Learning to Adapt: Bio-Inspired Gait Strategies for Versatile Quadruped Locomotion

Joseph Humphreys[1] and Chengxu Zhou[2]

Abstract—Deep reinforcement learning (DRL) has revolutionised quadruped robot locomotion, but existing control frameworks struggle to generalise beyond their training-induced observational scope, resulting in limited adaptability. In contrast, animals achieve exceptional adaptability through gait transition strategies, diverse gait utilisation, and seamless adjustment to immediate environmental demands. Inspired by these capabilities, we present a novel DRL framework that incorporates key attributes of animal locomotion: gait transition strategies, pseudo gait procedural memory, and adaptive motion adjustments. This approach enables our framework to achieve unparalleled adaptability, demonstrated through blind zero-shot deployment on complex terrains and recovery from critically unstable states. Our findings offer valuable insights into the biomechanics of animal locomotion, paving the way for robust, adaptable robotic systems.

Index Terms—Quadruped Locomotion; Bio-inspired Robotics; Deep Reinforcement Learning; Adaptive Gait Transition

Originally inspired by the remarkable adaptability of quadruped mammal locomotion—an ability shaped by innate and environmentally induced factors [1], [2], [3]—the field of quadruped robotics has made significant efforts to develop equally proficient locomotion frameworks. Presently, the most advanced systems rely on end-to-end deep reinforcement learning (DRL), which involves training a multilayer perceptron (MLP) [4] capable of navigating diverse environments. These frameworks demonstrate robustness in traversing real-world [5] and urban terrains [6], resisting perturbations [7], jumping between platforms [8], overcoming deformable surfaces [9], and recovering from falls [10]. Despite these achievements, their adaptability remains constrained, as most systems are limited to deploying a single targeted gait or locomotion strategy.

In contrast, biomechanics research has shown that no single gait is universally effective across all scenarios [11], [12], [13]. Animals adapt their locomotion by employing nominal gaits such as ambling, trotting, and running [14], while switching to specialised gaits like hopping, pronking, and bounding for off-nominal tasks such as predator evasion or obstacle navigation [15]. Current DRL frameworks fall short of replicating this level of versatility. To address this limitation, some approaches have focused on training DRL policies to learn multiple gaits by providing reference motions during training [16], [17], [18], [19], or by learning from policies that specialise in specific

gaits [20]. However, these methods remain insufficient when compared to the extensive capabilities observed in animal locomotion, which include:
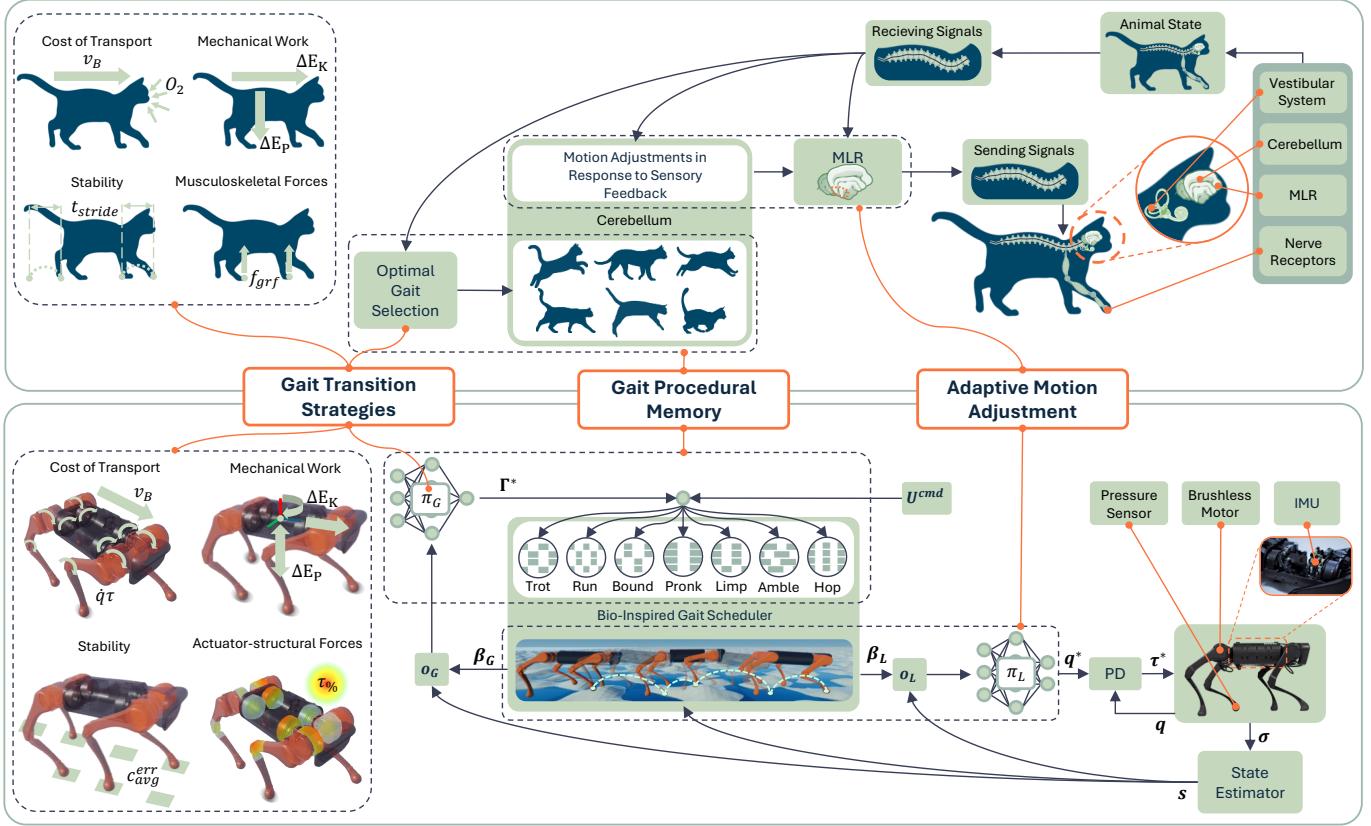
- Adaptation of gait style for optimal performance in response to challenging terrains and perturbations.
  - Enabled by advanced gait selection strategies.
- Rapid deployment of a diverse set of task- and state-specific gaits.
  - Attributable to gait procedural memory.
- Seamless deviation from nominal gait motions to address off-nominal contact states.
  - Achieved through precise motion adjustments tailored to the environment.

Although existing DRL frameworks have demonstrated progress in implementing learned gaits, none successfully integrate all three attributes simultaneously. This gap highlights the significant potential of biomechanics-inspired approaches to advance robotic locomotion.

While bio-inspired methods that leverage central pattern generators (CPGs) [21] have realised spontaneous gait transitions [22] and mimic certain animal behaviours [23], their performance in real-world applications is often limited. Typically, such experiments are constrained to controlled environments [24], [23] or, when conducted outdoors, are restricted by low velocities and simple tasks [18]. These limitations suggest that instead of attempting to replicate animal locomotion mechanisms precisely, state-of-the-art DRL frameworks should be augmented with high-level attributes derived from animal locomotion to instil the proficiency observed in nature.

Animal gait transition strategies, which contribute to optimal performance and enable navigation of challenging environments, are believed to emerge from the minimisation of metrics related to energy consumption [25], [26], [27], mechanical work [28], [29], [30], instability [31], [32], and musculoskeletal forces [33], [34], [35]. However, no singular metric has been definitively identified as the sole driver of these transitions. Instead, it is hypothesised that a combination of these factors influences gait transition strategies [31], [36], [37].

The concept of gait procedural memory, which facilitates the rapid deployment of a range of gaits, is thought to reside within the cerebellum of the animal brain. This region governs the coordination of limb movements for each gait learned by the animal [38], [39]. Similarly, adaptive motion adjustments—crucial for seamless adaptation to off-nominal contact states—are achieved through coordination between the mesencephalic locomotor region (MLR), which oversees

[1]School of Mechanical Engineering, University of Leeds, UK, LS2 9JT. el20jeh@leeds.ac.uk

[2]Department of Computer Science, University College London, UK, WC1E 6BT. chengxu.zhou@ucl.ac.uk

Fig. 1: **Instilling the core animal locomotion proficiency attributes within a DRL locomotion framework.** From taking an abstracted view of animal locomotion, the attributes of proficient locomotion are gait transition strategies, gait procedural memory, and adaptive motion adjustment. Taking on a similar structure, within the DRL locomotion framework these attributes are realised by the gait selection policy, π_G, bio-inspired gait scheduler, and locomotion policy, π_L. π_G has been trained to minimse the animal gait transition metrics applied to the quadruped robot based on the current robot state, s, and relevant bio-inspired gait scheduler output, β_G, to select the optimal gait, Γ^*. The bio-inspired gait scheduler then generates the gait references informed by s from encoded high-level gait parameters. The gait references, β_L, are then passed to π_L to inform of any adjustments to the nominal gait motions.

locomotion execution [40], and the cerebellum. These adjustments rely on sensory feedback to modify limb movements in response to the animal's current state [41].

Despite these insights, there has been no prior attempt to simultaneously integrate all these attributes—gait transition strategies, procedural memory, and adaptive motion adjustments—within a DRL locomotion framework. This leads to the following research questions:

1) How can the roles of the MLR and cerebellum inspire the augmentation of an end-to-end DRL locomotion policy to adapt to off-nominal contact states?
2) Can a DRL locomotion policy, inspired by gait procedural memory, learn to deploy a diverse set of gaits and perform rapid gait transitions?
3) How can metrics that characterise animal gait transition strategies be effectively leveraged within a DRL policy for optimal gait selection? Does the resulting behaviour align with that observed in animals?
4) Can the developed framework exhibit exemplary adaptability to traverse real-world terrains not encountered during training? What is the contribution of each metric

to this adaptability?

To address these questions, we propose a novel locomotion framework (see Fig. 1) designed to incorporate these key animal locomotion attributes. The framework demonstrates exceptional adaptability through zero-shot deployment in complex, real-world environments, relying solely on intra-perceptive sensors.

I. RESULTS

Within the framework, presented in Fig. 1, a gait selection policy, π_G, is trained for optimal gait selection through minimising gait transition metrics adopted from biomechanics to generate the output $\Gamma^* \in [0, 7]$ which maps to a specific gait within $[stand, trot, run, bound, pronk, limp, amble, hop]$. This selected gait, coupled with the velocity command within $U^{\text{cmd}} = [v_x^{\text{cmd}}, v_y^{\text{cmd}}, \omega_z^{\text{cmd}}, \Gamma^*] \in \mathbb{R}^4$, where v_x^{cmd}, v_y^{cmd} and ω_z^{cmd} are base velocities in x, y and yaw respectively, is passed to the bio-inspired gait scheduler (BGS) to generate gait references, as detailed in Section III-B. These gait references are contained within the BGS outputs β_L and β_G for the locomotion policy, π_L, and π_G respectively through inclusion
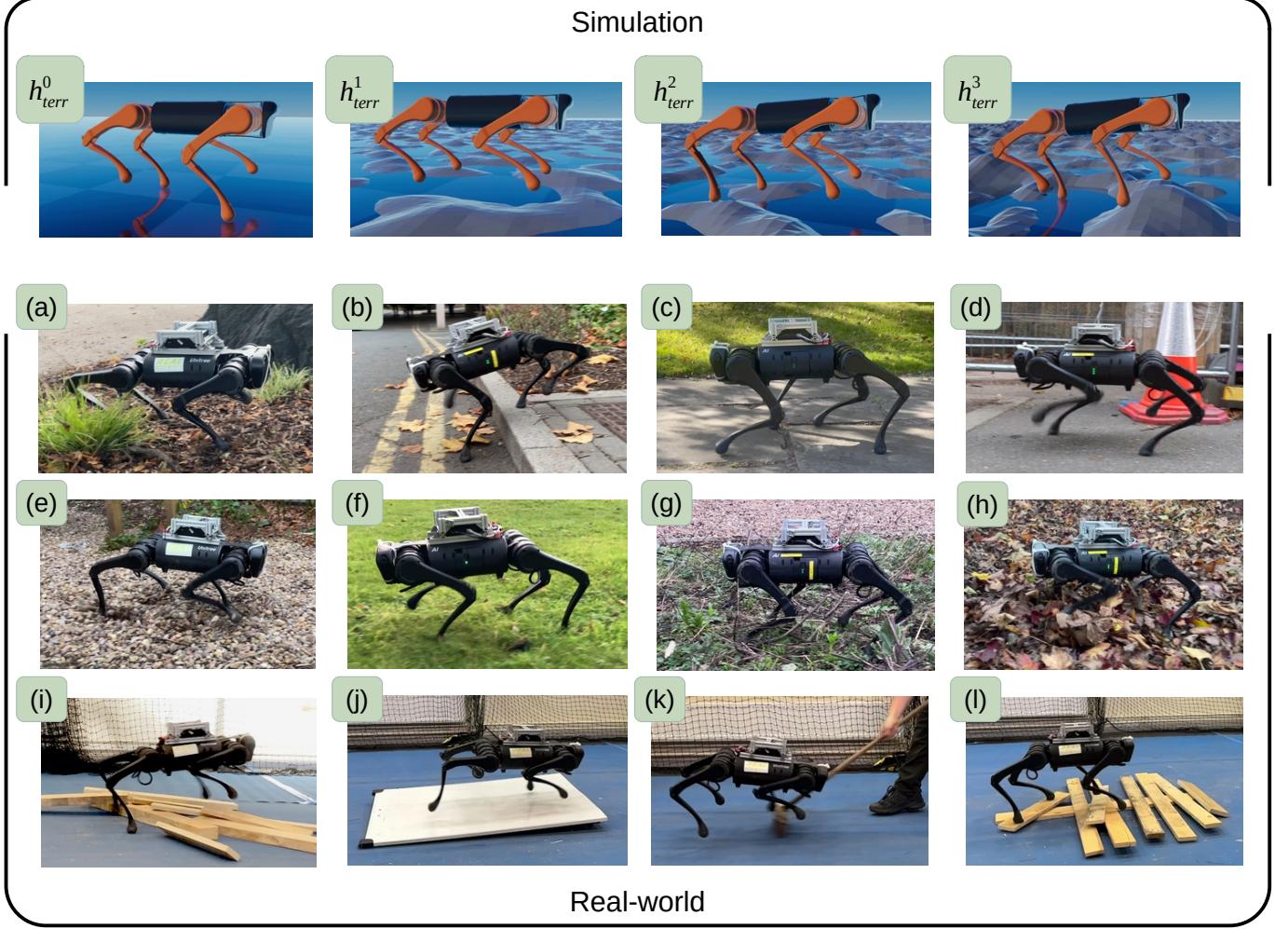
Fig. 2: **Snapshots of our framework being deployed in different environments.** Simulated terrains h_{terr}^0 through h_{terr}^3 are generated with fractal noise with maximum heights of 0 m, 0.06 m, 0.13 m and 0.2 m. To demonstrate the adaptability of our framework is transferable to the real world it has been deployed on (a) wood-chip, (b) a large step, (c) concrete slabs with large cracks, (d) tarmac, (e) deep rocks, (f) grassy terrain, (g) overgrown roots, (h) fallen leaves, (i) loose timber, (j) low-friction ramp, (k) flat terrain with perturbations, and (i) balanced timber.

within their observation vectors $\boldsymbol{o}_L$ and $\boldsymbol{o}_G$. In this respect, the BGS acts as pseudo gait procedural memory. The gait references are adjusted based on the robot's state, $\boldsymbol{s}$, generated by the state estimator (SE) from the sensor feedback vector, $\boldsymbol{\sigma}$, which in turn reflects the relationship between the cerebellum and MLR. To realise the output joint positions of π_L, $\boldsymbol{q}^*$, they are passed through a PD controller to generate joint torque commands $\boldsymbol{\tau}^*$.
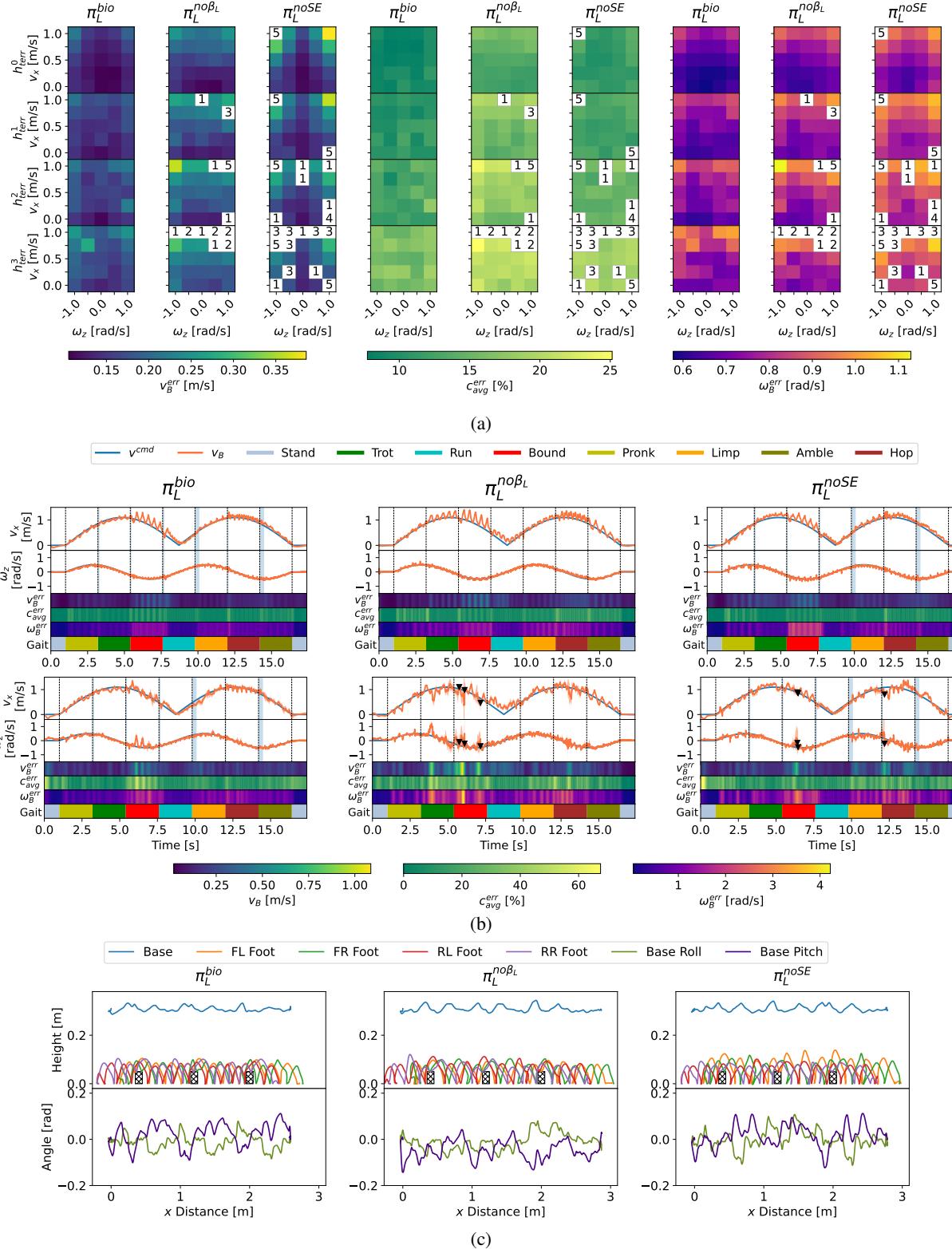
We evaluate the proposed framework through a set of studies, demonstrating it outperforms others by exhibiting quadruped animal locomotion strategies, and validating this gained proficiency on real-world terrain, as presented in Fig. 2 and Supplementary Video 1 (Appendix B).

A. Achieving Adaptive Motion Adjustment with a Diverse Set of Gaits

To evaluate our method of instilling adaptive motion adjustment and procedural memory for diverse gait deployment, a comparison study is completed between our bio-inspired locomotion policy π_L^{bio}, a standard multi-gait locomotion policy with no pseudo procedural memory $\pi_L^{\text{no}\boldsymbol{\beta}_L}$, and a policy trained which also uses $\boldsymbol{\beta}_L$ within $\boldsymbol{o}_L$ but implements the standard approach of extracting the observations from the simulator, π_L^{noSE}, for which a video summary can be found in Supplementary Video 2 (Appendix C). From the results of this study, shown in Fig. 3, the proficiency of π_L^{bio} over $\pi_L^{\text{no}\boldsymbol{\beta}_L}$ and π_L^{noSE} in terms of velocity tracking error v_B^{err}, contact schedule tracking error $c_{\text{avg}}^{\text{err}}$, and base stability (magnitude of undesirable base angular velocities) ω_B^{err}, is stark.

As illustrated by Fig. 3a, on flat terrain π_L^{bio} exhibits lower v_B^{err}, $c_{\text{avg}}^{\text{err}}$ and ω_B^{err} compared to $\pi_L^{\text{no}\boldsymbol{\beta}_L}$ and π_L^{noSE}, on average by 15%, 21% and 10% respectively (not accounting for failure cases), with this error increasing at higher velocity command magnitudes. However, this difference in performance is only exacerbated when rough terrain is introduced, in which both $\pi_L^{\text{no}\boldsymbol{\beta}_L}$ and π_L^{noSE} fail repeatedly, particularly at higher velocities and rougher terrain as indicated within Fig. 3a by the failure occurrence counts within the heatmap, whereas

Fig. 3: **Locomotion comparison study experiments** (a) each policy follows a set of command velocities in x and yaw between 0 to 1 m/s and -1 to 1 rad/s respectively. During each velocity pair, the commanded gait is cycled through all gaits, switching every 1 s. This repeated 5 times over flat, h_{terr}^0, to very rough terrain, h_{terr}^3 (shown in Fig. 2) with the average performance being plotted. A number rather than a magnitude indicates the count of experiments that the policy failed. (b) each policy follows a sinusoidal type trajectory in v_x and ω_z while switching gaits every 2 s, which is repeated 5 times over h_{terr}^0 and h_{terr}^3 terrain, with blue highlighted areas indicating transition phases and black triangles representing points where the robot failed. (c) policies follow a command of just $v_x = 0.5$ m/s while using a trot gait while encountering rectangular steps with a height of 0.05 m.

π_L^{bio} completes every experiment without fail despite only having experienced flat terrain during training. In turn, this demonstrates its impressive adaptability to new environments (see Section III-C for justification of omitting rough terrain from π_L training).

This incompetence of $\pi_L^{\text{no}\boldsymbol{\beta}_L}$ and π_L^{noSE} is caused by a lack of adaptive swing foot motion adjustments captured within $\boldsymbol{\beta}_L$ based on $\boldsymbol{s}$ and the accumulation of error within the SE respectively. Both of these factors are substantially affected by the instabilities rough terrain inflicts upon the robot. Considering that the nominal swing foot peak height is defined as 25% of the nominal base height and for h_{terr}^2 and h_{terr}^3 (as defined in Fig. 2) the peak terrain height is 44% and 67% of the base height, having no data or strategy to account for this harsh terrain results in the rapid deterioration of proficiency.

For $\pi_L^{\text{no}\boldsymbol{\beta}_L}$, Fig. 3b shows large spikes in $c_{\text{avg}}^{\text{err}}$ and ω_B^{err} when encountering h_{terr}^3, highlighting its inability to adapt swing foot trajectories and overcome an unrefined solution space. Fig. 3c further shows sustained instabilities in base height, roll, and pitch after contact with steps at 17% of the nominal base height. For π_L^{noSE}, Figs. 3b and 3c show that poor reference tracking and stability worsen with velocity command magnitude and time, as it lacks strategies to mitigate error build-up in the SE. With π_L^{bio} not experiencing any of these limitations, this explicitly demonstrates the effectiveness of implementing $\boldsymbol{\beta}_L$ and $\boldsymbol{s}$ within $\boldsymbol{o}_L$; π_L^{bio} can deploy a diverse set of gaits while exhibiting highly adaptive behaviour over previously unobserved terrain, in turn demonstrating successful instillation of adaptive motion adjustment and gait procedural memory.

B. Applying Biomechanics Metrics For Optimal Gait Selection

Directly applying biomechanics metrics to instil animal gait transition strategies is unsuitable due to differences between animals and robots and π_G training requirements. Therefore, we instead apply cost of transport, CoT, torque saturation, $\tau_\%$, external work, W_{ext}, and foot contact reference tracking error, $c_{\text{avg}}^{\text{err}}$, for the minimisation of energy consumption, musculoskeletal forces (generically referred to as actuator-structural forces), mechanical work, and instability respectively. Full details and justification of these adapted metrics are detailed in Section III-E.

In accordance with Section III-F, these metrics are unified within the training of the gait selection policy π_G^{uni} to instil the strategies animals use for optimal gait selection for exemplary adaptability. To investigate whether π_G^{uni} effectively minimises these metrics through gait selection, the results of competing the highly demanding velocity command trajectory presented in Fig. 4 and Supplementary Video 3 (Appendix D) for π_G^{uni} paired with π_L^{bio} are collected, along with that for all individual gaits deployable by π_L^{bio}, for flat terrain and terrain h_{terr}^2.

From the gait selection results in the bottom time series subplot presented in Fig. 4, it is clear that π_G^{uni} almost exclusively utilises trotting at low speeds and running at high speeds. When accelerating from low to high speeds, π_G^{uni} oscillates between trotting and running, with an increasing bias towards

running, to increase the stride frequency as visualised by the foot contact data in Fig. 4. Consequently, π_G^{uni} not only reliably tracks the optimal gait but outperforms individual gaits in metrics and velocity tracking by oscillating between trotting and running during acceleration. This behaviour, although never targeted during training, is reflected in animal locomotion strategies, where gait stride frequency increases with speed, and transitional phases blend gaits to minimize energy use [29].

However, introducing rough terrain and high acceleration causes π_G^{uni} to utilise other gaits as auxiliary tools to overcome instability. Hence a gait classification arises: trot and run gaits are nominal performance gaits at slower and faster speeds respectively, while bound, pronk, limp, amble and hop gaits are auxiliary gaits for off-nominal scenarios such as stability recovery.

When inspecting the radar charts in Fig. 4, which depict π_G^{uni} and each gaits relative performance in terms of the metrics, the origin of this emerged gait selection strategy becomes clear. On flat terrain, trot and run gaits achieve a good performance across the metrics with the gaits either exhibiting relatively worse performance. However, when it comes to overcoming rough terrain bound, hop and limp gaits all gain relative performance in $\tau_\%$, W_{ext} and $c_{\text{avg}}^{\text{err}}$, while trot and run gaits exhibit a reduced dominance in relative proficiency. In addition to this observation providing an insight as to why π_G^{uni} chooses to utilise these auxiliary gaits it also provides evidence to suggest that $\tau_\%$, W_{ext} and $c_{\text{avg}}^{\text{err}}$ can effectively characterise stability. This observation is further investigated and discussed in the following sections. Overall, π_G^{uni} outperforms all individual gaits across all metrics, with the exception of CoT for trot and run gaits where the difference is negligible, demonstrating the successful minimisation of the metrics and successful instillation of gait procedural memory of how to utilise each gait given the robot's state and task.

C. Comparison Between Robot and Animal Gait Selection

When developing metrics to characterise gait transitions in animals, data is collected over intervals of increasing forward velocity on flat terrain [42], [33], [29], [31]. Hence, within Fig. 5a we took the same approach. We also repeat this experiment with h_{terr}^3 to investigate correlations between the metrics and the effects of introducing rough terrain, as presented in Fig. 5b. We also train four additional π_G policies that individually minimise energy consumption π_G^{CoT}, actuator-structural forces $\pi_G^{\tau_\%}$, mechanical work $\pi_G^{W_{\text{ext}}}$ and stability $\pi_G^{c^{\text{err}}}$, in accordance with Section III-E, to compare their performance with π_G^{uni}. One unanimous observation across Fig. 5a is that animals experience a gait transition phase over a range of velocities [31], [42], [43]. This behaviour is reflected in π_G^{uni}, where we class a transition phase as where no individual gait occupies more than 75% of the gaits used at a specific speed.

1) Energy Expenditure – Cost of Transport: An animal's gait transition phase overlaps with the optimal point of transition for CoT, λ^{CoT}, to minimise CoT [27], [29], as depicted in Fig. 5a. This behaviour is reflected by π_G^{uni} as it tracks the lowest CoT gait and has a clear transition phase centred
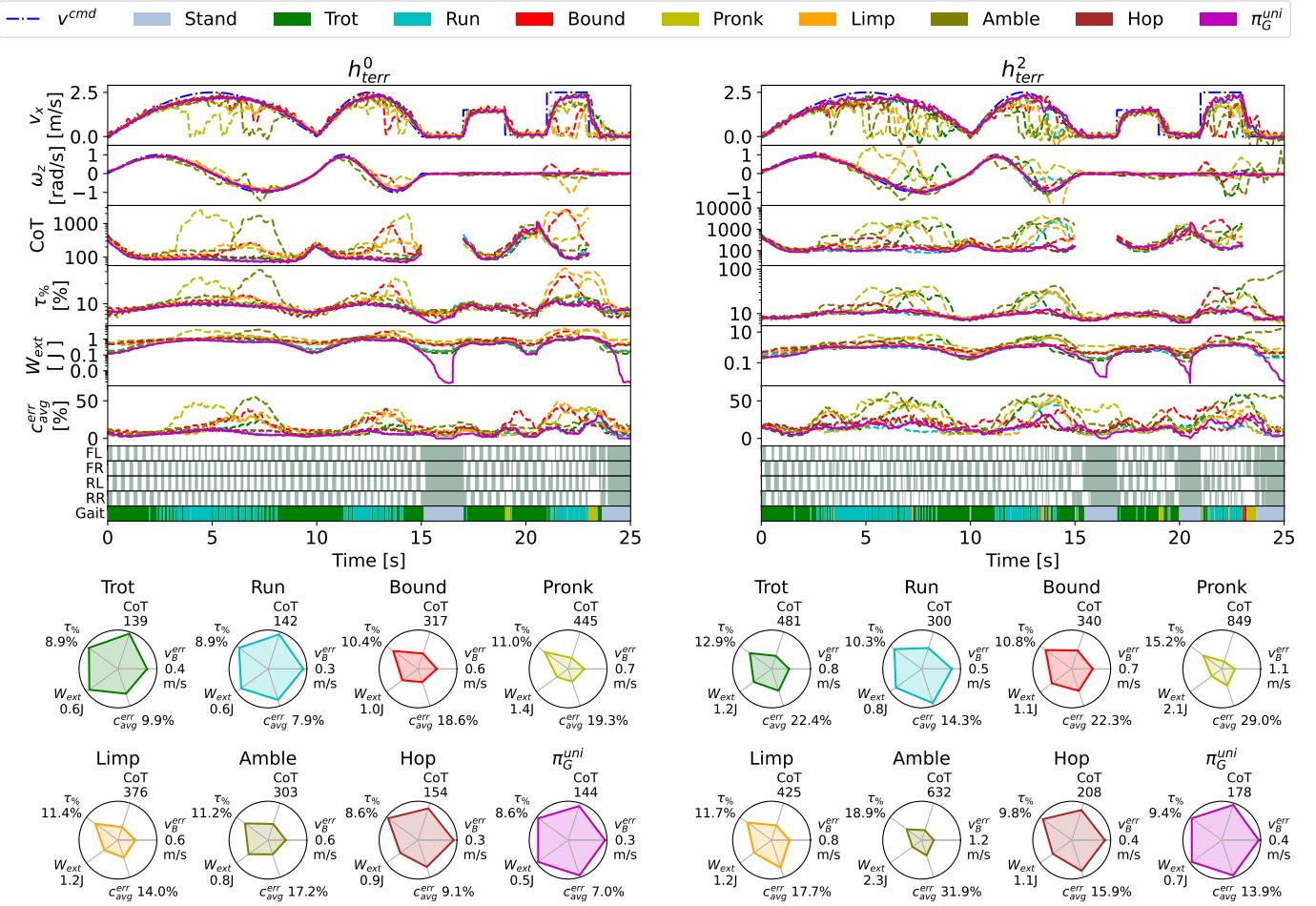
Fig. 4: **Comparative study between each gait and** π_G^{uni}. For terrains h_{terr}^0 and h_{terr}^2, each isolated gait and π_G^{uni} are given a velocity command to follow to assess their performance in terms of CoT, $\tau_\%$, W_{ext}, and $c_{\mathrm{avg}}^{\mathrm{err}}$; for each individual gait π_L^{bio} just run with the gait statically selected and for π_G^{uni} it is coupled with π_L^{bio} for autonomous optimal gait selection. Additionally at the bottom of these time series plots, the contact state of the feet and the gait that π_G^{uni} is also displayed. The average relative performance in terms of these metrics is displayed in the radar plots in the bottom of the figure, where each gait's performance is normalised to that of the best performer for each metric; the higher the value within the radar plot, the more effectively that metric has been minimised.
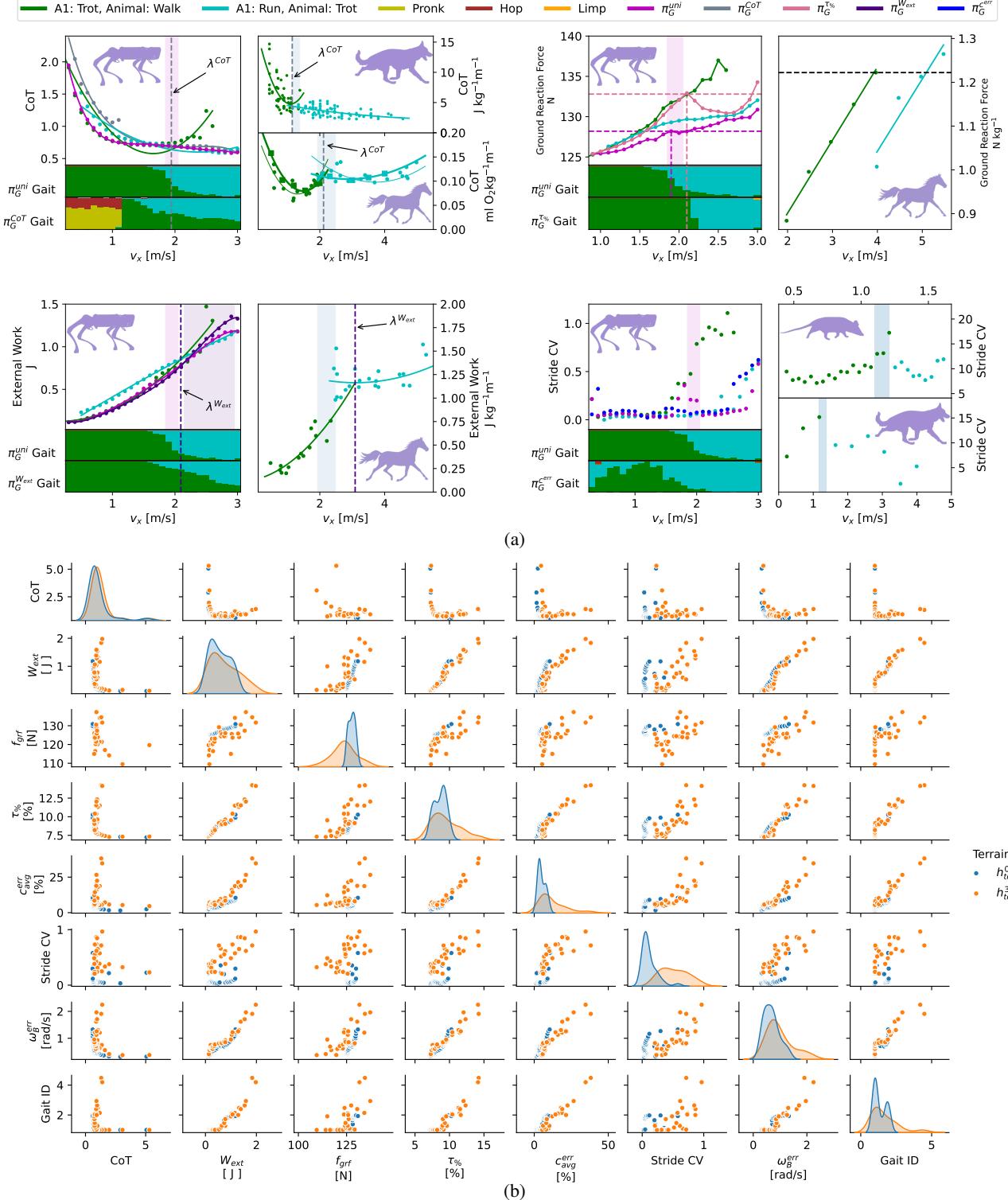
around λ^{CoT}. However, π_G^{CoT} initially fails to establish a transition phase and the transition is earlier than λ^{CoT}. This is a result of all π_G policies experiencing rough terrain during training where a hopping gait has improved CoT efficiency, as presented in Fig. 4. This is further supported by Fig. 5b, where the introduction of h_{terr}^3 leads to a very similar CoT distribution with a much higher variance in gait ID as utility gaits become more effective at metric minimisation. This suggests that training a policy that only considers CoT decreases generality and fails to resemble animal gait selection strategies.

2) Actuator-structural Forces – Foot Contact Forces: Animals are observed to change gait to minimise actuator-structural forces (i.e. musculoskeletal forces) [33], which in biomechanics is measured through ground reaction force, f_{grf}. Similarly, π_G^{uni} and $\pi_G^{\tau_\%}$ reduce f_{grf} through selecting the optimal gait for minimising $\tau_\%$. This supports that $\tau_\%$ is a suitable alternative to f_{grf}, which is further validated by a

strong correlation between them within Fig. 5b. However, not only does $\pi_G^{\tau_\%}$ maintain a trotting gait past optimal f_{grf}, but also the transition itself is instantaneous. In turn, $\pi_G^{\tau_\%}$ better reflects the animal data from [33] than π_G^{uni}. This could be explained by the metric being misinterpreted in [33]; with a high correlation between $\tau_\%$ and f_{grf} with stability metric $c_{\mathrm{avg}}^{\mathrm{err}}$, it suggests instability causes transition, which is rare on flat terrain, as discussed in Section II.

3) Mechanical Work – External Work: Animals are observed to transition to preserve external mechanical work, W_{ext}, but unlike with CoT, the transition happens before $\lambda^{W_{\mathrm{ext}}}$, as shown in Fig. 5a, demonstrating that minimisation of W_{ext} is relatively relaxed. π_G^{uni} reflects this behaviour as the transition occurs before $\lambda^{W_{\mathrm{ext}}}$ yet minimal W_{ext} is preserved. While $\pi_G^{W_{\mathrm{ext}}}$ can reduce W_{ext}, its transition phase is extended over a larger range of velocities compared to animals, occurring just after $\lambda^{W_{\mathrm{ext}}}$. In turn, this suggests that switching gaits between trotting and running offers minimal reductions in W_{ext}

Fig. 5: **Comparison between animal and robot gait selection policy strategies** where across all plots both animal and robot locomote at increasing linear forward velocity. (a) The bottom two plots of all robot data indicates the percentage of each gait utilised at that velocity. Magenta, purple and blue shaded regions indicate the transition phases of π_G^{uni}, $\pi_G^{W_{\text{ext}}}$ and animals respectively. This study compares transition strategies of (top left) π_G^{uni} and π_G^{CoT} to data collected from dogs [31] and horses [42], [29] in terms of CoT, (top right) π_G^{uni} and $\pi_G^{\tau_\%}$ to data collected from horses [33] in terms of foot ground reaction forces, (bottom left) π_G^{uni} and $\pi_G^{W_{\text{ext}}}$ to horses [29] in terms of external work, and (bottom right) π_G^{uni} and $\pi_G^{c_{\text{err}}}$ to opossums and dogs [31]. (b) Mapping the correlation between different all metrics and the average gait ID selected across all velocities for terrains h_{terr}^0 and h_{terr}^3.

resulting in a less definitive transition. However, this metric also seems to capture stability due to its high correlation with the stability metrics on h_{terr}^3 in Fig. 5b, providing insight into explaining its reduced role in Fig. 5a where only flat terrain is present.

4) Stability – Stride Duration Coefficient of Variation: Within [31] animals are shown to reduce their stride duration coefficient of variation (CV) to preserve stability as high gait periodicity indicates stability. This behaviour is presented in Fig. 5a, where animals are seen to initiate a transition phase when there is a significant increase in stride CV, consequently leading to a decrease in CV and an increase in stability. π_G^{uni} inherits the same strategy as only when a spike in stride CV is experienced does a transition phase begin that results in improved stability through lowering CV. However, this is not the case with $\pi_G^{c^{\text{err}}}$ as it acts to reduce stride CV much more aggressively by mixing both slow, fast and auxiliary gaits which results in no clear transition phase being produced.

Overall, only π_G^{uni} consistently reflects the behaviour across all animal data sets. In turn, this supports the notion that no singular metric can characterise animal gait selection behaviour and only when unifying them can similar behaviour in robots arise; the minimisation of the metrics is expected and is seen across all π_G policies, but the intricacies of animal gait transition behaviour are only seen in π_G^{uni}. Additionally, we have also verified that this behaviour transfers to real-world deployment as explored in Supplementary Information A (Appendix A-A).

D. Adaption to Real World Terrain

Although we have instilled animal gait transition strategies, gait procedural memory, and adaptive motion into our framework, its real-world proficiency remains uncertain. Grassy terrain may trap swing feet, and the ground often features irregularities. Despite π_L^{bio} only observing flat terrain during training, it can deploy all seven gaits on this terrain, as illustrated in Fig. 6a, demonstrating that gait procedural memory and adaptive motion adjustment successfully transfer to real-world environments, providing a high level of adaptability, as shown in Supplementary Video 4 (Appendix E).

Terrain that causes states of instability presents a significant risk to the robot, hence we test the limits of our framework through deployment on loose timber, muddy grass, and a low-friction board, as presented in Fig. 6b, Fig. 6c and Fig. 6d respectively and consolidated in Supplementary Video 5 (Appendix F), with an additional experiment for external perturbations included within Supplementary Information B (Appendix A-B). Each of the presented experiments showcases an off-nominal stability recovery event; however in the nominal scenario, the framework can maintain stability without changing gaits. In the case of loose timber, critical instability is caused when one back foot slips on a plank, causing it to collide with another. In response, π_G^{uni} utilises auxiliary gaits pronk and bound to recover, as depicted in Fig. 6b. This strategy of utilising the auxiliary gaits for stability recovery is seen across all experiments and reflected in animals as highlighted in Fig. 6e where a horse is observed to utilise bounding and limping gaits to traverse down complex rock formations.
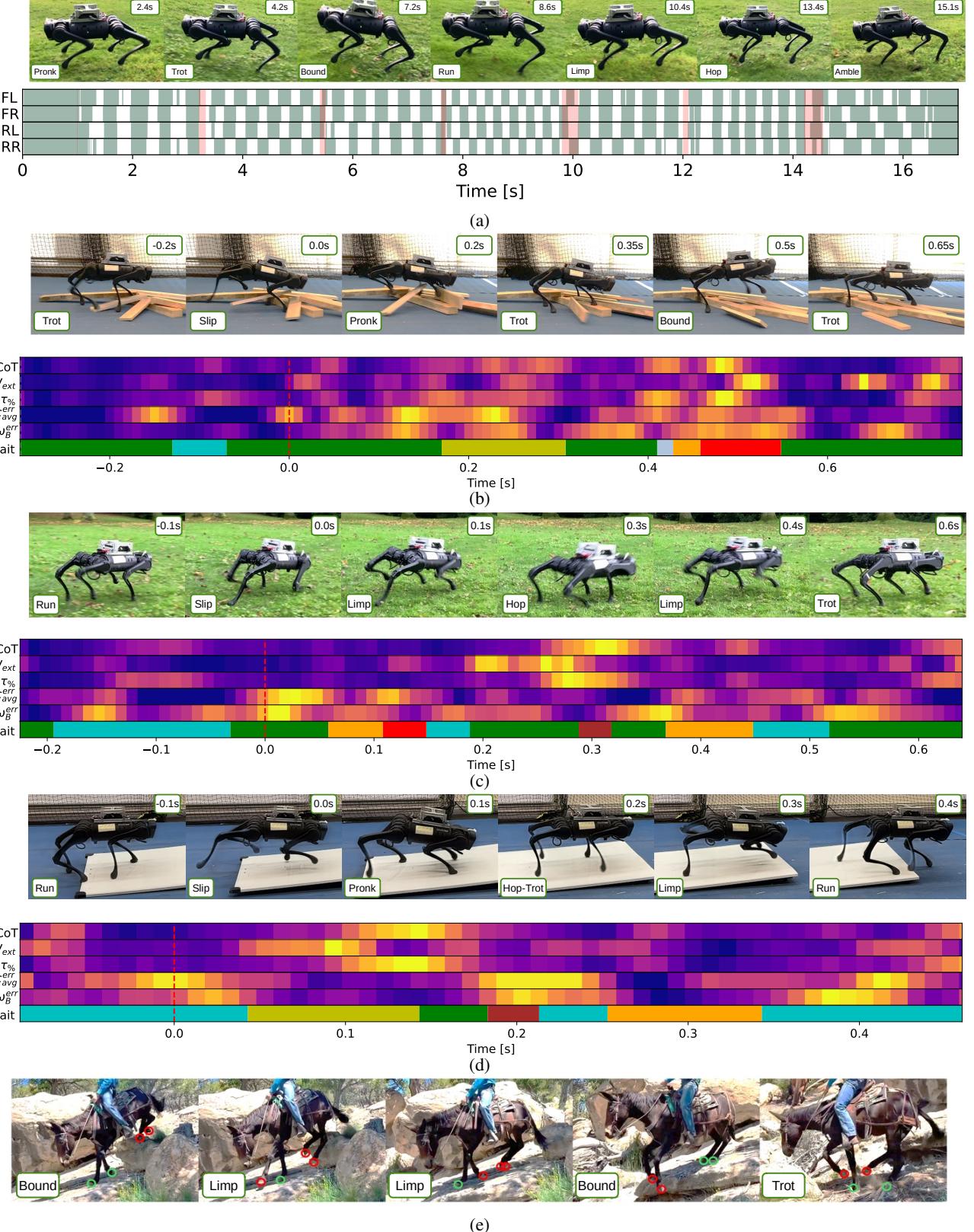
In all experiments presented, a considerable increase in a combination of W_{ext}, $\tau_\%$ and $c_{\text{avg}}^{\text{err}}$ is experienced by the robot before a gait transition, while a weaker correlation is observed with CoT; spikes in W_{ext}, $\tau_\%$ and $c_{\text{avg}}^{\text{err}}$ directly coincide or even preempt gait changes while CoT peaks lag. This is expected with $c_{\text{avg}}^{\text{err}}$ and W_{ext} as they capture periodicity and base height respectively. However, this is less expected for $\tau_\%$ to correlate with stability; this was never a factor investigated within [33], however, the correlation observed in these experiments and in Fig. 5b provides strong evidence that this is the case.

II. Discussion

From taking inspiration from animal locomotion proficiency attributes, we have developed a locomotion framework capable of traversing complex and high-risk terrain despite the robot not utilising extra-perceptive sensors nor experiencing any rough terrain during the training of π_L^{bio}. For π_L^{bio}, this is achieved through including the BGS output $\boldsymbol{\beta}_L$ (which encodes state dependent pseudo gait procedural memory and adaptive motion adjustments) within the observation space $\boldsymbol{o}_L$. This proves to be effective at overcoming rough terrain within Fig. 3 as without the presence of $\boldsymbol{\beta}_L$ within $\boldsymbol{o}_L$, increased failure and instability are observed. This is the equivalent of removing an animal's cerebellum functionality, which would result in reduced limb coordination and stability [41], in turn supporting the claim that $\boldsymbol{\beta}_L$ effectively encodes pseudo gait procedural memory.

π_G^{uni} for optimal gait selection greatly expands adaptability through instilling gait selection strategies used by animals. As demonstrated in Fig. 6, π_G^{uni} can maintain stability in the event of the terrain undergoing radical structural or friction coefficient adjustments. These scenarios pose risks to robots with vision systems, as they typically cannot detect ground friction or terrain changes beyond their front legs. Through the use of π_G^{uni}, this limitation is mitigated without implementing resource-heavy extra-perceptive sensors. Comparing Fig. 6b through 6d to Fig. 6e showcases that animals and π_G^{uni} utilise multiple auxiliary gaits to prevent failure. This behavior, untargeted during training, suggests that unifying these metrics encodes the intricacies of animal gait transitions.

One provoking observation from this work is that actuator-structural forces appear to characterise instability. Considering that [33] validates by applying increased loads that could cause instability, it suggests this metric was initially misunderstood. Another observation is that despite the employed biomechanics metrics only being tested on animals completing a linear forward trajectory on flat terrain, π_G^{uni} upholds animal gait transition strategies across a wide range of terrains and base velocity commands. This supports that these metrics successfully characterise gait transitions and the notion that robots can test biomechanics hypotheses, avoiding the resource, compatibility, and ethical challenges of animal testing. Moving forward, we aim to integrate extra-preceptive sensors for preemptive gait planning to reduce stability risks.

Fig. 6: **Framework deployment in real-world environments to evaluate adaptability** (a) The deployment of π_L^{bio} on uneven grassy terrain with manual gait selection to cycle through all gaits with a v_x^{cmd} of 0.5 m/s where red shaded regions indicate a transition period. For subfigures (b), (c) and (d) our framework is deployed on loose timber, muddy grass and a low-friction board where the event of critical loss of stability is indicated by the red dashed line and the bottom subplot uses the same gait colour code as Fig. 4. (e) Snapshots showing how animals also employ the strategy of using a mixture of auxiliary gaits to overcome challenging terrains, where red circles are swing feet and green circles are stance feet.

III. METHODS

A. Control Framework Overview

At the core of this work, the Unitree A1 quadruped robot used within all experiments features 12 degrees of freedom, n, which are all modeled as revolute joints, with their angular positions denoted as $\boldsymbol{q} \in \mathbb{R}^n$ and its base orientation represented as a rotation matrix $\boldsymbol{R}_B \in SO(3)$. As discussed in Section I-A and outlined in Fig. 1, both π_L and π_G are integrated within a control framework and supported by the SE and BGS for generation of the robot's state data and gait references respectively. The final output of π_L is target joint positions, $\boldsymbol{q}^*$, which then get converted into joint torques, $\boldsymbol{\tau}^*$, through the following PD controller to get sent to the motors,

$$\boldsymbol{\tau}^* = K_p(\boldsymbol{q}^* - \boldsymbol{q}) - K_d\dot{\boldsymbol{q}}, \tag{1}$$

where K_p and K_p are the PD controller gains. Throughout this work, a constant $K_p = 25$ N/m and $K_d = 1$ Ns/m are used while running at 1000 Hz, while π_L and π_G are run at 500 Hz and 100 Hz respectively.

B. Bio-inspired Gait Scheduler

The BGS primary output, $\boldsymbol{\beta}_L = [\boldsymbol{c}^{\text{ref}}, \boldsymbol{p}_x^{\text{ref}}, \boldsymbol{p}_y^{\text{ref}}, \boldsymbol{p}_z^{\text{ref}}] \in \mathbb{R}^{16}$, defines the reference contact state of each foot, $\boldsymbol{c}^{\text{ref}} \in \mathbb{B}^4$, and their reference Cartesian position in the world frame x-axis, $\boldsymbol{p}_x^{\text{ref}} \in \mathbb{R}^4$, y-axis, $\boldsymbol{p}_y^{\text{ref}} \in \mathbb{R}^4$, and z-axis, $\boldsymbol{p}_z^{\text{ref}} \in \mathbb{R}^4$ which are calculated online using the Raibert heuristic [44] to account for the current state of the robot. An adjusted version of the BGS output, $\boldsymbol{\beta}_G$, is used for π_G as not all the information in $\boldsymbol{\beta}_L$ is required. This has the form of $\boldsymbol{\beta}_G = [\boldsymbol{c}^{\text{ref}}, \boldsymbol{p}_z^{\text{ref}}, \Omega_{\text{stab}}, \kappa] \in \mathbb{R}^{10}$ where $\Omega_{\text{stab}} \in \mathbb{R}$ characterises the inherent stability of a gait [27], and $\kappa \in \mathbb{B}$ is a logical flag to indicate a state of gait transition. We originally developed the BGS within [45] where the Froude number [27], Ω, is used to trigger gait transition based exclusively on CoT which results in a set order of transitions. However, when applied to this work this method is not entirely suitable as now multiple biomechanics metrics and a set of auxiliary gaits need to be considered. One issue is that Ω values over 1 are not compatible when calculating how many gait cycles, C, should a transition occur over. With this work investigating higher velocities than in [45], this has been resolved through calculating C through

$$C = e^{-2\Omega} \tag{2}$$

This relationship ensures an almost instantaneous transition at a Ω of ≥ 2, which is the typical value that quadruped animals transition to a run [27] instantaneously. Another limitation is that the calculation of the transition resolution, n, (how quickly a transition should be progressed each time step) only enables the transition between set gait pairs; this was not an issue in [45] as CoT efficiency was the only metric considered. As π_G requires any gait transition pair to be possible, $\Omega_{\text{stab}} = g/hf^2$ [27] is utilised, where g is gravity, h is hip height and f is gait frequency. Through the use of Ω_{stab}, we are able to determine an indication of the inherent stability of any gait, hence a transition between a higher Ω_{stab} gait to a lower one should have smaller values of n to increase the smoothness of

the transition to promote stability. In the reverse scenario, a more harsh transition is more feasible hence larger values of n should be produce for rapid transition. As such, n is now calculated by

$$n = 1 + \frac{\Omega_{\text{stab}}}{\Omega_{\text{stab}}^{\text{next}}} \tag{3}$$

where $\Omega_{\text{stab}}^{\text{next}}$ is the Ω_{stab} of the gait that's being transitioned to. In essence, f of the current and next gait dictates the harshness of the transition. This behaviour is also reflected in animal gait transitions, where transitioning from running (higher f) to trotting (lower f) the transition is slower compared to the opposite scenario [46]. Overall, this augmented version of the BGS can achieve transition between any designed gait, while considering the inherent stability of the transition. Complete details of how $\boldsymbol{c}^{\text{ref}}$ is generated for each gait can be found in Supplementary Information C (Appendix A-C).

C. Policy Training

To simplify the training process, for both the locomotion policy, π_L, and gait selection policy, π_G, the training method, environment and network architecture are kept constant. Both policies are modelled as an MLP with hidden layer sizes $[512, 256, 128]$ and LeakyReLU activations. Subscripts L and G represent the specific parameters for the locomotion policy and gait selection policy respectively. The model-free DRL training problem for the policies is represented as a sequential Markov decision process (MDP), which aims to produce a policy that maximises the expected return of the policy π,

$$J(\pi) = \mathbb{E}_{\xi \sim p(\xi|\pi)} \left[\sum_{t=0}^{N-1} \gamma^t r \right], \tag{4}$$

in which $\gamma \in [0, 1)$ is the discount factor, ξ is a finite-horizon trajectory dependent on π with length N, $p(\xi|\pi)$ is the likelihood of ξ, and r is the reward function. The proximal policy optimization (PPO) algorithm [47] is used to train the locomotion policy and the hyperparameters used are detailed in Supplementary Information D (Appendix A-D). As discussed in Section I-A, we estimate the state of the robot during training using an SE. Hence, in terms of applying state feedback noise for domain randomisation to improved sim-to-real transfer we only need to implement this on the input sensor data vector of the SE, $\boldsymbol{\sigma} = [\boldsymbol{\omega_B}, \dot{\boldsymbol{v}}_B, \boldsymbol{q}, \dot{\boldsymbol{q}}, \boldsymbol{\tau}, \boldsymbol{f}_{\text{grf}}]$. This vector includes base angular velocity, $\boldsymbol{\omega_B} \in \mathbb{R}^3$, base linear acceleration, $\dot{\boldsymbol{v}}_B \in \mathbb{R}^3$, joint positions, $\boldsymbol{q}$, joint velocities, $\dot{\boldsymbol{q}}$, joint torques, $\boldsymbol{\tau}$, and foot ground reaction forces, $\boldsymbol{f}_{\text{grf}} \in \mathbb{R}^4$. As the initial state of the robot and its performance can never be guaranteed during real-world deployment, we also randomise the initial configuration of the robot, the mass of the robot's base, K_p and K_d. Additionally, to ensure that a rich variation of $\boldsymbol{U}^{\text{cmd}}$ is experienced during training randomly sampled gaits, velocity commands and velocity change durations (to achieve random acceleration) are implemented during training. For all details regarding the noise and sampling used within training please refer to Supplementary Information E (Appendix A-E). The environment itself is constructed using RaiSim [48], as the vectorized environment setup allows for

efficient training of policies. Additionally, the observation normalisation functionality offered by RaiSim is also used for improved training.

During the training of π_L only flat terrain is present within the environment to isolate and highlight the effect of implementing β_L; a core claim of this work is that the implementation of β_L aims to impart gait procedural memory within π_L^{bio} hence if rough terrain was observed during training it will become ambiguous if the improved performance is a direct result of implementing β_L. However, for training π_G, flat to very rough terrain is implemented using fractal noise, enabling the policy to learn to employ the use of each gait minimising biomechanics metrics on a variety of terrains. We train all variations of π_L and π_G for 20k iterations, taking 6 and 9 hours respectively, on a standard desktop computer with one Nvidia RTX3090 GPU with a training frequency of 100Hz. It is also important to note that the training of all π_G policies only utilise our final proposed bio-inspired locomotion framework π_L^{bio}.

D. Locomotion Policy

The goal of the locomotion policy π_L is to realise the input $\boldsymbol{U}^{\text{cmd}}$ while exhibiting stable and versatile behaviour. As such, π_L is trained to generate the action, $\boldsymbol{q}^*$, from an input observation, $\boldsymbol{o}_L = [\beta_L, \boldsymbol{s}, \boldsymbol{v}_B^{\text{cmd}}] \in \mathbb{R}^{69}$, where $\boldsymbol{v}_B^{\text{cmd}} = [v_x^{\text{cmd}}, v_y^{\text{cmd}}, \omega_z^{\text{cmd}}] \in \mathbb{R}^3$ is the high-level velocity command of the robot's base within $\boldsymbol{U}^{\text{cmd}}$, as outlined in Fig. 1. $\boldsymbol{s}$ is generated from the output of the SE and is defined as $\boldsymbol{s} = [\boldsymbol{\alpha R}_B^T, \boldsymbol{q}, \boldsymbol{\omega}_B, \dot{\boldsymbol{q}}, \boldsymbol{v}_B, z_B, \boldsymbol{\tau}, \boldsymbol{c}] \in \mathbb{R}^{50}$, where $\boldsymbol{\alpha} = [0,0,1]^T$ is used to select the vertical z-axis, $\boldsymbol{\omega}_B \in \mathbb{R}^3$ is the base angular velocity, $\boldsymbol{v}_B \in \mathbb{R}^3$ is the base linear velocity, z_B is the current base height, and $\boldsymbol{c} \in \mathbb{B}^4$ is the contact state of the feet. The locomotion reward function, r_L, is formulated so that the the output of the policy can realise the reference gait patterns and velocity commands stably, smoothly and accurately,

$$r_L = \text{w}_\eta r_\eta + \text{w}_{v^{\text{cmd}}} r_{v^{\text{cmd}}} + \text{w}_f r_f + \text{w}_{\text{stab}} r_{\text{stab}}, \quad (5)$$

where r_η, $r_{v^{\text{cmd}}}$, r_f and r_{stab} are the grouped reward terms focusing on efficiency, velocity command tracking, gait reference tracking and stability respectively. w_η, $\text{w}_{v^{\text{cmd}}}$, w_f and w_{stab} are the weights of each reward and are valued at -1.5, 15, -10, and -5 respectively. r_η aims to minimise joint jerk, $\dddot{\boldsymbol{q}}$, joint torque, and the difference between $\boldsymbol{q}^*$ and the previous action, $\boldsymbol{q}_{t-1}^*$,

$$r_\eta = \|\dddot{\boldsymbol{q}}\|^2 + \|\boldsymbol{\tau}\|^2 + \|\boldsymbol{q}^* - \boldsymbol{q}_{t-1}^*\| \quad (6)$$

$r_{v^{\text{cmd}}}$ minimises the difference between the commanded base velocity and the current base velocity,

$$r_{v^{\text{cmd}}} = \psi\left(\|\boldsymbol{v}_B - \boldsymbol{v}_B^{\text{cmd}}\|^2\right), \quad (7)$$

in which the function $\psi : x \to 1 - \tanh\left(x^2\right)$ is used to normalise the reward term so that their maximum value is 1 to prevent bias towards individual rewards, $\boldsymbol{v}_B = [v_x, v_y, \omega_z] \in \mathbb{R}^3$ is the current base x, y and yaw velocities. r_f ensures the robot realises the commanded gait references within β_L,

$$r_f = |\boldsymbol{c}^{\text{err}}| + \sum_{i=1}^{4} \|\boldsymbol{p}_i - \boldsymbol{p}_i^{\text{ref}}\|^2, \quad (8)$$

where $\boldsymbol{c}^{\text{err}} \in \mathbb{B}^4$ defines the feet that do not meet the desired contact state, with $\boldsymbol{p}_i \in \mathbb{R}^3$ and $\boldsymbol{p}_i^{\text{ref}} \in \mathbb{R}^3$ being the current and reference Cartesian positions of the i-th foot. r_{stab} aims to prevent contact foot slip, large hip joint motions and undesirable base orientations,

$$r_{\text{stab}} = \sum_{i=1}^{F} \|\dot{\boldsymbol{p}}_i\|^2 + \|\boldsymbol{\omega}_B\|^2 + \psi\left(\|\boldsymbol{\alpha R}_B - \boldsymbol{\alpha R}_B^{\text{des}}\|^2\right) \\ -\psi\left((z_B - z_B^{\text{nom}})^2\right) + \|\boldsymbol{q}_{\text{hip}}\|^2, \quad (9)$$

where $\boldsymbol{p}_i \in \mathbb{R}^3$ is the velocity of the i-th foot scheduled to be in stance, F is the number of stance feet, $\boldsymbol{\omega}_B = [\omega_x, \omega_y] \in \mathbb{R}^2$ where ω_x and ω_y are roll and pitch base velocities respectively, $\boldsymbol{R}_B^{\text{des}} \in SO(3)$ is the desired base orientation, z_B^{nom} is the nominal base heights respectively, and $\boldsymbol{q}_{\text{hip}} \in \mathbb{R}^4$ is the hip angular joint positions.

E. Biomechanics Gait Transition Metrics

As the set of biomechanics metrics applied in this work were originally designed for analysing animal locomotion, several adjustments to how they are calculated needed to be implemented; for example, energy consumption in animals is often measured through the rate of consumption of O_2, hence unsuitable for the application of robotics. Additionally, as robots provide a wide array of feedback data some of the metrics have also been augmented to better reflect the characteristic that these biomechanics metrics are attempting to characterise. That being said, for Fig. 5a only the original biomechanics metrics are applied to allow for direct comparison between robot and animal data.

1) Energy Efficiency: The calculation of CoT takes the general form of

$$\text{CoT} = \frac{P}{mgv}, \quad (10)$$

where P is power consumed, m is the system's mass, and g is gravity. When studying animal locomotion, P is found through measuring how much CO_2 is generated and O_2 is consumed and v is assumed to be the speed of the treadmill the animal is running on [25], [26], [27]. For the case of the robot, we calculate P from $\boldsymbol{\tau}$ and $\dot{\boldsymbol{q}}$ with an adjustment term, adopted from [49], and v is assumed to be the magnitude of the robot base velocity command to take a similar approach to animal studies and for consistent metric use between simulation and real-world deployment; completely accurate measurement of the robot's linear base velocity is impossible during real-world deployment due to the accumulation of error within the SE. As such, calculation of the robot's CoT is formulated as

$$\text{CoT} = \sum_{i=1}^{n} \frac{\max(\tau_i \dot{q}_i + 0.3\tau_i^2, 0)}{mg|\boldsymbol{v}_B|}, \quad (11)$$

where m is the robot's mass and g is gravity. It should be noted that CoT is only calculated and applicable when $|\boldsymbol{v}_B^{\text{cmd}}| > 0$.

2) Actuator-structural Forces: As gaining an exact understanding of the actuator-structural forces within animals, researchers have opted instead to measure the peak ground reaction forces of the animal's stance feet during locomotion using force plates [33]. Other methods include adding strain gauges to the bones of the animals [34]. However, in the case of robots we have the privilege of having access to joint state feedback while also knowing the exact limitations of the hardware. Therefore, when considering the biomechanics hypothesis that animals aim to minimise actuator-structural forces to prevent injury and that torque is proportional to strain and force, in the case of the robot we chose to characterise the actuator-structural forces through joint torque saturation, $\tau_\%$, which is calculated by

$$\tau_\% = \left| \frac{\boldsymbol{\tau}}{\boldsymbol{\tau}_{\text{lim}}} \right| \frac{1}{n}, \tag{12}$$

where $\boldsymbol{\tau}_{\text{lim}} \in \mathbb{R}^n$ is the joint torque limits (assumed based on manufacturers specification), which proves particularly usefully when considering that the hip joints of most quadruped robots, including the A1, are often more sensitive to forces at the foot due to their distance from the point of ground impact and the only motor of the leg in its set alignment; this would not be considered if just ground reaction force was used to characterise actuator-structural forces.

3) Mechanical Work Efficiency: During animal locomotion, if they are to have perfect mechanical work efficiency there would be a net zero change in external work over the duration of a gait cycle as there would be perfect exchange between kinetic and potential energy [28]. As expected perfect mechanical work is never seen in nature, hence mechanical work efficiency in animals is characterised by the sum of the change in kinetic and potential energy [28] or the sum of the external work of the animal [29] over the duration of a gait cycle. As this is typically calculated through measuring the O_2 uptake, for the case of robots we formulate the calculation of the external work, W_{ext}, through

$$W_{\text{ext}} = \sum_{i=0}^{t_{\text{gait}}} (\Delta E_{\text{k},i} - \Delta E_{\text{p},i}) \tag{13}$$

where t_{gait} is the duration of the current gait cycle, and $\Delta E_{\text{k},i}$ and $\Delta E_{\text{p},i}$ are the changes in kinetic and potential energy over a control time step respectively. The primary difference between the metrics seen in biomechancis to out formulation of W_{ext} is that $\Delta E_{\text{k},i}$ accounts for not only forward linear velocity but also lateral and angular velocity whereas originally only forward linear velocity is considered.

4) Stability: The best indication of stability in animals is their stride duration CV. This metric characterises periodicity, which is a primary indication of stable locomotion [31]. However, to accurately calculate this, the mean and standard deviation of the stride duration needs to be taken over an extended period of time for appropriate data generation. This is sufficient for undertaking analysis similar to that presented in Fig. 5a, but this presents an issue when it comes to analysing the performance of the proposed control framework as it is common for multiple speed commands being used within the

duration of one stride. Hence, to overcome this limitation we instead use $c_{\text{avg}}^{\text{err}} = |\boldsymbol{c}^{\text{err}}|/4$, which can be measured every time step rather than just each foot touchdown event; the gait references generated by the BGS have a constant and periodic stride duration, therefore an accurate tracking of this reference would in turn indicate high periodicity, which is further supported by the correlation between the two metrics within Fig. 5b.

F. Gait Selection Policy

To achieve optimal gait selection for a given state, we leverage the biomechanics metrics within the reward function of π_G, r_G. For the different variations of π_G used within Fig. 5a, each policy's reward function only features the metric that its focusing on within r_G but π_G^{uni} unifies all metrics hence uses the full form of r_G with all metrics. Additionally, as the biomechanics metrics all describe a characteristic that animals try to minimise through changing gaits, they can be directly applied within r_G with some normalisation where appropriate. The full form of r_G is

$$r_G = \text{w}_{\text{u}} r_{\text{u}} + \psi(\text{CoT}) + \psi(\tau_\%) + \psi(c_{\text{avg}}^{\text{err}}) + \psi(W_{\text{ext}}), \tag{14}$$

where r_{u} is the utility reward term which all π_G use and w_{u} is its weight with a value of 0.4. r_{u} aims to ensure the smoothness of the output Γ^*, the standing gait is only used when appropriate, and any select gait is still able to follow $\boldsymbol{v}_B^{\text{cmd}}$. To achieve this, r_{u} has the form of

$$r_{\text{u}} = r_{v^{\text{cmd}}} + r_{\text{stand}} + r_{\text{smooth}}, \tag{15}$$

in which $r_{v^{\text{cmd}}}$ is taken from (7), and r_{stand} is set to -10 if a stand gait is used when $|\boldsymbol{v}_B^{\text{cmd}}| > 0$ or not used when $|\boldsymbol{v}_B^{\text{cmd}}| = 0$. For r_{smooth}, the reward aims to penalise unnecessary changes in Γ^* to remove rapid gait changes when two gaits could achieve similar metric minimisation for a given task and state. As such, if there is a gait change between time steps it is calculated as $r_{\text{smooth}} = -\psi(\text{CoT} + \tau_\% + c_{\text{avg}}^{\text{err}} + W_{\text{ext}})$ otherwise it is set to 0. To generate Γ^*, π_G takes in input observation vector $\boldsymbol{o}_G = [\boldsymbol{s}, \boldsymbol{\beta}_G, \boldsymbol{v}_B^{\text{cmd}}, \dot{\boldsymbol{v}}_B^{\text{cmd}}, \Gamma_{t-1}^*] \in \mathbb{R}^{66}$ in which Γ_{t-1}^* is the previous output action to aid in action smoothing. Appropriate selection of the data provided to π_G is critical in order to achieve targeted minimisation of the biomechanics metrics. As such, the inclusion of $\boldsymbol{s}$ coupled with $\boldsymbol{c}^{\text{ref}}$, $\boldsymbol{p}_z^{\text{ref}}$, $\boldsymbol{v}_B^{\text{cmd}}$, $\dot{\boldsymbol{v}}_B^{\text{cmd}}$ and Ω_{stab} informs the policy of its current and demanded stability, while the terms $\boldsymbol{\tau}$ and $\dot{\boldsymbol{q}}$ within $\boldsymbol{s}$ capture the power consumption of the robot and the forces to which it is subjected.

APPENDIX A
SUPPLEMENTARY INFORMATION

A. Preservation of Biomechanics Metrics on Grass

When deploying π_L^{bio} with π_G^{uni} on a smooth low-friction floor and uneven grassy terrain π_G^{uni} can utilise the same behaviour animals demonstrate in minimising CoT; comparing the performance of maintaining a static gait of both trot and run gaits within Fig. A.1, π_G^{uni} can select the most energy efficient gait and even prevent failure as the trot gait cannot
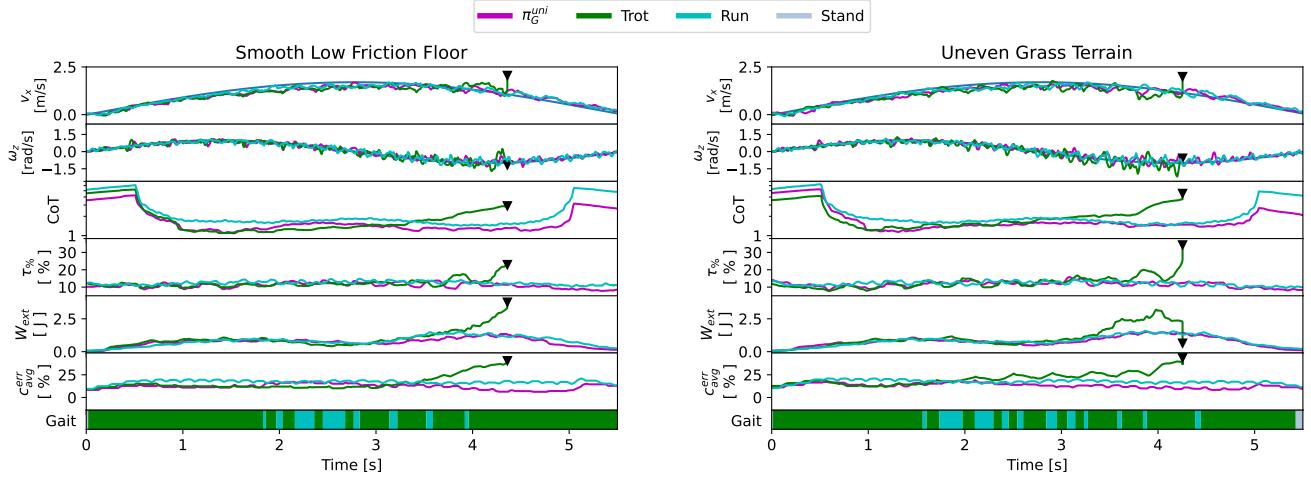
Fig. A.1: Deployment of π_L^{bio} and π_G^{uni} on smooth low-friction floor and uneven grassy terrain, with the black triangles indicating a point of failure.

maintain stability for the entire experiment. Additionally, the distribution of gait usage between trot and run gaits varies between the two terrains. This is due to the uneven grassy terrain causing reduced stability of the robot when trotting, as indicated by $c_{\text{avg}}^{\text{err}}$ in Fig. A.1, which in response π_G^{uni} exhibits increased modulation between trotting and running to preserve stability. This preservation of stability while minimising CoT is achieved through the resultant modulation of the stride frequency as gaits with higher stride frequency offer improved stability [31] and efficiency on rough terrain [29].

B. Stability Recovery from External Perturbations

An additional stability recovery experiment was completed where the robot was subjected to repeated external perturbations and in response utilised the utility gaits to recover, as presented in Fig. A.2, which in turn reflects the findings discussed in Sections I-D and II.

C. Generation Gait Contact References within the BGS

TABLE A.1: Gait design parameters.

Name	Period	Duty Factor	Phase Offset
Trot	0.40	0.50	0.00, 0.50, 0.50, 0.00
Trot Run	0.30	0.40	0.00, 0.50, 0.50, 0.00
Bound	0.40	0.40	0.00, 0.00, 0.50, 0.50
Pronk	0.50	0.50	0.00, 0.00, 0.00, 0.00
Amble	0.50	0.55	0.00, 0.50, 0.25, 0.75
Unnatural	0.40	0.50	0.05, 0.50, 0.50, 0.00
Hop	0.30	0.50	0.00, 0.00, 0.00, 0.00

In accordance to our work in [45], to generate c^{ref}, a set of phase variables, $\phi_i \in [0, 1)$, for each leg $\phi_1, \ldots, \phi_4$, are used to determine the progress along a gait pattern, and duty factor, $d_i \in [0, 1)$, sets the percentage of the phase that each leg is in stance. These encoded parameters for each gait are presented within Table A.1. When $\phi_i = d_i$, the contact state of the i-th

TABLE A.2: PPO Hyperparameters.

Parameter	Value
Number of Environments	240
Clip Range	0.2
Max Steps per Batch	400
GAE λ	0.95
Learning Epochs per Batch	4
Learning Rate	5e-4
Number of Mini-batches	4
Minimum Policy std	0.2
Reward Discount Factor	0.99
Optimizer	Adam

leg switches to swing. The phase of the i-th leg is calculated as

$$\phi_i = \frac{t - t_{i,0}}{T}, \tag{16}$$

in which t is the current time, $t_{i,0}$ is the start time of the current gait period of the i-th leg, and T is the gait period. The last parameter used to construct a gait pattern is the phase offset, $\theta_i \in [0, 1]$, that defines the difference in phase between the leading leg and all other legs through $\phi_i = \phi_1 + \theta_i$. As such, dependent on the phase and gait parameters, the i-th value of c^{ref} is either 1 or 0 to signify if the leg should be in stance or swing respectively.

D. PPO Hyperparameters

All PPO hyperparameters used within this work are detailed within Table A.2.

E. Noise and Sampling Distributions Used During Training

In the effort of improving sim-to-real transfer, domain randomisation utilised through randomly sampled noise from either uniform or normal distributions to all parameters within σ, the initial configuration of the robot in each episode, q_{init}, the mass of the robot's base, m_B, K_p and K_p. The details of this is presented in Table A.3. Additionally, to ensure that a rich variation of U^{cmd} is experienced during
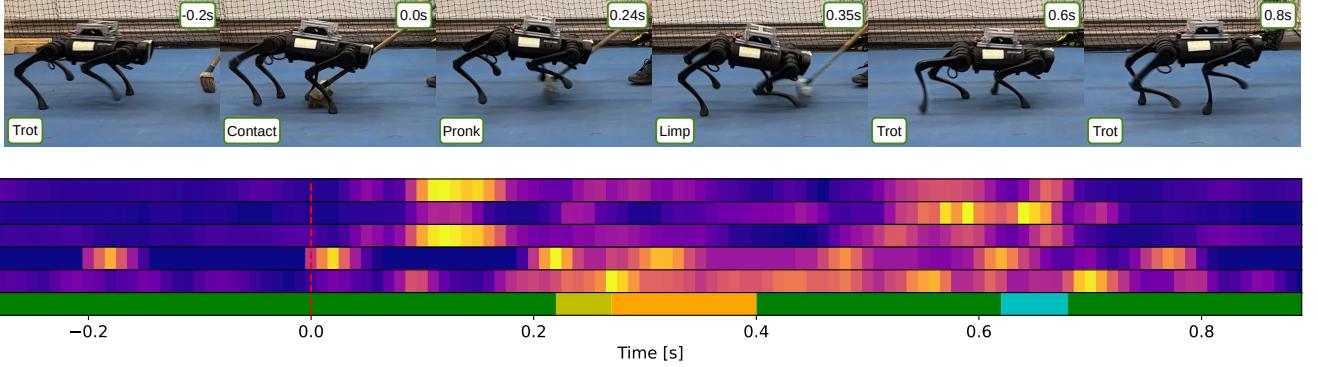
Fig. A.2: Enacting perturbations upon the legs of the robot to investigate how our framework can reject this instability.

TABLE A.3: Noise and Sampling Distributions.

Parameter	Distribution
$\boldsymbol{\omega}_B, \dot{\boldsymbol{v}}_B$	$0.015\mathcal{N}(0,1)$
$\boldsymbol{q}$	$0.005\mathcal{N}(0,1)$
$\dot{\boldsymbol{q}}$	$0.15\mathcal{N}(0,1)$
$\boldsymbol{\tau}, \boldsymbol{f}_{\text{frc}}$	$\mathcal{N}(0,1)$
m_B	$\max(-1, \min(\mathcal{N}(0,1), 3))$
K_p	$K_p \max(0.9, \min(1 + 0.05\mathcal{N}(0,1), 1.1))$
K_d	$K_d \max(0.9, \min(1 + 0.05\mathcal{N}(0,1), 1.1))$
v_x^{cmd}	$\mathcal{U}(0, 1.5)$
ω_z^{cmd}	$\mathcal{U}(-1, 1)$
Γ	$\mathcal{U}(0, 6)$
t_{acc}	$\mathcal{U}(t_{\text{acc}}^{\min}, t_{\text{acc}}^{\max})$

training, randomly sampled gaits, velocity commands and velocity change durations (to achieve random acceleration), t_{acc}, are implemented during training. These are within defined maximum, $t_{\text{acc}}^{\max} = 0.5$s, and minimum, $t_{\text{acc}}^{\min} = 0$s, acceleration durations. Further sampling details are also presented in Table A.3.

APPENDIX B
SUPPLEMENTARY VIDEO 1

Link: https://youtu.be/NwHoB7pErYQ

APPENDIX C
SUPPLEMENTARY VIDEO 2

Link: https://youtu.be/-DfkDFA3KkI

APPENDIX D
SUPPLEMENTARY VIDEO 3

Link: https://youtu.be/y4KnzMEdf78

APPENDIX E
SUPPLEMENTARY VIDEO 4

Link: https://youtu.be/I02DQ1RGdyw

APPENDIX F
SUPPLEMENTARY VIDEO 5

Link: https://youtu.be/f6CqJ7gb3ZM

REFERENCES

[1] C. Vanden Hole, J. Goyens, S. Prims, E. Fransen, M. Ayuso Hernando, S. Van Cruchten, P. Aerts, and C. Van Ginneken, "How innate is locomotion in precocial animals? A study on the early development of spatio-temporal gait variables and gait symmetry in piglets," *Journal of Experimental Biology*, vol. 220, no. 15, pp. 2706–2716, 08 2017.

[2] E. Avital and E. Jablonka, *Animal traditions: Behavioural inheritance in evolution*. Cambridge University Press, 2000.

[3] A. N. Wimberly, G. J. Slater, and M. C. Granatosky, "Evolutionary history of quadrupedal walking gaits shows mammalian release from locomotor constraint," *Proceedings of the Royal Society B: Biological Sciences*, vol. 288, no. 1957, p. 20210937, 2021.

[4] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *nature*, vol. 323, no. 6088, pp. 533–536, 1986.

[5] T. Miki, J. Lee, J. Hwangbo, L. Wellhausen, V. Koltun, and M. Hutter, "Learning robust perceptive locomotion for quadrupedal robots in the wild," *Science Robotics*, vol. 7, no. 62, p. eabk2822, 2022.

[6] I. M. Aswin Nahrendra, B. Yu, and H. Myung, "Dreamwaq: Learning robust quadrupedal locomotion with implicit terrain imagination via deep reinforcement learning," in *2023 IEEE International Conference on Robotics and Automation (ICRA)*, 2023, pp. 5078–5084.

[7] S. Chen, B. Zhang, M. W. Mueller, A. Rai, and K. Sreenath, "Learning torque control for quadrupedal locomotion," in *2023 IEEE-RAS 22nd International Conference on Humanoid Robots (Humanoids)*, 2023, pp. 1–8.

[8] V. Atanassov, J. Ding, J. Kober, I. Havoutis, and C. D. Santina, "Curriculum-based reinforcement learning for quadrupedal jumping: A reference-free design," 2024. [Online]. Available: https://arxiv.org/abs/2401.16337

[9] 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.

[10] J. Hwangbo, J. Lee, A. Dosovitskiy, D. Bellicoso, V. Tsounis, V. Koltun, and M. Hutter, "Learning agile and dynamic motor skills for legged robots," *Science Robotics*, vol. 4, no. 26, p. eaau5872, 2019.

[11] N. A. Curtin, H. L. Bartlam-Brooks, T. Y. Hubel, J. C. Lowe, A. R. Gardner-Medwin, E. Bennitt, S. J. Amos, M. Lorenc, T. G. West, and A. M. Wilson, "Remarkable muscles, remarkable locomotion in desert-dwelling wildebeest," *Nature*, vol. 563, no. 7731, pp. 393–396, 2018.

[12] T. Y. Hubel, K. A. Golabek, K. Rafiq, J. W. McNutt, and A. M. Wilson, "Movement patterns and athletic performance of leopards in the okavango delta," *Proceedings of the Royal Society B: Biological Sciences*, vol. 285, no. 1877, p. 20172622, 2018.

[13] S. Wilshin, M. A. Reeve, G. C. Haynes, S. Revzen, D. E. Koditschek, and A. J. Spence, "Longitudinal quasi-static stability predicts changes in dog gait on rough terrain," *Journal of Experimental Biology*, vol. 220, no. 10, pp. 1864–1874, 05 2017.

[14] E. Muybridge, L. Brown, L. Brown, L. Brown, and A. History, *Animals in motion*. Dover Publications, 1957.

[15] F. Righini, M. Carpineti, F. Giavazzi, and A. Vailati, "Pronking and bounding allow a fast escape across a grassland populated by scattered obstacles," *Royal Society Open Science*, vol. 10, no. 9, p. 230587, 2023.

[16] G. Feng, H. Zhang, Z. Li, X. B. Peng, B. Basireddy, L. Yue, Z. SONG, L. Yang, Y. Liu, K. Sreenath, and S. Levine, "Genloco: Generalized lo-

comotion controllers for quadrupedal robots," in *Proceedings of The 6th Conference on Robot Learning*, ser. Proceedings of Machine Learning Research, K. Liu, D. Kulic, and J. Ichnowski, Eds., vol. 205. PMLR, 14–18 Dec 2023, pp. 1893–1903.

[17] D. Kang, J. Cheng, M. Zamora, F. Zargarbashi, and S. Coros, "Rl + model-based control: Using on-demand optimal control to learn versatile legged locomotion," *IEEE Robotics and Automation Letters*, vol. 8, no. 10, pp. 6619–6626, 2023.

[18] Y. Shao, Y. Jin, X. Liu, W. He, H. Wang, and W. Yang, "Learning free gait transition for quadruped robots via phase-guided controller," *IEEE Robotics and Automation Letters*, vol. 7, no. 2, pp. 1230–1237, 2022.

[19] G. B. Margolis and P. Agrawal, "Walk these ways: Tuning robot control for generalization with multiplicity of behavior," in *6th Annual Conference on Robot Learning*, 2022.

[20] Z. Fu, A. Kumar, J. Malik, and D. Pathak, "Minimizing energy consumption leads to the emergence of gaits in legged robots," in *Conference on Robot Learning (CoRL)*, 2021.

[21] S. Dutta, A. Parihar, A. Khanna, J. Gomez, W. Chakraborty, M. Jerry, B. Grisafe, A. Raychowdhury, and S. Datta, "Programmable coupled oscillators for synchronized locomotion," *Nature communications*, vol. 10, no. 1, p. 3299, 2019.

[22] D. Owaki and A. Ishiguro, "A quadruped robot exhibiting spontaneous gait transitions from walking to trotting to galloping," *Scientific reports*, vol. 7, no. 1, p. 277, 2017.

[23] M. Shafiee, G. Bellegarda, and A. Ijspeert, "Viability leads to the emergence of gait transitions in learning agile quadrupedal locomotion on challenging terrains," *Nature Communications*, vol. 15, no. 1, p. 3073, 2024.

[24] F. Ruppert and A. Badri-Spröwitz, "Learning plastic matching of robot dynamics in closed-loop central pattern generators," *Nature Machine Intelligence*, vol. 4, no. 7, pp. 652–660, 2022.

[25] F. J. Diedrich and W. H. W. Jr., "The dynamics of gait transitions: Effects of grade and load," *Journal of Motor Behavior*, vol. 30, no. 1, pp. 60–78, 1998.

[26] S. J. Wickler, D. F. Hoyt, E. A. Cogger, and G. Myers, "The energetics of the trot–gallop transition," *Journal of Experimental Biology*, vol. 206, no. 9, pp. 1557–1564, 05 2003.

[27] R. M. Alexander, "The gaits of bipedal and quadrupedal animals," *Int J Rob Res*, vol. 3, no. 2, pp. 49–59, 1984.

[28] G. A. Cavagna, N. C. Heglund, R. K. Taylor, C. Richard, and T. Mechanical, "Mechanical work in terrestrial locomotion: two basic mechanisms for minimizing energy expenditure." *The American journal of physiology*, vol. 233 5, pp. R243–61, 1977.

[29] A. E. Minetti, L. P. Ardigò, E. Reinach, and F. Saibene, "The relationship between mechanical work and energy expenditure of locomotion in horses," *Journal of Experimental Biology*, vol. 202, no. 17, pp. 2329–2338, 09 1999.

[30] F. Saibene and A. E. Minetti, "Biomechanical and physiological aspects of legged locomotion in humans," *European journal of applied physiology*, vol. 88, pp. 297–316, 2003.

[31] M. C. Granatosky, C. M. Bryce, J. Hanna, A. Fitzsimons, M. F. Laird, K. Stilson, C. E. Wall, and C. F. Ross, "Inter-stride variability triggers gait transitions in mammals and birds," *Proceedings of the Royal Society B: Biological Sciences*, vol. 285, no. 1893, p. 20181766, 2018.

[32] M. Lemieux, N. Josset, M. Roussel, S. Couraud, and F. Bretzner, "Speed-dependent modulation of the locomotor behavior in adult mice reveals attractor and transitional gaits," *Frontiers in Neuroscience*, vol. 10, 2016.

[33] C. T. Farley and C. R. Taylor, "A mechanical trigger for the trot-gallop transition in horses," *Science*, vol. 253, no. 5017, pp. 306–308, 1991.

[34] A. A. Biewener and C. R. Taylor, "Bone Strain: A Determinant of Gait and Speed?" *Journal of Experimental Biology*, vol. 123, no. 1, pp. 383–400, 07 1986.

[35] "Muscle-tendon stresses and elastic energy storage during locomotion in the horse," *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology*, vol. 120, no. 1, pp. 73–87, 1998.

[36] J. R. Usherwood, "An extension to the collisional model of the energetic cost of support qualitatively explains trotting and the trot–canter transition," *Journal of Experimental Zoology Part A: Ecological and Integrative Physiology*, vol. 333, no. 1, pp. 9–19, 2020.

[37] M. A. Daley, A. J. Channon, G. S. Nolan, and J. Hall, "Preferred gait and walk–run transition speeds in ostriches measured using gps-imu sensors," *Journal of Experimental Biology*, vol. 219, no. 20, pp. 3301–3308, 10 2016.

[38] O. Kiehn, "Decoding the organization of spinal circuits that control locomotion," *Nature Reviews Neuroscience*, vol. 17, no. 4, pp. 224–238, 2016.

[39] K. Takakusaki, "Functional neuroanatomy for posture and gait control," *Journal of movement disorders*, vol. 10, no. 1, p. 1, 2017.

[40] B. R. Noga and P. J. Whelan, "The mesencephalic locomotor region: Beyond locomotor control," *Frontiers in Neural Circuits*, vol. 16, 2022.

[41] S. M. Morton and A. J. Bastian, "Cerebellar control of balance and locomotion," *The neuroscientist*, vol. 10, no. 3, pp. 247–259, 2004.

[42] T. Griffin, R. Kram, S. J. Wickler, and D. F. Hoyt, "Biomechanical and energetic determinants of walk–trot transition in horses," *J Exp Biol*, vol. 207, no. 24, pp. 4215–4223, 2004.

[43] D. F. Hoyt and C. R. Taylor, "Gait and the energetics of locomotion in horses," *Nature*, vol. 292, no. 5820, pp. 239–240, 1981.

[44] M. H. Raibert, *Legged robots that balance*. MIT press, 1986.

[45] J. Humphreys, J. Li, Y. Wan, H. Gao, and C. Zhou, "Bio-inspired gait transitions for quadruped locomotion," *IEEE Robotics and Automation Letters*, vol. 8, no. 10, pp. 6131–6138, 2023.

[46] Z. Afelt, J. Błaszczyk, and C. Dobrzecka, "Speed control in animal locomotion: transitions between symmetrical and nonsymmetrical gaits in the dog," *Acta neurobiologiae experimentalis*, vol. 43, no. 4-5, pp. 235–250, 1983.

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

[48] J. Hwangbo, J. Lee, and M. Hutter, "Per-contact iteration method for solving contact dynamics," *IEEE Robotics and Automation Letters*, vol. 3, no. 2, pp. 895–902, 2018. [Online]. Available: www.raisim.com

[49] Y. Yang, T. Zhang, E. Coumans, J. Tan, and B. Boots, "Fast and efficient locomotion via learned gait transitions," in *Conference on Robot Learning*, 2022, pp. 773–783.