

Whleaper: A 10-DOF Flexible Bipedal Wheeled Robot

Yinglei Zhu^{1†}, Sixiao He^{1†}, Zhenghao Qi¹, Zhuoyuan Yong¹, Yihua Qin¹ and Jianyu Chen^{2*}

Abstract—Wheel-legged robots combine the advantages of both wheeled robots and legged robots, offering versatile locomotion capabilities with excellent stability on challenging terrains and high efficiency on flat surfaces. However, existing wheel-legged robots typically have limited hip joint mobility compared to humans, while hip joint plays a crucial role in locomotion. In this paper, we introduce Whleaper, a novel 10-degree-of-freedom (DOF) bipedal wheeled robot, with 3 DOFs at the hip of each leg. Its humanoid joint design enables adaptable motion in complex scenarios, ensuring stability and flexibility. This paper introduces the details of Whleaper, with a focus on innovative mechanical design, control algorithms and system implementation. Firstly, stability stems from the increased DOFs at the hip, which expand the range of possible postures and improve the robot's foot-ground contact. Secondly, the extra DOFs also augment its mobility. During walking or sliding, more complex movements can be adopted to execute obstacle avoidance tasks. Thirdly, we utilize two control algorithms to implement multimodal motion for walking and sliding. By controlling specific DOFs of the robot, we conducted a series of simulations and practical experiments, demonstrating that a high-DOF hip joint design can effectively enhance the stability and flexibility of wheel-legged robots. Whleaper shows its capability to perform actions such as squatting, obstacle avoidance sliding, and rapid turning in real-world scenarios.

I. INTRODUCTION

Mobile robots can be divided into aerial and ground-based systems. Aerial robots are susceptible to changes in external airflow [1] and have limitations in endurance and positioning accuracy [2]. Ground robots are classified according to their mobility mechanisms as wheeled, legged, or crawler-based [3]. Wheeled robots are fast and efficient on flat terrain but struggle with obstacle traversal, terrain adaptability, and turning efficiency [4], [5]. On the other hand, legged robots, particularly bipedal ones, demonstrate exceptional adaptability across diverse terrains and excel in navigating obstacles [6], [7], [8], [9]. Legged robots provide improved flexibility, maneuverability, and stability [10], [11]. Nevertheless, their movement speeds are typically lower, and they are prone to rollovers. Wheel-legged robots have attracted considerable research interest in recent years. They combine the strengths of wheeled robots with legged robots, enabling fast and agile locomotion on flat terrain while effectively navigating obstacles [12]. Wheel-legged robots combine the high efficiency of wheels with the terrain

adaptability of legs. However, existing wheel-legged robots often have only one [13], [14], [15] or two [16], [17] DOFs at the hip joint. This limitation impedes tasks such as lateral walking and obstacle avoidance sliding. To enhance motion flexibility, it is imperative to increase the robot's DOFs.

Our goal is to design a 10-DOF bipedal wheeled robot, Whleaper, with 3 DOFs at each hip, similar to humans. Using optimal control theory and reinforcement learning control, we have developed control methods for our robot to perform multimodal motion. Compared to other robots, including legged and wheeled ones, and existing bipedal wheeled robots such as Ascento from ETH Zurich [18], our robot distinguishes itself with more DOFs at hip, a complex structure, and enhanced movement stability and flexibility.

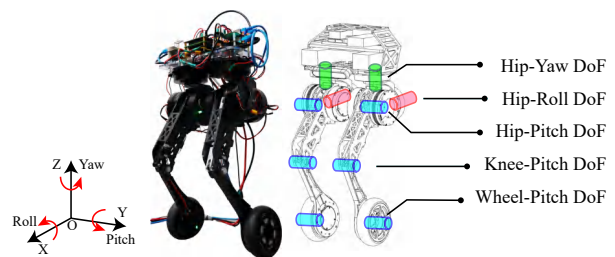


Fig. 1: Prototype of Whleaper robot and schematic diagram of DOF configuration.

The main contributions of this paper can be summarized as follows:

- Mechanical design of a 10-DOF bipedal wheel-legged robot, Whleaper, with terrain adaptability and high-performance mobility.
- Development of tailored control methods, including Linear Quadratic Regulator (LQR) and Reinforcement Learning (RL), which enable smooth transitions between sliding and walking modes.
- Establishment of the physical system for Whleaper, including both hardware and software.
- Validation of the mechanical design and algorithms via a series of simulations and real-world experiments.

The paper follows the structure outlined below:

- Section II introduces the robot's system, including mechanical design, hardware, and software.
- Section III explains the dynamic model and control methodology, involving the LQR equation, balancing control, and reinforcement learning strategy.
- Section IV presents an overview of the experiment settings, results from simulations and real-world scenarios.

[†]Contributed equally to this work.

¹Y. Zhu, Z. Yong and Y. Qin are with Weiyang College, S. He is with Department of mechanical engineering, and Z. Qi is with Department of Automation, Tsinghua University, Beijing, China, 100083.

²is with Institute for Interdisciplinary Information Sciences, Tsinghua University, Beijing, China, 100083.

*Corresponding author: Jianyu Chen, jianyuchen@tsinghua.edu.cn.

- Finally, in section V, we conclude the paper and outline avenues for future work.

II. SYSTEM DESCRIPTION

A. Mechanical Design

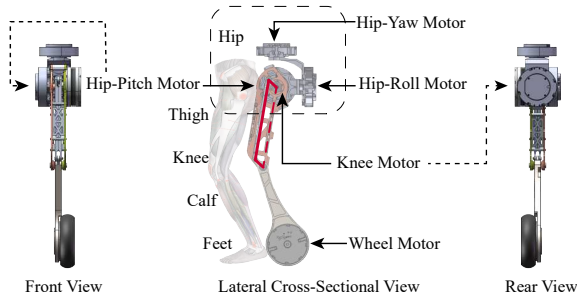


Fig. 2: Schematic diagram of joint space.

Whleaper’s legs have 5 DOFs each. As mentioned above, we aim to improve the stability and flexibility of the wheel-legged robot by increasing the DOFs in hip joints, and maintain its capabilities for both wheeled and legged locomotion.

It is necessary to demonstrate the rationale for this increase in DOFs. From a biological perspective, the hip of a human has 3 DOFs to ensure smooth walking [19]. The distinctive difference between the wheel-legged robot and humans is the presence of wheels for ground contact. A well-designed system should allow the wheels to have a wide range of orientations relative to the ground, thereby providing a greater variety of foot placement options.

Considering a wheel in contact with the ground, its orientation can be described using just two coordinates due to its axial symmetry (namely roll and yaw. Pitch represents the rotational direction of the wheel). Similar to ETH’s Ascento and Tencent’s Ollie [20], our design incorporates hip-pitch and knee-pitch motions for maintaining ground contact and improving adaptability to uneven terrain. Therefore, in our design, the orientation of the wheel is achieved by adding hip-roll and hip-yaw DOFs, while hip-pitch and knee-pitch DOFs ensure the walking motion.

The robot is composed of a base, hip joints, thighs, calves, and wheels. The hip joints are constructed with ball-wrist structures, providing 2 DOFs, yaw and roll, for thighs. Meanwhile, the hip joints’ ends and the thighs are connected by motors, which offering pitch DOFs to lift the robot’s legs.

Furthermore, we have implemented a series-parallel hybrid mechanical connection structure for the legs. The four-bar linkage reduces knee joint weight, lowers leg inertia, and enables precise leg control. Moreover, it efficiently transfers torque from the hip to the knee joint. So the overall leg structure is kinematically equivalent to a serial manipulator. This configuration is advantageous for both kinematic and dynamic solutions. At the end, a motor is directly attached to drive the rotation of wheels.

From the practical perspective, the hip-yaw and hip-roll movements play a crucial role in establishing the grounding posture of the wheels, while the hip-pitch motion facilitates

the lifting motion of the thigh. The knee motion ensures continuous contact between the wheels and the ground, fulfilling the required constraints.

TABLE I: Comparison of joint ranges (deg.) between Whleaper and the human body

Joint	Human Range	Robot Range
Hip Roll	-48,+31	-20,+5
Hip Pitch	-113,+28	-60,+0
Hip Yaw	-45,+45	-5,+5
Knee Pitch	-0,+134	-0,+120

Table I illustrates that Whleaper has a human-like range of motion in the knee, with 3 DOFs in its hip, but its overall range of motion still falls short of that of humans.

B. Hardware

Tailored to its unique mechanical structure, we have designed a specialized electrical and communication system for Whleaper, as depicted in Figure 3.

Because of concentrated motors and aluminum structure at the base and hip, Whleaper has a high center of gravity (COG). We use THU-8108 [21] motors for efficient control of multiple joints, delivering peak torques of 18 Nm. Specifically, THU-8112 [21] motors with peak torques of 50 Nm are chosen for the calves to withstand significant knee forces. Each motor is equipped with dual encoders for precise feedback on position, velocity, and torque, adaptable to closed-loop control methods.

We use the MicroStrain 3DM-GX5-AHRS IMU for high-precision state capture, featuring 1000 Hz sampling.

In terms of communication, two microcontrollers serve as the bridges between the motors and the computer. To enhance real-time performance, a hybrid communication approach is employed between the computer, microcontrollers, and motors, utilizing both Ethernet for Control Automation Technology (EtherCAT) and Controller Area Network (CAN).

For stable power supply, we use a 48V lithium-ion polymer (LiPo) battery pack.

Finally, a list of components can be found in Table II.

TABLE II: Components list

Component	Name
Computer	HP OMEN 16 GAMING LAPTOP
Knee Motors	THU-A8112-09
Other Motors	THU-A8108-06
Microcontrollers	XMC4300-F100F256
IMU	MicroStrain 3DM-GX5-AHRS
Joystick	Xbox Elite Wireless Controller Series 2
Battery	GS-5300mAh-30C-6S1P
Miniature Circuit Breaker	SIEMENS 5SJ52

C. Software

As depicted in Figure 4, the software architecture deployed on Whleaper is structured into three layers: the instruction layer, the control layer, and the hardware layer.

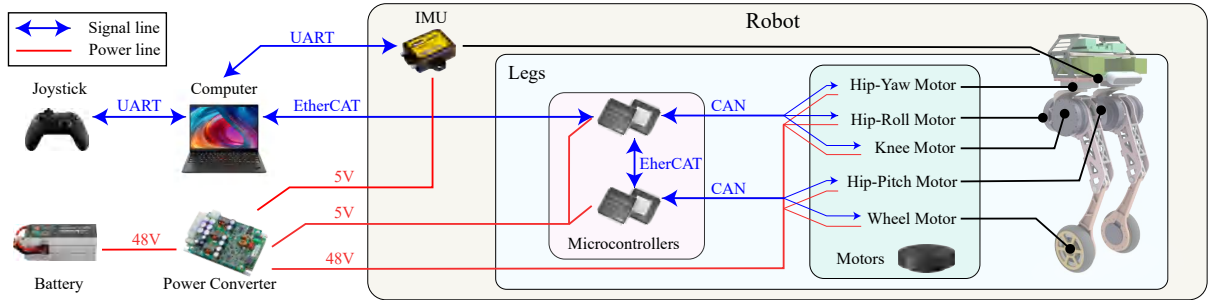


Fig. 3: Hardware connection diagram.

The instruction layer is responsible for user interaction and command transmission. Users can input desired pose and speed into the control layer by manipulating a joystick, while simultaneously observing real-time feedback of overall posture and joint states through the user interface (UI).

As a critical link in feedback control, the control layer is tasked with processing information and issuing commands. It calculates the estimated state vector by integrating data from the IMU and motors. Subsequently, the designated controller, such as LQR, computes motor torque commands. These commands are then transmitted to the hardware layer, and smoothly adjust the robot to the desired state.

The hardware layer is responsible for low-level data acquisition and control of hardware components including the IMU and motors. At the motor's low-level control, a PID controller is employed. The hardware layer also forwards overall attitude information from the IMU and joint states from motor encoders to the control layer.

III. ALGORITHM

In order to enable multiple modes of locomotion, namely sliding and walking, we have developed dedicated control algorithms for each. We utilize an optimal control algorithm (LQR) to achieve balanced sliding, while a reinforcement learning method based on Proximal Policy Optimization (PPO) [22] is employed to facilitate walking, jumping, etc.

A. Sliding: LQR Control for Balancing

1) *Coordinate System and Symbols:* The coordinate system is established as depicted in Figure 5.

We designate the rotations along the roll, pitch, and yaw axes of the hip as α , β , and γ respectively. The rotation of the knee is denoted by ϕ , and $\theta_{r/l}$ represents the rotation of wheel, while the subscripts r and l represent right and left. Considering the design specifications of the mechanism,

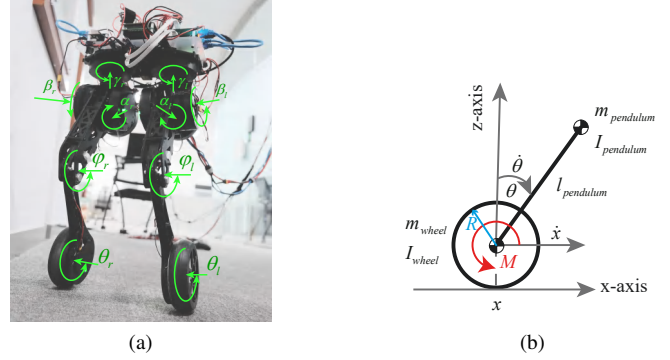


Fig. 5: Schematic representation of dynamics model.

where the range of motion is limited, we assume a momentarily constant mass for the base, the hip, and the connected thigh. Additionally, considering that the angles α and γ fall within a narrow range, we make the assumption that the leg can be effectively approximated as a planar rigid body in motion. We abstract the system as a wheeled inverted pendulum system (WIPS) [23] and establish the coordinate system as shown in Figure 5. We represent the LQR state vector as $\mathbf{x} = [x, \dot{x}, \theta, \dot{\theta}]^T$. θ represents the tilt angle. With the robot's pose provided, we can calculate the equivalent mass, inertia, and length:

$$m, I, l = m, I, l_{pendulum}(\alpha_l, \alpha_r, \beta_l, \beta_r, \gamma_l, \gamma_r, \phi_l, \phi_r) \quad (1)$$

2) *Dynamic Equations Derivation:* For the given model, the Lagrangian equations of motion are given by

$$L = T - V \quad (2)$$

$$\frac{d}{dt} \left(\frac{\partial L}{\partial \dot{q}_i} \right) - \frac{\partial L}{\partial q_i} = Q_i \quad (3)$$

where q_i represents each generalized coordinate and Q_i stands for generalized force term. According to Kim [23], the dynamic equation of WIPS can be driven as below:

$$(m_w + m_p + \frac{I_w}{R^2})\ddot{x} + m_p l_p \cos \theta \ddot{\theta} - M_p l_p \sin \theta \dot{\theta}^2 = \frac{M}{R} \quad (4)$$

$$(m_p l_p^2 + I_p)\ddot{\theta} + m_p \dot{x} l_p \cos \theta - m_p g l_p \sin \theta = -M \quad (5)$$

where the subscripts p and w represent pendulum and wheel respectively.

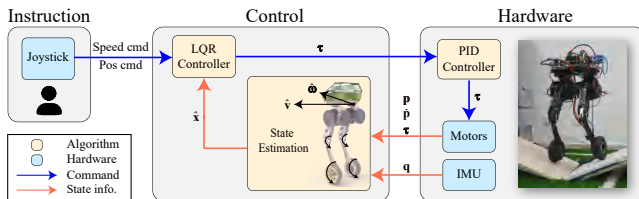


Fig. 4: Control system architecture.

3) *Stabilizing Control*: To achieve stable sliding motion, a robust stabilization algorithm is required. The target of the algorithm is to make the robot handle external disturbances to keep its equilibrium, using as little time and energy as possible. However, time saving and energy saving are contradictory in most cases. To solve this problem, LQR, an optimal control method for controlling a linear system with minimal cost, is used in the robot. Li, Yang, and Fan [24] demonstrated that employing an LQR controller effectively ensures strong reliability and robustness in stabilization scenarios involving a two-wheeled inverted pendulum.

The state equation for the LTI (Linear Time-Invariant) discrete system is as follows:

$$\mathbf{x}(k+1) = \mathbf{A}\mathbf{x}(k) + \mathbf{B}\mathbf{u}(k). \quad (6)$$

In the equation, \mathbf{x} is the state vector, \mathbf{u} is the input vector. The control matrices \mathbf{A} and \mathbf{B} define the discrete-time state-space representation of the system, which is derived from the linearization of the robot model (in formula 4 and 5) around the state vector $\mathbf{x} = [0, 0, 0, 0]^T$. The quadratic performance index J is denoted as

$$J = \sum_{k=0}^{\infty} \mathbf{x}^T(k)\mathbf{Q}\mathbf{x}(k) + \mathbf{u}^T(k)\mathbf{R}\mathbf{u}(k) \quad (7)$$

In the given equation, \mathbf{Q} and \mathbf{R} represents the weighting matrices for the state covariance. The optimal gain matrix \mathbf{K} can be determined by solving the discrete time algebraic Riccati equation. The robot is dynamically adjusted by the following control law, where \mathbf{u} contains M in Figure 5

$$\mathbf{u} = -\mathbf{K}\mathbf{x} \quad (8)$$

where the state vector \mathbf{x} can be derived from the kinematics.

Besides, for primary pose control of the legs, position control is employed, corresponding to the modification of equivalent mass parameters in the inverted pendulum model.

B. Walking: Reinforcement Learning Control

For robust walking and jumping control of 10-DOF robot, we conduct the PPO algorithm [22] in the simulation. PPO is a reinforcement learning algorithm that applies proximal policy optimization to train policies. It ensures stability by constraining the magnitude of policy updates and uses a clipping mechanism to control the extent of updates.

The target function is given by

$$L^{CLIP}(\theta) = \hat{E}_t[\min(r_t(\theta)\bar{A}_t, \text{clip}(r_t(\theta), 1 - \epsilon, 1 + \epsilon)\bar{A}_t)] \quad (9)$$

The first term of the minimum function is L^{CLIP} . The second term, $\text{clip}(r_t(\theta), 1 - \epsilon, 1 + \epsilon)\bar{A}_t$, is used to restrict the range of the probability ratio $r_t(\theta)$ by clipping it between $1 - \epsilon$ and $1 + \epsilon$ (where $\epsilon = 0.2$), ensuring that it does not deviate too far from 1.

As for the reward, we use a fine-tuned reward function to regulate the robot's actions, enabling it to walk and jump. We set termination and joint position limits to negative values, with large absolute magnitudes. The purpose is to expedite the robot's convergence to a stable state where it maintains balance and adheres to joint constraints. Taking jumping as

an example, we reward the time when its feet are off the ground, the velocity along the positive direction of the Z-axis, and the base height. Meanwhile, rotation and angular velocity are penalized. These lower-value reward functions ensure that the robot's jumping occurs in place.

IV. EXPERIMENTS

A. Simulation Platform

We have established a distributed simulation platform comprising three components:

- 1) Simulator: Simulates environments and robot motion.
- 2) Controller: Executes robot motion control algorithms.
- 3) Communicator: Links the controller and simulator via Robot Operating System (ROS).

This framework decouples simulator, control algorithms, and communication protocols. It facilitates the adaptation of these algorithms for real-world applications.

Drawing from the methodology presented in Legged Gym [25], we utilize IsaacGym [26] for GPU-based parallel training. We use PhysX and apply LQR and PPO algorithms.

B. Experimental Results

We designed experiments to showcase the advantages of Whleaper in terms of stability (IV-B.1), agility (IV-B.2), and flexibility (IV-B.3).

1) *Stability*: The experimental design is illustrated in Figure 6. Within the simulation, we initially restrict certain DOFs at the robot's hip to create different configurations of the wheel-legged robot. Subsequently, we attempt to control the robot using either LQR or PPO algorithms and compare their performance in executing specific motion tasks. Among these tasks, testing the walking stability stands out as particularly significant. In this experiment, we employ RL to guide the robot in a forward walking motion while concurrently monitoring the fluctuations in base attitude.

The base deviation depicted in the Figure 7 is quantified by the Root Mean Square (RMS) of the Roll and Pitch angles and angular velocities. The result illustrates that as the number of DOFs increases, the variations in base angles and angular velocities tend to decrease. Specifically, the hip-pitch DOF has the most significant impact, with the other two DOFs also exhibiting discernible effects.

With reference to the video, the increase in DOFs enhances the selection of foot placement during walking and expands the robot's state space, which may improve walking stability. It is noteworthy that while locking the roll and yaw DOFs results in smaller changes in angle and angular velocity, the gait appears slower and less natural according to the video.

2) *Agility*: In real-world scenarios, we coordinate multiple joints of Whleaper by LQR control, for achieving yaw angle variation in place, referred to as 'twist'. As shown in Figure 8 and the accompanying video, Whleaper with 10 DOFs exhibits smoother and faster yaw angle changes compared to configurations with 8 DOFs. This improvement stems from the 10-DOF robot's ability to adjust base rotation primarily through changes in leg posture with minimal wheel movements. In contrast, robots with 6 DOFs (e.g., Ascento)

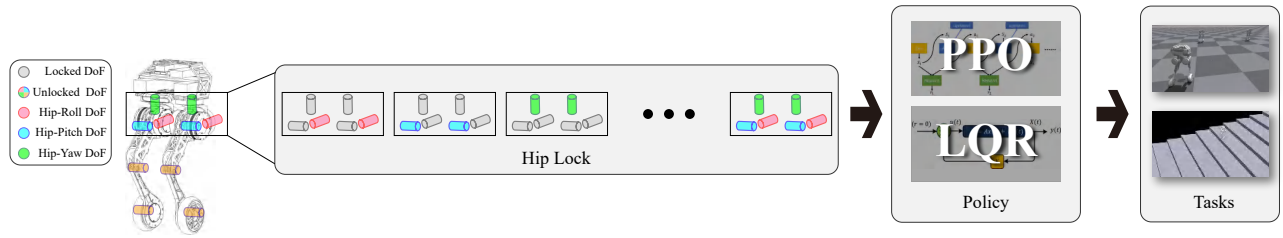


Fig. 6: Schematic diagram of the balance experiment.

rely solely on differential wheel speeds for rotation. Thus, the addition of hip joints can significantly enhance the execution of specific maneuvers.

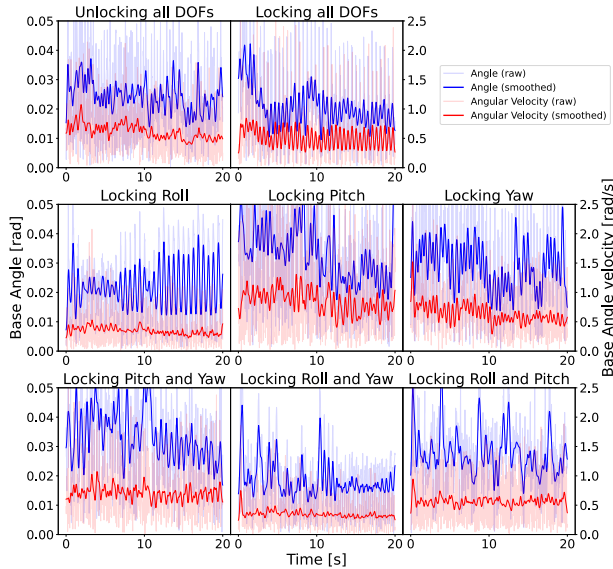


Fig. 7: Base variation under different DOFs in the balance experiment. The dark curve represents the actual data, while the light curve indicates the data after Gaussian smoothing.

3) *Flexibility*: As depicted in Figures 9, Wheaper demonstrates various human-like actions, highlighting its high-DOF advantage.

In narrow indoor environments, in addition to reaching destinations through a combination of straight and turning movements, it may be necessary to control the position and posture of the robot’s foothold to avoid obstacles. Therefore, for the two modes of movement, walking and sliding, we conducted tests using unconventional trajectories respectively. Lateral walking, as shown in Figure 9(c), requires coordinated motion of the three hip joints, a capability not achievable by existing wheel-legged robots. Obstacle avoidance sliding, as shown in Figure 9(d), utilizes the hip-roll joint to bifurcate the legs while simultaneously using the hip-yaw joint to alter the wheel’s heading. This coordinated motion helps to reduce ground resistance during leg bifurcation. Therefore, the high DOFs at the hip of Wheaper is necessary for precisely controlling the wheel trajectory.

Furthermore, movements like skate-like turn (Figure 9(e)) and knee swing (Figure 9(f)) also require all of the hip joints,

demonstrating Wheaper’s advantages over 6/8-DOF robots.

It’s worth noting that, besides flexibility in walking and sliding, in a simulated environment, by altering policies, we can freely switch the robot between LQR-controlled sliding and RL-controlled walking. This motion switching enables the robot to adopt different motion modes when dealing with various scenarios, further enhancing the flexibility of wheel-legged robots. Watching the video provides a comprehensive understanding of the robot’s capabilities.

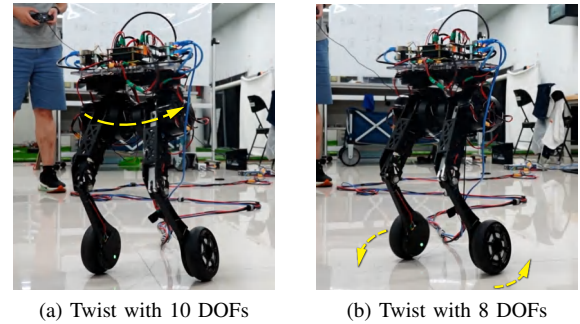


Fig. 8: Photos of twist motion under different DOFs.

V. CONCLUSIONS

This paper introduces Wheaper, a bipedal wheeled robot with 10 DOFs. Wheaper is equipped with a flexible and robust mechanical structure that enables it to perform a range of tasks, such as walking, jumping, and obstacle avoidance sliding. The robot’s versatile design allows for seamless transitions between different modes of locomotion, making it suitable for diverse applications.

We have successfully implemented LQR and reinforcement learning control in the simulation. It showcases the enhanced stability, agility, and flexibility of the wheel-legged robot by increasing hip DOFs. However, we only utilized LQR in limited scope of real-world experiments. Bridging the sim2real gap requires additional effort. In the future, we aim to transfer the reinforcement learning methods into real-world applications, and implement control strategies such as Model Predictive Control (MPC) and Whole-Body Control (WBC) to improve the precision and reliability of wheel-leg cooperative control.

ACKNOWLEDGMENT

This work was supported by the Intelligent Systems and Robotics Lab (ISR) at IIIS, Tsinghua University (THU). We

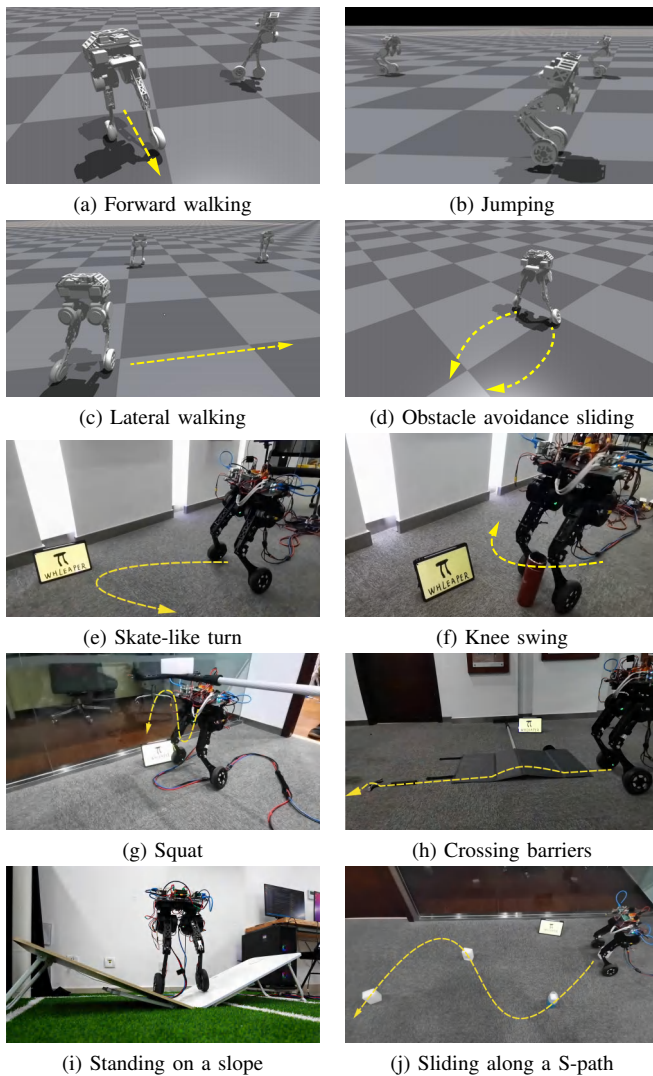


Fig. 9: Photos of the robot in a series of motions, in simulation and reality.

thank Merlin, Xiang Zhu, Zhi Zhang, Qiuhe Liu, and the ISR Lab team for helps in system setup and experiments.

Gratitude goes to Prof. Huichan Zhao, Prof. Chuxiong Hu, Prof. Mingguo Zhao, and colleagues from Weiyang College, Mechanical, Electrical, and Automation departments, along with Yang Cheng and Xuguang Dong, for their support.

Finally, we sincerely thank to the 41st THU Challenge Cup Match, for inspiring our dedication to robotics research.

REFERENCES

- [1] S. Park, J. Her, J. Kim, and D. Lee, "Design, modeling and control of omni-directional aerial robot," in *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2016, pp. 1570–1575.
- [2] A. E. Jimenez-Cano, J. Martin, G. Heredia, A. Ollero, and R. Cano, "Control of an aerial robot with multi-link arm for assembly tasks," in *2013 IEEE International Conference on Robotics and Automation*. IEEE, 2013, pp. 4916–4921.
- [3] L. C. Santos, F. N. Santos, E. S. Pires, A. Valente, P. Costa, and S. Magalhães, "Path planning for ground robots in agriculture: A short review," in *2020 IEEE International Conference on Autonomous Robot Systems and Competitions (ICARSC)*. IEEE, 2020, pp. 61–66.

- [4] Gita. [Online]. Available: <https://piaggiofastforward.com/>
- [5] J. Roos and M. Sütterlin, "Rezero with arm: A three degrees of freedom manipulator balancing on a ball," Master's thesis, ETH Zurich, 2017.
- [6] M. Hutter, C. Gehring, A. Lauber, F. Gunther, C. D. Bellicoso, V. Tsounis, P. Fankhauser, R. Diethelm, S. Bachmann, M. Blösch, et al., "Anymal-toward legged robots for harsh environments," *Advanced Robotics*, vol. 31, no. 17, pp. 918–931, 2017.
- [7] Spot. [Online]. Available: <https://bostondynamics.com/products/spot/>
- [8] Atlas. [Online]. Available: <https://bostondynamics.com/atlas/>
- [9] Honda. [Online]. Available: <https://asimo.honda.com/>
- [10] R. Armour, K. Paskins, A. Bowyer, J. Vincent, and W. Megill, "Jumping robots: a biomimetic solution to locomotion across rough terrain," *Bioinspiration & biomimetics*, vol. 2, no. 3, p. S65, 2007.
- [11] J. Mi, Y. Wang, and C. Li, "Omni-roach: A legged robot capable of traversing multiple types of large obstacles and self-righting," in *2022 International Conference on Robotics and Automation (ICRA)*. IEEE, 2022, pp. 235–242.
- [12] S. Wang, L. Cui, J. Zhang, J. Lai, D. Zhang, K. Chen, Y. Zheng, Z. Zhang, and Z.-P. Jiang, "Balance control of a novel wheel-legged robot: Design and experiments," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2021, pp. 6782–6788.
- [13] F. Raza and M. Hayashibe, "Towards robust wheel-legged biped robot system: Combining feedforward and feedback control," in *2021 IEEE/SICE International Symposium on System Integration (SII)*. IEEE, 2021, pp. 606–612.
- [14] L. Cui, S. Wang, J. Zhang, D. Zhang, J. Lai, Y. Zheng, Z. Zhang, and Z.-P. Jiang, "Learning-based balance control of wheel-legged robots," *IEEE Robotics and Automation Letters*, vol. 6, no. 4, pp. 7667–7674, 2021.
- [15] J. Jo, J. An, Y. Lee, and Y. Oh, "Whole-body balancing and differential driving control using a relative vector-based lqr to address non-holonomy of a wheel-legged robot," in *2023 20th International Conference on Ubiquitous Robots (UR)*. IEEE, 2023, pp. 302–307.
- [16] F. Raza, D. Owaki, and M. Hayashibe, "Modeling and control of a hybrid wheeled legged robot: Disturbance analysis," in *2020 IEEE/ASME International Conference on Advanced Intelligent Mechatronics (AIM)*. IEEE, 2020, pp. 466–473.
- [17] X. Li, H. Zhou, H. Feng, S. Zhang, and Y. Fu, "Design and experiments of a novel hydraulic wheel-legged robot (wlr)," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2018, pp. 3292–3297.
- [18] L. Zhao, Z. Yu, X. Chen, G. Huang, W. Wang, L. Han, X. Qiu, X. Zhang, and Q. Huang, "System design and balance control of a novel electrically-driven wheel-legged humanoid robot," in *2021 IEEE International Conference on Unmanned Systems (ICUS)*. IEEE, 2021, pp. 742–747.
- [19] H.-J. Lee, S. Lee, W. H. Chang, K. Seo, Y. Shim, B.-O. Choi, G.-H. Ryu, and Y.-H. Kim, "A wearable hip assist robot can improve gait function and cardiopulmonary metabolic efficiency in elderly adults," *IEEE transactions on neural systems and rehabilitation engineering*, vol. 25, no. 9, pp. 1549–1557, 2017.
- [20] J. Zhang, S. Wang, H. Wang, J. Lai, Z. Bing, Y. Jiang, Y. Zheng, and Z. Zhang, "An adaptive approach to whole-body balance control of wheel-bipedal robot ollie," in *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2022, pp. 12 835–12 842.
- [21] Encos. [Online]. Available: <http://encos.cn/Product.aspx>
- [22] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal policy optimization algorithms," *arXiv preprint arXiv:1707.06347*, 2017.
- [23] Y. Kim, S.-H. Lee, and D. H. Kim, "Dynamic equations of a wheeled inverted pendulum with changing its center of gravity," in *2011 11th International Conference on Control, Automation and Systems*. IEEE, 2011, pp. 853–854.
- [24] Z. Li, C. Yang, and L. Fan, *Advanced control of wheeled inverted pendulum systems*. Springer Science & Business Media, 2012.
- [25] 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.
- [26] V. Makovychuk, L. Wawrzyniak, Y. Guo, M. Lu, K. Storey, M. Macklin, D. Hoeller, N. Rudin, A. Allshire, A. Handa, et al., "Isaac gym: High performance gpu-based physics simulation for robot learning," *arXiv preprint arXiv:2108.10470*, 2021.