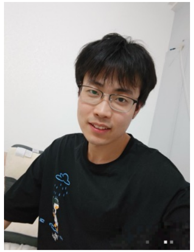
BUILDING GENERALIST ROBOT AUTONOMY IN THE WILD

Skill-aware Mutual Information Optimisation for Generalisation in Reinforcement Learning

Xuehui Yu
Harbin Institute of Technology
University of Edinburgh

Mhairi Dunion
University of Edinburgh



Xin Li
Harbin Institute of Technology

Stefano V. Albrecht
University of Edinburgh

THE UNIVERSITY *of* EDINBURGH
informatics

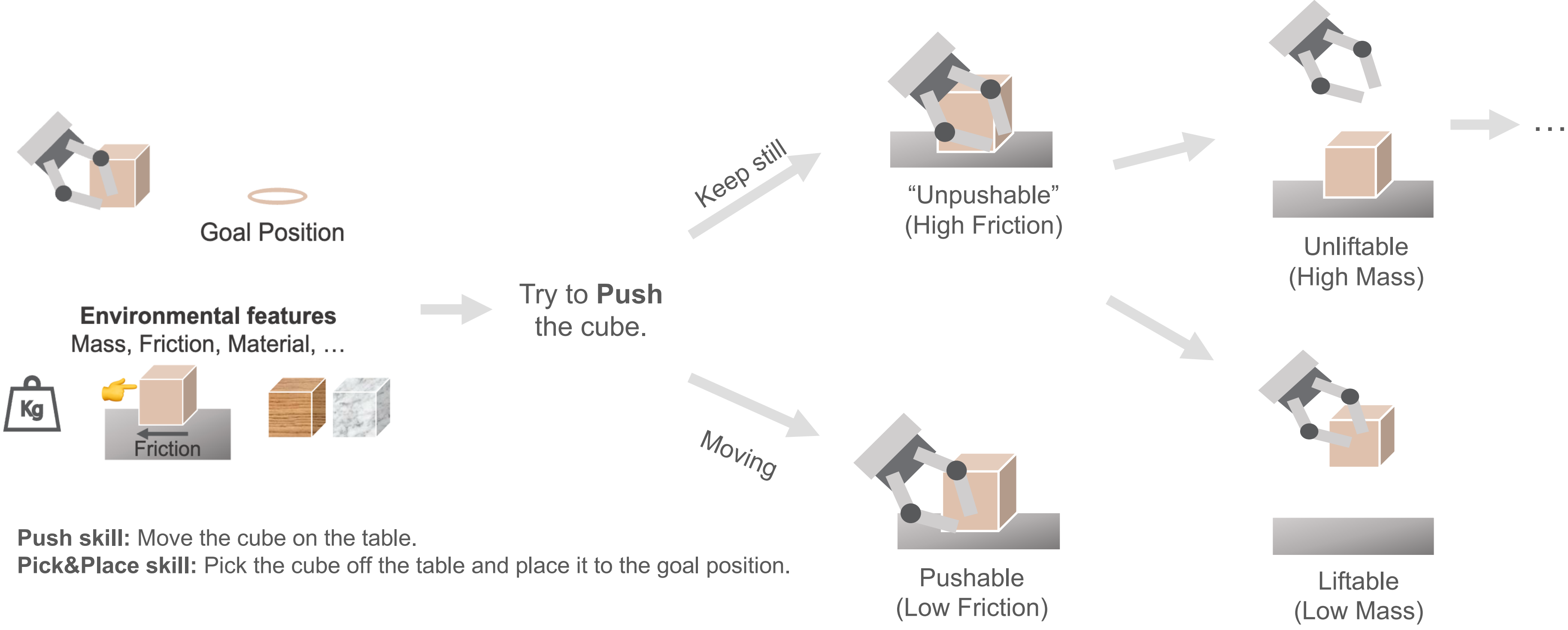
Autonomous Agents
Research Group

哈尔滨工业大学 计算学部
Faculty Of Computing, Harbin Institute Of Technology

语言技术研究中心·网络智能研究室
Web Intelligence Laboratory,Language Technology Research Center

MOTIVATION

- Meta-Reinforcement Learning (Meta-RL) agents can struggle to generalise across tasks with varying environmental features that require different optimal skills (i.e., different modes of behaviors).

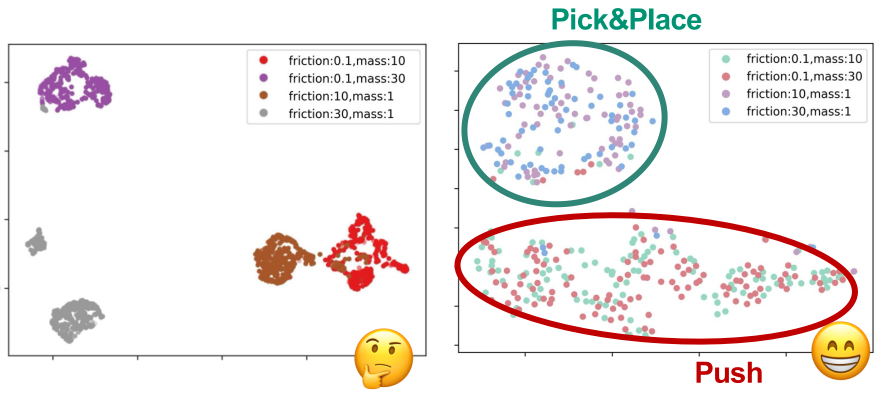


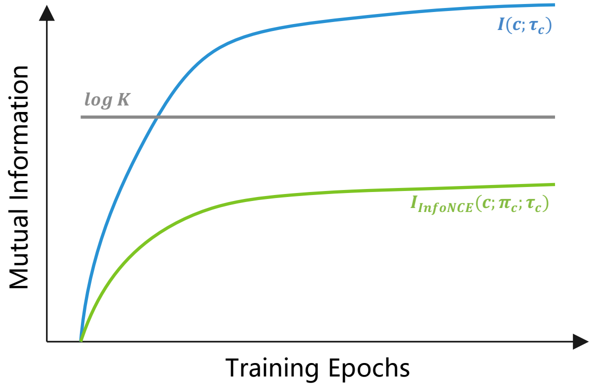
Push skill: Move the cube on the table.
Pick&Place skill: Pick the cube off the table and place it to the goal position.

CONTRASTIVE LEARNING

Integrating contrastive learning with Meta-RL brings significant advances, but:

- **Issue (i):** Existing context encoders based on contrastive learning do not distinguish tasks that require different skills.
- **Issue (ii):** Existing K-sample MI estimators, such as InfoNCE, are sensitive to the sample size K (i.e., the $\log$-K curse).





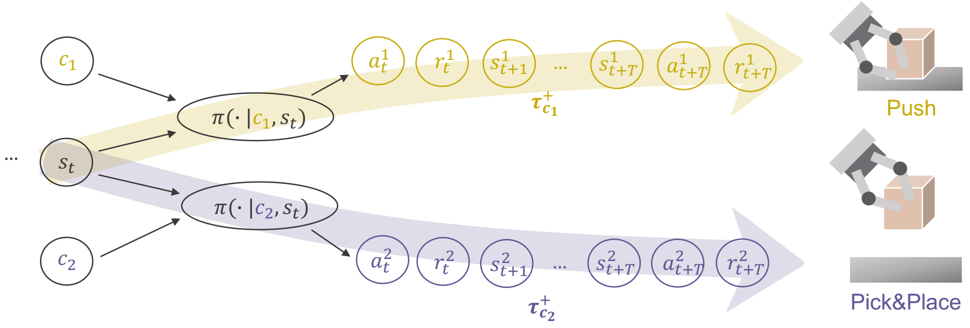
SKILL-AWARE MUTUAL INFORMATION

Step 1: An objective for a context encoder -- Skill-aware Mutual Information (SaMI):

SaMI is a generalised form of MI objective between context embeddings, skills, and trajectories:

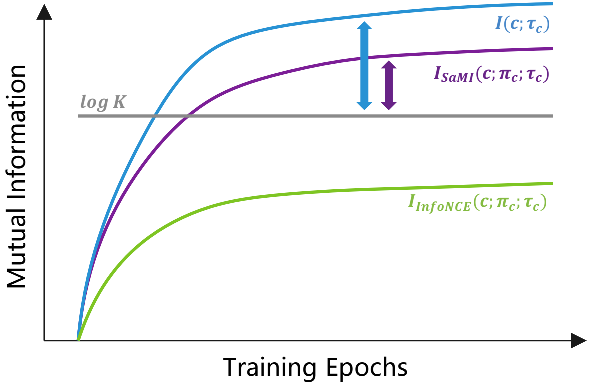
$$I_{SaMI}(c;\pi_c;\tau_c) = I(c;\tau_c) - I(c;\tau_c|\pi_c) \leq I(c;\tau_c)$$

✨ **Compress skill-related information** from trajectories



A policy π conditioned on a fixed context embedding c is defined as a skill $\pi(\cdot|c)$ (shortened as π_c).

✨ Smaller and easier to optimise



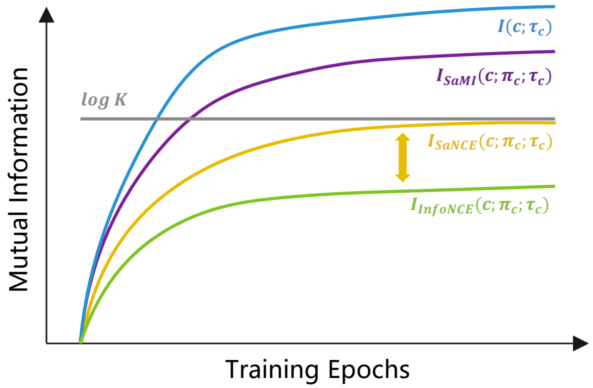
SKILL-AWARE NOISE CONTRASTIVE ESTIMATION

Step 2: A K-sample estimator for I_{SaMI} -- Skill-aware Noise Contrastive Estimation (SaNCE):

$$I_{SaNCE}(c;\pi_c;\tau_c|\psi,K)$$
$$= \mathbb{E}_{p(c_1,\pi_{c_1},\tau^+_{c_1})p(\tau^-_{c_1,2:K})}\left[\log\left(\frac{K\cdot f_\psi(c_1,\pi_{c_1},\tau^+_{c_1})}{f_\psi(c_1,\pi_{c_1},\tau^+_{c_1})+\sum_{k=2}^{K}f_\psi(c_1,\pi_{c_1},\tau^-_{c_1,k})}\right)\right]$$
$$\leq I_{SaMI}(c;\pi_c;\tau_c)$$

How to sample positive/negative samples c_1, π_{c_1}, τ_{c_1}?

✨ With the same training epochs, I_{SaNCE} is **closer to I_{SaMI}** compared to $I_{InfoNCE}$.

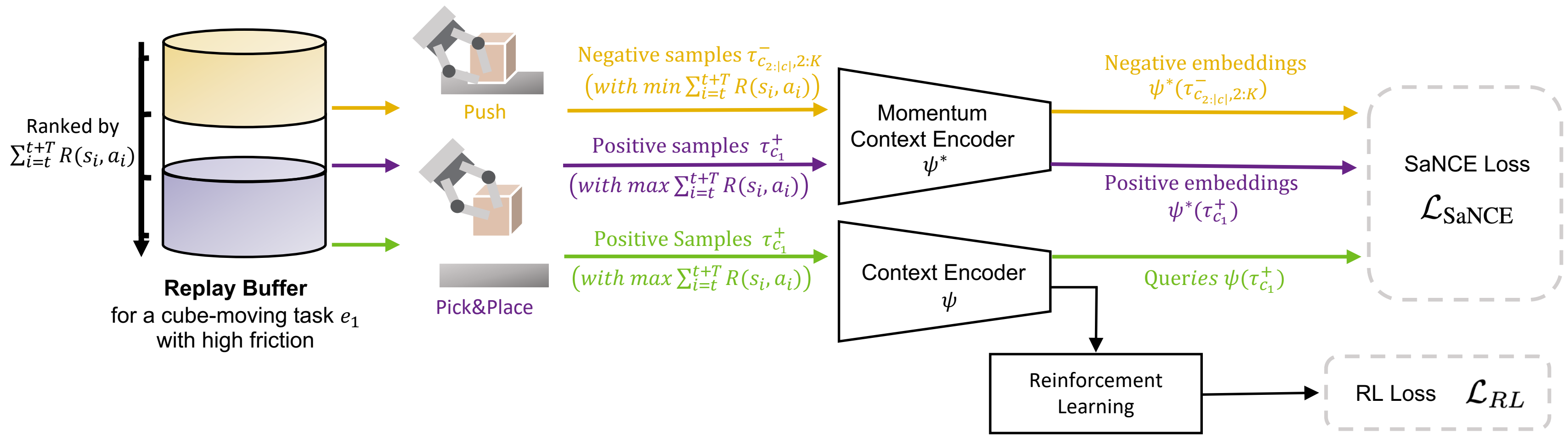


SKILL-AWARE TRAJECTORY SAMPLING STRATEGY

Step 3: Skill-aware trajectory sampling strategy

positive skills π_c^+ are defined as optimal skills achieving highest return;

negative skills π_c^- are those that result in lower returns.



A practical framework for using SaNCE in the meta-training phase.

SKILL-AWARE TRAJECTORY SAMPLING STRATEGY

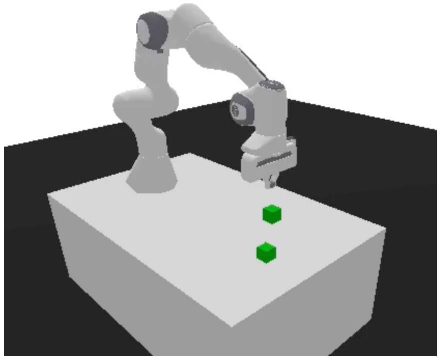
Step 3: Skill-aware trajectory sampling strategy

EXPERIMENT SETUP

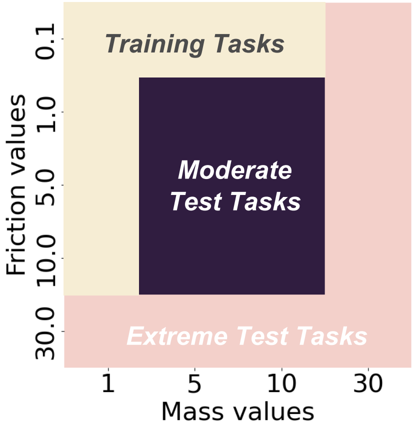
Zero-shot generalisation:

Moderate test tasks: interpolation

Extreme test tasks: extrapolation (unseen mass/friction values)

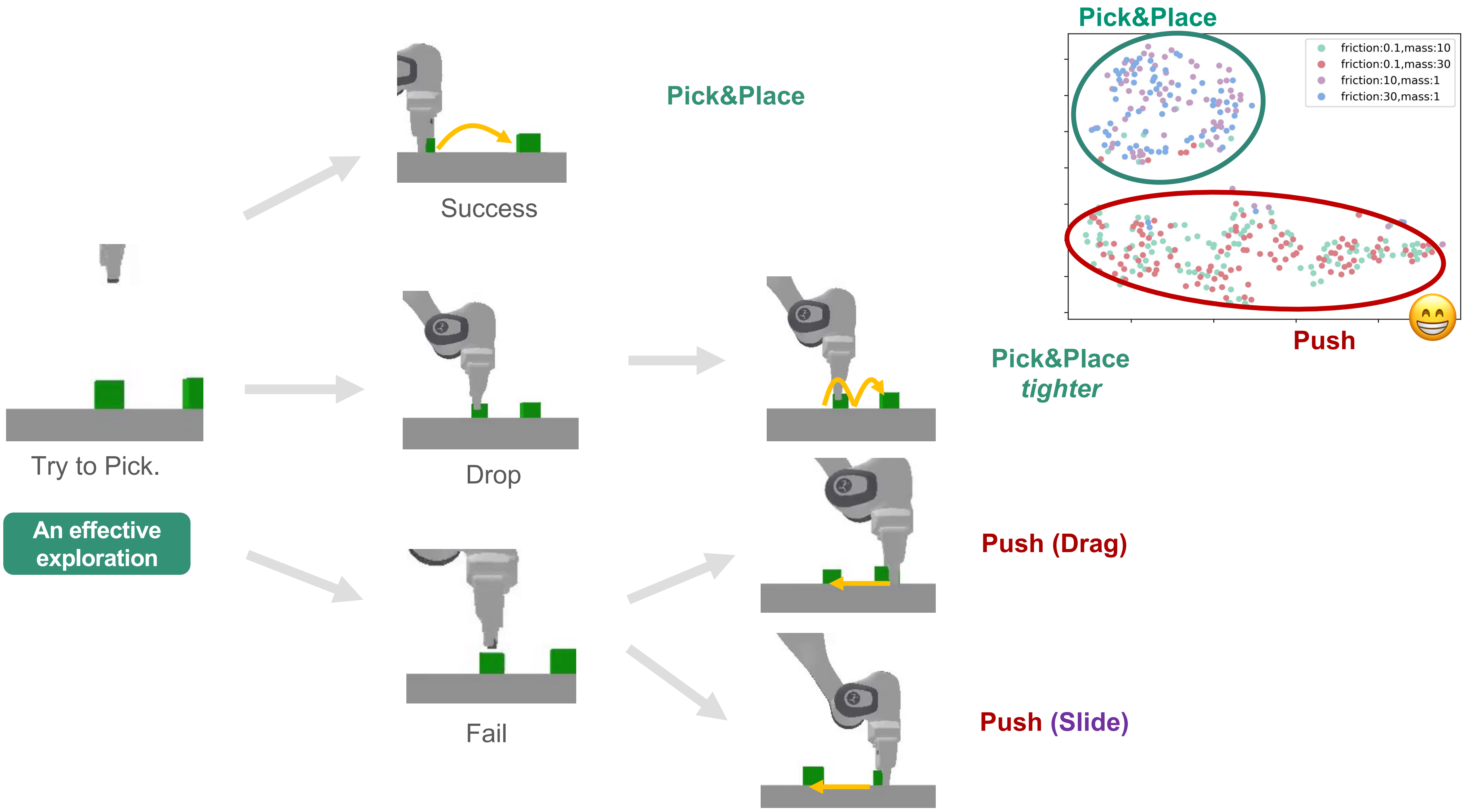


(a) Panda-gym

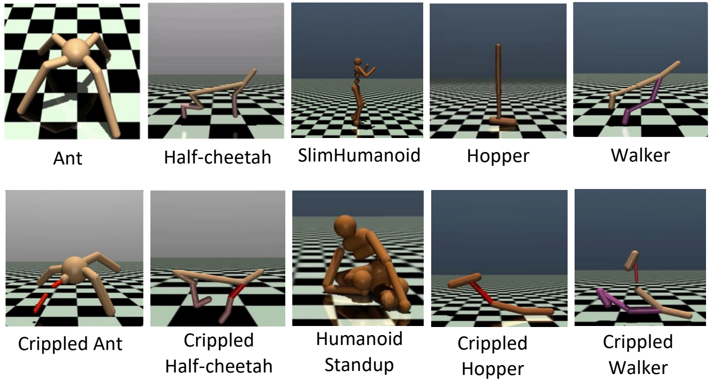


(b) Task setting

Friction
Friction = 5.0
Friction = 1.0
Friction = 0.1
Step 1: Clamp, and try to Pick
Clamp, Pick&Place
Clamp, Drop, Pick&Place tighter
Clamp, Pick&Place
Clamp, Drop, Pick&Place tighter
Clamp, Fail, Push (Drag)
Clamp, Pick&Place
Clamp, Fail, Push (Slide)
Clamp, Fail, Push (Slide)
Mass=1.0
Mass=10.0
Mass=30.0
Mass



Pick&Place
friction:0.1,mass:10
friction:0.1,mass:30
friction:10,mass:1
friction:30,mass:1
Push
Success
Pick&Place
Try to Pick.
An effective exploration
Drop
Pick&Place tighter
Fail
Push (Drag)
Push (Slide)

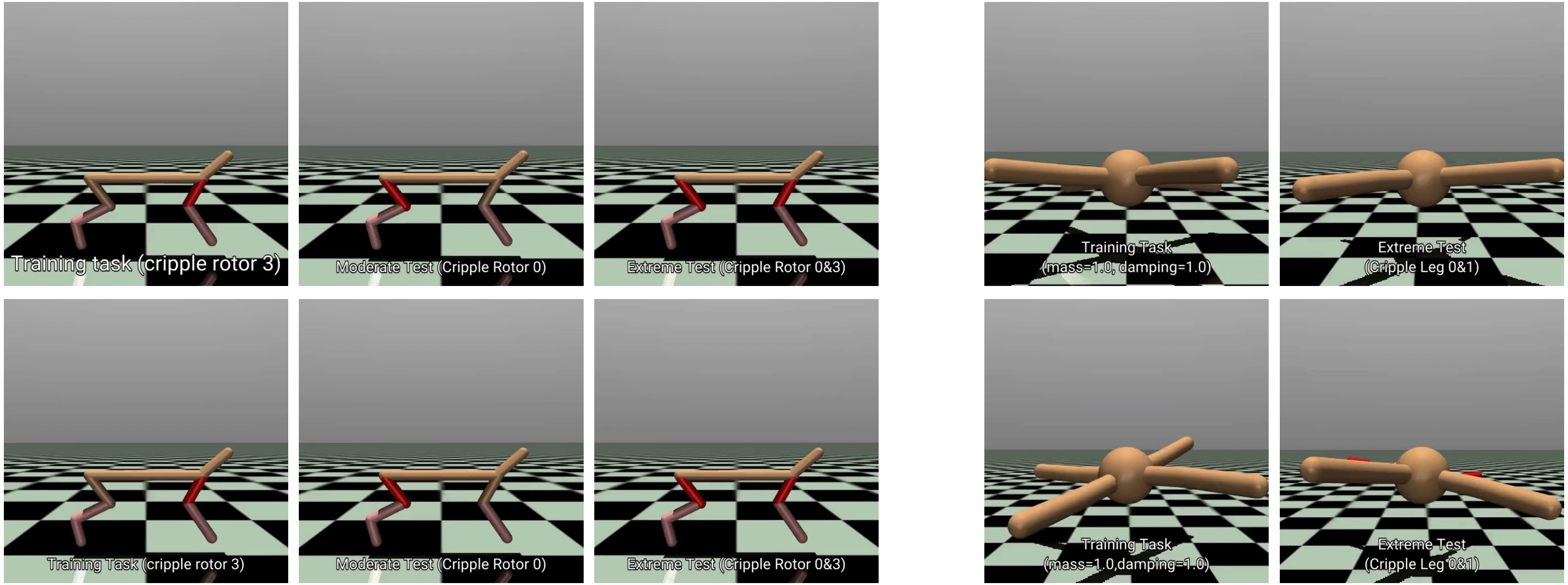


CODE

Our code, video demos and experimental data.

SaCCM (ours)

CCM



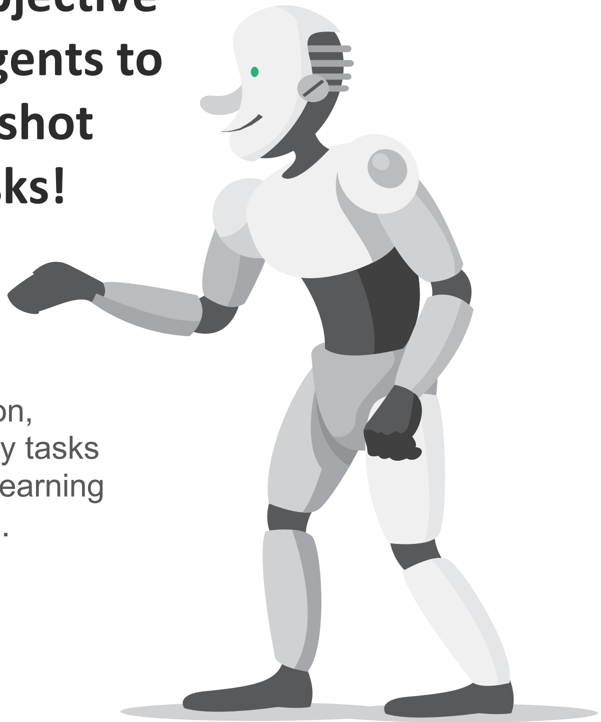
arXiv PAPER

CODE

Our code, video demos
and experimental data.

BENCHMARK

Use our SaMI learning objective to incentivise Meta-RL agents to be versatile and zero-shot generalise across tasks!



Without any prior skill distribution,
a set of skills applicable to many tasks
emerges solely from the SaMI learning
objective and the data provided.

NeurIPS 2024

👋 **Thu 12 Dec 4:30 p.m. PST -- Poster Session 4**