HyperMARL: Adaptive Hypernetworks for Multi-Agent RL

THE UNIVERSITY of EDINBURGH

TEXAS The University of Texas at Austin

Kale-ab Abebe Tessera, Arrasy Rahman, Amos Storkey, Stefano V. Albrecht

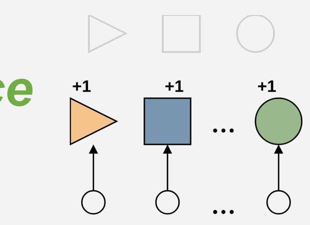
*Can we design a **parameter-shared MARL architecture** that flexibly supports both **specialised** and **homogeneous** behaviours—without modifying the **learning objective**, **preset diversity levels** or **sequential updates**?*

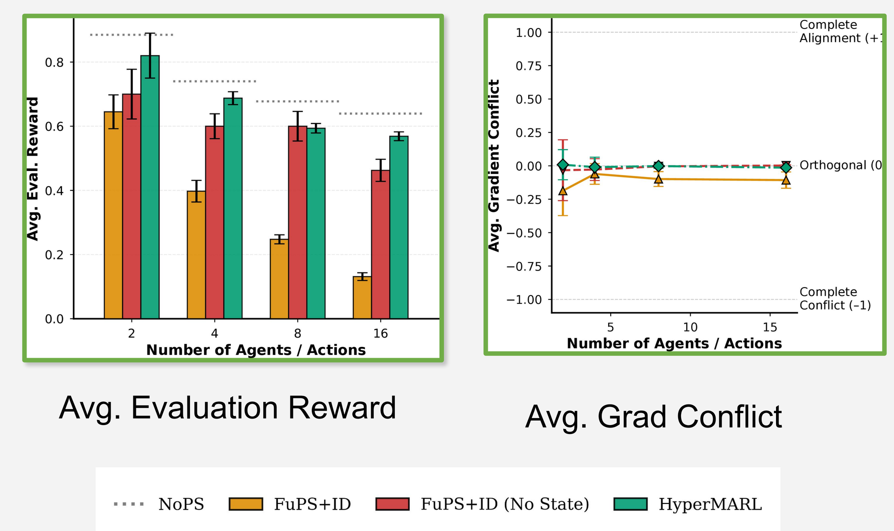
1) Adaptability

- ***Adaptability***, being able to learn ***specialised*** or ***homogeneous*** behaviours, is important in multi-agent systems.
- Efficient adaptability is critical in MARL – parameter sharing (FuPS: full PS) vs independent parameters per agent (NoPS).
- **Problem: *FuPS*** methods are efficient, but struggle with **specialisation**.

2) Challenges: Parameter Sharing + Specialisation

Coupling agent-IDs and **observations** = higher ***gradient interference***

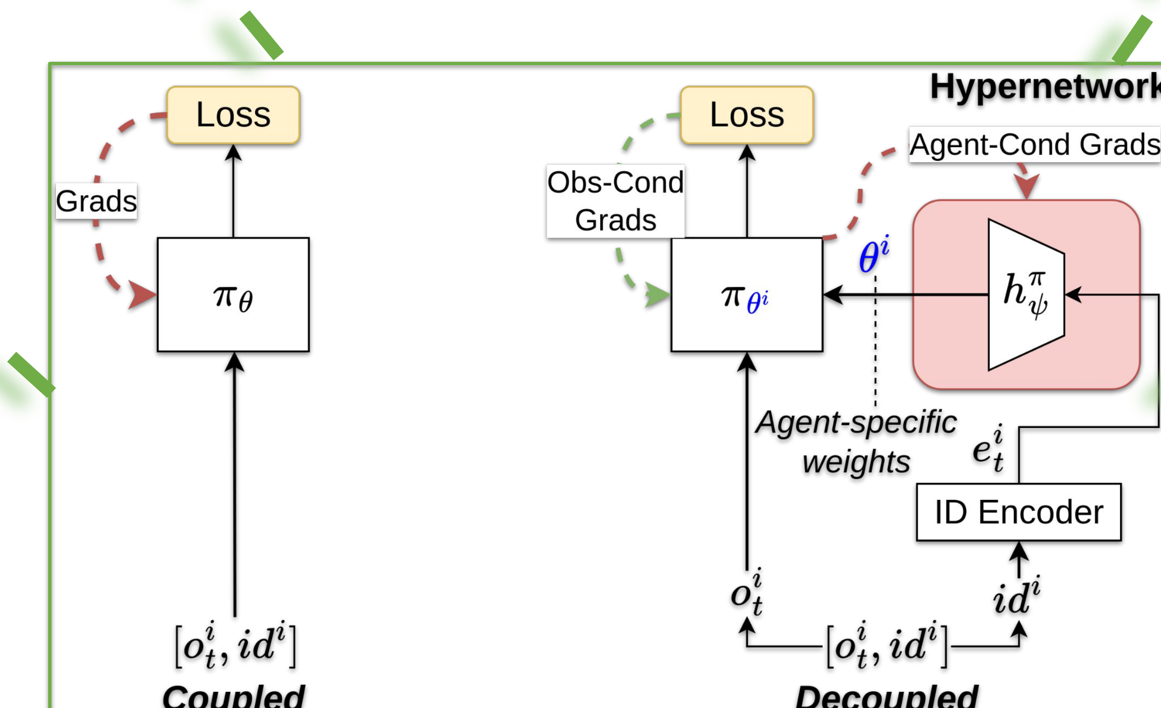




Avg. Evaluation Reward　　Avg. Grad Conflict

- Related methods – *altered objectives*, *preset diversity levels*, or sequential updates.
- Can shared policies adapt ***without these complexities***? Can we ***decouple*** agent IDs and observations?

HyperMARL



3) HyperMARL

★ Uses ***agent-conditioned* hypernetworks** to adaptively learn diverse, homogeneous or mixed behaviours.

★ Enabling specialisation through the decoupling of ***observation-*** and ***agent-conditioned*** grads.

★ **Input**: Agent ID or learned embedding.

★ **Output**: per-agent policy and critic weights.

4) Gradient Decoupling

$$\nabla_\psi J(\psi) = \sum_{i=1}^{I} \underbrace{\nabla_\psi h_\psi^\pi(e^i)}_{\mathbf{J}_i \text{ (agent-conditioned)}} \underbrace{\mathbb{E}_{\mathbf{h}_t,\mathbf{a}_t\sim\pi}\left[A(\mathbf{h}_t,\mathbf{a}_t)\nabla_{\theta^i}\log\pi_{\theta^i}(a_t^i \mid h_t^i)\right]}_{Z_i \text{ (observation-conditioned)}}.$$

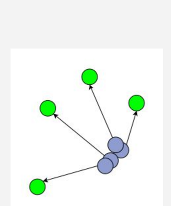
❖ ***Agent-conditioned:*** deterministic w.r.t to mini-batch samples, separating agent identity from traj noise.

❖ ***Observation-conditioned***: averages trajectory noise **per agent**.

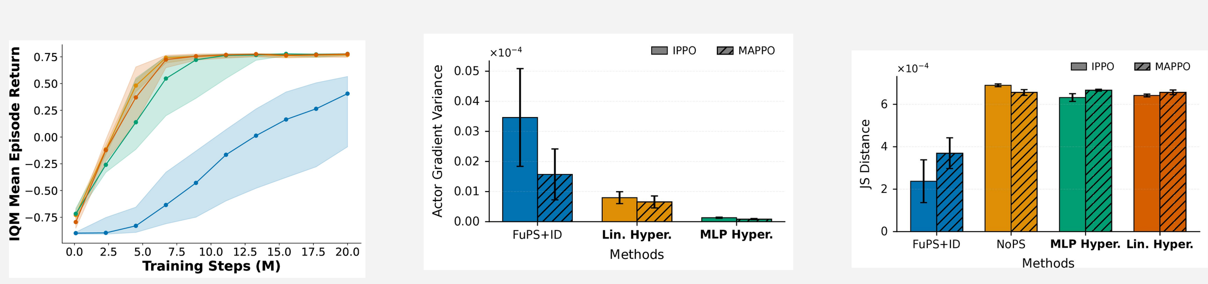
5) Results

5.1 Competitive to NoPS and Baselines in Specialised Tasks

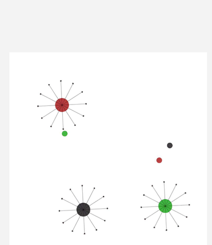
➢ Learns **diverse** behaviours with a **lower policy variance**.



Dispersion (Bettini et al., 2022)



➢ **Adapts** to specialised, homogeneous and mixed tasks.

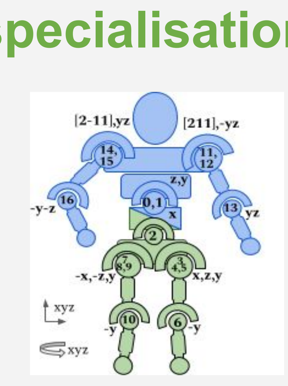


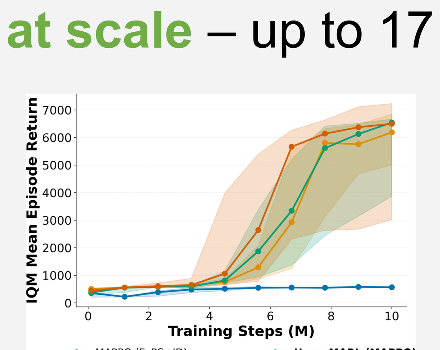
Navigation (Bettini et al., 2022)

a) Different Goals　b) Same Goals　c) Mixed Goals

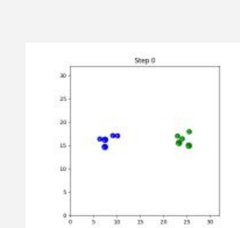
➢ Handles **specialisation at scale** – up to 17 agents.

MAMuJoCo (Peng et al., 2021)

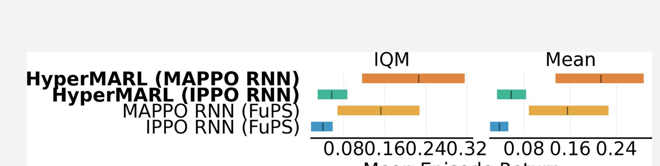




5.2 Competitive in homogeneous scenarios



SMAX (Rutherford et al., 2024)

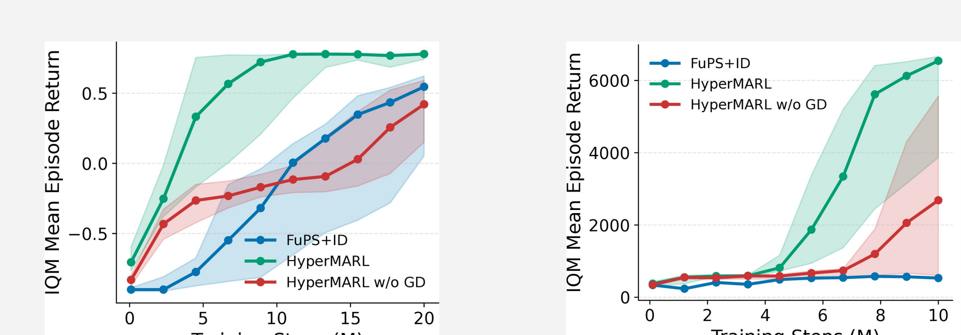


Smax v2 20 agents.

Many more results in the paper (e.g. off-policy, ***22 scenarios, and 6 baselines***).

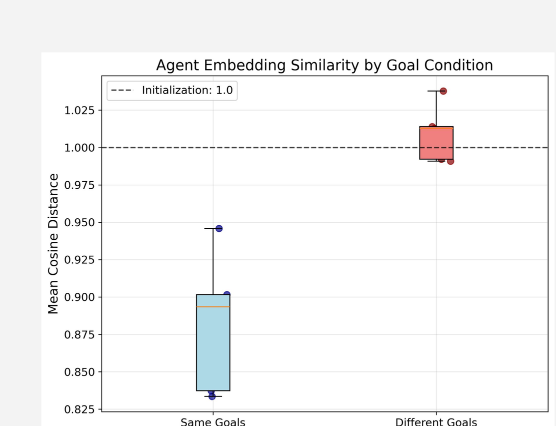
6) Ablations + Embeddings

★ ***Gradient decoupling matters***



Dispersion　Humanoid MAMuJoCo

★ ***Agent embeddings** move **closer** when task requires **similar** behaviour and **further apart** when task requires **different** behaviour.*



7) Main Takeaways

- ***Gradient decoupling*** is a key ingredient in mitigating cross-agent interference -- can be achieved through agent-conditioned hypernets.
- Using ***HyperMARL***, we showed it is possible to learn ***adaptive*** behaviour, across diverse tasks, without altering the *learning objective, preset diversity levels or sequential updates*.

University of Edinburgh, University of Texas at Austin