Running A.I

정원석

1st 함께하는 딥러닝 컨퍼런스
06.28

소개

정원석

뉴욕시립대 - Baruch college (Data Science Major)

ConnexionAI Freelance Researcher

모두의연구소 CTRL (Contest in RL) 랩장

DeepLearningCollege 강화학습 연구원

Github:
https://github.com/wonseokjung

Facebook:
https://www.facebook.com/ws.jung.798

Blog:
https://wonseokjung.github.io/

순서

1. Reinforcement Learning
2. Atari
3. SuperMario
4. Sonic
5. Prosthetics
6. Latest trend

The Rise of Reinforcemet Learning

By Wonseok Jung

Reinforcement learning

아기는 어떻게 배울까?

바라본다

만진다

웃는다

운다

아기가 자라면서, 주위를 바라보고, 팔을 들고, 노는 행동을
누군가가 가르쳐서 하는 것은 아닐것이다.

https://goo.gl/images/RQgtpW

아이가 학습하는 과정

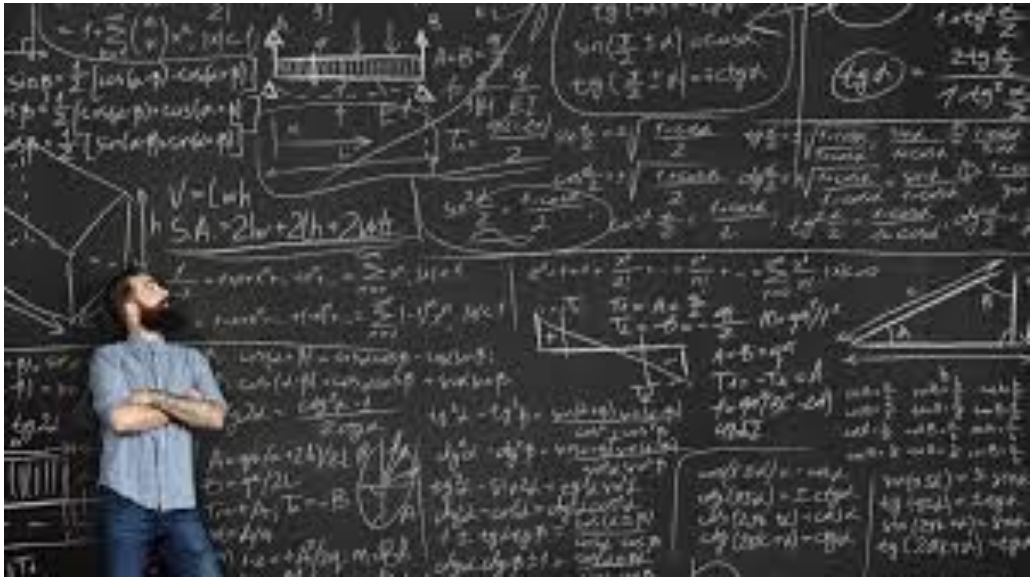
일어나서 다가온다 **칭찬**

일어나서 다가온다

흙을 먹으려 한다. **꾸중**

흙을 먹지 않는다

수학적인분석, 계산 실험



아이가 환경과 상호작용을 하며 배우는 방법과 같이
수학적인 분석과, computation 실험으로 학습하는 방법을
'Reinforcement learning(강화학습)'이라고 한다.

Reward를 최대로 하는 action 선택

선택

1. 반지를 준다
2. 영화를 보자고 한다.
3. 생각에 잠긴다.
4. 택시를 잡는다.

Reinforcement learning은 Reward(보상)을 최대화 하는 action(행동)을 선택한다.

https://goo.gl/images/HpBRJT

Fail and Success

Learner(배우는자)는 여러 action을 해보며,
reward를 가장 높게 받는 action을 찾는다.

https://goo.gl/images/GoHQYh

Reinforcement Learning

선택된 action이 당장의 reward 뿐만 아닌,

다음의 상황 또는 다음 일어나게 될 reward에도 영향을 끼칠수도 있다.

Exploration and Exploitation

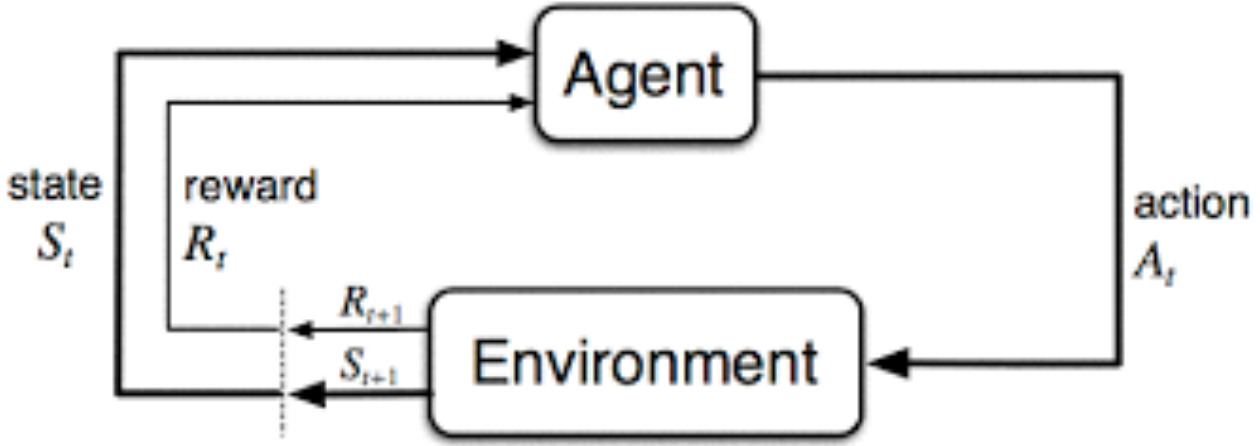
Exploitation

Agent

Exploration

Agent는 reward를 더 많이 받는 action을 선택하기 위해 exploitation을 해야 하지만, 여러가지 action을 골고루 해보며 많은 상황을 경험하기 위해서는 exploration을 해야한다.

Markov Decision process



S_t **Agent** A_t **action** **Env** S_{t+1} R_{t+1} S_{t+1} **Agent** ⇨ ····

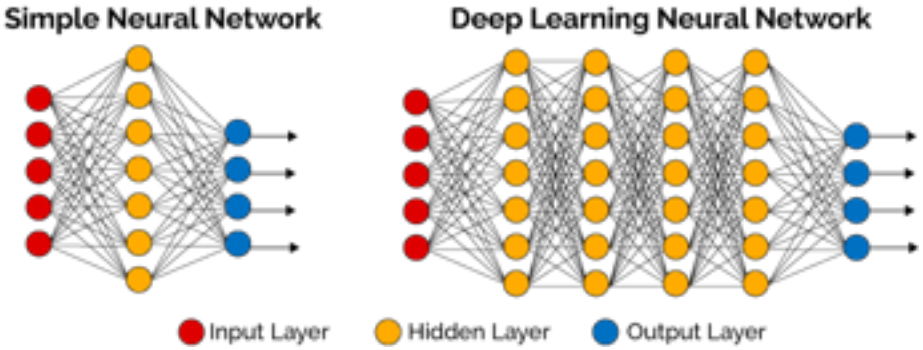
Agent는 MDP를 통해 env와 상호작용을 하며 배운다.

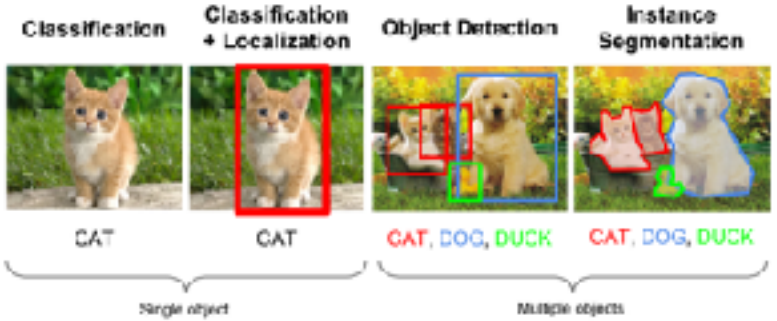
Atari

High dimensional state

Discrete actions

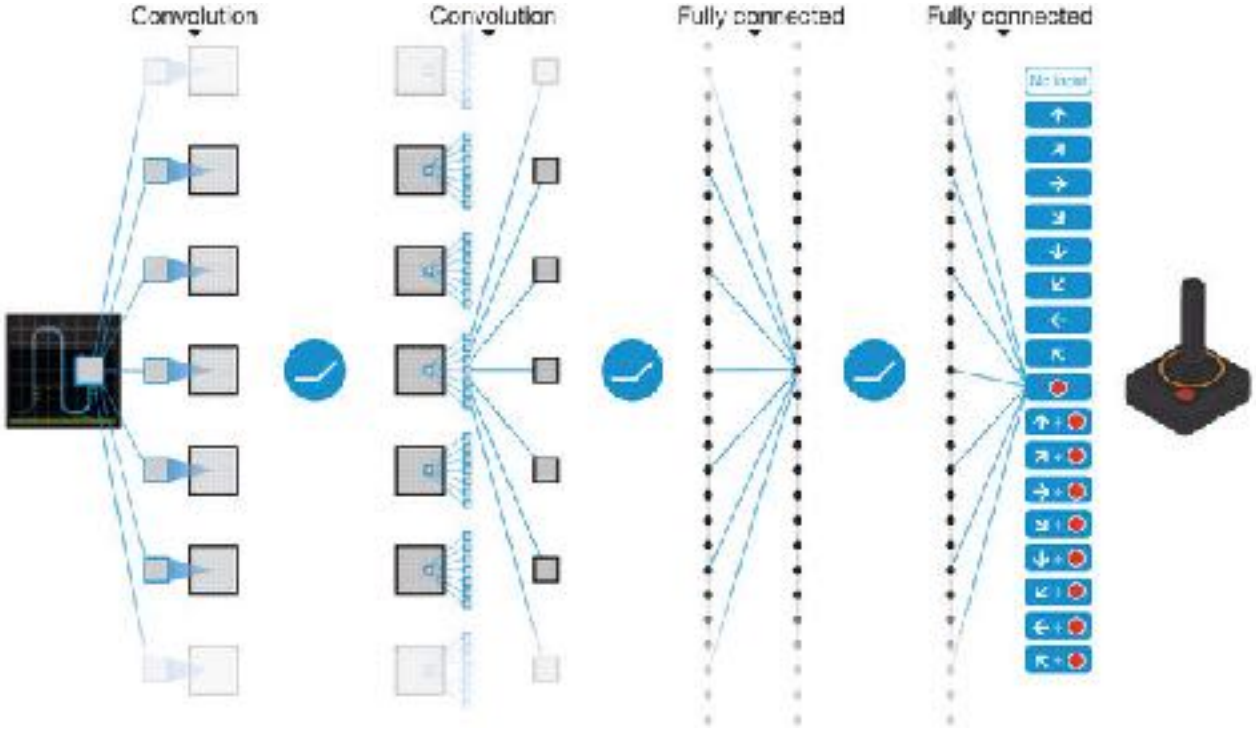
Deeplearning





딥러닝의 등장인해 으로 high dimensional data를 input으로 받는것이 가능해졌다.

Deep learning+Reinforcement Learning



deep network와 reinforcement learning이 결합한 알고리즘

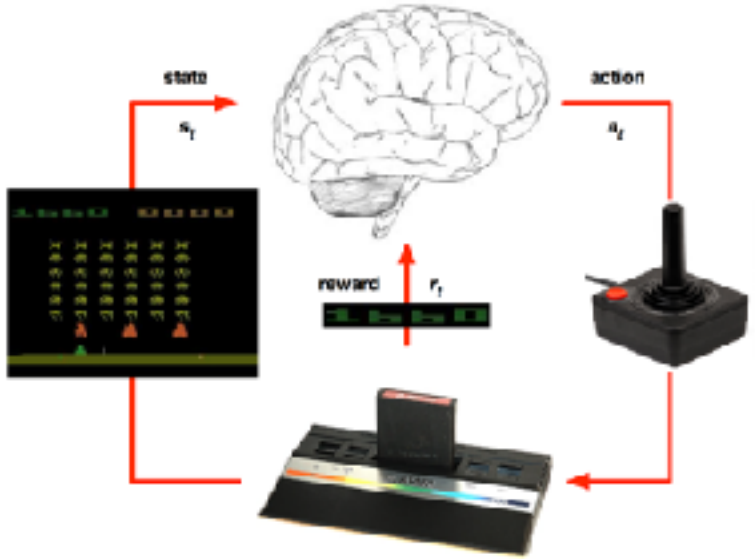
https://goo.gl/images/oNu5Gr

Deepmind, DQN

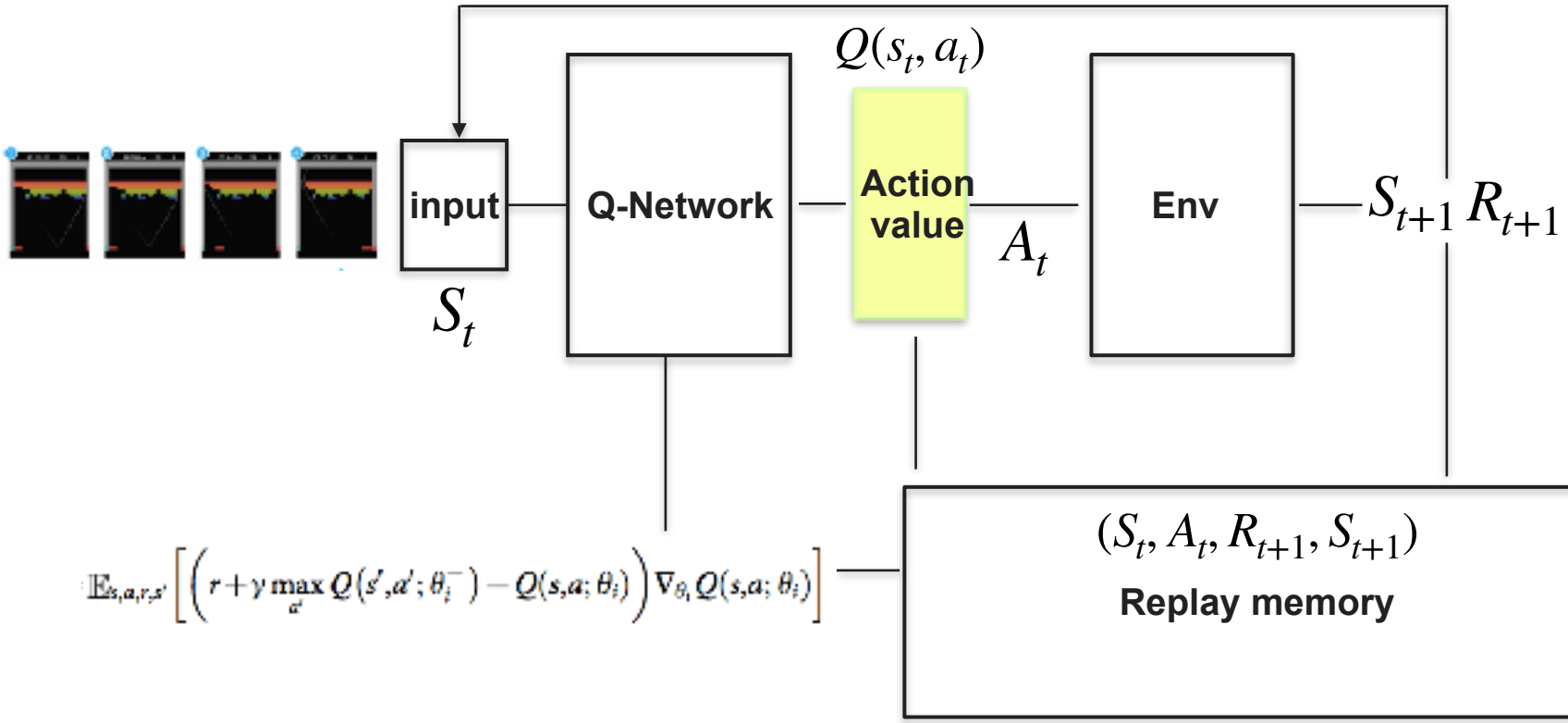
Deeplearning을 강화학습에 적용하여, 사람보다 플레이를 잘하는 인공지능을 만듦

Human-level control through deep reinforcement learning

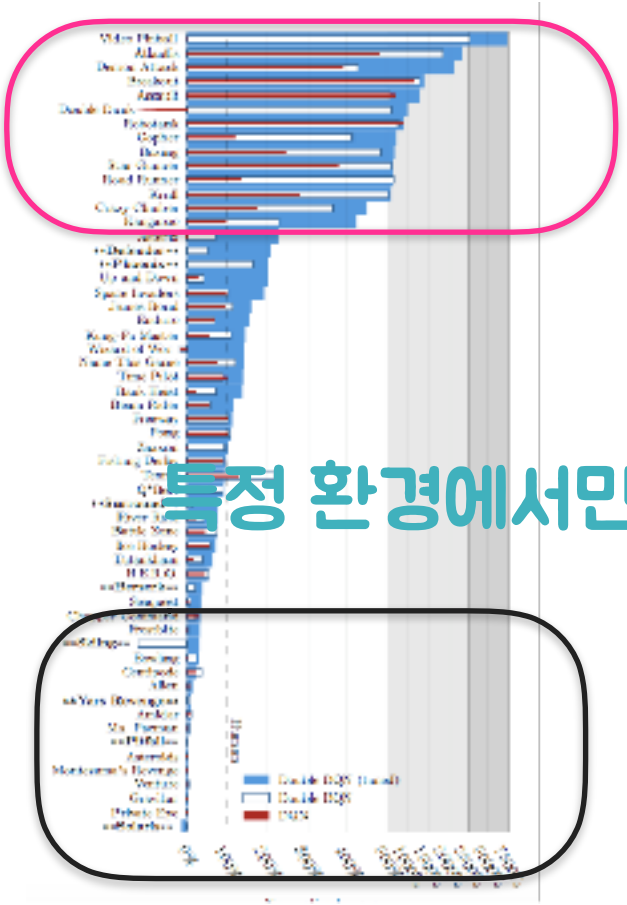
Volodymyr Mnih[1]*, Koray Kavukcuoglu[1]*, David Silver[1]*, Andrei A. Rusu[1], Joel Veness[1], Marc G. Bellemare[1], Alex Graves[1], Martin Riedmiller[1], Andreas K. Fidjeland[1], Georg Ostrovski[1], Stig Petersen[1], Charles Beattie[1], Amir Sadik[1], Ioannis Antonoglou[1], Helen King[1], Dharshan Kumaran[1], Daan Wierstra[1], Shane Legg[1] & Demis Hassabis[1]



Deep Q network Architecture



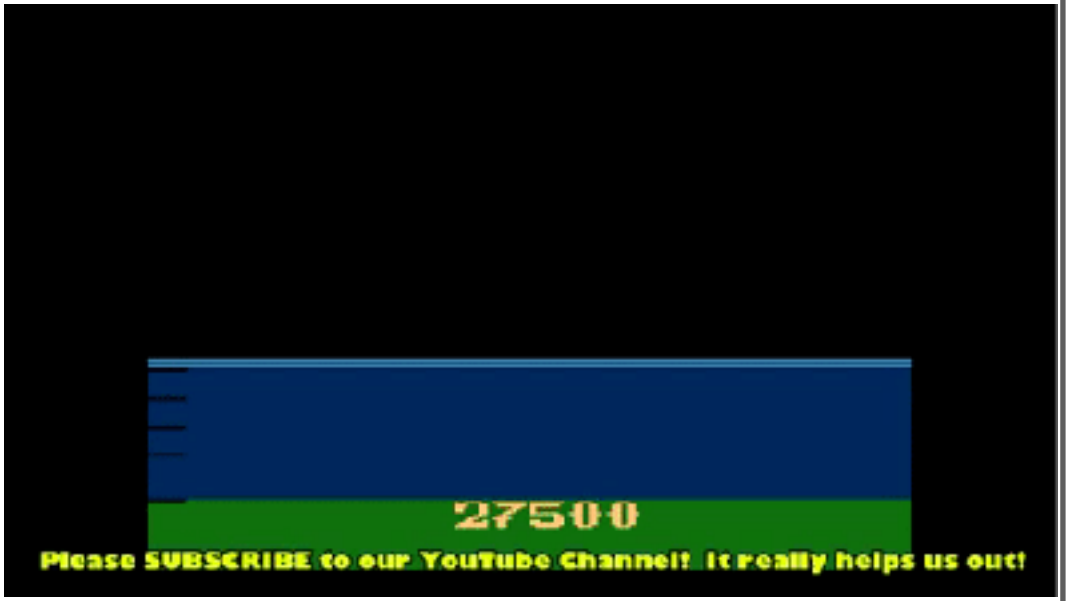
Atari에서 DQN의 한계



특정 환경에서만 퍼포먼스가 좋다

성능이 상위권인 환경


22500
4


27500
Please SUBSCRIBE to our YouTube Channel! It really helps us out!

성능이 중 하위권인 환경

JamesBond

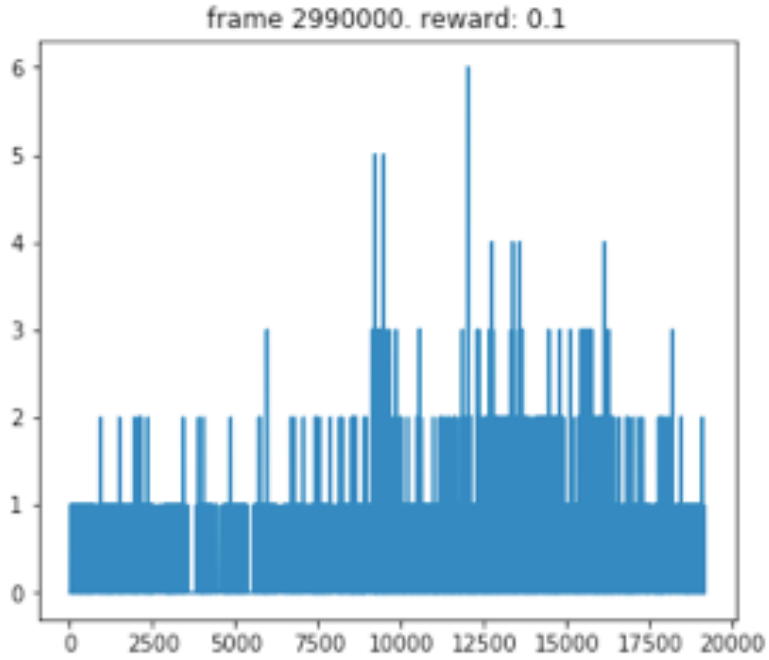
Skiing

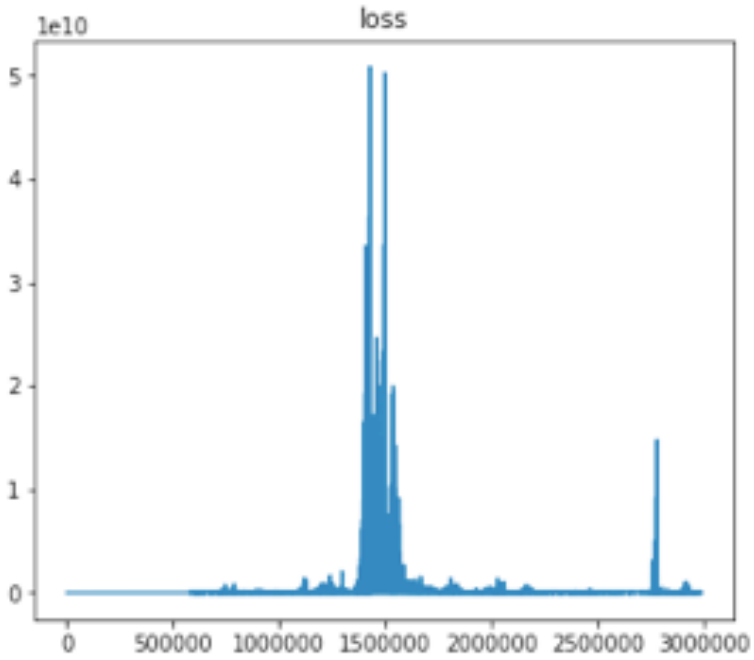
Chopper command

배경이 변하는 환경에서는 학습 성능이 떨어진다.

Result

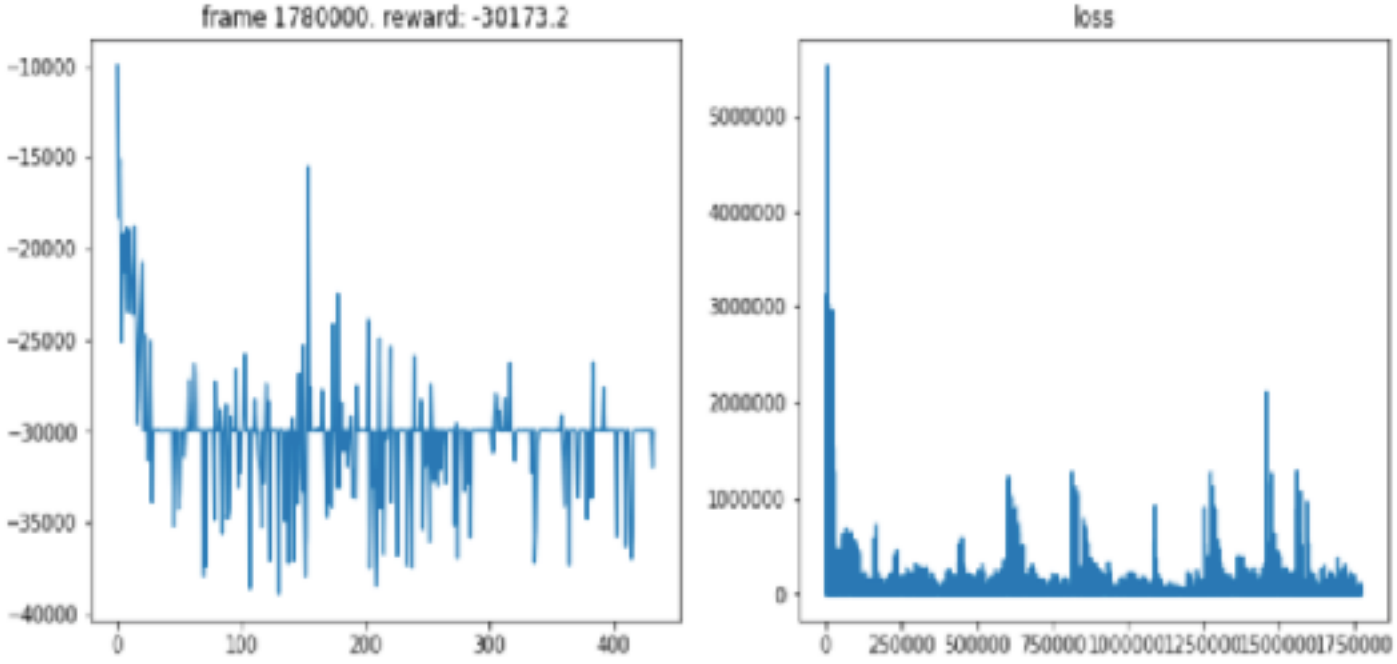
JamesBond





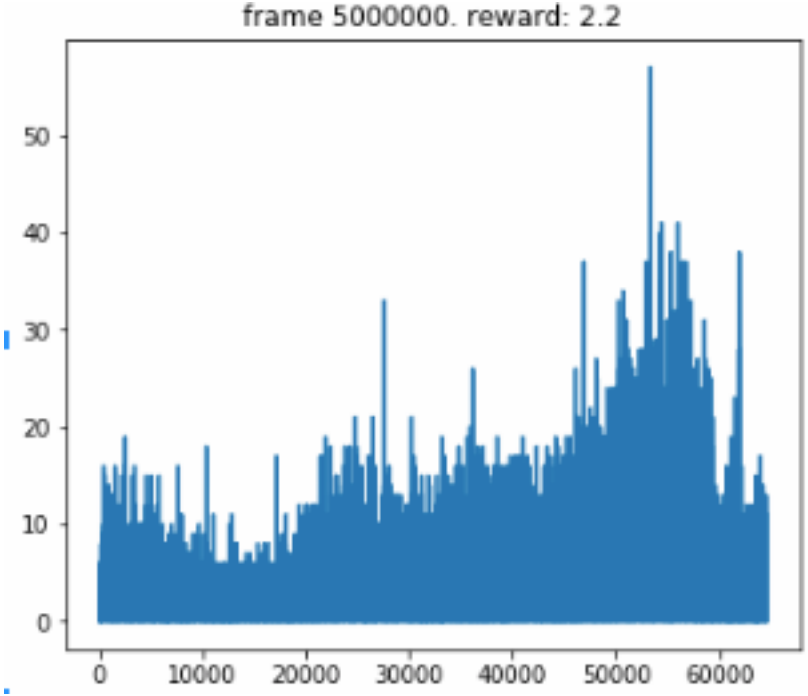
Result

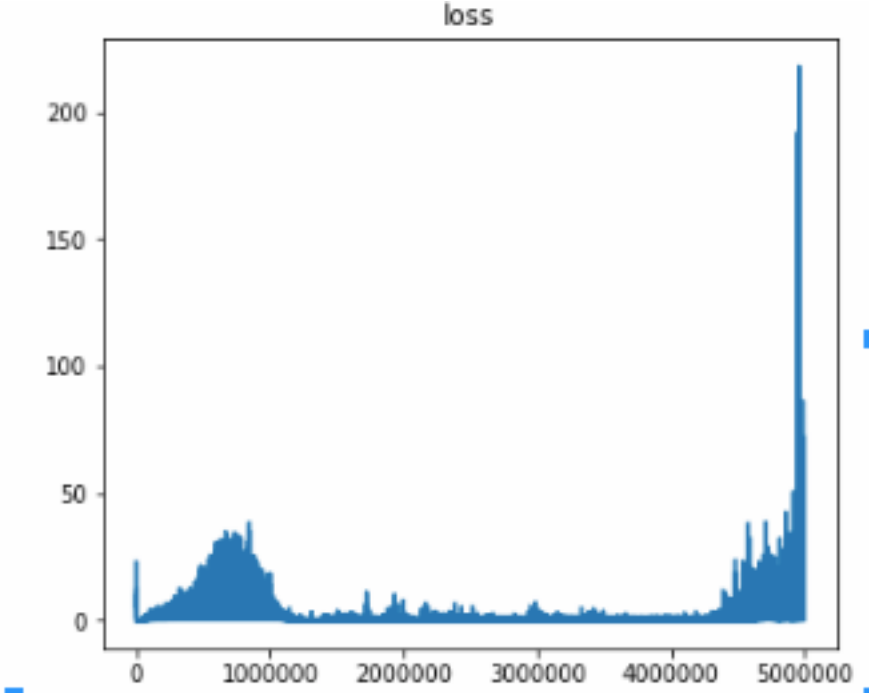
Skiing



Result

Command chopper





더 복잡한 state와 더 많은 action이 있는 환경에서는 ?

SuperMario

High dimensional state

Discrete actions

Complex Environment

First challenge - SuperMario Bros

1985 Nintendo

강화학습으로 똑똑한 Mario를 만들어보자

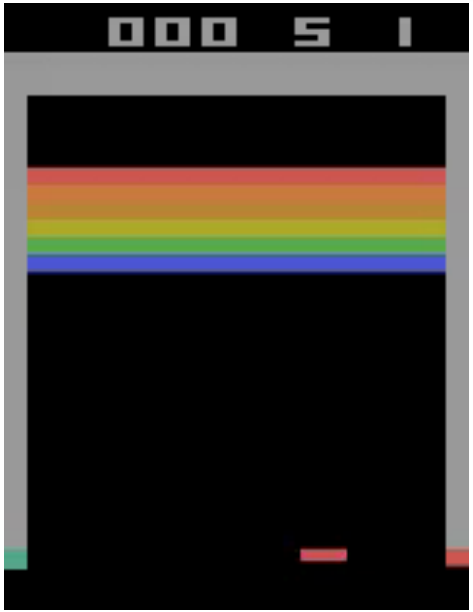
벽돌깨기와 슈퍼마리오의 Goal의 비교

벽돌을 모두 없애는 것이 목표

슈퍼마리오는 깃발을 잡는것이 목표

Reward - Breakout

State



State : 화면, [210, 260 , 3]

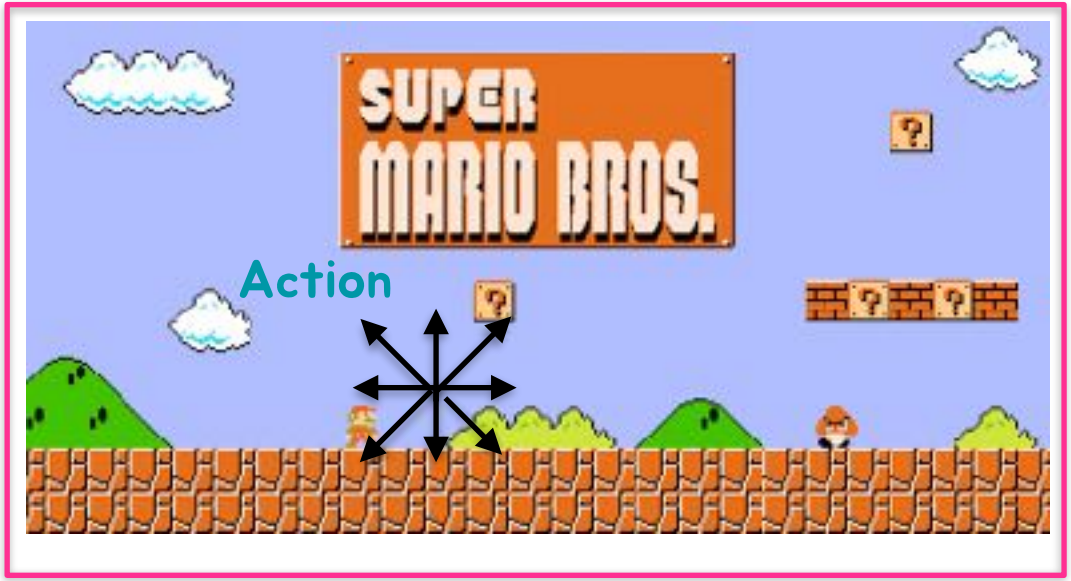
Action : None, 왼쪽, 오른쪽

Reward : 벽돌 격파

벽돌을 없앨수록 높은 Reward를 받는다.

Reward - 슈퍼마리오

State



State : 화면

Action : 상, 하 , 좌, 우,점프,달리기, action의 조합

Reward : 앞으로 전진할때 Reward +1, 뒤로가면 -1

Transition Probability : 1

도착지인 깃발에 가까이 갈수록 높은 reward를 받는다.

DQN을 사용하여 학습

input

s

Q-Network

Q(s,a)

Action value

a

Env

s’ r

$$\mathbb{E}_{s,a,r,s'}\left[\left(r+\gamma\max_{a'}Q(s',a';\theta_i^-)-Q(s,a;\theta_i)\right)\nabla_{\theta_i}Q(s,a;\theta_i)\right]$$

$(S_t, A_t, R_{t+1}, S_{t+1})$

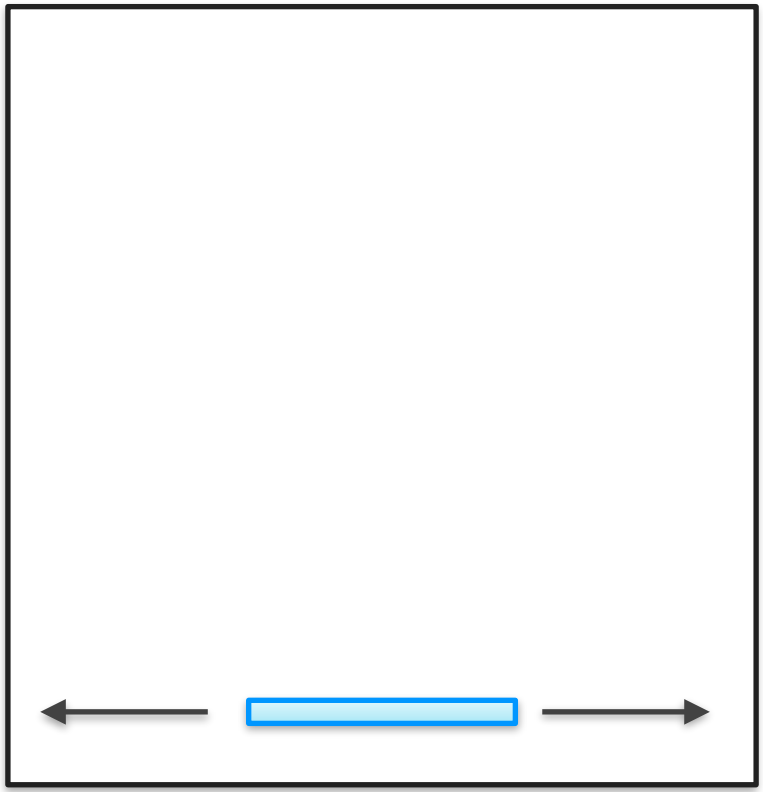
Replay memory

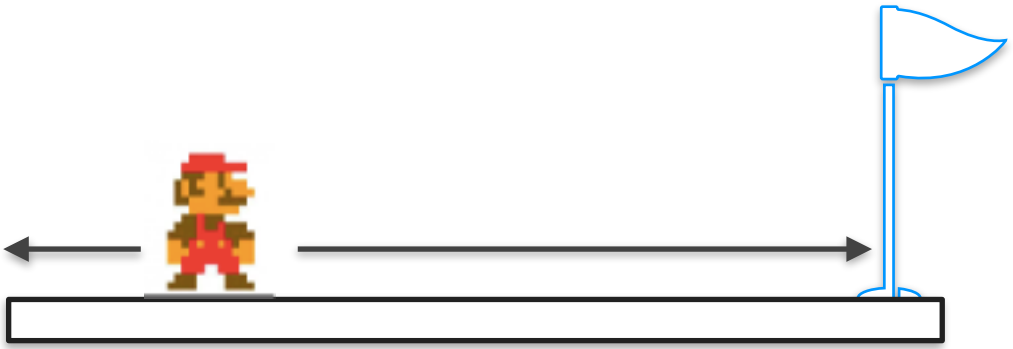
계속되는 실패...

https://youtu.be/zRf_7Xa_MSE

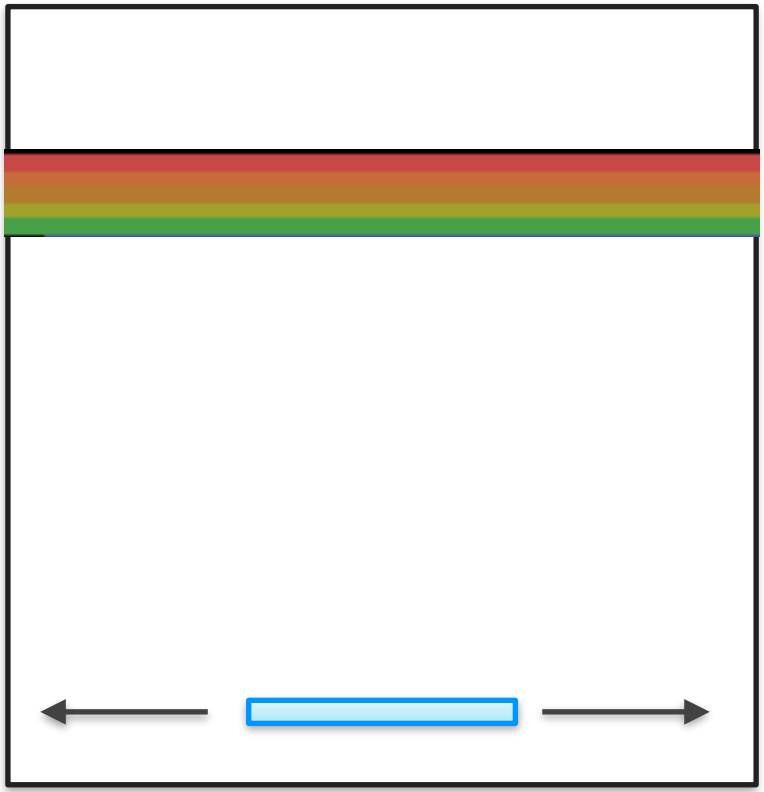
원인이 무엇일까?

Action의 범위

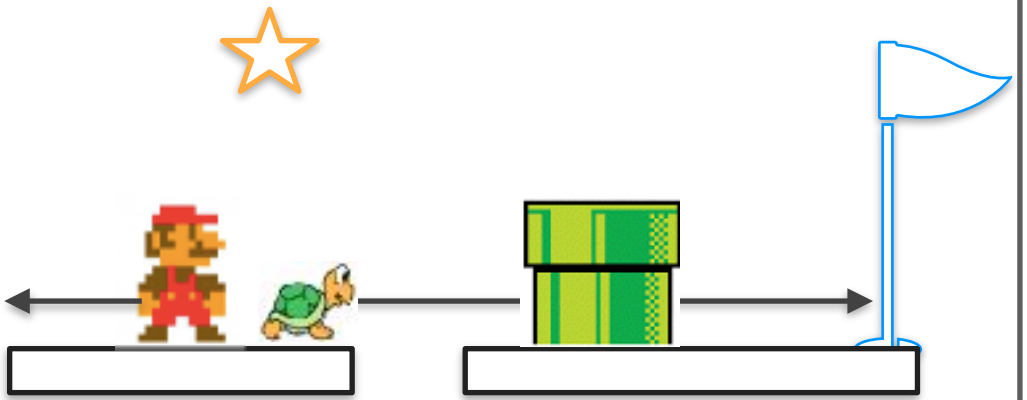




Complexity



복잡성으로 인해 학습이 더욱 어렵다.



Reward 설정

목표달성하지 못하면 -

시간이 지날때마다 -

깃발에서 멀어지면 -

깃발에 가까워지면 +

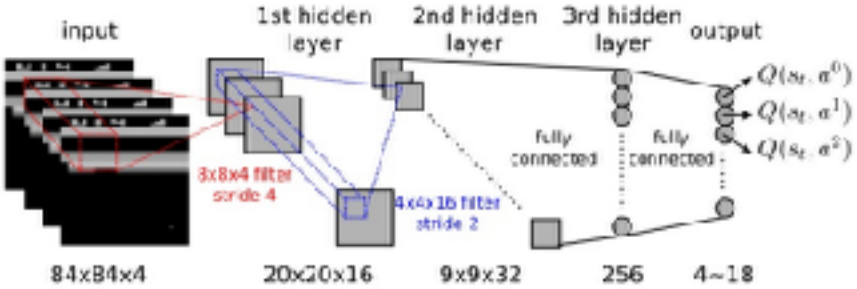
목표에 도착하면 +

Penalty, Bonus reward추가

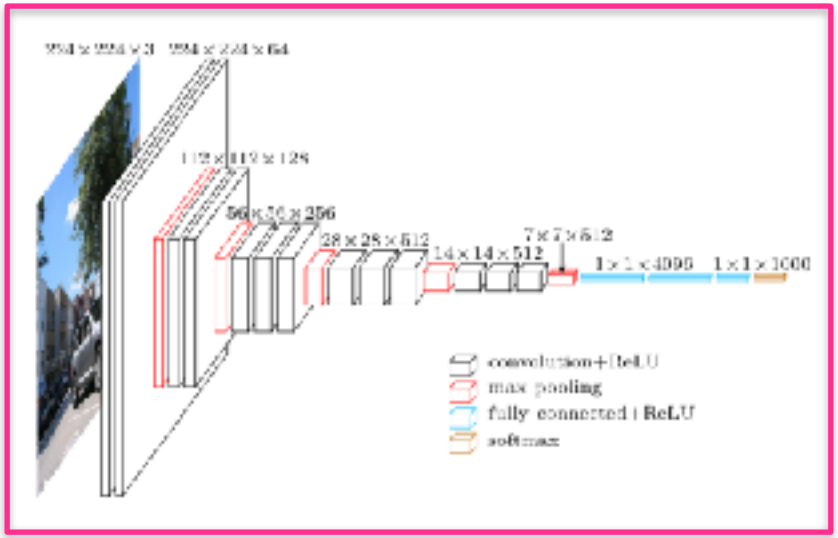
Deep learning model

VGG model and regular 비교

더 깊게 쌓아보자



https://goo.gl/images/s8XrCK



https://goo.gl/images/eoXooC

Level 1 통과!

After

7000Episodes

6 Days

https://youtu.be/WlLBRsgSFt8

풀리지 않은 문제들

각 Level의 화면이 다르기 때문에 General agent를
만들기가 어렵다.

레벨 2를 학습시키는 도중..

Exploration??

https://youtu.be/EvyM4ZUhDpE

Sonic

High dimensional state

Discrete actions

More Complex Environment

Skills

OpenAI Retro challenge

OpenAI에서 개최한 Sonic Contest에 참여

더 어려워진 난이도와 많아진 action 조합

action의 조합 + skill

또한 복잡성이 높아짐

최신 DQN 알고리즘을 사용해보자

To the Rainbow

Rainbow: Combining Improvements in Deep Reinforcement Learning

Matteo Hessel DeepMind
Joseph Modayil DeepMind
Hado van Hasselt DeepMind
Tom Schaul DeepMind
Georg Ostrovski DeepMind

Will Dabney DeepMind
Dan Horgan DeepMind
Bilal Piot DeepMind
Mohammad Azar DeepMind
David Silver DeepMind

2017년 10월 Deepmind에서 Rainbow DQN을 발표

Extension to DQN

1.Double Q-learning.

2.Prioritized replay.

3.Dueling networks.

4.Multi-step learning.

5.Noisy Nets.

6.Distributional RL.

+ 7.A3C

참고 :

https://github.com/wonseokjung/wonseokjung.github.io/blob/master/_posts/2018-05-23-RL-Totherb7.md

https://wonseokjung.github.io//reinforcementlearning/update/RL-Totherb7/

Deep Q network

Multi-step learning

input

s

Double DQN ,
DIS

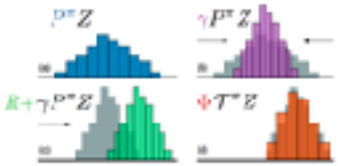
Noisy Q(s,a)

a

Env

s’ r

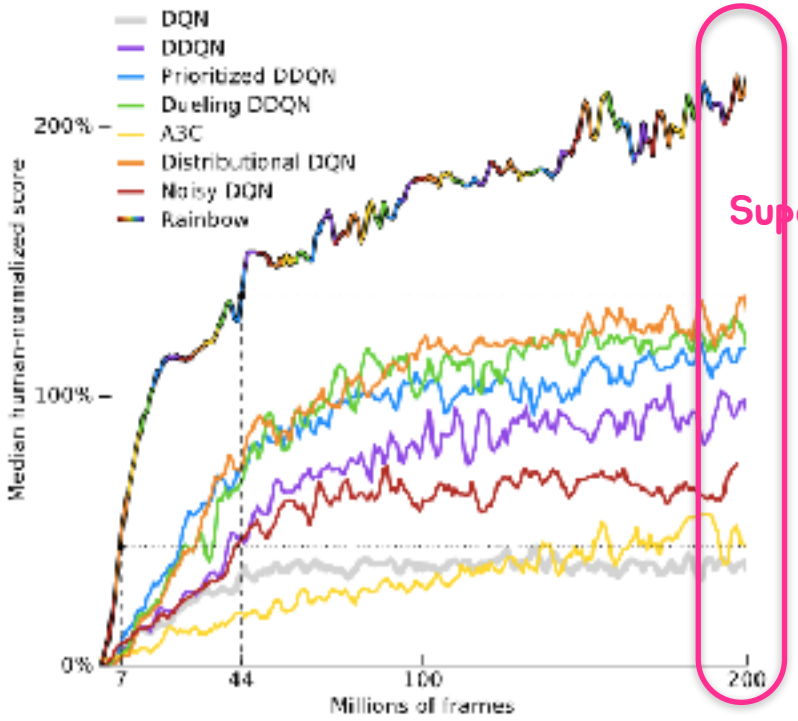
$$(R_{t+1} + \gamma_{t+1} q_{\bar{\theta}}(S_{t+1}, argmax_{a'} q(St) - q_{\theta}(S_t, A_t))^2$$



Prioritized
replay
Replay memory

학습이 더 필요한
transitions을
sample

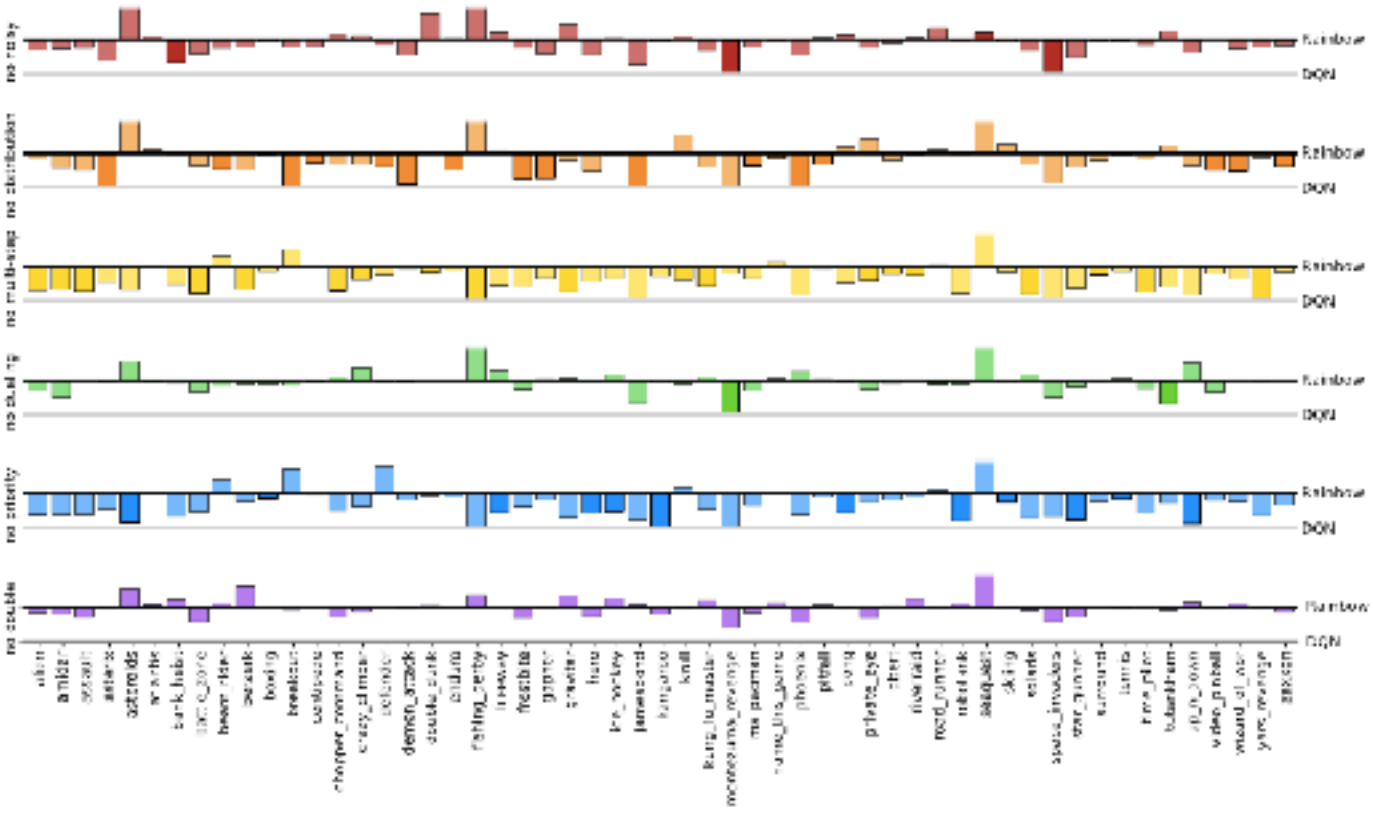
To the Rainbow-2



SuperMario에 적용한 DQN알고리즘
대비 엄청난 상승

DQN계열의 알고리즘6개와 A3C를 조합하여 만든 강화학습 알고리즘이다.

Atari환경에서의 성능비교



Rainbow를 사용하여 Sonic을 학습

Sonic -Rainbow DQN(with noisy network, epsilon =0)

https://contest.openai.com/videos/132.mp4

상위 10%로 OpenAI 대회 마무리!

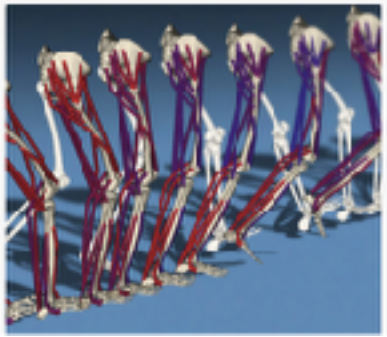
게임이 아닌 더 많은 action을 가진 agent도
강화학습으로 학습이 가능할까 ?

A.I Prosthetics

High dimensional state

Continuous actions

NIPS 2018 : AI for Prosthetics Challenge



NIPS 2018: AI for Prosthetics Challenge

Reinforcement learning with musculoskeletal models

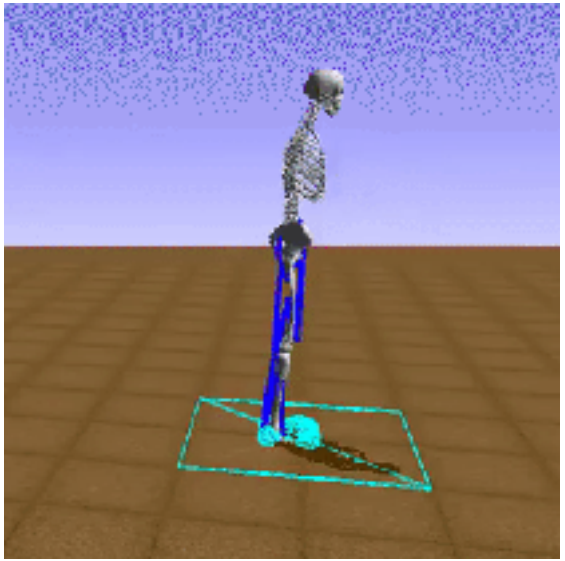
Stanford Neuromuscular Biomechanics Laboratory

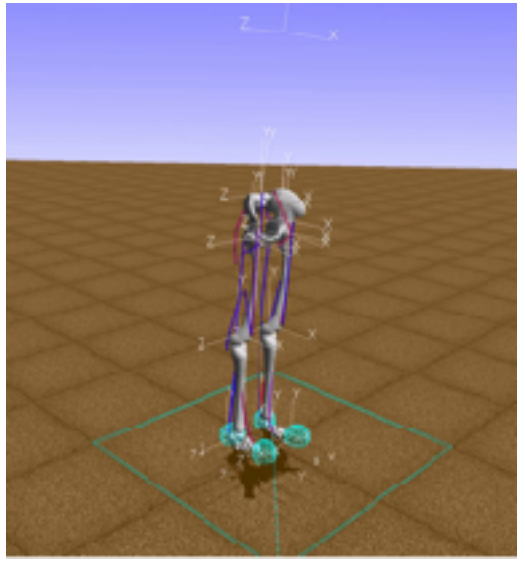
crowdAI

EPFL Digital EpidemiologyLab

84 days left

6674 Views | 82 Participants | 317 Submissions

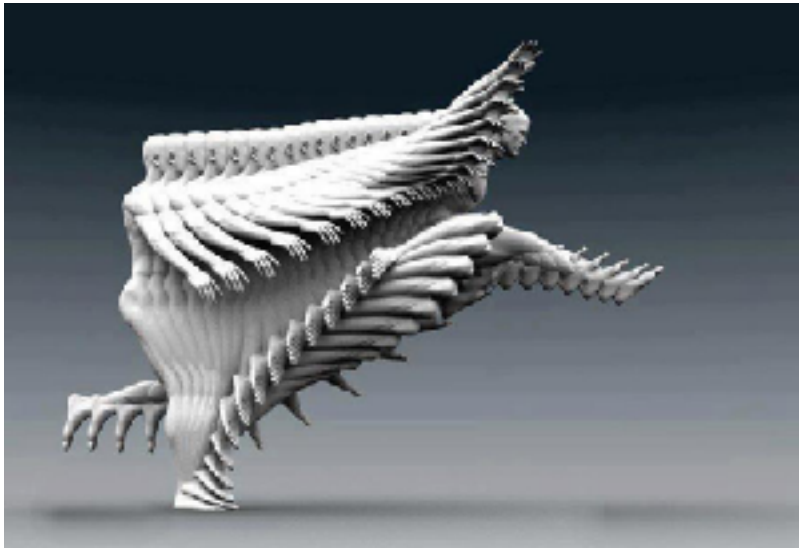




Action in Real world

Discrete Action

Continuos Action



https://twitter.com/iamruj

DQN solved High dimensional state, but not continues action

Two methods of choosing action

1. action-value :
   - Learning the action value
   - Estimate action value을 바탕으로 action을 선택한다.
   - Policies would not even exist without the action-value estimates

$$q^{\pi}(s,a) = E_{\pi}[G_t \mid S_t = s, A_t = a]$$

2. Parameterized policy :
   - select actions without consulting value function
   - Value function still be used to learn policy parameter
   - Value function이 action을 선택하는 기준으로 사용되지 않는다

$J(\theta)$: Performance measure

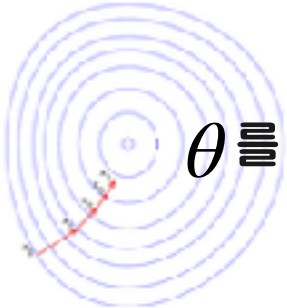
$$\theta_{t+1} = \theta_t + \alpha \widehat{\nabla J(\theta)}$$

Select action using PG Method

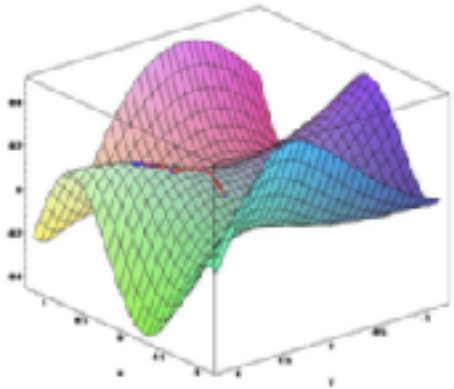
Discrete Action

https://www.cs.ubc.ca/~gberseth/blog/demystifying-the-many-deep-reinforcement-learning-algorithms.html

Continuos Action

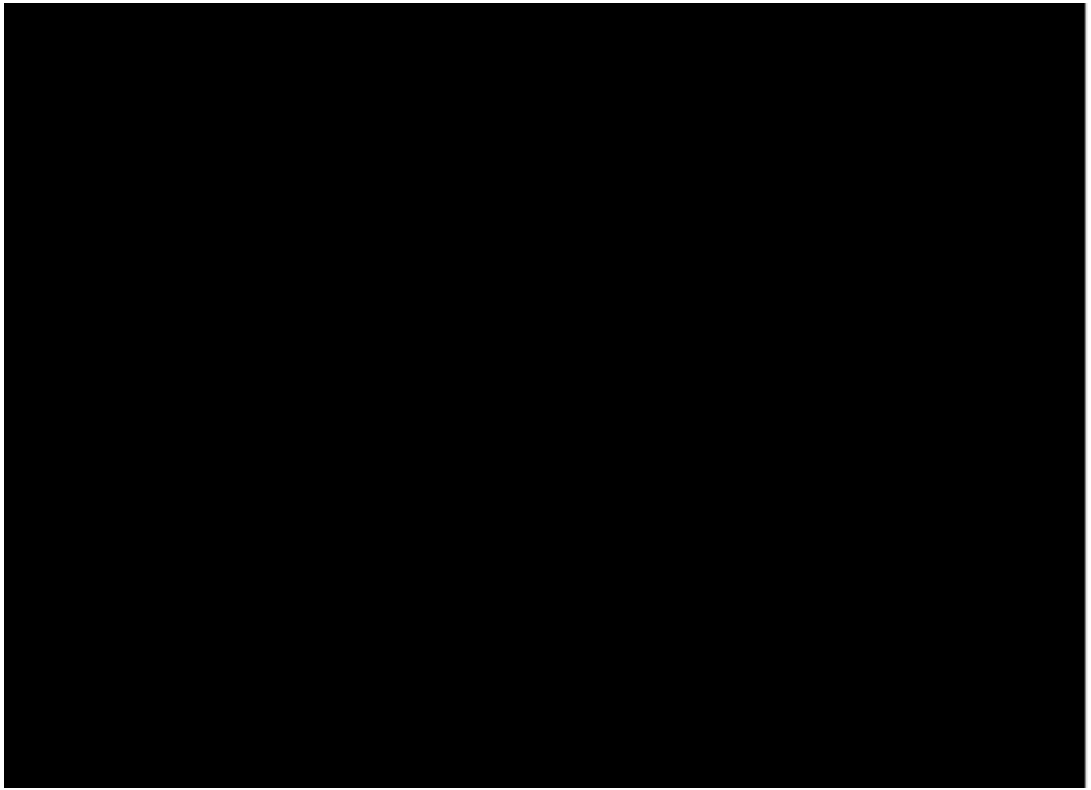


θ를 업데이트



http://www0.cs.ucl.ac.uk/staff/d.silver/web/Teaching_files/pg.pdf

Emergence of Locomotion Behaviours in Rich Environments



https://www.youtube.com/watch?v=hx_bgoTF7bs&t=98s

너무 웃기잖아...

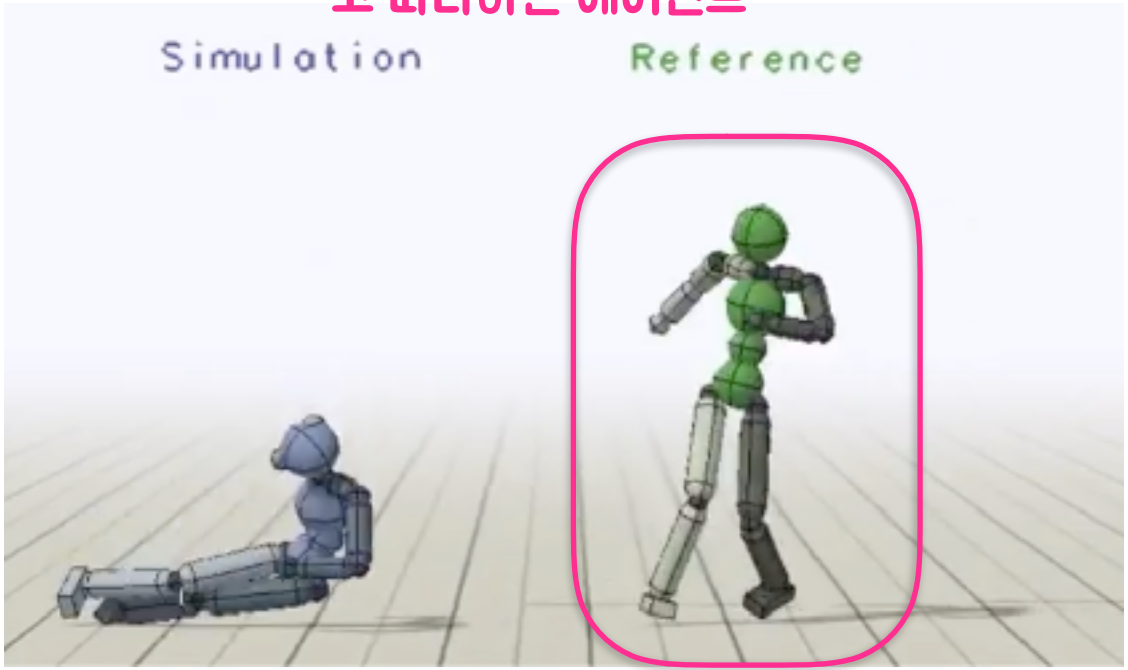
Community에 올라온 글중..

강화학습에서 풀어야할 문제들

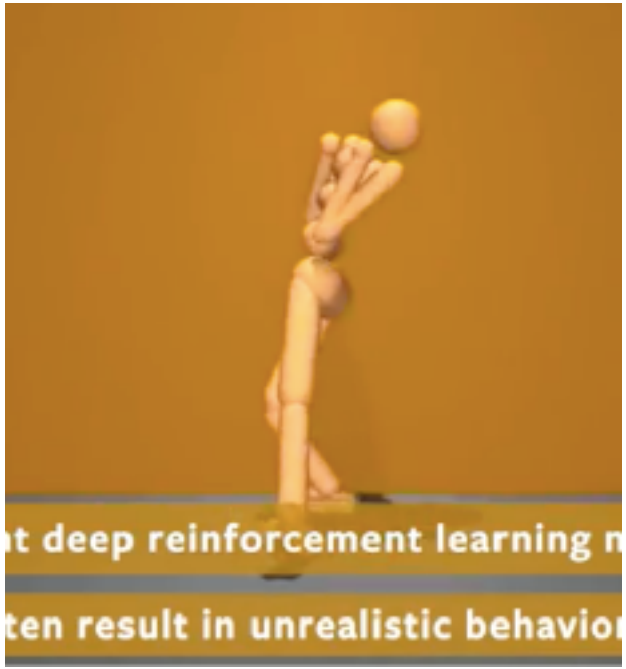
Latest trends

DeepMimic

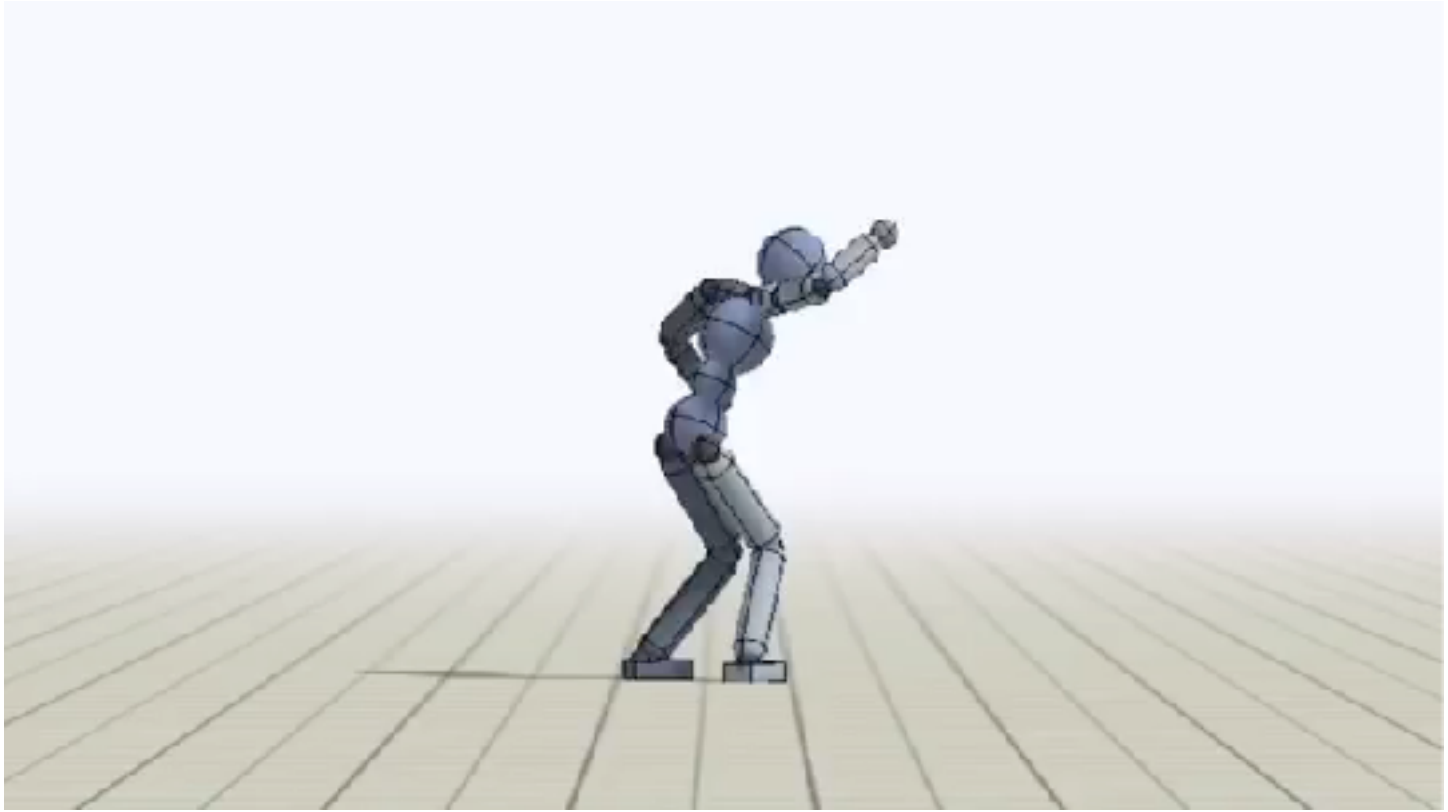
Reference Motion을 보고 따라하는 에이전트



더이상 과음은 하지 않도록..



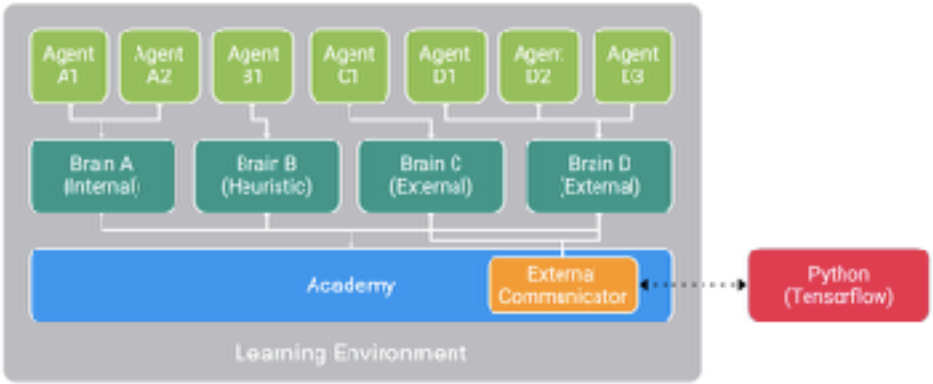
This virtual stuntman could improve video game physics



https://www.youtube.com/watch?v=XCLSkFKTWyg

이런 시뮬레이션 환경을 개인이 만들수 있을까?

Unity ml-agent

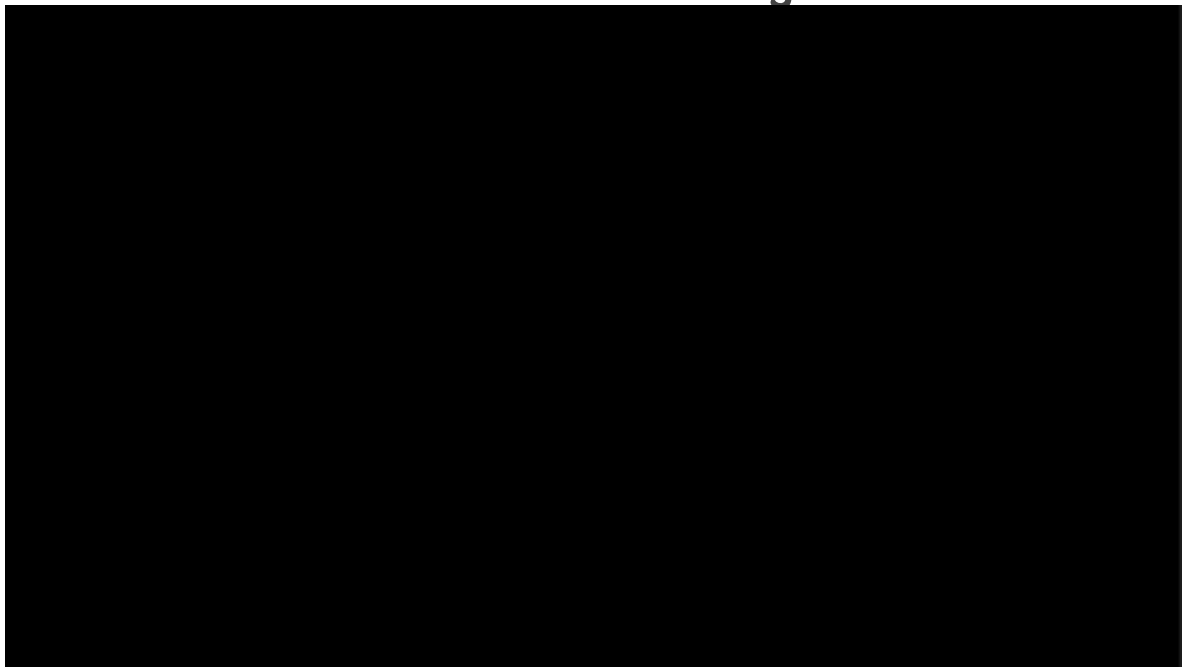


Unity Machine Learning Agents를 사용하여 개인이 환경을 제작하는 것이 가능

Unity ml-agent

사람이 플레이한것을
정답으로 학습

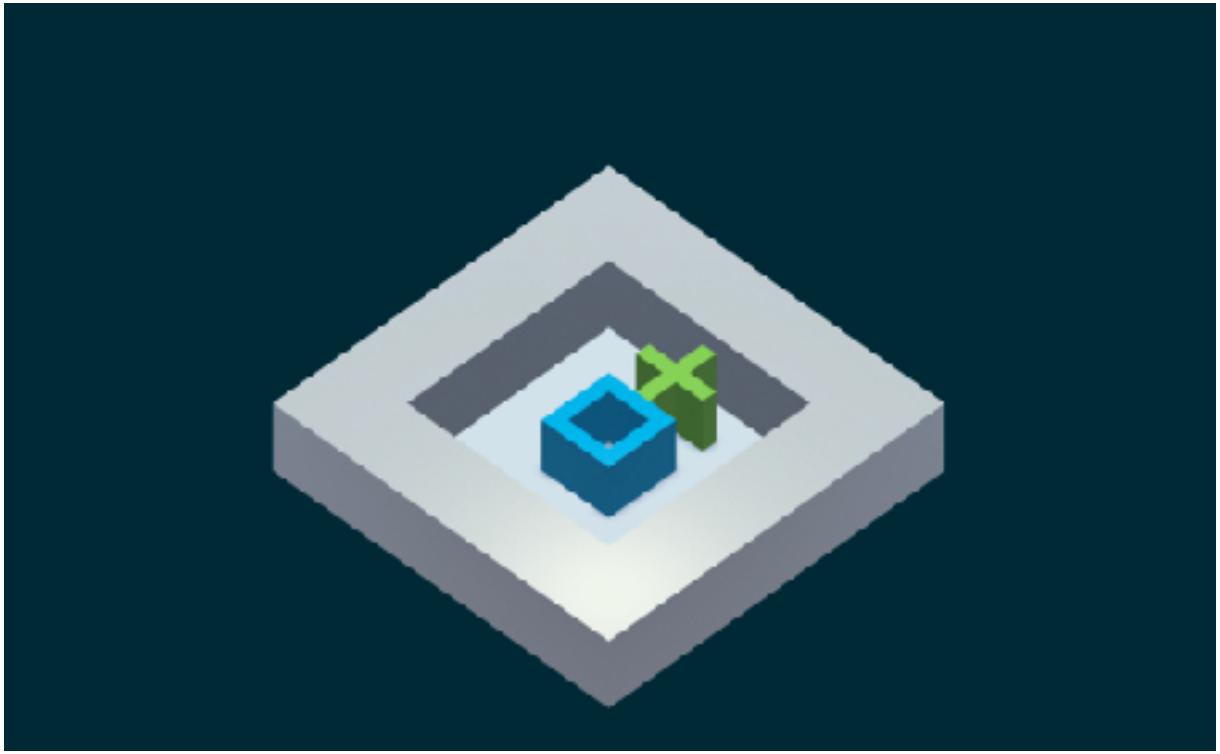
Imitation learning



https://www.youtube.com/watch?v=kpb8ZkMBFYs&feature=youtu.be

Unity ml-agent

Curriculum learning



https://youtu.be/vRPJAefVYEQ

Very easy

⇩

Easy

⇩

Medium

⇩

Hard

⇩

Very hard

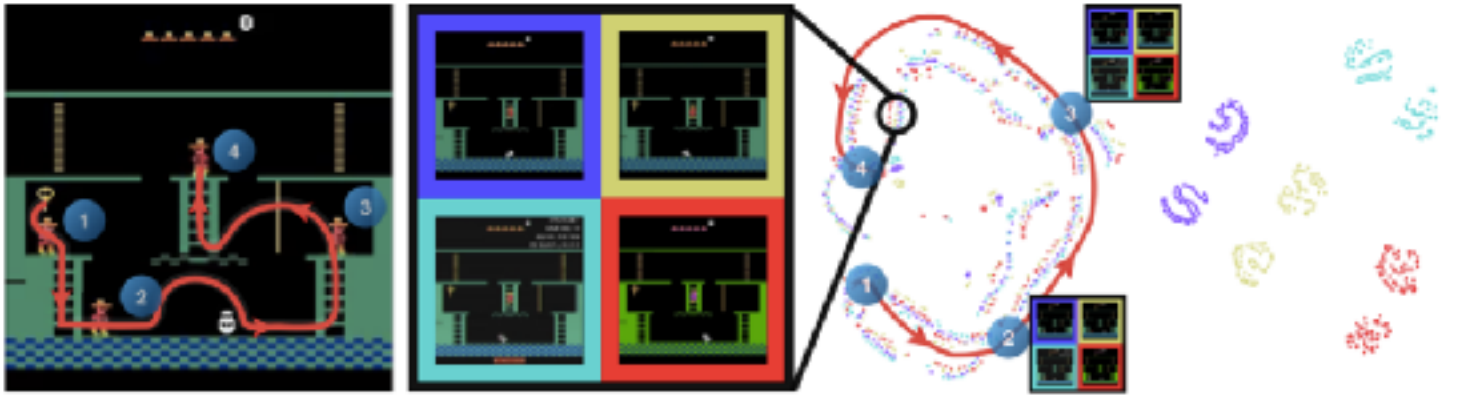
Exploration ?
Sparse Reward?

Exploration

Playing hard exploration games by watching YouTube

Yusuf Aytar*, Tobias Pfaff*, David Budden, Tom Le Paine, Ziyu Wang, Nando de Freitas

DeepMind, London, UK
{yusufaytar,tpfaff,budden,tpaine,ziyu,nandodefreitas}@google.com



Summary

1. Reinforcement Learning
2. Atari
3. SuperMario
4. Sonic
5. Prosthetics
6. Latest trend

감사합니다.

Github:
https://github.com/wonseokjung

Facebook:
https://www.facebook.com/ws.jung.798

Blog:
https://wonseokjung.github.io/