Latent Representations Encoding Behavior and Modulating Adaptivity in Locomotion Learning

Janosch Bajorath · Malte Schilling · Autonomous Intelligent Systems, University of Münster

Universität Münster

Autonomous Intelligent Systems-Group

Problem

Learned quadruped locomotion is structured from the top down, through gait rewards, reference motions and pattern generators, or from the bottom up, through an auxiliary objective that shapes the latent representation. In either case the policy compresses the proprioceptive history into a latent z_t, from which the locomotion controller infers the robot's state and its deviation under domain changes. Adaptive locomotion results from these two readings. What z_t organizes is not specified by either design. Whether behavior and domain occupy separable parts of z_t, how the locomotion controller reads these parts, and whether a curriculum over the command and domain distributions alters this organization remain open.

Contributions

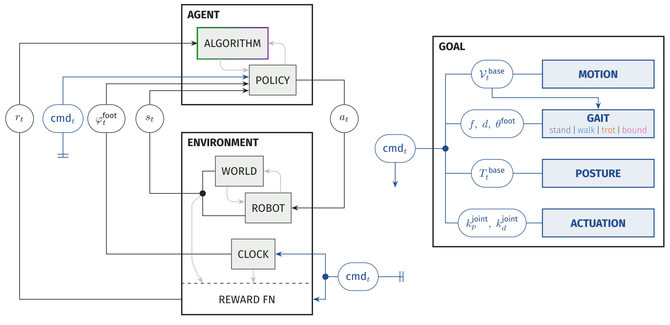
Three contributions address these points, for a policy trained with PPO alone, with the AUX gradients added, and with a curriculum conditioned on command and domain.

1 Latent Organization. A geometric characterization of z_t in its encoding of behavior and domain.

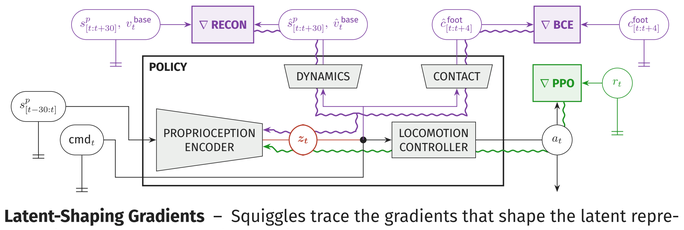
2 Partition under Domain Perturbation. A component-wise partition of z_t into behavior and domain adaptation.

3 Adaptivity under Masking. A masking of z_t that tests the partition in closed loop.

RL Training Setup

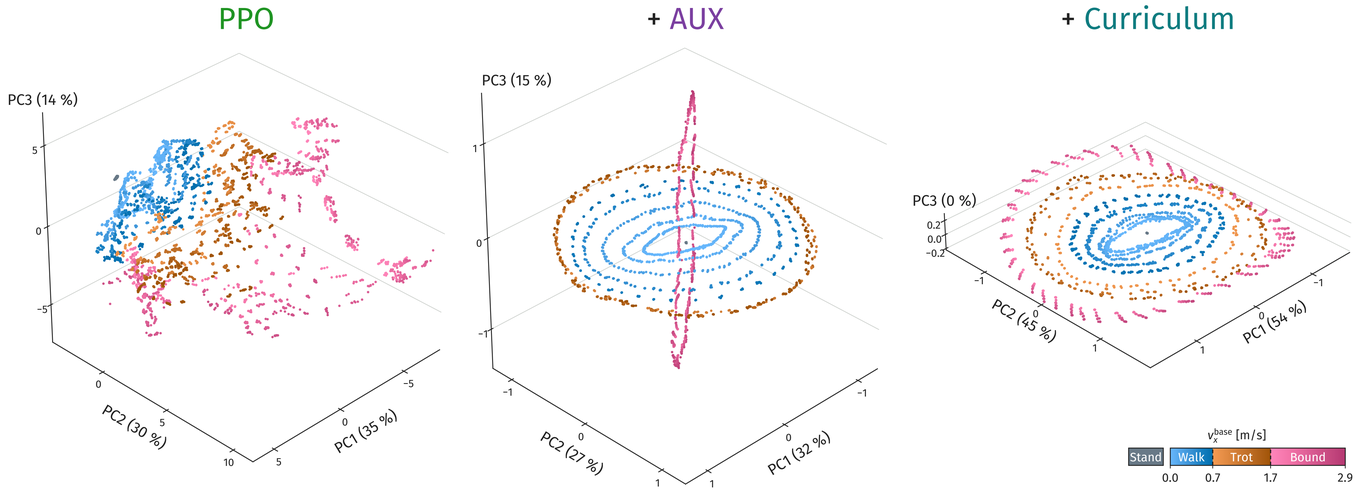


Goal-Conditioned RL – The agent-environment loop is extended by the cmd_t and the foot phases φ_t^{foot} that enter the policy and the reward function. MOTION sets the base twist $\mathcal{V}^{\text{base}} = (v_x, v_y, \omega_z)$, GAIT the stepping frequency, duty and footfall pattern, which follow from the commanded speed, POSTURE the base height and attitude, and ACTUATION the PD gains of the joints. The CLOCK maps the commanded gait to the prescribed foot phases φ_t^{foot}. The reward functions track the twist and pose and score the contact schedule given by φ_t^{foot}.

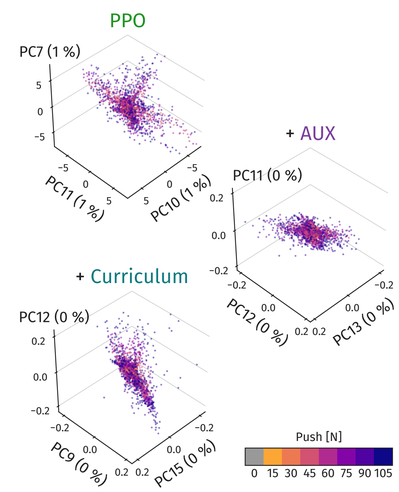


Latent-Shaping Gradients – Squiggles trace the gradients that shape the latent representation. PPO reaches the locomotion controller and the encoder, while the auxiliary gradients AUX (RECON and BCE) reach only the encoder, through the DYNAMICS and CONTACT decoders. The proprioception encoder compresses the noisy history $s^p_{[t-30:t]}$ into $z_t \in \mathbb{R}^{16}$, and the locomotion controller maps z_t and cmd_t to the joint targets a_t.

1 Latent Organization

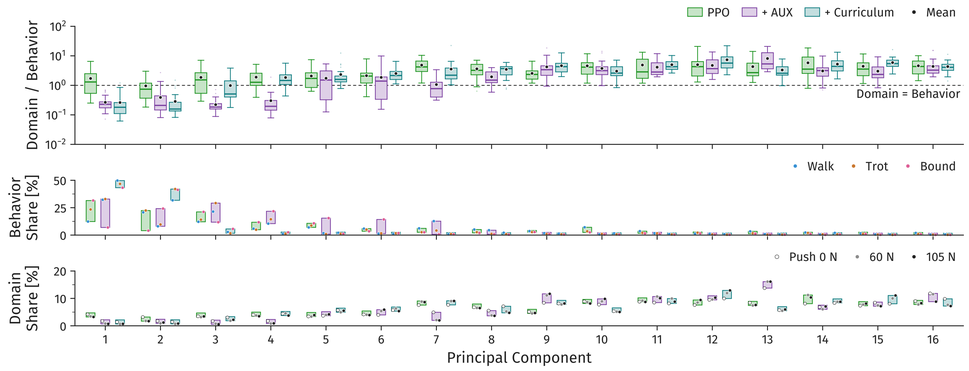


Behavior Encoding – PCA of z_t per training condition (% variance), each panel on its own scale, color the commanded speed and gait. Under PPO the gait cycles form displaced, irregular point clouds without a common geometry. The AUX gradients organize them into thin concentric rings centered on the stand command, with speed as radius and bound set apart from the walk and trot plane. Under the curriculum all gaits lie on concentric rings in a single plane.

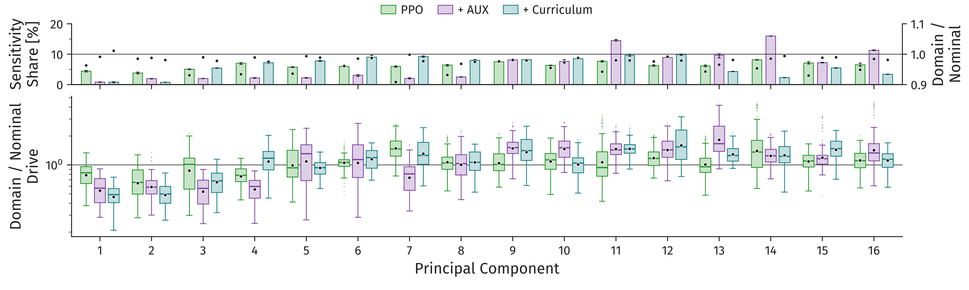


Domain Encoding – Latent z_t under pushes of 0 to 105 N, colored by push level, on the three components the push raises most (% variance).

2 Partition under Domain Perturbation

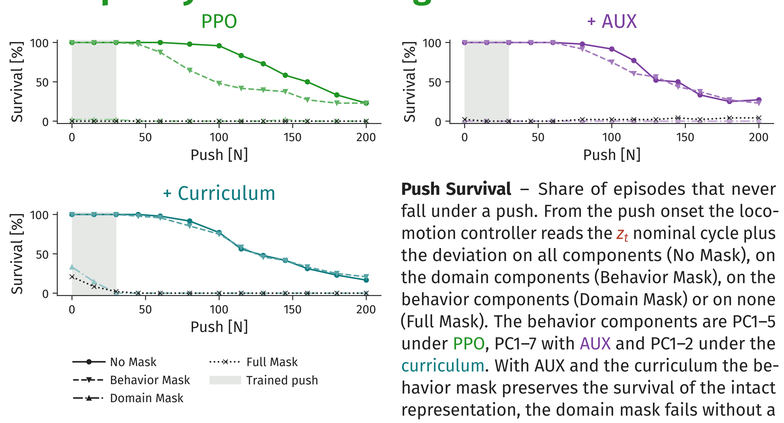


Latent Partition – *Domain / Behavior* is the displacement of z_t from the nominal cycle (the command's latent progression without a push) under a 105 N push over its nominal amplitude, over all gaits. *Behavior Share* is its part of the nominal amplitude, per gait, and *Domain Share* its part of the push displacement, per push level. With AUX and the curriculum the leading components carry the behavior and the trailing ones the push, under PPO components contain both.



Drive under Perturbation – *Sensitivity Share* is the nominal gain of the joint targets on each component, the dots the gain at 105 N over the nominal one. The drive of a component is the change of the joint targets when the component is held at its nominal cycle. *Domain / Nominal Drive* is its share of the drive under the domain perturbation over its share under nominal locomotion. With AUX and the curriculum the drive moves to the domain components, under PPO it stays spread.

3 Adaptivity under Masking



Push Survival – Share of episodes that never fall under a push. From the push onset the locomotion controller reads the z_t nominal cycle plus the deviation on all components (No Mask), on the behavior components (Behavior Mask), on the domain components (Domain Mask) or on none (Full Mask). The behavior components are PC1–5 under PPO, PC1–7 with AUX and PC1–2 under the curriculum. With AUX and the curriculum the behavior mask preserves the survival of the intact representation, the domain mask fails without a push.

Conclusion and Outlook

Conclusion. The locomotion controller computes the joint targets from a feedforward reference, the command and the foot phase, and from the state feedback in z_t. Behavior components of z_t carry the commanded locomotion as closed cycles ordered by gait and speed and are read at low gain, while domain components respond to a push with an excursion that encodes the deviation and are read at high gain. Under masking the behavior components can be replaced by their nominal cycle, the domain components cannot. The AUX gradients produce this partition, PPO alone does not, and the curriculum concentrates the behavior on two components.

Outlook. The proprioception encoder is split into one for behavior and one for domain. The clock is then replaced by a phase the policy derives itself, from a periodic encoder, a pattern generator or interoception. Finally the gait rewards are removed, so that the gait follows from optimization, steered by a curriculum the latent representation drives.