



Sim-to-Real Transfer for Legged Locomotion: Progress and Persistent Gaps

Pawel Wolde^{1*}, Magdalena Girard², Camila Tadesse³

¹Department of Mechanical Engineering, Lapland University Centre for Mechanical Engineering, Ethiopia

²Department of Mechanical Engineering, Pannonian Institute for Advanced Mechanical Engineering, Canada

³Department of Mechanical Engineering, Carpathian Graduate Institute of Mechanical Engineering, Canada

Article Info

Volume 1, Issue 1, 2026

doi: [10.62715/ijrra.2026.v1.i1.pp7-13](https://doi.org/10.62715/ijrra.2026.v1.i1.pp7-13)

Received: 29 September 2025

Accepted: 27 November 2025

Published: 27 January 2026

Abstract

Controllers for legged robots are now routinely trained by reinforcement learning in simulation and run on hardware without further tuning. A decade ago this was regarded as impractical, because policies exploited the inaccuracies of the simulator and failed on contact with the physical machine. This narrative review traces how the change came about and where transfer remains unreliable. It first sets out the sources of the mismatch between simulated and real legged systems: actuator dynamics, contact and friction, sensing delay and noise, and synthetic perception. It then examines the main families of remedy, namely identified and learned actuator models, randomisation of dynamics, online adaptation from proprioceptive history, privileged teacher and student training, and massively parallel simulation. Published demonstrations on quadrupeds, bipeds and humanoids are discussed as evidence of what these methods have delivered. The review then turns to the gaps that persist: deformable and moving terrain, the realism of simulated exteroception, the lack of shared evaluation protocols, the dependence on hand-shaped rewards, and the treatment of hardware wear and safe failure. Learning directly on the robot is considered as a complement. The article closes with directions for work that would make transfer a measurable property instead of a reported success.

Keywords: *Sim-to-real transfer, Legged locomotion, Reinforcement learning, Domain randomisation, Reality gap, Quadruped robots*

© 2026 Pawel Wolde et al. This is an open access article under the CC BY license (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

1. Introduction

A legged robot is an awkward subject for trial-and-error learning. It falls, and falls damage it. Its dynamics are hybrid, switching at every footfall, and the forces that matter most pass through contacts that are difficult to measure. Reinforcement learning needs a volume of experience that no physical machine can supply in reasonable time. Simulation supplies it cheaply, and the price is the mismatch between the simulated and the physical robot. The term for that mismatch, the reality gap, comes from evolutionary robotics, where Jakobi et al. (1995) observed that controllers evolved in simulation transferred only when the simulation contained appropriate noise.

* Corresponding author: Pawel Wolde, Department of Mechanical Engineering, Lapland University Centre for Mechanical Engineering, Ethiopia. E-mail: pawel.wolde@lapland.edu

© 2026 Pawel Wolde et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

For many years the strongest legged controllers were model based. Whole-body and model predictive control gave quadrupeds such as the MIT Mini Cheetah fast and dynamic gaits (Kim et al., 2019), and learned policies were a laboratory curiosity by comparison. That changed quickly. Within a few years policies trained entirely in simulation were shown trotting and galloping on a small quadruped (Tan et al., 2018), recovering a larger one from falls (Hwangbo et al., 2019) and carrying it over rough natural ground (Lee et al., 2020).

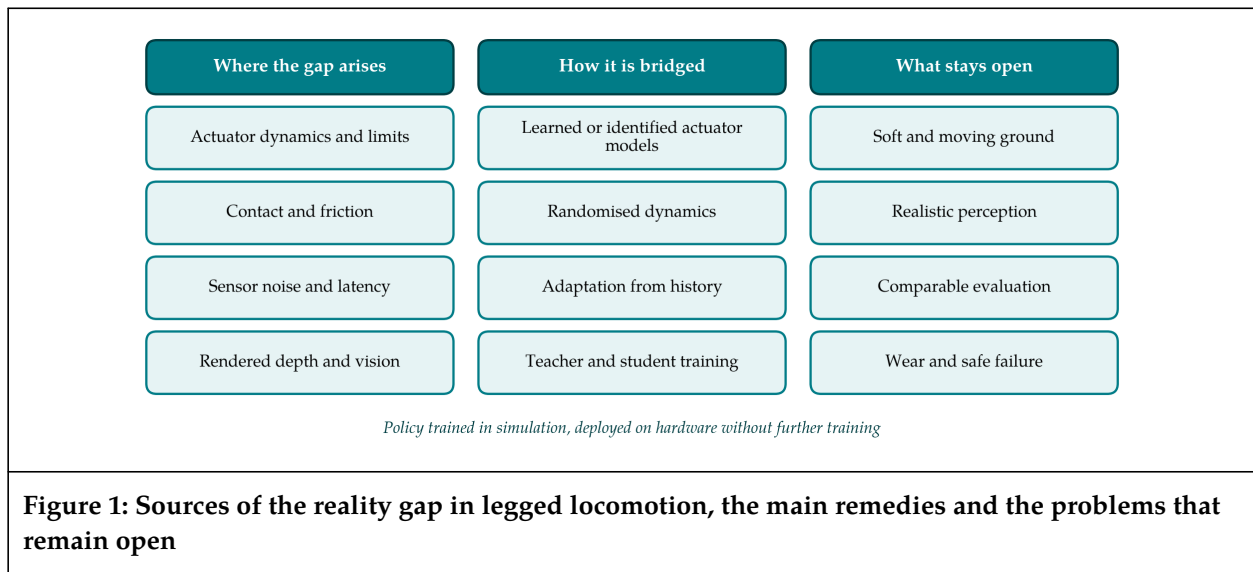
This review asks two questions of that literature. Which ideas account for the progress, and which parts of the gap remain open once the demonstrations are set aside? The second question receives the more attention, because the successes are well publicised and the failures are seldom written up.

2. Scope and Approach

The article is a narrative review of the published literature on transferring learned locomotion controllers from simulation to legged hardware. It concentrates on reinforcement learning for quadrupeds and bipeds, and it draws on peer-reviewed articles, conference papers and widely cited preprints known to the field. It does not attempt an exhaustive or systematic search, and it does not rank methods numerically, since the reported experiments differ in robot, terrain and measure. General surveys of sim-to-real methods in robotics are available elsewhere (Zhao et al., 2020; Muratore et al., 2022); the present review is narrower and limited to legs.

3. Where the Reality Gap Comes From

Figure 1 groups the sources of mismatch, the remedies and the open problems that organise the rest of this review. Four sources recur in the literature.



3.1. Actuators

Rigid-body simulators treat a joint as an ideal torque source. Real actuators are nothing of the kind. Gear trains add friction and backlash, motor drivers saturate, series elastic elements store energy, and the low-level control loop introduces delay. Hwangbo et al. (2019) identified the actuator as the dominant error for the ANYmal quadruped, whose joints are series elastic (Hutter et al., 2016), and Tan et al. (2018) reached the same conclusion for the direct-drive Minitaur, where an accurate motor model and a model of latency were both needed before policies transferred.

3.2. Contact

Simulators resolve contact with simplified models: rigid or lightly compliant bodies, Coulomb friction with a single coefficient, and solvers that soften constraints for numerical stability, as in MuJoCo

(Todorov et al., 2012). Real feet slip, sink, roll on their edges and strike surfaces whose properties change from step to step. A policy can learn to rely on a contact artefact, such as a small penetration that the solver permits, and that reliance is invisible until the hardware trial.

3.3. Sensing and perception

Joint encoders are accurate, but base velocity is estimated, inertial measurements drift and are noisy, and every signal arrives late. Exteroception is worse. Depth cameras and lidar return missing data on reflective or dark surfaces, treat tall grass as an obstacle and lose their pose estimate when the robot slips, none of which appears in a cleanly rendered height map. Miki et al. (2022) made this unreliability the centre of their design, training a policy to decide for itself when the perceived terrain should be distrusted.

4. Methods for Crossing the Gap

The remedies fall into five families, which are usually combined. Table 1 summarises them with representative sources.

Table 1: Main families of sim-to-real method for legged locomotion, with representative sources and limits			
Family	Idea	Representative sources	Limit
Actuator modelling	Identify or learn the map from command to joint torque	Tan et al., 2018; Hwangbo et al., 2019	Specific to one robot and its wear state
Dynamics randomisation	Vary mass, friction, delays and gains during training	Peng et al., 2018; Tobin et al., 2017	Ranges are hand-set; wide ranges cost performance
Randomisation tuned from data	Fit the simulation parameter distribution to real rollouts	Chebatar et al., 2019	Needs repeated hardware trials
Online adaptation	Infer hidden properties from recent proprioceptive history	Yu et al., 2017; Kumar et al., 2021	Adapts only within the trained range
Privileged teacher and student	Teacher sees true state; student imitates it from sensors	Lee et al., 2020; Miki et al., 2022	Two-stage training; student may lag the teacher
Massively parallel simulation	Thousands of robots simulated at once on a graphics processor	Makoviychuk et al., 2021; Rudin et al., 2022	Speed gained by simplifying contact

4.1. Better models of the robot

The most direct remedy is to improve the simulator where it is worst. Tan et al. (2018) did so analytically, with a motor model identified on the bench. Hwangbo et al. (2019) took a different course and trained a neural network on logged data to predict joint torque from a short history of position errors and velocities, then placed that network inside the rigid-body simulation. The approach captures effects that are tedious to model by hand, and it ties the trained policy to a particular actuator in a particular condition.

4.2. Randomisation

If the true parameters cannot be known, a policy can be trained across a distribution of them. Tobin et al. (2017) introduced domain randomisation for visual inputs, and Peng et al. (2018) applied the same idea to dynamics, randomising masses, friction and delays so that the real system would appear as one more sample. The idea has an antecedent in the noise-based argument of Jakobi et al. (1995). Randomisation is now universal in legged work. Its weakness is that the ranges are chosen by the engineer: too narrow and the policy overfits, too wide and it becomes conservative. Chebatar et al. (2019) addressed this by

adjusting the randomisation distribution from real trajectories.

4.3. Adaptation and privileged learning

A policy that is merely robust treats every surface alike. An adaptive one infers what it is standing on. Yu et al. (2017) paired a universal policy with an online estimator of physical parameters. Lee et al. (2020) trained a teacher with access to privileged simulator state, such as contact forces and terrain profile, and then a student that reproduced the teacher's actions from a history of joint measurements alone. Rapid motor adaptation, proposed by Kumar et al. (2021), follows similar reasoning with a module that estimates a compact description of the environment from recent states and actions. Peng et al. (2020) adapted policies that imitate recorded animal motion by searching a learned latent space of dynamics on the real robot.

4.4. Faster simulation

The last ingredient is throughput. Simulators that run thousands of environments in parallel on a graphics processor (Makoviychuk et al., 2021) reduced the training of a rough-terrain walking policy from days to minutes (Rudin et al., 2022), using a standard on-policy learning algorithm. Speed changes practice: with short iterations, engineers tune rewards and randomisation ranges against hardware behaviour, so that the designer becomes part of the transfer loop.

5. What Has Been Demonstrated

The evidence for these methods consists of hardware demonstrations. On quadrupeds the sequence runs from flat-ground gaits (Tan et al., 2018) through fall recovery and fast running (Hwangbo et al., 2019) to blind traversal of mud, snow and rubble (Lee et al., 2020), perceptive hiking on an alpine trail (Miki et al., 2022) and parkour-style climbing and jumping (Hoeller et al., 2024). Bipedes, with a smaller support polygon and less margin, followed. Policies for the Cassie biped were transferred after iterative refinement of the training design (Xie et al., 2020) and later climbed stairs without vision (Siekmann et al., 2021). Radosavovic et al. (2024) reported a full-size humanoid walking outdoors under a policy trained only in simulation.

These results are persuasive, and they should be read for what they are. Each is a demonstration on one robot, selected by its authors, in conditions they chose. Success rates are reported by some groups and not by others, and negative results are rare. The literature shows that transfer can work. It says much less about how often it does, or by what margin.

6. Persistent Gaps

Terrain that deforms or moves. Training terrains are rigid height fields. Sand, deep snow, vegetation, loose scree and floating or swaying platforms violate that assumption. Blind policies cope with some of these through adaptation, but they do so by extrapolation, and no widely used simulator models granular or yielding ground at the speed that learning requires.

Perception. Policies that use cameras or lidar are trained on rendered depth or sampled height maps with synthetic noise. The failure modes of real sensors are structured, not random, and correlated with the very situations in which the robot most needs to see. The belief-state design of Miki et al. (2022) mitigates the problem by falling back on proprioception; it does not make the simulated sensor realistic.

Evaluation. There is no accepted test course, no standard perturbation set and no common reporting format for sim-to-real locomotion. Papers report different robots on different ground with different measures. Whether one method transfers better than another is in most cases impossible to establish from the published record, a weakness that surveys of the wider field also note (Muratore et al., 2022).

Reward and curriculum engineering. Policies that transfer well are trained with long lists of shaped reward terms that penalise foot slip, joint acceleration, torque and action rate. These terms are partly a disguised remedy for the gap, since they discourage behaviour the hardware cannot reproduce, and they are tuned by hand against the real robot. The knowledge involved is rarely documented in full.

Hardware condition and safe failure. A learned actuator model describes a robot on the day the data were logged. Gearboxes wear, batteries sag and feet are replaced. Little published work examines how transferred policies degrade as hardware ages, or how a policy should behave when the situation lies outside anything it met in training.

Learning on the robot itself is one response. Haarnoja et al. (2019) trained a small quadruped to walk in the real world in about two hours, Smith et al. (2022) reduced this to around twenty minutes with a carefully tuned off-policy method, and Wu et al. (2023) used a learned world model for the same purpose. Such methods avoid the gap by never entering simulation, but so far for simple gaits on benign ground. The likelier role for real-world learning is to refine a policy pre-trained in simulation.

7. Directions for Future Work

Three lines of work would address the gaps identified above. The first is measurement. A shared set of physical test conditions, with agreed perturbations and reporting of failures as well as successes, would let the field compare methods and would expose which sources of mismatch still matter. The second is simulation of what is currently omitted: compliant and granular ground, sensor models built from recorded failure cases, and actuator models that include temperature and wear. The third is a principled account of the residual gap at run time, so that a policy can recognise when its training no longer applies and hand over to a conservative fallback.

A further question concerns the division of labour between learning and models. Model-based controllers come with constraints that can be checked, and learned policies with robustness that cannot yet be certified. Hybrid schemes, in which a learned policy proposes and a model-based layer verifies, are an obvious meeting point and are likely to matter most where legged robots work near people.

8. Conclusion

Sim-to-real transfer for legged locomotion has moved from doubtful to routine through a small number of ideas applied together: honest actuator models, randomised dynamics, adaptation from proprioceptive history, privileged training and very fast simulation. The demonstrations that followed are real achievements. They are also selected ones, and the published record does not yet support quantitative claims about reliability. Deformable terrain, realistic perception, comparable evaluation, undocumented reward engineering and ageing hardware are where the gap persists. Progress on those fronts depends less on new learning algorithms than on better measurement and on simulators that include what has so far been left out.

Conflict of interest

To be declared by the author(s) before publication.

Funding

To be declared by the author(s) before publication.

References

- Chebatar, Y., Handa, A., Makoviyshuk, V., Macklin, M., Issac, J., Ratliff, N. and Fox, D. (2019). [Closing the sim-to-real loop: Adapting simulation randomization with real world experience. *Proceedings of the 2019 IEEE International Conference on Robotics and Automation \(ICRA\)*](#), 8973-8979.
- Haarnoja, T., Ha, S., Zhou, A., Tan, J., Tucker, G. and Levine, S. (2019). [Learning to walk via deep reinforcement learning. *Robotics: Science and Systems XV*](#).
- Hoeller, D., Rudin, N., Sako, D. and Hutter, M. (2024). [ANYmal parkour: Learning agile navigation for quadrupedal robots. *Science Robotics*](#), 9(88): eadi7566. <https://doi.org/10.1126/scirobotics.adi7566>
- Hutter, M., Gehring, C., Jud, D., Lauber, A., Bellicoso, C.D., Tsounis, V., Hwangbo, J., Bodie, K., Fankhauser, P., Bloesch, M., Diethelm, R., Bachmann, S., Melzer, A. and Hoepflinger, M. (2016).

- ANYmal - a highly mobile and dynamic quadrupedal robot. *Proceedings of the 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 38-44.
<https://doi.org/10.1109/IROS.2016.7758092>
- Hwangbo, J., Lee, J., Dosovitskiy, A., Bellicoso, D., Tsounis, V., Koltun, V. and Hutter, M. (2019). *Learning agile and dynamic motor skills for legged robots. Science Robotics*, 4(26): eaau5872.
<https://doi.org/10.1126/scirobotics.aau5872>
- Jakobi, N., Husbands, P. and Harvey, I. (1995). *Noise and the reality gap: The use of simulation in evolutionary robotics. Advances in Artificial Life (ECAL 1995), Lecture Notes in Computer Science*, 929: 704-720. Springer, Berlin. https://doi.org/10.1007/3-540-59496-5_337
- Kim, D., Di Carlo, J., Katz, B., Bledt, G. and Kim, S. (2019). *Highly dynamic quadruped locomotion via whole-body impulse control and model predictive control.* arXiv preprint arXiv:1909.06586.
- Kumar, A., Fu, Z., Pathak, D. and Malik, J. (2021). *RMA: Rapid motor adaptation for legged robots. Robotics: Science and Systems XVII.* <https://doi.org/10.15607/RSS.2021.XVII.011>
- Lee, J., Hwangbo, J., Wellhausen, L., Koltun, V. and Hutter, M. (2020). *Learning quadrupedal locomotion over challenging terrain. Science Robotics*, 5(47): eabc5986. <https://doi.org/10.1126/scirobotics.abc5986>
- Makoviychuk, V., Wawrzyniak, L., Guo, Y., Lu, M., Storey, K., Macklin, M., Hoeller, D., Rudin, N., Allshire, A., Handa, A. and State, G. (2021). *Isaac Gym: High performance GPU-based physics simulation for robot learning.* arXiv preprint arXiv:2108.10470.
- Miki, T., Lee, J., Hwangbo, J., Wellhausen, L., Koltun, V. and Hutter, M. (2022). *Learning robust perceptive locomotion for quadrupedal robots in the wild. Science Robotics*, 7(62): eabk2822. <https://doi.org/10.1126/scirobotics.abk2822>
- Muratore, F., Ramos, F., Turk, G., Yu, W., Gienger, M. and Peters, J. (2022). *Robot learning from randomized simulations: A review. Frontiers in Robotics and AI*, 9: 799893.
<https://doi.org/10.3389/frobt.2022.799893>
- Peng, X.B., Andrychowicz, M., Zaremba, W. and Abbeel, P. (2018). *Sim-to-real transfer of robotic control with dynamics randomization. Proceedings of the 2018 IEEE International Conference on Robotics and Automation (ICRA)*, 3803-3810. <https://doi.org/10.1109/ICRA.2018.8460528>
- Peng, X.B., Coumans, E., Zhang, T., Lee, T.-W., Tan, J. and Levine, S. (2020). *Learning agile robotic locomotion skills by imitating animals. Robotics: Science and Systems XVI.* <https://doi.org/10.15607/RSS.2020.XVI.064>
- Radosavovic, I., Xiao, T., Zhang, B., Darrell, T., Malik, J. and Sreenath, K. (2024). *Real-world humanoid locomotion with reinforcement learning. Science Robotics*, 9(89): eadi9579.
<https://doi.org/10.1126/scirobotics.adi9579>
- Rudin, N., Hoeller, D., Reist, P. and Hutter, M. (2022). *Learning to walk in minutes using massively parallel deep reinforcement learning. Proceedings of the 5th Conference on Robot Learning, Proceedings of Machine Learning Research*, 164: 91-100.
- Siekmann, J., Green, K., Warila, J., Fern, A. and Hurst, J. (2021). *Blind bipedal stair traversal via sim-to-real reinforcement learning. Robotics: Science and Systems XVII.*
- Smith, L., Kostrikov, I. and Levine, S. (2022). *A walk in the park: Learning to walk in 20 minutes with model-free reinforcement learning.* arXiv preprint arXiv:2208.07860.
- Tan, J., Zhang, T., Coumans, E., Iscen, A., Bai, Y., Hafner, D., Bohez, S. and Vanhoucke, V. (2018). *Sim-to-real: Learning agile locomotion for quadruped robots. Robotics: Science and Systems XIV.* <https://doi.org/10.15607/RSS.2018.XIV.010>
- Tobin, J., Fong, R., Ray, A., Schneider, J., Zaremba, W. and Abbeel, P. (2017). *Domain randomization for transferring deep neural networks from simulation to the real world. Proceedings of the 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 23-30.
<https://doi.org/10.1109/IROS.2017.8202133>

- Todorov, E., Erez, T. and Tassa, Y. (2012). [MuJoCo: A physics engine for model-based control](#). *Proceedings of the 2012 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 5026-5033. <https://doi.org/10.1109/IROS.2012.6386109>
- Wu, P., Escontrela, A., Hafner, D., Abbeel, P. and Goldberg, K. (2023). [DayDreamer: World models for physical robot learning](#). *Proceedings of the 6th Conference on Robot Learning, Proceedings of Machine Learning Research*, 205.
- Xie, Z., Clary, P., Dao, J., Morais, P., Hurst, J. and van de Panne, M. (2020). [Learning locomotion skills for Cassie: Iterative design and sim-to-real](#). *Proceedings of the Conference on Robot Learning, Proceedings of Machine Learning Research*, 100: 317-329.
- Yu, W., Tan, J., Liu, C.K. and Turk, G. (2017). [Preparing for the unknown: Learning a universal policy with online system identification](#). *Robotics: Science and Systems XIII*.
- Zhao, W., Queralta, J.P. and Westerlund, T. (2020). [Sim-to-real transfer in deep reinforcement learning for robotics: A survey](#). *Proceedings of the 2020 IEEE Symposium Series on Computational Intelligence (SSCI)*, 737-744.