Interpretable Soil-Robust Bipedal Locomotion on Tilled Soils with Material Curriculum and a Stance-Gated Slip Reward

Anhar Risnumawan[*†], Achmad Fahrul Aji[*‡], Yoshimasa Nakamura[*],
Teuku Zikri Fatahillah[*], Hasan Imaduddin[*], Naoyuki Kubota[*]
[*]*Department of Mechanical Systems Engineering, Graduate School of Systems Design*
Tokyo Metropolitan University, 6-6 Asahigaoka, Hino, Tokyo 191-0065, Japan
e-mail: {risnumawan-anhar, achmad-aji, nakamura-yoshimasa1, teuku-fatahillah, hasan-imaduddin}@ed.tmu.ac.jp, kubota@tmu.ac.jp
[†]*Mechatronics Engineering Division*, *Politeknik Elektronika Negeri Surabaya*, Indonesia
[‡]*Department of Electrical Engineering*, *Politeknik Negeri Semarang*, Indonesia

***Abstract*—Autonomous systems for precision agriculture increasingly support crop monitoring, targeted spraying, harvesting, and in-field soil assessment, yet the wheeled and tracked platforms used most often can lose traction on tilled soils and exacerbate soil compaction. We study bipedal locomotion under this soil uncertainty using reinforcement learning with per-episode sampling of friction and compliant-contact parameters. Training follows a multi-stage curriculum that progressively expands the friction and stiffness ranges and employs a slip-averse objective that is gated to loaded stance by a small normal-force threshold. We introduce a fuzzy soil-state model that maps the sampled parameters to traction and support membership degrees and a fuzzy soil difficulty index that summarizes curriculum progression in interpretable terms. We also present a smooth stance-membership relaxation of the hard contact gate used in the slip penalty, yielding a continuous measure of stance confidence. Simulation experiments demonstrate reduced stance-phase slip, improved stability, and better disturbance recovery relative to standard baselines under the tested conditions, while the fuzzy formulation provides an interpretable bridge between soil mechanics, curriculum design, and reward shaping.**

I. INTRODUCTION

The growing demand for autonomous field operations in precision agriculture has renewed interest in field-deployable mobile robots. Most deployed agricultural systems remain wheeled or tracked, which provides throughput and stability but can limit terrain adaptability and soil preservation on soft ground [1], [2]. These platforms can struggle in tilled fields where traction varies with moisture, compaction, and surface preparation. Legged systems, and bipeds in particular, can step across discontinuities, negotiate furrows and wheel ruts, and temporally concentrate contact forces to reduce compaction. Realizing these advantages requires locomotion that remains stable when friction and support stiffness change unpredictably across a field. Fig. 1 shows our biped in a simulated agricultural scenario.

While bipedal robots have demonstrated robust mobility on rocky surfaces, stairs, and sandy beaches [3], [4], [5], tilled agricultural soils remain comparatively underexplored. The combination of variable traction, variable compliance, and structured surface features creates failure modes that differ from the rigid or uniformly compliant surfaces commonly studied in legged robotics.

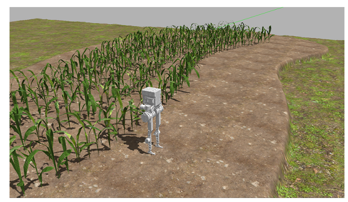


Fig. 1: Hunter biped in a Gazebo-simulated agricultural field with crop rows, exhibiting variable friction, wet or dry, and compliance typical of tilled soils.

Tilled soils exhibit strong spatial and temporal variation in friction and compliance due to moisture, tillage depth, and compaction history [6]. Foot soil interactions can range from shallow deformation to substantial sinkage, and consecutive footsteps can encounter markedly different support, which makes rigid-body assumptions inadequate. Furrows, wheel-track compaction, and patchy preparation further couple geometric obstacles with material variation [7], requiring policies to adapt simultaneously to topography and substrate properties in agricultural settings.

Soil conditions are often communicated with linguistic descriptors such as *loose*, *packed*, or *slick*. We employ fuzzy sets to map sampled material parameters to interpretable traction and support memberships, which provides an auditable description of curriculum progression and clarifies where slip penalties apply. This fuzzy interface bridges terramechanics-inspired parameterization, curriculum design, and reward shaping through interpretable traction descriptors without introducing a separate soil classifier [8].

Current approaches to robust legged locomotion demonstrate impressive capabilities but leave gaps that are pronounced on tilled agricultural soils. Learning-based methods often rely on broad domain randomization and treat slip as an emergent effect rather than an explicit objective [9]. Existing curricula commonly escalate geometric complexity, while the transition from high-traction to low-traction substrates is less directly targeted [4]. Slip-aware controllers frequently detect and respond after slip occurs [10], which can be inadequate for bipeds on wet, slip-prone fields where stability margins are small. These considerations motivate proactive, material-aware training strategies that anticipate and reduce excessive slippage under the tested conditions.

We train a Proximal Policy Optimization (PPO) locomotion policy with a multi-stage curriculum that expands friction and compliance ranges and combines them with a stance-gated slip objective. This framework improves traction management while preserving an interpretable description of the training soil conditions. Our contributions are as follows.

- We propose a material curriculum over friction and compliance to address slip and sinkage on tilled soils.
- We introduce a stance-gated slip penalty that focuses learning on loaded contacts and sharpens reward attribution.
- We define a fuzzy soil-state interpretation of sampled parameters through traction and support memberships and a Mamdani-style difficulty index for auditable curriculum descriptions.
- We present a smooth relaxation of the stance gate that recovers the hard threshold used in our experiments and provides continuous stance confidence.

Simulation results show lower stance-phase slip, better stability, and stronger disturbance recovery than standard baselines under the tested conditions.

II. RELATED WORK

Curriculum learning and robustness training have improved legged locomotion by gradually increasing task difficulty, most commonly through geometric complexity such as obstacle height and surface irregularity [5], [4]. Complementary strategies rely on broad domain randomization to enable transfer across deformable and granular terrains by varying contact and terrain parameters during training [9], [11], and terramechanics-inspired models have been incorporated to better represent deformable ground interactions [12]. While these approaches provide strong generalization, slip and sinkage are often treated as emergent outcomes rather than as objectives explicitly tied to material properties that vary with moisture and tillage in agricultural fields.

Slip-aware locomotion has been addressed primarily through reactive detection and mitigation, including proprioception-based slip estimation with subsequent impedance or gait adjustments [13], [10]. Related bio-inspired designs and biomechanical analyses emphasize compliance and coordination for stability on uneven terrain [14], [15].

Fuzzy systems offer interpretable tools for reasoning under terrain and traction uncertainty using linguistic categories and graded membership. Prior work introduced fuzzy traversability indices and safety modulation rules for navigation and risk-aware control [16], [17], and fuzzy knowledge-based approaches have been applied to soil and agricultural traction to predict and regulate slip under uncertain ground conditions [8], [18]. More recent studies explored fuzzy contact reasoning for contact and sliding detection [19] and fuzzy reward representations for explainable or more stable reinforcement learning [20], [21]. Building on this literature, we use fuzzy sets as an interpretability layer for describing sampled soil parameters and for providing a graded stance membership used in slip-averse reward shaping, while the control policy remains a standard PPO policy trained from data.

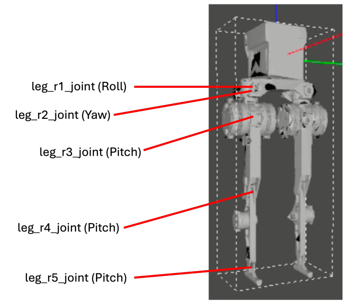


Fig. 2: Hunter biped platform used in this study.

III. METHODOLOGY

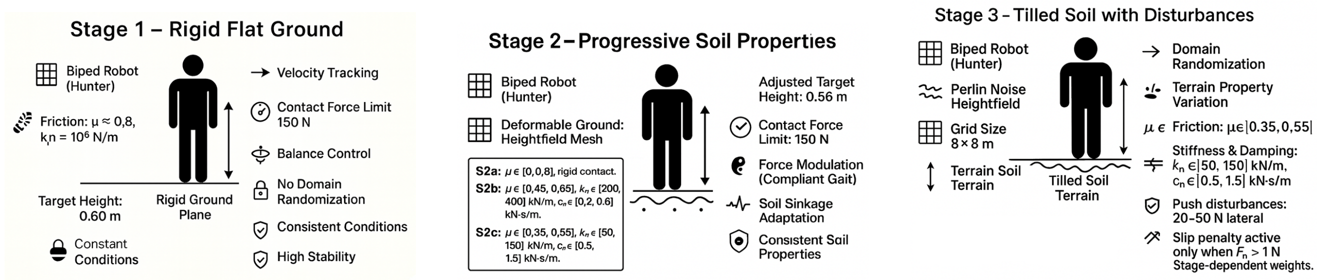
We develop a multi-stage reinforcement learning actor-critic pipeline for bipedal locomotion on tilled agricultural soils that explicitly targets traction management through material-property progression and a contact-gated slip-averse objective.

A. Robot Model and Task

We train a Hunter biped robot to track commanded forward velocities over tilled soil using proprioceptive feedback, task command inputs, and PD-assisted torque control. The robot model comprises a 10-degree-of-freedom lower body with 5 joints per leg: a 3-DOF hip, 1-DOF knee, and 1-DOF ankle. Each joint employs position-controlled servos with uniform torque limits of 23.7 N·m to reflect realistic actuator constraints. The observation space encompasses joint states (positions $\mathbf{q} \in \mathbb{R}^{10}$ and velocities $\dot{\mathbf{q}} \in \mathbb{R}^{10}$), base orientation and angular velocity, base linear velocity, binary contact signals from each foot, and the command inputs $(\mathbf{v}_{cmd,xy}, \omega_{cmd,z})$ used by the tracking objective. The action space consists of normalized joint position targets $\mathbf{a} \in [-1, 1]^{10}$ commanded at 50 Hz (control timestep $dt = 0.020$ s). These targets are maintained for 4 physics substeps while a joint-space PD controller operates at 200 Hz (simulation timestep `sim_dt` $= 0.005$ s), with output torques clipped to the 23.7 N·m limit. The PD gains are specifically tuned for compliant, slip-prone soil conditions: proportional gains K_p of 76, 80, and 60 N·m/rad for hip, knee, and ankle joints respectively, with corresponding derivative gains K_d of 2.6, 2.7, and 2.0 N·m·s/rad. Joint limits prevent self-collision while constraining joint velocities to 30.1 rad/s. Fig. 2 illustrates the Hunter platform configuration and contact points used for proprioceptive sensing. We intentionally exclude external sensors such as cameras to isolate the effects of slip-aware learning from perception-based approaches.

B. Terrain and Soil Modeling

We model tilled soil in NVIDIA Isaac Sim (PhysX) using compliant normal contact and Coulomb friction. Here, μ denotes the friction coefficient, k_n denotes the normal



(a) Stage 1: Rigid Foundation (b) Stage 2: Progressive Tilled Soil (c) Stage 3: Tilled Soil with Disturbances

Fig. 3: Stage environments used throughout the curriculum. The subfigures show S1 on a rigid foundation, S2 with progressive tilled-soil conditions, and S3 with lateral disturbances applied on the S2c terrain distribution in Isaac Sim.

stiffness, and c_n denotes the normal damping. The normal force follows a linear spring–damper (Kelvin–Voigt) model, while lateral forces follow Coulomb friction with separate static and dynamic coefficients. During the soil curriculum stages (S2a–S3), we sample the material parameters at the beginning of each episode as $\mu \sim \mathcal{U}(\mu_{\min}^{(s)}, \mu_{\max}^{(s)})$, $k_n \sim \mathcal{U}(k_{\min}^{(s)}, k_{\max}^{(s)})$ kN/m, and $c_n \sim \mathcal{U}(c_{\min}^{(s)}, c_{\max}^{(s)})$ kN·s/m, where s denotes the curriculum stage defined in the multi-stage curriculum design subsection. Unless otherwise noted, PhysX uses SI units, so we pass k_n in N/m and c_n in N·s/m. For readability, we report stiffness and damping in kN/m and kN·s/m.

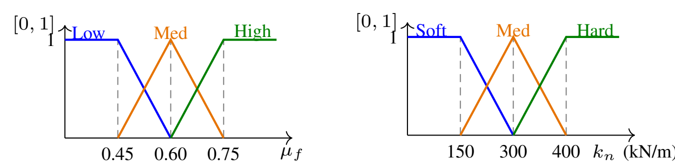
For tilled soils, friction varies strongly with moisture and has been reported to range approximately from $\mu = 0.3$ for wet soil to $\mu = 0.8$ for dry soil [6]. Across stages S2a–S3, the sampled parameters span $\mu \in [0.35, 0.80]$, $k_n \in [50, 400]$ kN/m, and $c_n \in [0.2, 1.5]$ kN·s/m. Stage S1 uses a rigid-contact setting with fixed $\mu \approx 0.8$ and $k_n = 10^6$ N/m. We instantiate the normal model as $F_n = k_n\,\delta + c_n\,\dot{\delta}$, where δ denotes the normal penetration depth and F_n is reported in newtons (N). We map the sampled friction to the static and dynamic coefficients as $\mu_s = \mu$ and $\mu_d = 0.8\,\mu$ unless stated otherwise.

In practice, we randomize the robot-soil contact by applying the sampled friction to the robot foot material, while the terrain-side friction is set by the selected terrain preset. PhysX uses a "min" combine mode for friction, so the effective coefficient is $\min(\mu_{\text{foot}}, \mu_{\text{terrain}})$. This preserves consistent terrain presets across scenes while enabling per-episode variation of the contact pair. The sampled parameter ranges were chosen to cover wet-to-dry loam and loose-to-packed tilled soil while balancing realism and RL stability, consistent with soft-soil simplifications in prior work [12], [11]. Terrain geometry includes randomized furrow depths (5–15 cm) and spacing (0.8–1.2 m), wheel-track compaction bands, and low-μ patches covering 10–30% of the surface to emulate field heterogeneity following tillage [7].

C. Fuzzy Soil-State Modeling

In each episode, we sample soil contact parameters including the friction coefficient μ and the compliant normal-contact parameters (k_n, c_n). We then introduce a fuzzy soil-state layer that maps these continuous variables to overlapping linguistic descriptors of traction and support. This layer provides an interpretable view of the existing parameter randomization and curriculum without modifying the PPO policy architecture or adding new observations, which is inline with prior fuzzy terrain and safety models for robotic navigation [16], [17].

1) Traction and support membership functions: In this study, traction refers to the friction available at foot-ground contact, while support refers to the normal stiffness of the soil. We denote the episode-level friction coefficient assigned to the foot material by μ_f, with $\mu_f \equiv \mu$ in Sec. III-B. This parameter represents the commanded foot-side friction in each episode. The actual traction at contact, however, also depends on the terrain friction coefficient μ_{terrain}. Because PhysX uses a "min" friction combine mode, the effective contact coefficient becomes $\mu_{\text{eff}} = \min(\mu_f, \mu_{\text{terrain}})$. Therefore, the realized traction can be lower than μ_f whenever the terrain preset or localized low-μ patches impose a smaller μ_{terrain}.



(a) Traction (b) Support

Fig. 4: Piecewise-linear fuzzy membership functions used for traction and support.

Fig. 4 shows the traction and support membership functions. For support, the breakpoints are chosen to match the curriculum stiffness bands used in this study. We denote the corresponding membership degrees by $\mu^{i}_{\text{trac}}(\mu_f)$ and $\mu^{j}_{\text{sup}}(k_n)$, where $i \in \{\text{Low}, \text{Med}, \text{High}\}$ and $j \in \{\text{Soft}, \text{Med}, \text{Hard}\}$. The damping parameter c_n can be fuzzified in the same manner, for example into under-damped, nominal, and over-damped sets. In this paper, however, we use k_n as the main proxy for support because it directly determines the penetration-to-force relation in the Kelvin–Voigt normal model.

2) Fuzzy soil difficulty rule base and index: We summarize joint traction and support conditions through a fuzzy soil difficulty output $D \in \{\text{Easy}, \text{Moderate}, \text{Hard}\}$. This output is computed with a Mamdani-style rule base, using $\min(\cdot)$ for

AND and $\max(\cdot)$ for aggregation. The formulation follows fuzzy traversability methods that compress multiple terrain attributes into a single interpretable score [16].

In this study, the rule base follows an intuitive monotonic relation between traction, support, and difficulty. Low traction leads to Hard difficulty across support levels. High traction combined with medium or high support leads to Easy difficulty. Intermediate cases map to Moderate, capturing conditions with either moderate traction or limited support.

For a rule indexed by traction set i and support set j, the firing strength is

$$\alpha_{ij} = \min\left(\mu^{i}_{\text{trac}}(\mu_f), \mu^{j}_{\text{sup}}(k_n)\right). \tag{1}$$

Let $(\gamma_{\text{Easy}}, \gamma_{\text{Mod}}, \gamma_{\text{Hard}})$ denote the aggregated activations for the three consequents. Each activation is obtained by taking the maximum over all rules that lead to the corresponding output.

We then report a scalar difficulty index $d \in [0.2, 0.8]$ by singleton defuzzification:

$$d = \frac{0.2\,\gamma_{\text{Easy}} + 0.5\,\gamma_{\text{Mod}} + 0.8\,\gamma_{\text{Hard}}}{\gamma_{\text{Easy}} + \gamma_{\text{Mod}} + \gamma_{\text{Hard}} + \varepsilon} \tag{2}$$

Here $\varepsilon > 0$ is a small constant for numerical safety. This index provides an interpretable description of curriculum progression. Episodes with lower μ_f and lower k_n yield higher d, which is consistent with the staged parameter ranges used in the curriculum.

3) Fuzzified stance gate for slip penalty: The slip penalty in Eq. (6) is activated only when a foot bears load, using a normal-force threshold F_{th}. We express this stance condition through a fuzzy membership $\mu_{\text{stance}}(F_n) \in [0, 1]$ and write the generalized form

$$R_{\text{slip}} = -w_{slip} \sum_{f \in \{\text{feet}\}} ||\mathbf{v}_{\tan,f}||\, \mu_{\text{stance}}(F_{n,f}). \tag{3}$$

A smooth, memoryless stance membership that relaxes the hard gate is the logistic function

$$\mu_{\text{stance}}(F_n) = \frac{1}{1 + \exp(-\beta(F_n - F_{\text{th}}))}, \tag{4}$$

with slope parameter $\beta > 0$. As $\beta \to \infty$, $\mu_{\text{stance}}(F_n)$ approaches the crisp indicator $\mathbb{1}_{F_n > F_{\text{th}}}$, which recovers the stance-gated formulation used in this study.

This fuzzy interpretation clarifies the design intent because the penalty weight is modulated by a degree of stance confidence rather than by a discontinuous switch. Related fuzzy membership approaches have been used for contact and sliding detection on humanoids, which supports the feasibility of continuous stance confidence estimates for legged systems [19].

D. Multi-Stage Curriculum Design

We stage learning from rigid, high-traction conditions to increasingly deformable and slip-prone soils before introducing external disturbances. Stage S1 uses rigid flat ground ($\mu \approx 0.8$, effectively rigid contact with $k_n = 10^6$ N/m) to bootstrap the gait. Stage S2a then introduces packed soil with $\mu \in [0.6, 0.8]$.

TABLE I: Reward terms and stage-specific settings used across the curriculum. Stage 3 retains the S2c reward schedule and adds lateral pushes through the environment.

Term / setting	**S1**	**S2a–S2c**
Velocity tracking R_{track}	Active	Active
Base motion penalties ($w_{lin\_z}, w_{ang\_xy}, w_{orient}$)	$(2.0, 2.0, 2.0)$	$(1.0, 1.0, 1.0)$
Torque penalty w_{torq}	10^{-5}	10^{-4}
Smoothness penalties	$\|\ddot{\mathbf{q}}\|^2$, $\|\Delta\mathbf{a}\|^2$	Off
Air-time reward	$w_{air} = 1.0$, $t_{\text{target}} = 0.5$ s, $t_{\text{cap}} = 0.2$ s	Same as S1
Base height	$z_{\text{des}} = 0.60$ m, $w_{height} = 10.0$	$z_{\text{des}} = 0.56$ m, $w_{height} = 4.0$
Joint-limit penalty w_{limits}	10.0	10.0
Foot orientation penalty w_{ori}	2.0	0.5
Slip penalty	Inactive	$w_{slip} = 0.5$, $F_{\text{th}} = 1$ N
Contact-force penalty	Inactive	$w_{force} = 0.05$, $F_{\max} = 150$ N

Stage S2b broadens the material variation to moderately tilled soil with $\mu \in [0.45, 0.65]$ and increased compliance. Stage S2c represents fully tilled soil with $\mu \in [0.35, 0.55]$ and $k_n \in [50, 150]$ kN/m under full randomization. Stage S3 keeps the S2c terrain distribution and adds lateral pushes of 20–50 N to improve disturbance robustness. This progression follows established curriculum learning principles [4].

Stage changes are initiated manually between runs. We monitor episode returns, slip rates, and stability metrics, then start the next stage with the latest saved checkpoint once performance has stabilized under the current conditions.

E. Reward Formulation

The reward is a weighted sum of standard locomotion terms and two soil-specific contact terms. To keep the presentation compact, Table I lists the active terms and stage-specific weights. In brief, the policy is rewarded for commanded linear and yaw velocity tracking and penalized for unstable base motion, orientation error, excessive torque, unsafe joint excursions, large contact forces, and stance-phase slip. Stage 1 emphasizes gait bootstrapping on rigid ground, Stages 2a–2c retain the tracking and stability terms while enabling soil-specific penalties for compliant terrain, and Stage 3 keeps the S2c weights while adding lateral pushes.

Linear and angular tracking are shaped by an exponential reward,

$$R_{\text{track}} = \exp\left(-\frac{\|\mathbf{e}_v\|^2}{\sigma_{lin}}\right) + \exp\left(-\frac{e_\omega^2}{\sigma_{ang}}\right). \tag{5}$$

Here $\mathbf{e}_v = \mathbf{v}_{cmd,xy} - \mathbf{v}_{base,xy}$ denotes planar velocity error and $e_\omega = \omega_{cmd,z} - \omega_{base,z}$ denotes yaw-rate error. We use $\sigma_{lin} = \sigma_{ang} = 0.25$, so a velocity error of 0.5 m/s or 0.5 rad/s yields $\exp(-1) \approx 0.37$. This form provides near-maximum reward near the commanded velocity while decaying smoothly as tracking error grows.

Our main soil-specific term is the stance-gated slip reward

$$R_{\text{slip}} = -w_{slip} \sum_{f\in\{\text{feet}\}} \|\mathbf{v}_{\tan,f}\| \mathbb{1}_{F_{n,f}>F_{\text{th}}} \tag{6}$$

where $\mathbf{v}_{\tan,f}$ is the tangential velocity of foot f and $F_{n,f}$ is its normal ground-reaction force. More explicitly, $\mathbf{v}_{\tan,f} = \mathbf{v}_{foot,f} - (\mathbf{v}_{foot,f} \cdot \mathbf{n}_f)\,\mathbf{n}_f$ (m/s) is the tangential component of foot velocity along the local contact plane, with $\mathbf{n}_f$ the local surface normal, and $\mathbb{1}_{F_{n,f}>F_{\text{th}}}$ is active only when the foot is in loaded stance. This term is therefore inactive during swing, so it does not penalize intended foot motion and improves reward attribution for true stance-phase slip. Recent deep reinforcement learning studies have also used fuzzy reward components to provide graded feedback during curriculum training, which supports the broader view that smoother shaping can improve learning stability when hard thresholds would otherwise yield brittle signals [21].

We complement this term with a contact-force penalty,

$$R_{\text{contact}} = -w_{force} \sum_{f} \max(F_{n,f} - F_{\max}, 0). \tag{7}$$

In the soil stages we use $w_{slip} = 0.5$, $w_{force} = 0.05$, $F_{\text{th}} = 1$ N, and $F_{\max} = 150$ N. For the 12.667 kg Hunter robot, the nominal per-foot load in double support is approximated as $F_{\text{foot}} \approx mg/2 = 12.667\times 9.81/2 \approx 62.1$ N, while the nominal single-support load is $mg = 12.667 \times 9.81 \approx 124.2$ N. Hence $F_{\text{th}} = 1$ N is only about 1.6% of the nominal per-foot stance load and provides a conservative loaded-contact detector. We set $F_{\max} = 150$ N, which places the cap about 26 N above nominal single-support loading, so it mainly penalizes impacts or aggressive single-leg loading while leaving typical stance forces unpenalized. These thresholds were tuned for our platform and can be scaled for robots with different body masses or contact-sensing characteristics.

The remaining terms are standard gait-shaping and safety regularizers: tracking reward, base-motion and orientation penalties, torque penalty, bounded air-time reward, base-height tracking, and penalties on joint-limit violations and undesirable foot orientation. Base-motion and orientation penalties suppress excessive vertical motion, roll, and pitch, while the torque term discourages unnecessarily large actuator effort. The air-time reward remains capped by t_{cap} so it encourages consistent stepping without rewarding excessively long aerial phases. The S1 smoothness terms $\|\ddot{\mathbf{q}}\|^2$ and $\|\Delta\mathbf{a}\|^2$ penalize joint acceleration and step-to-step action changes, respectively, which helps bootstrap a regular gait on rigid ground. These terms are disabled once the policy enters deformable-soil stages, where larger corrective adjustments become necessary.

The base-height target is reduced from 0.60 m on rigid terrain to 0.56 m in soil stages to accommodate the modest sinkage expected on compliant ground. The foot-orientation penalty is also softened beyond S1 so the feet can adapt to uneven support without incurring excessive cost, whereas the joint-limit penalty remains strong to prevent unsafe excursions near the hardware range of motion. Table I summarizes the full weight schedule, including the S1-only smoothness terms and the reuse of the S2c reward setting in Stage 3.

F. Policy Training Algorithm

The PPO configuration uses separate actor and critic MLPs with three hidden layers of sizes [512, 256, 128] and ELU activations. The actor network takes actor observations o_t^{actor} as input and outputs the Gaussian policy mean with dimension equal to the robot action space. The action standard deviation is a learned, state-independent parameter. The critic network takes critic observations o_t^{critic} and outputs a scalar value estimate. Key PPO settings follow standard practice: clip parameter ϵ=0.2, discount γ=0.99, GAE λ=0.95, and entropy coefficient 0.01. Rollout collection and optimization are configured through Hydra. By default the number of parallel environments is N=4096 and the per-environment rollout length is T=24 steps. The number of PPO epochs per iteration is E=5. Minibatching uses a configured number of mini-batches M, which implies an effective minibatch size of $\frac{NT}{M}$. The total samples per iteration equal $N \times T$.

The policy is trained sequentially across S1 $\rightarrow$ S2a $\rightarrow$ S2b $\rightarrow$ S2c $\rightarrow$ S3, with each stage using the previous checkpoint and stage-specific parameters. Advancement is manual based on monitoring learning curves and predefined targets, without automatic threshold-based progression.

We vary soil properties across curriculum stages and runs within the uniform ranges defined in Section III-B, which provides coverage over the target domain while avoiding the substantial cost of high-fidelity granular simulation [11], [12].

IV. EXPERIMENTS

All experiments use NVIDIA Isaac Sim 4.2 with a 200 Hz physics rate and a 50 Hz control rate (four physics substeps per action). Runs use Ubuntu 22.04 with an NVIDIA RTX 6000 Ada GPU and CUDA 12.9. Our simulator stack builds on HumanoidVerse [22], extended with the Hunter robot model and the proposed soil curriculum. For the Gazebo agricultural field we use an open-source model [23].

We evaluate four training variants: *Ours (Full)* with the complete curriculum, *Smaller-Net (Full)* with an MLP [256, 128, 64], *No-Curriculum* trained directly on S3, and *Flat-Only (S1)* trained only on S1. Unless otherwise stated, evaluation uses three seeds with 100 episodes of 60 s per seed. We report stance-phase slip (m/100 m), falls (falls/100 m), pitch variability σ_{pitch} (deg), specific mechanical work ("CoT"), and forward velocity tracking error, averaging metrics across seeds and reporting seed-level standard deviation where shown.

As shown in Table II, Ours (Full) reduces slip distance by 95.5% relative to Smaller-Net (Full) and by over an order of magnitude relative to Flat-Only (S1) and No-Curriculum. Under the same condition, Ours (Full) achieves zero falls while the ablated baselines exhibit nonzero fall rates. Ours (Full) also improves cost of transport and velocity tracking error relative to the tested baselines.

TABLE II: Performance on tilled soil without disturbance. Values are computed per seed from 100 episodes and then averaged across three seeds. "CoT" denotes specific mechanical work (J/kg·m) and is not normalized by g.

Method	**Slip Distance (m/100 m)**	**Fall Rate (falls/100 m)**	**Stability (σ_{pitch}, deg)**	**CoT (J/kg·m)**	**Velocity Error (m/s)**
Tilled Soil without Disturbance					
Ours (Full)	10.92	0.00	0.55 ± 0.02	0.18 ± 0.00	0.098 ± 0.005
Smaller-Net (Full)	244.00	1.73	0.46 ± 0.03	0.32 ± 0.10	0.184 ± 0.002
Flat-Only (S1)	407.92	74.86	9.71 ± 0.85	4.25 ± 2.86	0.451 ± 0.025
No-Curriculum	1037.45	54.28	4.73 ± 0.33	3.99 ± 0.71	0.237 ± 0.013

The poor transfer of Flat-Only (S1) supports the hypothesis that staged exposure to friction and compliance is necessary for traction-aware behaviors. Fig. 5 reports representative training traces for each stage. In S1, the tracking reward increases rapidly and episode length approaches the horizon, which is consistent with learning a stable gait on rigid ground. In S2, randomized friction and compliance increase variance in tracking and stability and activate the slip penalty, followed by gradual recovery as the policy adapts to the broader soil range. In S3, external pushes further perturb training, yet the traces indicate partial recovery in episode length and slip penalty under the tested settings. This progression shows a friction-cone interpretation of stance control, where the policy selects contact placements and push directions that keep $\|\mathbf{F}_t\| \leq \mu F_n$, with $\mathbf{F}_t$ and F_n the tangential and normal ground-reaction components.

Per-term reward traces indicate smooth acquisition of foundational gait components across stages. Base height tracking saturates early as upright posture is mastered, while tracking and energy terms improve gradually without premature plateaus. Torque penalties decrease early and increase slightly in later stages, showing a tradeoff between stronger step impulses for slip mitigation and strict energy minimization.

Table III reports S3 evaluation on fully tilled soil with lateral pushes. Ours (Full) achieves the lowest falls and slip, while Smaller-Net (Full), No-Curriculum, and Flat-Only degrade in transfer. These trends suggest that staged exposure and sufficient model capacity are important for traction-aware contact strategies on tilled soils.

To limit gradient spikes from sparse slip penalties at the onset of soil randomization, we keep the slip term inactive in S1 and activate it in S2 with a small force threshold ($F_{\text{th}} = 1$ N) and a moderate weight ($w_{\text{slip}} = 0.5$). Larger slip weights ($w_{\text{slip}} > 0.6$) yielded conservative gaits with lower speed and higher energy. Normalizing the slip penalty by F_n reduced over-penalization during light contacts, though we retain the simpler stance-gated form in Eq. (6) for clarity.

Despite reasonable in-distribution performance, failures still occur at very low out-of-distribution friction ($\mu = 0.25$), where the policy sometimes overextends steps during packed-to-loose soil transitions, causing large slips and occasional falls. This suggests terrain anticipation or slip estimation may improve safety margins.

Fig. 6(a)–(c) shows representative snapshots under Easy, Moderate, and Hard simulated soil conditions. Across these

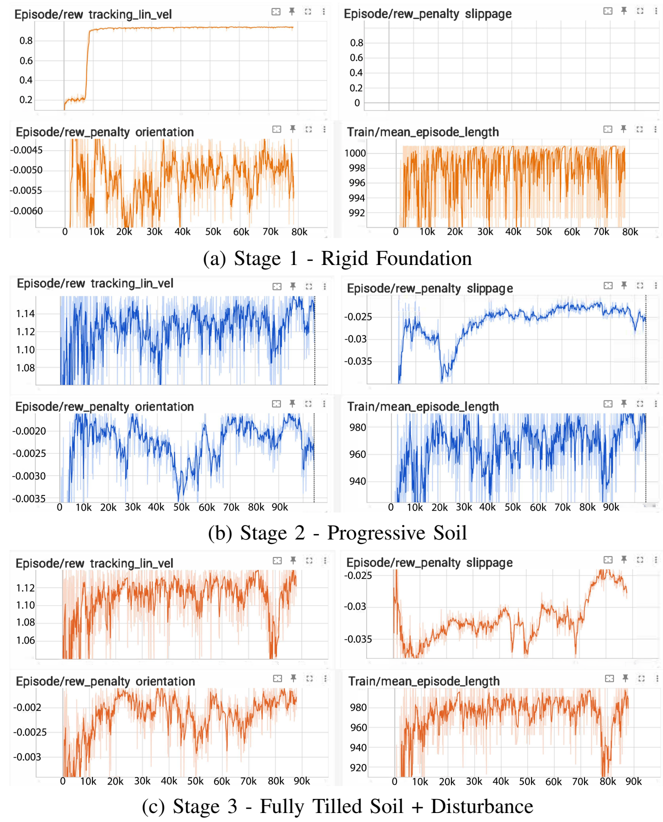


Fig. 5: Representative per-term training traces for each curriculum stage.

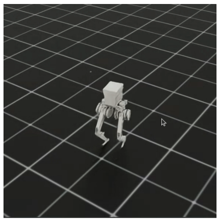
tested settings, the learned policy maintains forward progression and upright posture while adopting more conservative stance and body motions on harder surfaces. Fig. 6(d) illustrates a planned intermediate validation step on the physical Hunter robot using an indoor setup with soft carpet, surface dots, and a glossy finish to approximate compliant, uneven, and slippery support, allowing safer and more repeatable analysis of failure modes before outdoor trials. Validation on real agricultural soil, together with on-robot soil estimation, closed-loop slip regulation, and soil models informed by field measurements, remains future work.

V. CONCLUSION

This work presents a training strategy for bipedal locomotion on agricultural soils that combines curriculum learning of soil properties with a contact-gated objective discouraging stance-phase slip. We also introduce a fuzzy soil-state interpretation of the sampled material parameters that provides interpretable traction and support memberships, a compact difficulty index for curriculum description, and a smooth stance-membership relaxation that recovers the hard contact gate used in our experiments. Results demonstrate reduced slip, improved stability, and lower fall rates compared to baselines, suggesting that explicit slip mitigation shapes contact timing and force direction consistent with traction limits under the tested conditions.

TABLE III: Evaluation on tilled soil with disturbances. Values are averaged over three seeds with 100 episodes per seed. "Total Dist." and "Total Falls" are per-seed totals over 100 episodes. Slip is reported as meters per 100 meters traveled.

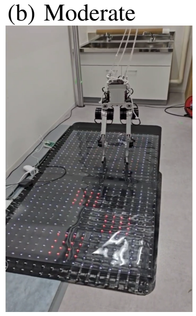
Setup	Total Dist. (m)	Total Falls	Avg Len (steps)	Avg Dist/Ep (m)	Falls/100 m	Slip (m/100 m)
Ours (Full)	1076.51	5	2852.3	10.77	0.46	12.69
Smaller-Net (Full)	512.17	20	2408.8	5.12	3.90	726.28
No-Curriculum	345.30	28	2169.0	3.45	8.11	300.01
Flat-Only	111.85	47	1606.4	1.12	42.02	885.69



(a) Easy

(b) Moderate

(c) Hard



(d) Future work

Fig. 6: Representative snapshots and a future work direction. Panels (a)–(c) show Easy, Moderate, and Hard simulated soil conditions, where the learned controller maintains locomotion under the tested settings. Panel (d) illustrates preliminary real-robot validation with Hunter as future work.

ACKNOWLEDGMENT

This work was supported by the Japan Science and Technology Agency (JST), Moonshot R&D Program, under grant number JPMJMS2034 and by the Indonesian Education Scholarship (BPI), Center for Higher Education Funding and Assessment (PPAPT), and the Indonesian Endowment Fund for Education (LPDP).

REFERENCES

[1] G. Reina, A. Milella, and R. Galati, "Terrain assessment for precision agriculture using vehicle dynamic modelling," *Biosystems Engineering*, vol. 162, pp. 124–139, 2017.

[2] S. Fountas, N. Mylonas, I. Malounas, E. Rodias, C. H. Santos, and E. Pekkeriet, "Agricultural robotics for field operations," *Sensors*, vol. 20, no. 9, p. 2672, 2020.

[3] H. Duan, B. Pandit, M. S. Gadde, B. Van Marum, J. Dao, C. Kim, and A. Fern, "Learning vision-based bipedal locomotion for challenging terrain," in *2024 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2024, pp. 56–62.

[4] R. P. Singh, M. Morisawa, M. Benallegue, Z. Xie, and F. Kanehiro, "Robust humanoid walking on compliant and uneven terrain with deep reinforcement learning," in *2024 IEEE-RAS 23rd International Conference on Humanoid Robots (Humanoids)*, 2024, pp. 497–504.

[5] B. van Marum, M. Sabatelli, and H. Kasaei, "Learning perceptive bipedal locomotion over irregular terrain," *arXiv preprint arXiv:2304.07236*, 2023.

[6] C. Hua, W. Zhang, H. Fu, Y. Zhang, B. Yu, C. Jiang, Y. Wei, Z. Chen, and X. Kuang, "The prediction method and application of off-road mobility for ground vehicles: A review," *World Electric Vehicle Journal*, vol. 16, no. 1, p. 47, 2025.

[7] O. A. Ani, B. B. Uzoejinwa, A. O. Ezeama, A. P. Onwualu, S. N. Ugwu, and C. J. Ohagwu, "Overview of soil-machine interaction studies in soil bins," *Soil and Tillage Research*, vol. 175, pp. 13–27, 2018.

[8] A. Rahman, A. Hossain, A. H. M. Z. Alam, and M. Rashid, "Fuzzy knowledge-based model for prediction of traction force of an electric golf car," *Journal of Terramechanics*, vol. 49, no. 1, pp. 13–25, 2012.

[9] J. Lee, J. Hwangbo, L. Wellhausen, V. Koltun, and M. Hutter, "Learning quadrupedal locomotion over challenging terrain," *Science Robotics*, vol. 5, no. 47, p. eabc5986, 2020.

[10] Y. Nisticò, S. Fahmi, L. Pallottino, C. Semini, and G. Fink, "On slip detection for quadruped robots," *Sensors*, vol. 22, no. 8, p. 2967, 2022.

[11] 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.

[12] G. Bhardwaj, S. Dasgupta, N. Sukavanam, and R. Balasubramanian, "Soft soil gait planning and control for biped robot using deep deterministic policy gradient," *arXiv preprint arXiv:2306.08063*, 2023.

[13] F. Jenelten, J. Hwangbo, F. Tresoldi, C. D. Bellicoso, and M. Hutter, "Dynamic locomotion on slippery ground," *IEEE Robotics and Automation Letters*, vol. 4, no. 4, pp. 4170–4176, 2019.

[14] A. A. Saputra, J. Botzheim, A. J. Ijspeert, and N. Kubota, "Combining reflexes and external sensory information in a neuromusculoskeletal model to control a quadruped robot," *IEEE Transactions on Cybernetics*, vol. 52, no. 8, pp. 7981–7994, 2021.

[15] H. Asai, T. Noda, and J. Morimoto, "Exploring the role of thigh muscle coordination with biarticular muscles for stable running," 2023, preprint (Research Square), November 2023.

[16] H. Seraji, "Fuzzy Traversability Index: A new concept for terrain-based navigation," *Journal of Robotic Systems*, vol. 17, no. 2, pp. 75–91, 2000.

[17] E. Tunstel, A. M. Howard, and H. Seraji, "Fuzzy Rule-Based Reasoning for Rover Safety and Survivability," in *Proceedings of the 2001 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2001, pp. 1413–1420.

[18] S. Soylu and K.
cCarman, "Fuzzy logic based automatic slip control system for agricultural tractors," *Journal of Terramechanics*, vol. 95, pp. 25–32, 2021.

[19] T. Xu, D. Xie, L. Zhu, and H. Ding, "Hierarchical Fuzzy-Based Contact and Sliding Detection for Humanoid Robots with Foot IMU," in *2025 IEEE-RAS 24th International Conference on Humanoid Robots (Humanoids)*. IEEE, 2025, pp. 1155–1162.

[20] E. Zander, B. van Oostendorp, and B. Bede, "Reinforcement learning with Takagi–Sugeno–Kang fuzzy systems," *Complex Engineering Systems*, vol. 3, no. 1, p. 9, 2023.

[21] C.-Y. Huang, A. Efendi, and Y.-C. Wang, "Curriculum-guided deep reinforcement learning with fuzzy rewards for autonomous push–grasp manipulation in cluttered environments," *Robotics and Autonomous Systems*, vol. 198, p. 105325, 2026.

[22] CMU LeCAR Lab, "Humanoidverse: A multi-simulator framework for humanoid robot sim-to-real learning," https://github.com/LeCAR-Lab/HumanoidVerse, 2025. [Online]. Available: https://github.com/LeCAR-Lab/HumanoidVerse

[23] J. Barthel, J. Bier, T. Friedel, T. Ruigrok, L. Tuschla, and R. van Essen, "Virtual maize field," 2023, rOS package to procedurally generate randomized fields with rows of plants for Gazebo. [Online]. Available: https://github.com/FieldRobotEvent/virtual_maize_field