Sparse Reward

Hung-yi Lee

Sparse Reward

Reward Shaping

Reward Shaping

Take “Play”,
$r_{t+1} = 1, r_{t+100} = -100$

Take “Study”,
~~$r_{t+1} = -1$~~, $r_{t+100} = 100$

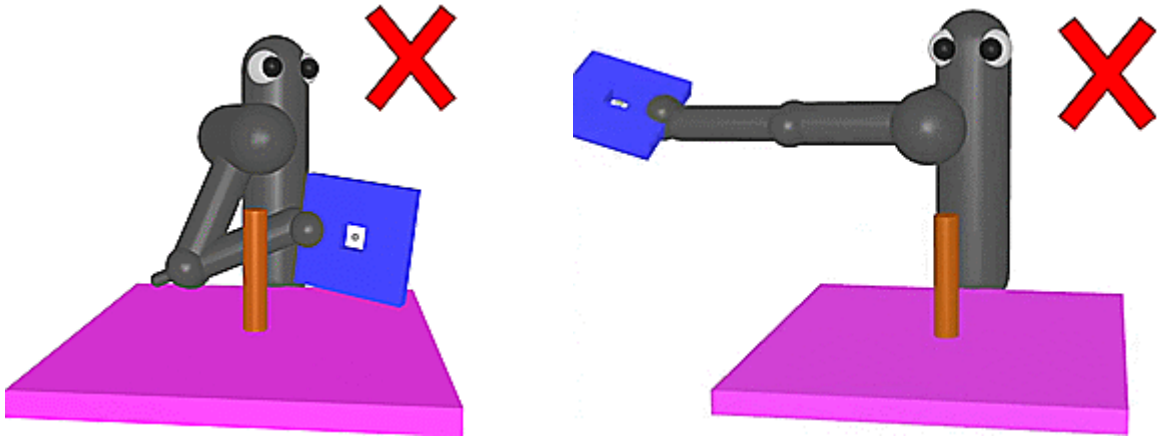
$r_{t+1} = 1$

Reward Shaping

VizDoom

https://openreview.net/forum?id=Hk3mPK5gg&noteId=Hk3mPK5gg

Parameters	Description	FlatMap | CIGTrack1
living	Penalize agent who just lives	-0.008 / action
health_loss	Penalize health decrement	-0.05 / unit
ammo_loss	Penalize ammunition decrement	-0.04 / unit
health_pickup	Reward for medkit pickup	0.04 / unit
ammo_pickup	Reward for ammunition pickup	0.15 / unit
dist_penalty	Penalize the agent when it stays	-0.03 / action
dist_reward	Reward the agent when it moves	9e-5 / unit distance



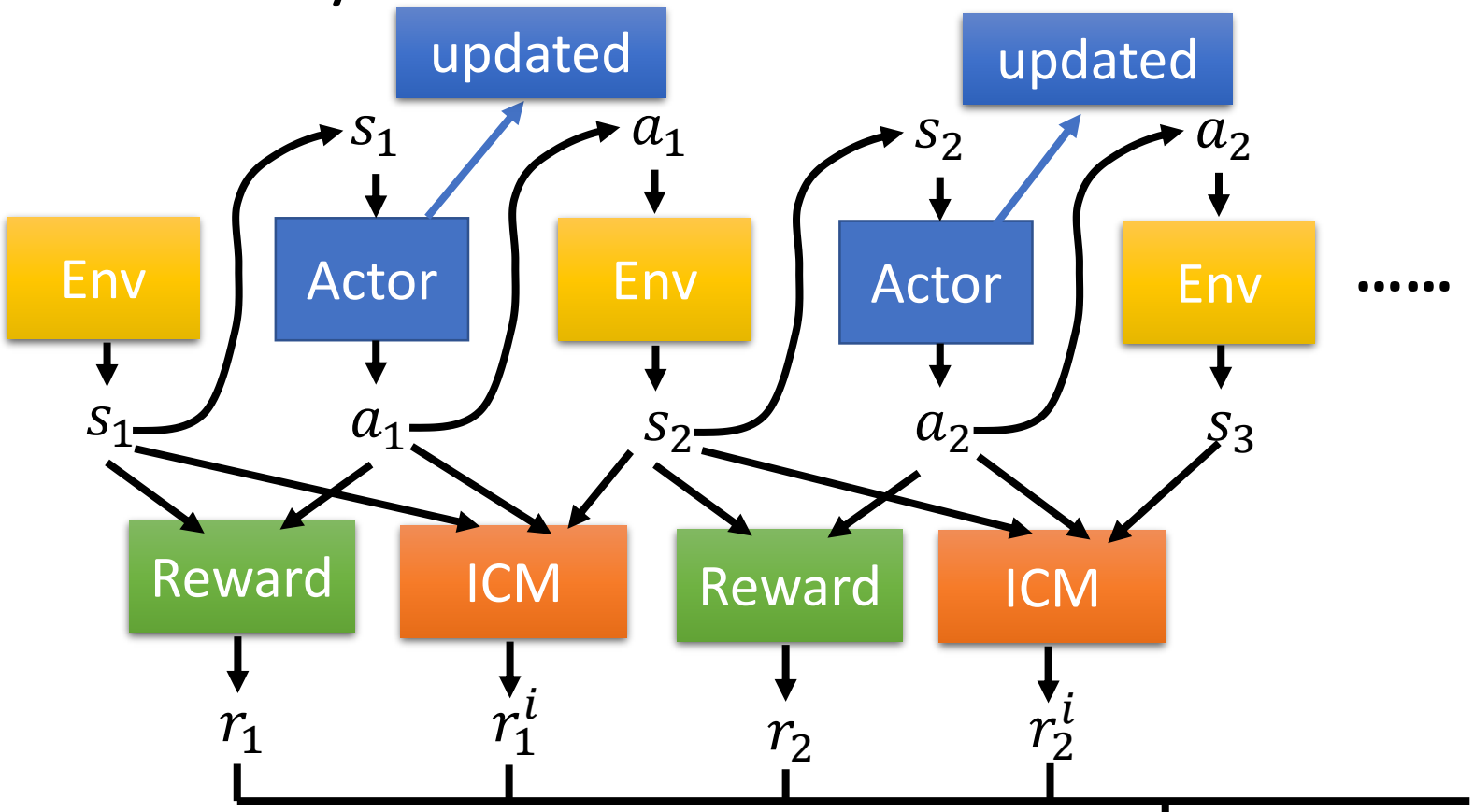
Get reward, when closer

Need domain knowledge

https://openreview.net/pdf?id=Hk3mPK5gg

https://arxiv.org/abs/1705.05363

Curiosity

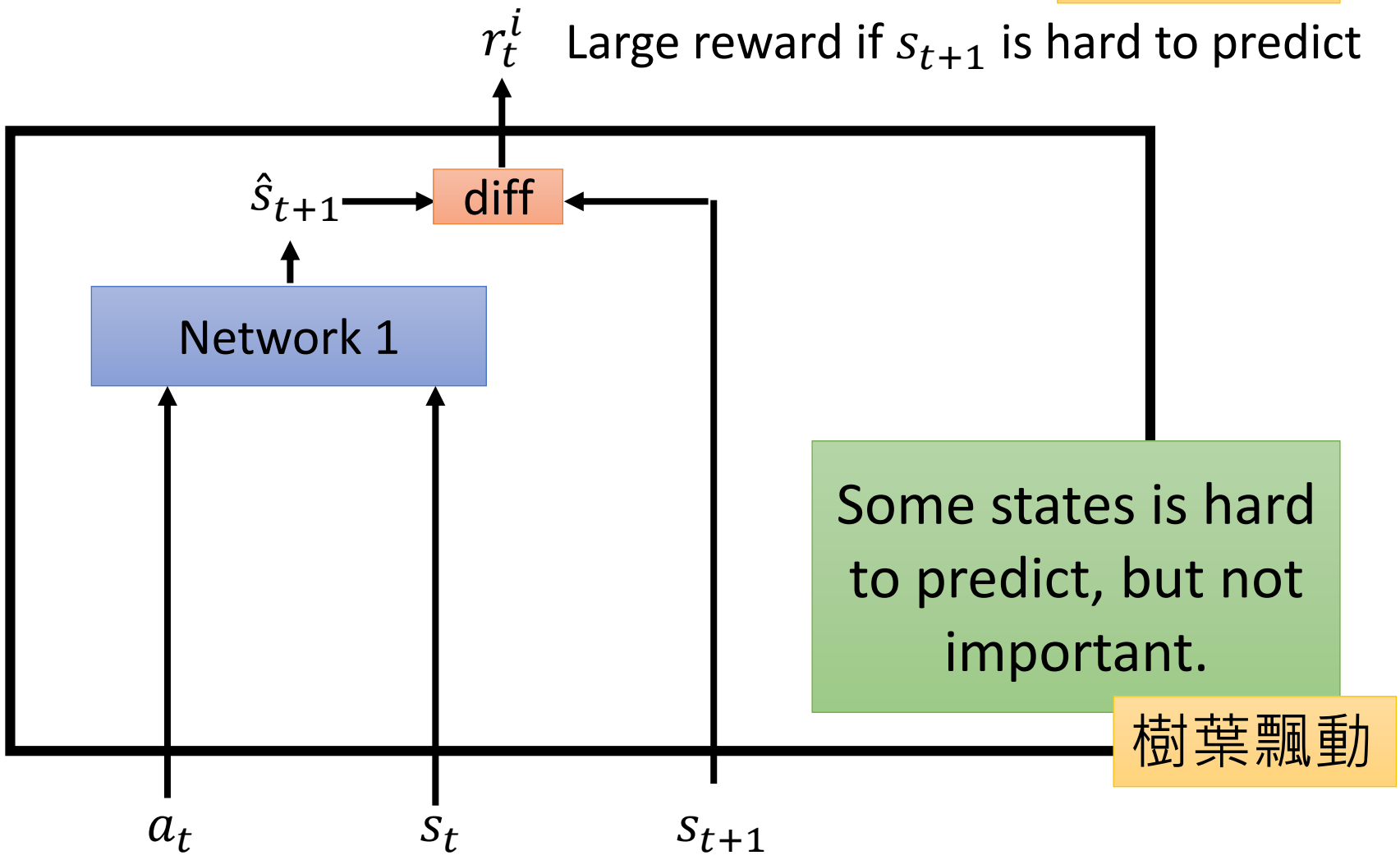


ICM = intrinsic curiosity module

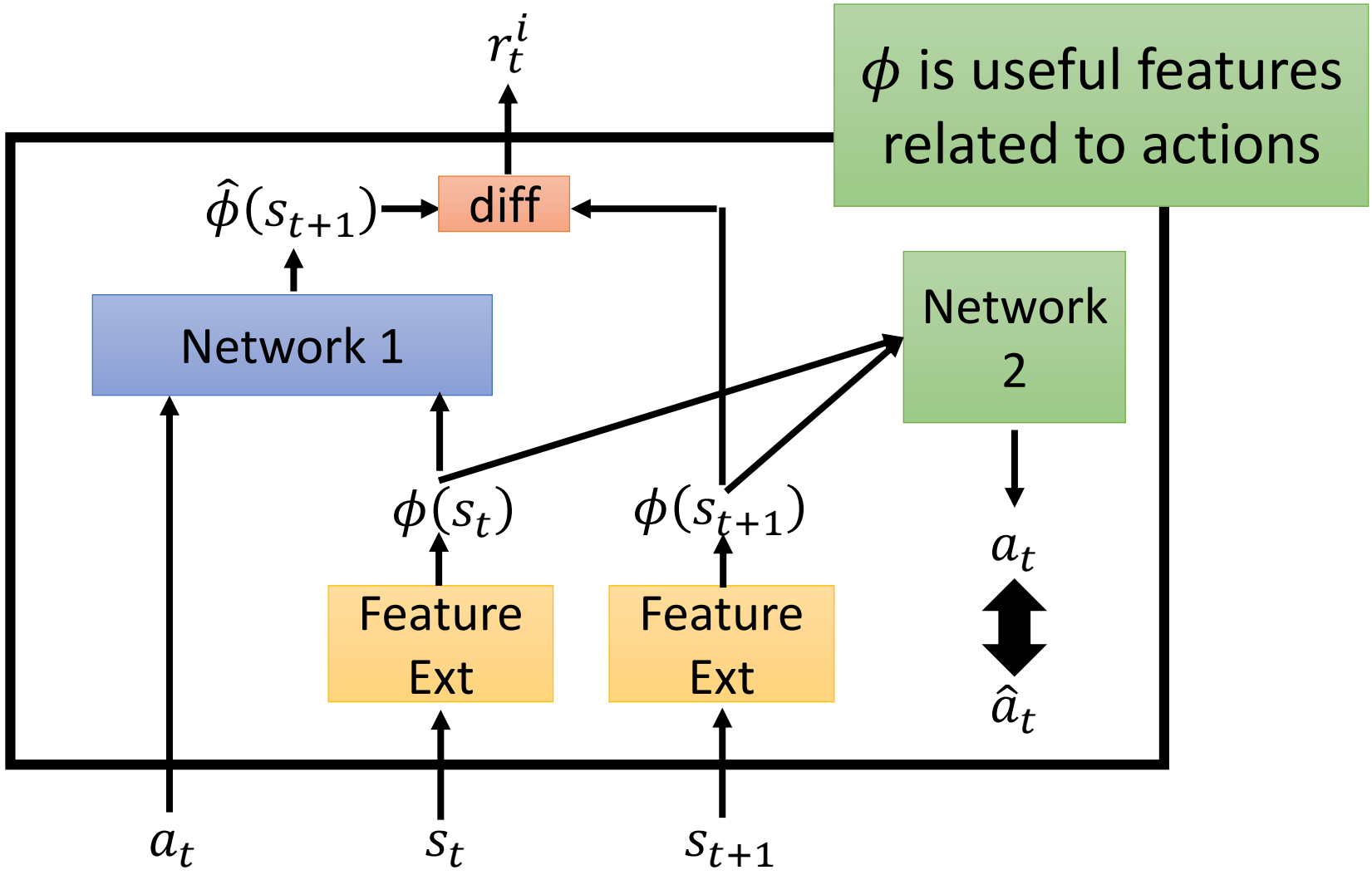
$$R(\tau) = \sum_{t=1}^{T} r_t + r_t^i$$

Intrinsic Curiosity Module

鼓勵冒險

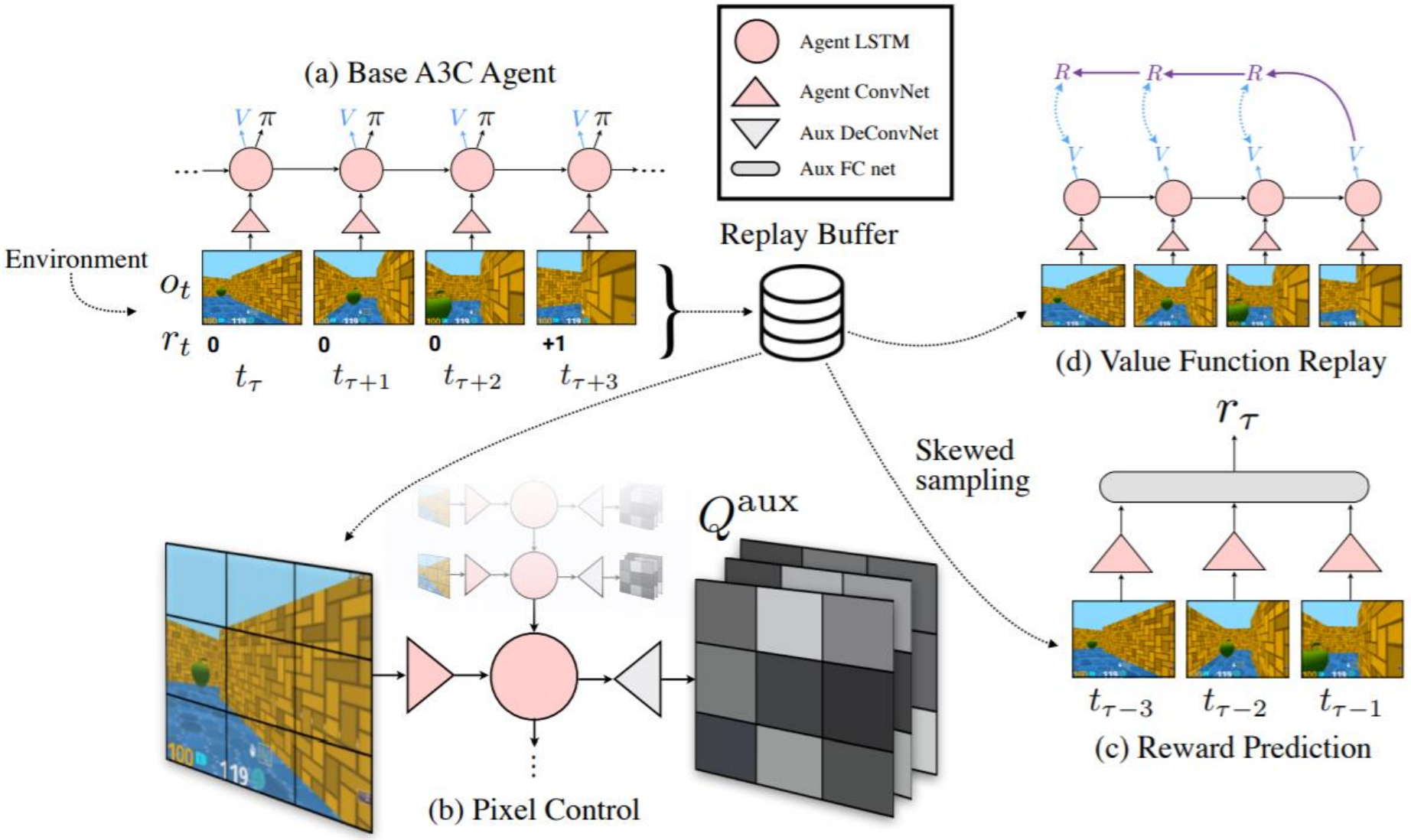


Intrinsic Curiosity Module



https://arxiv.org/abs/1611.05397

Reward from Auxiliary Task



Demo

Sparse Reward

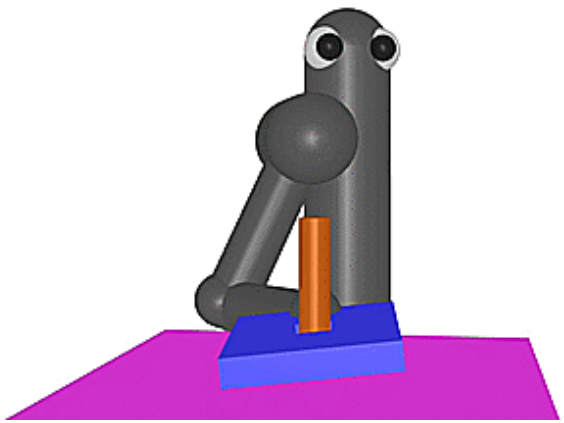
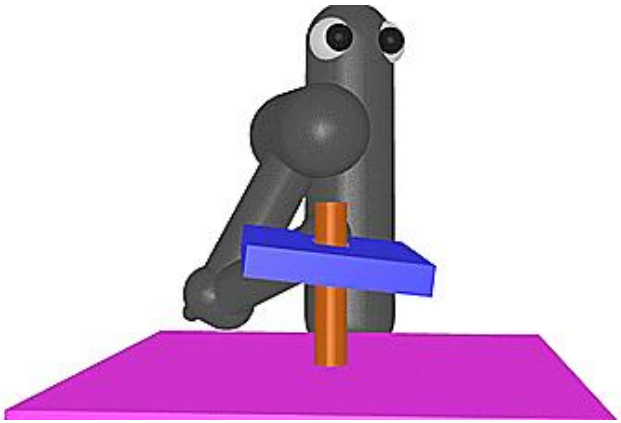
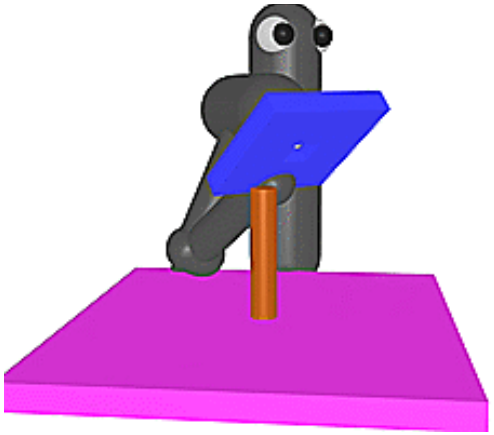
Curriculum Learning

Curriculum Learning

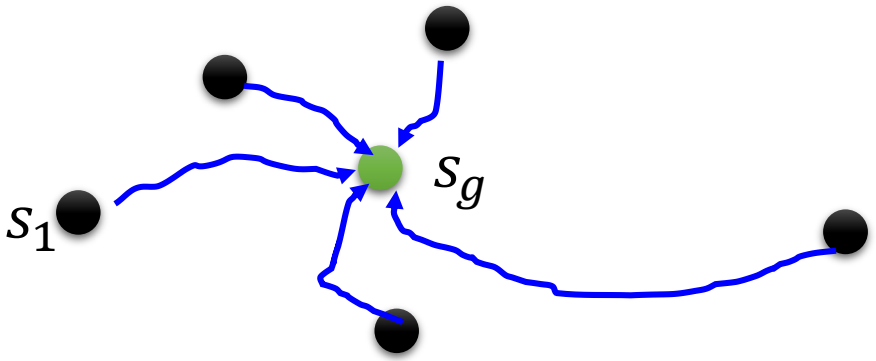
- Starting from simple training examples, and then becoming harder and harder.

VizDoom

	Class 0	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6	Class 7
Speed	0.2	0.2	0.4	0.4	0.6	0.8	0.8	1.0
Health	40	40	40	60	60	60	80	100

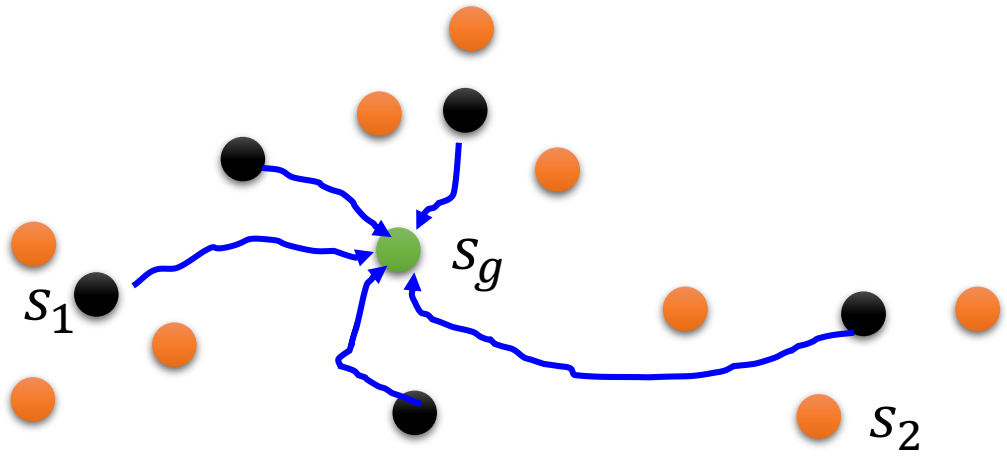




Reverse Curriculum Generation



- Given a goal state s_g.
- Sample some states s_1 “close” to s_g
- Start from states s_1, each trajectory has reward $\mathrm{R}(s_1)$

Reverse Curriculum Generation



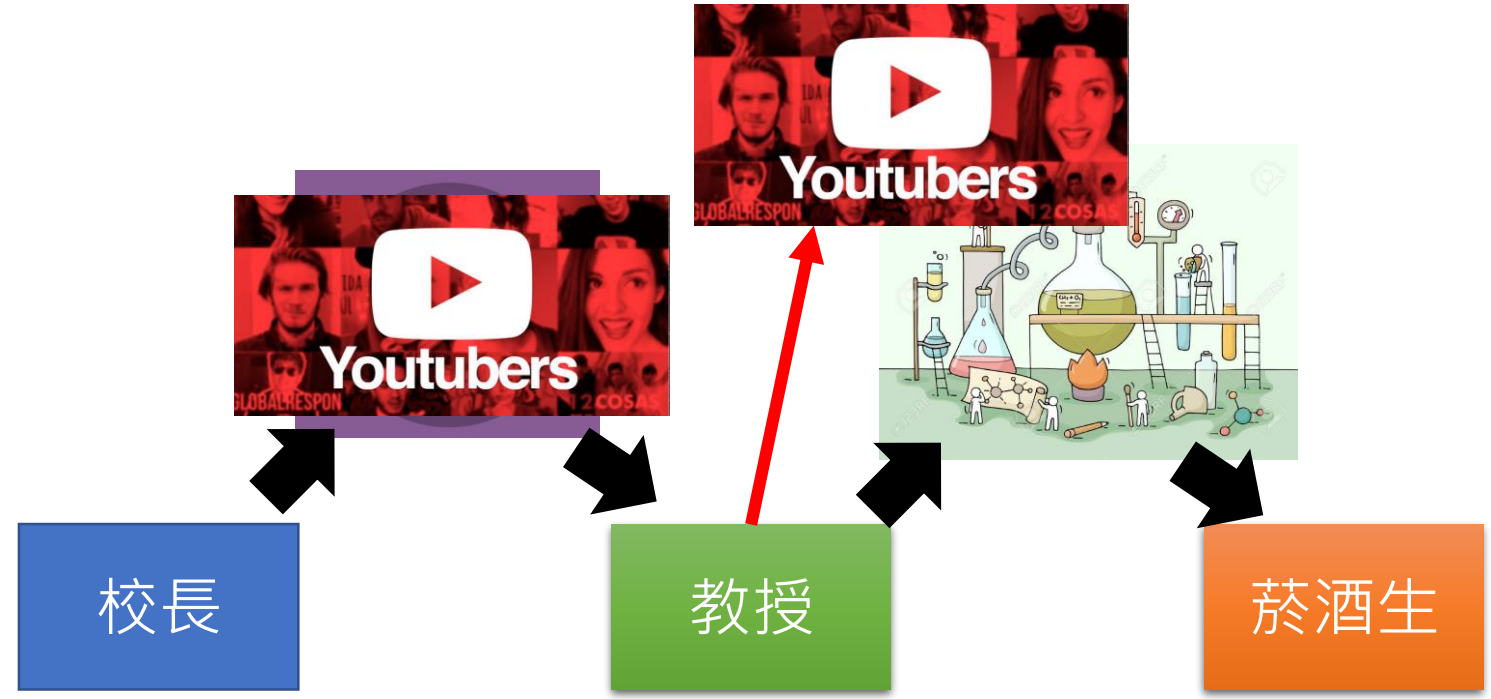
- Delete s_1 whose reward is too large (already learned) or too small (too difficult at this moment)
- Sample s_2 from s_1, start from s_2

Sparse Reward

Hierarchical Reinforcement Learning

Hierarchical RL

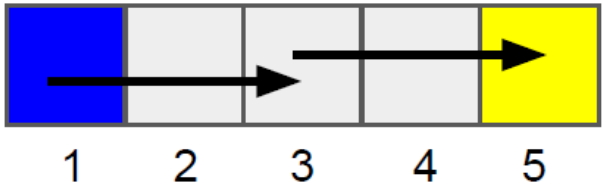
下面這個例子純屬虛構，
跟真實的狀況完全不同



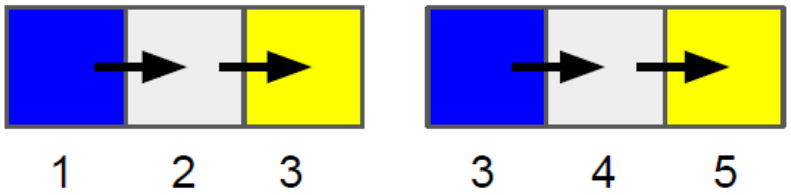
- If lower agent cannot achieve the goal, the upper agent would get penalty.
- If an agent get to the wrong goal, assume the original goal is the wrong one.

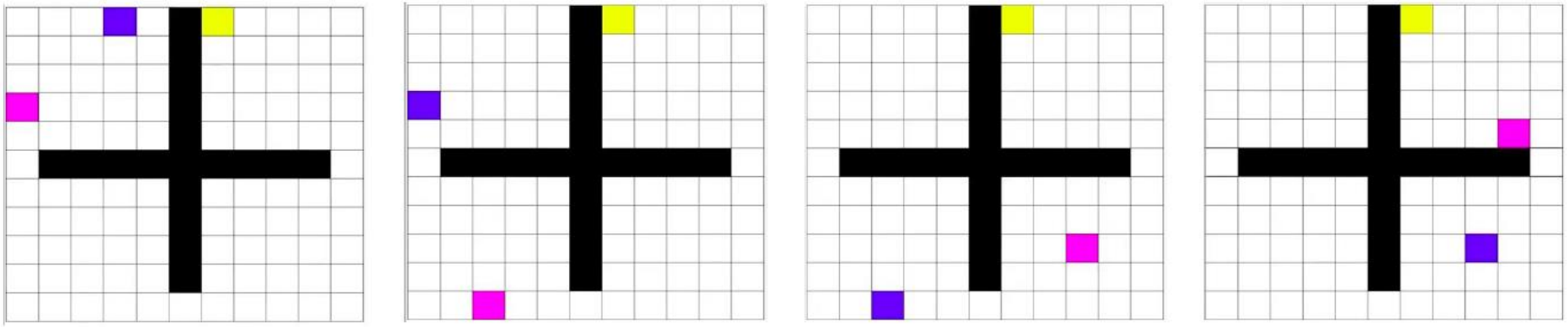
https://arxiv.org/abs/1805.08180

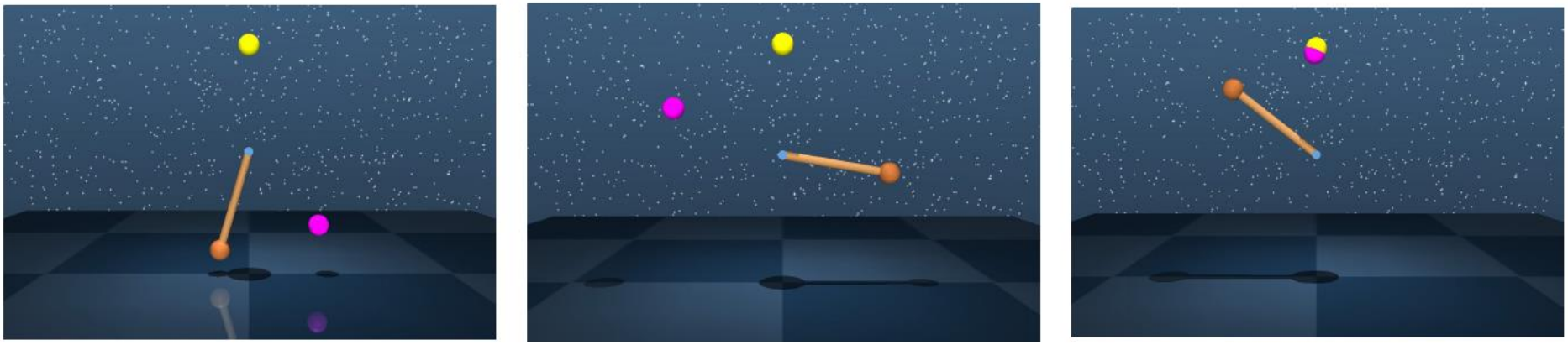
High-Level



Low-Level







https://arxiv.org/abs/1805.08180

Acknowledgement

- 感謝芮祥麟博士發現課程網頁上拼字的錯誤