Safe Locomotion Skills Learning Across Robot Scales

Xinlei Pan, Ronald Fearing, Stella Yu {xinleipan, ronf, stellayu}@berkeley.edu
Wenhao Yu, Jie Tan (magicmelon, jietan)@google.com

1. Summary

Robot safety is a key bottleneck for learning legged locomotion. During initial exploration, robots could burn out motors, fall, or damage hardware. To ensure safety, many learning methods limit explorations resulting in suboptimal performance. We draw inspiration from animal learning: baby animals explore freely due to their small sizes and light weights; this learning carries through subsequent size and morphological changes with growth. We hypothesize that small robots can learn a policy with reduced safety concerns, and the policy can be adapted towards larger robots. In our study, we showed successful learning of a scale-normalized state/action representation, combined with a scale-invariant universal policy, to control a simple dynamic system cart-pole. We extend our effort towards a more complicated locomotion robot, and enable safe policy transfer to real robots by combining policy transfer across scales and sim-to-real policy transfer.

2. Related Works

[1] transfers policies among multiple robots with known physical parameters. [2] proposes a policy domain transfer with different observation spaces but not different scales of robots. [3] transfers skills between robots of different morphologies but requires each robot to learn skills. [4] adapts locomotion policies to new environments through latent space optimization. Our previous UCB-Google collaboration [5] transfers human demonstrations to real robots in navigation settings. Our previous work [6] learns to co-optimize morphology and policy for grasping purposes and learns a representation that can transfer across robots of different morphologies. This work aims to transfer locomotion skills from small robots to large robots in the real world while leveraging simulated robots for learning transfer rules across scales.

3. Research Approach

Our ultimate goal is to obtain policies for real large legged robots, where direct training on these robots can be dangerous. Direct policy transfer from small robots to a large robot in the real world or sim-to-real transfer on large robots may not work due to limited amounts of real large robot data. However, we have access to legged robots of small scales and have simulation platforms that can learn policies across scales. We combine *policy transfer across scales* with *sim-to-real policy transfer* to learn policies that can work on real large robots with a small amount of real robot data.

Policy transfer across scales. We achieved policy transfer across scales within simulation on the simple cart-pole problem by using an encoder/decoder representation learning framework (Figure 1 left), where the encoder maps robot states of different scales to a normalized state. A universal policy then uses the normalized state to output a normalized action, and the decoder maps the normalized action back to the raw action for a particular scale. The regularization loss is added to encode scale invariant information in the encoded latent space. This encoder/decoder framework learns a scale-invariant latent space that makes it much easier to transfer policies across scales. Our results on a simulated canonical cartpole system demonstrate significantly better results than traditional approaches such as domain randomization and are on par with the performance of typical control based approaches such as linear quadratic regulator. As a proof-of-concept, the encoder/decoder approach was successful for normalized states for cart-pole. The non-linear dynamics (and environment interactions) for legged robots are unlikely to scale in a straight-forward fashion as in cart-pole. This makes it nontrivial to extend our success on cartpole towards legged robots. We explore novel network designs that enable it to approximate specific physical rules of legged robots.

Sim-to-real policy transfer. Our **next step** includes sim-to-real policy transfer for locomotion robots. We use a latent variable model approach for achieving the sim-to-real adaptation part. We add an additional latent variable “env info” representing the factors that change across environments such as surface frictions or joint properties. The latent variable model will be learned across small robots in simulation and small robots in the real world. The ultimate transfer happens by using small real robots (abundant data) to estimate the environment specific information, and reusing the scale invariant

policy with scale information estimated from a large simulated robot, enabling few shot generalization on large real robots (Figure 1 right).

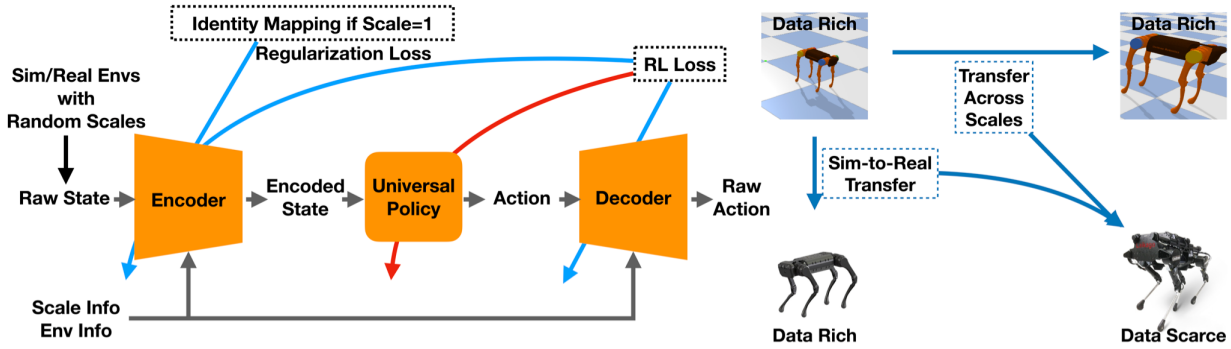

Figure 1. Left: system diagram. Right: Policy transfer pipeline design.

4. Current Results on Policy Transfer Across Scales on Laikago

We evaluate our proposed approach and several baselines, including (1) Domain randomization (DR): A PPO policy that is trained across different scales on the Laikago robot for scales uniformly sampled from 0.5-1.5, and then tested on scale 0.5, 1.0, 1.5, 2.0; (2) DR with a feature encoder: domain randomization but with the feature encoder as in our method, trained on Laikago across scales from 0.5-1.5, and tested on scale 0.5, 1.0, 1.5, 2.0; (3) Our method: PPO with randomized scales between 0.5-1.5, and with a feature encoder and the regularization loss. (4) PPO policy on individual scales (Oracle). We show the testing time reward characterizing moving distance (normalized by body length) in the following table.

Method\Scale	0.5	1.0	1.5	2.0
DR	1.70	2.64	2.51	2.19
DR With Feature	1.64	2.60	1.74	1.84
Ours	1.82	6.59	3.39	2.92
PPO (Oracle)	8.3	7.9	5.4	4.5

These results show the unique challenge that policies that are trained on different scales cannot be easily transferred with the domain randomization method. Especially for out-of-distribution scales (scale 2.0), the DR method fails significantly compared with the PPO method. Our method gains a slight edge compared with the baselines but there are rooms for improvement.

References

[1] C.Tao, et al. Hardware conditioned policies for multi-robot transfer learning. NIPS 2018.

[2] Z.Qiang et al. Learning Cross-Domain Correspondence for Control with Dynamics Cycle-Consistency. ICLR 2021

[3] G. Abhishek, et al. Learning invariant feature spaces to transfer skills with reinforcement learning. ICLR 2017

[4] Yu et al. Learning fast adaptation with meta strategy optimization. *RA-L 2019.*

[5] Pan et al. Zero-shot Imitation Learning from Demonstrations for Legged Robot Visual Navigation. ICRA 2020.

[6] Pan et al. Emergent Hand Morphology and Control from Optimizing Robust Grasps of Diverse Objects. ICRA 2021

[7] Duan et al. Learning Task Space Actions for Bipedal Locomotion. *arXiv preprint arXiv:2011.04741* (2020).

[8] Nagabandi et al. Neural network dynamics models for control of under-actuated legged millirobots. IROS 2018.