ICML 2026 · Accepted

SPHERE

Mitigating the Loss of Spectral Plasticity in Mixture-of-Experts for Deep Reinforcement Learning

Lirui Luo[1,2], Guoxi Zhang[2], Hongming Xu[2], Cong Fang[1], Qing Li[2]

[1] Peking University · [2] Beijing Institute for General Artificial Intelligence (BIGAI)

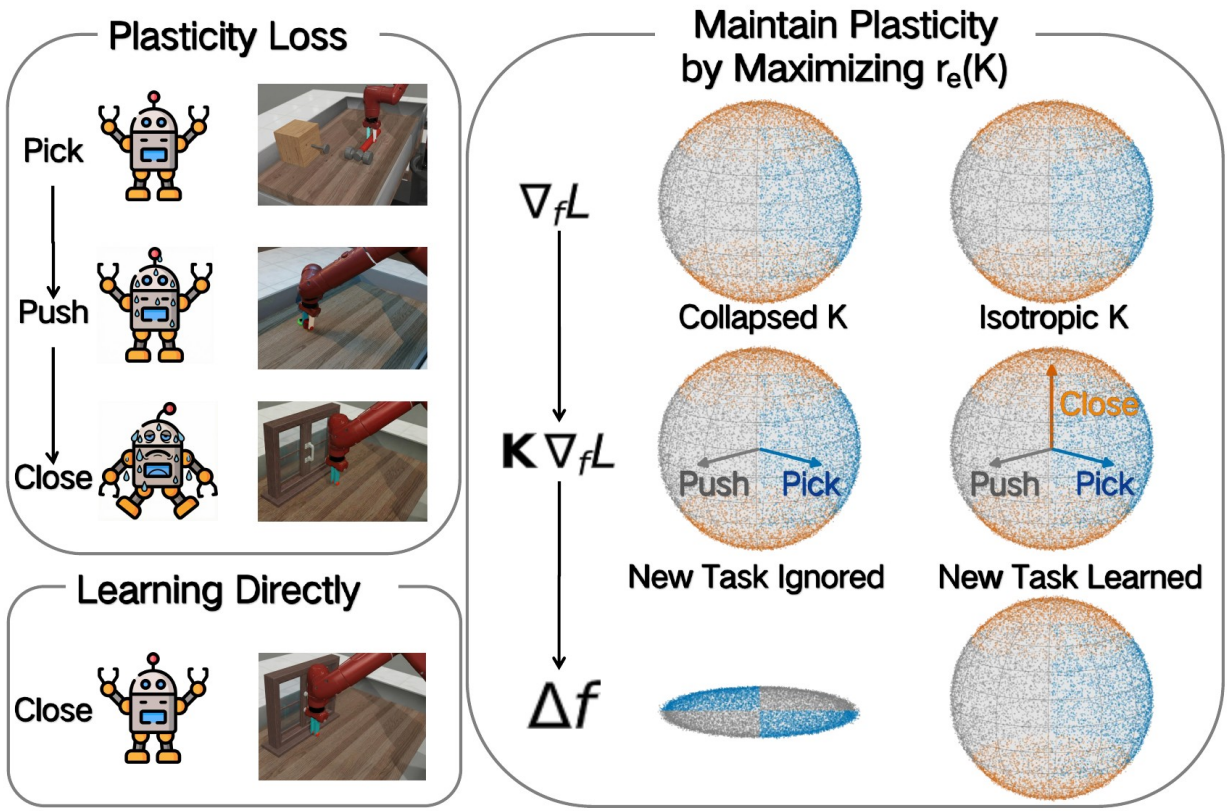
sphere-rl.github.io

github.com/sphere-rl/sphere

北京通用人工智能研究院
Beijing Institute for General Artificial Intelligence

Motivation

Continual RL exposes a different kind of failure: the policy can keep receiving data but stop adapting.



What goes wrong?

Later-task updates become low-dimensional, so new skills are ignored even when they are learnable.

Why MoE matters

MoE policies scale control capacity, but sparse routing does not automatically prevent continual plasticity loss.

SPHERE's entry point

Diagnose the eNTK spectrum, then regularize weighted expert features as a tractable proxy.

Plasticity Loss as Loss of Spectral Plasticity

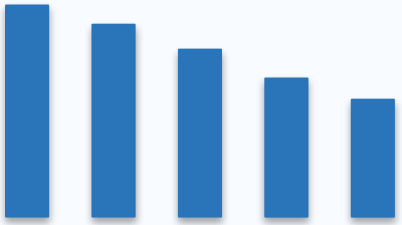
Functional update geometry

$$\Delta f = -\eta\, K\, \nabla f\, L$$

Collapsed spectrum

Few usable update directions

Higher spectral plasticity



Diverse functional directions

Key idea

- Measure plasticity by the spectral-entropy effective rank of the empirical NTK.
- Low rank means updates concentrate in a few directions.
- A practical proxy avoids forming the full eNTK.

Method

SPHERE regularizes the expert features that control the MoE update geometry.

1. Route

Top-K MoE selects a small expert set for each state.

2. Weight features

Concatenate routing-weighted expert features into A_last.

3. Regularize

Apply a practical Parseval penalty so the Gram stays well-conditioned.

What makes it practical

- Acts on forward-pass feature matrices, not the full eNTK.
- Targets the last hidden expert layer by default.
- Keeps the penalty local to the MoE representation used by the policy.

Evaluation Setup

Two continuous-control suites test whether policies keep learning across a task stream.

MetaWorld CW10

10 manipulation tasks

1M steps per task

5 seeds

- Shared observation/action spaces
- Continual World task subset
- Report average final success under RL vs CRL

HumanoidBench H1

5 humanoid tasks

10M steps per task

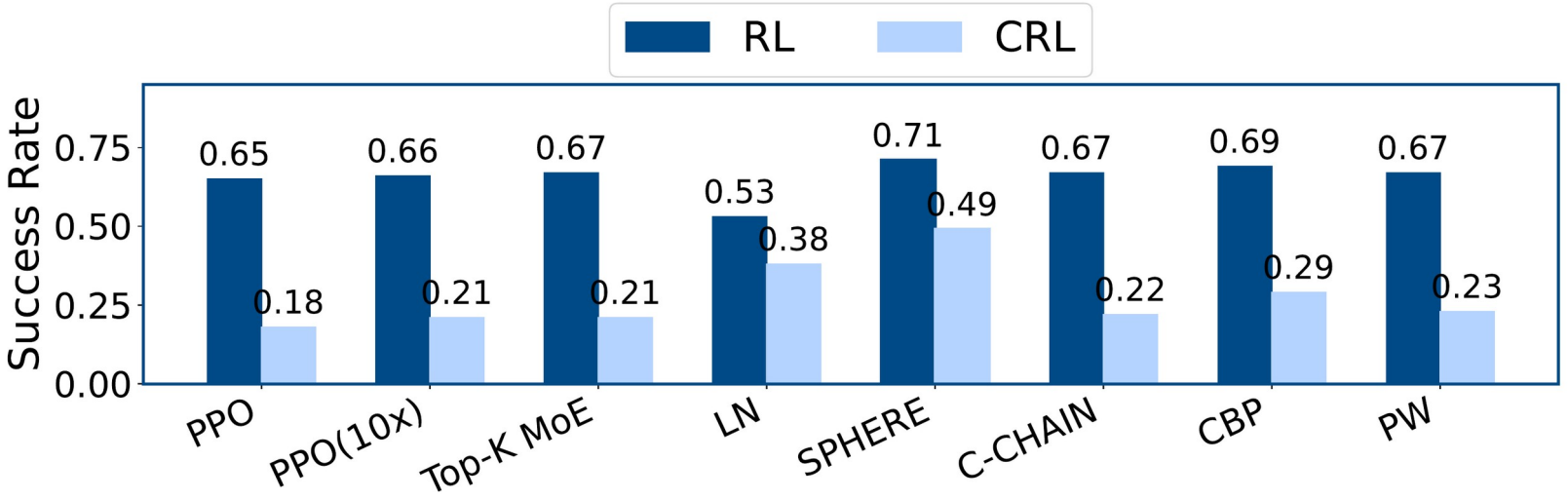
5 seeds

- Stand, Walk, Pole, Slide, Run
- Same H1 observation/action spaces
- Harder long-horizon continual-control sequence

Comparison target: unregularized Top-K MoE plus dense, widened, and continual-learning mitigation baselines.

MetaWorld Result

SPHERE narrows the gap between per-task RL and continual RL.



+133%

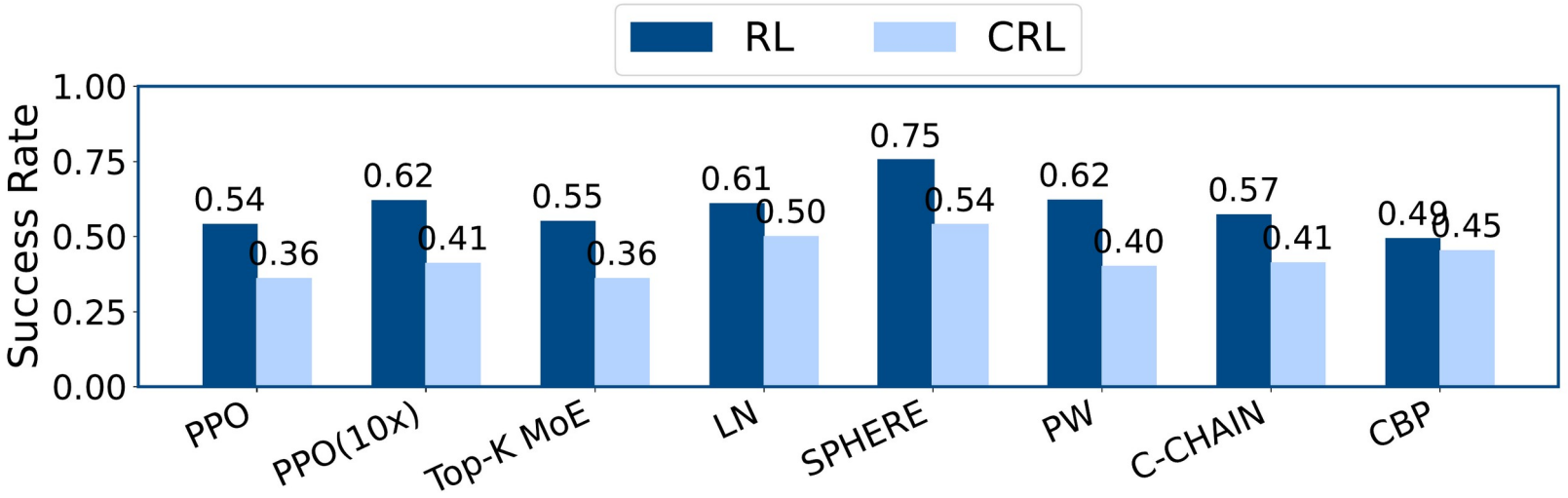
CRL average success vs. Top-K MoE

RL–CRL gap reduced by 52%

- The strongest gain appears exactly where plasticity loss matters: continual RL.
- SPHERE outperforms prior mitigation baselines while preserving the MoE policy form.
- The result supports the spectral-plasticity framing beyond a single benchmark.

HumanoidBench Result

The same spectral-plasticity intervention transfers to harder humanoid tasks.



+50%

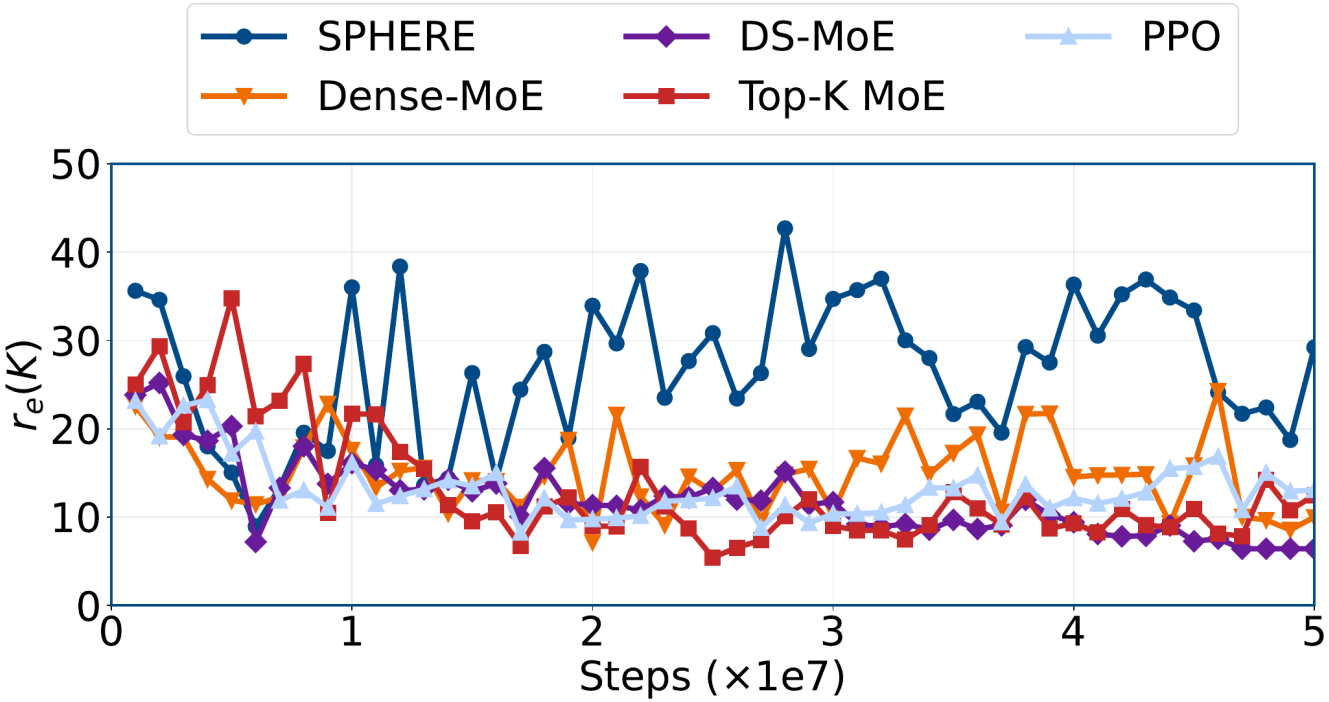
CRL average success vs. Top-K MoE

Also improves single-task RL by 36%

- Long-horizon humanoid training makes within-task and across-task plasticity loss more visible.
- SPHERE improves the Top-K MoE actor against dense and continual-learning baselines.
- The paper reports five H1 tasks: Stand, Walk, Pole, Slide, and Run.

Mechanism Evidence

Performance gains align with sustained spectral plasticity.



Readout

- Baselines' effective rank decays during CRL.
- SPHERE keeps r_e(K) higher throughout training.
- Higher rank means a less collapsed functional update space.

This is not just a performance table: it links the observed degradation to the paper's proposed update-geometry diagnostic.

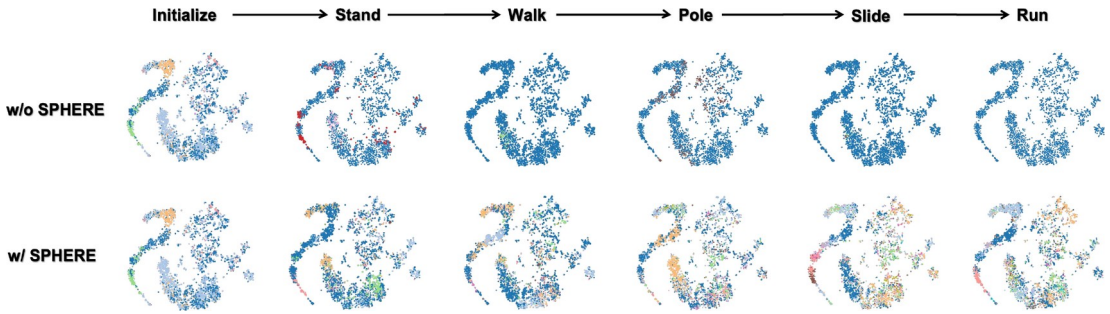
Design and Proxy Checks

The best version is the routing-weighted, cross-expert last-layer feature Gram.

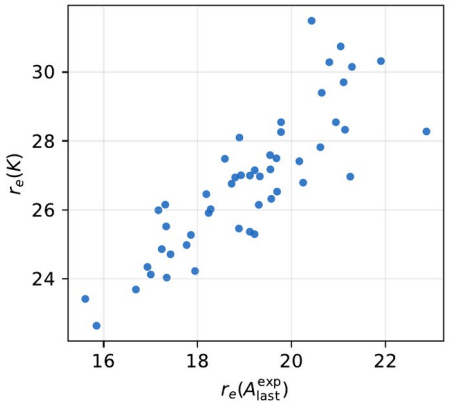
HumanoidBench CRL ablation

Variant	Avg. success
w/o SPHERE	0.36 ± 0.08
w/ SPHERE	**0.54 ± 0.12**
All hidden expert layers	0.42 ± 0.07
Per-expert loss sum	0.40 ± 0.08
Gradient-factor regularization	0.43 ± 0.09

Alternatives help, but none matches the full routing-weighted SPHERE setup.



Qualitative feature-space view



Proxy check

Weighted expert features track the eNTK rank trend, supporting the practical regularizer.

Takeaways

Framing Plasticity loss in MoE RL can be understood as spectral collapse in functional update directions.

Method SPHERE turns an intractable eNTK target into a practical Parseval penalty on weighted expert features.

Evidence Across MetaWorld and HumanoidBench, SPHERE improves continual-control success and maintains higher spectral plasticity.

Project: sphere-rl.github.io · Code: github.com/sphere-rl/sphere