



Team 4

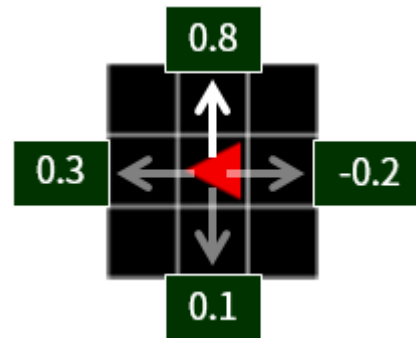
Project Summary



Deep Q-networks

Review-Q function

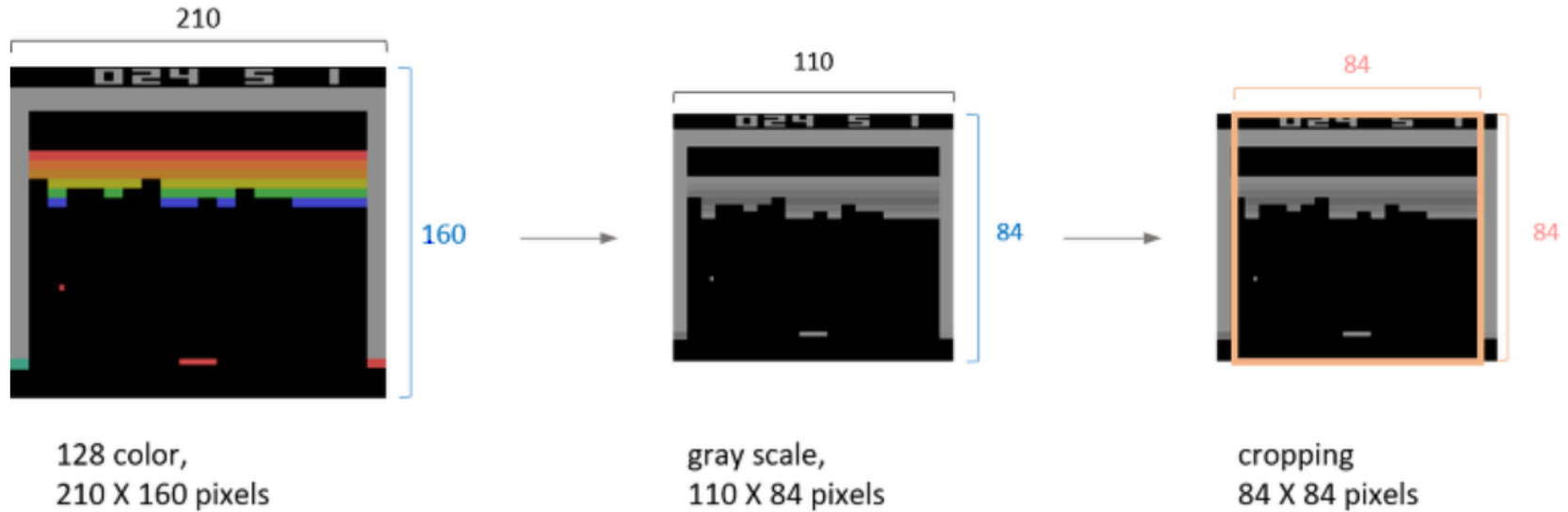
- ✓ 상태를 행동으로 치환하는 함수
- ✓ 각 상태에 대해 행동의 값을 각각 구하고,
가장 높은 Q값을 가진 행동을 선택하여
움직임
- ✓ Policy (정책)



Deep Q-network

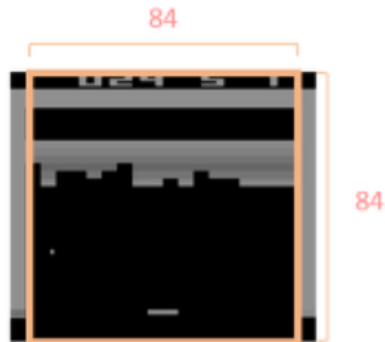
- ✓ 정의: convolutional networks를 이용하여
차원 축소하여 learning하는 방법
- ✓ Why?
CNN을 통해 픽셀을 받아와 차원 축소

Deep Q-network



Deep Q-network with 벽돌깨기

- ✓ 행동(action): 제자리 좌 우(발사) $\rightarrow a$
- ✓ 보상(reward): 벽돌 깨 때마다 점수를 받으며 위 층의 벽돌을 깨수록 더 큰 점수를 받음 $\rightarrow r$
- ✓ 상황(statement): 한 프레임에서의 environment $\rightarrow s$



Deep Q-network with 벽돌깨기

- ✓ Bellman equation := 최적의 규칙을 정의하는 함수

$$Q(s, a) = r + \gamma \max_{a'} Q(s', a')$$

r : 현재 상태의 최고의 보상

$\gamma \max Q(s', a')$: 다음 상태에서의 미래 보상의 최대값

γ : 감가율(미래의 보상 < 현재의 보상) ex) 0.7

Deep Q-network with 벽돌깨기

✓ 알고리즘

1. 먼저 $Q(s, a)$ 행렬을 랜덤값으로 초기화한다.
2. 초기 상태 s 에서 시작한다.
3. 아래 과정을 반복적으로 수행한다.

1. 액션 a 를 선택하여 실행한다.

2. 액션에 따른 보상 r 과 변경된 상태 s' 를 확인한다.

3. Q값을 갱신한다. $\rightarrow q(s, a) = q(s, a) + \alpha(r + \gamma \max_{a'} q(s', a') - q(s, a))$

4. 현재 상태를 변경한다

Deep Q-network with 벽돌깨기

- ✓ Q function

$$q(s, a) = q(s, a) + \alpha(r + \gamma \max_{a'} q(s', a') - q(s, a))$$

- ✓ Q function을 인공신경망으로 근사(차원의 문제)

$$q_{\theta}(s, a) = q_{\theta}(s, a) + \alpha(r + \gamma \max_{a'} q_{\theta}(s', a') - q_{\theta}(s, a))$$

감사합니다

벽돌깨기를 잘하려면?

공을 위로 올리세요~~~~~