2026 IEEE International Conference on Robotics and Automation (ICRA 2026)
June 1-5, 2026. Vienna, Austria

DeepSkate: reinforcement learning of a robust controller for energy efficient quadruped skating

James Florin Petri[1,2] and Gerard Lacey[1]

***Abstract*—Wheeled-legged hybrid robots have generated growing interest in the research community due to the need for more efficient and versatile locomotion. Most recent research has focused on active wheels, but passive wheeled systems have great potential in improving energy efficiency. However, skating remains highly complex due to the difficulties of balancing dynamic motion, managing wheel-ground interactions, achieving precise torque control for smooth rolling, and adapting to unpredictable terrain while maintaining stability. We present an end-to-end model-free reinforcement learning approach that enables quadrupedal robots to skate efficiently, achieving agile and robust locomotion on both flat and rough terrain. Our skating-specific policy and sim-to-real pipeline are validated on a physical quadruped across diverse terrains with varying roughness, slopes, and features, consistently demonstrating controlled and efficient traversal. The robot achieves velocities up to 1.5 m/s with a cost of transport 40.9% lower than the skating state of the art and 70.9% lower than standard legged locomotion. These results establish skating as a feasible and efficient alternative mode of urban locomotion for quadrupedal robots, setting a foundation for future wheeled-legged research.**

***Index Terms*—Reinforcement Learning, Legged Robots, Sensorimotor Learning**

I. Introduction

In recent years legged robots have become increasingly common, with bipedal and quadrupedal robots being developed for both research and commercial purposes [1]–[5]. Recent achievements in control, learning, and the quadruped's mechanics allow for navigating complex terrains with greater autonomy, however challenges in energy efficiency, real-world robustness, and cost continue to limit their widespread adoption beyond niche applications. Due to the low energy efficiency typical of legged robots, lengthy autonomous missions become impractical in unfamiliar environments without docking stations or battery swaps.

Interest in hybrid wheeled-legged locomotion has grown significantly due to its potential to combine the best attributes of both legs and wheels. Legged locomotion offers superior adaptability and manoeuvrability on complex and uneven terrains, while wheeled locomotion provides energy efficient and fast movement on flat or predictable surfaces. The integration of these two forms of locomotion into a single system creates a versatile robotic platform capable of traversing a wide range of environments [6]–[12].

The merger of the discrete walking and continuous wheeled locomotion domains enables for a more versatile platform at the cost of controller model complexity [6]–[10]. Most well-known wheel-legged hybrid robots are active (also referred to as actuated or driven) [9], [10], [13]–[15], but recently, interest in passive wheeled variants (also referred to as unactuated or skating) has increased [6], [8], [16]–[18]. Relative to standard quadrupeds, skating robots provide greater energy efficiency, measured using Cost-of-Transport (CoT), and more efficient cruising [6], [16], [19]. Compared to their active wheel counterparts, skating quadrupeds are cheaper, less complex and have a lower limb mass, resulting in a lower torque requirement for the leg motors [8], [16].

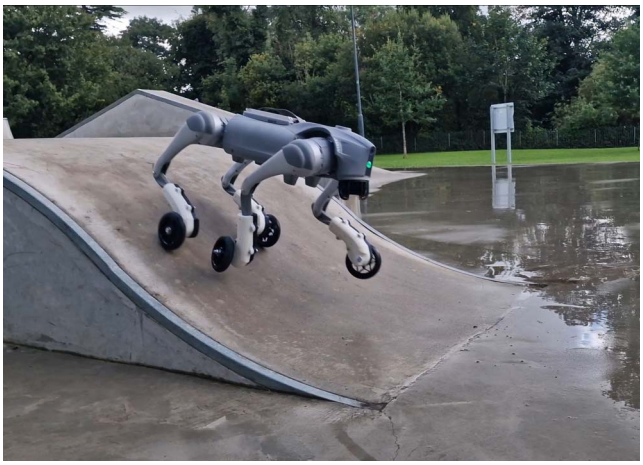

Fig. 1. Our passive-wheel modified Go2 skating down a ramp in a skatepark.

However, due to their reliance on body dynamics and ground interactions, passive wheeled-legged quadrupeds are harder to control relative to their active wheeled counterparts. Additionally, once at speed, the passive platforms present challenges in maintaining stability and safety while also allowing for rapid and controlled braking. Reinforcement Learning (RL) has seen many developments recently, becoming widely used in robotic locomotion [20], [21] as, with the proper implementation, end-to-end learning of a highly adaptable model with good generalization and the ability to handle non-linear behaviour can be rapidly trained [22]–[24].

Achieving high-performance skating, characterized by high speeds, adaptability and terrain-optimized locomotion, could pave the way for cementing hybrid legged robots for prolonged operation periods, addressing one of the main limitations of standard legged robots.

This work proposes a novel learning strategy for achieving effective skating on a quadruped with non-steerable wheels

[1]Department of Electronic Engineering, Maynooth University, Ireland
[2]Corresponding author: james.petri@mu.ie

for both flat and rough terrain. Our policy was successfully transferred to a modified Unitree Go2, achieving speeds of 1.5 m/s while maintaining stability and velocity control and a considerably lower cost-of-transport than the state of the art. The main contributions of this paper are summarized below:

1) A novel end-to-end model-free RL policy that learns skating specific dynamics, enabling high-performance quadruped skating without the need for steerable ankles.
2) Our skating quadruped achieves a significantly lower cost of transport, higher sustained speeds, and robustness to off-road terrain, surpassing prior skating approaches and highlighting skating as an efficient locomotion strategy.
3) A benchmark for skating quadrupeds, covering both flat and rough terrain traversal at effective speeds, providing a strong foundation for future research in efficient skating.

II. Related Work

Traditional approaches for legged robot control often rely on rigorous mathematical modelling combined with optimization. Model Predictive Control (MPC) is a popular framework that solves a finite-horizon optimal control problem while respecting system constraints, making it robust to disturbances and effective for multi-contact scenarios [25]. Other methods include Zero Moment Point (ZMP) control, which maintains stability by regulating the centre of pressure [6], [26], and Virtual Model Control (VMC), which uses virtual mechanical components to produce compliant motions [8], [27]. These modelled approaches offer interpretability and stability guarantees but can be limited by simplified dynamics and high computational cost.

Learning-based methods, particularly Reinforcement Learning, address these limitations by discovering control policies directly from experience. Algorithms such as PPO [28] and SAC [29] enable robots to learn by interacting with the environment, often aided by sim-to-real transfer techniques [23], [30]. While learning-based methods can handle complex unstructured environments better than classical approaches, they face challenges including large sample complexity, potential instability during training, and difficulty in providing formal safety guarantees.

There are numerous configurations and implementations for hybrid wheel-legged locomotion, each providing different solutions to the resultant locomotion space. Several skating quadruped implementations have emerged, primarily utilizing passive non-steerable or partially steerable ankle joints. Early platforms like ANYmal [6] and QSkater [17] demonstrated effective skating-walking gaits with no wheel steering and with the wheels aligned with the body. These rely on a force-controlled stance, often using claws or compliant covers for propulsion. The QSkater platform further explored hybrid configurations where some legs walk and others skate [17]. Both platforms achieving significant Cost of Transport (CoT) reductions (80% for ANYmal and 70% for QSkater) relative to their standard legged counterparts, however both implementations achieve slow speeds with little dynamic room for error handling or adaptability for uneven terrain. QSkater-E [31] expands on its predecessor [17] by adding an additional 4th DoF to the legs, allowing for a steerable ankle joint. This dexterous configuration allowed formation of statically stable gaits and the opportunity to compare the performance of an additional ankle joint. At a speed of 0.4 m/s, the 3DoF configuration had a lower CoT (0.136) relative to its 4DoF variant (0.18), which is likely a result of the increased leg mass and energy requirements of the additional motors [17], [31]. A quad–wheel–legged robot was demonstrated with a cooperative skating gait, achieving stable simulated locomotion at speeds up to 1 m/s, the fastest reported for a non-steerable skating quadruped to date, however no CoT metrics or deployment metrics were provided [32].

A 6-legged robot was adapted for skating by connecting the front and back legs with what are essentially skateboards, while the middle legs are used for propulsion, velocity control and stopping [8]. Notably, with their touchdown-assist gait [8], they can achieve speeds of up to 2.17 m/s and 95% velocity command accuracy by using the middle legs efficiently for pushing while the front and back legs control direction and stability. Energy metrics for the efficiency of the platform were not presented, but they succeeded in demonstrating the speed, versatility and viability of a skating hybrid platform leveraging the hybrid locomotion domain.

As a robot's speed increases, the controller design and complexity becomes more critical. The performance of model-based control is highly dependent on the designer's knowledge and the ability to model the robot's dynamics effectively [33]. Reinforcement Learning has proved to be a powerful tool for learning the dynamics of robot locomotion with many demonstrations of training reliable gaits in relatively short times with good generalization that reduce the sim-to-real gap [20], [21], [33], [34].

RL has been used for generating skating motions on JPL's highly dexterous quadruped Robosimian [35] by having the agent operate in Cartesian (task) space rather than in the joint space. They embedded forward and inverse kinematics knowledge directly into the model, which reduced training time, and successfully achieved a locomotion policies [35]. Passive skating on quadrupeds with non-steerable wheels was previously achieved using a model-based RL approach [16]. Propulsion was generated by pushing against the ground with the hind legs, which are set at a fixed ankle angle of 20°, with the direction and magnitude of the ground reaction force determined based on a simplified Single Rigid Body Dynamics (SRBD) model. Their model was validated in Isaac Gym and tested on a Unitree Go1, achieving speeds of 0.83 m/s and stable constant motion in two gaits. A CoT decrease of 73.9% was also observed relative to the Go1's standard trotting gait (1.38 to 0.36) [16].

III. Methodology

This section focuses on the design and implementation of our skating quadruped, with the control and training architecture presented in Fig. 2.

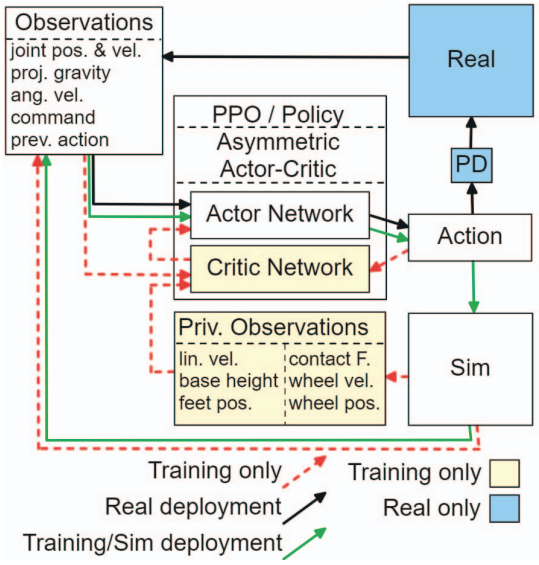

Fig. 2. Control loop diagram for the skating policy for training and sim/real deployment.

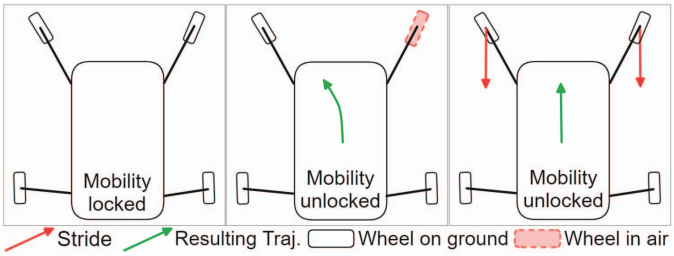

Fig. 3. A visualization of kinematically locking and unlocking the mobility of a quadruped with non-steerable wheels. Additionally, the ankle angle $\gamma \neq 0$ is visible on the front wheels.

A. Mechanical design and considerations

Motion characteristics vary based on the number of degrees of freedom per leg. When at least 4DoF are present, more dynamic skating motions and velocity control can be implemented [14], [36], [37], however, for a skating robot with 3 controllable DoF, braking or mobility locking techniques need to be considered [6], [7], [16]. As a mechanically simple system was desired, no active braking mechanism was added. By rotating the non-steerable ankle angle γ_i (relative to the robot frame x-axis), the robot's mobility can be locked and unlocked by lifting and rolling the wheels appropriately, technique demonstrated in Fig. 3. This technique can be used for propulsion and deceleration.

When skating, humans predominantly vary the wheel to body angle by approximately 0° to 30°, allowing for acceleration, cruising and slowing down by rotating the ankle [38]. As we wanted more leverage, γ_i was set to $\pm 30°$ on the forelegs and 0° on the hind legs, for which a URDF and USD model was derived with the Go2-W as the reference. On the real robot the feet were removed and 2 ASA 3D printed moulds clamp around the tibias. The wheels, commonly found

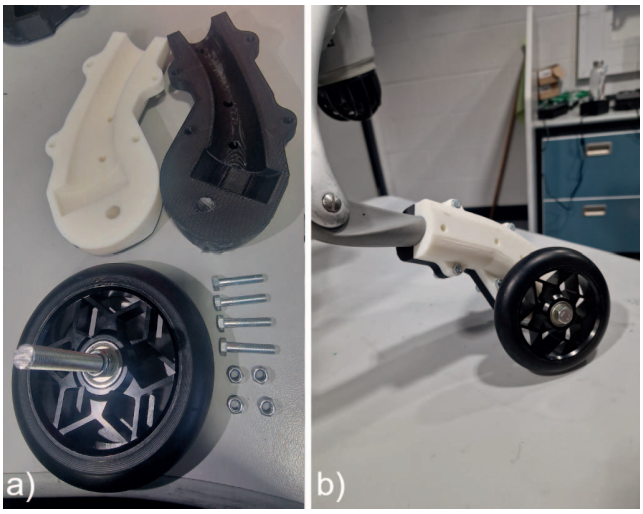

Fig. 4. a) Disassembled skate, including the 3D printed moulds. b) The skate assembly on the FR leg of the Go2 (The foot needs to be removed before clamping the moulds).

on scooters, feature a durable lightweight aluminium frame, a rubber exterior for grip, and an internal bearing that enables low-friction rotation. Each wheel has a diameter of 110 mm, with the mounting design presented in Fig. 4. As the feet were removed, the pneumatic force sensors are deactivated, resulting in no ground contact feedback.

B. Learning Environment

IsaacLab [39] was chosen as the training platform for GPU-based RL framework and the inclusion of many tools for training and for facilitating the sim-to-real transfer.

The RL training algorithm used is Proximal Policy Optimization (PPO) [28] in an asymmetric actor-critic structure. Our implementation is based on the open-source PPO implementation published by ETH [21]. PPO was selected as it provides stable and reliable policy updates, preventing performance collapse during the training of complex locomotion skills. The key hyperparameters are covered in Table I.

A terrain generator that seed randomizes sloped (hills and valleys), boxed and rough sub-environments was used for the training curriculum. The traversal difficulty, i.e. the roughness, or slope, increases with the curriculum level achieved, with the level increase being dictated by successfully navigating from the centre outwards. For going uphill, the maximum slope is set to 20%, while going downhill is set to a maximum of 40%. The boxes have a variation of 100 mm at the highest level, and the noise increases linearly for the rough environment. Flat sub-terrains were also added to promote a more universal gait. A visualization of some of the terrains and the robot skating through them can be seen in Fig. 5.

C. Task Formulation

The tasks are formalized as continuous Partially Observable Markov Decision Processes (asymmetric actor-critic), $(S, A, P, R, \Omega, O, \gamma)$, with the objective of training a policy

TABLE I
POLICY CONFIGURATION AND PPO HYPER-PARAMETERS.

Parameter	Value
Number of environment steps per training batch	24
Learning epoch per training batch	5
Number of mini-batches per training batch	4
Reward discount factor, γ	0.99
GAE discount factor, λ	0.95
PPO clipping	0.2
Value loss coefficient	1.0
Entropy coefficient	0.01
Episode length	20s
Actor hidden layers	[512, 256, 128]
Critic hidden layers	[512, 256, 128]
Activation function	elu

TABLE II
SKATING SPECIFIC CUSTOM REWARDS

Term	Equation
$r_{\text{a\_rate}}$	$\sum_{i=1}^{N_{\text{actions}}} \begin{cases} 0.5(\Delta a_i)^2 & \text{if } \lvert\Delta a_i\rvert < 1 \\ \lvert\Delta a_i\rvert - 0.5 & \text{otherwise} \end{cases}$
r_{support}	$\sum_{i \in \mathcal{J}} \lvert q_i - q_{0,i}\rvert$
r_{still}	$\sum_{j \in \mathcal{J}} \left(\lvert q_i - q_{0,i}\rvert \text{ if } c = 0\right)$
$r_{\text{j\_mirror}}$	$\frac{1}{K}\sum_k^K \max\left(0, \lVert q_{k,1} - q_{k,2}\rVert - \xi\right)$
$r_{\text{step\_height)}}^{\text{B/W}}$	$\sum_i^M \left(p_{f,i}^{z,B/W} - h_{target}\right)^2 \tanh\left(\alpha \left\lVert v_{f,i}^{xy,B/W}\right\rVert\right)$
r_{straight}	$\left(1 - \lvert g_z^B\rvert\right)^2$
r_{tilt}	$\left(\arctan\frac{g_y^B}{g_z^B} - \arctan\frac{v_x \omega_z}{g}\right)^2$
$r_{\text{track\_lin\_vel}}$	$\exp\left(-\frac{\lVert \mathbf{v}_{\text{cmd}}^{xy} - \mathbf{v}_{\text{curr}}^{xy}\rVert_2^2}{\sigma_v^2}\right) \cdot \exp\left(-\frac{\sum_{j \in \mathcal{J}} \lvert \tau_i \dot{q}_i\rvert}{\sigma_e\left(mg \cdot \lVert v_{\text{curr}}^{xy}\rVert_2 + \epsilon\right)}\right)$

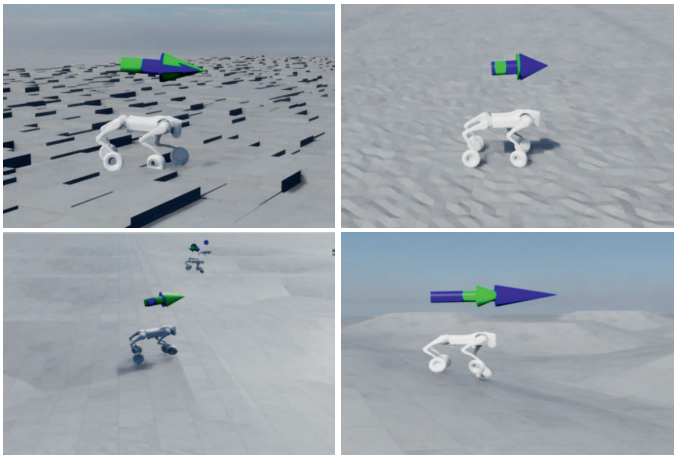

Fig. 5. The skating sub-environments: boxed, rough, uphill, downhill (top left to bottom right). Those environments are used for both training and sim validation of the trained models.

that maps proprioceptive observations to joint actions, enabling a quadruped robot to follow desired skating velocity commands while maintaining stability. The MDP framework is structured by: state space, action space, reward and termination functions, and domain randomization, each designed to align with the capabilities of Isaac Lab and enable an efficient sim-to-real transfer.

Each leg has three drivable joints, therefore the action vector $a_t \in A$ is given by the joint target positions for the agent at time t, $a_t \in \mathbb{R}^{12}$. The observation vector, $o \in \Omega$, is defined at time t as $o_t = [\omega_b, g_b, c, q_{rel}, \dot{q}_{rel}, a_{t-1}] \in \mathbb{R}^{45}$, where $\omega_b \in \mathbb{R}^3$ is the angular velocity in the robot frame, $g_b \in \mathbb{R}^3$ is the projected gravity vector in the robot frame, $c \in \mathbb{R}^3$ is the velocity command (desired x/y linear velocity, z angular velocity), $q_{rel} \in \mathbb{R}^{12}$ is the joint positions relative to the default offsets, $\dot{q}_{rel} \in \mathbb{R}^{12}$ is the relative joint velocities and $a_{t-1} \in \mathbb{R}^{12}$ is the previous action. The critic expands on the observations with privileged information by adding wheel position and velocity observation ($q_{rel} \in \mathbb{R}^{16}$ and $\dot{q}_{rel} \in \mathbb{R}^{16}$) and by adding linear velocity observations ($v_b \in \mathbb{R}^3$), forming $o_t^c = [v_b, \omega_b, g_b, c, q_{rel}, \dot{q}_{rel}, a_{t-1}] \in \mathbb{R}^{56}$.

The policy trains using a set of 21 rewards, expanding on the RSL-RL framework's root and joint penalties [21] with skating specific rewards, covered in Table II. Notably, the first reward equation, r_{a_rate}, represents the action rate computation through Huber Loss. This implementation provided us better control over the smoothness of the skating actions. Different gait characteristics imply different terrain traversal capabilities, with the following rewards used for shaping it. In some skating gaits, such as the tripod and gallop skating gaits, at least two legs are required to keep a stance while one or two legs generate the pushing motion required for propulsion. r_{support} penalizes the joints for leaving the default stance position. As the hip flexion/extension and the knee can act as a light suspension while skating, they are penalized less compared to the hip abduction/adduction. The reward function is similar to r_{still} which is used for penalizing not staying still with a zero command. r_{j_mirror} introduces threshold mirroring (ξ) to selected joint pairs, which can be desirable for achieving certain gaits. Notably, in our implementation, the mirroring is implemented in K groups of 2. The swing height is controlled by $r_{step\_height}^{B/W}$, which can be used either in the world or body frame, and is only activated when the target command is at least 0.1 m/s and the robot did not fall over. When $\{v_x \neq 0, v_y = 0, \omega_z \neq 0\}$, r_{tilt} rewards inclining into the curve, following the centripetal force, while r_{straight} rewards keeping a straight pose in all the other cases. Efficient skating minimizes the cost-of-transport by alternating between acceleration and cruising periods. To achieve this the linear velocity tracking reward, $r_{\text{track\_lin\_vel}}$, was formulated as the product of 2 exponentials. The first exponential penalizes deviations from the desired velocity, promoting precise motion, while the second exponential penalizes high actuation effort relative to the kinetic scale ($mg \cdot \lVert v_{\text{curr}}^{xy}\rVert_2$), incentivizing energy-efficient cruising and smooth acceleration bursts. Both σ_v and σ_e need tuning for balancing the 2 exponents.

D. Sim-to-real Transfer

To create a more robust skating policy and narrow the sim-to-real gap, noise was added during training to the observations and the ground friction was randomized within a set of margins. Additionally, sudden velocity changes were added to

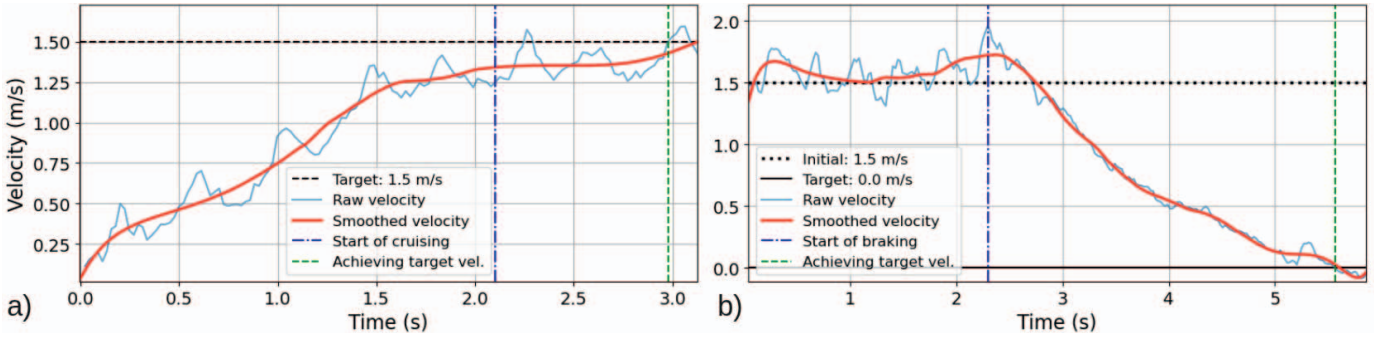

Fig. 6. a) Acceleration and b) deceleration graphs recorded in IsaacSim for validating the velocity control abilities of the trained skating policy. The cruising period refers to a more energy efficient motion where the robot slowly adjusts or maintains the velocity. Braking is achieved by actively locking and unlocking the mobility of the system using the front wheels (no active braking mechanism used).

the base to simulate external forces perturbing the robot. When skating on dirt roads or pebbles, rapid lateral slip is commonly encountered. Based on the ground friction and robot mass, a force was applied perpendicular to the wheels mimicking the slip force overcoming the ground friction, $F > \mu N$, as seen in Eq. 1. The force is applied, based on a set probability, p_{slip}, for short time instances, allowing for tuning. Additionally, as real-world control loops present slight delay, states from the previous time step are set for the current state based on a tunable probability [40].

$$F_{slip} \in \left(\mu\frac{mg}{4}, 1.2\mu\frac{mg}{4}\right], \quad \text{for} \quad t_{slip} \in [0.1; 0.25] \qquad (1)$$

When deployed on the real robot, the observation vector o_t is constructed from proprioceptive IMU and joint encoder readings, with the previous action being loaded from memory. The actions applied to the robot are converted into joint torques with a PD controller with fixed gains. Joint specific scaling factors and torque limits are applied to ensure safety and stability. The ROS2 Humble based low-level control loop is executed at 200Hz, while the policy inference loop is runs every 4 control steps (decimation of 4) at 50Hz.

IV. Experiments & Results

A. Simulation results

Before transferring the policy to the real robot, its performance, behaviour and efficiency need to be tested and evaluated on different terrain types and experiments.

The velocity control of the simulated robot is evaluated for maintaining constant velocity, acceleration to the trained maximum velocity, and rapid deceleration to a full stop. The linear velocity readings for these are graphed in Fig. 6. When accelerating from a standstill, a velocity of 1.2 m/s was reached in 2.01 s, with the target velocity of 1.5 m/s achieved in 2.97 s. Deceleration to a full stop from 1.5 m/s was achieved in 3.26 s. In all of those experiments, the simulated robot successfully maintained its balance and and target direction, with an average command accuracy of 95%. Across five independent training runs with different random seeds, the mean final cumulative reward was 45.44±2.84 (mean and standard deviation over seeds), corresponding to a variance of 8.07, which indicates consistent convergence.

Next, we iteratively increased the target velocity during simulated deployment with the aim of finding the experimental maximum speed. It should be noted that the velocity ranges during training for the linear velocity were $v_x = [-1.0, 1.5]$ and $v_y = [-1.0, 1.0]$. By doing this, a maximum average velocity of 2.16 m/s and a maximum peak velocity of 2.46 m/s were achieved, with a 0 to 1.6 m/s acceleration time of 3.41 s and 0 to peak acceleration time of 8.9 s. While the simulated results proved the possibility of achieving such speeds, the resultant gait tended to stumble and struggle to maintain a straight line, behaviour which appears at command velocities over 1.9 m/s and becomes common at over 2.1 m/s. We theorize that the achieved maximum speed is higher relative to the maximum training speed as a result of the agent learning to control the decent while going downhill and the periodic velocity characteristic to skating. With proper tuning and a more targetted reward set, the feasibility of reaching such speeds could increase.

The simulated CoT for the platform was measured at different speeds selected to match the speeds used in other related works [6], [16], [31], [32] and other increments up to 2 m/s, which are tabulated in Table III (Average CoT over 100 m). For all the speeds, the simulated CoT of our robot was lower compared to the benchmark platforms. We also computed the CoT ignoring the acceleration period and it was noticed that the locomotion becomes more efficient once cruising begins at the target speed, with an approximate observed drop of 18%. While the simulated CoT does not tell the entire story for the actual real world performance, it is worth examining as it is one of the main drivers during policy training.

Our policy's off-road capabilities were tested, successfully going up-hill on 20% slopes and downhill on 40% slopes. Additionally, the robot was tolerant to high terrain noise presented as steps, bumps and holes. In our 50 experiment runs, the robot did fall, presenting a stable gait. Fig. 5 presents the simulated skating robot in the different rough terrains.

The learned policy shows a hybrid skating gait that combines the behaviour of many standard legged gaits. During

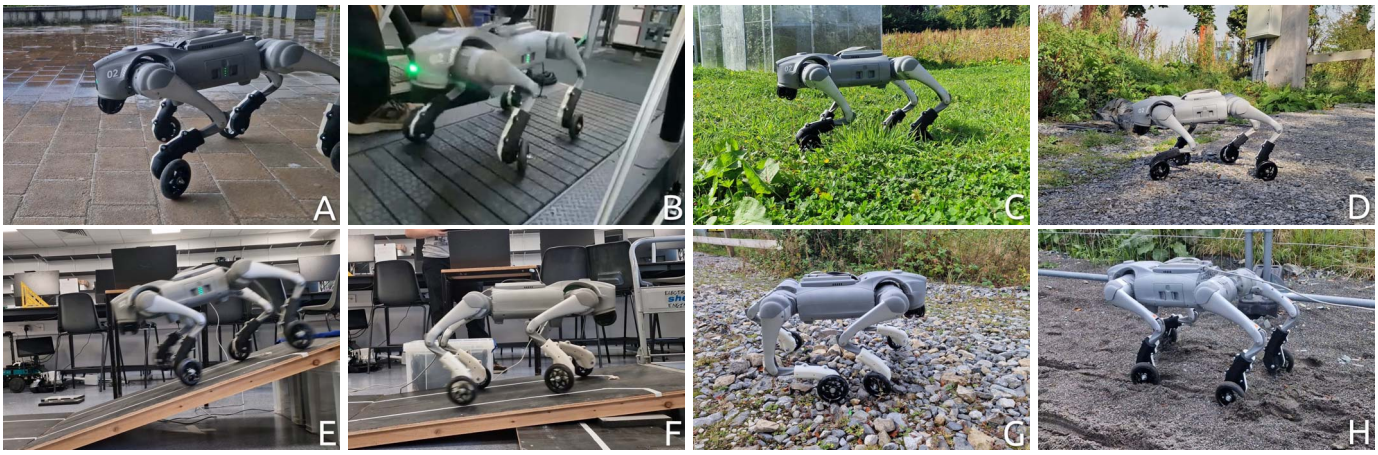

Fig. 7. The experiments for our skating quadruped in different environments: a) pavement, b) treadmill CoT tests, c) grass, d) gravel, e) going downhill (30%) f) going uphill (10%), g) rough hardcore, and h) sand.

TABLE III
THE SIMULATED COT AVERAGE AT DIFFERENT VELOCITIES. THE RESULTS ARE COMPUTED AS THE AVERAGE OF FIVE 100M RUNS.

Velocity	Simulated Cost-of-Transport
0.343 [6]	0.086 ± 0.018
0.4 [17], [31]	0.096 ± 0.041
0.83 [16]	0.108 ± 0.005
1.0 [32]	0.091 ± 0.016
1.5	0.098 ± 0.026
2.0	0.099 ± 0.033

forward and curved forward trajectories, a merged tripod-trot gait is formed, while in all the other cases a trot-like gait is formed, which is expected due to the lack of wheel steering. At higher command velocities, a gallop-like pattern forms in which the hind legs remain in constant ground contact. Notably, in both the gallop-like and the merged tripod-trot gaits, the hind wheels maintain near constant ground contact, stabilizing the gait.

An ablation study was conducted to evaluate the contribution of the proposed reward terms to the formed behaviour (Table II). First, a policy was trained without any of the proposed rewards, relying on the default rewards provided in IsaacLab's RSL-RL implementation. This baseline produced a more erratic gallop-like gait with a reduced peak velocity of 1.7 m/s. The off-road capabilities were significantly degraded, as the policy failed to complete the full curriculum. The resulting simulated CoT also increased to 0.184 ± 0.045 (measured at 1 m/s, averaged over five runs with different seeds). We attribute this increase to the more erratic and highly actuated gait, particularly at lower speeds. As this configuration serves as the baseline, the proposed reward structure demonstrates improvements in off-road robustness, reduction of CoT, and promotion of a more structured gait for velocity control. The improved gait also enables higher speeds, as the reward design constrains the search space towards a valid stable gait.

To further validate these findings, we re-enabled the gait specific rewards while removing only r_{a_rate} and $r_{track\_lin\_vel}$, replacing the latter with the standard velocity tracking term. Under this configuration, the initial skating gait forms, however the actuation increased, with greater joint travel observed. Stride and swing phases were shortened, resulting in an increased step frequency relative to the full policy. Off-road performance was preserved, but CoT increased to 0.12 ± 0.031, together with a slight reduction in stability, with occasional stumbling and rare falls. Interestingly, the maximum average speed increased to 2.21 m/s, with a peak velocity of 2.53 m/s. We hypothesize that this increase, while small, is due to the action smoothing and CoT-based velocity tracking constraining high speed behaviour in preference of stability and energy efficiency. With further reward tuning or a conditional reward formulation, the reduction in maximum speed induced by these terms could be limited. Upon selectively reintroducing r_{a_rate} and $r_{track\_lin\_vel}$, we observed that both terms contribute to reductions in CoT and peak velocity.

B. Real robot results

As with every sim-to-real transfer, tuning was required. For our transfer, we used a PD controller on the joint commands with action scaling and clipping. The time required to tune our skating Go2 was considerably longer and harder relative to tuning the standard Go2, but the transfer was successful.

We recorded the platform's speed on flat ground and on a treadmill, with the maximum safe velocity being 1.5 m/s. We achieved 1.6 m/s, however the gait stability was poor and the robot struggled keeping a straight line. Notably, the maximum speed matches the maximum speed of the active wheeled version, Unitree Go2-W, and exceeds the standard trot gait of the Unitree Go2 of around 1.42 m/s (not in running mode). When the target velocity was under around 0.15 m/s the robot faced challenges maintaining the speed, either struggling to start off, or overshooting considerably.

The cost of transport of our robot was evaluated over a distance of 100 m, calculated from the mechanical energy consumption of the controllable joints using joint torques τ_i and

TABLE IV
COT METRICS FOR DIFFERENT SKATING AND NON-SKATING PLATFORMS AT DIFFERENT VELOCITIES (M/S) OVER 100 M. OUR SKATING ROBOT ACHIEVED THE LOWEST RECORDED COT AT 0.83 M/S AND THE HIGHEST REPORTED SPEED AMONG SKATING QUADRUPEDS AT 1.5 M/S.

S/N	Platform	Vel (m/s)	CoT
S	Anymal [6]	0.342	0.121 ± 0.031
S	QSkater [17]	0.4	0.136 ± 0.028
S	QSkater-E [31]	0.4	0.18 ± 0.02
S	Skating Go1 [16]	0.83	0.36
S	Our Skating Go2	0.278	0.1033 ± 0.0184
		0.342	0.0931 ± 0.0080
		0.4	0.0722 ± 0.0029
		0.83	$\mathbf{0.0715 \pm 0.0048}$
		1.0	0.0772 ± 0.0106
		1.25	0.1155 ± 0.0616
		1.39	0.1295 ± 0.0497
		1.5	$\mathbf{0.1757 \pm 0.0611}$
N	Unitree Go2	0.83	0.3525 ± 0.0103
		1.25	0.5294 ± 0.0950
		1.42	0.5574 ± 0.1153

angular velocities $\dot{q}_i$ (Eq.2). Results from three full treadmill runs at different speeds are reported in Table IV, covering both the capabilities of our platform and speeds reported in the literature. For all trials below 1.39 m/s, our robot achieved lower CoTs than any previously reported skating robots, with the most efficient velocity observed at 0.83 m/s (3.0 km/h). At this optimal velocity, our skating platform demonstrated a 40.9% reduction in CoT compared to other skating quadrupeds and 70.9% decrease relative to the standard Go2. At 1.39 and 1.5 m/s, the CoTs remained competitive with benchmark platforms while achieving and sustaining significantly higher speeds. Relative to the standard Go2 at 1.42 m/s, we observe a 76.7% drop in CoT, while if we compare the maximum speeds of both gaits, we observe a 68.5% reduction. Together, these results highlight the efficiency gains of skating over walking and the hybrid platform's ability to sustain competitive efficiency at substantially higher speeds. Additionally, the results reinforce that RL-trained policies achieve a lower CoT compared to model-based controllers reported in the literature as they learn to exploit the natural dynamics and passive compliance of the system, discovering energy-efficient coordination strategies that are difficult to obtain through manual design.

$$\text{CoT} = \frac{\int \sum_{i=1}^{12} |\tau_i \dot{q}_i| dt}{mgd} \tag{2}$$

The robot's terrain capabilities were measured on pavement, grass, gravel, sand, hardcore and ramps. The robot can be seen skating through all those in Fig. 7. The robot successfully traversed all terrain types, however for pavement, grass and gravel the speed had to be reduced based on the terrain to around 0.6 to 1.0 m/s otherwise the robot would trip forwards and fall. The ramp performance diminished with the transfer to the real robot, with an upwards slope of 10% and a downhill incline of 30%. For the downhill experiment, the decreased direction and velocity control due to the speed achieved resulted deciding that any higher incline was a safety risk. The 10% slope is the absolute maximum slope achieved, however the ascension was difficult. An 8% slope is more achievable for normal use, while any higher than 10% would result in the robot stumbling and falling over. For the gravel, grass and hardcore (all uneven), the robot successfully traversed them with no notable observations other than the decrease of the target speed. Finally, the sand experiment was successfully traversed, but starting took longer as grip had to be established. Despite continuous impact and loading, the skates and moulds exhibited only superficial scratches, with no structural damage. Overall, we deem the off-road performance successful for light rough terrain.

Finally, the overall gait performance is evaluated. During hardware experimentation, the same gait patterns observed in simulation transferred successfully to the real robot, preserving their characteristic stability and hind-leg structure. Based on gait stability, dynamic response, and experimental results, it is demonstrated that the implementation achieved strong real-world performance. Mechanically, mobility locking via fixed wheel orientations (30° outwards for the front legs and 0° for the hind legs) enabled effective locomotion without active braking or steerable wheels, resulting in a mechanically simple and lightweight system that remains suitable for extension to other low-friction environments such as ice or snow. As traversed slopes and velocities increase, an active braking system may be required to further enhance performance.

V. CONCLUSION

This paper presents a novel end-to-end RL strategy for training highly dynamic systems, such as skating quadrupeds. Our model was successfully transferred to a real skating quadruped, achieving and maintaining speeds of up to 1.5 m/s, equal to its active wheeled counterpart, and, to our knowledge, the highest speed achieved on a skating quadruped of this class. Additionally, our policy obtained a cost-of-transport 70.9% lower than the standard Go2, and, at the most efficient speed, 40.9% lower compared to other skating quadrupeds. The robot successfully traversed both flat and light rough terrain, which demonstrates the feasibility of skating for urban environments. One limiting factor in the rough terrain is stumbling due to tripping as a result of blind locomotion. By adding exteroceptive sensing, such as vision, and transformable legs or a braking system, robots could leverage cruising and increase their travel range through skating, which could be applicable in delivery robots and other applications where range and scarce charging points are considerations. Adding braking would also increase skating safety around humans and in risky environments. While our policy was designed for skating, we believe that many concepts could be transferred to other configurations to achieve a lower cost of transport or low friction terrain traversal (snow/ice).

VI. ACKNOWLEDGEMENTS

The authors would like to thank the reviewers for their time and valuable insights, Maynooth University for providing the equipment and funding, and Gregory Heinrich Mantel for his help in the CAD development.

REFERENCES

[1] M. Staniaszek, T. Flatscher, J. Rowell, H. Niu, W. Liu, Y. You, R. Skilton, M. Fallon, and N. Hawes, "AutoInspect: Towards Long-Term Autonomous Industrial Inspection," Apr. 2024.

[2] M. Hutter, C. Gehring, D. Jud, A. Lauber, C. D. Bellicoso, V. Tsounis, J. Hwangbo, K. Bodie, P. Fankhauser, M. Bloesch, R. Diethelm, S. Bachmann, A. Melzer, and M. Hoepflinger, "ANYmal - a highly mobile and dynamic quadrupedal robot," in *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 38–44, Oct. 2016.

[3] M. Kulkarni, M. Dharmadhikari, M. Tranzatto, S. Zimmermann, V. Reijgwart, P. De Petris, H. Nguyen, N. Khedekar, C. Papachristos, L. Ott, R. Siegwart, M. Hutter, and K. Alexis, "Autonomous Teamed Exploration of Subterranean Environments using Legged and Aerial Robots," in *2022 International Conference on Robotics and Automation (ICRA)*, pp. 3306–3313, May 2022.

[4] M. Fiorucci, G. Schillaci, M. Tannous, L. Bianchi, L. Salusti, S. Cioncolini, A. Politano, and G. De Magistris, "Evaluating the Robustness of Autonomous Inspections in the Energy Industry with a Quadruped Robot," in *ADIPEC*, OnePetro, Oct. 2023.

[5] Z. Bi, K. Chen, C. Zheng, Y. Li, H. Li, and J. Ma, "Interactive Navigation for Legged Manipulators with Learned Arm-Pushing Controller," Mar. 2025.

[6] M. Bjelonic, C. Dario Bellicoso, M. Efe Tiryaki, and M. Hutter, "Skating with a Force Controlled Quadrupedal Robot," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 7555–7561, Oct. 2018.

[7] J. Chen, K. Xu, and X. Ding, "Roller-Skating of Mammalian Quadrupedal Robot With Passive Wheels Inspired by Human," *IEEE/ASME Transactions on Mechatronics*, vol. 26, pp. 1624–1634, June 2021.

[8] L. Yang, Y. Yin, F. Gao, Z. Wang, L. Wang, Q. Sun, and H. Gao, "Design and Control of a Novel Six-Legged Skating Robot With Skateboards," *IEEE/ASME Transactions on Mechatronics*, pp. 1–12, 2023.

[9] M. Geilinger, S. Winberg, and S. Coros, "A Computational Framework for Designing Skilled Legged-Wheeled Robots," *IEEE Robotics and Automation Letters*, vol. 5, pp. 3674–3681, Apr. 2020.

[10] Y. Pan, R. A. I. Khan, C. Zhang, A. Zhang, and H. Shang, "Pegasus: A Novel Bio-inspired Quadruped Robot with Underactuated Wheeled-Legged Mechanism *," in *2024 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 1464–1469, May 2024.

[11] W.-T. Chen, E.-C. Tsui, W.-S. Yu, and P.-C. Lin, "Indoor and Outdoor Multi-Terrain Stair-Climbing Robot Design," in *2025 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 14419–14425, May 2025.

[12] Y.-S. Shen, W.-S. Yu, and P.-C. Lin, "Contact Force Estimation for a Leg-Wheel Transformable Robot With Varying Contact Points," in *2025 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 684–690, May 2025.

[13] M. Bjelonic, R. Grandia, O. Harley, C. Galliard, S. Zimmermann, and M. Hutter, "Whole-Body MPC and Online Gait Sequence Generation for Wheeled-Legged Robots," in *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 8388–8395, Sept. 2021.

[14] I. Belli, M. P. Polverini, A. Laurenzi, E. M. Hoffman, P. Rocco, and N. Tsagarakis, "Optimization-Based Quadrupedal Hybrid Wheeled-Legged Locomotion," July 2021.

[15] M. Arnold, L. Hildebrandt, K. Janssen, E. Ongan, P. Bürge, Á. G. Gábriel, J. Kennedy, R. Lolla, Q. Oppliger, M. Schaaf, J. Church, M. Fritsche, V. Klemm, T. Tuna, G. Valsecchi, C. Weibel, M. Wüthrich, and M. Hutter, "LEVA: A High-Mobility Logistic Vehicle with Legged Suspension," in *2025 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 7438–7444, May 2025.

[16] J.-Y. Cho and J.-Y. Kim, "Learning-Based Dynamic Roller Skating for Quadruped Robots Using Passive Wheels," in *Korean Society of Mechanical Engineers 2024 Conference*, pp. 1403–1404, Nov. 2024.

[17] J. Chen, K. Xu, R. Qin, and X. Ding, "Locomotion Control of Quadrupedal Robot With Passive Wheels Based on CoI Dynamics on SE(3)," *IEEE Transactions on Industrial Electronics*, vol. 71, pp. 7551–7560, Aug. 2023.

[18] A. J. Hung, C. E. Adu, and T. Y. Moore, "SKOOTR: A Skating, Omni-Oriented, Tripedal Robot," in *2025 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 15921–15928, May 2025.

[19] G. Endo and S. Hirose, "Study on Roller-Walker — Improvement of Locomotive Efficiency of Quadruped Robots by Passive Wheels," *Advanced Robotics*, vol. 26, pp. 969–988, May 2012.

[20] G. B. Margolis and P. Agrawal, "Walk These Ways: Tuning Robot Control for Generalization with Multiplicity of Behavior," Dec. 2022.

[21] N. Rudin, D. Hoeller, P. Reist, and M. Hutter, "Learning to Walk in Minutes Using Massively Parallel Deep Reinforcement Learning," Aug. 2022.

[22] H. Fu, K. Tang, P. Li, G. Deng, and C. Chen, "Multi-agent Reinforcement Learning with Hybrid Action Space for Free Gait Motion Planning of Hexapod Robots," in *8th Annual Conference on Robot Learning*, Sept. 2024.

[23] S. Ha, P. Xu, Z. Tan, S. Levine, and J. Tan, "Learning to Walk in the Real World with Minimal Human Effort," Nov. 2020.

[24] J. Shi, C. Bai, H. He, L. Han, D. Wang, B. Zhao, M. Zhao, X. Li, and X. Li, "Robust Quadrupedal Locomotion via Risk-Averse Policy Learning," Sept. 2023.

[25] J. Di Carlo, P. M. Wensing, B. Katz, G. Bledt, and S. Kim, "Dynamic Locomotion in the MIT Cheetah 3 Through Convex Model-Predictive Control," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 1–9, Oct. 2018.

[26] S. Kajita, F. Kanehiro, K. Kaneko, K. Fujiwara, K. Harada, K. Yokoi, and H. Hirukawa, "Biped walking pattern generation by using preview control of zero-moment point," in *2003 IEEE International Conference on Robotics and Automation (Cat. No.03CH37422)*, vol. 2, pp. 1620–1626 vol.2, Sept. 2003.

[27] J. Pratt, P. Dilworth, and G. Pratt, "Virtual model control of a bipedal walking robot," in *Proceedings of International Conference on Robotics and Automation*, vol. 1, pp. 193–198 vol.1, Apr. 1997.

[28] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal Policy Optimization Algorithms," Aug. 2017.

[29] T. Haarnoja, A. Zhou, P. Abbeel, and S. Levine, "Soft Actor-Critic: Off-Policy Maximum Entropy Deep Reinforcement Learning with a Stochastic Actor," Aug. 2018.

[30] J. Tan, T. Zhang, E. Coumans, A. Iscen, Y. Bai, D. Hafner, S. Bohez, and V. Vanhoucke, "Sim-to-Real: Learning Agile Locomotion For Quadruped Robots," May 2018.

[31] J. Chen, R. Qin, L. Huang, Z. He, K. Xu, and X. Ding, "Unlocking Versatile Locomotion: A Novel Quadrupedal Robot with 4-DoFs Legs for Roller Skating," in *2024 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 8037–8043, May 2024.

[32] G. Chen, Q. Meng, and Y. Lin, "Cooperative skating motion control for a quad-wheel-legged robot," *Results in Engineering*, vol. 27, p. 106792, Sept. 2025.

[33] G. B. Margolis, G. Yang, K. Paigwar, T. Chen, and P. Agrawal, "Rapid locomotion via reinforcement learning," *The International Journal of Robotics Research*, vol. 43, pp. 572–587, Apr. 2024.

[34] H. Liu, S. Teng, B. Liu, W. Zhang, and M. Ghaffari, "Discrete-Time Hybrid Automata Learning: Legged Locomotion Meets Skateboarding," Apr. 2025.

[35] G. Bellegarda and K. Byl, "Training in Task Space to Speed Up and Guide Reinforcement Learning," in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 2693–2699, Nov. 2019.

[36] G. Bellegarda, K. van Teeffelen, and K. Byl, "Design and Evaluation of Skating Motions for a Dexterous Quadruped," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 1703–1709, May 2018.

[37] M. Geilinger, R. Poranne, R. Desai, B. Thomaszewski, and S. Coros, "Skaterbots: Optimization-based design and motion synthesis for robotic creatures with legs and wheels," *ACM Transactions on Graphics*, vol. 37, pp. 160:1–160:12, July 2018.

[38] J. Chen, K. Xu, H. Ma, and X. L. Ding, "Motion characteristics of human roller skating," *Biology Open*, p. bio.037713, Jan. 2019.

[39] M. Mittal, C. Yu, Q. Yu, J. Liu, N. Rudin, D. Hoeller, J. L. Yuan, R. Singh, Y. Guo, H. Mazhar, A. Mandlekar, B. Babich, G. State, M. Hutter, and A. Garg, "Orbit: A Unified Simulation Framework for Interactive Robot Learning Environments," *IEEE Robotics and Automation Letters*, vol. 8, pp. 3740–3747, June 2023.

[40] S. Chamorro, V. Klemm, M. de La Iglesia Valls, C. Pal, and R. Siegwart, "Reinforcement Learning for Blind Stair Climbing with Legged and Wheeled-Legged Robots," in *2024 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 8081–8087, May 2024.