nature machine intelligence

Article

https://doi.org/10.1038/s42256-023-00701-w

Identifying important sensory feedback for learning locomotion skills

Received: 3 December 2022

Accepted: 30 June 2023

Published online: 21 August 2023

Check for updates

Wanming Yu[1,5], Chuanyu Yang[1,2,5], Christopher McGreavy[1], Eleftherios Triantafyllidis[1], Guillaume Bellegarda[3], Milad Shafiee[3], Auke Jan Ijspeert[3] & Zhibin Li[4]

Robot motor skills can be acquired by deep reinforcement learning as neural networks to reflect state–action mapping. The selection of states has been demonstrated to be crucial for successful robot motor learning. However, because of the complexity of neural networks, human insights and engineering efforts are often required to select appropriate states through qualitative approaches, such as ablation studies, without a quantitative analysis of the state importance. Here we present a systematic saliency analysis that quantitatively evaluates the relative importance of different feedback states for motor skills learned through deep reinforcement learning. Our approach provides a guideline to identify the most essential feedback states for robot motor learning. By using only the important states including joint positions, gravity vector and base linear and angular velocities, we demonstrate that a simulated quadruped robot can learn various robust locomotion skills. We find that locomotion skills learned only with important states can achieve task performance comparable to the performance of those with more states. This work provides quantitative insights into the impacts of state observations on specific types of motor skills, enabling the learning of a wide range of motor skills with minimal sensing dependencies.

The notion of learning machines predates the origins of cybernetics, control theories and apparatus in the 1940s[1], with a long-standing interest in creating functioning replicas of living organisms. Robots with morphologies similar to their biological counterparts provide unique opportunities to develop machines with motion capabilities comparable to those of animals. As easy-to-control platforms, robots allow scientists to study sensorimotor learning, providing opportunities to conduct control experiments and generate quantitative data analysis[2–5].

In robot learning, a large part of physical motor skills can be formulated as feedback control policies, that is, control policies represented as the state–action mapping that can be learned in the form of neural networks. The selection of feedback states becomes critical for the effective learning of robot skills. If key feedback is missing, the robot would not be able to achieve the desired performance. Physics simulation allows access to as many ideal feedback states as possible, potentially leading to better results[6]. However, certain states are not directly measurable in real robots as in the simulation and, therefore, require state estimation that is subject to uncertainties or errors[7,8], making the performance susceptible to uncertainties. Hence, it is desirable to reduce sensing dependencies at the stage of policy training, where only the most task-relevant feedback states are used for learning state–action mapping.

Deep reinforcement learning (DRL) has been successful in achieving various locomotion skills, such as trotting[9–12], pacing, spinning[11], walking[13], galloping[14], balance recovery[9] and multi-skill locomotion[12]. Deep neural networks show success in acquiring complex motor skills.

[1]School of Informatics, University of Edinburgh, Edinburgh, UK. [2]Shenzhen Amigaga Technology, Shenzhen, China. [3]Biorobotics Laboratory, École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland. [4]Department of Computer Science, University College London, London, UK. [5]These authors contributed equally: Wanming Yu, Chuanyu Yang. e-mail: alex.li@ucl.ac.uk

However, because of lack of interpretability[15,16], it is unclear how to determine the relative importance of different feedback states or which state observations are most effective. Therefore, a lot of human insights and engineering efforts are required to empirically select appropriate feedback states in robot learning[6,10,17–20].

Existing studies have used different combinations of feedback signals for learning quadruped locomotion. For example, joint positions, joint velocities, angular velocities and body orientation were used to learn gait transitions[21]. Additionally, history of body orientation, joint positions and previous actions were used to adjust locomotion policies in a quadruped robot[22]. High-speed locomotion on natural terrains was achieved by using joint positions and velocities, body orientation and previous actions[23]. However, currently the selection of these feedback states is empirical and lacks a systematic approach to determine the importance of various states for diverse motor tasks.

Biological insights and motivation

Biological studies find that animals use multimodal sensory information collected from various sensory organs to achieve different locomotion tasks, including visual, mechanical, chemical and thermal sensation, all in unison to render feedback from their own movements and the surroundings[24–27]. All this sensory information inherits redundancy that ensures robustness for movement control in animal locomotion[28]. To study the importance of different sensing information from various sensory organs, lesion studies or ablation studies have been used[29,30]. However, conducting such experiments on live animals is challenging because of ethical limitations and the difficulty of selectively stimulating and/or removing different receptors[31]. Figure 1 illustrates the similar functionality of sensory feedback in quadrupedal animals and robots. Using robots with morphologies resembling their biological counterparts offers the opportunity to investigate motion intelligence in artificial systems, and facilitates controlled and quantitative analysis of sensorimotor skills learned through machine learning, providing meaningful insights and implications for further studies in biology and neuroscience.

Related work

Ablation experiments, which focus on the impact of removing a single feedback state on performance, are commonly adopted to study the qualitative importance of individual feedback states. Experiments on a lamprey-like robot demonstrated that the distributed hydrodynamic-force feedback contributes to the generation and coordination of rhythmic undulatory swimming motion[20]. Ablation studies showed that including the Cartesian joint position or contact information has different influences on learning robot locomotion behaviours[18]. For learning central pattern generator (CPG)-based quadruped locomotion, foot-contact Booleans and CPG states are selected from various combinations of states through ablation studies[32]. Ablation studies examine the difference in performance between the inclusion and exclusion of signals of interest, and a type of feedback state is considered to be important if the performance downgrades after its removal. However, ablation studies provide only qualitative importance of individual states, without addressing the relative quantitative importance of individual sensing signals when compared with the whole set of sensory feedback.

There have also been limited attempts to compare the importance of a certain type of feedback at different time steps quantitatively. For example, the influence of proprioceptive states on foot-height commands was quantified and compared at different time steps[10]. However, this quantitative approach is used for analysing the foot-trapping behaviour during trotting and has not been extended to studying other motor skills.

In robot learning, understanding the quantitative relative importance of different sensory feedback is crucial for learning approaches to produce desired behaviours, which is yet missing in the field. Therefore, our research aims to answer the following open questions in the context of robot learning: What is the relative importance of different sensory feedback signals for a given motor task and various quadrupedal gaits? Which sensory feedback signals are essential, necessary and sufficient for learning quadrupedal locomotion? Which redundant feedback is beneficial to have but not entirely necessary?

Contribution

We present a systematic approach to quantify the relative importance of different sensory feedback in learning quadruped locomotion skills. Through distinct quadrupedal tasks, that is, balance recovery, trotting and bounding, feedback control policies are learned by neural networks as differentiable state–action mapping. We formulate a systematic saliency-analysis method to rank the level of importance of sensory feedback and identified a common set of essential feedback states for general quadruped locomotion. Further, we demonstrate the effectiveness of learning new motor skills, such as pacing and galloping, using only the most essential key states identified by our proposed approach.

In summary, our main contributions are the (1) development of a systematic saliency-analysis method to specifically quantify and rank the importance of each sensory feedback for a specific motor task, (2) identification of a common set of essential feedback states for general quadruped locomotion based on the saliency analysis of representative locomotion skills and (3) successful robot learning of new motor skills using only essential feedback states, demonstrating the efficacy of a minimal set of sensors.

Our study contributes to identifying the most essential feedback, that is, key states, in a task-specific manner, enabling robust motor learning using only the key states. The results provide new insights into the quantitative relative importance of different feedback states in locomotion behaviours. The identification of essential sensory feedback guides the selection of a minimum and necessary set of sensors, allowing robots to learn and perform robust motor skills with minimal sensing dependencies.

Results

Here we investigate the quantitative relative importance of common feedback states for three representative and distinct locomotion skills (balance recovery, trotting and bounding) and identify a set of key feedback states that are consistently more important than others: joint positions, gravity vector (that is, body orientation) and base linear and angular velocities. Our results show that learning locomotion skills with only the key feedback states achieves performance comparable to that when using all available states. Furthermore, we demonstrate the effectiveness of these key feedback states when used in learning new locomotion skills, such as pacing and galloping. More results can be found in Supplementary Videos 1–4.

Identifying key feedback states for quadruped locomotion

Quantifying the relative importance of feedback states. We formulate a systematic saliency analysis for quantifying the relative importance of various feedback states for a desired motor skill. We first determine a collection of nine feedback states commonly used in quadruped locomotion[10,12]: base position, gravity vector, base angular velocity, base linear velocity, joint position or angle, joint velocity, joint torque, foot position and foot contact (contact status or forces; see Methods for definitions). Using this full set of states, we obtain the neural-network policies for locomotion skills on the A1 quadruped robot[33] in PyBullet simulation via the DRL framework detailed in Fig. 2c and Supplementary Note 1. At each time step, we compute the saliency values of each dimension of the feedback states with respect to the associated actions using the integrated-gradients method[34] (Methods). The saliency value measures the influence of the input signal on generated actions. For each feedback dimension, it quantifies the number

https://doi.org/10.1038/s42256-023-00701-w

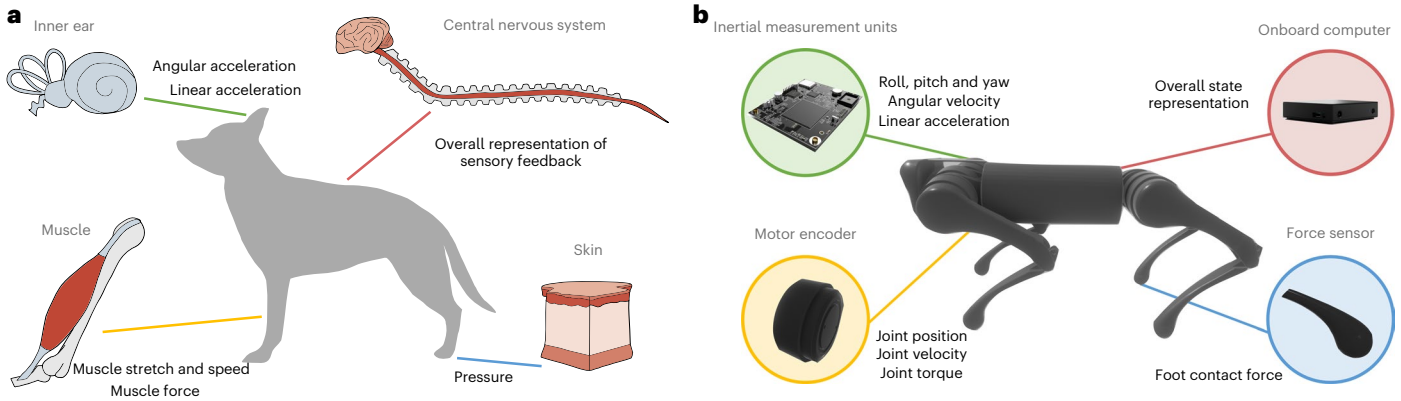


Fig. 1 | Comparison of the functional sensory feedback between quadrupedal animals and their robotic counterparts. a, Sensory organs and feedback of a dog. The central nervous system fuses sensory information from various organs, such as the inner ear, muscle and skin, and then produces motor commands. The vestibular system of vertebrates senses angular and linear accelerations, the muscle spindles measure the stretch and stretch speed of skeletal muscles, the Golgi tendon organs measure exerted muscular forces and the skin feels pressures. **b**, Sensors and feedback on a legged robot. An onboard computer processes measurements of signals from different sensors, such as an inertial measurement unit, motor encoders and force sensors, and then generates actions for electric motors.

of changes in output actions as the input signal varies at a certain time step. The higher the saliency value, the more associated actions change as the input signal varies, indicating a higher level of importance and task relevance of a particular signal. Finally, we formulate the relative importance of a given feedback state as the percentage of its quantified importance over the entire period of motions, with respect to the sum of importance of all the feedback states (Methods).

Key feedback states for quadruped locomotion. Here we identify key feedback states from the collection of nine feedback states according to the ranking of relative importance for three representative and distinct locomotion tasks: balance recovery, trotting and bounding. We visualize the saliency values and relative importance as saliency maps (Fig. 3a) and doughnut charts (Fig. 3b) delineating the contribution of each state over time and their overall impacts to the above skills, which reveals key feedback states with around 80% relative importance in total for the three skills compared with 20% relative importance for task-irrelevant states.

From the time plots of relative importance for the nine feedback states in Extended Data Fig. 1, we found that relative importance varies depending on the robot posture or phase over time. During balance recovery, gravity vector and joint positions are found to be the most important feedback states during body flipping and standing up, respectively. During periodic trotting and bounding, within each gait cycle, the most crucial feedback state alternates between joint positions and base linear velocity.

Our study reveals that each joint has a different level of contribution to distinct locomotion tasks (Fig. 4). For example, sagittal movement states and joints that have a large range of motions usually have a higher contribution to locomotion than others. During trotting, as the rear legs deliver more power to propel the robot forward and overcome energy loss caused by friction and impacts, the rear hip pitch joints move in larger ranges, resulting in higher importance. During bounding, the front knee joint positions are more important than the rear knee joint positions, as bounding requires the front knee joints to buffer landing impacts more and provide stable body weight support during the pre-landing and stance phase.

To summarize, our proposed systematic approach has identified a common set of key states that are consistently more important across the three quadrupedal locomotion skills on flat ground with a fixed gait frequency: joint positions, gravity vector and base linear and angular velocities.

Key feedback states under various circumstances. Rough terrain. We train new trotting and bounding policies on rough terrain in a 6.4 m × 6.4 m area, which consists of 4,096 cubes each with 0.1 m length and width, and heights sampled from 0 to 3 cm. Results in Extended Data Fig. 2 show that the key states for trotting and bounding remain the same as those on a flat ground.

Gait frequencies. Trotting and bounding policies were trained with 1 Hz, 2 Hz and 5 Hz, respectively. Saliency maps and doughnut charts in Supplementary Fig. 8 indicate that key feedback states for trotting and bounding are consistent across low, medium and high gait frequencies. Thus, we conclude that the identified key feedback states are not affected by the gait frequency within such a range.

Foot-contact status versus foot-contact forces. The previous analysis uses sigmoid contact, which is a continuous signal indicating the contact status based on the norm of contact forces (Methods). Here we replace the sigmoid foot contact with normalized foot-contact forces for learning trotting and bounding, where the foot-contact forces are normalized by the body weight and capped between zero and one. We conclude that using either of the foot-contact formulations as feedback states renders the same conclusions regarding the importance of foot contact (Supplementary Fig. 9).

Importance of history states. To investigate the impact of historical information on sensory feedback selection, we trained a new policy for balance recovery with an expanded set of state input, including states at the current time step and two history steps. The saliency map and the bar plot in Extended Data Fig. 3 show that (1) for each type of feedback state, current information is more important than history information, (2) important states remain important within each set of history states and (3) the overall importance of all states at the current time step is much higher than that at both history steps.

Quantifying the relative importance of the feedforward input. Importance of the feedforward phase vector. For trotting and bounding, the phase vector is used to enforce the cyclic pattern as a 'feedforward' input to the neural network besides the feedback states (Fig. 2c and 'Methods'). Note that the phase vector is included in the training of all feedback control policies for trotting and bounding but excluded from the ranking of state importance, except this section. Without

https://doi.org/10.1038/s42256-023-00701-w

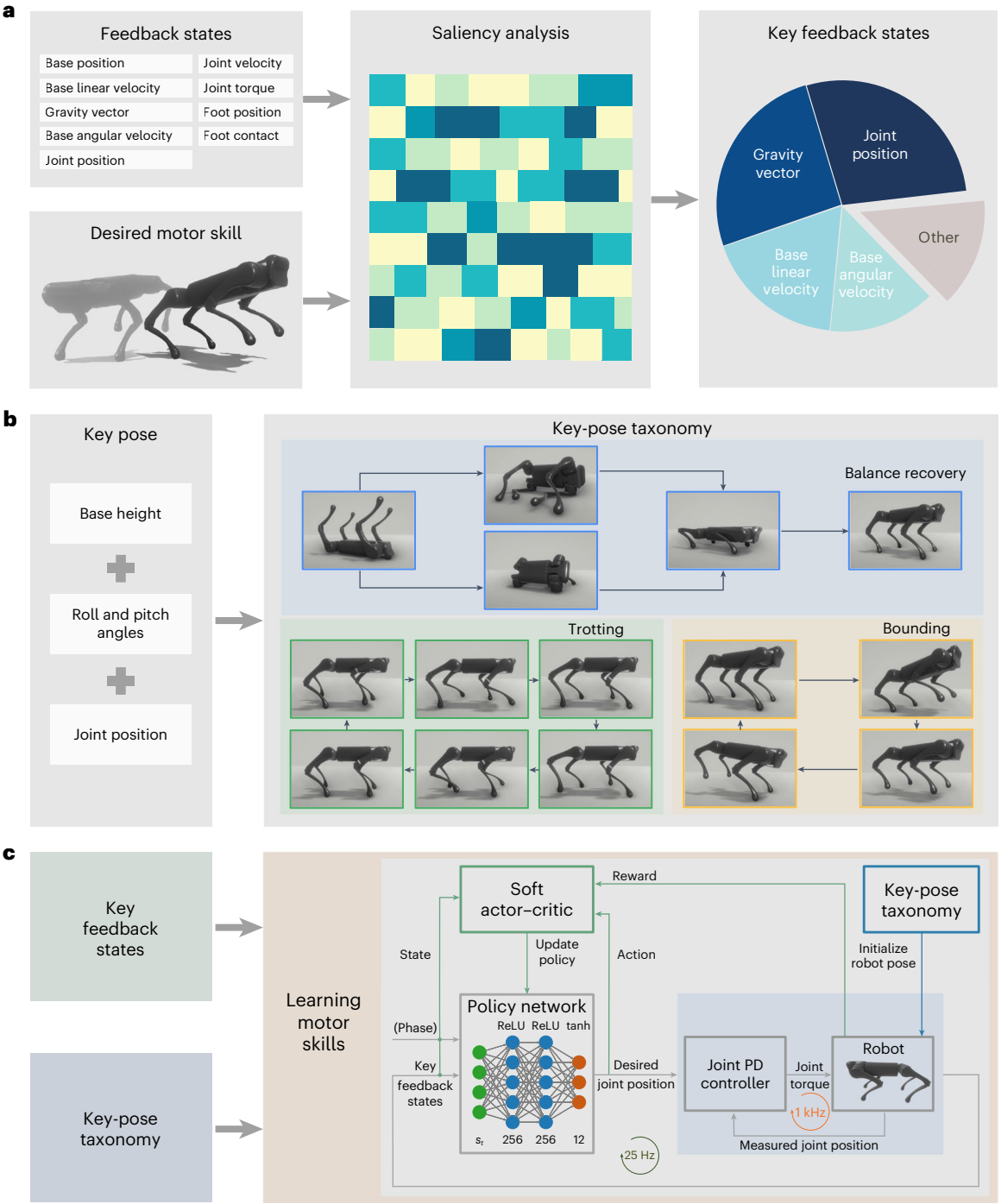


Fig. 2 | Proposed approach for identifying key feedback states used to learn effective locomotion skills. a, The proposed saliency analysis can rank the importance among different feedback states, given a set of feedback states and a targeted task-level skill. The pie chart shows our findings on the essential feedback states for quadruped locomotion, including joint position, gravity vector, base linear velocity and base angular velocity. **b**, Key-pose taxonomy used in robot-pose initialization for learning effective balance recovery, trotting and bounding. Each key pose is represented by an array of body height, body orientation (roll and pitch angles) and joint positions, which determines the pose of a floating-base robot and is categorized by distinct contact patterns that are unique to the targeted type of gait. **c**, The DRL framework utilizing key feedback states and key-pose taxonomy that are sufficient for successful and effective learning of robot motor skills, where the phase vector ($\sin 2\pi\phi$, $\cos 2\pi\phi$) inputs to the policy network in parallel with key feedback states for periodic locomotion skills (ϕ represents 0–100% phase over a gait period). PD, proportional-derivative; ReLU, rectified linear unit; s_t, state observations at time step t.

the phase vector as input, the robot would fail to learn cyclic motion. The analysis in Extended Data Fig. 4 shows that the feedforward phase vector counts for the relative importance of 28% and 38% for trotting and bounding, respectively, which is more important than any type of feedback states.

Importance of the phase vector during swing and stance. From the time plot of saliency value for phase vector in Extended Data Fig. 5, we found that the importance of the phase vector follows a cyclic pattern, reaching the highest value twice within each gait period around the transitions between swing and stance during trotting and bounding.

https://doi.org/10.1038/s42256-023-00701-w

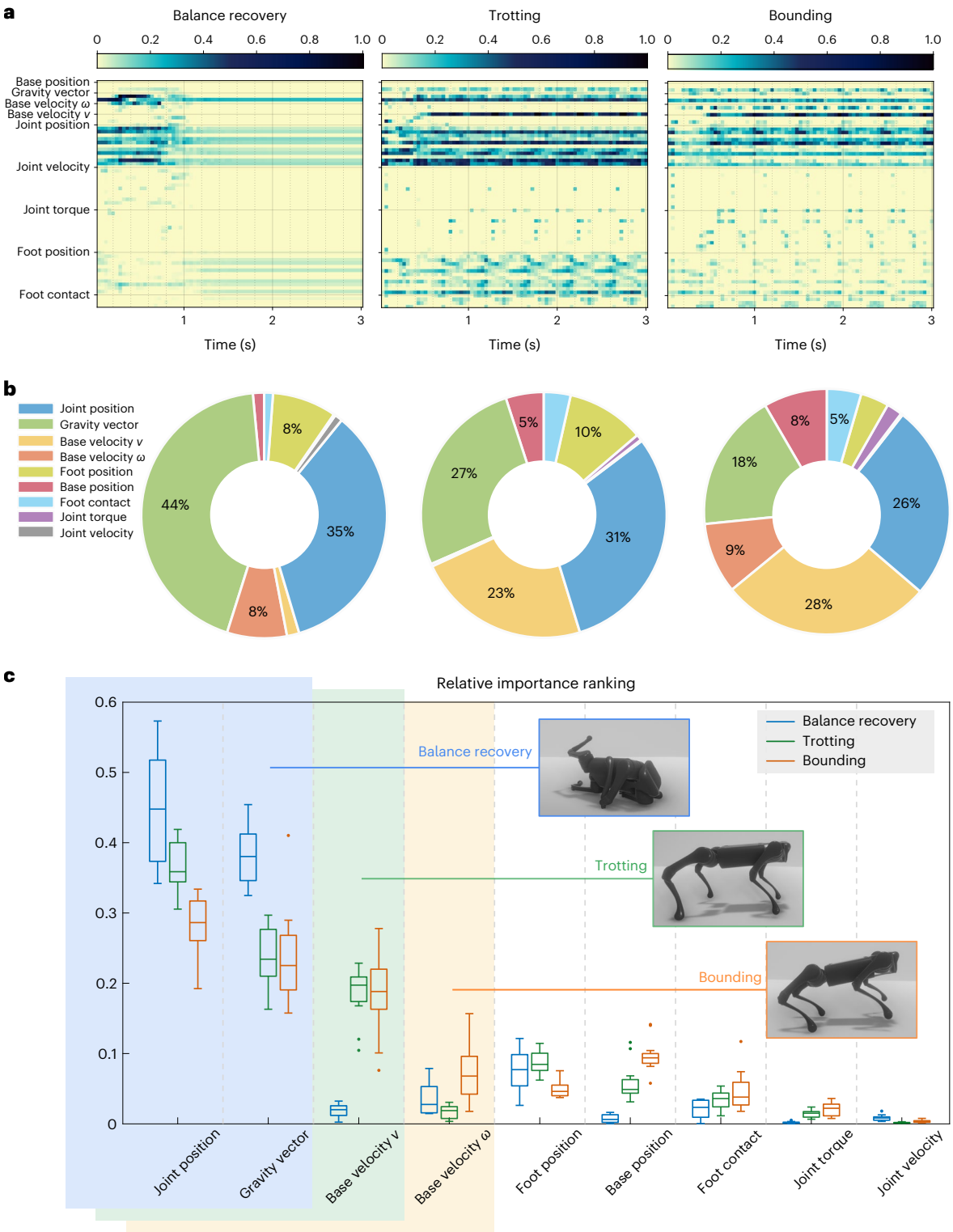


Fig. 3 | Ranking of feedback states with top four identified key states for balance recovery, trotting and bounding skills through saliency analysis. **a**, Saliency maps showing the importance of the 64-dimensional feedback for three learned skills, respectively. At each time instance in the saliency map, a darker pixel indicates that the corresponding feedback signal has more importance and influence on the generated actions, compared to others. **b**, Doughnut charts showing the relative importance of the nine feedback states for three learned skills, summarizing the overall averaged importance based on Fig. 3a. **c**, A boxplot showing statistics of relative importance of each feedback state for balance recovery, trotting and bounding, respectively. Each box shows the median (middle line of box), 25th and 75th percentiles (lower and upper bounds of box, respectively), minimum and maximum (lower and upper whiskers, respectively) and outliers (dots) of the relative importance of the corresponding state with $n = 12$ samples (Supplementary Figs. 1–7). The full set of nine feedback states are ranked in an order of relative importance from high to low, suggesting the key feedback states for quadruped locomotion include joint positions, gravity vector (that is, body orientation), base linear velocity and base angular velocity. Blue-, green- and yellow-shaded areas enclose key feedback states for the three locomotion skills.

https://doi.org/10.1038/s42256-023-00701-w

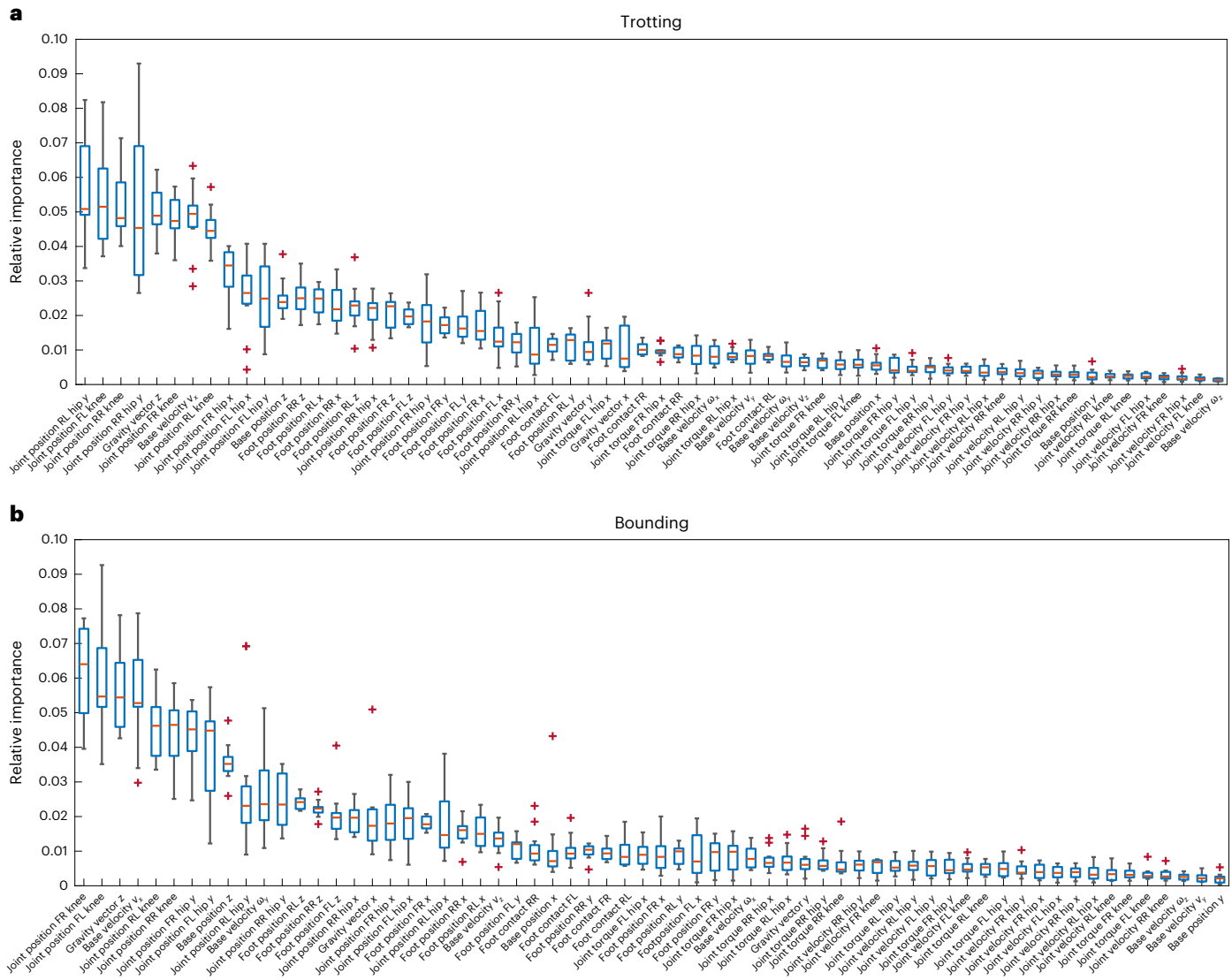


Fig. 4 | Comparison of the relative importance of 64 dimensions of feedback state for trotting and bounding. a, A boxplot expanding the boxplot in Supplementary Fig. 7b, which shows the relative importance ranking among 64 state dimensions for trotting on a flat ground ($n = 12$ samples). **b**, A boxplot expanding the boxplot in Supplementary Fig. 7c, which shows the relative importance ranking among 64 state dimensions for bounding on a flat ground ($n = 12$ samples). Each box shows the median (red horizontal line), 25th and 75th percentiles (lower and upper blue horizontal lines, respectively), minimum and maximum (lower and upper grey horizontal lines, respectively) and outliers (red plus symbol) of the relative importance of the corresponding state dimension from 12 different trials for the corresponding locomotion task. FL, front left; FR, front right; RR, rear right; RL, rear left; x, y, z, variables projected in the corresponding Cartesian coordinate; v, linear velocity; ω, angular velocity.

The cyclic pattern of the phase vector is synchronized with foot–ground contact. Such high importance during the contact transitions indicates that the phase vector regulates the timing of establishing and breaking foot–ground contact, resulting in a synchronized cyclic pattern with the transitions between swing and stance.

Benchmarking of learned motor skills

We formulate task-related performance metrics for the quantitative evaluation of three locomotion tasks (Methods) and benchmark the following five settings in ten scenarios with the same DRL framework: (1) 'full-state policies' learned with the nine feedback states (Fig. 3a) and random robot-pose initialization, (2) 'key-state policies' learned with four key states and key-pose initialization, (3) 'irrelevant-state policies' that only use five less important states, (4) open-loop trajectories from full-state policies and (5) open-loop trajectories from key-state policies. Open-loop trajectories repeat the desired joint positions for two gait periods generated by the feedback policies.

We found that the average performance of key-state policies over all five performance metrics (Methods) is comparable to that of full-state policies for three skills (Fig. 5d), and if key feedback states are missing, there would be a substantial drop in task performance (forward velocity and heading accuracy for trotting and bounding) or learning success rate (balance recovery). More details can be found in Fig. 5a–c, Supplementary Fig. 12 and Supplementary Note 2.

Furthermore, the performance benchmark of the closed-loop control policies versus the open-loop trajectories (Supplementary Figs. 13 and 14) demonstrates the importance of utilizing feedback states to correct robot behaviours. Also, compared with the open-loop trajectories from the full-state policies with random explorations, results indicate that trajectories learned by the key-state policies are typically more stable when executed in an open-loop manner. This

suggests the neural network tends to discover more conservative and stable trajectory patterns surrounding the solution space initialized by the key poses.

Applicability to learning new skills

We further validate our approach to learn new locomotion skills. Using the key feedback states identified from three representative locomotion skills and the newly designed task-specific key poses (Supplementary Fig. 17), the A1 quadruped robot successfully learns robust pacing and galloping gaits (Fig. 6a,b and Supplementary Video 4).

Based on the *t*-distributed stochastic neighbour embedding[35] plot in Fig. 6d, we found that the trajectories of key-state pacing and galloping are located within the circle formed by the trajectories of balance recovery, trotting and bounding policies. This suggests that the studied skills encapsulate the common patterns of leg movements with sufficient diversity and the newly learned skills are closely related neighbourhood skills to the studied skills. Therefore, using the same set of key states can help in effectively learning both new gait patterns. However, learning more distinct new locomotion skills may require identifying new key feedback states, which can be achieved by reapplying our approach.

Correlation between feedback states

Heatmaps were generated to reflect the non-linear correlation across feedback states by mutual information[36] (Fig. 6e and Extended Data Fig. 3c). The average correlation coefficients between any two types of current feedback states were visualized using chord diagrams in Fig. 6f and Extended Data Fig. 3d. These findings reveal that the key states identified for balance recovery are correlated with all other task-irrelevant states to varying degrees. Moreover, the measurements of these key states are usually less noisy, making them more suitable to be used. Therefore, these results suggest that selecting key states with underlying correlations with other signals can effectively reduce the number of sensors required for feedback in the closed-loop control.

Discussion

This work has developed a quantitative analysis method for selecting feedback states for learning-based closed-loop control. As a result of motor learning through neural networks, the importance of states is indirectly encapsulated by a large amount of learned neural-network weights. Our method contributes to the interpretation, comparison and validation of the state importance by a direct ranking of quantitative relative importance over common sensory feedback for quadruped locomotion. The study suggests that joint positions, gravity vector and base linear and angular velocities are the essential states, composing 80% total relative importance for motor learning, whereas foot positions, foot contact, base position, joint torque and velocities are better to have but not necessary and thus can be excluded from motor learning without significantly affecting robot performance.

Benefits for robot-control design

Our method provides a quantitative ranking of feedback states and hence identifies the level of their importance, guiding the selection of a minimum and suitable set of sensors to learn robust quadruped locomotion. Different combinations of sensors can be chosen based on the hardware availability for particular applications. For example, we found that motor encoders, inertial measurement unit (IMU) and estimation of body linear velocities suffice to achieve common quadruped locomotion skills.

To summarize, this research benefits the design of robot control in several aspects:

(1) Improves design efficiency of the learning framework by selecting important feedback states through one single training session, which is more efficient than empirical trial-and-error or ablation studies that require multiple iterative processes.
(2) Promotes lightweight and cost-effective robot design by equipping only task-relevant sensors.
(3) Reduces the need for developing state estimation for task-irrelevant or unimportant states, thus reducing the dependencies of task success on sensing and state estimation uncertainties that helps to mitigate simulation-to-real mismatch.

Our quantitative analysis approach requires a successful sensorimotor policy, that is, a differentiable state–action mapping of the motor skills, to be used for identifying the importance level of states. In cases where motion learning is initially infeasible via reinforcement learning, all commonly available sensing shall be included to facilitate motor learning, or other approaches like supervised learning or imitation learning can be used if demonstrations are available.

Feedforward pathways

For simplicity, the periodic feedforward signals are implemented as a two-dimensional phase vector ($\sin 2\pi\phi, \cos 2\pi\phi$) generated directly using ϕ (temporal information, that is, 0–100% phase over a gait period; Fig. 2c), representing continuous periodicity as a form of priors or prior knowledge[37,38]. Thus, they are not modulated by sensory feedback signals in our study. In principle, these feedforward signals can be implemented as CPG networks and be modulated in various ways, for example, with phase resetting mechanisms or with feedback terms that continuously modulate the phase and amplitude of CPG signals, as in [32,39–42]. In future work, it would be interesting to include such feedback mechanisms and investigate whether the relative importance of different sensory modalities would change. For instance, it might lead to a higher importance of load feedback, which has been suggested to be important for cat locomotion[43] and shown to be a sufficient source of information for interlimb coordination[39].

Relation to biology

In general, our findings agree with biological findings and hypotheses. The key sensory feedback found by our studies on quadrupedal robots maps to the vestibular system and muscle spindles that have been proved to be critical for postural control and goal-directed vertebrate locomotion on the biological counterparts[25,44]. Our analysis also reveals that important states vary with tasks and certain sensory signals are more critical than others at different moments during a gait cycle, consistent with existing biological findings and hypotheses[25,45]. It is important to note that our conclusions on important states stem from common locomotion skills on a mechanical robot and may differ from the general findings in animals, for example, the importance of contact forces and limb loading as discussed in the previous section. Our contribution to biology is to provide biologists with the computational approach to identify important states on simulated animals, for example, neuromechanical simulations[46], when ethical or technical limitations prevent certain biological studies on live animals.

Limitations and future work

To obtain general conclusions on state importance with minimal human bias and good statistical characteristics, we designed basic reward functions that encourage the exploration of feasible but slightly different motor policies across training sessions. In contrast with state-of-the-art locomotion[10–12], we do not rely on reference trajectories or heavily fine-tuned reward terms to achieve natural-looking gaits. This allows the feedback control policy to fully exploit the solution space, drawing general conclusions about the importance of different states.

The identified key important states enabled successful learning of motor skills, demonstrating robustness to uncertainties in sensing, environment and robot dynamics (Supplementary Videos 1–4), which suggests that our saliency analysis based on the neural network, that is, the mapping from filtered state input to unfiltered action output, has captured essential features of the overall policy for the identification of key feedback states. As for future work, further investigation of other

https://doi.org/10.1038/s42256-023-00701-w

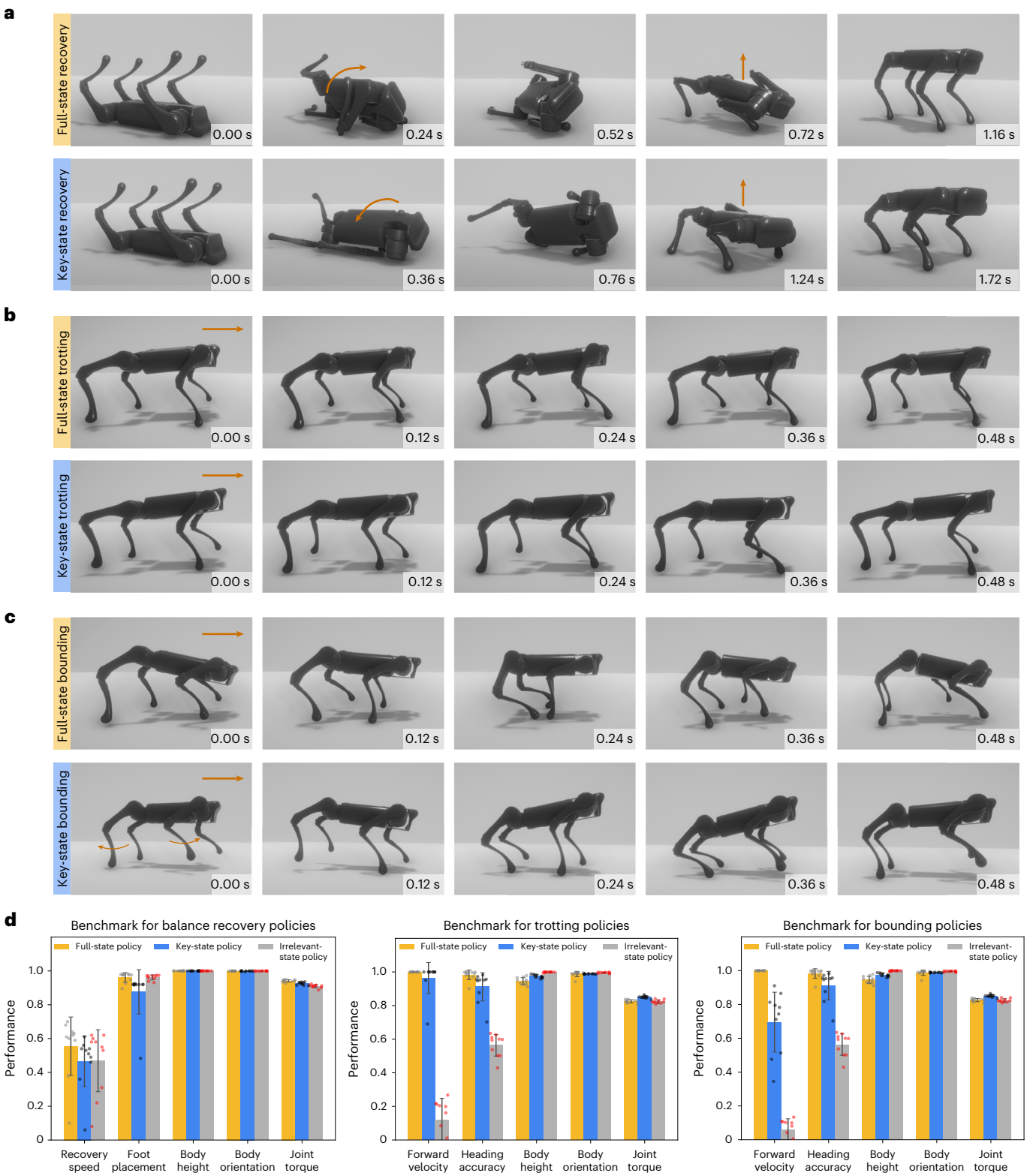
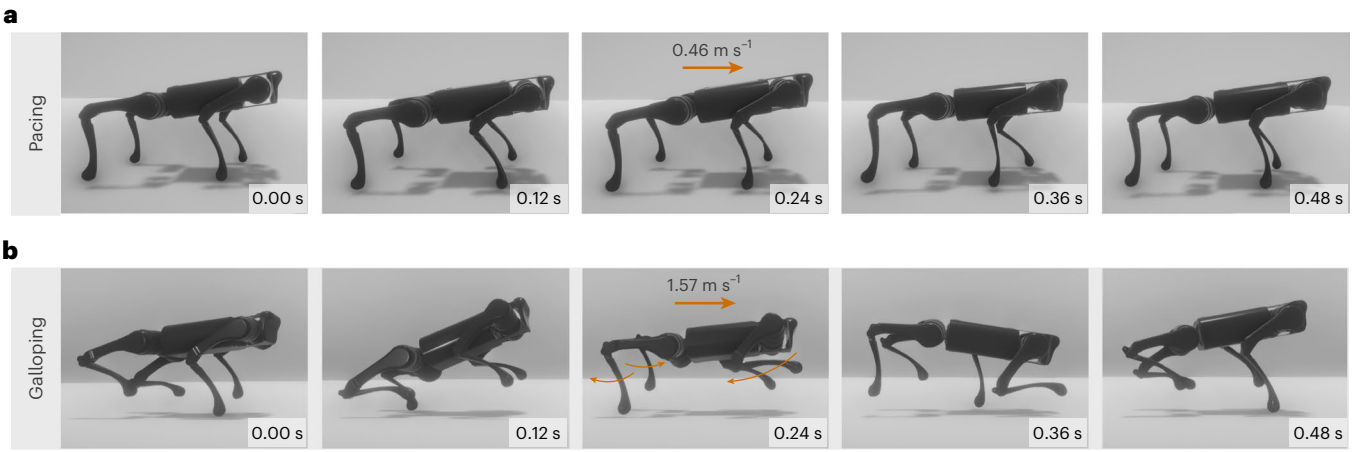


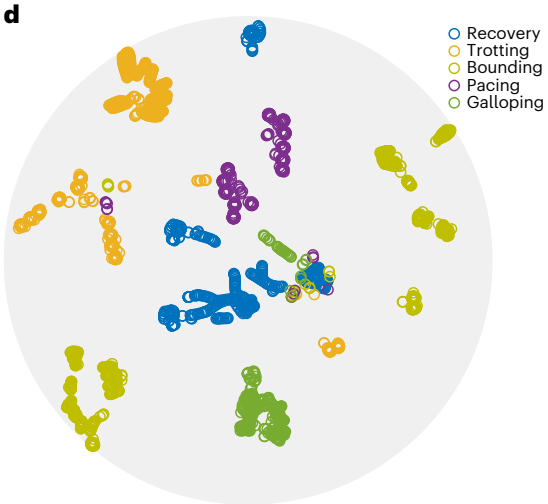
Fig. 5 | Benchmarking of task performance of full-state, key-state and irrelevant-state policies for balance recovery, trotting and bounding. **a**, Balance recovery from lying down on the back by a full-state policy (top) and a key-state policy (bottom). **b**, Learned trotting gait by a full-state policy (top) and a key-state policy (bottom). **c**, Learned bounding gait by a full-state policy (top) and a key-state policy (bottom). **d**, Comparison of task performance using metrics for full-state, key-state and irrelevant-state policies of balance recovery, trotting and bounding. Data are presented as mean values ± standard deviation (error bars) of $n = 10$ samples (scenarios). The dots on top of each bar plot are assigned different colours to distinguish samples from three policies. Higher values of metrics indicate better performance. Details of testing scenarios are in Supplementary Note 3 and Extended Data Fig. 6. The key-state policies achieved 94.7%, 99.1% and 93.7% of the task performance on average, respectively, with respect to the full-state policies. Irrelevant-state trotting and bounding policies achieve 11.7% and 5.9%, respectively, in forward velocity and 57.3% and 57.3%, respectively, in heading accuracy with respect to full-state policies. Although the selected irrelevant-state balance recovery policy achieves similar performance to full-state and key-state policies, the success rate of learning such policy drops substantially.

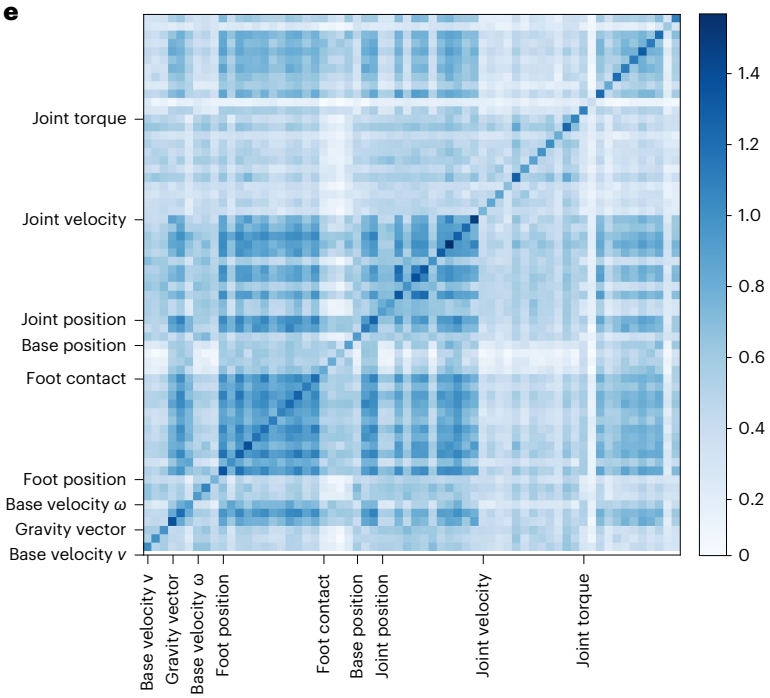
https://doi.org/10.1038/s42256-023-00701-w

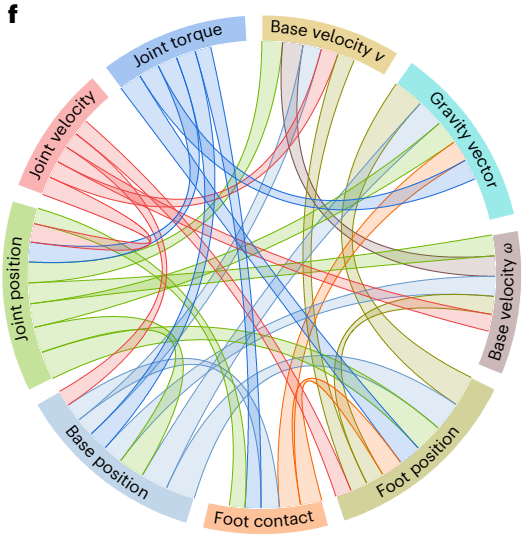


c

Feedback state	Recovery	Trotting	Bounding
Joint position	●	●	●
Gravity vector	●	●	●
Base linear velocity		●	●
Base angular velocity			●







components of the framework and the properties of the overall policy would be an interesting direction. When considering the accuracy of sensory feedback, the importance ranking of feedback states can be further refined by composing the saliency map and sensitivity matrix of sensor noise levels (Supplementary Note 4).

In future applications, in case certain states are identified as important but unreliable because of hardware limitations, one potential solution is to train a neural network to infer the estimation of such error-prone states from more reliable feedback. Prior work[47] has demonstrated the feasibility of such an approach, for example, estimating

Fig. 6 | Applicability to learning new locomotion skills using the key feedback states. a, Successful pacing gait learned by the A1 robot with an average forward velocity of 0.46 m s^{-1}. **b**, Successful galloping gait learned by the A1 robot with an average forward velocity of 1.57 m s^{-1}. **c**, Key feedback states used for balance recovery, trotting and bounding skills. Each colour represents one type of feedback state: blue, green, yellow and orange dots represent joint position, gravity vector and base linear and angular velocities, respectively. **d**, The two-dimensional projections of the 18-dimensional trajectories (body height, body linear velocity, roll and pitch angles and joint positions) sampled from key-state pacing and galloping policies, full-state and key-state balance recovery, trotting and bounding policies using *t*-distributed stochastic neighbour embedding. **e**, Heatmap showing the non-linear correlation between any two dimensions of the nine feedback states from full-state balance recovery policy, where a darker colour indicates stronger correlation. **f**, Chord diagram summarizing the correlation between any two feedback states for full-state balance recovery (self-correlation and percentage <25% are removed for clarity), where the wider the link between any two states, the stronger they correlate with each other.

body velocity and foot contact from joint positions measured by motor encoders, and the gravity vector plus base angular velocities obtained from IMU measurements.

Methods

Robot platform

We chose the A1 quadruped robot[33] for our study, which approximates a small dog and is commonly used for locomotion research (see robot specifications in Supplementary Table 4). Extensive simulation validations were conducted in a physics-based simulation—PyBullet[48]. All the robot locomotion tasks were simulated in PyBullet with high-fidelity physics, and the resulting robot motions and data were rendered in Unity[49] for high-resolution snapshots and videos, which provide better visualization quality of the physical interactions and movements.

Robot motor skills

In this framework, the saliency analysis approach requires a differentiable state–action mapping of the motor skills, in the form of a neural network. Therefore, we trained the neural network on physics simulation data to map the robot states to the corresponding actions. Thus, the trained neural network represents a motor skill and computes the robot actions in response to given input signals, which allows us to apply quantitative analysis and identify the importance level of each feedback state in such state–action mapping.

Saliency analysis

Integrated gradients. Our use of the saliency analysis is inspired by feature attribution methods in image classification and explainable artificial intelligence[15,16,34,50]. There are several saliency methods or attribution methods, such as integrated gradients[34], guided backpropagation[51], DeepLift[52] and gradient-weighted class activation mapping[53]. Here, we use an integrated-gradients method to define the saliency values of feedback states for a learned policy. Integrated gradients satisfies two axioms that are desirable for attribution methods[34]: (1) 'sensitivity', that is, the attribution should be non-zero if each input and baseline lead to different outputs; and (2) 'implementation invariance', that is, the attributions should be the same for two networks if the outputs of both networks are identical for all the inputs, regardless of the detailed implementation.

Some other common saliency methods are not able to satisfy both. For example, vanilla gradients of the output with respect to the input and guided backpropagation break the axiom sensitivity[52], which will result in the gradients focusing on irrelevant features. Another common technique DeepLift breaks the axiom implementation invariance, where results may differ for the networks with same functionalities but different implementation. In summary, the use of integrated gradients allows us to identify feedback states that are truly relevant. The same conclusion holds for each type of locomotion skill regardless of the implementation of deep neural networks, as long as the outputs of two implementations are the same for the identical inputs.

It shall be noted that the integrated-gradients method is applicable to the state–action mapping that is differentiable, meaning that it can analyse the influence of feedback states of motor skills that can be represented by differentiable machine learning models. However, it cannot be applied to underlying state–action policies that are non-differentiable. At time step t, for the ith dimension of feedback states $x \in \mathbb{R}^n$, integrated gradients $G(x_{i,t})$ are defined as follows:

$$G(x_{i,t}) = \sum_{j=1}^{m} \left| \frac{(x_{i,t} - \hat{x}_{i,t})}{p} \sum_{k=1}^{p} \frac{\partial F_j(\hat{x}_t + k/p(x_t - \hat{x}_t))}{\partial x_{i,t}} \right| \quad (1)$$

where $F(x_t) \in \mathbb{R}^m$ represents the generated actions at time step t, $\hat{x}_{i,t}$ is the baseline input zero and $p = 25$ is the number of steps in the Riemann approximation of the integral. The partial derivative is computed through backpropagation by calling tf.gradients() in TensorFlow[54]. For a better visualization to reveal the relative importance among feedback states via saliency maps, we define the raw saliency value $S_d(x_{i,t})$ as follows instead of directly using the computed integrated gradients:

$$\epsilon = \frac{1}{nN} \sum_{t=1}^{N} \sum_{i=1}^{n} G(x_{i,t}) \quad (2)$$

$$S_d(x_{i,t}) = \begin{cases} G(x_{i,t}) - \epsilon, & G(x_{i,t}) > \epsilon \\ 0, & \text{else} \end{cases} \quad (3)$$

where N is the number of total time steps during the entire motion. The raw saliency value $S_d(x_{i,t})$ is further normalized to the range of [0, 1] as follows:

$$S(x_{i,t}) = S_d(x_{i,t}) / \max_{\substack{i \in \{1,2,\ldots,n\} \\ t \in \{1,2,\ldots,N\}}} S_d(x_{i,t}) \quad (4)$$

Relative importance of feedback states. For the ith dimension of feedback states $x \in \mathbb{R}^n$, overall importance during the entire motion I_i is computed as follows:

$$I_i = \sum_{t=1}^{N} S(x_{i,t}) \quad (5)$$

where $S(x_{i,t})$ is the saliency value for x_i at time step t and N is the number of total time steps during the entire motion.

For feedback state $o \in \mathbb{R}^h$, $(h \leq n)$, overall importance I_o is computed as follows:

$$I_o = \frac{1}{h} \sum_{q=1}^{h} I_{i(o,q)} \quad (6)$$

where $i(o, q)$ is the index of the dimension of feedback states x that maps to the qth dimension of feedback state o (see Supplementary Table 1).

This work considers nine feedback states in total. For feedback state o, relative importance r_o is defined as follows:

$$r_o = \frac{I_o}{\sum_{o=1}^{9} I_o} \quad (7)$$

Feedback states for learning locomotion skills

Here we introduce a longlist of nine candidate states used in legged locomotion and the way to measure or estimate them on a real robot: (1) base position in the world frame that can be estimated using visual inertial odometry[55], the three-dimensional base position was used rather than the base height alone for a fair comparison with other states; (2) normalized gravity vector in the robot's local frame that reflects the body orientation of the robot and can be computed using roll and pitch angle measurements from IMU; (3) base angular velocity measured by IMU; (4) base linear velocity in the robot heading frame estimated by fusing leg kinematics and the acceleration from IMU; (5) joint position measured by motor encoders; (6) joint velocity that is measured by motor encoders and further normalized by maximum joint velocity; (7) joint torque that is measured by torque sensors and further normalized by maximum joint torque; (8) foot position relative to base in the robot heading frame that can be computed through forward kinematics; and (9) foot contact with the ground that is computed by applying sigmoid function to the L^2 norm of contact force F_i measured by the force sensor at the end of the ith foot as follows:

$$\frac{1}{1+e^{-c_1(F_i-c_2)}}, i=1,2,3,4 \tag{8}$$

where $c_1 = c_2 = 2.0$. Thus, foot contact is continuous within the range of [0, 1] (an example of continuous foot contact is in Extended Data Fig. 5) without a discontinuous switch between zero and one as in a threshold function that may affect the differentiability of the neural network for applying our analysis. For learning periodic locomotion tasks, such as trotting and bounding, we included a two-dimensional feedforward phase vector ($\sin 2\pi\phi$, $\cos 2\pi\phi$) on top of the above set of feedback states to represent continuous temporal information that encodes phase ϕ from 0% to 100% of a gait period. At each time step, the phase increases by a constant increment without any phase-resetting mechanisms and is computed as follows:

$$\phi = \frac{k \bmod (Tf_c)}{Tf_c} \tag{9}$$

where k is the control step counter, T is desired gait period and f_c is the control frequency. The phase vector is the feedforward term and, thus, was excluded for the quantification and comparison of the state importance, because the focus of this study is on the feedback terms.

For full-state policies, this complete set of feedback states was used for balance recovery (without the feedforward phase vector), trotting and bounding. For key-state policies, the states used were different for the three locomotion skills (Fig. 6c). Specifically, states (2) and (5) were used for balance recovery, states (2), (4), (5) and phase vector were used for trotting and states (2), (3), (4), (5) and phase vector were used for bounding. Learning pacing and galloping skills used the same set of states as for bounding.

Key-pose taxonomy for effective exploration and learning

During training control policies using full feedback states, the robot pose is initialized with a random configuration at each training episode to encourage the exploration of diverse states, which is a technique commonly used in robot learning[9,18]. However, random initialization is not data efficient in exploring the state space for a type of locomotion task, as most robot configurations are a priori invalid, because of the physical feasibility of the balance criteria. In other words, most of the robot configurations are not balanced or very far away from the desired locomotion, and therefore, the collected samples are skewed by invalid exploration and less efficient for learning. To this end, on top of the key feedback states, we propose key-pose taxonomy to initialize the robot configuration at each training episode, as seeding conditions to enable more effective exploration and learning.

Inspired by animal locomotion[56,57] and whole-body support pose taxonomy from humanoid robots[58,59], given a specific locomotion task, we can design key-pose taxonomy that consists of representative robot–ground contact configurations and distinct robot poses. The robot–ground contact configurations are straightforward to obtain because quadrupedal locomotion is well studied in biology and we can easily obtain representative contact phases, for example, trotting of dogs and horses. Compared to the existing pose taxonomy[58,59] that is for the classification and inter-transitions of loco-manipulation, our proposed key-pose taxonomy here aims at task-specific effective learning.

Specifically, we use the configuration space of a floating-base robot to define each key pose, which is composed of body height, body orientation (roll and pitch angles) and joint angles. The base linear velocity and base angular velocity were set as zero at the start of each episode. Given the same contact configuration, we shall note that a quadruped robot may have multiple poses. For example, crouching and standing share the same robot–ground contact by four feet. Therefore, to balance the aspect of diversity, based on each ground contact or gait phase, we can use the robot configuration space to define multiple distinct key poses, so as to increase the number of initial poses that can sparsely cover the feasible motions related to a task.

Following this principle, we designed five key poses for balance recovery, six for trotting, four for bounding, four for pacing and five for galloping as shown in Fig. 2b and Supplementary Fig. 17. The key poses within the designed taxonomy were sampled to initialize the robot pose at each training episode for each task, and the detailed transitions between the key poses will be explored during the learning process and, thus, obtained as a natural outcome. Using key-pose taxonomy as initial posture setting makes learning more efficient by narrowing the solution space for learning compared to random exploration.

In this work, we assigned the key poses within the taxonomy with equal probability for the DRL agent to encounter and explore upon. It shall be noted that the probability of each key pose can be more flexible. For example, we can assign higher probabilities to the key poses that are task relevant but less likely to be encountered in natural interactions with the environment.

Quantitative metrics of performance evaluation

To quantify the performance for comprehensive comparison, a set of performance metrics, S, is designed for each task. For balance recovery, the performance metric set $S_{\text{recovery}} = \{s_\tau, s_r, s_f, s_{h_N}, s_{\phi_N}\}$. For trotting and bounding, the performance metric sets S_{trotting} and S_{bounding} are the same, that is, $S = \{s_\tau, s_v, s_\psi, s_h, s_\phi\}$. Note that the performance metrics are used for post-learning performance evaluations, which are not the same as the reward terms designed for learning in terms of formulations and weights for each physical quantity.

The metric value for each physical quantity is in the range of [0, 1], and N is the number of total time steps of an episode. Joint torque is used for the performance evaluation across all the three locomotion tasks. The performance metrics for this physical quantity are evaluated as follows. A higher value of joint torque metric indicates a more energy-efficient motion.

$$s_\tau = 1 - \frac{1}{12N}\sum_{i=1}^{12}\sum_{t=1}^{N}|\tau_{i,t}|/\hat{\tau} \tag{10}$$

where $\tau_{i,t}$ is the joint torque of the ith joint at time step t and $\hat{\tau} = 33.5$ Nm is the maximum joint torque.

Performance metrics for balance recovery. For balance recovery, the performance metrics for the other four physical quantities are evaluated as follows:

(1) Recovery speed metric

$$s_r = 1 - T/\hat{T} \tag{11}$$

where T is the time duration of recovery to a standing posture and $\hat{T}$ is the time duration from the start of recovery to the end of an episode. A higher value indicates the recovery is completed within a shorter time period.

(2) Final foot placement metric

$$s_f = 1 - \frac{1}{8}\sum_{i=1}^{8} |\mathbf{p}_{f,i} - \hat{\mathbf{p}}_{f,i}|/\hat{d} \tag{12}$$

where $\mathbf{p}_f$ and $\hat{\mathbf{p}}_f$ are vectors of final and nominal foot positions, respectively, in the horizontal plane of the robot heading frame. $\hat{\mathbf{p}}_f = [0.18\,\text{m}, 0.13\,\text{m}, -0.18\,\text{m}, -0.13\,\text{m}, -0.18\,\text{m}, 0.13\,\text{m}, 0.18\,\text{m}, -0.13\,\text{m}]$ and $\hat{d} = 0.3\,\text{m}$ for A1 quadruped robot. A higher value indicates that the four feet are closer to the nominal foot positions at the end of recovery.

(3) Final body height metric

$$s_{h_N} = \min\left(h_N, \hat{h}\right)/\hat{h} \tag{13}$$

where h_N is the body height at the end of recovery and $\hat{h} = 0.25\,\text{m}$ is the nominal standing height of the robot. A higher value means that the final body height is closer to the nominal standing height of the robot.

(4) Final body orientation metric

$$s_{\phi_N} = (\mathbf{g}_N\hat{\mathbf{g}} + 1)/2 \tag{14}$$

where $\mathbf{g}_N$ is the gravity vector of the robot at the end of recovery and $\hat{\mathbf{g}} = [0, 0, -1]$ is the nominal gravity vector of the robot. A higher value indicates that the final body orientation is closer to the nominal body orientation, that is, zero roll angle and pitch angle.

Performance metrics for trotting and bounding. For trotting and bounding, the other four physical quantities are the same and the performance metrics are evaluated as follows:

(1) Forward velocity metric

$$s_v = \min\left(\frac{1}{N}\sum_{t=1}^{N} V_t, \hat{V}\right)/\hat{V} \tag{15}$$

where V_t is the forward velocity in the horizontal plane at time step t, $\hat{V}$ is the nominal forward velocity, and $\hat{V} = 0.5\,\text{m}\,\text{s}^{-1}$ for trotting and $\hat{V} = 1.0\,\text{m}\,\text{s}^{-1}$ for bounding. A higher value indicates a faster average forward velocity during the entire episode. It shall be noted that a lower value for bounding does not indicate worse learning performance. The reason is that we did not set a desired forward velocity strictly in the reward for training bounding as for trotting. We set the desired velocity for training bounding as $1.0\,\text{m}\,\text{s}^{-1}$. However, we do not penalize velocity higher than $1.0\,\text{m}\,\text{s}^{-1}$ to encourage higher velocity if possible. As a result, the robot may learn bounding policies with different average forward velocities with the same training settings.

(2) Heading accuracy metric

$$s_\psi = \left(\frac{1}{N}\sum_{t=1}^{N} \frac{\mathbf{v}_{\text{h},t}\hat{\mathbf{v}}_\text{h}}{|\mathbf{v}_{\text{h},t}||\hat{\mathbf{v}}_\text{h}|} + 1\right)/2 \tag{16}$$

where $\mathbf{v}_{\text{h},t}$ is the velocity vector of the robot in the horizontal plane of the robot heading frame at time step t, $\hat{\mathbf{v}}_\text{h}$ is the nominal velocity vector in the horizontal plane, $\hat{\mathbf{v}}_\text{h} = [0.5\,\text{m}\,\text{s}^{-1}, 0\,\text{m}\,\text{s}^{-1}]$ for trotting and $\hat{\mathbf{v}}_\text{h} = [1.0\,\text{m}\,\text{s}^{-1}, 0\,\text{m}\,\text{s}^{-1}]$ for bounding, and $|\cdot|$ is the magnitude of the vector. A higher value indicates better tracking of the nominal heading during the entire episode.

(3) Body height metric

$$s_h = \frac{1}{N}\sum_{t=1}^{N} \min(h_t, \hat{h})/\hat{h} \tag{17}$$

where h_t is the body height at time step t and $\hat{h} = 0.3\,\text{m}$ is the nominal height of the robot. A higher value means that the body height is closer to the nominal height during the entire episode.

(4) Body orientation metric

$$s_\phi = \left(\frac{1}{N}\sum_{t=1}^{N} \mathbf{g}_t\hat{\mathbf{g}} + 1\right)/2 \tag{18}$$

where $\mathbf{g}_t$ is the gravity vector of the robot at time step t and $\hat{\mathbf{g}} = [0, 0, -1]$ is the nominal gravity vector of the robot. A higher value indicates that the body orientation is closer to the nominal body orientation during the entire episode, that is, zero roll angle and pitch angle.

Metrics for task-performance evaluation. Given the five individual performance metrics for each locomotion task, we can further evaluate overall performance of key-state policies with respect to full-state policies. Consider a task-related physical quantity i, metric for the key-state policy $s_{i,\text{key}}$ and metric for the full-state policy $s_{i,\text{full}}$, overall performance of the key-state policy s_key is defined as follows:

$$s_\text{key} = \frac{1}{5}\sum_{i=1}^{5} \frac{s_{i,\text{key}}}{s_{i,\text{full}}}. \tag{19}$$

Furthermore, we can compute the mean of s_key for multiple tasks to evaluate the overall performance (balance recovery, trotting and bounding) in a statistical manner. The overall performance of irrelevant-state policies and open-loop trajectories are evaluated in the same way.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Source data for main text figures are available at https://github.com/yuwanming/feedback_importance_data. All other data that support the plots within this paper and other findings of this study are available from the corresponding author upon reasonable request.

Code availability

The code for training locomotion skills and saliency analysis is available at https://github.com/yuwanming/A1_quadruped_env (ref. 60).

References

1. Wiener, N. *Cybernetics or Control and Communication in the Animal and the Machine* (MIT Press, 2019).
2. Ijspeert, A. J. Biorobotics: using robots to emulate and investigate agile locomotion. *Science* **346**, 196–203 (2014).
3. Karakasiliotis, K. et al. From cineradiography to biorobots: an approach for designing robots to emulate and study animal locomotion. *J. R. Soc. Interface* **13**, 20151089 (2016).
4. Nyakatura, J. A. et al. Reverse-engineering the locomotion of a stem amniote. *Nature* **565**, 351–355 (2019).
5. Cheng, G., Ehrlich, S. K., Lebedev, M. & Nicolelis, M. A. Neuroengineering challenges of fusing robotics and neuroscience. *Sci. Robot.* **5**, 7–10 (2020).
6. Kalashnikov, D. et al. Scalable deep reinforcement learning for vision-based robotic manipulation. In *Proc. of The 2nd Conference on Robot Learning* (eds Billard, A., Dragan, A., Peters, J. & Morimoto, J.) 651–673 (PMLR, 2018).
7. Xie, Z., Da, X., van de Panne, M., Babich, B. & Garg, A. Dynamics randomization revisited: a case study for quadrupedal locomotion. In *2021 IEEE International Conference on Robotics and Automation* 4955–4961 (IEEE, 2021).

8. Ibarz, J. et al. How to train your robot with deep reinforcement learning: lessons we have learned. *Int. J. Rob. Res.* **40**, 698–721 (2021).
9. Hwangbo, J. et al. Learning agile and dynamic motor skills for legged robots. *Sci. Robot.* **4**, eaau5872 (2019).
10. Lee, J., Hwangbo, J., Wellhausen, L., Koltun, V. & Hutter, M. Learning quadrupedal locomotion over challenging terrain. *Sci. Robot.* **5**, eabc5986 (2020).
11. Peng, X. B. et al. Learning agile robotic locomotion skills by imitating animals. In *Proc. Robotics: Science and Systems* (eds Toussaint, M. Bicchi, A. & Hermans, T.) (2020).
12. Yang, C., Yuan, K., Zhu, Q., Yu, W. & Li, Z. Multi-expert learning of adaptive legged locomotion. *Sci. Robot.* **5**, eabb2174 (2020).
13. Haarnoja, T. et al. Learning to walk via deep reinforcement learning. In *Proc. Robotics: Science and Systems* (eds Bicchi, A., Kress-Gazit, H. & Hutchinson, S.) (2019).
14. Tan, J. et al. Sim-to-real: learning agile locomotion for quadruped robots. In *Proc. Robotics: Science and Systems* (eds Kress-Gazit, H., Srinivasa, S., Howard, T. & Atanasov, N.) (2018).
15. Yosinski, J., Clune, J., Nguyen, A., Fuchs, T. & Lipson, H. Understanding neural networks through deep visualization. Deep Learning Workshop at *International Conference on Machine Learning* (2015).
16. Jiménez-Luna, J., Grisoni, F. & Schneider, G. Drug discovery with explainable artificial intelligence. *Nat. Mach. Intell.* **2**, 573–584 (2020).
17. Tassa, Y. et al. Deepmind control suite. Preprint at https://arxiv.org/abs/1801.00690 (2018).
18. Reda, D., Tao, T. & van de Panne, M. Learning to locomote: understanding how environment design matters for deep reinforcement learning. In *Proc. 13th ACM SIGGRAPH Conference on Motion, Interaction and Games* (ACM, 2020).
19. Marasco, P. D. et al. Neurorobotic fusion of prosthetic touch, kinesthesia, and movement in bionic upper limbs promotes intrinsic brain behaviors. *Sci. Robot.* **6**, eabf3368 (2021).
20. Thandiackal, R. et al. Emergence of robust self-organized undulatory swimming based on local hydrodynamic force sensing. *Sci. Robot.* **6**, eabf6354 (2021).
21. Shao, Y. et al. Learning free gait transition for quadruped robots via phase-guided controller. *IEEE Robot. Autom. Lett.* **7**, 1230–1237 (2021).
22. Smith, L. et al. Legged robots that keep on learning: fine-tuning locomotion policies in the real world. In *2022 International Conference on Robotics and Automation* 1593–1599 (IEEE, 2022).
23. Margolis, G. B., Yang, G., Paigwar, K., Chen, T. & Agrawal, P. Rapid locomotion via reinforcement learning. In *Proc. Robotics: Science and Systems* (eds Hauser, K., Shell, D. & Huang, S.) (2022).
24. Dickinson, M. H. et al. How animals move: an integrative view. *Science* **288**, 100–106 (2000).
25. Rossignol, S., Dubuc, R. & Gossard, J.-P. Dynamic sensorimotor interactions in locomotion. *Physiol. Rev.* **86**, 89–154 (2006).
26. Taylor, G. K. & Krapp, H. G. Sensory systems and flight stability: what do insects measure and why? *Adv. Insect. Phys.* **34**, 231–316 (2007).
27. Carpenter, R. & Reddi, B. *Neurophysiology: A Conceptual Approach* (CRC Press, 2012).
28. Roth, E., Hall, R. W., Daniel, T. L. & Sponberg, S. Integration of parallel mechanosensory and visual pathways resolved through sensory conflict. *Proc. Natl Acad. Sci. USA* **113**, 12832–12837 (2016).
29. Cox, S., Ekstrom, L. & Gillis, G. The influence of visual, vestibular, and hindlimb proprioceptive ablations on landing preparation in cane toads. *Integr. Comp. Biol.* **58**, 894–905 (2018).
30. Sober, S. J. & Sabes, P. N. Flexible strategies for sensory integration during motor planning. *Nat. Neurosci.* **8**, 490–497 (2005).
31. Pearson, K., Ekeberg, Ö. & Büschges, A. Assessing sensory function in locomotor systems using neuro-mechanical simulations. *Trends Neurosci.* **29**, 625–631 (2006).
32. Bellegarda, G. & Ijspeert, A. CPG-RL: learning central pattern generators for quadruped locomotion. *IEEE Robot. Autom. Lett.* **7**, 12547–12554 (2022).
33. Unitree A1. *Unitree* https://www.unitree.com/products/a1 (accessed 2 December 2022).
34. Sundararajan, M., Taly, A. & Yan, Q. Axiomatic attribution for deep networks. In *International Conference on Machine Learning* (eds Precup, D. & Teh, Y. W.) 3319–3328 (PMLR, 2017).
35. Van der Maaten, L. & Hinton, G. Visualizing data using t-SNE. *J. Mach. Learn. Res.* **9**, 2579–2605 (2008).
36. Cover, T. M. & Thomas, J. A. In *Elements of Information Theory* Ch. 2, 13–55 (John Wiley and Sons, 2005).
37. Jonschkowski, R. & Brock, O. State representation learning in robotics: using prior knowledge about physical interaction. In *Proc. Robotics: Science and Systems* (eds Fox, D., Kavraki, L. E. & Kurniawati, H.) (2014).
38. Yang, C., Yuan, K., Heng, S., Komura, T. & Li, Z. Learning natural locomotion behaviors for humanoid robots using human bias. *IEEE Robot. Autom. Lett.* **5**, 2610–2617 (2020).
39. Owaki, D., Kano, T., Nagasawa, K., Tero, A. & Ishiguro, A. Simple robot suggests physical interlimb communication is essential for quadruped walking. *J. R. Soc. Interface* **10**, 20120669 (2013).
40. Aoi, S. & Tsuchiya, K. Stability analysis of a simple walking model driven by an oscillator with a phase reset using sensory feedback. *IEEE Trans. Robot.* **22**, 391–397 (2006).
41. Owaki, D. & Ishiguro, A. A quadruped robot exhibiting spontaneous gait transitions from walking to trotting to galloping. *Sci. Rep.* **7**, 277 (2017); https://doi.org/10.1038/s41598-017-00348-9
42. Fujiki, S. et al. Adaptive hindlimb split-belt treadmill walking in rats by controlling basic muscle activation patterns via phase resetting. *Sci. Rep.* **8**, 17341 (2018).
43. Ekeberg, O. & Pearson, K. Computer simulation of stepping in the hind legs of the cat: an examination of mechanisms regulating the stance-to-swing transition. *J. Neurophysiol.* **94**, 4256–4268 (2005).
44. Grillner, S., Wallén, P., Saitoh, K., Kozlov, A. & Robertson, B. Neural bases of goal-directed locomotion in vertebrates—an overview. *Brain Res. Rev.* **57**, 2–12 (2008).
45. Caggiano, V. et al. Midbrain circuits that set locomotor speed and gait selection. *Nature* **553**, 455–460 (2018).
46. Hase, K., Miyashita, K., Ok, S. & Arakawa, Y. Human gait simulation with a neuromusculoskeletal model and evolutionary computation. *J. Vis. Comput. Animat.* **14**, 73–92 (2003).
47. Ji, G., Mun, J., Kim, H. & Hwangbo, J. Concurrent training of a control policy and a state estimator for dynamic and robust legged locomotion. *IEEE Robot. Autom. Lett.* **7**, 4630–4637 (2022).
48. Coumans, E. & Bai, Y. Pybullet, a python module for physics simulation for games, robotics and machine learning. https://pybullet.org (2019).
49. Juliani, A. et al. Unity: a general platform for intelligent agents. Preprint at https://arxiv.org/abs/1809.02627 (2018).
50. Simonyan, K., Vedaldi, A. & Zisserman, A. Deep inside convolutional networks: visualising image classification models and saliency maps. Workshop at *International Conference on Learning Representations* (2014).
51. Springenberg, J. T., Dosovitskiy, A., Brox, T. & Riedmiller, M. Striving for simplicity: the all convolutional net. Workshop at *International Conference on Learning Representations* (2015).
52. Shrikumar, A., Greenside, P., Shcherbina, A. & Kundaje, A. Learning important features through propagating activation differences. In *International Conference on Machine Learning* (eds Precup, D. & Teh, Y. W.) 3145–3153 (PMLR, 2017).

53. Selvaraju, R. R. et al. Grad-CAM: visual explanations from deep networks via gradient-based localization. In *Proc. IEEE International Conference on Computer Vision* 618–626 (IEEE, 2017).
54. Abadi, M. et al. Tensorflow: a system for large-scale machine learning. In *12th USENIX Symposium on Operating Systems Design and Implementation* 265–283 (2016).
55. Mourikis, A. I. & Roumeliotis, S. I. A multi-state constraint Kalman filter for vision-aided inertial navigation. In *Proc. 2007 IEEE International Conference on Robotics and Automation* 3565–3572 (IEEE, 2007).
56. Alexander, R. M. *Principles of Animal Locomotion* (Princeton Univ. Press, 2013).
57. Biewener, A. & Patek, S. *Animal Locomotion* (Oxford Univ. Press, 2018).
58. Borras, J. & Asfour, T. A whole-body pose taxonomy for loco-manipulation tasks. In *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems* 1578–1585 (IEEE, 2015).
59. Borràs, J., Mandery, C. & Asfour, T. A whole-body support pose taxonomy for multi-contact humanoid robot motions. *Sci. Robot.* **2**, eaaq0560 (2017).
60. Yu, W. & Yang, C. A1 quadruped env. Zenodo https://doi.org/10.5281/zenodo.8006935 (2023).

Acknowledgements

We gratefully acknowledge Q. Rouxel for providing valuable suggestions to improve the technical quality of the early version of this manuscript.

Author contributions

W.Y., C.Y. and Z.L. conceived the study on systematic state selection. W.Y. and C.Y. implemented the DRL framework. W.Y. implemented the quantification of state importance and collected data. W.Y., C.Y., G.B., M.S., A.J.I. and Z.L. designed the simulation experiments and contributed to data analysis. Z.L. directed and managed the research and provided scientific and technical solutions. W.Y. and Z.L. wrote the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

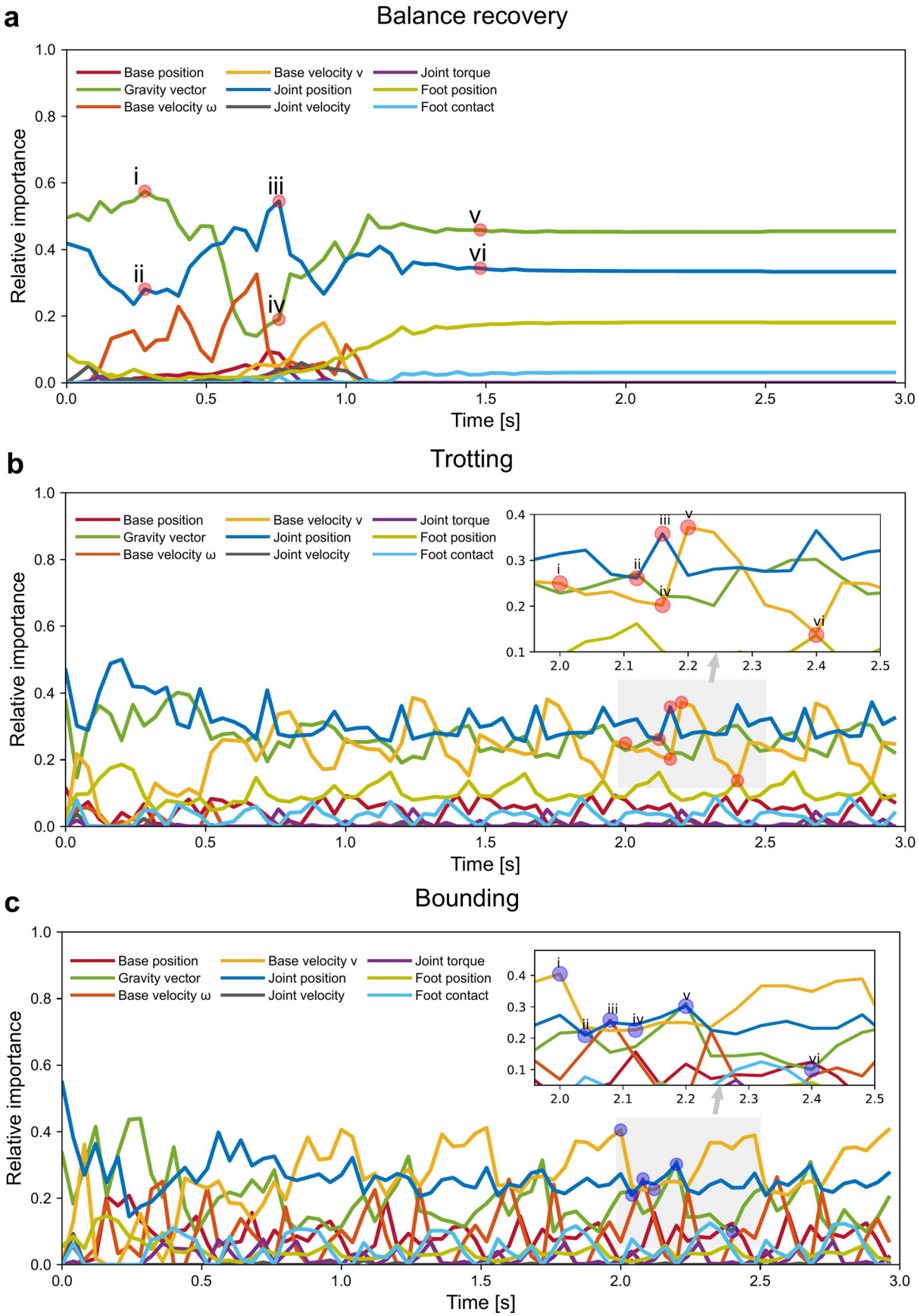
Extended data is available for this paper at https://doi.org/10.1038/s42256-023-00701-w.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s42256-023-00701-w.

Correspondence and requests for materials should be addressed to Zhibin Li.

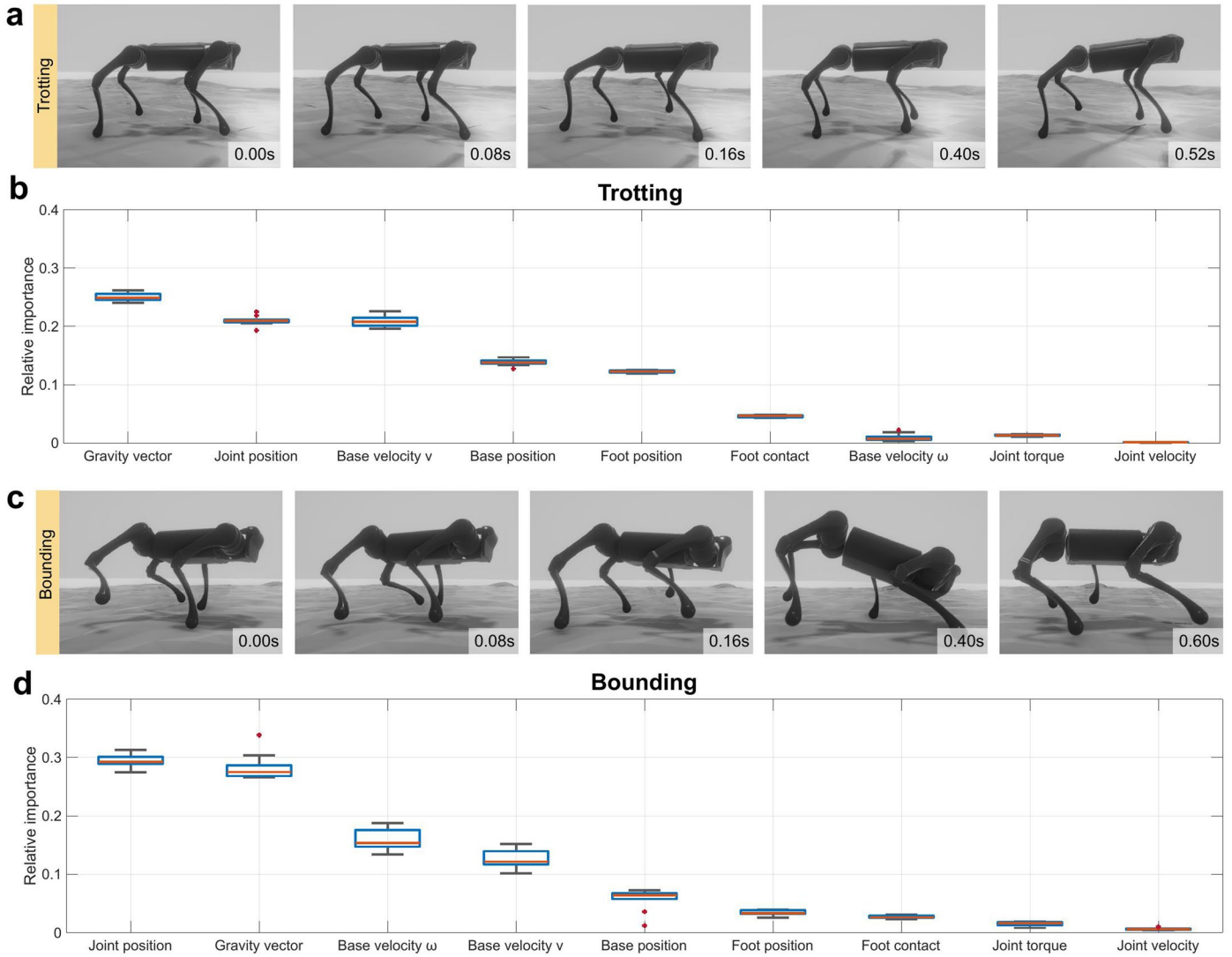
Peer review information *Nature Machine Intelligence* thanks Owaki Dai and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

https://doi.org/10.1038/s42256-023-00701-w



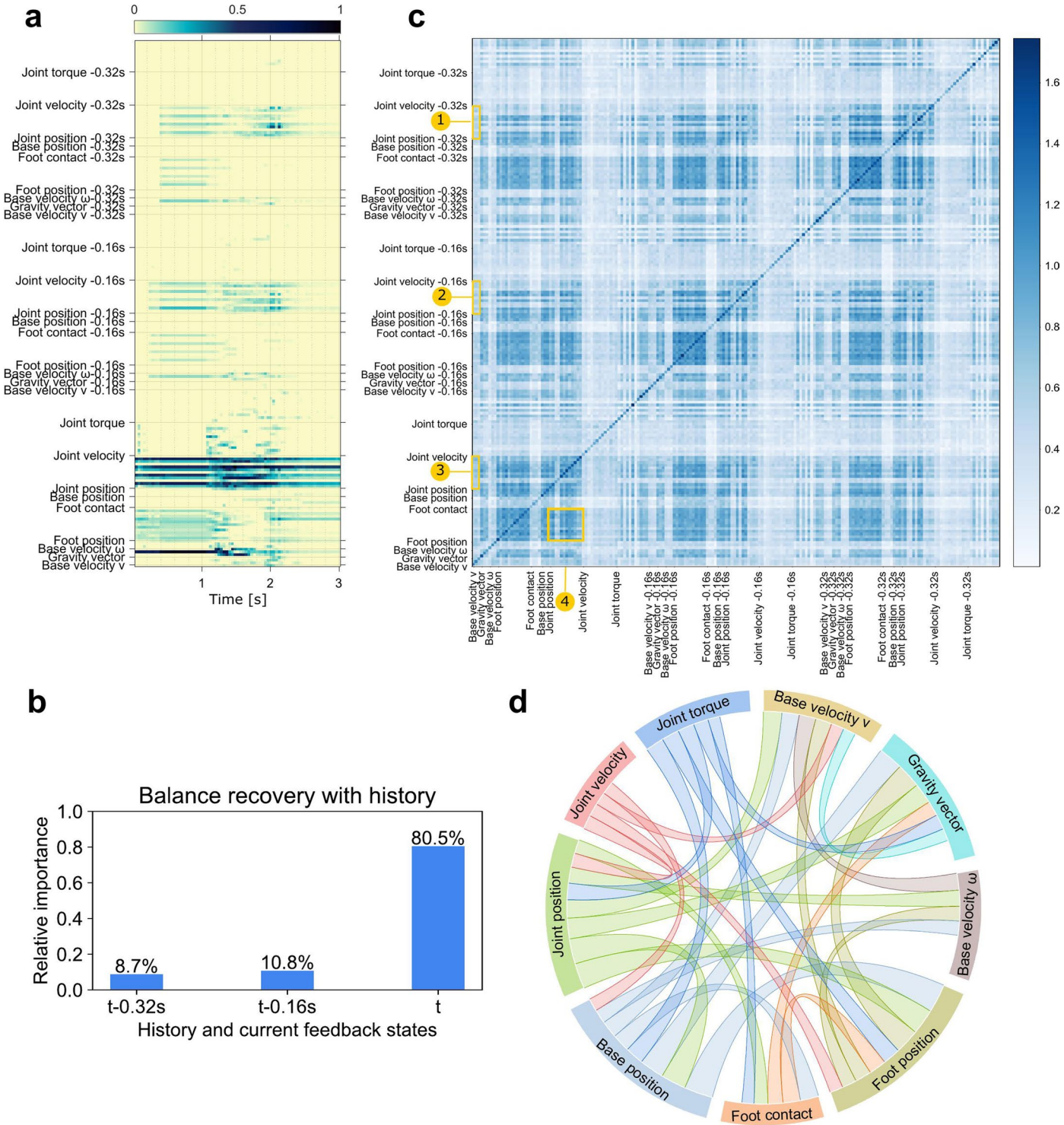
Extended Data Fig. 1 | Relative importance of nine feedback states over time. Relative importance of nine feedback states during 0-3 s. a, Balance recovery. b, Trotting. c, Bounding.

https://doi.org/10.1038/s42256-023-00701-w



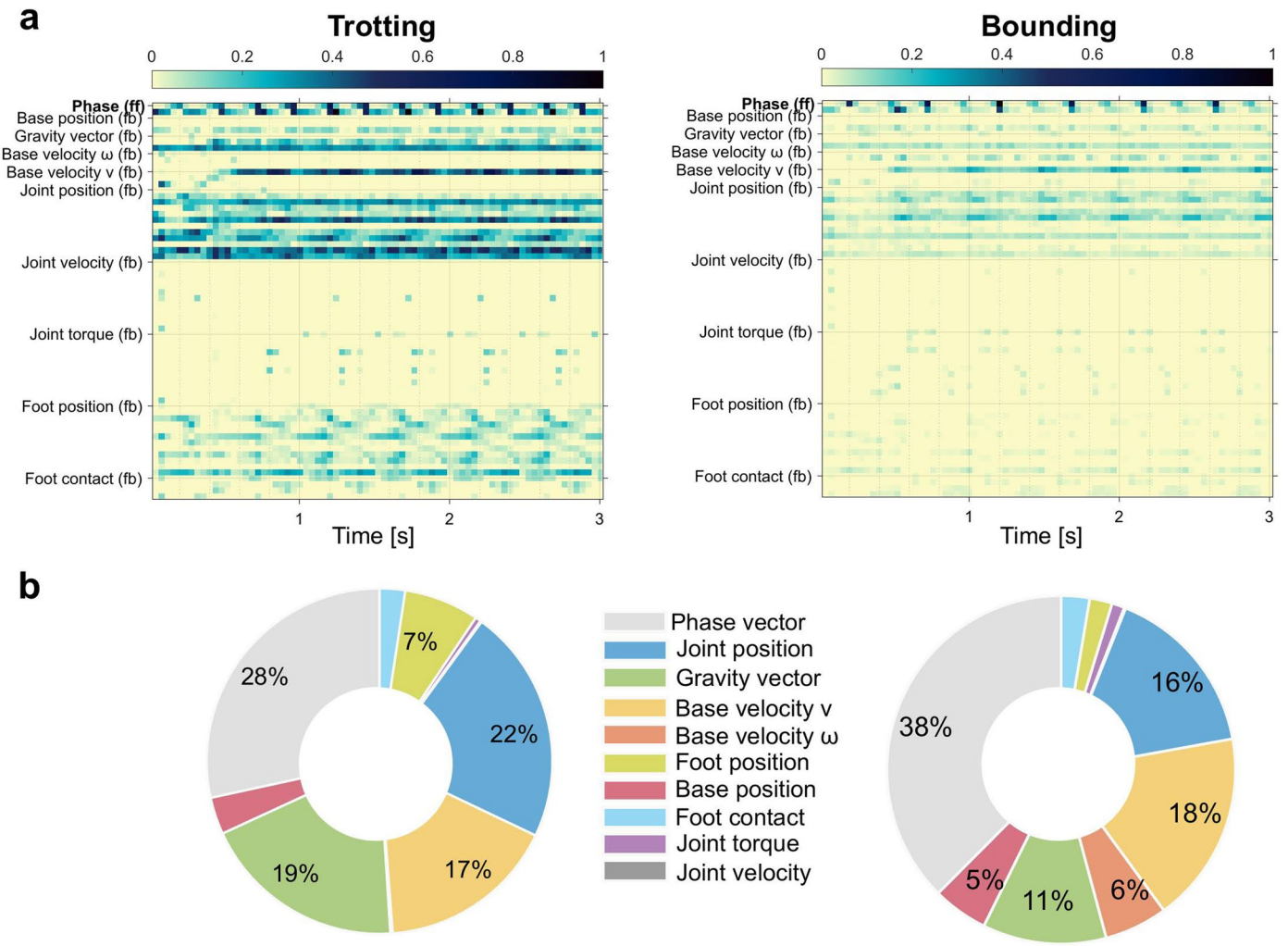
Extended Data Fig. 2 | Key feedback states for locomotion on uneven terrains. Key feedback states for locomotion on an uneven terrain with a maximum height of 3 cm of the randomly generated irregular surfaces. a, The learned trotting over the rough terrain. b, Boxplot showing the importance ranking of nine feedback states for trotting on the random terrain. Each box shows the median (red horizontal line), 25th and 75th percentiles (lower and upper blue horizontal lines), minimum and maximum (lower and upper grey horizontal lines), and outliers (red plus symbol) of the relative importance of the corresponding state with n = 10 samples (random seeds). c, The learned bounding over the rough terrain. d, Boxplot showing the importance ranking of nine feedback states for bounding on the random terrain. Each box shows the median (red horizontal line), 25th and 75th percentiles (lower and upper blue horizontal lines), minimum and maximum (lower and upper grey horizontal lines), and outliers (red plus symbol) of the relative importance of the corresponding state with n = 10 samples (random seeds).

https://doi.org/10.1038/s42256-023-00701-w



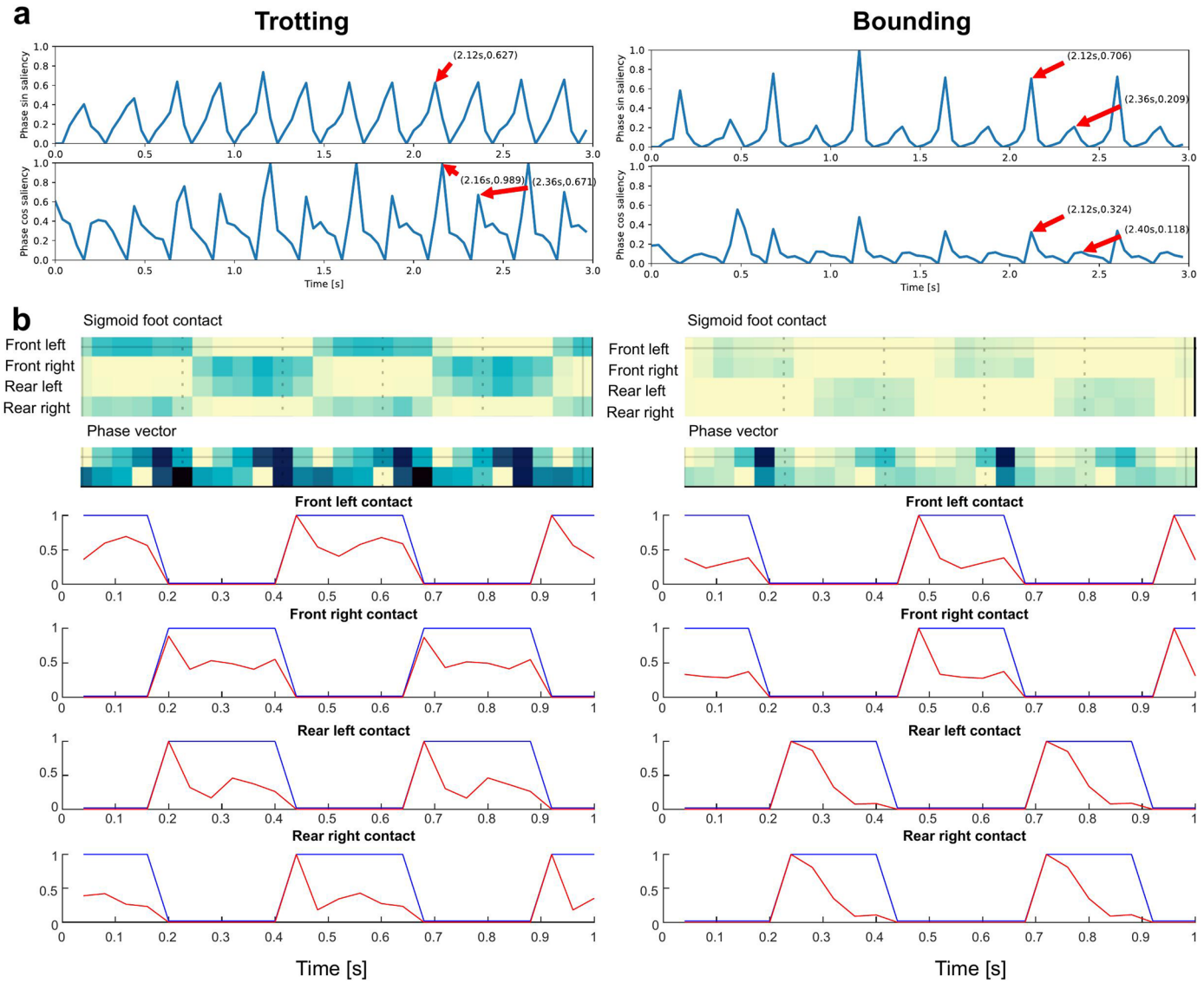
Extended Data Fig. 3 | Analysis of the impact of history state information for the learned balance recovery policy. a, Saliency map showing the importance of feedback states at current time step and two history steps (0.16 s and 0.32 s ago) during 0-3 s. b, Relative importance of all feedback states at the current time step and two history steps. c, Heatmap showing the non-linear correlation between any two dimensions of feedback states including current and two history steps, where a darker colour indicates a stronger correlation. Current body velocity has stronger correlation with current joint positions (label 3) than history joint positions (label 1&2). d, Chord diagram showing the correlation between any two feedback states at current time step (self-correlation and percentage < 25% are removed for clarity), where the wider the link between any two states, the stronger they correlate with each other.

https://doi.org/10.1038/s42256-023-00701-w



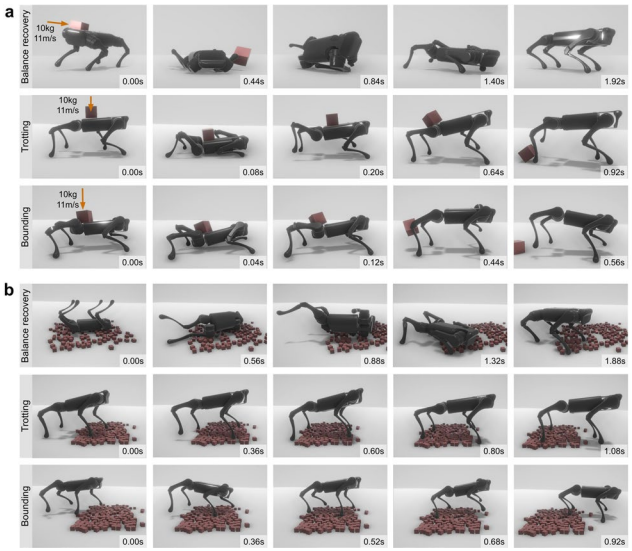
Extended Data Fig. 4 | Analysis of the relative importance of the feedforward phase vector for trotting and bounding. Analysis of the relative importance of the feedforward phase vector ($\sin 2\pi\phi$, $\cos 2\pi\phi$) for trotting and bounding. a, Saliency maps showing the variation of the importance of the feedforward (ff) phase vector and nine feedback states (fb). b, Doughnut charts showing the relative importance of the feedforward (ff) phase vector (28%, 38%) and the feedback states (fb) (72%, 62%) for trotting and bounding, respectively.

https://doi.org/10.1038/s42256-023-00701-w



Extended Data Fig. 5 | Analysis of the relative importance of the feedforward phase vector between swing and stance for trotting and bounding. Analysis of the relative importance of the feedforward phase vector ($\sin 2\pi\phi$, $\cos 2\pi\phi$) between swing and stance for trotting and bounding. a, Time plots of saliency values of the phase vector during 0-3 s. b, Saliency maps showing the variation of the importance of the phase vector and foot contact between swing and stance and time plots of sigmoid (blue) and normalized (red) foot contact feedback during 2-3 s (two gait periods).

https://doi.org/10.1038/s42256-023-00701-w



Extended Data Fig. 6 | Robustness tests of key-state policies for balance recovery, trotting and bounding against unexpected perturbations. a, Perturbation by a 10 kg flying box at 11 m s^{-1} initial velocity for balance recovery (top), trotting (middle) and bounding (bottom). b, Stable traversal over unseen rubble for balance recovery (top), trotting (middle) and bounding (bottom).

nature portfolio

Corresponding author(s): Zhibin Li

Last updated by author(s): Jun 8, 2023

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☒	☐	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☐	☒	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☐	☒	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	The code for training locomotion policies is available at https://github.com/yuwanming/A1_quadruped_env Software is listed in the file environment.yml in the above public repository.
Data analysis	The code for importance analysis is available at https://github.com/yuwanming/A1_quadruped_env Software is listed in the file environment.yml in the above public repository.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

Source data for main text figures are available at https://github.com/yuwanming/feedback_importance_data. Any additional data in this study are available upon request by contacting corresponding author alex.li@ucl.ac.uk.

Human research participants

Policy information about studies involving human research participants and Sex and Gender in Research.

Reporting on sex and gender	not involved in this research
Population characteristics	N/A
Recruitment	N/A
Ethics oversight	N/A

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☒ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	Quantitative method
Research sample	Neural network policies of a simulated quadruped robot A1 in PyBullet simulation engine as representative research samples to study state impacts on action for common quadruped robots. The robot is a common and representative quadruped with twelve degrees of freedom, and the policies reflecting state-action mapping of robot skills were obtained from various training sessions with different random seeds
Sampling strategy	Random sampling
Data collection	Collected and recorded by python code in the provided code repository. The researcher collected data alone and was blind about the condition and study hypothesis
Timing	Data were collected from computer simulation and not affected by the start and end dates
Data exclusions	No data were excluded from the analysis
Non-participation	None
Randomization	Robot data were not allocated into experimental groups

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging