

센서 노이즈 환경에서 입력 방식에 따른 강화학습 정책의 강인성 비교

이규석^{0,1}, 유리¹

아주대학교 소프트웨어학과¹

lgs8106@ajou.ac.kr, riyu@ajou.ac.kr

A Comparative Study on the Robustness of Reinforcement Learning Policies for Different Inputs under Sensor Noise

Gyu-Seok Yi^{0,1}, Ri Yu¹

Dept. of Software and Computer Engineering, Ajou University¹

요약

시뮬레이션에서 강화학습으로 학습된 로봇 정책을 현실 세계에 적용하는 Sim-to-Real 연구는 환경의 변화나 센서의 측정 오류와 같이, 시뮬레이션에는 없는 노이즈로 인해 성능의 저하를 겪는다. 본 연구는 이러한 현실 간극을 줄이기 위한 기초 연구로서, 어떤 시각 입력 방식이 여러 센서 노이즈에 더 강인한 정책을 생성하는지 탐구한다. 이를 위해 시뮬레이션 환경에서 로봇 팔 큐브 쌓기 과제를 대상으로, RGB 입력과 Depth 입력을 사용하는 강화학습 정책을 각각 학습시켰다. 이후, 가우시안 노이즈, 핵심 누락 같은 실제 센서가 겪을 수 있는 다양한 노이즈를 시뮬레이션에 적용하여 얻은 결과로 각 정책의 강인성을 비교 분석한다. 실험 결과, 이상적인 환경에서는 Depth 정책이 RGB 정책보다 높은 성능을 보였지만, 노이즈 환경에서는 RGB 정책이 Depth 정책보다 더 적은 성능 하락을 보여 높은 강인성을 나타냈다. 이는 실제 운용 환경의 노이즈 유형을 고려하여 시각 입력 방식을 선택하는 것이 Sim-to-Real 성공률을 높이는 요소임을 시사한다.

1. 서론

심층 강화학습은 복잡한 로봇 제어 문제를 해결하는 강력한 도구로 부상했으며, 특히 물리 시뮬레이션 환경에서 성과를 보여주었다. 이러한 성공에 힘입어 시뮬레이션에서 학습한 정책을 실제 로봇에 이전하는 Sim-to-Real 연구[1]가 활발히 진행되고 있는데, 시뮬레이션의 완벽한 환경과 달리, 실제 로봇이 마주하는 현실 세계는 예측 불가능한 변수가 많아 시뮬레이션에서 학습된 정책이 제대로 성능을 발휘하지 못하는 현실 간극 문제가 난관으로 남아있다.

이러한 현실 간극을 유발하는 원인 중 하나는 시각 센서 데이터의 불완전성이다. 실제 환경의 카메라는 조명, 그림자, 반사, 핵심 누락 등 정보를 정확히 측정하지 못하는 경우가 발생한다. 이처럼 각 시각 입력 방식은 고유한 노이즈 특성을 가지며, 이는 정책 안정성과 강인성에 직접적인 영향을 미친다.

따라서 본 논문에서는 현실 간극의 문제를 완화하기 위한 연구로서, 어떤 시각 입력 방식이 현실적인 센서 노이즈에 더 강인한 정책을 생성하는지 탐구하고자 한다. 이를 위해 시뮬레이터에서 로봇 팔의 큐브 쌓기 과제를 대상으로 RGB 입력과 Depth 입력을 사용하는 강화학습 정책을 각각 학습시키고, 이후 실제 센서가 겪을 수 있는 가우시안 노이즈, 핵심 누락 같은 노이즈를 시뮬레이션에 적용하여 각 정책의 성능 변화를 비교 분석한다.

2. 실험 설계

본 연구는 센서 노이즈 환경에서 강화학습 정책의 강인성이 시각 입력 방식에 따라 어떻게 달라지는지 비교 분석하고자 한다. 이를 위해 그림 2의 입력부에 해당하는 RGB와 Depth 입력을 각각 사용하는 두 가지 경우로 나누어 정책을 학습하고, 노이즈 유무에 따른 성능 변화를 측정하는 실험을 다음과 같이 설계했다.

2.1. 시뮬레이션 환경 및 과제

효율적인 병렬 학습을 위해 NVIDIA Isaac Gym[2] 시뮬레이터를 사용하였으며, 128개의 병렬 환경에서 학습을 동시에 진행했다. 실험 과제는 Franka Emika Panda 로봇 팔을 이용하여 테이블 위의 한 큐브를 다른 큐브 위에 쌓는 큐브 쌓기로 설정하였다 (그림 1 참조).

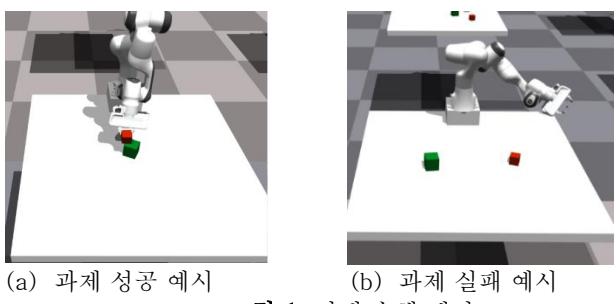


그림 1: 과제 수행 예시

2.2. 정책 학습 모델

정책 학습을 위해 Actor-Critic 모델을 사용하였으며, PPO(Proximal Policy Optimization)[3] 알고리즘을 통해 정책을 최적화하였다. 네트워크는 시각 특징을 추출하는 CNN을 액터와 크리틱이 공유하고, 이후 각자의 역할을

* 구두발표논문

* 학부생 주제자 논문임

* 본 연구는 2025년 과학기술정보통신부 및 정보통신기획평가원의 SW중심대학사업의 연구결과로 수행되었음(2022-0-01077)

수행하는 두 개의 MLP로 나뉘는 구조를 가진다.

액터 네트워크: 96x96 해상도의 RGB 또는 Depth를 입력으로 받아, 합성곱 계층으로 구성된 CNN을 통해 특징을 추출한다. MLP 계층을 거쳐 로봇의 7개 관절과 2개의 그리퍼를 제어하는 9차원의 연속 행동을 출력한다. 크리틱 네트워크: 공유된 CNN의 시각 특징과 더불어, 그림2의 State에 해당하는 로봇 관절 및 큐브의 위치와 회전 정보를 입력으로 받아 현재 상태의 가치를 평가하는 단일 스칼라 값을 출력한다.

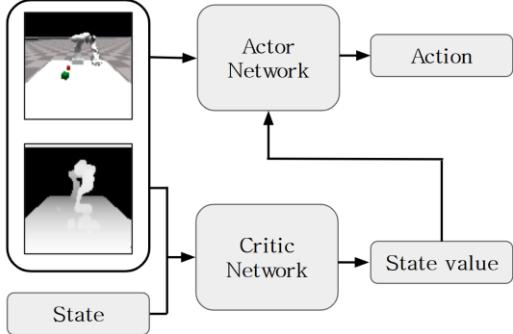


그림 2: 네트워크 구조

2.3. 센서 노이즈 모델링

현실 세계의 불완전한 센서 정보를 모사하기 위해 다음과 같은 노이즈 모델을 각 시각 입력에 적용하였다.

가우시안 노이즈: 일반적인 센서 측정 오류를 모사하기 위해 각 픽셀 값에 정규분포를 따르는 무작위 값을 추가하였다.

픽셀 누락: 센서 정보가 소실되는 현상을 모사하기 위해 특정 비율의 픽셀 값을 임의로 제거하였다.

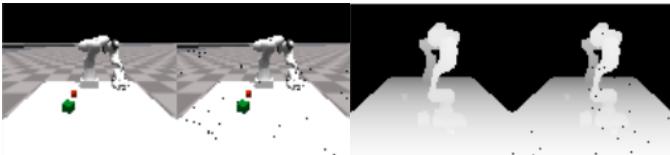


그림 3: 픽셀누락 적용 예시

3. 실험 및 결과

3.1. 강인성 평가 방법

정책의 강인성 측정을 위해, 노이즈가 없는 기준 환경에서 학습된 정책의 성능을 평가한다. 이후, 동일한 정책을 노이즈가 적용된 환경에서 테스트를 진행하여 성능의 변화를 관찰한다. 정책의 성능은 10스텝 이상 진행된 유효 에피소드에서 약 500개의 샘플을 수집하여 평균적인 보상을 계산하였고, 강인성은 기준 대비 성능 저하율로 측정하였다.

3.2. 보상 측정 방법

로봇 팔이 과제를 얼마나 성공적으로 수행하는지에 따라 여러 요소로 구성된다. 거리(그리퍼와 큐브 사이 거리에 반비례), 들어올리기(큐브가 일정 높이 이상 올라갔는지), 정렬(큐브 A가 큐브 B위에 정렬되었는지), 쌓기(큐브 A가 큐브 B위에 성공적으로 쌓였는지) 등을 고려하여 보상이 결정된다.

3.3. 실험 결과

입력 방식	환경	평균 보상	성능 저하율
RGB	기준	11.59	-
	픽셀누락(0.5%)	3.69	68.2%
	픽셀누락(0.25%)	7.31	36.9%
	가우시안(0.02)	2.87	75.2%
	가우시안(0.01)	6.02	48.1%
Depth	기준	13.97	-
	픽셀누락(0.5%)	2.05	85.3%
	픽셀누락(0.25%)	3.76	73.1%
	가우시안(0.02)	1.98	85.8%
	가우시안(0.01)	2.49	82.2%

표 1: 입력 방식 및 노이즈에 따른 실험 결과

표 1에서 ‘픽셀누락’의 팔호 안 숫자는 임의로 제거된 픽셀의 비율을 의미하며, ‘가우시안’의 팔호 안 숫자는 추가된 노이즈 정규분포의 표준편차를 의미한다.

4. 결론 및 한계점

Sim-to-Real 전환 과정에서 발생하는 현실 간극 문제에 대해, 본 연구는 RGB와 Depth 입력 방식의 강인성을 다양한 센서 노이즈 환경에서 비교 분석하였다. 실험 결과(표 1 참조), 노이즈가 없는 이상적인 환경에서는 기하학적 정보가 풍부한 Depth 정책의 성능(13.97)이 RGB 정책의 성능(11.59)보다 높았으나, 가우시안 노이즈나 정보 소실이 있는 픽셀 누락 같은 환경에서는 RGB 정책이 성능 저하율이 36.9%~75.2%인 반면, Depth 정책은 73.1%~85.8%에 달하는 성능 저하율을 보였다. 이는 특정 로봇 응용 분야에 센서를 선택할 때, 단순히 최고 성능에만 집중하기보다, 실제 운용 환경에서 예상되는 주요 노이즈 유형을 분석하고 이에 강인한 시각 입력 방식을 선택하는 것이 Sim-to-Real 성공률을 높이는 핵심 요소임을 시사한다. 특히, Depth 센서가 제공하는 3D 정보의 이점에도 불구하고, 실제 환경의 복잡한 노이즈 특성에 대한 민감성을 고려할 때, RGB 센서의 상대적인 노이즈 강인성이 특정 시나리오에서는 더 실용적인 대안이 될 수 있다.

하지만 본 연구는 다음과 같은 한계점을 가진다. 연산의 효율을 위해 학습에 96x96 저해상도 이미지를 사용하였다. 이로 인해 고해상도 이미지에서 나타날 수 있는 미세한 노이즈 영향이 특징 추출의 차이를 충분히 반영하지 못했을 수 있다. 또한 큐브 쌓기라는 제한된 환경에서의 단일 과제에 대해서만 실험이 이루어졌기 때문에 결과의 일반화에 한계가 있다.

참고문헌

- [1] J. Tobin, R. Fong, A. Ray, J. Schneider, W. Zaremba, and P. Abbeel. 2017. Domain Randomization for Transferring Deep Neural Networks from Simulation to the Real World. *arXiv preprint arXiv:1703.06907*, 2017
- [2] V. Makoviychuk et al. 2021. Isaac Gym: High Performance GPU-Based Physics Simulation For Robot Learning. *arXiv preprint arXiv:2108.10470*, 2021
- [3] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov. 2017. Proximal Policy Optimization Algorithms. *arXiv preprint arXiv:1707.06347*, 2017