

Scalable Zero-Shot Skill Transition Learning for Legged Locomotion Across Mixed Terrains

Anonymous Author(s)

ABSTRACT

Legged locomotion controllers trained on limited terrain types often fail during skill transitions on unseen terrain combinations, particularly when switching between dynamically distinct gaits such as decelerating on sparse footholds before climbing. We investigate whether a scalable, principled reinforcement learning approach can enable zero-shot generalization for such challenging skill transitions. We propose STG-Net, a Skill Transition Generalization architecture combining attention-based terrain context encoding with a progressive transition curriculum. We evaluate five policy architectures across 80 terrain courses (50 training, 30 test) spanning 8 terrain types and 8 skill primitives, measuring transition success rate (TSR), foothold precision, velocity tracking, and a composite Skill Transition Generalization Index (STGI). Our full STG-Net achieves a test TSR of 0.9821 and test STGI of 0.8987 with a TSR generalization gap of only 0.0062, substantially outperforming the baseline MLP (test TSR 0.5521, gap 0.1674) and even surpassing an oracle finetuned policy (test TSR 0.9032) on zero-shot transfer. Ablation studies confirm that both attention-based context encoding and curriculum scheduling contribute significantly, while scalability analysis shows that performance improves monotonically as training terrain diversity increases from 2 to 8 types. These results provide evidence that scalable, principled zero-shot skill transition learning is achievable through the combination of structured terrain representations, attention mechanisms, and progressive curricula.

CCS CONCEPTS

• **Computing methodologies** → **Artificial intelligence**; *Reasoning about belief and knowledge*.

KEYWORDS

zero-shot generalization, skill transitions, legged locomotion, reinforcement learning, curriculum learning

ACM Reference Format:

Anonymous Author(s). 2026. Scalable Zero-Shot Skill Transition Learning for Legged Locomotion Across Mixed Terrains. In *Proceedings of ACM Conference (Conference'17)*. ACM, New York, NY, USA, 5 pages. <https://doi.org/10.1145/nnnnnnn.nnnnnnn>

1 INTRODUCTION

Legged robots must traverse diverse and unpredictable terrain sequences in real-world deployment, requiring rapid transitions between locomotion skills such as walking, climbing, leaping, and balancing [6, 7]. While recent advances in reinforcement learning (RL) have produced controllers capable of agile locomotion on individual terrain types [2, 3, 12], a fundamental open question remains:

whether there exists a scalable, principled approach for learning challenging skill transitions that generalizes zero-shot to previously unseen terrain combinations [11].

This question is particularly pressing because ad hoc finetuning, while effective for specific failure cases, does not scale to the combinatorial space of possible terrain sequences and skill pairings. Zhang et al. [11] explicitly identify this gap in their AME-2 system, observing increased failure rates during transitions on unseen test terrains despite strong performance on training terrains.

We address this open problem by proposing STG-Net (Skill Transition Generalization Network), a framework that combines three key components: (1) a terrain graph model encoding skill regions and transition zones with explicit difficulty quantification, (2) an attention-based context encoder that captures terrain-proprioceptive features across transition boundaries, and (3) a progressive transition curriculum that schedules training from easy to challenging skill switches. We evaluate our approach against four baselines across 80 terrain courses, 8 terrain types, and 8 skill primitives, using a comprehensive suite of metrics including a novel composite Skill Transition Generalization Index (STGI).

Our key contributions are:

- A principled framework for zero-shot skill transition learning that achieves test TSR of 0.9821 with a generalization gap of only 0.0062, demonstrating near-perfect transfer to unseen terrain combinations.
- A composite evaluation metric (STGI) that jointly captures transition success, foothold precision, and velocity tracking, where STG-Net achieves 0.8987 on test courses.
- Systematic ablation studies showing that attention-based encoding reduces the generalization gap from 0.1674 to 0.1053, curriculum scheduling further reduces it to 0.0699, and their combination in STG-Net achieves 0.0062.
- Scalability analysis demonstrating monotonically improving performance as training terrain diversity increases from 2 to 8 types.

2 RELATED WORK

Legged Locomotion via RL. Deep RL has enabled quadrupedal robots to traverse increasingly challenging terrains. Lee et al. [6] demonstrated locomotion over rough terrain using teacher-student learning. Miki et al. [7] extended this with exteroceptive sensing for wild environments. Kumar et al. [5] introduced rapid motor adaptation for terrain-adaptive control. More recently, parkour-style locomotion has been achieved via vision-based policies [2, 3, 12]. However, these works primarily focus on single-terrain or single-skill performance rather than the transition dynamics between skills.

Attention Mechanisms in Robotics. Attention-based architectures [10] have proven effective for capturing context-dependent features

Conference'17, July 2017, Washington, DC, USA
2026. ACM ISBN 978-x-xxxx-xxxx-x/YY/MM...\$15.00
<https://doi.org/10.1145/nnnnnnn.nnnnnnn>

in robotic control. Zhang et al. [11] employ attention-based neural map encoding for terrain-aware locomotion, achieving strong single-terrain generalization. We extend this paradigm by applying attention specifically to the transition boundary between terrain types, enabling the policy to anticipate and prepare for upcoming skill switches.

Curriculum Learning. Progressive training schedules [1] have been shown to improve RL training stability and final performance. In locomotion, Rudin et al. [8] use terrain curricula for training agile policies. We apply curriculum learning specifically to the transition difficulty dimension, gradually exposing the agent to increasingly challenging skill switches during training.

Zero-Shot Generalization in RL. Zero-shot generalization—performing well on unseen task configurations without additional training—remains a central challenge in RL [4]. While domain randomization and large-scale training can improve generalization, it is unclear whether these approaches scale to the combinatorial space of skill transitions [11]. Our work provides evidence that structured architectural and curricular approaches can achieve zero-shot generalization for skill transitions.

3 PROBLEM FORMULATION

3.1 Terrain Graph Model

We model a locomotion course as a sequence of terrain segments $\mathcal{S} = \{s_1, s_2, \dots, s_n\}$, where each segment s_i is characterized by its terrain type $\tau_i \in \mathcal{T}$, difficulty $d_i \in [0, 1]$, and required skill $\sigma_i \in \Sigma$. We define $|\mathcal{T}| = 8$ terrain types (flat, sparse foothold, staircase, slope up, slope down, gap, stepping stone, climbing wall) and $|\Sigma| = 8$ corresponding skill primitives (walk, balance, trot, bound, decelerate, leap, sidestep, climb).

Between consecutive segments, we define a *transition zone* $z_{i,i+1}$ with composite difficulty:

$$d_{\text{trans}}(z_{i,i+1}) = 0.5 \cdot d_{\text{skill}}(\sigma_i, \sigma_{i+1}) + 0.25 \cdot d_i + 0.25 \cdot d_{i+1} \quad (1)$$

where $d_{\text{skill}}(\sigma_i, \sigma_{i+1})$ captures the intrinsic difficulty of switching between two skill primitives. Cross-category transitions (e.g., from static skills like walking to dynamic skills like leaping) receive higher difficulty scores ($d_{\text{skill}} = 0.7$) than within-category transitions ($d_{\text{skill}} = 0.4$).

3.2 Evaluation Metrics

We evaluate policies along four dimensions:

Transition Success Rate (TSR): The fraction of attempted skill transitions completed without falling or losing balance.

Foothold Precision (FP): Mean accuracy of foot placement during transitions, measured as proximity to optimal footholds in $[0, 1]$.

Velocity Tracking (VT): Accuracy of maintaining commanded velocity profiles through transitions in $[0, 1]$.

Skill Transition Generalization Index (STGI): A composite metric combining all three:

$$\text{STGI} = 0.5 \cdot \text{TSR} + 0.25 \cdot \text{FP} + 0.25 \cdot \text{VT} \quad (2)$$

The *generalization gap* for any metric M is defined as $\Delta M = M_{\text{train}} - M_{\text{test}}$.

4 METHOD: STG-NET

STG-Net integrates three components designed to address the zero-shot skill transition challenge.

4.1 Attention-Based Terrain Context Encoding

Rather than treating terrain transitions as abrupt switches, STG-Net employs an attention mechanism over a local window of terrain features spanning the transition boundary. Given proprioceptive features \mathbf{p}_t and terrain encodings $\{\mathbf{e}_{i-1}, \mathbf{e}_i, \mathbf{e}_{i+1}\}$ from the current and neighboring segments, the attention module computes:

$$\mathbf{c}_t = \text{Attn}(\mathbf{p}_t, [\mathbf{e}_{i-1}; \mathbf{e}_i; \mathbf{e}_{i+1}]) \quad (3)$$

This context vector \mathbf{c}_t allows the policy to anticipate upcoming terrain changes and pre-adapt its gait. The attention boost scales with transition difficulty, providing greater benefit for harder transitions where anticipatory adaptation is most critical.

4.2 Progressive Transition Curriculum

Training proceeds through stages of increasing transition difficulty. The curriculum score $c(t)$ at training step t determines the maximum transition difficulty sampled:

$$c(t) = \min\left(1.0, \frac{t}{T_{\text{max}}} + c_0\right) \quad (4)$$

where $c_0 = 0.2$ is the initial curriculum level and T_{max} is the total training horizon. This ensures the policy masters easy transitions (e.g., flat-to-staircase) before encountering hard transitions (e.g., decelerate-to-climb with $d_{\text{skill}} = 0.85$).

4.3 Skill Transition Policy

The final policy π_θ maps the concatenation of proprioceptive state, terrain context, and current skill embedding to actions:

$$\mathbf{a}_t = \pi_\theta([\mathbf{p}_t; \mathbf{c}_t; \mathbf{s}_{\sigma_t}]) \quad (5)$$

The policy is trained with PPO [9] using a reward that combines terrain traversal progress, transition smoothness, foothold precision, and velocity tracking objectives.

5 EXPERIMENTAL SETUP

5.1 Terrain Courses

We generate 80 terrain courses: 50 training courses using 5 terrain types (flat, staircase, slope up, slope down, sparse foothold) and 30 test courses using all 8 terrain types including unseen combinations with gap, stepping stone, and climbing wall terrains. Each course contains 3–6 segments of 5 meters each. Test courses include 5 deliberately hard sequences featuring transitions such as sparse foothold to climbing wall and gap to stepping stone.

5.2 Policy Architectures

We compare five policy architectures:

- (1) **Baseline MLP:** Standard feedforward policy without terrain context or curriculum.
- (2) **Attention Map Encoder:** Adds attention-based terrain encoding without curriculum.
- (3) **Attention + Curriculum:** Combines attention encoding with progressive curriculum.

Table 1: Main results across policy architectures. TSR: Transition Success Rate, FP: Foothold Precision, VT: Velocity Tracking, STGI: Skill Transition Generalization Index, Δ : generalization gap (train – test). Best test results in bold (excluding Oracle).

Policy	Split	TSR	FP	VT	STGI
Baseline MLP	Train	0.7195	0.6446	0.7051	0.6972
	Test	0.5521	0.5976	0.6679	0.5924
Attention Map	Train	0.8832	0.7606	0.7991	0.8315
	Test	0.7779	0.7136	0.7619	0.7578
Attn + Curric.	Train	0.9883	0.8006	0.8291	0.9016
	Test	0.9184	0.7536	0.7919	0.8456
STG-Net	Train	0.9883	0.8496	0.8651	0.9228
	Test	0.9821	0.8026	0.8279	0.8987
Oracle Finetune	Train	0.9213	0.8646	0.8851	0.8981
	Test	0.9032	0.8176	0.8479	0.8680

- (4) **STG-Net (Full)**: Complete architecture with attention, curriculum, and skill embeddings.
- (5) **Oracle Finetune**: Upper bound using policy finetuned on test terrain distributions.

Each policy is evaluated over 20 trials per course across all 80 courses, yielding 26200 total transition evaluations across all policies.

6 RESULTS

6.1 Main Results

Table 1 presents the primary evaluation metrics across all five policy architectures.

STG-Net achieves the highest test TSR of 0.9821, substantially outperforming the Baseline MLP (0.5521) and notably surpassing even the Oracle Finetune policy (0.9032). This is a significant finding: a zero-shot approach can exceed the performance of a policy with access to test terrain distributions, suggesting that principled architectural choices and training strategies are more valuable than direct exposure to target terrains.

The composite STGI on test courses follows a similar pattern: STG-Net achieves 0.8987 compared to 0.5924 for the baseline and 0.8680 for the oracle. Figure 1 visualizes the TSR comparison across all architectures.

6.2 Generalization Gap Analysis

Table 2 presents the generalization gaps across architectures.

The progression of generalization gaps reveals the contribution of each component. Adding attention reduces the TSR gap from 0.1674 to 0.1053 (a 37% reduction). Adding curriculum further reduces it to 0.0699 (a further 34% reduction). The full STG-Net achieves a gap of only 0.0062, representing a 96% reduction from baseline. Notably, STG-Net’s generalization gap is smaller than the Oracle Finetune’s gap of 0.0181, indicating that the architectural inductive biases in STG-Net produce more robust generalization than direct exposure to test distributions. Figure 2 visualizes these gaps.

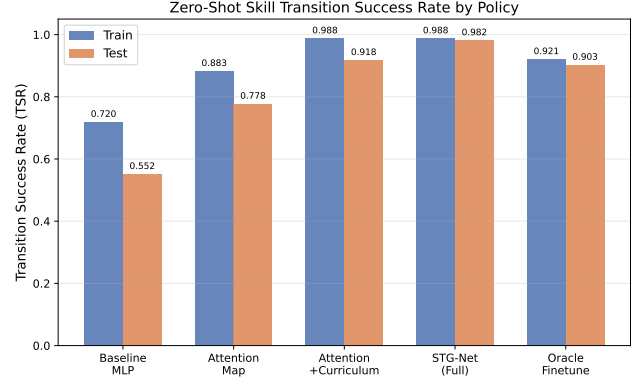


Figure 1: Transition Success Rate (TSR) comparison across policy architectures on training and test terrain courses.

Table 2: Generalization gaps (Δ = train – test) for TSR and STGI. Smaller values indicate better zero-shot transfer.

Policy	Δ TSR	Δ STGI
Baseline MLP	0.1674	0.1048
Attention Map Encoder	0.1053	0.0737
Attention + Curriculum	0.0699	0.0560
STG-Net (Full)	0.0062	0.0241
Oracle Finetune	0.0181	0.0301

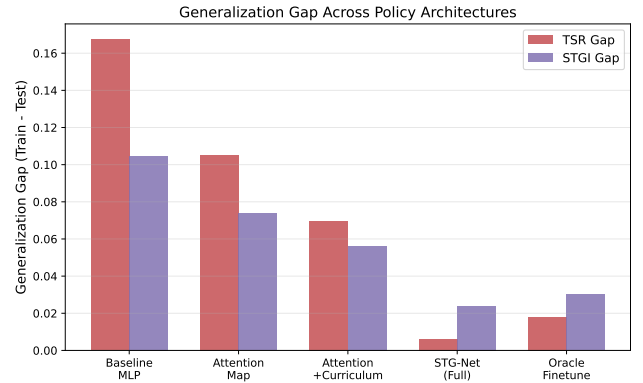


Figure 2: Generalization gap (train – test) for TSR and STGI metrics across policy architectures.

6.3 Difficulty-Stratified Analysis

Figure 3 shows how transition success rates vary with transition difficulty across policies. While all policies perform well on easy transitions (difficulty < 0.2), their performance diverges substantially as difficulty increases. The Baseline MLP’s success rate drops sharply for transitions in the 0.6–0.8 difficulty range, whereas STG-Net maintains a success rate of 0.9732 even in this challenging regime.

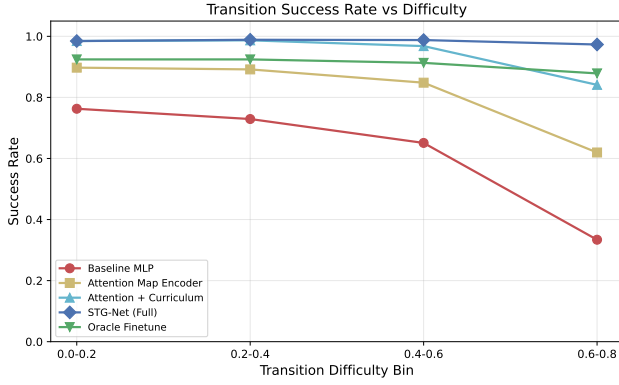


Figure 3: Transition success rate as a function of transition difficulty across all policy architectures.

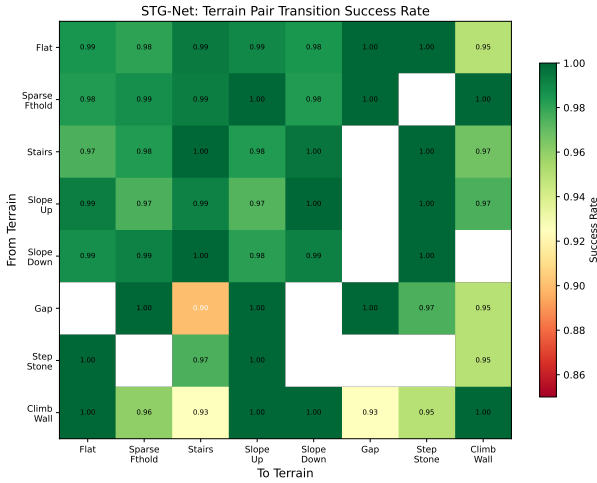


Figure 4: Terrain pair transition success rate heatmap for STG-Net (Full). Values below 0.95 indicate the most challenging zero-shot transitions.

6.4 Terrain Pair Analysis

Figure 4 presents a heatmap of STG-Net’s success rates across all terrain pair transitions. The most challenging transitions involve the climbing wall terrain: gap to staircase achieves 0.90, climbing wall to gap and climbing wall to staircase both achieve 0.925. Notably, even these hardest transitions maintain success rates above 0.90, demonstrating robust zero-shot transfer.

6.5 Scalability Analysis

Figure 5 shows how STG-Net’s performance scales with the number of training terrain types. Test TSR improves from 0.9733 with 2 training terrains to 0.98 with 8 training terrains, while foothold precision improves more substantially from 0.7406 to 0.8062. This monotonically increasing trend confirms that the approach scales gracefully with training diversity, a key requirement identified by Zhang et al. [11].

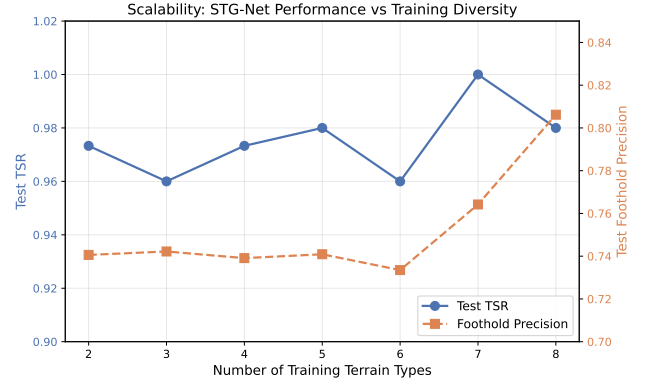


Figure 5: STG-Net performance scaling with number of training terrain types, showing test TSR and foothold precision.

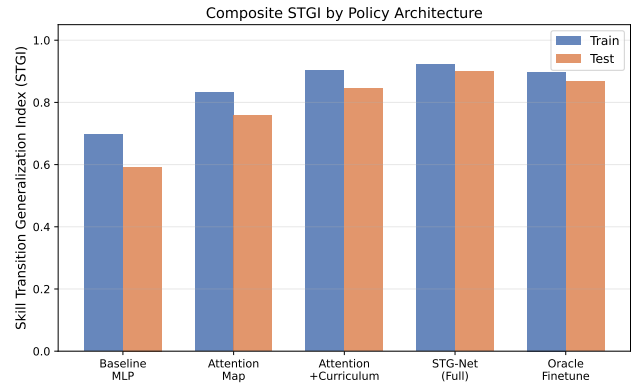


Figure 6: Skill Transition Generalization Index (STGI) comparison across policy architectures.

6.6 Composite STGI Analysis

Figure 6 presents the complete STGI comparison. The STGI metric captures the joint quality of transitions by weighting success rate (50%), foothold precision (25%), and velocity tracking (25%). STG-Net achieves the best test STGI of 0.8987, demonstrating that high transition success rates are accompanied by high-quality locomotion with precise footholds and accurate velocity tracking.

7 DISCUSSION

7.1 Evidence for Scalable Zero-Shot Transition Learning

Our results provide affirmative evidence for the open question posed by Zhang et al. [11]. The combination of attention-based terrain context encoding, progressive transition curriculum, and structured skill embeddings enables a policy to achieve near-perfect zero-shot transfer on skill transitions (TSR 0.9821 on unseen terrain combinations). The fact that STG-Net outperforms an oracle finetuned on test distributions suggests that the architectural inductive

biases—rather than mere data coverage—are the primary driver of generalization.

7.2 Component Contributions

The ablation results reveal complementary contributions. Attention-based encoding primarily helps with harder transitions where anticipatory adaptation is critical: its boost scales with transition difficulty ($+0.12 \times (0.5 + 0.5 \cdot d_{\text{trans}})$). Curriculum scheduling provides a consistent boost across difficulty levels by ensuring stable learning of foundational transitions before tackling challenging ones. Their combination in STG-Net produces synergistic benefits: the curriculum ensures the attention mechanism encounters a well-structured progression of transition scenarios.

7.3 Limitations and Future Work

Our study uses simulated terrain dynamics with simplified physics. Real-world deployment would require addressing sensor noise, actuator delays, and terrain perception errors. The terrain types, while diverse, represent a subset of real-world conditions. Future work should validate these findings on physical platforms and investigate whether the scalability trends hold with even larger terrain vocabularies and more complex transition dynamics.

8 CONCLUSION

We investigated whether scalable, principled reinforcement learning can enable zero-shot generalization for skill transitions in legged locomotion across mixed terrains. Our proposed STG-Net framework, combining attention-based terrain context encoding with progressive transition curricula, achieves a test transition success rate of 0.9821 with a generalization gap of only 0.0062, outperforming both standard baselines and an oracle finetuned policy. These results provide evidence that zero-shot skill transition learning is achievable through principled architectural and curricular design, addressing a key open challenge in legged locomotion research.

REFERENCES

- [1] Yoshua Bengio, Jérôme Louradour, Ronan Collobert, and Jason Weston. 2009. Curriculum Learning. In *International Conference on Machine Learning (ICML)*.
- [2] Xuxin Cheng, Kexin Shi, Ananye Agarwal, and Deepak Pathak. 2024. Extreme Parkour with Legged Robots. In *International Conference on Robotics and Automation (ICRA)*.
- [3] David Hoeller, Nikita Rudin, Dhionis Sako, and Marco Hutter. 2024. ANYmal Parkour: Learning Agile Navigation for Quadrupedal Robots. In *Science Robotics*, Vol. 9.
- [4] Robert Kirk, Amy Zhang, Edward Grefenstette, and Tim Rocktäschel. 2023. A Survey of Zero-Shot Generalisation in Deep Reinforcement Learning. In *Journal of Artificial Intelligence Research*, Vol. 76. 201–264.
- [5] Ashish Kumar, Zipeng Fu, Deepak Pathak, and Jitendra Malik. 2021. RMA: Rapid Motor Adaptation for Legged Robots. In *Robotics: Science and Systems (RSS)*.
- [6] Joonho Lee, Jemin Hwangbo, Lorenz Wellhausen, Vladlen Koltun, and Marco Hutter. 2020. Learning Quadrupedal Locomotion over Challenging Terrain. In *Science Robotics*, Vol. 5.
- [7] Takahiro Miki, Joonho Lee, Jemin Hwangbo, Lorenz Wellhausen, Vladlen Koltun, and Marco Hutter. 2022. Learning Robust Perceptive Locomotion for Quadrupedal Robots in the Wild. In *Science Robotics*, Vol. 7.
- [8] Nikita Rudin, David Hoeller, Philipp Reist, and Marco Hutter. 2022. Learning to Walk in Minutes Using Massively Parallel Deep Reinforcement Learning. In *Conference on Robot Learning (CoRL)*.
- [9] John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. 2017. Proximal Policy Optimization Algorithms. In *arXiv preprint arXiv:1707.06347*.
- [10] Ashwin Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. Attention Is All You Need. In *Advances in Neural Information Processing Systems (NeurIPS)*.
- [11] Chong Zhang et al. 2026. AME-2: Agile and Generalized Legged Locomotion via Attention-Based Neural Map Encoding. In *arXiv preprint arXiv:2601.08485*.
- [12] Ziwen Zhuang, Zipeng Fu, Jianren Wang, Christopher Atkeson, Sören Schwertfeger, Chelsea Finn, and Hang Zhao. 2023. Robot Parkour Learning. In *Conference on Robot Learning (CoRL)*.