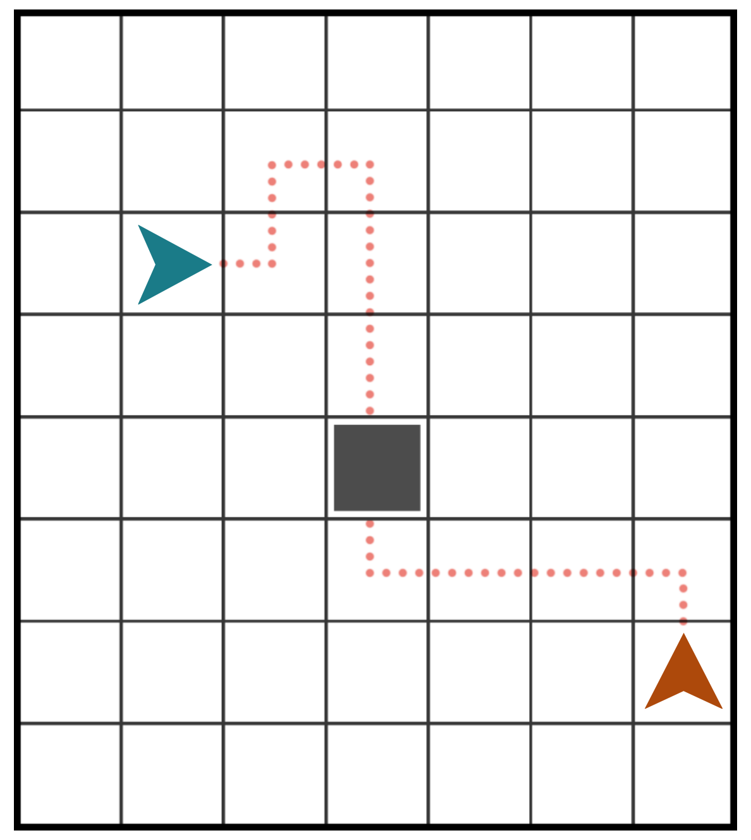
Shared Experience Actor-Critic for Multi-Agent Reinforcement Learning

Filippos Christianos, Lukas Schäfer, Stefano V. Albrecht

Summary

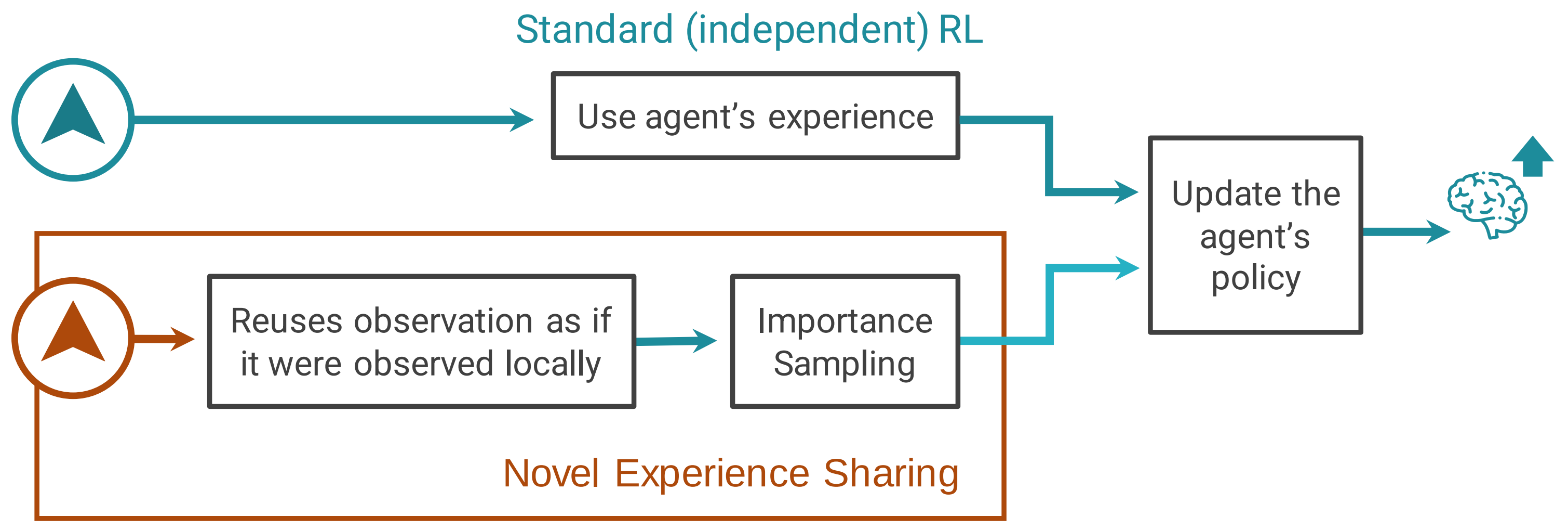
- **Problem:** Multi-agent reinforcement learning (MARL) with sparse rewards
- **Contribution:** We propose a novel experience sharing method (Shared Experience Actor-Critic or SEAC) that combines gradients of multiple agents to share experience between agents.
- **Evaluation:**
    - Evaluated in four sparse-reward multi-agent environments
    - Consistently outperforms baselines and three state-of-the-art MARL algorithms (MADDPG, QMIX, ROMA)
    - SEAC learns in fewer steps and converges to higher returns
    - In harder tasks, sharing experience makes the difference between not learning at all and learning

Motivation and Idea



- Consider the simple multi-agent game: two agents must simultaneously arrive at the goal
- This presents a difficult, sparsely rewarded exploration problem
- When agents finally succeed, the idea of sharing experience is appealing: both agents could learn to approach the goal from two different directions from a single, successful episode

Idea: Make use of both agents' exploration by combining their training gradients and correcting for off-policy



Standard (independent) RL

Use agent's experience

Reuses observation as if it were observed locally

Importance Sampling

Update the agent's policy

Novel Experience Sharing

Methodology

Policy Gradient Actor Loss:

Standard Actor Loss of agent i

$$\mathcal{L}(\phi_i) = \boxed{-\log \pi\left(a_t^i \middle| o_t^i; \phi_i\right) \left(r_t^i + \gamma V\left(o_{t+1}^i; \theta_i\right) - V\left(o_t^i; \theta_i\right)\right)}$$
$$\boxed{-\lambda \sum_{k \neq i} \frac{\pi(a_t^k | o_t^k; \phi_i)}{\pi(a_t^k | o_t^k; \phi_k)} \log \pi\left(a_t^k \middle| o_t^k; \phi_i\right) \left(r_t^k + \gamma V\left(o_{t+1}^k; \theta_i\right) - V\left(o_t^k; \theta_i\right)\right)}$$

Actor Loss using experience from other agents (k) with importance sampling correction

Value Function Critic Loss:

Starndard Value Loss of agent i

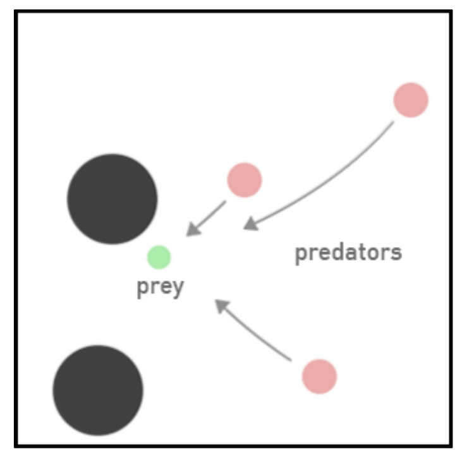
Value Loss using experience from other agents (k) with importance sampling correction

$$\mathcal{L}(\theta_i) = \boxed{\left\|V\left(o_t^i; \theta_i\right) - y_i^i\right\|^2} \boxed{+ \lambda \sum_{k \neq i} \frac{\pi(a_t^k | o_t^k; \phi_i)}{\pi(a_t^k | o_t^k; \phi_k)} \left\|V\left(o_t^k; \theta_i\right) - y_k^i\right\|^2}$$

$$y_k^i = r_t^k + \gamma V\left(o_{t+1}^k; \theta_i\right)$$

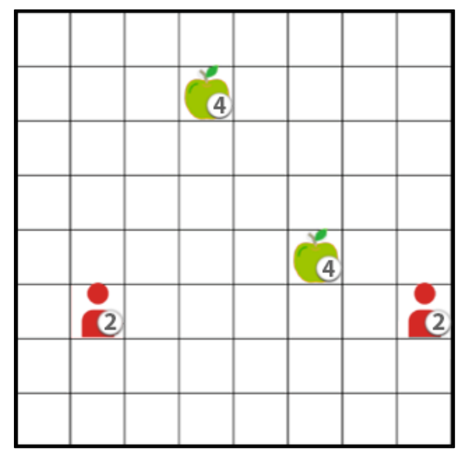
Experiments

We evaluate on ten sparsely rewarded tasks in four environments:

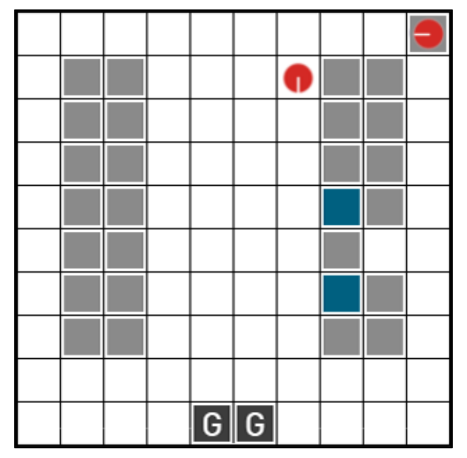


Predator Prey (sparse)

SMAC−3m (sparse)



Level-Based Foraging (LBF)



Multi-Robot Warehouse (RWARE)

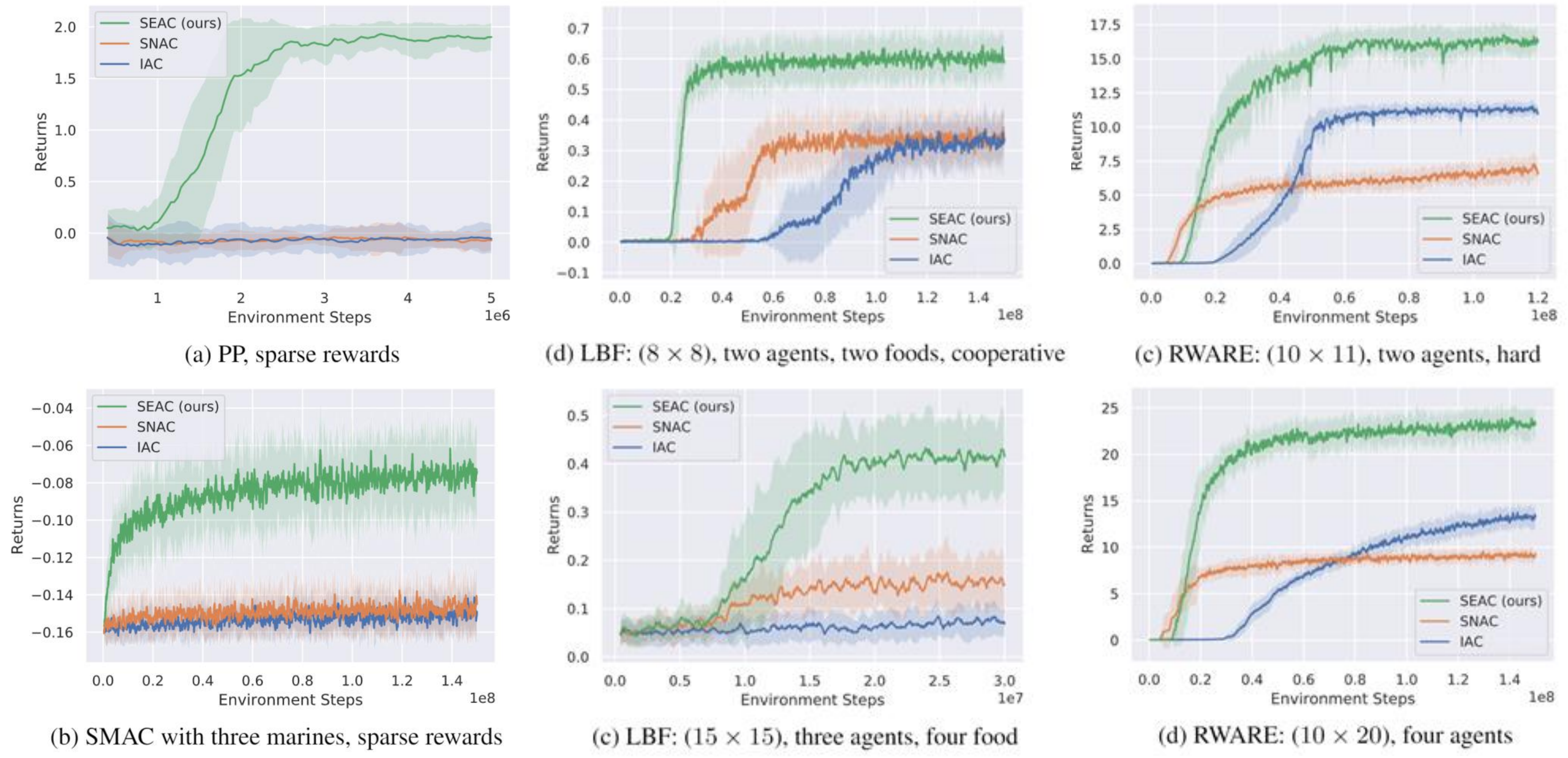
Baselines: (1) Independent Actor-Critic (IAC) and (2) Shared Network Actor-Critic (SNAC)

State-of-the-art MARL: (1) MADDPG, (2) QMIX and (3) ROMA

Contact
f.christianos@ed.ac.uk

SEAC RWARE LBF

Results



(a) PP, sparse rewards

(d) LBF: (8 × 8), two agents, two foods, cooperative

(c) RWARE: (10 × 11), two agents, hard

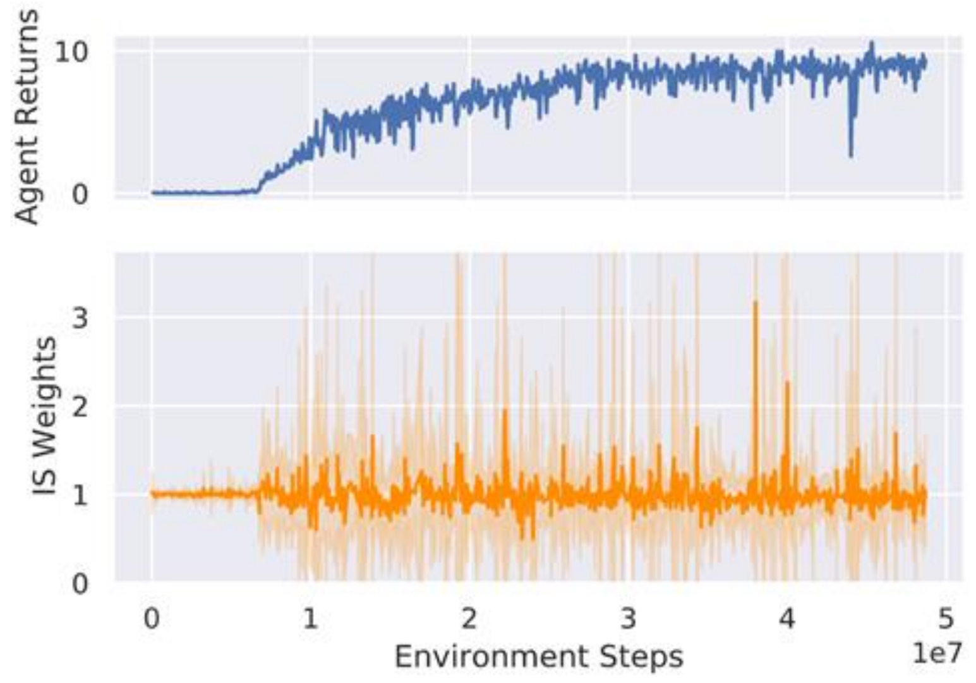
(b) SMAC with three marines, sparse rewards

(c) LBF: (15 × 15), three agents, four food

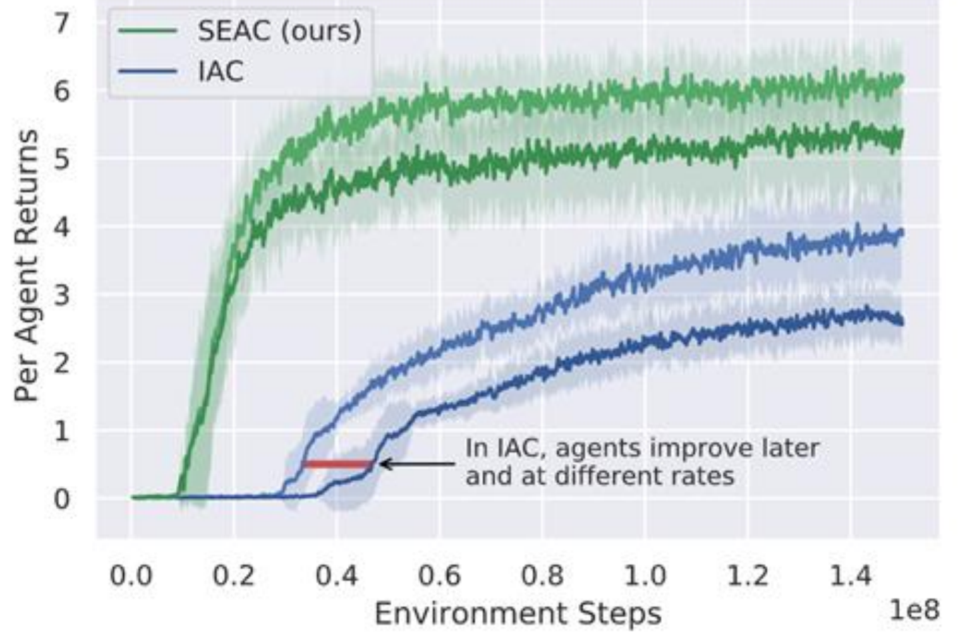
(d) RWARE: (10 × 20), four agents

	IAC	SNAC	SEAC (ours)	QMIX	MADDPG	ROMA
PP (sparse)	-0.04 ±0.13	-0.04 ±0.1	**1.93 ±0.13**	0.05 ±0.07	**2.04 ±0.08**	0.04 ±0.07
SMAC-3m (sparse)	-0.13 ±0.01	-0.14 ±0.02	**-0.03 ±0.03**	**0.00 ±0.00**	**-0.01 ±0.01**	**0.00 ±0.00**
LBF-(15x15)-3ag-4f	0.13 ±0.04	0.18 ±0.08	**0.43 ±0.09**	0.03 ±0.01	0.01 ±0.02	0.03 ±0.02
LBF-(8x8)-2ag-2f-coop	0.37 ±0.10	0.38 ±0.10	**0.64 ±0.08**	**0.79 ±0.31**	0.01 ±0.02	0.01 ±0.02
RWARE-(10x20)-4ag	13.75 ±1.26	9.53 ±0.83	**23.96 ±1.92**	0.00 ±0.00	0.00 ±0.00	0.00 ±0.00
RWARE-(10x11)-4ag	**40.10 ±5.60**	36.79 ±2.36	**45.11 ±2.90**	0.00 ±0.00	0.00 ±0.00	0.01 ±0.01

Analysis



Importance weights of one SEAC agent in RWARE, (10x11), two agents, hard



Best vs. Worst performing agents on RWARE, (10x20), four agents

Agents learn similar, but not identical policies which improves coordination

Agents learn simultaneously which helps in exploring promising joint actions more