as continuous wheel rotation demands less power than repetitive leg lifting. In contrast, negotiating obstacles or climbing stairs requires legged stepping, which is substantially more energy-intensive, particularly for heavier platforms. This characteristic further complicates the design of hybrid locomotion frameworks, as it necessitates the coordination of simultaneous walking and driving behaviors while ensuring agility, versatility, and energy efficiency.

Recent advances in reinforcement learning (RL) have enabled robots to learn sophisticated locomotion strategies directly from trial-and-error interactions, bypassing the need for handcrafted gait design. RL has shown strong performance on quadrupedal robots, generating agile, adaptive, and robust locomotion behaviors. However, directly transferring such approaches to wheeled-legged robots is non-trivial due to the expanded action space, modified contact dynamics, and the need for explicit energy-awareness during training. Without careful adaptation, learned policies tend to favor either overly aggressive locomotion, which wastes energy, or overly conservative behaviors, which sacrifice agility.

To address these challenges, we present *ATRos*, a reinforcement learning framework for agile and transferable locomotion on wheeled-legged robots with a focus on energy efficiency. ATRos extends RL-based locomotion learning by incorporating rolling dynamics into the control architecture and introducing an energy-aware reward design that explicitly penalizes excessive power consumption while rewarding agile, stable motion. This enables the robot to discover hybrid locomotion strategies that achieve both efficiency and versatility, leveraging wheels for sustained rolling and legs for robust stepping when necessary.

In summary, the contributions of this paper are:
- We highlight the unique challenges of training locomotion policies for wheeled-legged robots compared to legged systems, focusing on the trade-off between agility and energy efficiency.
- We introduce ATRos, a reinforcement learning frame-

*These authors contributed equally to this work.

†This work was performed during an internship at Huawei.

[1]Jingyuan Sun is with Shanghai Huawei Technologies Co., Ltd., Shanghai 201799, China (corresponding author to provide phone: 18603680666; fax: none; e-mail: `sunjingyuan1@huawei.com`).

[2]Hongyu Ji is with the College of Future Information Technology, Fudan University, Shanghai 200433, China, `hyji21@m.fudan.edu.cn`

[3]Zihan Qu is with the Robotics Institute, Carnegie Mellon University, Pittsburgh, PA 15213, USA.

[4]Chaoran Wang is with the School of Aeronautic and Astronautics, Zhejiang University, Hangzhou 310027, China.

[5]Mingyu Zhang is with the School of Electronics and Communication Engineering, Sun Yat-sen University, Shenzhen 518107, China.

work that achieves locomotion control on wheeled-legged platforms using only proprioceptive information, with an energy-aware reward design.
- We validate ATRos in both simulation and preliminary real-world experiments on a wheeled-legged quadruped robot, demonstrating improved energy efficiency and agile locomotion.

II. RELATED WORK

A. Hybrid Locomotion of Wheeled-Legged Robots

Wheeled locomotion has been extensively studied in both laboratory and commercial applications, benefiting from its simplicity in control and high energy efficiency. In contrast, legged locomotion remains more challenging due to the high degrees of freedom and the indirect nature of base motion through intermittent foot-ground contacts, which has motivated heuristic, bio-inspired, and optimization-based strategies. Wheeled-legged robots aim to combine the efficiency of wheels with the versatility of legs. However, research on fully integrated walking-driving motion remains relatively limited. Early efforts such as DRC-HUBO+ [1] and RoboSimian [2] primarily exploited wheels, with legs playing only auxiliary roles. More advanced system likes ANYmal with wheels [3] demonstrated driving-stepping capabilities for stairs and obstacles, highlighting the potential of hybrid designs. More recently, the FLORES project [4] proposed a redesigned wheel-leg architecture coupled with a tailored RL controller, enabling smooth mode transitions and improved adaptability across diverse terrains. Despite these advances, designing controllers that can seamlessly coordinate rolling and stepping while balancing agility, versatility, and energy efficiency remains an open challenge.

B. Learning Approaches for Locomotion

Although optimization-based methods have made notable progress in locomotion control, their reliance on simplified dynamics models often results in a trade-off between accuracy and computational efficiency [5]. These limitations have motivated a growing interest in learning-based approaches, which allow robots to acquire complex behaviors through a direct function approximation. Reinforcement learning algorithms such as PPO [6] have been widely adopted for 3D locomotion, and hierarchical DRL frameworks further extend this by decomposing high-level motion planning and low-level control, enabling long-horizon tasks and multi-behavior training [7]. Recent advances also emphasize robustness and sim-to-real transfer. Lee *et al.* [8] developed a blind locomotion controller for rough terrain, while Hwangbo *et al.* [9] demonstrated actuator network training that effectively transfers to real quadrupeds. Building on reinforcement learning, [10] introduced a hierarchical RL framework that integrates locomotion control with long-horizon navigation planning, successfully deployed in urban missions of the kilometer scale. Despite these successes, learning approaches still face challenges in achieving reliable real-world deployment across diverse environments.
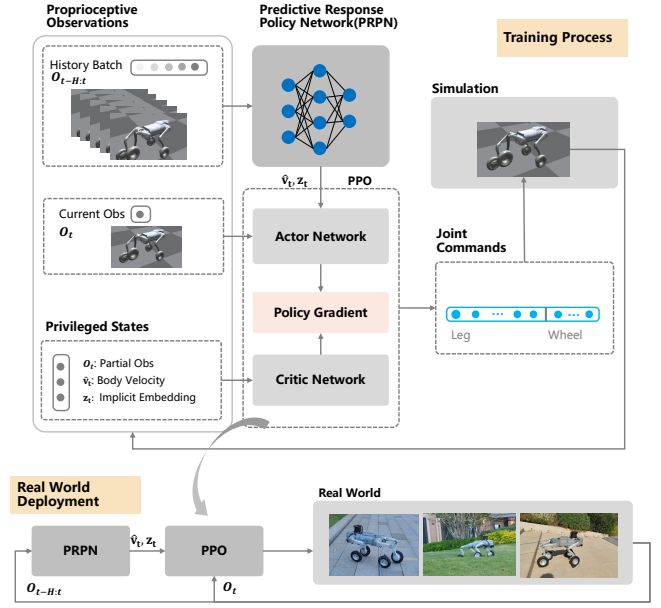

Fig. 1. The overview framework of **ATRos**. PRPN encodes partial observations into predictive signals for a PPO-based actor–critic framework, enabling adaptive and energy-efficient locomotion with zero-shot sim-to-real transfer.

III. WHEELED-LEGGED LOCOMOTION POLICY

A. Framework Overview

The wheeled-legged locomotion problem is formulated as a partially observable Markov decision process (POMDP), since critical factors such as terrain variations and external forces are not directly observable. At each timestep t, the agent receives a partial observation $\mathbf{o}_t \in \mathcal{O}$ and generates an action $\mathbf{a}_t \sim \boldsymbol{\pi}_\theta(\mathbf{a}_t|\mathbf{o}_t, \hat{\mathbf{v}}_t, \mathbf{z}_t)$ using the policy $\boldsymbol{\pi}_\theta$. To mitigate partial observability, we introduce a *Predictive Response Policy Network* (PRPN), which encodes a history of observations $\mathbf{o}_{t-H:t}$ to predict the body velocity $\hat{\mathbf{v}}_t$ and an implicit embedding $\mathbf{z}_t$, where H represents the number of the history batch steps. These predictive signals, together with the current observation $\mathbf{o}_t$, are passed to the actor network, which outputs the action $\mathbf{a}_t$. The critic network, in turn, receives both the privileged state $\mathbf{s}_t$ and the current observation $\mathbf{o}_t$ to provide value estimates for policy updates.

The actor and critic are jointly trained with proximal policy optimization (PPO) [6] to maximize the expected return while ensuring stable optimization. Our formulation explicitly integrates predictive dynamics $\hat{\mathbf{v}}_t$ and latent embeddings $\mathbf{z}_t$ as auxiliary signals. This design provides three advantages: (i) predictive modeling of body dynamics improves locomotion stability, (ii) implicit embeddings $\mathbf{z}_t$ offer a compact and transferable state representation for robust policy learning, and (iii) leveraging privileged information in the critic accelerates training and improves sample efficiency for sim-to-real transfer.
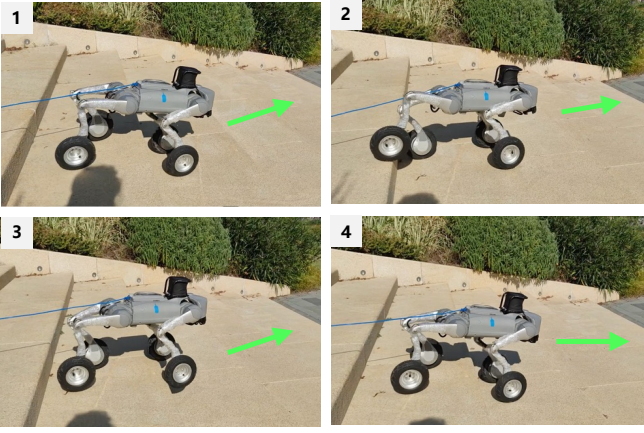
State Space. The state s_t is composed of proprioceptive observations:

$$\mathbf{s}_t = [\mathbf{c}_t, \ \omega_t, \ \mathbf{g}_t, \ \mathbf{q}_t, \ \dot{\mathbf{q}}_t, \ \mathbf{a}_{t-1}] \tag{1}$$
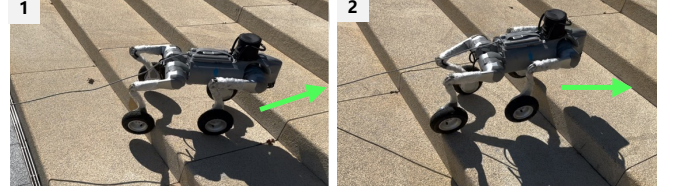
(a) Grassy terrain

(b) Going down stairs

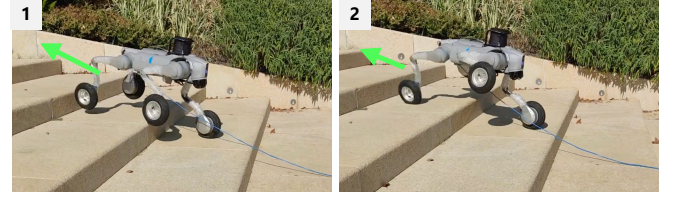(c) Going up stairs (forward)

(d) Going up stairs (backward)

Fig. 2. Snapshots of Go2W robot crossing different types of terrains: grass and stairs with the height of 13 cm. (a) Wheel-based traversal of grassy terrain and small slopes. (b) Wheel-based traversal of stair descent. (c) Forward stair ascent achieved through wheeled-legged locomotion. (d) Backward stair ascent through wheeled-legged locomotion.

where $\mathbf{c}_t \in \mathbb{R}^3$ represents the body velocity command, $\omega_t \in \mathbb{R}^3$ indicates the body angular velocity, $\mathbf{g}_t \in \mathbb{R}^3$ is the gravity vector, $\mathbf{q}_t \in \mathbb{R}^{16}$ and $\dot{\mathbf{q}}_t \in \mathbb{R}^{16}$ are the joint positions and joint velocities, $\mathbf{a}_{t-1} \in \mathbb{R}^{16}$ shows the previous action. Thus, the proprioceptive state dimension is $\mathbf{s}_t \in \mathbb{R}^{57}$.

Action Space. Different from pure legged locomotion, the policy here outputs actions for both legs $\mathbf{a}_t^{\mathrm{leg}} \in \mathbb{R}^{12}$ and wheels $\mathbf{a}_t^{\mathrm{wheel}} \in \mathbb{R}^4$. To accelerate the training process, the legged action space $\mathbf{a}_t^{\mathrm{leg}}$ is scaled by a factor k_l to constrain the output action and improve stability, and it is represented as an increment relative to a reference joint configuration. Thus, the target joint position of legs can be computed as:

$$q_t^{\mathrm{des}} = k_l \mathbf{a}_t^{\mathrm{leg}} + q_t^{\mathrm{ref}}, \qquad (2)$$

where q_t^{ref} is the reference joint configuration in the robot stand pose. In the wheeled action space, $k_w \mathbf{a}_t^{\mathrm{wheel}}$ is interpreted as the target velocity.

Reward Function. The overall reward is formulated as a weighted combination of multiple sub-rewards:

$$\mathbf{r}_t\left(\mathbf{s}_t, \mathbf{a}_t\right) = \sum \alpha_i r_i, \qquad (3)$$

where α_i denotes the weight coefficient assigned to each reward component, and r_i represents a specific reward term. We adopt the reward structure in [11] and additionally design wheel-related reward functions to better capture the characteristics of wheeled-legged locomotion. This extended reward formulation promotes efficient, stable, and terrain-adaptive locomotion, thereby facilitating robust sim-to-real transfer.

B. Training Details

To effectively train robust locomotion policies for wheeled-legged robots, we adopt a simulation framework combined with domain randomization and curriculum learning. The details are as follows.

- Simulation Setup: We train policies in Isaac Gym [12] using 4096 parallel environments with a rollout horizon of 100 steps. The framework allows fast, large-scale data collection, and stable performance is typically reached after several thousand iterations.
- Dynamics Randomization: To improve robustness and facilitate sim-to-real transfer, we randomize key phys-

ical parameters of both the robot and environment, including body and link masses, center of mass, payloads, ground properties, motor strength, PD gains, actuation delay, external perturbations, and initial configurations.

- Training Curriculum. Following [12], we employ a terrain curriculum to train the hybrid locomotion policy. Robots are initially trained on simple terrains, with difficulty gradually increasing based on performance. The terrain set contains 200 height field-based variations arranged by difficulty, and the command space is progressively expanded to cover complex locomotion scenarios such as slopes, stairs, and uneven grounds.

IV. EXPERIMENTS

In this section, we evaluate the proposed *ATRos* framework both in simulation and real-world environments. The experiments are designed to assess locomotion agility, stability, and energy efficiency across two representative terrains: (i) grass terrain, (ii) stairs. The details of the experimental setup and details are introduced as following.

A. Experimental setup

For hardware deployment, we use a Unitree Go2W EDU wheeled-legged robot, which possesses 16 DoFs (twelve for legged actuators and four for wheeled actuators) as shown in Fig. 2. The computation is provided by an NVIDIA RTX 4090 GPU, which enables real-time operation. The RL-based hybrid locomotion policy runs at 50 Hz, ensuring responsive execution of leg and wheel actions. The generated joint torques are transmitted from the computer to the robot actuators via a wired Ethernet connection, supporting low-latency and reliable deployment in physical environments.

B. Cross grass terrains

The grass terrain consist of soft, uneven outdoor grassland with natural slopes and scattered pits, designed to test adaptability, stability, and robustness under irregular conditions. On grass terrain, the robot tends to utilize its wheels for rolling, which highlights the energy efficiency of wheeled locomotion while simultaneously maintaining motion agility, thus achieving a balance between efficiency and flexibility. The snapshots of these experiments are shown in Fig. 2(a).

C. Cross stairs

To evaluate the obstacle adaptability performance of the hybrid locomotion system, ascending and descending experiments were carried out on an outdoor staircase with a step height of 13 cm. It was anticipated that during descent, the robot would predominantly use wheeled motion to improve energy efficiency, whereas during ascent, the legged versatility would be fully utilized to enhance obstacle traversal.

As illustrated in the accompanying figure, the robot employed exclusively wheeled locomotion throughout the entire descending phase, as shown in the test snapshot in Fig. 2(b).

During ascent, the robot effectively leveraged the combined wheeled-legged locomotion, with the corresponding test snapshot provided in Fig. 2(c).

Furthermore, the robot is tested in a reverse stair-ascent mode. The results demonstrate that the robot maintains excellent performance even when ascending stairs backward, successfully achieving stable motion. A snapshot from this experiment is presented in Fig. 2(d).

V. CONCLUSION

In this paper, a hybrid locomotion framework is proposed to address the coordinated collaboration between wheels and legs for realizing smooth locomotion. The experiments demonstrate that ATRos achieves both agile and energy-efficient locomotion over diverse terrains. ATRos prefers wheeled mode that significantly realizing energy efficient motions on rolling-dominated terrains while maintaining comparable agility on uneven and stair environments.

Our future work will focus on two main aspects. First, we aim to enhance the stability of obstacle negotiation. Then, we are going to improve the integration of wheels and legs to achieve more natural and reliable wheeled-legged motions.

REFERENCES

[1] J. Lim, I. Lee, I. Shim, H. Jung, H. M. Joe, H. Bae, O. Sim, J. Oh, T. Jung, S. Shin *et al.*, "Robot system of drc-hubo+ and control strategy of team kaist in darpa robotics challenge finals," *Journal of Field Robotics*, vol. 34, no. 4, pp. 802–829, 2017.

[2] P. Hebert, M. Bajracharya, J. Ma, N. Hudson, A. Aydemir, J. Reid, C. Bergh, J. Borders, M. Frost, M. Hagman *et al.*, "Mobile manipulation and mobility as manipulation—design and algorithms of robosimian," *Journal of Field Robotics*, vol. 32, no. 2, pp. 255–274, 2015.

[3] M. Bjelonic, C. D. Bellicoso, Y. de Viragh, D. Sako, F. D. Tresoldi, F. Jenelten, and M. Hutter, "Keep rollin'-whole-body motion control and planning for wheeled quadrupedal robots," *IEEE Robotics and Automation Letters*, 2019.

[4] Z. Song, J. Xu, C. Zheng, Y. Li, Z. Bi, and J. Ma, "Flores: A reconfigured wheel-legged robot for enhanced steering and adaptability," *arXiv preprint arXiv:2507.22345*, 2025.

[5] 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.

[6] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal policy optimization algorithms," *arXiv preprint arXiv:1707.06347*, 2017.

[7] 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.

[8] J. Lee, J. Hwangbo, L. Wellhausen, V. Koltun, and M. Hutter, "Learning quadrupedal locomotion over challenging terrain," *Science robotics*, vol. 5, no. 47, 2020.

[9] 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, 2019.

[10] J. Lee, M. Bjelonic, A. Reske, L. Wellhausen, T. Miki, and M. Hutter, "Learning robust autonomous navigation and locomotion for wheeled-legged robots," *Science Robotics*, 2024, integrated hierarchical RL for navigation and locomotion in urban settings.

[11] J. Long, Z. Wang, Q. Li, L. Cao, J. Gao, and J. Pang, "Hybrid internal model: Learning agile legged locomotion with simulated robot response," in *The Twelfth International Conference on Learning Representations*, 2024.

[12] N. Rudin, D. Hoeller, P. Reist, and M. Hutter, "Learning to walk in minutes using massively parallel deep reinforcement learning," in *Conference on Robot Learning*. PMLR, 2022, pp. 91–100.