# 시뮬레이션 환경에서의 로봇팔 이동 정책 학습과 도메인 랜덤화를 통한 실제 로봇 전이에 관한 연구

정하륜, 오현수, 김재호*

세종대학교

haryun.sejong@gmail.com, hyeonsu.sejong@gmail.com, *kimjh@sejong.ac.kr

# Study on Robotic Arm Motion Policy Learning in Simulation and Transfer to Real Robots Through Domain Randomization

Jeong Ha Ryun, Oh Hyeon Su, Kim Jae Ho*

Sejong Univ.

요 약

본 논문에서는 가상 환경 물리 시뮬레이터인 IsaacSim에서 PPO알고리즘을 사용하여 6관절 로봇팔인 Kinova Gen3의 목표점 이동 정책을 학습시켰고, 물리 매개변수를 무작위로 바꾸어가며 학습하는 방법인 도메인 랜덤화를 적용하여 학습된 정책과 도메인 랜덤화를 적용하지 않고 학습된 정책을 각각 실제 로봇에 전이했을 때의 동작 결과를 비교하고 분석한다. 실험 결과 도메인 랜덤화 적용 정책이 평균오차 0.02095m, 도메인 랜덤화 미적용 정책이 평균오차 0.04695m를 기록하여 도메인 랜덤화의 유효성을 입증하였다. 본 연구는 가상 환경 시뮬레이션 기반 정책의 실제 환경 전이 가능성을 높이는 방법을 검증하며, 향후 동적 장애물 회피 및 복합 작업으로의 확장 연구 방향을 제시한다.

## Ⅰ. 서 론

최근 로봇의 정책을 가상 환경에서 학습한 후 실제 환경에 전이하는 'Sim2Real' 기법이 활발히 연구되고 있다[1]. 해당 기법은 실제 환경에서 데이터를 수집하며 로봇의 정책을 학습시키는 방법에 비해 시간, 공간, 비용적 부담을 크게 절감한다[2]. 이 기법은 IsaacSim[3]과 같은 고정밀 물리, 광학 엔진을 통합한 시뮬레이터를 통해 가능하며 실제 환경과 유사한 조건으로 가상 환경에서 로봇 정책 학습을 할 수 있다.

그러나 가상 환경이 정교하더라도 실제 환경과 중력, 마찰 등 물리적 특성이 정확히 일치할 수 없다. 그리고 관절 모터의 측정 오차 그리고 링크의 질량, 무게중심 등 로봇의 물리적 특성 또한 정확히 일치할 수 없다. 또한 IsaacSim의 가상 로봇과 실제 로봇은 제어방식에도 차이가 존재한다. 이러한 차이로 인해 가상 환경 물리 시뮬레이션을 기반으로 학습된 정책을 실제 환경에 전이하면 시뮬레이션에서 현실로 전이되는 과정에서 물리, 제어, 센싱 차이로 성능이 저하되는 현상인 'Sim2Real Gap'이 발생한다[4]. 이를 해결하기 위한 방법으로 가상 환경의 물리적 특성을 다양화하여 에이전트가 다양한 물리 환경을 경험하도록 만드는 도메인 랜덤화 기법이 제안되었다[5]. 이 기법을 적용하면 학습된 정책이 특정 환경에 과적합되지 않고 다양한 환경에서 안정적으로 동작하도록 일반화할 수 있다.

본 논문에서는 IsaacSim에서 물리 시뮬레이션을 기반으로 학습된 도메인 랜덤화 적용 정책과 도메인 랜덤화 미적용 정책을 각각 실제 6관절 Kinova Gen3 로봇팔에 전이하여 목표점 이동 동작을 반복 수행한 후 두 정책의 동작 결과를 비교하고 분석함으로써 도메인 랜덤화의 효과를 실험적으로 검증한다.

## Ⅱ. 본론

가상 환경에서 학습된 로봇팔 이동 정책을 실제 로봇에 적용한 결과, 동작 정확도가 낮았다. 조사 결과 가상 시뮬레이션과 현실 간의 물리, 제어, 센싱 등의 매개변수 불일치로 인한 Sim2Real Gap 때문이었다. 본 논문에서는 이를 해소하기 위해 학습 단계에서 물리 매개변수를 일정 범위 내에서 무작위화하는 도메인 랜덤화를 도입하고, 그 효과를 정량적으로 분석한다. 실험에서는 가상 환경 IsaacSim에서 PPO알고리즘을 사용하여 6관절 로봇팔인 Kinova Gen3의 목표점 이동 정책을 학습하며, 매 에피소드가 시작될 때 로봇팔의 물리 매개변수를 다양화하는 방식으로 도메인 랜덤화를 진행하였다.

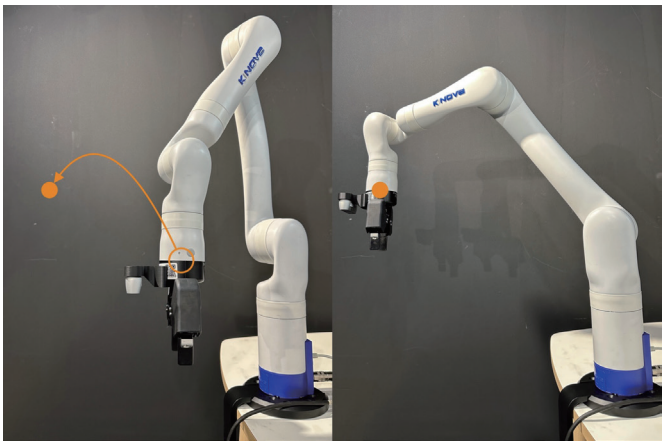

그림 1. 로봇팔의 목표점 이동 동작

IsaacSim에서는 정책을 병렬적으로 학습할 수 있다. 병렬환경의 갯수는 4096개인 것이 하드웨어 활용도와 수렴 속도 면에서 가장 효율적임을 확인하였다[6]. 따라서 본 연구는 4096개의 병렬환경으로 정책 학습을 진행하였고, 절차는 다음과 같다. 에피소드가 시작될 때 로봇팔의 물리 매개변수인 링크의 질량과 관절의 마찰, 강성, 감쇠를 독립적으로 무작위화하고

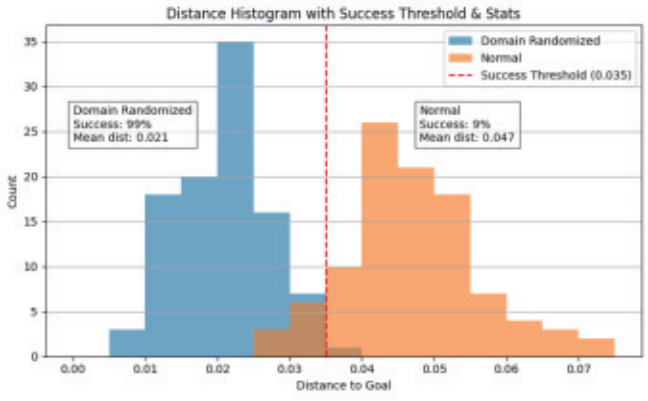


**그림 2. 도메인 랜덤화 적용 정책과 미적용 정책의 동작 수행 결과 비교 히스토그램**

목표점이 작업 공간에서 무작위로 선정된다. 이때 무작위 범위는 실험적으로 실제값의 0.7배에서 1.3배 사이로 설정하였다. 첫 번째 에피소드에서는 로봇팔이 초기 자세로 시작하고 이후 에피소드부터는 동작 성공 시 이전 에피소드의 마지막 자세, 동작 실패 시 초기 자세로 시작한다.

에이전트가 받는 주요 보상은 수식 (1)과 (2)와 같이 엔드이펙터와 목표점 사이의 거리와 엔드이펙터 링크의 정렬도로 구성하여 엔드이펙터 링크가 지면과 수직이고 베이스 링크 기준 정면을 바라보는 자세로 목표점에 도달하도록 유도하였다. [그림 1]은 목표점 이동 동작 시 유도한 엔드 이펙터 링크의 자세로 목표점에 도달한 로봇팔의 모습이다.

$$reward_d = \lambda_d\left(e^{\alpha_d(d+\beta_d)} + \gamma_d\right) \quad (1)$$

$$penalty_s = \lambda_s(\log_{10} + \alpha_s(s+\beta_s)) + \gamma_s) \quad (2)$$

이때 $d$는 엔드 이펙터와 목표점 사이의 거리를 최대 거리로 나누어 [0, 1]구간으로 정규화한 것이고 $\lambda_d$는 1.0, $\alpha_d$는 -2.3, $\beta_d$는 -0.176, $\gamma_d$는 -0.15로 실험적으로 설정하였다. $s$는 엔드 이펙터 링크의 TF와 전역 TF의 코사인 유사도를 이용하여 정렬도를 계산한 뒤 [0, 1] 구간으로 정규화한 것이고 $\lambda_s$는 1.0, $\alpha_s$는 -0.7, $\beta_s$는 -1.143, $\gamma_s$는 1.0 으로 실험적으로 설정하였다.

여기에 보조 보상으로 수식 (3)과 (4)와 같이 행동의 크기와 스텝 별 자세의 변화량을 패널티로 추가하여 움직임과 떨림을 억제하였다.

$$penalty_a = \lambda_a \sum_{i=1}^{6} a_i^2 \quad (3)$$

$$penalty_p = \lambda_p \sum_{i=1}^{6} (a_i - a_i')^2 \quad (4)$$

이때 $\lambda_a$는 0.02, $\lambda_p$는 0.07 으로 실험적으로 설정하였다. $i$는 관절의 번호이다. $a_i$는 현재 스텝의 행동 벡터이고, $a_i'$는 이전 스텝의 행동 벡터이다. 따라서 최종 보상함수는 수식 (5)가 된다.

$$reward = reward_d - \sum_{x \in \{s,a,p\}} penalty_x \quad (5)$$

실제 로봇으로 정책을 전이하여 100회의 목표점 이동 동작을 실험한 결과 도메인 랜덤화 적용 정책은 99.0%의 성공률과 0.02095m의 평균오차를 기록하였으며 도메인 랜덤화 미적용 정책은 9.0% 성공률과 0.04695m의 평균오차를 기록하였다. [그림 2]는 도메인 랜덤화 적용 여부에 따른 두 정책의 성공률과 오차를 시각적으로 비교한다. 두 정책을 비교하면 전자의 성공률은 후자의 11배이고 평균오차는 후자에 비해 55.4% 감소하여 향상된 성능을 보였다. 이는 정책이 학습 단계에서 다양한 물리 환경을 경험하여 일반화 능력을 확보한 덕분으로 해석된다. 이러한 결과는 가상 환경 시뮬레이션을 기반으로 학습된 정책을 실제 환경에 성공적으로 전이하기 위해서는 도메인 랜덤화가 필수적임을 보인다.

## Ⅲ. 결론

본 논문에서는 가상 환경 물리 시뮬레이션으로 학습된 정책을 실제 환경에 전이할 때 발생하는 성능 저하, 즉 Sim2Real Gap 문제의 해결책인 도메인 랜덤화 기법의 효과를 실험적으로 검증하였다. 이를 위해 로봇의 물리적 특성인 링크의 질량과 관절의 마찰, 강성, 감쇠를 독립적으로 바꾸어 가며 학습을 진행하였다. 정책을 실제 로봇에 전이하여 목표점 이동 동작을 100회 반복 시행한 실험에서 도메인 랜덤화 적용 정책의 평균 오차는 0.02095m로 도메인 랜덤화 미적용 정책의 평균오차인 0.04695m보다 55.4% 낮았다. 이는 학습 단계에서 다양한 물리 환경을 경험한 정책이 실제 환경에서도 안정적으로 동작할 수 있음을 의미하며, 도메인 랜덤화 기법의 유효성을 입증한다. 결과적으로 가상 환경 물리 시뮬레이션을 기반으로 학습된 정책을 실제 환경에 전이하기 위해서는 도메인 랜덤화가 필수적임을 보여준다. 향후 연구에서는 본 논문의 결과를 기반으로 동적 장애물이 존재하는 환경에서의 장애물 회피, 물체 운반 및 조립과 같은 보다 복합적인 작업으로 과업을 확장하여, 시뮬레이션 기반 학습의 실용성을 심층적으로 분석할 계획이다.

## ACKNOWLEDGMENT

이 논문은 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원-대학ICT연구센터(ITRC)의 지원(IITP-2025-RS-2021-II211816)과 산업통상자원부 및 산업기술평가관리원(KEIT)의 지원(RS-2022-00154678)과 과학기술정보통신부 및 정보통신기획평가원의 정보통신방송혁신인재양성(메타버스융합대학원)사업 연구 결과로 수행되었음(IITP-2025-RS-2023-00254529)

## 참 고 문 헌

[1] Radosavovic, Ilija, et al. "Real-world humanoid locomotion with reinforcement learning."Science Robotics9.89 (2024): eadi9579.

[2] Zhang, Xiang, et al. "Efficient sim-to-real transfer of contact-rich manipulation skills with online admittance residual learning."Conference on Robot Learning. PMLR, 2023.

[3] Zhou, Zhehua, et al. "Towards building AI-CPS with NVIDIA Isaac Sim: An industrial benchmark and case study for robotics manipulation."Proceedings of the 46th international conference on software engineering: software engineering in practice. 2024.

[4] Salvato, Erica, et al. "Crossing the reality gap: A survey on sim-to-real transferability of robot controllers in reinforcement learning."IEEE Access9 (2021): 153171-153187.

[5] Chen, Xiaoyu, et al. "Understanding domain randomization for sim-to-real transfer."arXiv preprint arXiv:2110.03239(2021).

[6] Rudin, Nikita, et al. "Learning to walk in minutes using massively parallel deep reinforcement learning."Conference on Robot Learning. PMLR, 2022.