Reinforcement Learning with Unsupervised Auxiliary Tasks, M.Jaderberg,2016.

Reinforcement Learning with Unsupervised Auxiliary Tasks

Lee Won-ho
gh9908@gmail.com

Reinforcement Learning with Unsupervised Auxiliary Tasks, M.Jaderberg,2016.

Index

- A3C
- Pseudo-reward
- Auxiliary Tasks
  - control task
  - reward prediction task
- **UNREAL** Agent
- Experiments

Reinforcement Learning with Unsupervised Auxiliary Tasks, M.Jaderberg,2016.

A3C(Asynchronous Advantage Actor-Critic)

- Advantage Actor-Critic
  - 기존 Actor-Critic에서는 2개 이상의 액션에 대해 비슷한 평가를 내릴 수 있음
  - 애초에 해당 상태의 가치가 너무 높게 책정했기 때문
  - 해당 액션을 통해 추가로 얻는 가치에 초점을 둠
- Asynchronous → n-step

Reinforcement Learning with Unsupervised Auxiliary Tasks, M.Jaderberg,2016.

Pseudo-reward

classic rl focuses on maximisation of **Extrinsic rewards**(too sparse)

↓

what and **how** to learn in their absence

↓

reconstruct targets(pixels, features),

maximisation of pseudo-reward functions

Reinforcement Learning with Unsupervised Auxiliary Tasks, M.Jaderberg,2016.

Auxiliary task

- Control task
  - reconstruct targets
- Reward Prediction task

Reinforcement Learning with Unsupervised Auxiliary Tasks, M.Jaderberg,2016.

Auxiliary task control task

- Pixel Changes
  - changes correspond to important events in an environment
- Network features
  - policy or value network be learned to extract task-relevant high level features of environment
  - useful quantities for the agent to learn to control
  - activation of any hidden unit of agent’s neural network can itself be an auxiliary reward

Reinforcement Learning with Unsupervised Auxiliary Tasks, M.Jaderberg,2016.

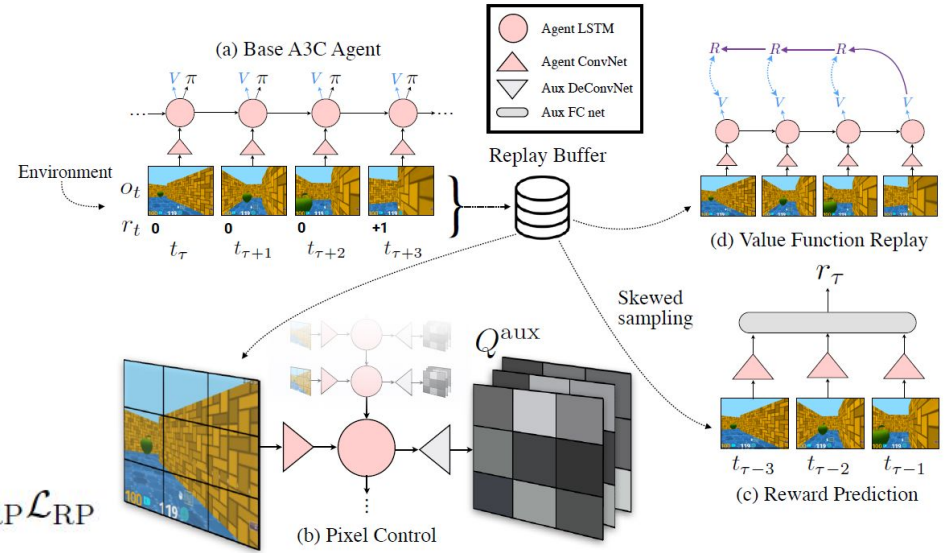
Auxiliary task reward prediction task

- using prioritised replay
  - oversampling rare rewarding states
  - replay buffer to perform *value function replay*

Reinforcement Learning with Unsupervised Auxiliary Tasks, M.Jaderberg,2016.

UNREAL Agent

- **UN**supervised **RE**inforcement and **A**uxiliary **L**earning
  - primary policy is trained with **A3C** → efficiency and stability
  - auxiliary tasks are trained with **prioritised experience replay** → efficiency
  - Auxiliary control
  - Auxiliary reward prediction



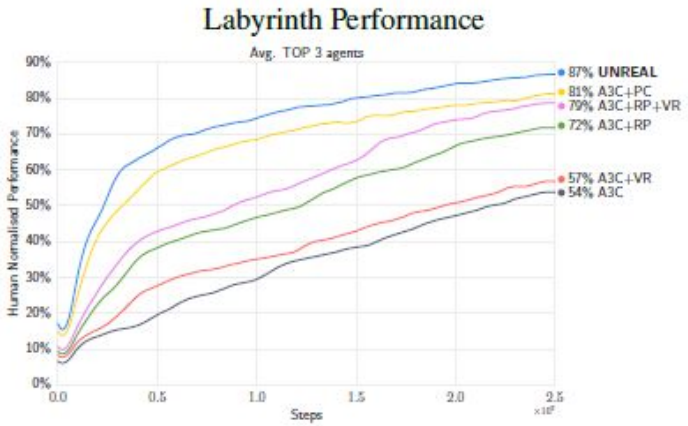
$$\mathcal{L}_{UNREAL}(\theta) = \mathcal{L}_{\text{A3C}} + \lambda_{\text{VR}}\mathcal{L}_{\text{VR}} + \lambda_{\text{PC}}\sum \mathcal{L}_{Q}^{(c)} + \lambda_{\text{RP}}\mathcal{L}_{\text{RP}}$$

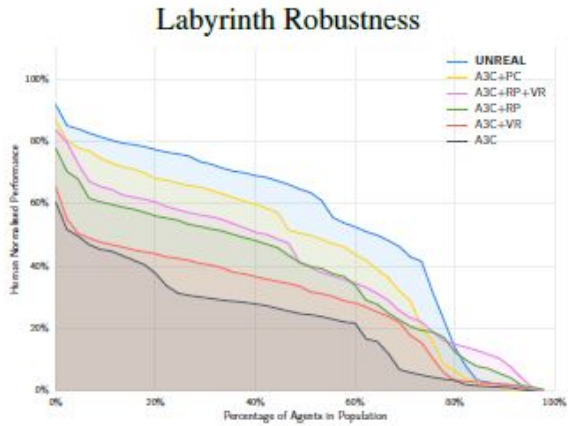
Experiments

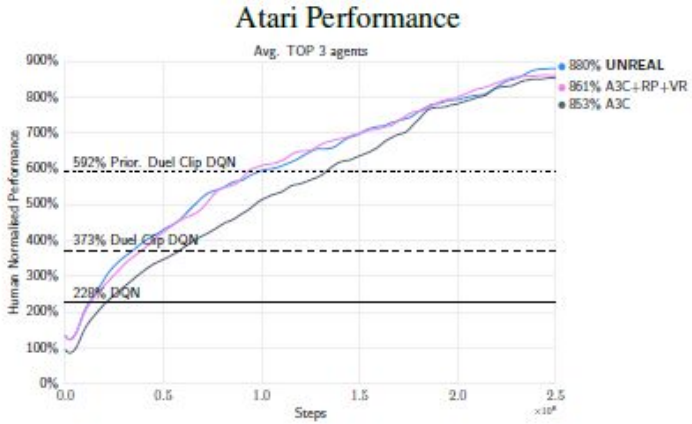
Performance
UNREAL 87% > A3C 54%

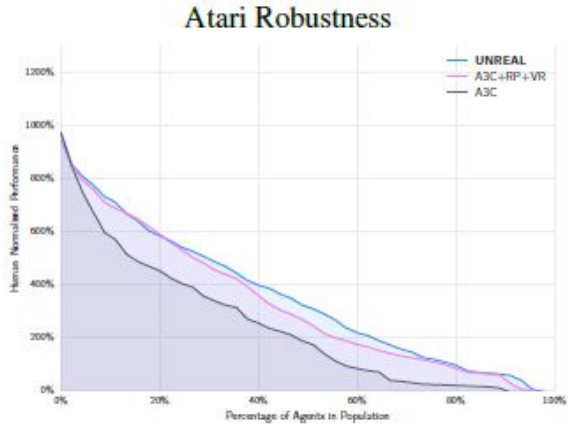
Robustness
50% of UNREAL agents > A3C

Reinforcement Learning with Unsupervised Auxiliary Tasks, M.Jaderberg,2016.







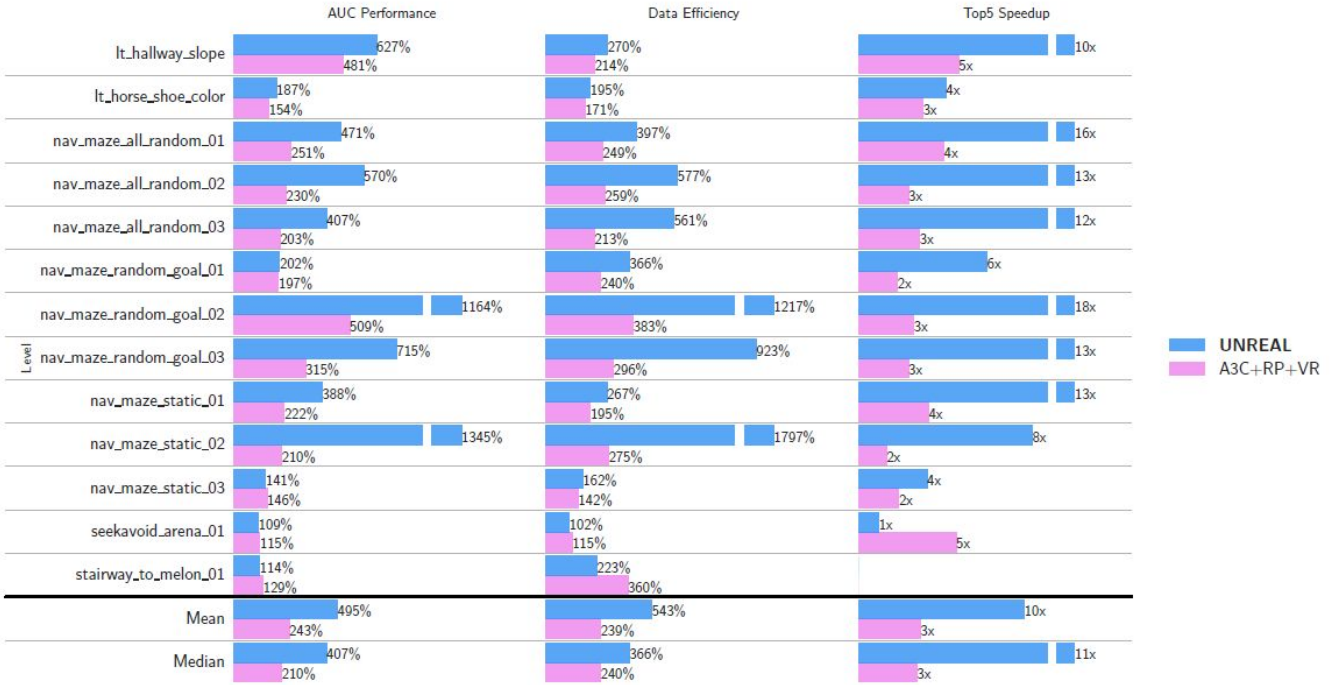


https://youtu.be/Uz-zGYrYEjA

Reinforcement Learning with Unsupervised Auxiliary Tasks, M.Jaderberg,2016.

Experiments

performance of **pixel control**



Experiments

input reconstruction → hurts final performance

pixel control > simply predicting immediate pixel changes

Reinforcement Learning with Unsupervised Auxiliary Tasks, M.Jaderberg,2016.