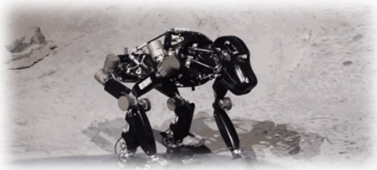
Subequivariant Graph Reinforcement Learning in 3D Environments

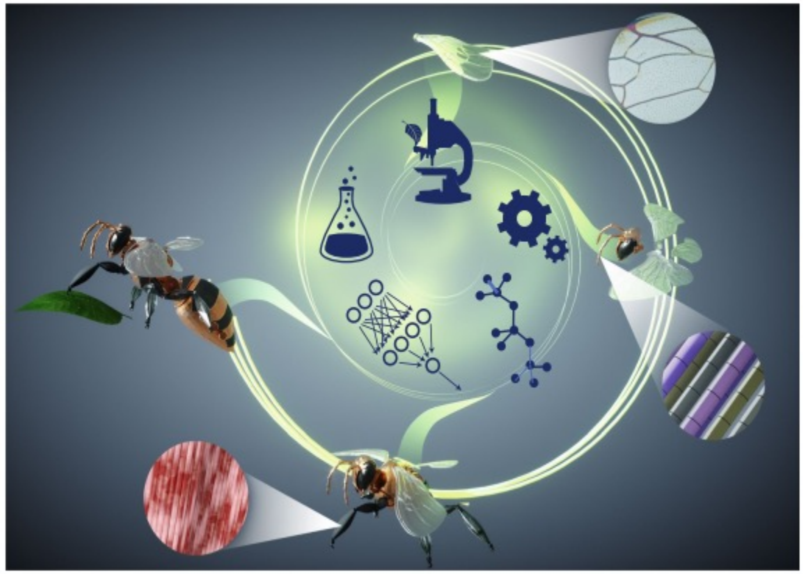
Project Page: https://alpc91.github.io/SGRL/

Runfa Chen*, Jiaqi Han*, Fuchun Sun, Wenbing Huang

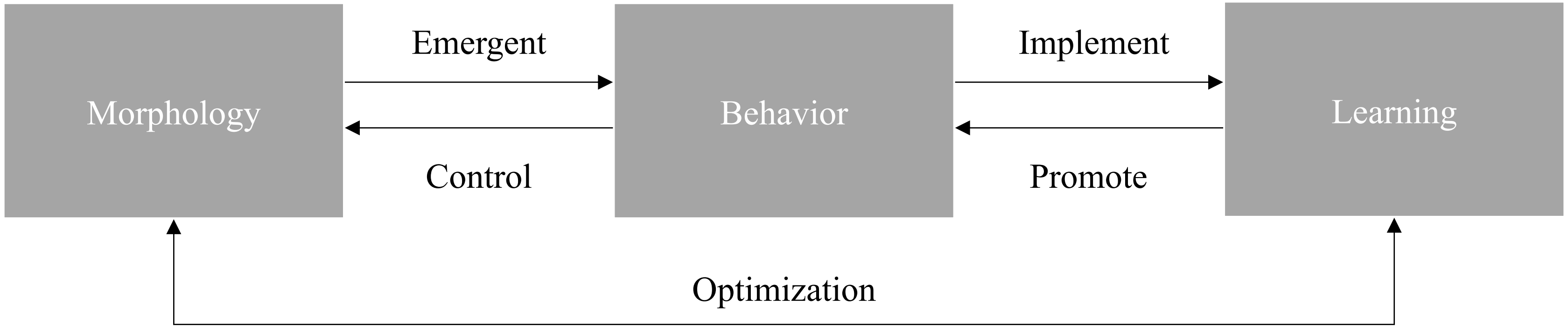
Background: Physical AI (PAI)


- Research in PAI involves close interaction between:
  - The structure of brain
  - Body morphology
  - Interaction with the environment



[Miriyev et.al. PAI. Nature Machine Intelligence 2020]



The architecture of morphology-based embodied intelligence

A. Miriyev and M. Kovač, “Skills for physical artificial intelligence,” Nature Machine Intelligence, vol. 2, pp. 658-660, 2020.

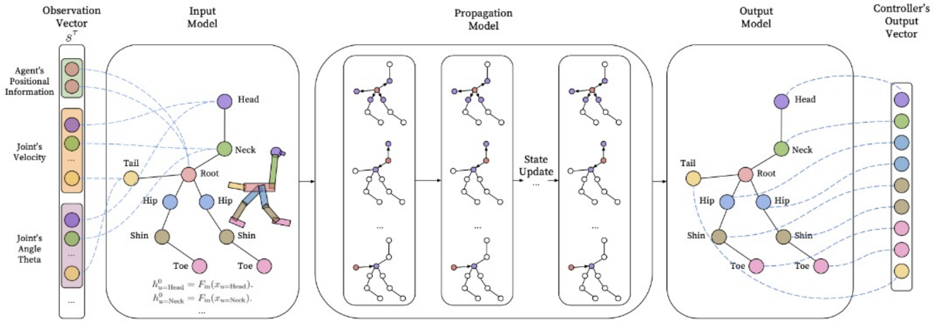
Background: Morphology

- Challenges in agent morphology

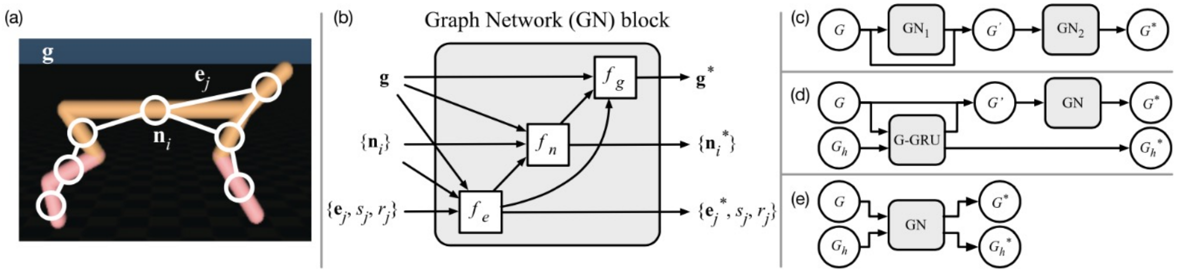
  - Each robot has a different morphology.
  - A separate policy is trained for each robotics setup.
  - Hard to generalize.

Prior Attempts

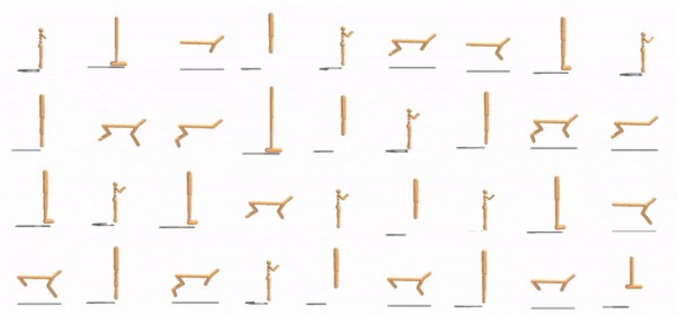
2D Planar!



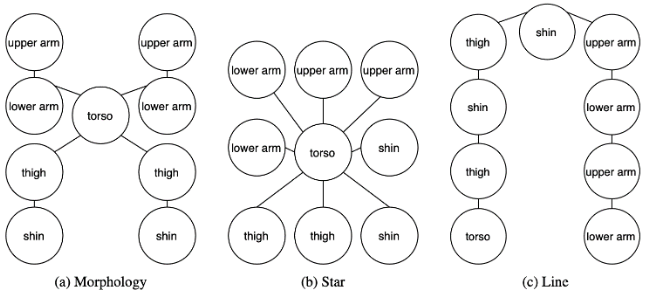
[Wang et.al. NerveNet. ICLR 2018]



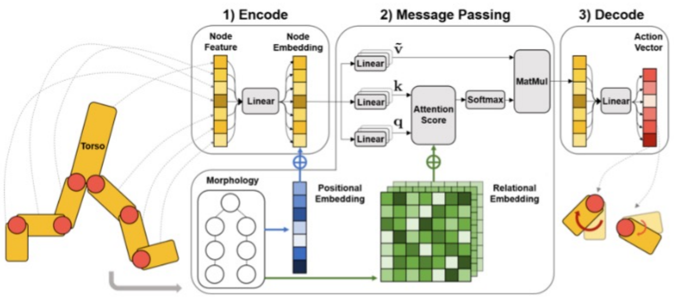
[Sanchez-Gonzalez et.al. GN. ICML 2018]



[Huang et.al. SMP. ICML 2020]

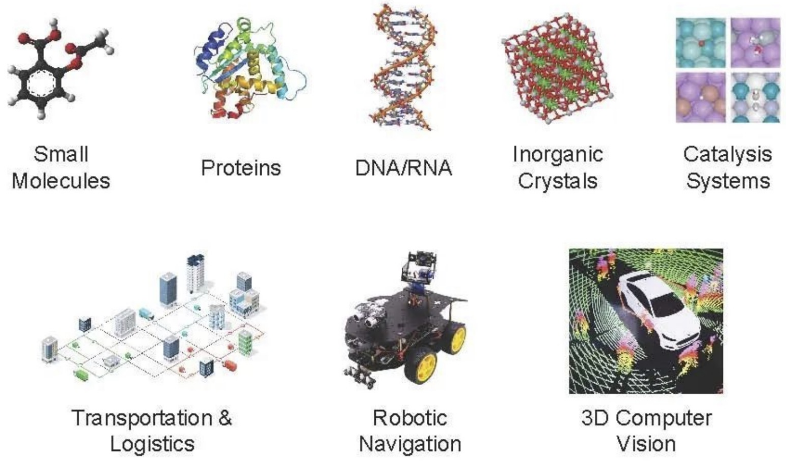


[Kurin et.al. AMORPHEUS. ICLR 2021]

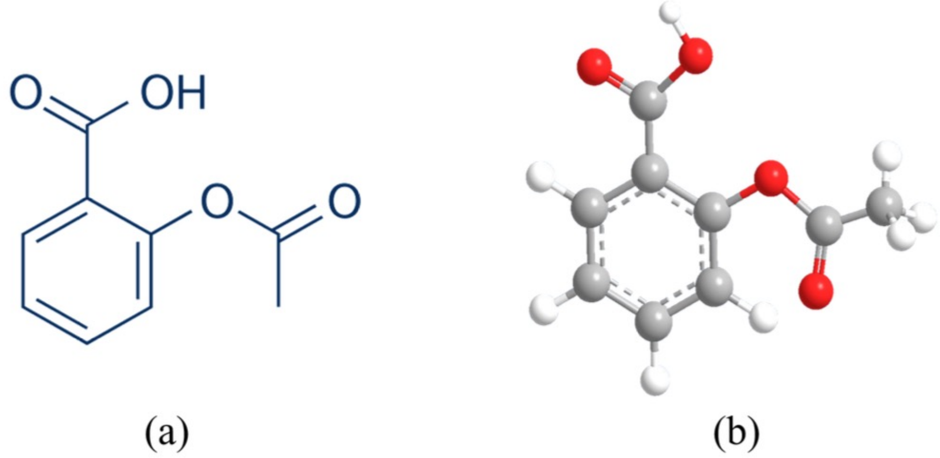


[Hong et.al. SWAT. ICLR 2022]

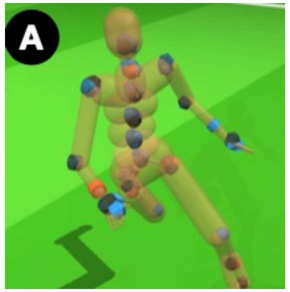
Motivation



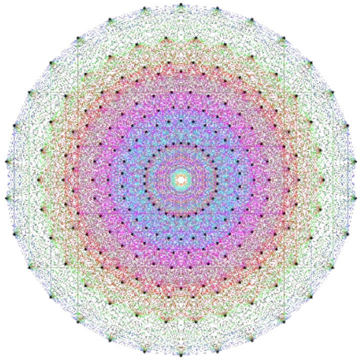
Geometric Structure and Systems
[Joshi et.al. Geometric Graph Neural Networks. NeurIPS 2022]



The examples in the figure correspond to Aspirin (acetylsalicylic acid) molecules:(a) topology graph and (b) geometric graph.



3D Geometric Graph!

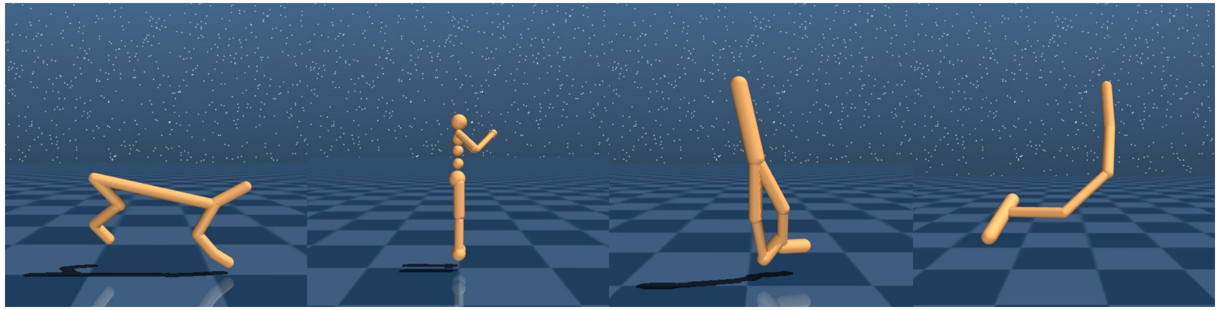


Symmetry!

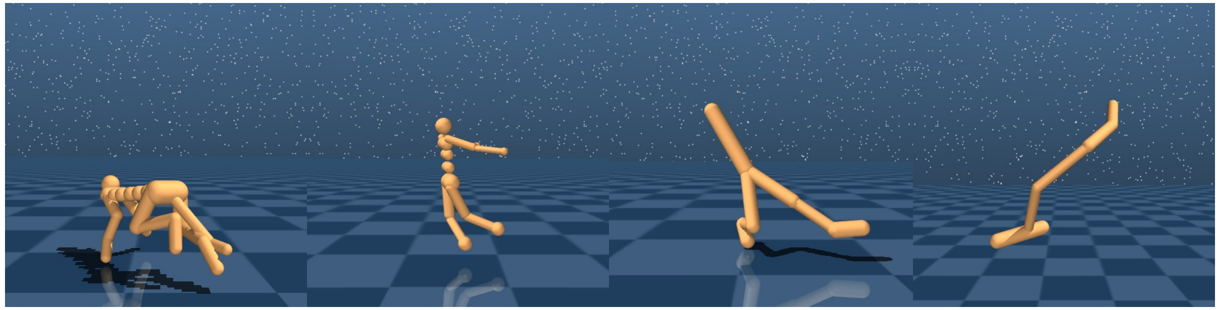
Our Setup: 3D-SGRL

Table 1. Comparison in the problem setup.

		2D-Planar	**Our 3D-SGRL**
State Space	Range	xoz-plane	3D space
	Initial	x^+-axis	Arbitrary direction
	Target	x^+-axis	Arbitrary direction
Action Space	# Actuators	1 per joint	3 per joint
	DoF	1 per joint	3 per joint
Symmetry	External Force	NULL	Gravity $\vec{\boldsymbol{g}}$, Target $\vec{\boldsymbol{d}}$
	Group	$\emptyset$	$\mathrm{O}_{\vec{\boldsymbol{g}}}(3)$

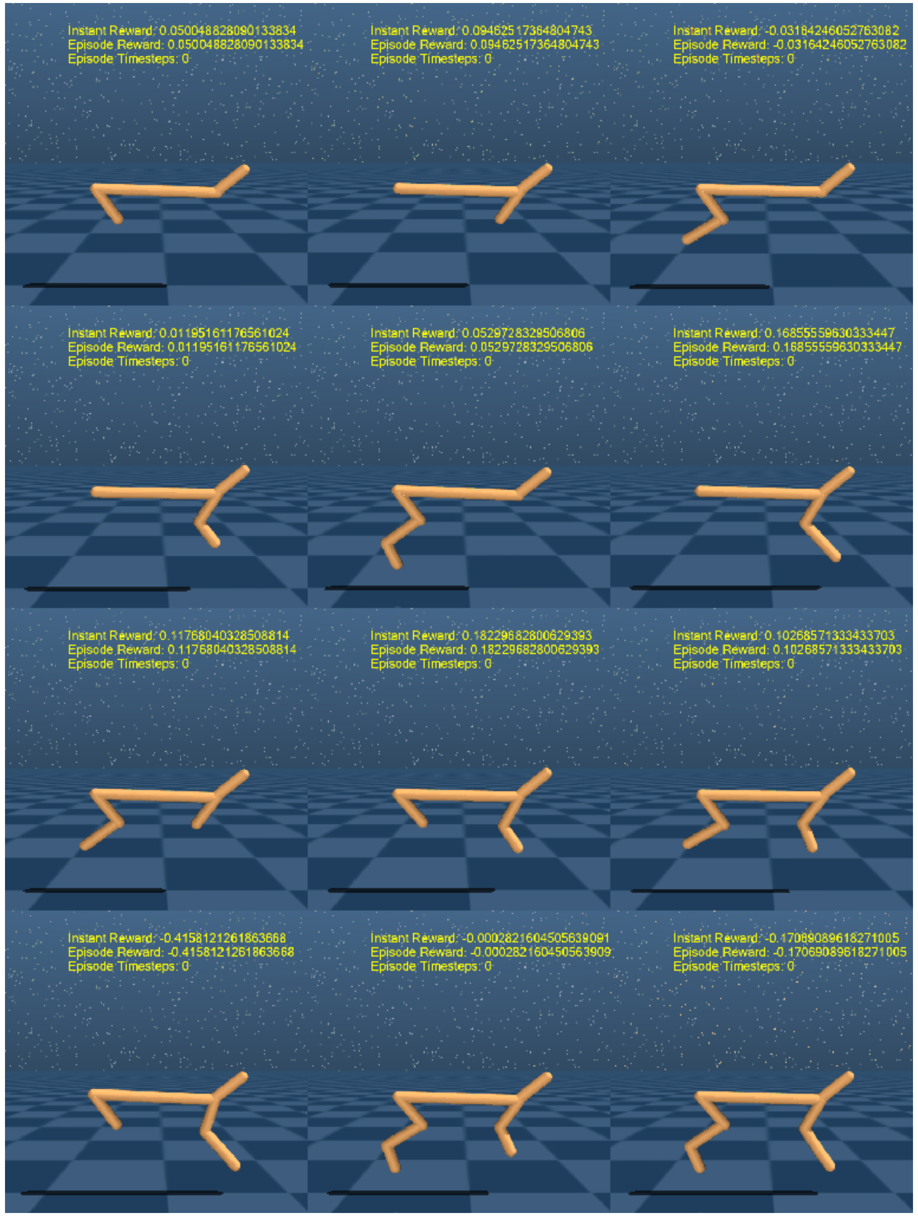


(a) 2D Planar Locomotion Environments

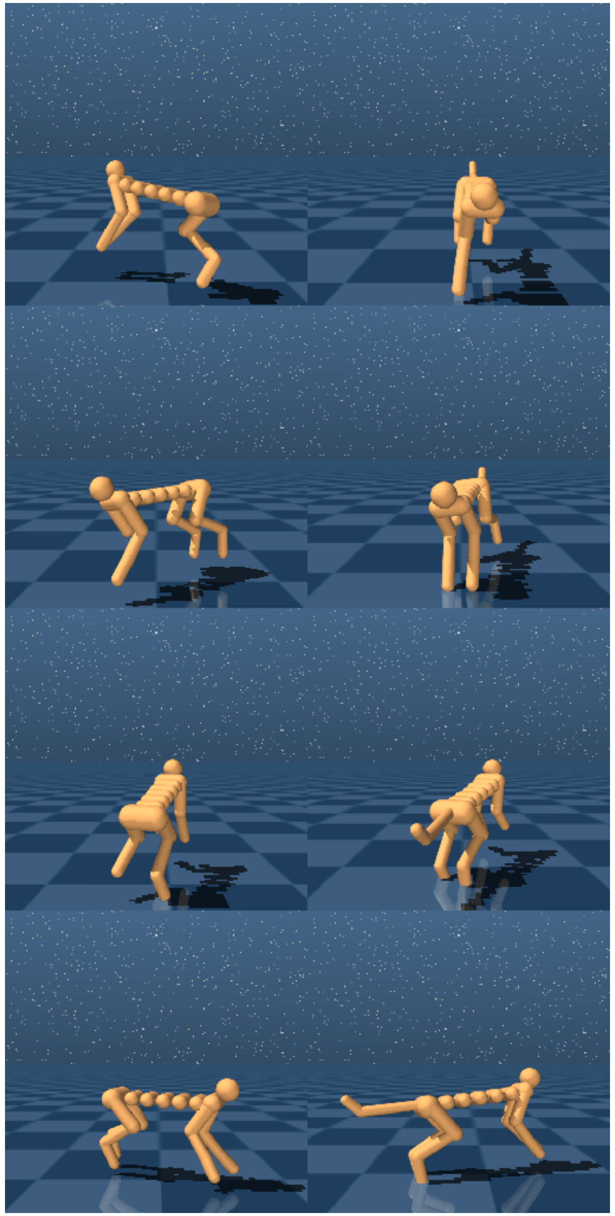


(b) 3D Subequivariant Locomotion Environments cheetah

2D-Planar



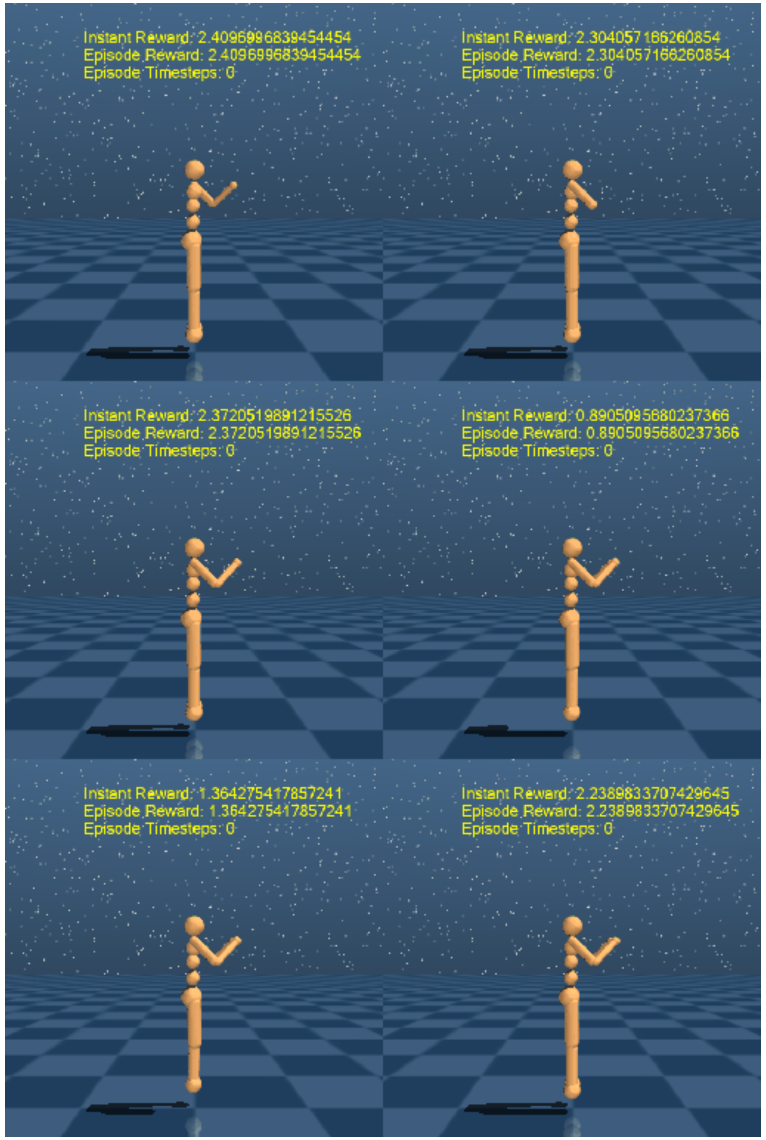
3D-SGRL

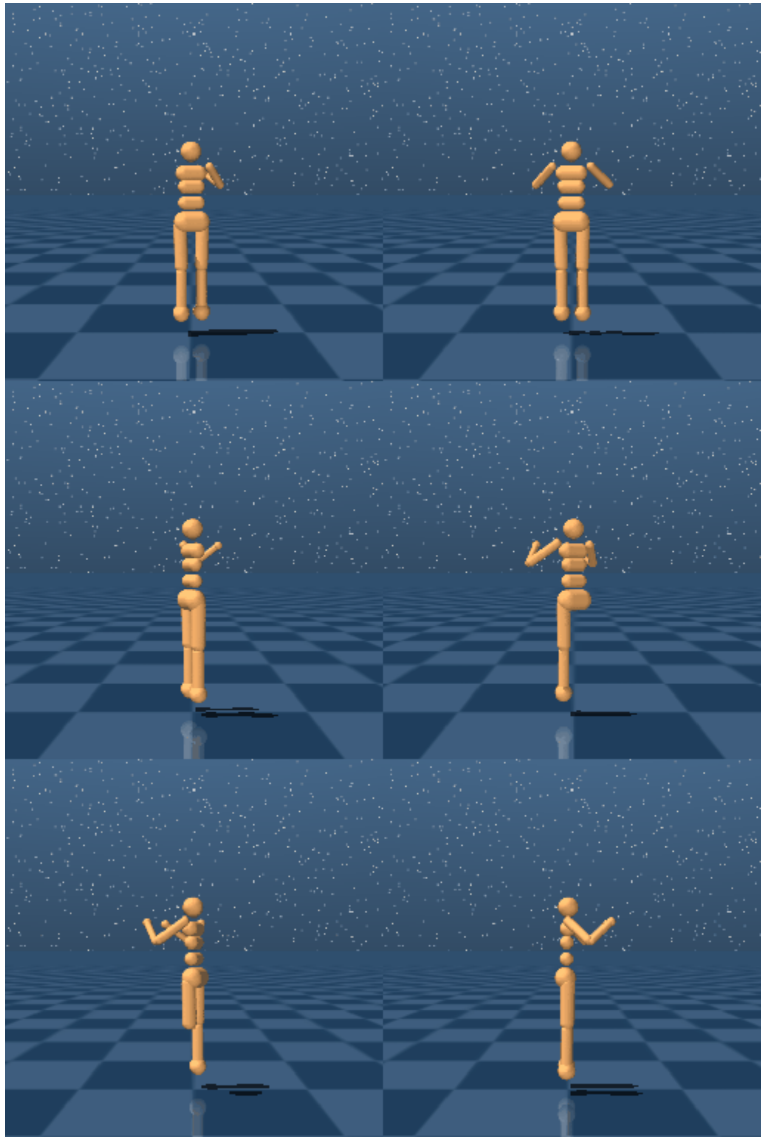


humanoid

2D-Planar

3D-SGRL

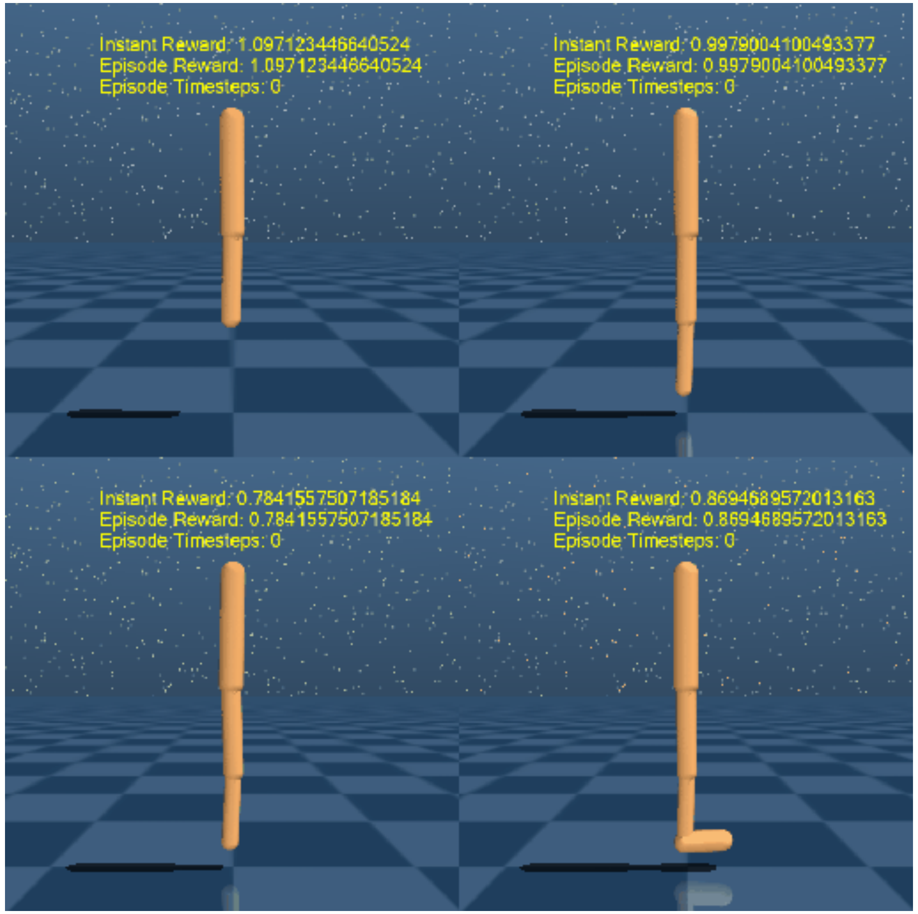


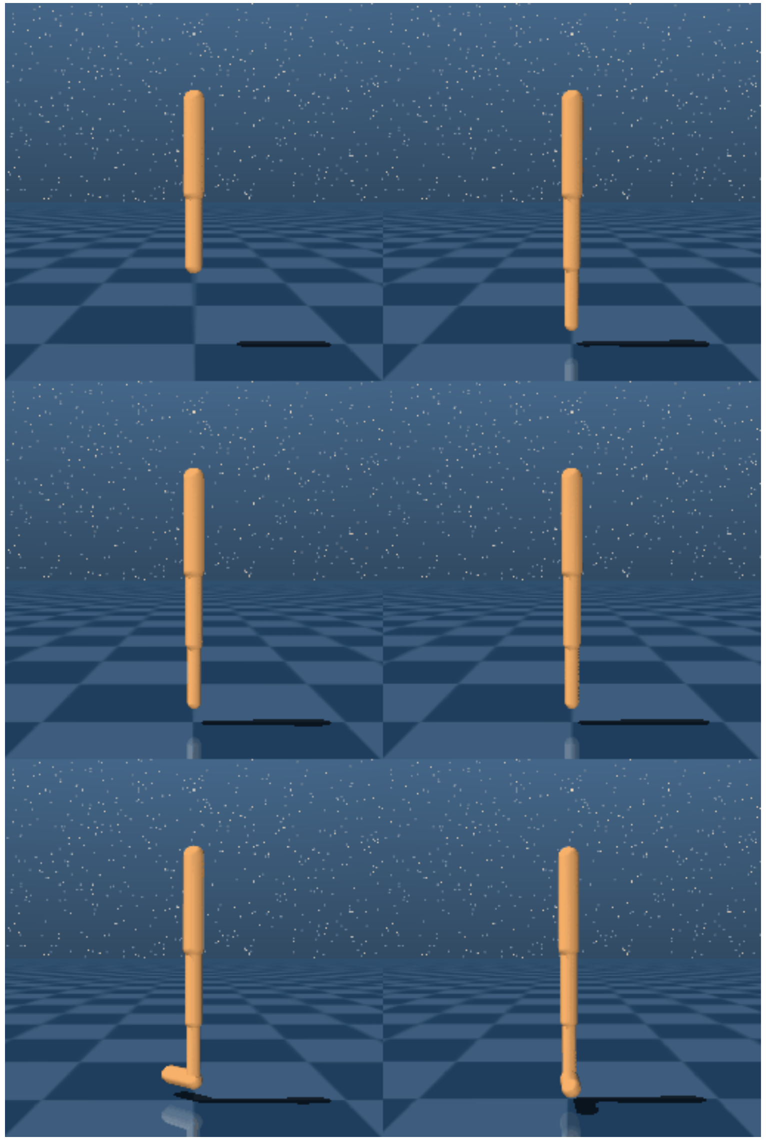


walker

2D-Planar

3D-SGRL

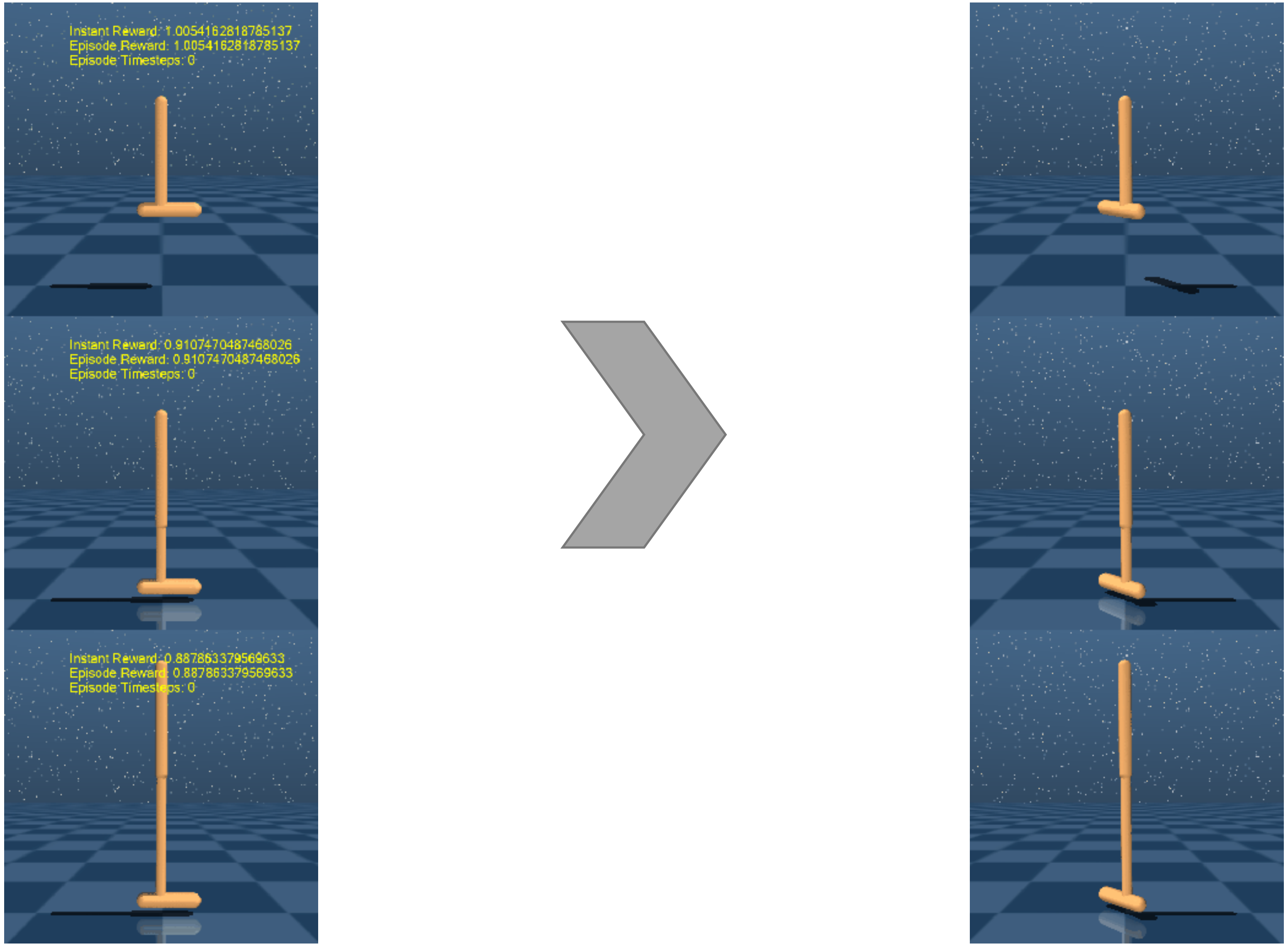




hopper

2D-Planar

3D-SGRL



Method: Subequivariant Transformer (SET)

➢ Equivariance and Subequivariance

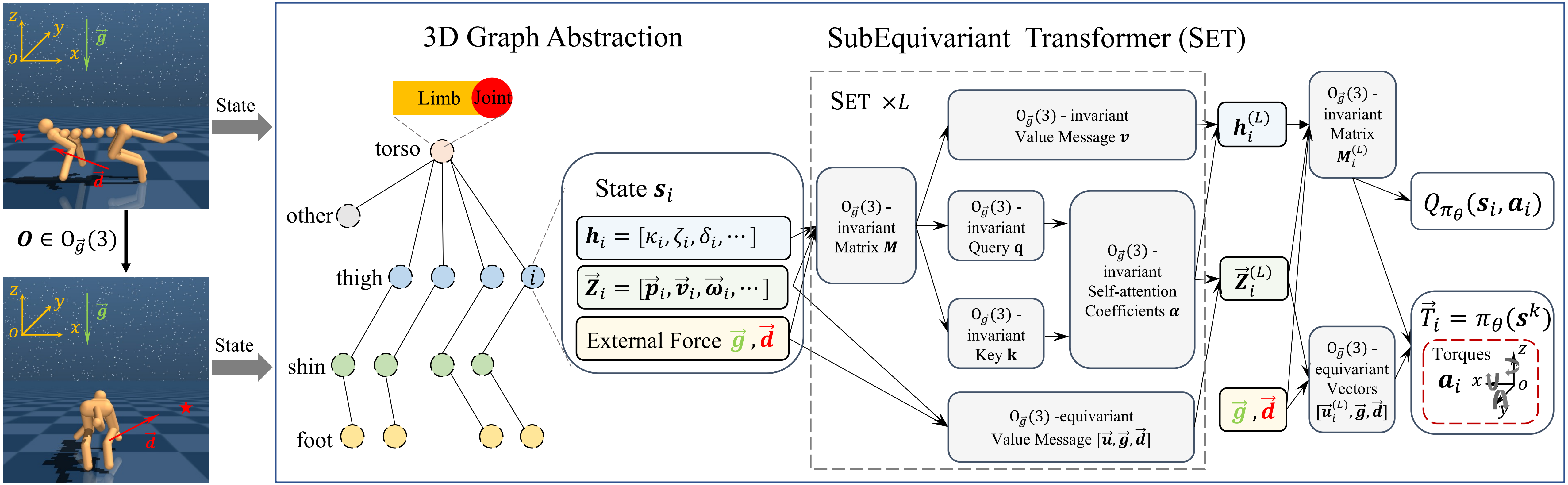
Definition 2.1 (E(3)-equivariance). Suppose $\vec{\boldsymbol{Z}}$ to be 3D geometric vectors (positions, velocities, etc) that are steerable by E(3) transformations, and $\boldsymbol{h}$ non-steerable features.

- The function f is E(3)-equivariant, if for any transformation $g \in \mathrm{E}(3), f(g \cdot \vec{\boldsymbol{Z}}, \boldsymbol{h}) = g \cdot f(\vec{\boldsymbol{Z}}, \boldsymbol{h}), \forall \vec{\boldsymbol{Z}} \in \mathbb{R}^{3\times m}, \boldsymbol{h} \in \mathbb{R}^{d}$.
- Similarly, f is invariant if $f(g \cdot \vec{\boldsymbol{Z}}, \boldsymbol{h}) = f(\vec{\boldsymbol{Z}}, \boldsymbol{h})$.

Han et al. (2022a) additionally considers equivariance on the subgroup of $\mathrm{O}(3)$, induced by the external force $\vec{\boldsymbol{g}} \in \mathbb{R}^3$ like gravity, defined as

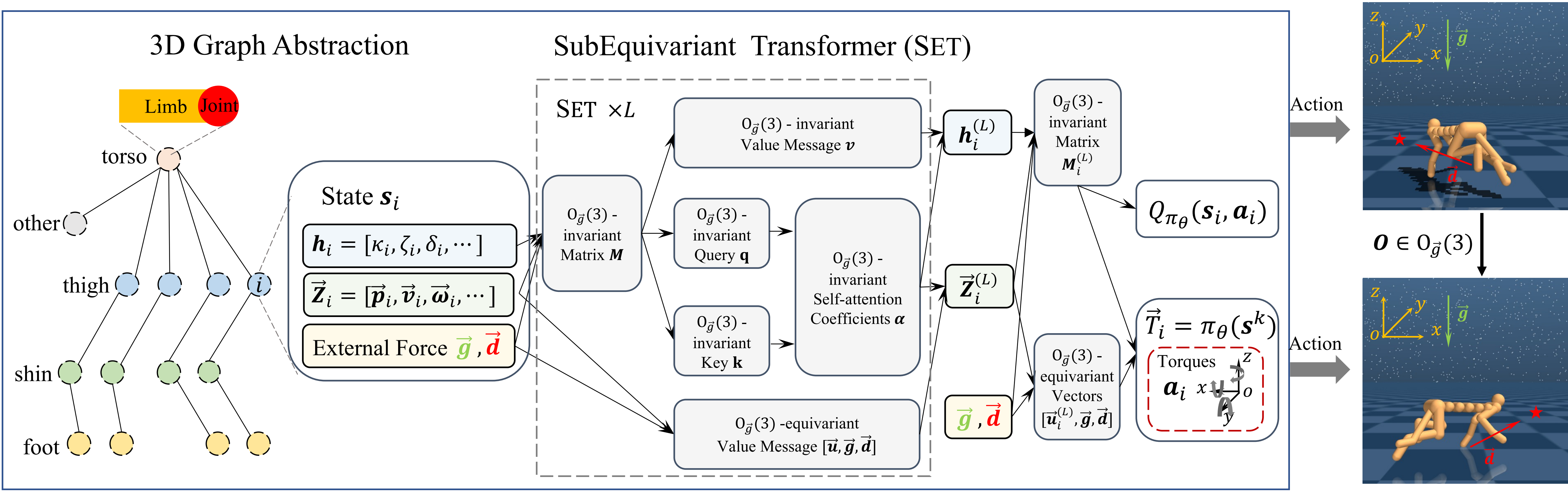
$$\mathrm{O}_{\vec{\boldsymbol{g}}}(3) := \{\boldsymbol{O} \in \mathbb{R}^{3\times 3} \mid \boldsymbol{O}^\top \boldsymbol{O} = \boldsymbol{I}, \boldsymbol{O}\vec{\boldsymbol{g}} = \vec{\boldsymbol{g}}\}$$

Method: SET



State
State
$\boldsymbol{O} \in \mathrm{O}_{\vec{g}}(3)$
3D Graph Abstraction
SubEquivariant Transformer (SET)
Limb
Joint
torso
other
thigh
shin
foot
i
State $\boldsymbol{s}_i$
$\boldsymbol{h}_i = [\kappa_i, \zeta_i, \delta_i, \cdots]$
$\vec{\boldsymbol{Z}}_i = [\vec{\boldsymbol{p}}_i, \vec{\boldsymbol{v}}_i, \vec{\boldsymbol{\omega}}_i, \cdots]$
External Force $\vec{g}$, $\vec{d}$
SET $\times L$
$\mathrm{O}_{\vec{g}}(3)$ - invariant Matrix $\boldsymbol{M}$
$\mathrm{O}_{\vec{g}}(3)$ - invariant Value Message $\boldsymbol{v}$
$\mathrm{O}_{\vec{g}}(3)$ - invariant Query $\mathbf{q}$
$\mathrm{O}_{\vec{g}}(3)$ - invariant Key $\mathbf{k}$
$\mathrm{O}_{\vec{g}}(3)$ - invariant Self-attention Coefficients $\boldsymbol{\alpha}$
$\mathrm{O}_{\vec{g}}(3)$ -equivariant Value Message $[\vec{\boldsymbol{u}}, \vec{\boldsymbol{g}}, \vec{\boldsymbol{d}}]$
$\boldsymbol{h}_i^{(L)}$
$\vec{\boldsymbol{Z}}_i^{(L)}$
$\vec{g}$, $\vec{d}$
$\mathrm{O}_{\vec{g}}(3)$ - invariant Matrix $\boldsymbol{M}_i^{(L)}$
$\mathrm{O}_{\vec{g}}(3)$ - equivariant Vectors $[\vec{\boldsymbol{u}}_i^{(L)}, \vec{\boldsymbol{g}}, \vec{\boldsymbol{d}}]$
$Q_{\pi_\theta}(\boldsymbol{s}_i, \boldsymbol{a}_i)$
$\vec{T}_i = \pi_\theta(\boldsymbol{s}^k)$
Torques
$\boldsymbol{a}_i$

Method: SET



3D Graph Abstraction
Limb
Joint
torso
other
thigh
shin
foot
i
State $\boldsymbol{s}_i$
$\boldsymbol{h}_i = [\kappa_i, \zeta_i, \delta_i, \cdots]$
$\vec{\boldsymbol{Z}}_i = [\vec{\boldsymbol{p}}_i, \vec{\boldsymbol{v}}_i, \vec{\boldsymbol{\omega}}_i, \cdots]$
External Force $\vec{g}$, $\vec{d}$
SubEquivariant Transformer (SET)
SET $\times L$
$\mathrm{O}_{\vec{g}}(3)$ - invariant Matrix $\boldsymbol{M}$
$\mathrm{O}_{\vec{g}}(3)$ - invariant Value Message $\boldsymbol{v}$
$\mathrm{O}_{\vec{g}}(3)$ - invariant Query $\mathbf{q}$
$\mathrm{O}_{\vec{g}}(3)$ - invariant Key $\mathbf{k}$
$\mathrm{O}_{\vec{g}}(3)$ - invariant Self-attention Coefficients $\boldsymbol{\alpha}$
$\mathrm{O}_{\vec{g}}(3)$ -equivariant Value Message $[\vec{\boldsymbol{u}}, \vec{\boldsymbol{g}}, \vec{\boldsymbol{d}}]$
$\boldsymbol{h}_i^{(L)}$
$\vec{\boldsymbol{Z}}_i^{(L)}$
$\vec{g}$, $\vec{d}$
$\mathrm{O}_{\vec{g}}(3)$ - invariant Matrix $\boldsymbol{M}_i^{(L)}$
$\mathrm{O}_{\vec{g}}(3)$ - equivariant Vectors $[\vec{\boldsymbol{u}}_i^{(L)}, \vec{\boldsymbol{g}}, \vec{\boldsymbol{d}}]$
$Q_{\pi_\theta}(\boldsymbol{s}_i, \boldsymbol{a}_i)$
$\vec{T}_i = \pi_\theta(\boldsymbol{s}^k)$
Torques
$\boldsymbol{a}_i$
Action
Action
$\boldsymbol{O} \in \mathrm{O}_{\vec{g}}(3)$
z y x o $\vec{g}$ $\vec{d}$

Experiments

Task definition

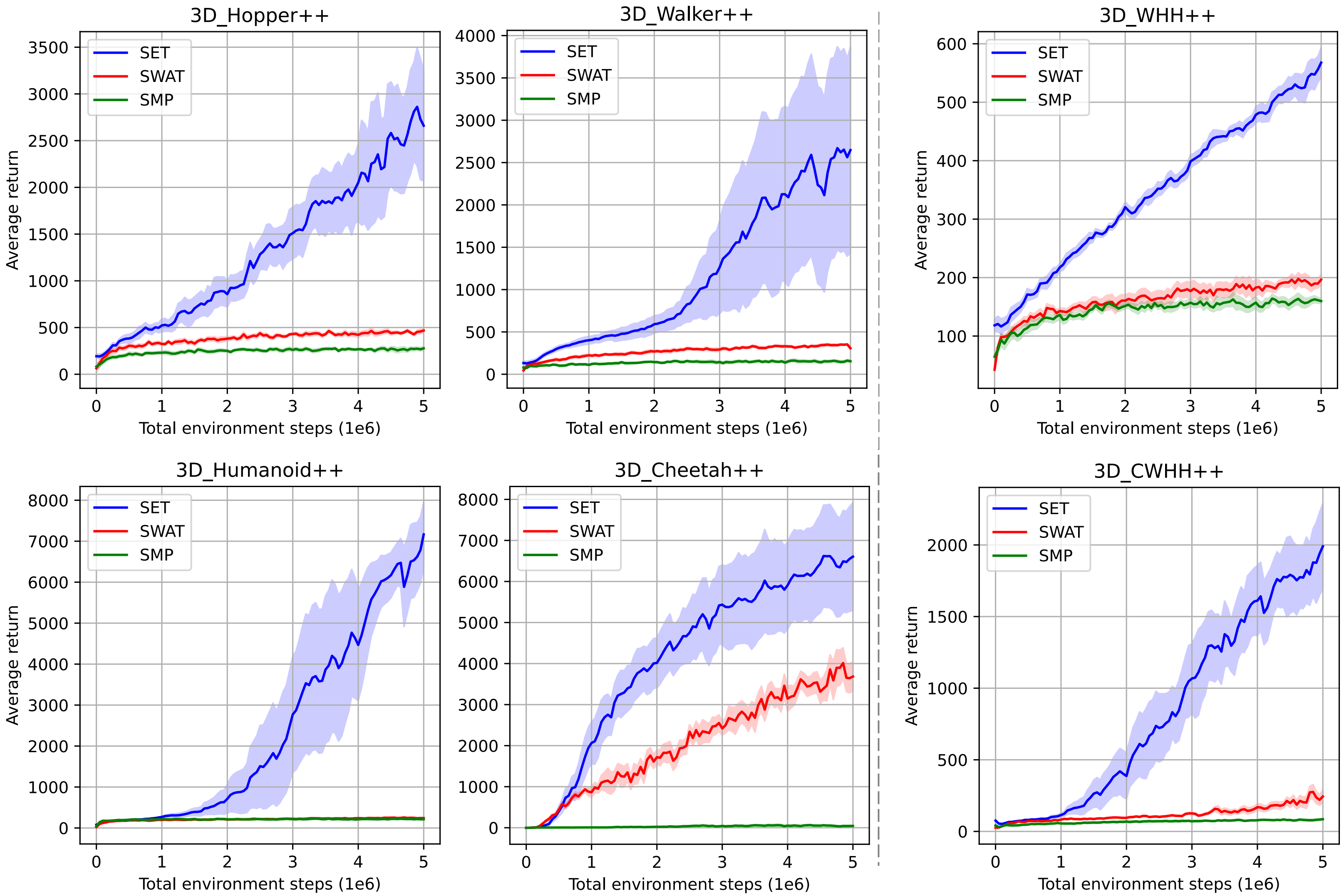
1. Multi-task with different morphologies: For each multi-task environment, a single policy is simultaneously trained on multiple variants.

2. Zero-Shot Generalization: We take the trained policies from multi-task and test on the unseen zeroshot testing variants.

3. Evaluation on v2-variants: We evaluate SET in a transfer learning setting where the trained policies from multi-task are tested and transferred on the v2-variants environments.

4. Single-task Learning: The policy in each plot is trained on one morphology variant and evaluated on this variant.

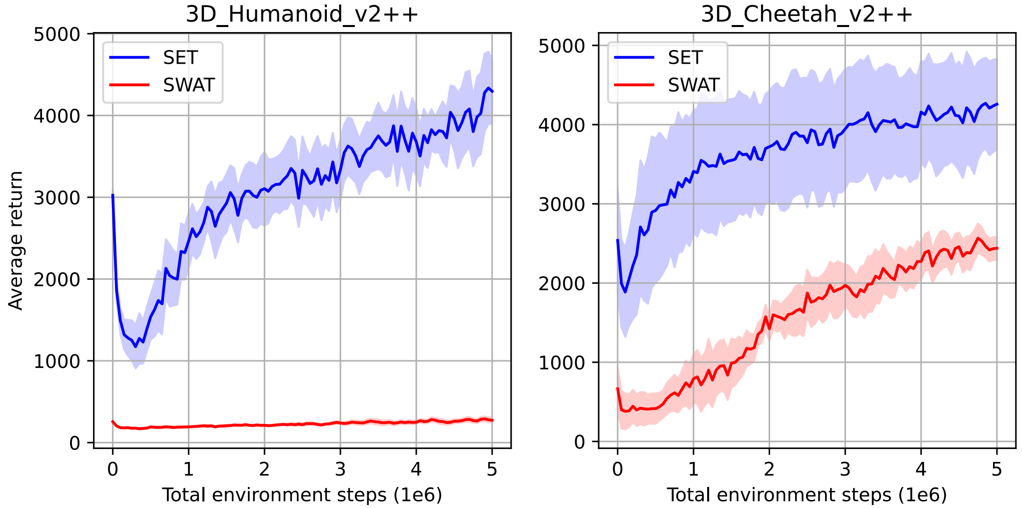
Experiments: Multi-Task



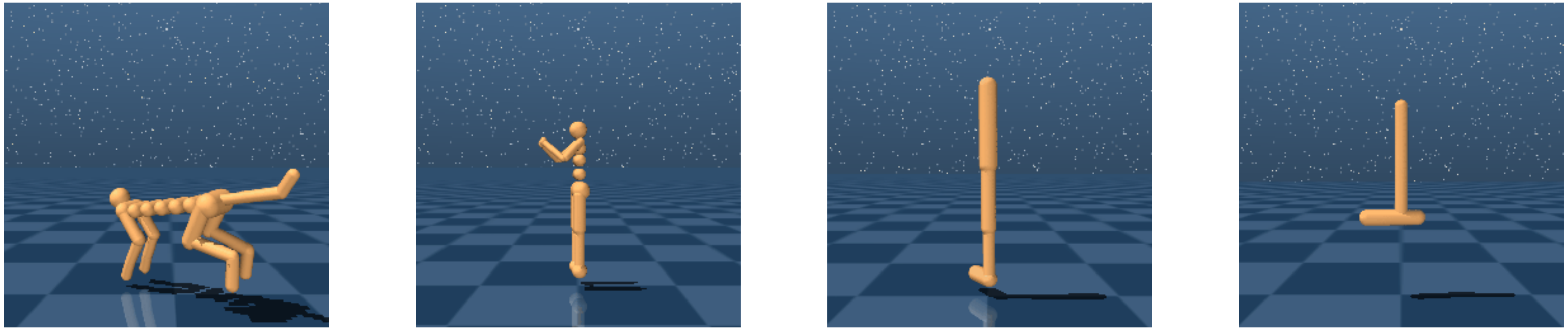
Experiments: Zero-Shot

Environment	SET	SWAT	SMP
in-domain (3D_Walker++, 3D_Humanoid++, 3D_Cheetah++)			
3d_walker_3	**276.2** ± 17.4	207.0 ± 52.7	56.8 ± 15.1
3d_walker_6	**431.3** ± 146.2	358.0 ± 58.9	143.4 ± 50.7
3d_humanoid_7	**244.8** ± 7.9	170.3 ± 51.7	190.9 ± 16.2
3d_humanoid_8	**299.6** ± 23.7	141.4 ± 22.1	185.4 ± 9.2
3d_cheetah_11	**4643.9** ± 292.6	1785.3 ± 999.3	2.0 ± 2.9
3d_cheetah_12	**916.0** ± 39.7	744.1 ± 317.1	29.8 ± 10.7
cross-domain (3D_CWHH++)			
3d_walker_3	**206.8** ± 37.4	17.9 ± 13.7	18.0 ± 22.9
3d_walker_6	**243.7** ± 32.3	114.9 ± 40.3	103.9 ± 1.8
3d_humanoid_7	**161.9** ± 3.4	152.0 ± 6.8	124.2 ± 15.7
3d_humanoid_8	**180.0** ± 6.5	156.6 ± 1.7	129.3 ± 0.1
3d_cheetah_11	**1078.1** ± 722.8	4.3 ± 1.6	6.2 ± 0.5
3d_cheetah_12	**3038.3** ± 2803.3	349.7 ± 304.3	6.6 ± 1.2

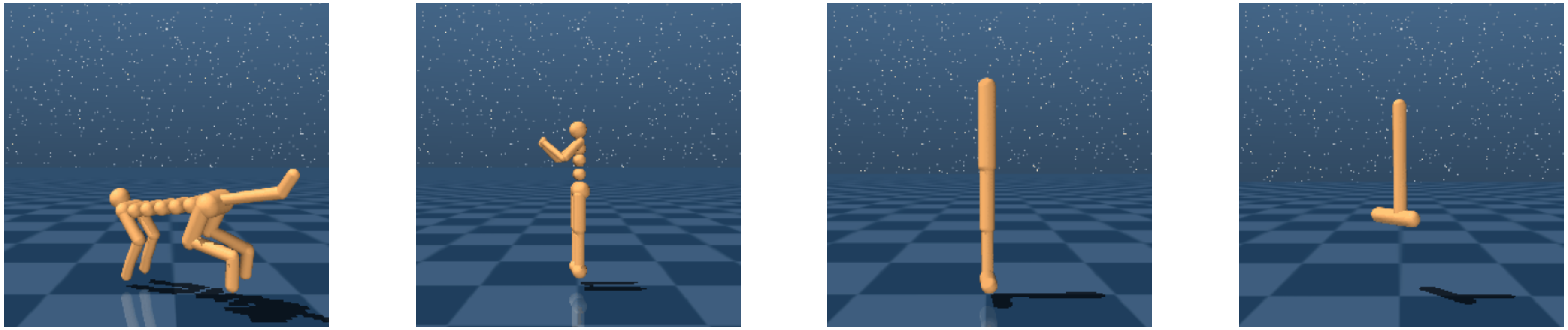
Experiments: v2-variants



SWAT



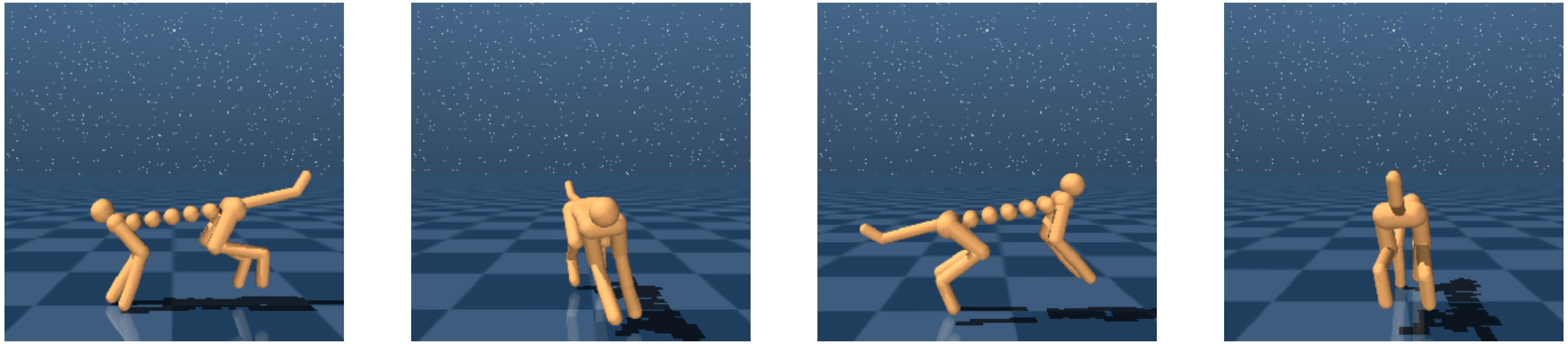
SET



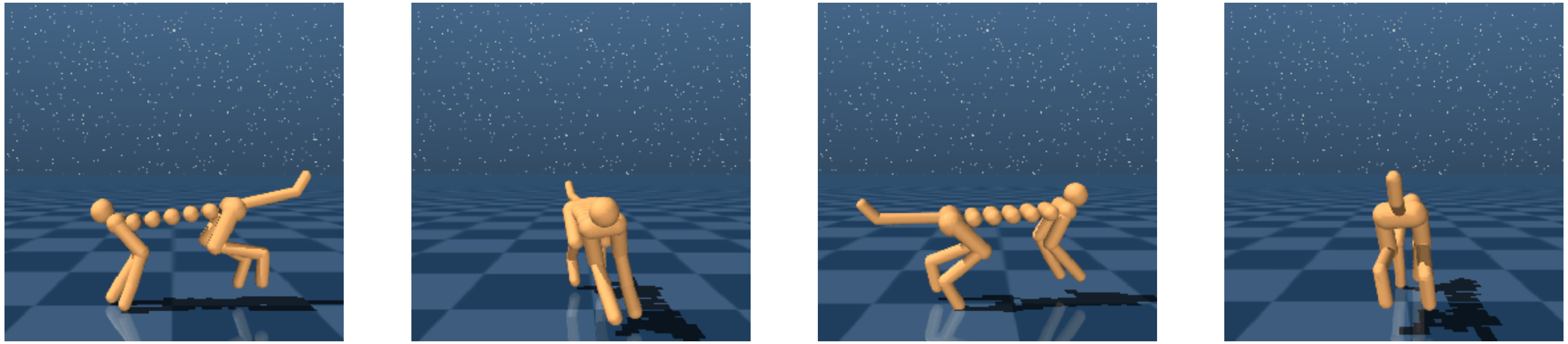
Experiments: Orientation Generalization

Methods	500k training steps 0°	90°	180°	270°	random	1M training steps 0°	90°	180°	270°	random
SWAT	1886.1 ± 148.9	1005.5 ± 615.3	$\mathbf{120.5} \pm 178.5$	791.0 ± 493.4	1232.3 ± 72.9	2592.6 ± 155.6	1340.2 ± 668.0	$\mathbf{-5.6} \pm 8.5$	1193.5 ± 345.2	1178.6 ± 674.9
SET	1587.4 ± 411.3	1695.6 ± 278.4	$\mathbf{1659.9} \pm 110.2$	1388.3 ± 173.8	1465.2 ± 161.0	4622.0 ± 292.8	4799.5 ± 172.9	$\mathbf{4756.3} \pm 103.4$	4899.8 ± 139.7	4902.8 ± 62.9

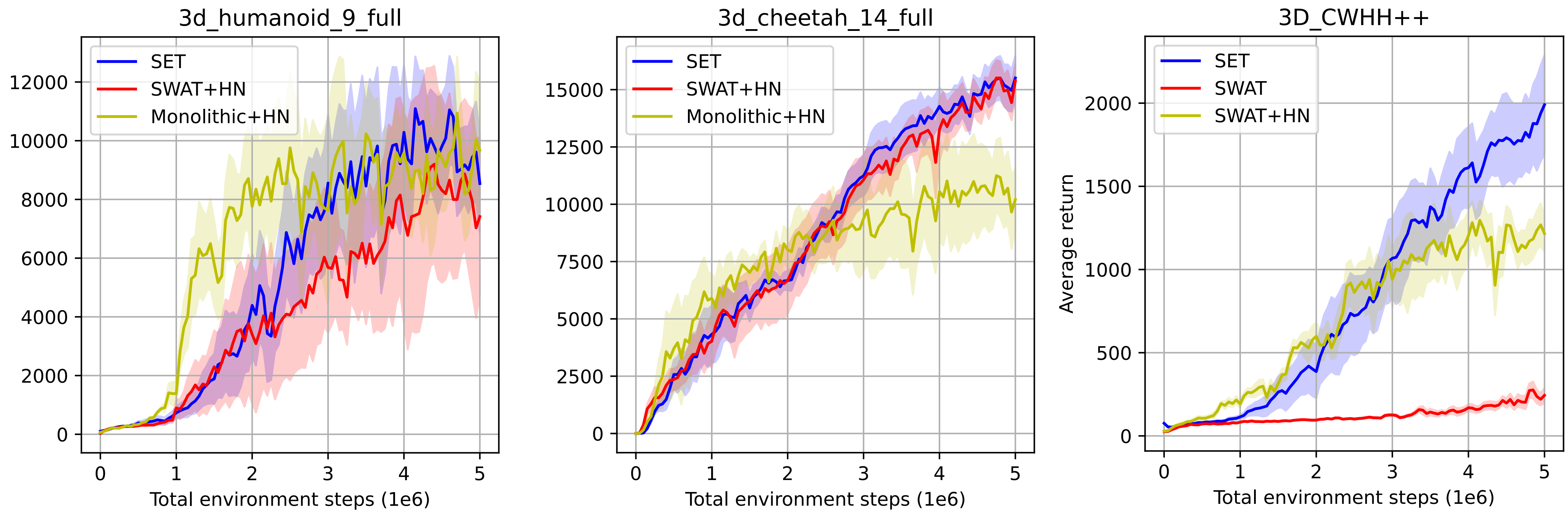
SWAT



SET



Experiments: Comparison with Heading Normalization (HN)



Take-away

- We introduce 3D-SGRL, a set of more practical yet highly challenging benchmarks for morphology-agnostic RL, where the agents are permitted to turn and move in the 3D environments with arbitrary starting configurations and arbitrary target directions.

- To effectively optimize the policy on such challenging benchmarks, we propose to enforce the policy network with geometric symmetry. We introduce a novel architecture dubbed SET that captures the rotation/translation equivariance particularly when external force fields like gravity exist in the environment.

- We verify the performance of the proposed method on the proposed 3D benchmarks, where it outperforms existing morphology-agnostic RL approaches by a significant margin in various scenarios.

Thanks!

For more information, welcome to visit our website:

https://alpc91.github.io/SGRL/