

인간-로봇 물리적 상호작용을 위한 센서리스 휴머노이드의 강화학습 기반 외력 순응 보행 — Sim-to-Real 외력 분산 보상

Reinforcement Learning-Based Compliant Locomotion for Sensorless Humanoids in Physical Human-Robot Interaction with Sim-to-Real Force Dispersion Compensation

박 연 수¹ · 김 건 우² · 박 성 관³ · 이 창 주² · 임 성 수[†]

Yeonsoo Park¹, Gunwoo Kim², Sung Gwan Park³, Chang Joo Lee², Sungsoo Rhim[†]

Abstract: Humanoid robots operating in human environments must safely respond to physical human-robot interactions while maintaining stable locomotion. Conventional reinforcement learning-based locomotion policies primarily focus on velocity tracking and typically treat external forces as disturbances, which can lead to motor oscillations and unstable behaviors under continuous physical contact. This study proposes a force-compliant locomotion framework for humanoid robot without dedicated force/torque sensors using body-wrench information and reinforcement learning based on the PPO algorithm. During training, external forces are estimated from privileged body-wrench information provided by physics-engine in IsaacLab simulation environment and a ground-reaction-force removal algorithm is applied to improve estimation robustness. To reduce the Sim-to-Real Gap caused by force dispersion during human-robot interaction, an effective force model and a compliance velocity generation method are introduced. The proposed reward structure prioritizes force compliance over velocity tracking when external forces are detected. Simulation results in IsaacLab and real-world experiments on a Unitree G1 humanoid robot indicate that the proposed policy can generate force-compliant locomotion without dedicated force sensors.

Keywords: Humanoid, Reinforcement Learning, Simulation, Locomotion, Force Estimation, Physical Human-Robot Interaction (pHRI)

1. 서 론

최근 강화학습과 Whole-Body Control 기술의 발전으로 휴머노이드 로봇은 안정적인 보행을 넘어 물체 조작, 물류 작업 등 다양한 작업을 수행할 수 있게 되었다. 특히 산업 현장과 인간 생활 공간에서 휴머노이드 활용이 확대됨에 따라 로봇이 사람과 동일한 공간에서 인간-로봇 상호작용을 위해 안전하게 동작하는 능력이 요구되고 있다.

대부분의 휴머노이드 보행 제어 알고리즘은 목표 속도나 목표 자세 추종을 중심으로 설계되어 있다^[1,2]. 따라서 외부 접촉이나 충돌에 의해 발생하는 외력은 추종 성능을 저해하는 외란으로 취급되며, 원래의 명령을 유지하기 위해 외력의 영향을 억제하도록 동작한다. 이러한 방식은 외력에 의해 발생하는 편차를 최소화하는 데에는 효과적이지만, 지속적인 접촉이 유지되거나 사람이 로봇을 직접 밀거나 당기는 협업 작업 중에는 여러 문제가 발생한다. 외력에 저항하기 위해 과도한 관절 토크가 발생하거나, 모터 진동 및 자세 불안정 현상이 나타날 수 있어 사람이 상해를 입거나 로봇이 파손되는 등 안전성이 저하될 수 있다.

이러한 문제를 해결하기 위해 Impedance Control^[3], Admittance Control^[4], Force/Torque Sensor (F/T Sensor) 기반 Compliance Control^[5], 외란 관측기 기반 순응 제어^[6], 물리적 상호작용 기반 보행 제어^[7] 등이 제안되었다. 최근에는 강화학습을 이용하여

Received : Jun. 15. 2026; Revised : Aug. 12. 2026; Accepted : Aug. 14. 2026

1. M.S. Student, Department of Mechanical Engineering, Kyung Hee University, Yongin, Korea (ysally77@khu.ac.kr)

2. Researcher, Safetics, Inc., Seoul, Korea (gwkim, cilee@safetics.io)

3. Ph.D. Student, Department of Mechanical Engineering, Kyung Hee University, Yongin, Korea (qkrtjdrhks21@khu.ac.kr)

† Professor, Corresponding author: Mechanical Engineering, Kyung Hee University, Yongin, Korea (ssrhim@khu.ac.kr)

외력에 적응적인 보행정책을 생성하거나 힘과 위치를 통합하여 정책을 학습하는 연구도 수행되고 있다^[8-11]. 신호 조건부 다목적 강화학습을 이용하여 보행 명령 추종과 외력 순응을 동시에 최적화하고^[12], 외력 추정 정보를 활용하여 지속적인 외란 환경에서의 휴머노이드 보행 안정성을 향상시키는 선행연구^[13]도 있다. 한편 강화학습으로 순응 거동을 학습하는 연구도 제안되었다. 예를 들어 시연 데이터로부터 전신 순응 제어를 학습하는 접근^[14]은 접촉 위치·방향에 대한 명시적 가정 없이 순응성을 획득한다. 본 연구는 이와 달리 강화학습으로 힘·토크 센서 없이 추정한 외력으로부터 외력 순응 속도를 명시적으로 생성한다는 점에서 구별된다.

선행연구들은 주로 외력 순응과 보행 성능의 균형 또는 지속적인 외란에 대한 대응에 초점을 맞추고 있으며, 실제 인간-로봇 상호작용(pHRI, Physical Human-Robot Interaction) 환경에서 사람과 로봇의 신장 차이에 따라 접촉 방향이 달라져 발생할 수 있는 외력 성분의 차이를 명시적으로 고려하지 않았다. 시물레이션에서는 외력을 특정 링크에 특정 방향으로 직접 인가할 수 있지만, 실제 pHRI 환경에서는 사람과 로봇의 신장 차이로 인해 동일한 의도의 수평 방향 접촉이라도 접촉 방향이 달라지면서 수직 방향의 힘 성분이 함께 발생할 수 있다. 이로 인해 실제 로봇에 작용하는 수평 외력 성분이 시물레이션에서 설정한 외력의 수평 성분보다 감소할 수 있으며, 결과적으로 시물레이션에서 학습한 외력 조건과 실제 환경에서 로봇에 작용하는 외력 조건 사이에 차이가 발생할 수 있다. 이러한 차이는 외력에 대한 정책의 입력과 실제 환경에서의 외력 조건 사이의 불일치를 유발하여 Sim-to-Real 성능에 영향을 미칠 수 있다.

원리적으로 접촉점과 힘의 방향을 알면 분산되는 힘은 로봇 모델로 계산할 수 있다. 그러나 실제 pHRI 환경에서는 접촉 기하와 신장 차이의 개인차, 그리고 외력 추정 오차로 인해 이 분산을 정확히 복원하기 어렵고, 이로 인해 학습 분포와 실제 분포 사이의 불일치가 남는다. 본 연구에서는 실제 pHRI 환경에서의 외력 성분 차이를 보상하기 위해 수평 방향 외력과 수직 방향 외력 중 큰 값을 선택하는 경험적 유효 힘(Effective Force)을 정의하고, 이를 목표 외력 순응 속도 생성 과정에 직접 반영하여 시물레이션과 실제 환경에서의 외력 조건 차이를 완화하고자 한다.

본 연구의 기여는 다음 세 가지로 요약된다. 첫째, 실환경으로의 배포 단계에서 힘·토크 센서나 별도의 외력 추정 모듈 없이 고유수용성 정보만으로 외력에 순응하는 센서리스 외력 순응 보행 정책을 학습한다. 비대칭 Actor-Critic 구조 기반의 특권 정보 학습을 통해 시물레이션 학습 단계에서만 알 수 있는 특권 정보를 이용한다. 둘째, 인간과 로봇의 신장 차이로 인해 실제 환경에서 발생하는 수직 방향 외력 분산을 보상하기 위한 경험적 유효 힘 모델을 정의하고, 이를 목표 외력 순응 속도 생성에 직접 연결하여 Sim-to-Real Gap을 완화한다. 셋째, 외력이

감지될 때 보행 명령 추종보다 외력 순응을 우선하도록 보상 가중치를 설계하고, 이를 29개의 자유도를 가지는 실물 휴머노이드에 추가 튜닝 없이 전이하여 순응 거동을 확인한다.

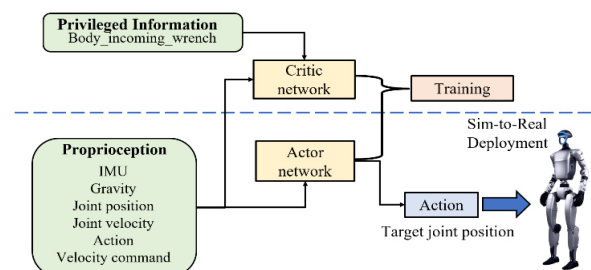
학습된 정책은 IsaacLab에서 PPO알고리즘을 통해 학습되었으며, 실제 Unitree G1 로봇에 적용하였다. 실험 결과, 로봇은 외력이 없는 경우 기존 보행 명령을 안정적으로 추종하였으며, 외력이 발생할 경우 보행 명령이 있어도 외력 방향으로 순응하여 이동하는 동작을 보였다. 이를 통해 본 연구는 별도의 힘토크 센서 없이도 실물 휴머노이드에서 외력 순응 보행이 가능함을 확인하였다.

2장에서는 본 연구에서 사용된 외력 순응 보행을 위해 특권 정보를 이용하여 외력을 추정하고 그 정확도를 측정한다. 추정된 힘을 기반으로 목표 순응 속도를 생성하는 알고리즘에 대해 설명한다. 3장에서는 IsaacLab을 이용한 강화학습을 위한 행동 공간(action space), 명령, 보상 함수, 도메인 랜덤화에 대한 설계를 제시한다. 4장에서는 학습의 수렴을 파악하기 위한 보상 함수 추이와 그 결과를 로봇의 속도와 보행 명령 속도의 RMSE (Root Mean Square Error, 평균 제곱근 오차), 로봇의 속도와 목표 순응 속도의 RMSE로 평가한다. 성능을 비교할 모델은 외력 순응 학습을 진행하지 않고 일반 보행을 학습한 모델이다. 또한 관절 토크의 RMS (Root Mean Square)도 함께 비교하여 관절 구동 토크도 비교한다. 마지막으로 5장에서는 본 연구의 결론과 향후 연구 방향을 제시한다.

2. 외력 순응 보행 알고리즘

2.1 비대칭 구조를 이용한 학습 프레임워크

본 연구에서는 시물레이션 전용 특권 정보를 학습에 활용하기 위해, 선행 연구에서^[13] 사용된 비대칭 Actor-Critic 구조를 채용하였다. 본 연구의 기여는 이 구조 자체가 아니라, 특권 정보로부터 추정한 외력을 유효 힘 기반 순응 속도로 변환하여 보상 함수에 결합하는 방식에 있다. 비대칭 Actor-Critic 구조를 통해 학습하고 Sim-to-Real 배포를 진행하는 프레임워크는 [Fig. 1]과 같다. Actor 네트워크와 Critic 네트워크에 사용된 관측 정보는 [Table 1]과 같다.



[Fig. 1] Asymmetric Actor-Critic structure for training and Sim-to-Real deployment

[Table 1] Observation terms

Components	Dimension	Critic	Actor
Base_ang_vel	3	✓	✓
Projected_gravity	3	✓	✓
Joint_pos	29	✓	✓
Joint_vel	29	✓	✓
Actions	29	✓	✓
Velocity_commands	3	✓	✓
Base_lin_vel	3	✓	
Force_commands	3	✓	
Body_incoming_wrench	18	✓	

Actor 네트워크에는 실제 로봇에서 획득 가능한 고유수용성 정보와 보행 명령만 입력하였다. 여기에는 로봇 몸체 각속도(Base_ang_vel), 중력 벡터(Projected_gravity), 관절 위치(Joint_pos), 관절 속도(Joint_vel), 행동(Actions), 보행 명령(Velocity_commands)이 포함된다. 따라서 실제 로봇에 정책을 배포할 때에도 추가적인 함토크 센서나 별도의 외력 추정 모듈은 필요하지 않다.

Critic 네트워크에는 Actor 네트워크에 입력한 정보들에 로봇 몸체 선속도(Base_lin_vel), 외력 부여 명령(Force_commands), Body_incoming_wrench 정보를 추가로 입력하였다. 이 정보들은 시뮬레이션 상에서만 얻을 수 있고 실제 로봇에서는 얻을 수 없는 정보이거나 노이즈가 심해 정확한 값을 얻기 어려운 값이기 때문에 특권 정보에 해당한다. Base_lin_vel은 로봇 몸체의 선속도 벡터 정보로 IMU를 통해 실물 로봇에서도 얻을 수 있는 정보지만 노이즈가 심해 정확도가 낮아 Critic 네트워크의 입력으로 사용하였다. 외력 부여 명령은 시뮬레이션 상에서 로봇에게 부여하는 외력 명령으로 정확한 외력 벡터를 알 수 있지만 실제 로봇에는 힘센서가 없어 알 수 없는 정보이기 때문에 Critic 네트워크에 입력하였다. Body_incoming_wrench는 IsaacLab의 물리 엔진 내부에서 계산되는 링크별 6축 힘 및 토크 정보로 로봇 base frame을 기준으로 한다. 마찬가지로 실제 로봇에서는 얻을 수 없는 정보이기 때문에 Critic 네트워크에만 입력하였다.

이 중 body_incoming_wrench를 이용하여 외력에 대한 정보를 얻었다. 외력 부여 명령을 사용하여 외력에 대한 정보를 얻을 수도 있었지만 이는 시뮬레이션 상에서 로봇에게 외력과 토크 정보를 직접 제공하기 때문에 시뮬레이션 환경과 실제 로봇 환경의 차이를 만들어 낼 수 있어 로봇에게 외력을 부여하는 역할로만 사용하였다. Body_incoming_wrench는 링크가 받는 힘 및 토크 정보로 중력, 지면 반력 등이 힘 정보에 포함되어 있어 순수 외력을 얻는 과정이 필요하다. Body_incoming_wrench에서 지면 반력 및 중력 성분을 제거하여 얻은 외력 정보는 가상의 목표 외력 순응 속도를 생성하는 데 사용된다. 생성된 목표 순응 속도는 보상 함수 내부에서 계산되며, 로봇이 외력 방향으

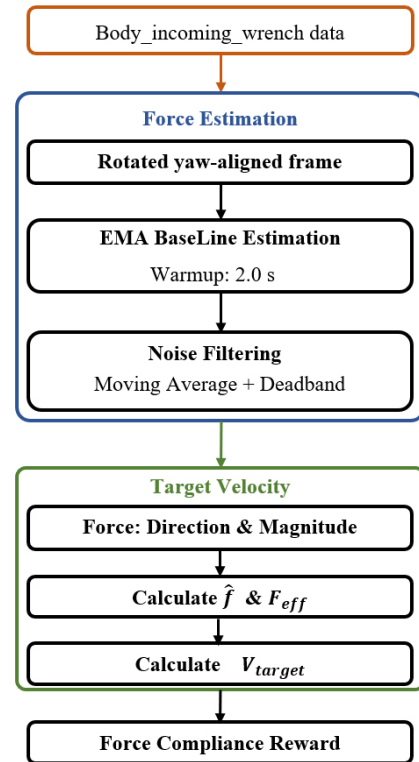
로 이동하며 순응하도록 정책을 유도하는 기준으로 활용된다.

이러한 비대칭 구조를 통해 정책 네트워크는 학습 과정에서 Critic 네트워크가 제공하는 외력 관련 정보를 활용하여 가치 함수를 학습하는 동시에, Actor 네트워크는 고유수용성 정보만으로 외력 인가 시 발생하는 관절 상태 변화와 로봇 동역학 변화를 암시적으로 학습하게 된다. 결과적으로 실제 환경에서는 별도의 외력 측정 없이도 고유수용성 정보만을 이용하여 외력에 순응하는 행동을 생성할 수 있다.

2.2 외력 추정 및 목표 순응 속도 생성

2.2.1 외력 추정

본 연구에서는 외력 순응 정책이 외력의 존재와 방향을 인식할 수 있도록, IsaacLab 환경에서 제공되는 body_incoming_wrench 정보를 이용하여 외력을 추정한다. 추정된 외력은 이후 목표 순응 속도를 생성하는 입력으로 사용된다. 그러나 body_incoming_wrench에는 순수한 외력뿐만 아니라 지면 반력과 중력에 의한 힘 정보도 함께 포함하고 있어 해당 정보를 그대로 사용할 경우, 실제 외력의 크기와 방향을 정확하게 반영하기 어렵다. 따라서 본 연구에서는 body_incoming_wrench에서 외력 이외의 성분을 제거한 후, 외력의 존재와 방향을 나타내는 추정 외력을 얻는 과정을 수행하였다. 외력 순응 보상 함수를 계산하는 과정은 [Fig. 2]와 같다.



[Fig. 2] Force estimation and target velocity generation process

먼저 `body_incoming_wrench`를 로봇의 yaw 방향을 기준으로 정렬된 좌표계로 변환하여 외력의 방향을 일관되게 표현하였다. 이후 외력과 무관하게 지속적으로 발생하는 중력 성분과 보행 과정에서 발생하는 지면 반력 성분을 제거하기 위해 외력이 작용하지 않는 상태의 기준값을 추정하였다. 기준값은 지수 이동평균(Exponential Moving Average, EMA)를 이용하여 계산하였다. 특히, 초기에는 기준값이 안정적으로 수렴하지 않을 수 있으므로 처음 2초 동안의 warm-up 구간을 두어 기준값을 초기화하였다. 이를 통해 외력과 무관하게 중력 및 지면 반력에 의해 발생하는 지속적인 성분을 기준값으로 추정하고, `body_incoming_wrench`에서 기준값을 제거하여 외력에 해당하는 성분을 추출하였다.

기준값을 제거한 후에도 수치적 노이즈 및 순간적인 진동 성분이 남을 수 있으므로, 이동 평균 필터를 적용하여 고주파 성분과 미세한 진동을 완화하였다. 또한 설정된 범위를 적용하여 미세한 잔류 신호를 제거하였다. 이러한 처리 과정을 통해 최종적으로 외력의 크기와 방향을 나타내는 추정 외력 벡터를 얻는다. 추정 외력은 보상 함수의 계산 과정에서 사용되며 외력 순응 보행에 필요한 목표 순응 속도를 계산하기 위한 목적으로 사용된다.

외력 추정기의 성능은 지면 반력 잔차의 영향을 배제하여 추정기 자체의 성능을 평가하기 위해 로봇 속도가 0인 정지 상태에서 우선 수행되었으며, 이후 지면 반력 변동이 반영되는 보행 중 조건에서도 동일한 방식으로 추가 평가하였다. 정지·보행 조건 비교는 [Table 2] 및 이하에서 다룬다. 추정된 외력의 정확도를 평가하기 위해 시뮬레이션에서 인가한 외력 명령 F_{cmd} 와 추정 외력 F_{est} 을 비교하였다. 평가는 로봇의 속도가 0일 때 수행하였으며, 수평면의 단일 축 방향 외력과 수직·수평 성분이 동시에 존재하는 다축 외력을 모두 포함하도록 구성하였다. 단일 축 외력은 x축 또는 y축 방향 40 N을 인가하였으며, 다축 외력은 (x, y, z) 순서로 (40, 0, 40) N 및 (0, 40, 40) N 조건을 사용하였다. 로봇의 `torso_link`와 `left/right_wrist_yaw_link`에 각각 4개의 외력 조건을 적용하여 총 12개의 조건으로 설정하고 평가하였다.

보행 중 조건은 정지 상태일 때와 동일한 외력 조건으로 로봇이 1 m/s로 x방향 또는 y방향으로 보행 중인 상태를 가정하였

다. 이때 외력은 보행 방향과 반대 방향으로 인가하여, 사람이 로봇의 진행을 저지하며 미는 특정한 pHRI 상황을 모사하였다.

추정 성능은 외력 감지율($r_{f,d}$, Force Detection Rate), 방향 유사도(S_d , Directional Similarity), 평균 절대 오차(MAE, Mean Absolute Error)를 사용하여 성능을 분석하였다. 외력 감지율은 실제 외력이 존재하는 구간 중 추정기가 외력의 존재를 올바르게 검출한 비율을 의미하며 식 (1)과 같이 표현된다. 식 (1)에서 TP (True Positive)는 실제 외력이 존재하고 추정기 또한 외력이 존재한다고 판단한 경우를 의미하며, FN (False Negative)은 실제 외력이 존재하지만 추정기가 외력이 존재하지 않는다고 판단한 경우를 의미한다. 방향 유사도는 식 (2)처럼 정의되며 실제 외력 벡터 F_{cmd} 와 추정 외력 벡터 F_{est} 사이의 코사인 유사도로 정의되며 값이 1에 가까울수록 추정된 외력의 방향이 실제 외력의 방향과 일치함을 의미한다. 식 (3)은 평균 절대 오차를 의미하고 명령된 x, y, z 성분별 외력 $F_{cmd,dir}$ 과 추정된 x, y, z 성분별 외력 $F_{est,dir}$ 사이의 절대 오차를 평균한 값으로 정의한다.

$$r_{f,d} = \frac{TP}{TP + FN} \quad (1)$$

$$S_d = \frac{F_{cmd} \cdot F_{est}}{\|F_{cmd}\| \|F_{est}\|} \quad (2)$$

$$MAE = \frac{1}{3} \sum |F_{cmd,dir} - F_{est,dir}| \quad (3)$$

[Table 2]는 전체 12개의 외력 조건에 대한 외력 추정 성능의 평균과 표준편차를 정지 및 보행 조건별로 나타낸다. 정지 조건에서 외력 감지율은 평균 96.958±0.588%, 방향 유사도는 평균 0.879±0.058로 나타나 대부분의 외력 구간을 안정적으로 검출하고 방향을 정확하게 추정함을 확인하였다. 보행 조건에서도 외력 감지율은 평균 99.731±0.568%로 오히려 소폭 높게 나타났으며, 방향 유사도는 평균 0.859±0.148로 정지 조건과 유사한 수준을 유지하였다. 다만 방향 유사도의 표준편차는 정지 조건에 비해 보행 조건에서 상대적으로 크게 나타나, 보행 중 지면 반력 변동으로 인해 조건별 편차가 커질 수 있음을 보여준다.

평균 절대 오차는 정지 조건에서 9.818±3.127 N, 보행 조건에서 10.125±4.892 N으로 두 조건 간 차이가 크지 않았다. 이는 지면 반력 노이즈가 가장 극심한 동적 보행 상황에서도 추정 오차가 크게 증가하지 않음을 보여주는 결과로, 정지 상태에서만 검증되었던 기존의 한계를 보완한다. 다만 표준편차는 보행 조건에서 다소 크게 나타났는데, 이는 걸음 주기에 따른 지면 반력 변동이 추정 오차의 변동성에 일부 영향을 미치는 것으로 판단된다.

[Table 2] Overall statistics of external force estimation performance under static and dynamic walking conditions (simulation). (Values are represented as mean ± standard deviation)

Metric	Static Walking	Dynamic Walking
$r_{f,d}$ [%]	96.958 ± 0.588	99.731 ± 0.568
S_d	0.8790 ± 0.058	0.8590 ± 0.148
MAE[N]	9.8180 ± 3.127	10.125 ± 4.892

본 연구의 외력 추정기는 외력 순응 보상 함수의 계산 과정에서 사용되며 외력 순응 보행을 위한 외력 존재 여부와 방향 정보를 제공하는 것을 목적으로 하며, 정지·보행 두 조건 모두에서 일관되게 높은 감지율과 방향 유사도를 보인 점을 고려할 때 추정된 외력은 목표 순응 속도 생성에 충분히 활용 가능함을 확인하였다. [Table 2]의 동적 보행 평가는 개선된 추정기로 추가 수행한 것이며, 4장의 정책 성능은 제출 시점 추정기로 학습·평가된 결과이다. 개선된 추정기의 오차가 더 작으므로 2.2.2절의 강건성 논의는 보수적으로 성립한다.

추정된 외력은 다음 절에서 설명하는 목표 순응 속도 v_{tgt} 생성 과정의 입력으로 사용되며, 로봇이 외력 방향으로 자연스럽게 이동하도록 유도하는 기준 정보로 활용된다.

2.2.2 Sim-to-Real 힘 분산을 고려한 목표 순응 속도 생성

본 절에서 제안하는 유효 힘 모델과 목표 순응 속도 생성 방법은 사람이 로봇의 한 접촉 지점에 준정적인 힘을 가하는 대표적인 pHRI 상황, 특히 사람과 로봇의 신장 차이로 인해 힘이 수평 및 수직 방향으로 분산되는 단일 접촉 시나리오를 대상으로 분석한 것이다.

시뮬레이션 환경에서는 외력 부여 명령에 의해 생성된 힘이 지정된 방향과 크기로 로봇에 직접 적용되므로 의도한 외력 벡터가 정확하게 전달된다. 그러나 실제 환경에서는 사람이 휴머노이드 로봇을 밀거나 당기는 과정에서 사람과 로봇의 신장 차이로 인해 힘이 의도한대로 정확하게 전달되지 않는다. 본 연구에서 사용한 Unitree G1은 일반적인 성인보다 신장이 낮기 때문에 사람이 로봇에게 가하는 힘은 수평 성분만이 아니라 수직 성분으로도 분산된다.

이러한 힘 분산의 차이는 시뮬레이션과 실제 환경에서 정책이 경험하는 외력 조건의 차이를 발생시켜 Sim-to-Real Gap을 유발하는 요인으로 작용할 수 있다. 사람의 신장을 h_h , 로봇의 신장을 h_r 이라 하고 사람과 로봇의 거리를 d 라고 한다. 이때 사람과 로봇의 높이 차에 의해 생기는 외력 벡터는 xy 수평면에 대해 각도 θ 를 가지고 θ 는 식 (4)와 같이 표현된다.

$$\theta = \tan^{-1}\left(\frac{h_h - h_r}{d}\right) \quad (4)$$

전체 외력의 크기를 F 라고 하면 F 의 xy 수평 성분 F_{xy} 와 수직 성분 F_z 는 각각 식 (5)와 식 (6)으로 분해된다.

$$F_{xy} = F \cos \theta \quad (5)$$

$$F_z = F \sin \theta \quad (6)$$

로봇은 점프를 하거나 몸을 숙이고 이동하지 않으므로 z방향으로는 이동하지 않는다고 가정한다. 따라서 외력에 대해 순응하는 보행 동작을 생성할 때 로봇은 F_z 에 대해서는 반응하지 않기 때문에 F_z 는 무시된다. 시뮬레이션에서는 특히 수평 방향으로 외력을 가할 때, 외력이 의도한 방향으로 정확하게 인가되어 $F_z \approx 0$ 인 경우가 대부분이었지만, 실제 환경에서는 신장차이로 인해 F_z 가 0이 아니게 되면서 시뮬레이션과 동일한 크기의 외력을 인가해도 F_{xy} 가 감소하게 된다. 신장차가 클수록 θ 가 커지게 되고, F_{xy} 의 크기도 감소되어 로봇은 시뮬레이션과 동일한 크기의 외력에 대해서 더 작은 수평 외력만을 인식하게 되어 결과적으로 실제 환경에서는 충분한 순응 동작을 생성하지 못하는 문제가 발생하였다.

분산된 외력의 실제 크기는 $\sqrt{\|F_{xy}\|^2 + \|F_z\|^2} = F$ 로 L2 norm을 통해 완전히 복원할 수 있다. 그러나 실제 pHRI 환경에서는 접촉 기하와 신장 차이의 개인차로 인해 이 분산을 정확히 복원하기 어렵고, 이로 인해 학습 분포와 실제 분포 사이의 불일치가 남는다. 2.2.1절에서도 기술하였듯이 추정 외력에는 수치적 노이즈와 지면 반력 잔차가 포함되어 있어, 두 축의 오차가 함께 증폭되어 유효 힘이 과대평가되고, 이에 비례하여 생성되는 목표 순응 속도가 급격하게 증가되어 보행 안정성을 해칠 수 있다. 이에 본 연구에서는 노이즈에 강건하면서도 분산된 힘을 일정 수준 이상 반영하기 위해 수평 방향 외력 F_{xy} 과 수직 방향 외력 F_z 중 큰 값을 선택하는 경험적 기준을 적용하여 유효 힘 F_{eff} 를 정의한다.

$$F_{eff} = \max(\|F_{xy}\|, \|F_z\|) \quad (7)$$

식 (7)의 유효 힘을 정의하면 신장차에 의한 θ 의 크기가 어떠한 경우에도 식 (8)와 같은 관계가 성립한다. 즉, 단일 축 추정 대비 분산된 힘 크기를 최소 71 % 이상 보존하면서도, L2 norm과 달리 단일 축 노이즈만 반영하여 과대평가를 억제한다. 이를 통해 외력 크기에 비례하는 안정적인 순응 속도 생성이 가능하다.

$$F_{eff} \geq \frac{F}{\sqrt{2}} = 0.71F \quad (8)$$

$$v_{tgt} = \text{clip}\left\{V_{\min} + V_{\max} \cdot \tanh\left(\frac{F_{eff}}{20 + \epsilon}\right), 0, V_{\max}\right\} \quad (9)$$

$$\text{clip}\{a, b, c\} = \begin{cases} b & \text{if } a < b \\ a & \text{if } b \leq a \leq c \\ c & \text{if } c < a \end{cases} \quad (10)$$

식 (9)는 유효 힘의 크기에 따라 목표 순응 속도의 크기 v_{tgt}

를 결정하며, 식 (10)의 clip 연산을 통해 v_{tgt} 가 설정된 범위를 벗어나지 않도록 제한한다. 학습 환경에서 인가한 외력의 최대 크기가 50 N임을 고려하여 정규화 상수는 20으로 설정하였으며, ϵ 는 수치적 안정성을 확보하기 위한 유효상수로 $1e-6$ 의 값을 적용하였다. v_{tgt} 은 유효 힘의 크기에 비례하여 증가하도록 설계되되, 작은 외력에 대해서 충분한 순응 동작이 발생하도록 최소 속도를 보장하고, 큰 외력에 대해서는 보행 안전성을 고려하여 최대 속도를 제한하였다. 이에 따라 최소 속도인 V_{min} 는 0.2 m/s로 설정하였고, 최대 속도인 V_{max} 는 0.8 m/s로 설정하였다. 외력이 없는 경우 $F_{eff} = 0$ 이므로 tanh항의 값이 0이 되어 v_{tgt} 은 $V_{min} = 0.2$ m/s에서 시작한다. 이후 유효 힘의 크기가 증가함에 따라 tanh항의 출력이 증가하면서 v_{tgt} 또한 단조 증가하며 $V_{max} = 0.8$ m/s를 초과하지 않도록 제한한다. 특히, 정규화 상수 20을 적용함으로써 유효 힘이 증가함에 따른 속도의 증가율을 조절하였으며, 약 19.5 N 이상의 유효 힘에서는 v_{tgt} 이 최대값인 0.8 m/s에 제한되도록 설계하였다. 이를 통해 큰 외력이 인가되는 상황에서도 과도한 순응 속도가 생성되는 것을 방지하고, 실제 환경에서 발생하는 외력의 분산을 고려하면서 로봇에 실제로 작용하는 외력에 대한 순응 동작이 안정적으로 생성될 수 있도록 하였다.

식 (9)의 목표 순응 속도 생성 함수는 유효 힘 추정 오차에 대해 구조적으로 강건하도록 설계되었다. tanh 함수와 clip 연산으로 인해 목표 순응 속도의 유효 힘에 대한 민감도($\partial v_{tgt} / \partial F_{eff}$)는 유효 힘이 커질수록 감소하며, 본 연구에서 설정한 $V_{min} = 0.2$ m/s, $V_{max} = 0.8$ m/s, 정규화 상수 20을 대입하면 F_{eff} 가 약 19.46 N을 넘는 순간부터 v_{tgt} 은 V_{max} 에서 포화되어 민감도가 0에 수렴한다. 포화 이전 구간 ($F_{eff} < 19.46$ N)에서는 유효 힘의 크기가 순응 속도에 반영되므로, 유효 힘의 크기 추정은 접촉 시작 및 해제 구간의 속도 조절에 기여한다. 2.2.1절에서 외력 추정 성능을 평가한 실험 조건은 40 N의 외력이었으며, 이는 포화 임계값 19.46 N을 크게 상회한다. 정지 조건에서 측정된 평균 절대 오차 9.818 ± 3.127 N을 적용하면, 평균+2표준편차 수준인 16.07 N의 오차가 더해지더라도 유효 힘 추정치는 포화 구간 내에 위치하여 목표 순응 속도는 변화 없이 0.8 m/s로 유지된다. 보행 조건에서 측정된 평균 절대 오차 10.125 ± 4.892 N을 적용해도 평균+2표준편차 수준인 19.91 N까지는 여전히 포화 구간 내에 머무르며, 이는 40 N 조건에서 포화가 해제되는 데 필요한 오차인 20.54 N에 근접하지만 이를 넘지 않는 수준이다. 다만 이 분석은 측정된 오차의 평균, 표준편차를 식 (9)에 대입하여 얻은 정적인 수치적 상한이며, 오차가 정책에 미치는 영향을 시뮬레이션이나 실험 로봇 실험을 직접 관찰하여 검증한 것은 아니다. 따라서 외력 추정치의 오차가 클 때 정책이 실제

로 어떤 반응을 보이는지에 대한 정량적인 검증은 추가 학습 없이 기존 시뮬레이션 환경에서 추정 외력에 노이즈를 주입하는 방식으로 수행할 수 있을 것으로 판단되며, 이를 향후 과제로 남긴다.

외력을 따라 순응하는 동작이 생성되기 위해서는 외력의 방향과 일치하는 방향으로 이동해야 한다. 로봇은 xy평면에서 움직이므로 외력 방향 단위벡터 $\hat{f}$ 는 식 (11)과 같다.

$$\hat{f} = \frac{F_{xy}}{\|F_{xy}\| + \epsilon} \quad (11)$$

$$V_{tgt} = v_{tgt} \hat{f} \quad (12)$$

로봇이 따라야 할 목표 순응 속도 V_{tgt} 를 생성하는 것은 식 (12)과 같다. 생성된 V_{tgt} 은 외력이 인가되었을 때 로봇이 추종해야 하는 가상의 목표 속도이며, 실제 명령으로 입력되는 값이 아니라 보상 함수 내부에서 계산되어 정책이 외력 방향으로 이동하도록 유도하는 기준 속도로 사용된다.

$$R_{comp} = \exp\left\{-\frac{(\|V_r - V_{tgt}\|)^2}{0.25}\right\} \quad (13)$$

식 (13)는 보상 함수를 계산할 때 사용되는 식으로 로봇의 속도 V_r 이 V_{tgt} 에 가까워질수록 높은 보상을 얻도록 설계되었다.

본 연구에서 목표 순응 속도를 식 (9)과 같이 명시적인 수식으로 설계한 것은, 실제 로봇에 적용할 때 속도 범위를 제한할 수 있는 해석 가능한 형태로 상위 레벨 목표를 제한하기 위함이다. 목표 속도 생성을 정책 네트워크에 완전히 맡길 경우, 학습 시 다루지 못한 외력 조건이나 도메인 랜덤화 범위를 벗어난 상황에서 정책이 예측하기 어려운 속도 명령을 생성할 위험이 있다. 이에 본 연구에서는 최소-최대 속도가 보장되는 목표 순응 속도를 명시적으로 설계하고, 이 목표를 실제로 달성하는 데 필요한 균형 유지, 발 디딤 조정, 관절별 토크 배분 등 전신 협응 동작은 강화학습 정책이 스스로 학습하도록 역할을 분리하였다. 4.4절에서 확인한 바와 같이, 제한한 정책은 목표 속도를 단순히 추종하는 데 그치지 않고 관절 평균 RMS 토크를 31.63 % 감소시키는 등 목표 속도 수식만으로는 설명되지 않는 순응 동작을 함께 학습하였으며, 이는 강화학습이 목표 속도 추종을 넘어 안전하고 효율적인 전신 순응 거동을 함께 학습하였음을 보여준다. 다만 현재의 유효 힘 모델과 목표 속도 생성 방식은 앞서 언급한 단일 접촉 시나리오에 한정되어 있으며, 다중 접촉 상황까지 일반화하고 목표 속도를 사전에 설계하는 대신 정책이 힘-속도 관계 자체를 직접 학습하도록 확장하는 것은 향후 연구로 남긴다.

3. 강화학습

3.1 행동 공간

본 연구에서는 정책 네트워크의 행동 공간을 각 관절의 목표 위치로 정의하였다. 정책 네트워크는 매 제어 주기마다 휴머노이드의 각 관절에 대한 목표 위치 q_d 를 출력하며, 출력된 행동은 저수준 PD 제어기의 참조 입력으로 사용된다.

정책이 생성한 목표 위치는 현재 관절 상태와 함께 PD 제어기에 입력되어 최종 구동 토크를 계산한다. 직접 토크를 출력하는 방식은 학습 초기 단계에서 불안정한 거동을 유발할 수 있으며 실제 로봇 적용 시 안전성 문제가 발생할 수 있다. 반면 목표 위치 기반 행동 공간은 저수준 제어기를 통해 관절의 움직임을 안정적으로 제한할 수 있으므로 시뮬레이션 학습과 실제 로봇 배포 모두에 유리하다.

q_d 는 정책이 생성한 목표 관절 위치, q 는 현재 관절 위치, $\dot{q}$ 는 현재 관절 속도, K_p 와 K_d 는 각각 비례 및 미분 계인을 의미한다. Unitree G1은 총 29 자유도를 가지며 정책 네트워크는 각 자유도에 대응하는 29차원의 목표 관절 위치 벡터 a_t 를 출력한다.

$$\tau = K_p(q_d - q) - K_d(\dot{q}) \quad (14)$$

$$q_d = q_{init} + \alpha a_t \quad (15)$$

이때, 모터의 급격한 고주파 진동을 억제하고 부드러운 순응 궤적을 생성하기 위해, 식 (15)을 통해 정책의 가공되지 않은 출력값 a_t 에 스케일링 인자 α 를 곱하여 초기 관절 각도 q_{init} 에 더해지는 변위 offset 형태로 q_d 가 계산된다. 스케일링 인자 α 는 0.25로 결정하였다.

3.2 외력 부여 명령

본 연구에서는 휴머노이드가 실제 환경에서 발생할 수 있는 다양한 외력 상황에 대해 순응하도록 학습하기 위해 IsaacLab의 Manager-Based 환경에 이벤트를 통해 구동되는 외력 명령을 새롭게 구성하였다. 휴머노이드가 외력을 받는 상황은 로봇의 몸통에 해당하는 torso_link와 양 쪽 손목에 해당하는 left/right_wrist_yaw_link을 지정하여 로봇이 몸통과 팔 부분에 외력을 받는 상황을 가정하였다. 각 링크에 작용하는 외력 벡터의 x, y, z 성분은 미리 설정된 범위 내에서 균등분포를 이용하여 무작위로 샘플링된다. Torso_link에 부여하는 외력의 크기는 전방향에서 -50~50 N이고, wrist_yaw_link에 부여하는 외력의 크기는 전방향에서 -40~40 N이다.

각 학습 단계에서 전체 병렬 환경의 70%에는 외력을 인가하고, 나머지 30%에는 외력을 인가하지 않은 일반 보행 조건을 적용하였다. 이를 통해 외력 순응 행동을 충분히 학습하는 동시에 외력이 없는 상황에서는 기존 보행 명령을 추종하는 능력을 함께 유지하도록 구성하였다.

외력은 순간적인 충돌로만 이루어진 형태가 아니라 수초 동안 지속되는 준정적 외력으로 가정하였다. 이는 사람이 로봇을 밀거나 당기는 상황과 같이 일정 시간 동안 힘이 지속적으로 작용하는 상호작용을 고려하기 위함이다. 또한 정책이 외력의 존재 여부를 명확히 인식하고 외력 방향으로의 순응 행동을 충분히 학습할 수 있도록 외력 이벤트는 Interval, Ramp, Active, Settling의 네 단계로 구성된다. 각 에피소드 내에서 Interval 단계는 2.5초, 활성 유지 단계는 1-5초 동안 무작위로 지속되도록 스케줄링하여 환경의 다양성을 확보하였다. Interval 단계에서는 외력이 작용하지 않으며 로봇은 기본 보행 명령을 수행한다. 이후 Ramp 단계에서는 외력의 크기가 급변하여 계단형으로 증가한다. 이러한 순간적인 단계별 외력 변동은 시스템에 강한 불연속적 과도 충격을 가함으로써 제어 시스템의 순간 강건성을 시험한다. Active 단계에서는 목표 외력이 일정 시간 유지되어 정적 순응 능력을 학습시키고, 마지막으로 Settling 단계에서는 외력이 ramp단계에서처럼 계단형으로 감소하여 제거된다.

이러한 구조를 통해 로봇은 갑작스러운 충격뿐만 아니라 지속적으로 작용하는 외력과 외력이 사라지는 상황까지 학습할 수 있으며, 외력이 부여되는 상황에 따른 순응 행동과 원래 보행 명령으로의 복귀 행동을 함께 학습할 수 있도록 하였다.

3.3 보상 함수

본 연구에서는 휴머노이드가 기본적인 보행 능력을 유지하면서도 외력이 인가되는 경우 외력 방향으로 순응하도록 학습하기 위해 보상 함수를 크게 두 부류로 나눠 구성하였다. 보행 명령 추종, 외력 순응을 위한 동작 생성 보상과 안정적인 동작 수행 및 실제 로봇 적용을 위한 안정성 및 규제항 보상으로 구분하였다.

3.3.1 동작 생성 보상 함수

동작 생성 보상 함수는 [Table 3]와 같다. 로봇이 기본적인 보행 능력을 유지하면서도 외력에 순응하는 행동을 학습하도록 설계하였다. 외력이 인가되지 않는 상황에서는 xy평면 선속도 추종 보상인 Track_lin_vel_xy와 yaw 각속도 추종 보상인 Track_ang_vel_z가 주된 역할을 수행한다. Track_lin_vel_xy는 로봇의 선속도 v_{xy}^r 가 선속도 명령 v_{xy}^{cmd} 을 추종하도록 하여 로봇이 부여된 선속도 명령을 따라 보행하도록 한다. Track_ang_

[Table 3] Motion generation rewards

Reward	Expression	Weight
Track_lin_vel_xy	$\exp\left(-\frac{\ v_{xy}^{cmd} - v_{xy}^r\ ^2}{1.0}\right)$	1.0
Track_ang_vel_z	$\exp\left(-\frac{\ w_z^{cmd} - w_z^r\ ^2}{1.0}\right)$	1.0
Foot_clearance	$\exp\left(-\frac{\ h_{foot} - 0.18\ ^2}{0.05}\right)$	0.5
Double_support	$G(v_{cmd}, v_{tgt}) \exp\left(-\frac{\ V_r\ ^2}{0.05^2}\right)$	0.5
Force_compliance	$\exp\left\{-\frac{(\ V_r - V_{tgt}\)^2}{0.25}\right\}$	3.5

vel_z는 로봇의 각속도 w_z^r 이 각속도 명령 w_z^{cmd} 을 추종하도록 하여 로봇이 부여된 각속도 명령을 따라 보행하도록 한다.

이때 안정적인 보행 동작을 학습하기 위해 발을 들어올리는 보상 함수인 Foot_clearance를 정의하였다. h_{foot} 는 로봇의 한 쪽 발의 높이를 의미하고, 보상 함수는 한 발만 공중에 있을 때만 계산된다. 또한 명령 속도가 0일 때 제자리에 정지하기 위해 보행 속도 명령과 순응 속도 명령이 모두 정지 조건일때, 양 발을 지면에 붙이고 서있는 자세를 만드는 Double_support도 함께 정의하였다. Double_support에서 보행 속도 명령과 순응 속도 명령이 모두 정지 조건일 때만 활성화되기 위해 게이트 함수 G를 사용하였다. 보행 명령 속도의 크기 $v_{cmd} \leq 0.1$ m/s이고, 목표 외력 순응 속도의 크기 $v_{tgt} \leq 0.1$ m/s인 경우에 G는 1을 가지고 그 외에는 G는 0을 가진다. 외력이 인가되는 상황에서는 2.2절에서 설명한 목표 순응 속도 V_{tgt} 으로 외력 순응 보상인 Force_compliance를 계산한다. 외력 순응 보상은 로봇의 속도 V_r 와 목표 순응 속도 V_{tgt} 사이의 오차를 최소화하도록 설계하여 로봇은 외력이 부여되면 외력 방향으로 이동하는 순응 행동을 학습하게 된다. 특히, 실제 로봇의 파손 방지 및 pHRI 환경에서의 안전성을 향상시키기 위해서 외력 순응 보상의 가중치는 보행 명령 추종 보상의 가중치보다 3.5배 높게 설정하였다.

3.3.2 안정성 및 규제함

로봇은 학습하면서 과도한 자세 변화나 불안정한 동작을 생성할 수 있어 실물 휴머노이드에 적용 가능한 정책을 학습하기 위해 [Table 4]의 안정성 및 규제함 보상 함수를 함께 사용하여 로봇의 자세와 동작 상태를 제한하여 외력에 순응하는 과정에서 기본적인 안정성을 유지하도록 설계하였다. Flat_orientation은 중력 벡터의 xy평면 성분 g_{xy}^r 을 억제하여 몸통 기울어짐을 방지하고, Lin_vel_z와 Ang_vel_xy는 각각 몸통의 연직 방향 선 속도 v_z^r 과 xy평면 각속도 w_{xy}^r 를 억제하여 불필요한 상체의 흔들림을 방지한다. 또한 Base_height를 통해 보행 중 로봇 베이스의 높이 $h^{r,base}$ 가 기준 높이인 0.78에서 크게 벗어나지 않도록 하여 주저앉거나 과도하게 낮아지는 동작을 방지하였다.

[Table 4] Stability and regularization rewards

Reward	Expression	Weight
Flat_orientation	$(g_{xy}^r)^2$	-5.0
Lin_vel_z	$(v_z^r)^2$	-0.2
Ang_vel_xy	$(w_{xy}^r)^2$	-0.05
Base_height	$(h^{r,base} - 0.78)^2$	-100.0
Undesired_contact	$\max(f_i^{con})$	-1.0
Feet_slide	$\sum_{i=1}^2 (v_{xy}^{foot} \cdot \ F_i^{foot}\)$	-2.0
Contact_force	$(F_i^{foot} - 0.01)^2$	-5.0e-6
Joint_deviation	$ q_{cur,j} - q_{limit,j} $	-0.5
Dof_torque	$\sum_j (\tau_j)^2$	-1.0e-8
Dof_acc	$\sum_j (acc_j)^2$	-1.0e-8
Dof_pos_limit	$\sum_{j=1}^{29} \max(lim_j - q_j)$	-1.0
Action_rate	$\ a_s - a_{s-1}\ ^2$	-0.05
Termination_penalty	terminated_buffer	-200

들림을 방지한다. 또한 Base_height를 통해 보행 중 로봇 베이스의 높이 $h^{r,base}$ 가 기준 높이인 0.78에서 크게 벗어나지 않도록 하여 주저앉거나 과도하게 낮아지는 동작을 방지하였다.

발 접촉과 관련된 보상함은 안정적인 보행과 불필요한 접촉을 억제하기 위해 적용하였다. Undesired_contact는 발 이외의 신체 부위에서 의도하지 않은 접촉 f_i^{con} 이 발생하는 것을 방지하며, Feet_slide는 발의 평면 방향 속도 v_{xy}^{foot} 와 발의 접촉력 F_i^{foot} 의 내적을 이용하여 지면과 접촉한 상태에서 발이 미끄러지는 동작을 억제한다. Contact_force는 발에 과도한 접촉력이 발생하는 것을 제한하여 외력에 대응하는 과정과 보행 시 발생할 수 있는 불필요한 충격을 완화하도록 설계하였다. 또한 Joint_deviation을 통해 각 관절의 자세 $q_{cur,j}$ 가 초기 기준 자세 $q_{limit,j}$ 에서 과도하게 벗어나는 것을 억제하였다. 이때 j 는 로봇의 29개의 관절 인덱스를 의미한다.

실제 로봇의 구동 및 제어 안정성을 고려하여 관절에 대한 규제함도 적용하였다. Dof_torque와 Dof_acc는 각각 관절 토크 τ_j 와 관절 가속도 acc_j 의 크기를 최소화하여 과도한 구동력과 급격한 관절 운동을 억제한다. Dof_pos_limit은 관절 위치 q_j 가 허용된 가동 범위 lim_j 에 접근하거나 이를 벗어나는 것을 방지한다.

학습이 이루어지는 동안 정책이 비정상적인 행동을 만들어 내지 않도록 Action_rate에서 현재 스텝에서 정책이 출력한 행동 a_s 과 이전 스텝에서 정책이 출력한 행동 a_{s-1} 의 변화량을 최소화하여 제어 입력의 급격한 변화를 억제하고 관절의 진동

을 완화하도록 설계하였다. Termination_penalty는 에피소드가 초기에 종료되는 경우를 terminated_buffer로 판별하여 비정상적으로 학습이 종료되었을 때 큰 페널티를 부여하여, 정책이 안정적인 상태를 지속적으로 유지하도록 유도하였다.

이러한 안정성 및 규제항은 순응 동작이 생성되는 과정에서 발생할 수 있는 과도한 자세 변화, 불필요한 접촉, 관절의 과도한 구동 및 급격한 제어 입력 변화를 억제하는 역할을 한다. 이를 통해 외력 순응 능력과 기본적인 보행 안정성 및 로봇의 물리적 제약을 동시에 고려한 정책을 학습하도록 하였다.

3.4 도메인 랜덤화

강화학습 기반 제어 정책은 시뮬레이션 환경에서 학습되기 때문에 실제 로봇에 배포할 경우 모델링 오차, 마찰 특성 차이, 질량 분포 오차 및 구동기 특성 차이로 인해 성능이 저하될 수 있다. 이러한 Sim-to-Real Gap은 특히 외력 순응 보행과 같이 외부 환경과의 상호작용이 중요한 문제에서 더욱 크게 나타난다. 본 연구에서는 실제 환경 변화에 대한 정책의 강건성을 향상시키고 실제 Unitree G1 환경으로의 전이를 용이하게 하기 위해 도메인 랜덤화 기법을 적용하였다. [Table 5]에서의 operation은 학습 환경의 물리적 파라미터를 변경하는 연산 방식을 의미하며 가산(Offset), 배율(Scale), 직접 할당(Direct Assignment)으로 구분된다. 가산은 기본값에 임의의 값을 더하거나 빼는 방식이고, 배율은 기본값에 임의의 배율을 곱하는 방식이며, 직접 할당은 설정된 값을 파라미터에 그대로 적용하는 방식이다.

[Table 5]에 있는 것처럼 환경의 물리적 특성 변화를 모사하기 위해 매 에피소드마다 로봇의 구동기 특성 및 초기 상태를 변화시켰다. 관절 PD 게인(Pd_gains-stiffness, damping), 초기 베이스 위치(Reset_base-xy, yaw), 모터 영점 오프셋(Motor_zero_offset), 관절 마찰 계수(Joint_param-friction, viscous friction), 관절 위치(Joint_param-xy)를 범위 내에서 무작위로 변경함으로써 동일한 초기 조건에 대한 과적합을 방지하였다.

에피소드 시작 시 정적 지면 마찰 계수(Static friction), 동적 지면 마찰 계수(Dynamic friction), 링크 질량(Link_mass), 베이스 질량(Base_mass), 질량 중심 위치(Base_com)를 무작위로 변경하였다. 이는 실제 로봇에서 발생할 수 있는 하드웨어 편차 및 환경 변화에 대해 정책이 과적합되는 것을 방지하기 위함이다.

일정 간격마다 물리적 특성을 변화시킬 수 있는 Interval 요소에는 3.2절에서 설명한 외력 부여 명령을 실질적으로 적용시키기 위해 apply_force_command를 입력하고 적용 간격을 0으로 설정하여 딜레이 없이 설정한 외력 생성 명령이 로봇에게 적용되도록 하였다. 이를 통해 로봇에게 외력을 부여하는 상황을 만들어 낼 수 있었다.

[Table 5] Domain randomization

Components	Range	Unit	Operation
Reset Every episode			
Pd_gains(stiffness)	(0.8, 1.2)	-	scale
Pd_gains(damping)	(0.5, 1.5)	-	scale
Reset_base (xy)	(-0.5, 0.5)	m	offset
Reset_base (yaw)	(-0.5, 0.5)	rad	offset
Motor_zero_offset	(-0.05, 0.05)	rad	offset
Joint_param (friction)	(0.01, 1.0)	μ	direct assignment
Joint_param (viscous friction)	(0.3, 1.5)	μ	direct assignment
Joint_param (xy)	(-0.5, 0.5)	m	offset
Startup			
Static_friction	(0.2, 2.0)	μ	direct assignment
Dynamic_friction	(0.2, 2.0)	μ	direct assignment
Link_mass	(0.8, 1.2)	-	scale
Base_mass	(-1, 3)	kg	offset
Base_com	(-0.12, 0.12)	m	offset
Interval			
apply_force_command	(0.0, 0.0)	sec	-

4. 결 과

4.1 학습환경

본 연구의 강화학습 정책은 IsaacLab 시뮬레이션 환경에서 학습되었다. 물리 시뮬레이션 주기는 0.005 s, 제어 주기는 0.02 s로 설정하였다. 학습 과정에서는 총 4096개의 병렬 환경을 사용하여 데이터 수집 효율을 높였으며, 각 환경은 4.0 m 간격으로 배치하여 환경 간 간섭이 발생하지 않도록 구성하였다. 정책 네트워크는 MLP 구조를 사용하였고, 2.1절에서 언급한 비대칭 Actor-Critic 네트워크 구조를 사용하였다.

본 연구에서는 외력에 대한 순응 행동을 단계적으로 학습하기 위해 커리큘럼 학습(curriculum learning)을 적용하였다. 학습 초기에는 외력이 없는 조건에서 시작하고, 학습이 진행됨에 따라 인가되는 외력의 크기를 점진적으로 증가시켜 최대 50 N 까지 확대하였다.

정책 학습에는 [Table 6]에 수록된 것처럼 PPO 알고리즘의 주요 학습 파라미터를 적용하였다. 정책 및 가치 함수의 파라미터를 갱신하는 학습률 Learning Rate은 1.0e-3, 현재 시점에서 미래 보상의 가치를 결정하는 할인 계수 Discount Factor는 0.99, 미래 보상의 영향과 추정 분산 사이의 균형을 조절하는 파

[Table 6] PPO parameter

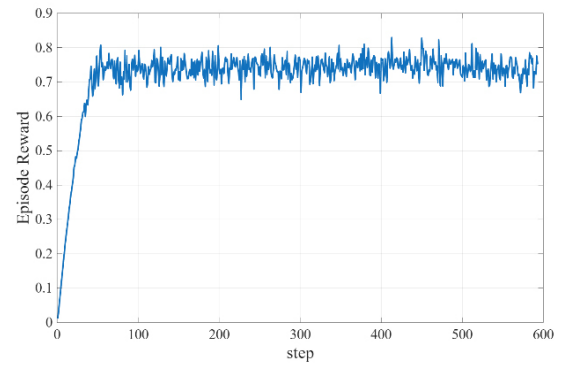
Parameter	Value
Learning Rate	1.0e-3
Discount Factor	0.99
GAE Parameter	0.95
Clip Parameter	0.2
Entropy Coefficient	0.008
Value Loss Coefficient	1.0
Number of Learning Epochs	5
Number of Mini-batches	4
Desired KL Divergence	0.01
Max Gradient Norm	1.0
Learning Rate Schedule	Adaptive
Empirical Normalization	False

라미터 GAE(Generalized Advantage Estimation) Parameter는 0.95, 정책 업데이트 과정에서 정책 변화의 범위를 제한하여 학습의 안정성을 확보하기 위한 Clip Parameter는 0.2로 설정하였다. 정책의 탐색을 유도하는 Entropy Coefficient는 0.008, 가치 함수 손실 가중치 Value Loss Coefficient는 1.0으로 설정하였다. 하나의 정책 업데이트에서 동일한 수집 데이터를 반복하여 학습하는 횟수를 나타내는 Number of Learning Epochs는 5회로 설정하였으며, 수집된 데이터를 나누어 학습에 사용하는 묶음의 수를 나타내는 Number of Mini-batches는 4개로 설정하였다. 이전 정책과 갱신된 정책 사이의 분포 차이를 나타내는 지표인 Desired KL Divergence는 0.01로 설정하였으며, 이를 기준으로 학습률을 조절하는 adaptive learning rate scheduling을 적용하였다. 정책 업데이트 과정에서 gradient의 최대 크기를 제한하는 Max Gradient Norm은 1.0으로 설정하였다. 학습률을 학습 과정에서 자동으로 조절하는 방법을 나타내는 Learning Rate Schedule은 Adaptive로 설정하였으며, 학습 데이터의 경험적 분포를 이용하여 입력값 또는 학습 신호를 정규화하는 Empirical Normalization은 적용하지 않았다.

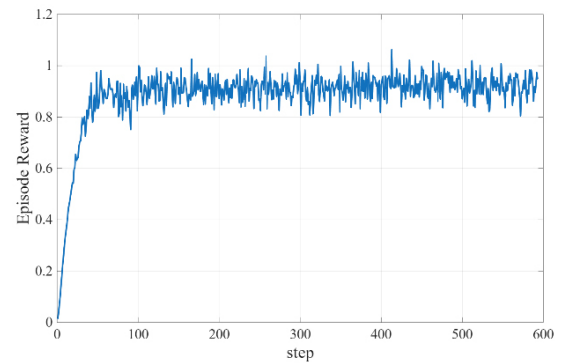
4.2 학습 결과 수렴 분석

본 연구의 강화학습은 IsaacLab 환경에서 PPO 알고리즘을 사용하여 수행하였으며, 학습을 위해 RSL_RL 라이브러리를 사용하였다. RSL_RL은 로봇 강화학습을 위한 PPO 기반의 학습 및 실행 기능을 제공하는 경량 강화학습 라이브러리로, 다수의 병렬 환경에서 수집된 경험을 이용하여 정책을 학습할 수 있다.

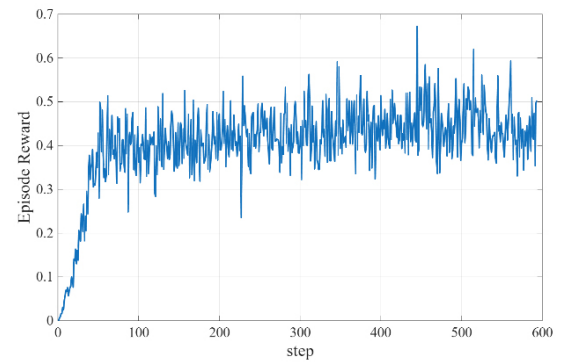
본 연구에서는 학습 과정의 안정성과 수렴 여부를 확인하기 위해 주요 보상 함수와 전체 평균 에피소드 보상의 변화를 [Fig. 3]에 나타내었다. 학습 과정에서 생성되는 RSL_RL의 학습 로



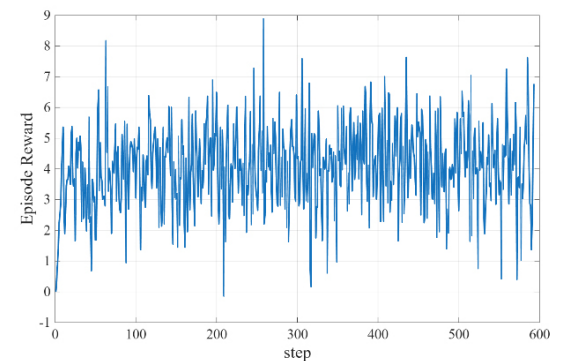
(a) Track_ang_vel_z episode reward learning curve



(b) Track_lin_vel_xy episode reward learning curve



(c) Force_compliance episode reward learning curve



(d) mean episode reward during training

[Fig. 3] Reward learning curve during the policy training

그를 Tensorboard를 통해 시각화하여 학습 과정에서의 보상 변화를 분석하였다. RSL-RL은 학습을 반복(iteration) 단위로 수

행하며, 각 반복에서 수집된 경험을 이용하여 정책을 갱신한다. 따라서 [Fig. 3]의 가로축은 학습 반복 스텝(learning iteration step)을 나타낸다.

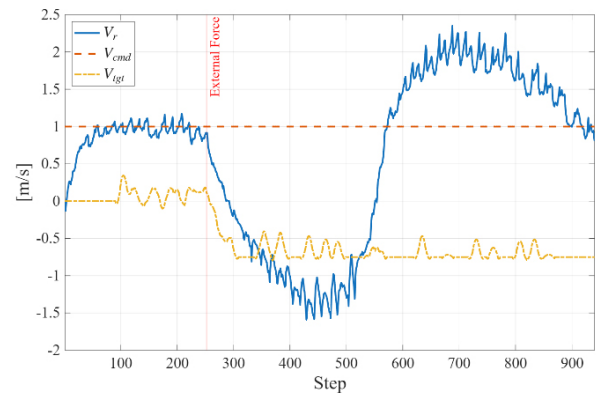
학습이 안정적으로 진행되는지를 확인하기 위해 동작 생성 보상 함수 중 yaw 각속도 명령 추종 보상인 [Fig. 3(a)] $Track_ang_vel_z$, xy 선속도 명령 추종 보상인 [Fig. 3(b)] $Track_lin_vel_xy$, 외력 순응 보상인 [Fig. 3(c)] $Force_compliance$ 의 변화를 분석하였다. $Track_ang_vel_z$ 는 학습 과정에서 보상이 증가한 후 약 0.75의 에피소드 보상으로 수렴하였으며, $Track_lin_vel_xy$ 는 약 0.90의 에피소드 보상으로 수렴하였다. 또한 외력에 대한 순응 동작 생성을 유도하는 $Force_compliance$ 는 학습이 진행됨에 따라 증가하여 약 0.45의 에피소드 보상으로 수렴하였다. 이를 통해 보행 명령 추종과 외력 순응에 관련된 주요 보상 함수가 학습 과정에서 일정한 범위로 안정화되는 것을 확인하였다.

전체 학습 과정의 추이를 확인하기 위해 모든 보상 항목을 종합한 Mean Episode Reward의 변화도 함께 분석하여 [Fig. 3(d)]로 나타냈다. Mean Episode Reward는 학습 초기 단계에서 빠르게 증가한 이후 변동 폭이 감소하면서 약 4.5의 에피소드 보상으로 수렴하는 경향을 나타냈다. 이는 학습이 진행됨에 따라 정책이 설정된 보상 함수의 목표를 지속적으로 만족하며 안정화되었음을 의미한다. 각 보상 함수와 Mean Episode Reward가 학습 후반부에서 일정한 범위로 안정화되고 수렴하는 경향을 보인다는 점에서, 본 연구의 강화학습 정책은 설정된 학습 환경에서 안정적으로 수렴한 것으로 판단하였다.

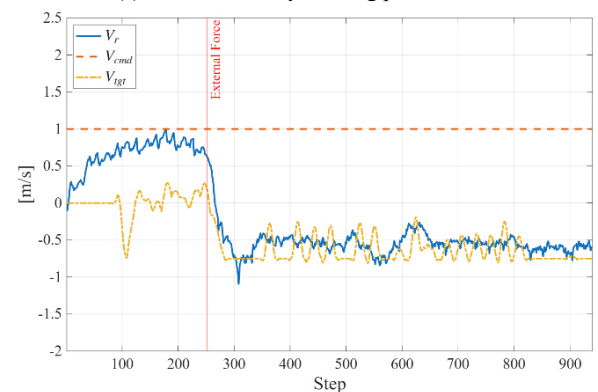
4.3 외력 순응 속도 성능 평가

제한한 외력 순응 정책이 다양한 보행과 외력 조건에서 적절한 순응 동작을 생성하는지 평가하였다. 평가는 외력 순응이 없는 일반 보행 모델(Baseline)과 외력 순응 동작을 학습한 모델을 비교하여 진행하였다. 평가를 위해 외력을 인가하기 전후의 로봇의 속도 V_r , 목표 순응 속도 V_{tgt} , 보행 명령 속도 V_{cmd} 를 비교하였다. 외력이 인가되기 전에는 보행 명령 속도와 로봇의 속도의 RMS 차인 Tracking RMSE를 이용하여 평가하였다. 외력이 인가된 후에는 목표 순응 속도와 로봇의 속도의 RMS 차인 Compliance RMSE로 평가하였다.

[Fig. 4]는 torso_link에 x 방향 -40 N, z 방향 30 N의 외력이 인가된 대표적인 실험 결과를 나타낸다. 외력은 시뮬레이션 시작 후 약 250 step부터 인가되었다. [Fig. 4(a)]는 외력 순응이 없는 일반 보행 모델의 결과이며, [Fig. 4(b)]는 외력 순응 동작이 학습된 모델의 결과이다. 외력이 인가되기 전에는 두 모델 모두 보행 명령 속도를 안정적으로 추종하는 것을 확인할 수 있다. [Fig. 4(a)]에서 외력이 인가된 이후 일반 보행 모델은 외력에 의



(a) Baseline velocity tracking performance



(b) Force Compliance tracking performance

[Fig. 4] Representative velocity tracking results under external force (Simulation)

해 로봇의 속도가 보행 명령으로부터 일시적으로 벗어나며, 이후 외력에 대한 반응을 보이지 않고 기존 보행 명령을 추종하기 위해 속도를 크게 변화시키는 결과를 보였다. 반면, [Fig. 4(b)]에서는 외력이 인가된 구간에서 기존의 보행 명령 속도를 지속적으로 추종하기보다, 인가된 외력에 따라 생성된 목표 순응 속도를 추종하는 방향으로 동작한다. 이는 제한한 정책이 외력 발생 시 기존 보행 명령을 무리하게 유지하기보다 외력에 따른 순응 동작으로 전환하여 외력에 대응함을 보여주며, 과도한 속도 변화를 방지하여 안정성을 가진다는 것을 볼 수 있다.

보다 일반적인 성능 평가를 위해 직진, 후진, 좌측 이동 및 우측 이동 명령에 대해 각 반대 방향의 외력을 torso와 wrist에 인가하는 총 8개의 실험 조건을 구성하였다. 보행 명령은 1 m/s의 속도 크기를 가지고 외력은 수평 방향으로 40 N, 수직 방향으로 30 N을 인가하였다.

제한한 정책과 일반 보행 정책의 속도 추종 성능은 [Table 7]에 나타난 것과 같다. 외력 순응 모델의 평균 Tracking RMSE는 0.3022 ± 0.1581 로 일반 보행 정책의 Tracking RMSE인 0.2686 ± 0.1514 보다 약 11 % 증가하였다. 이는 외력 순응 정책이 외력 순응 행동을 생성하는 과정에서 기존 보행 명령 추종 성능의 일

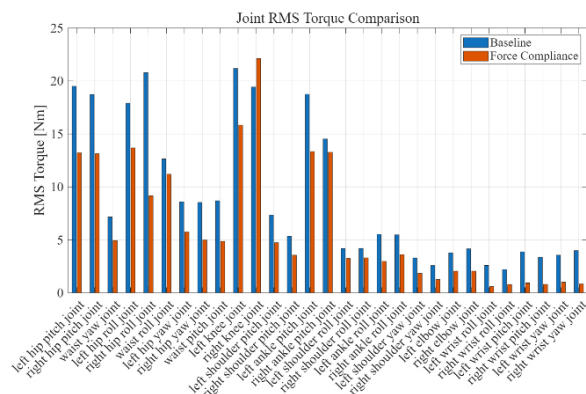
[Table 7] Average tracking RMSE and compliance RMSE under various velocity and force conditions in simulation. (Values are represented as mean $\pm$ standard deviation)

Model	Tracking RMSE	Compliance RMSE
Baseline	0.2686 ± 0.1514	0.7886 ± 0.4232
Force Compliance	0.3022 ± 0.1581	0.3445 ± 0.1163

부를 희생하였기 때문이다. 반면 외력 순응 모델의 Compliance RMSE는 0.3445 ± 0.1163 으로 일반 보행 정책의 Compliance RMSE인 0.7886 ± 0.4232 대비 약 56.3 % 감소하였다. 외력 순응 모델의 Compliance RMSE의 표준편차 역시 크게 감소하여 다양한 외력 조건에서 보다 일관된 순응 성능을 나타냄을 확인하였다. 이 결과는 제안한 외력 순응 정책이 보행 명령 추종 성능을 일부 저하를 허용하는 대신 외력 순응 성능을 크게 향상시켰다.

4.4 관절 구동 토크 비교

[Fig. 5]는 관절별 RMS 토크를 비교한 결과를 나타낸다. 4.3 절에서 제시된 일반적인 성능 평가를 위해 진행한 실험 조건과 동일한 조건으로 실험을 수행하였다. 순응 정책을 학습한 모델의 RMS 토크가 일반 보행 정책만을 학습한 모델에 비해 감소하였다. 이는 제안한 외력 순응 정책이 외력을 저항하는 대신 순응 행동을 생성함으로써 관절에 요구되는 구동 토크를 효과적으로 감소시켰음을 의미한다. [Fig. 5]의 데이터를 기반으로 일반 보행 정책과 외력 순응 정책의 전체 관절 평균 RMS 토크를 비교하였다.



[Fig. 5] Joint torque RMS Comparison with Baseline and Force compliance policy (Simulation)

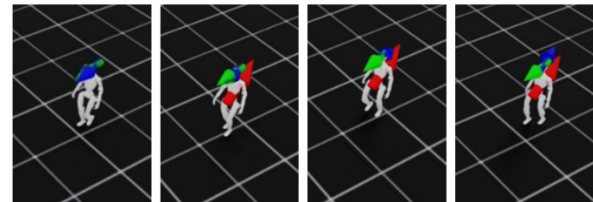
[Table 8] Average joint torque RMS (Simulation)

Model	Mean RMS	Reduction
Baseline	9.0288	-
Force Compliance	6.1734	31.63%

[Table 8]에서 확인할 수 있듯이 전체 관절 평균 RMS 토크는 일반 보행 정책에서는 9.0288 Nm을 가지고 외력 순응 모델에서는 6.1734 Nm로 감소하여 약 31.63 %의 개선을 보였다.

4.5 시뮬레이션 및 Sim-to-Real

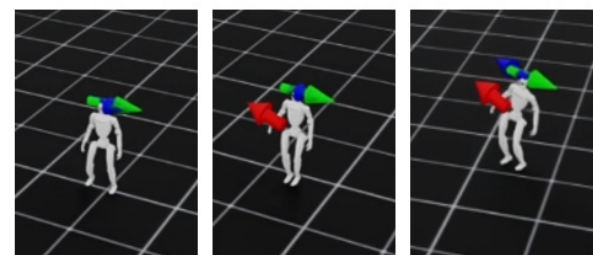
[Fig. 6]은 시뮬레이션과 실제 로봇 환경에서 제안한 정책의 순응 보행 결과를 비교한 것이다. [Fig. 6(a)]와 [Fig. 6(b)]는 로봇에게 x방향 직진 보행 명령을 부여한 상태에서 로봇의 torso_link에 직진 명령과 반대 방향의 외력을 인가한 경우의 시뮬레이션과 실험 로봇의 결과다. 시뮬레이션에서 로봇은 보행 명령을 수행하여 직진하다가 외력이 인가되자 외력에 순응하



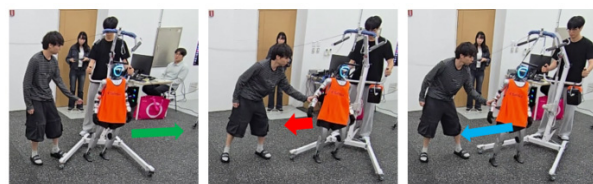
(a) Backward external force under forward walking command in simulation



(b) Backward external force under forward walking command in real



(c) Rightside external force under leftside walking command in simulation



(d) Rightside external force under leftside walking command in real

[Fig. 6] Simulation result and real robot result. Red arrow means external force, blue arrow means velocity of robot, and green arrow means commanded velocity

여 외력 방향으로 이동 방향이 변화하는 모습을 확인할 수 있었다. 실제 로봇도 시뮬레이션과 유사하게 보행 명령 수행 중 외력 작용 시 진행 방향이 변화하며 외력 방향으로 순응하는 행동이 관찰되었다. [Fig. 6(c)]와 [Fig. 6(d)]는 로봇에게 좌측 이동 명령을 부여한 상태에서 로봇의 오른팔에 외력을 인가한 경우를 나타낸다. 시뮬레이션과 실제 로봇 모두 보행 명령을 수행하여 옆으로 이동하던 중 외력이 인가되자 외력 방향으로 이동하는 모습을 보였다. 또한 외력이 제거된 이후에는 다시 보행 명령을 수행하여 원래의 이동 방향을 유지하는 거동을 확인할 수 있었다. 이를 통해 시뮬레이션에서 학습된 정책이 실제 로봇 환경에서도 추가적인 튜닝 없이 적용 가능하며, 외력 작용 여부에 따라 이동 방향을 조절하는 순응 보행 특성을 나타냄을 보여준다.

5. 결 론

본 논문은 별도의 외력 측정용 힘·토크 센서가 탑재되지 않은 휴머노이드 로봇을 대상으로, 실제 pHRI 환경에서 발생하는 외력 분산 현상을 보상하고 강건한 순응 보행을 달성하기 위한 강화학습 기반의 제어 프레임워크를 제안하였다. 시뮬레이션 학습 단계에서는 비대칭 Actor-Critic 구조를 도입하여 물리 엔진의 특권 정보를 효과적으로 활용하였으며, EMA 필터 기반의 지면 반력, 중력 제거 알고리즘을 통해 외력을 추정하였다. 추정된 외력을 기반으로 실제 환경에서 인간과 로봇의 신장 차이로 인해 발생하는 수직 방향 힘 분산 문제를 극복하기 위해 로봇에게 실질적으로 작용하는 경험적 유효 힘 모델을 새롭게 정의하였다. 현재 유효 힘 모델은 수평 방향 외력이 주로 인가되는 특정 pHRI 상황을 기반으로 설계되었으며, 다양한 접촉 위치와 접촉 방향에 대해 일반화된 모델은 아니지만, 이를 목표 외력 순응 속도로 변환하여 안전성을 우선으로 고려한 보상 함수를 구축하였다.

IsaacLab 시뮬레이션 환경에서의 평가는 로봇의 속도와 외력 순응 속도의 RMS 차로 진행된 결과, 제안한 순응 보행 정책은 외력 인가 시 목표 순응 속도를 높은 정확도(Average Compliance RMSE: 0.3445)로 추종하는 것을 확인하였다. 또한, 외력에 저항하던 기존 일반 보행 정책 대비 관절의 평균 RMS 구동 토크를 약 31.63% 감소시켰으므로써 외력에 대응하는데 필요한 관절 구동 부담이 감소함을 확인하였다. 최종적으로 시뮬레이션에서 학습된 정책을 추가적인 튜닝 없이 29개의 자유도를 가지는 실물 Unitree G1에 적용하였으며, 힘 센서가 없는 환경에서도 외부 충격 및 지속적인 당김 외력에 대해 순응하는 거동이 나타냄을 확인하였다.

본 연구는 몇 가지 한계점을 가지고 있다. 첫째, 외력 추정 정확도와 순응 보행 성능은 시뮬레이션 환경에서 정지 및 보행 조건 모두에 대해 검증되었고 실물 로봇에서의 정성적 순응 거동

도 확인하였으나, 실물 로봇 환경에서 외력 추정 정확도와 순응 성능을 함께 정량적으로 검증하는 작업은 외부 센서 구축을 하는데 시간이 소요되어 아직 수행되지 않아 이는 향후 과제이다. 둘째, 외력이 인가되는 시간이 길어질 경우, 추정 오차가 누적과 관성 효과로 외력 순응 성능이 일부 저하되는 현상이 확인되었다. 셋째, 목표 순응 속도 생성 함수의 강건성 논의는 측정된 오차의 평균·표준편차를 식 (9)에 대입하여 얻은 정적인 수치적 상황에 기반한 것으로, 오차가 정책에 미치는 영향을 시뮬레이션이나 실물 로봇 실험을 통해 검증하는 작업은 향후 과제로 남아있다. 넷째, 현재의 유효 힘 모델과 목표 순응 속도 생성 방식은 사람과 로봇 간 단일 접촉 지점을 가정한 pHRI 상황에 한정되어 있으며, 다중 접촉 상황까지 일반화하거나 목표 속도를 사전에 설계하는 대신 정책이 힘·속도 관계 자체를 직접 학습하도록 확장하는 것은 향후 연구로 남긴다. 특히 SoftMimic^[14]과 같이 시연데이터로부터 순응 거동을 암묵적으로 학습하는 접근은 접촉 위치나 방향에 대한 명시적 가정 없이도 순응성을 획득할 수 있다는 장점이 있어, 본 연구의 목표 속도 기반 설계와 결합하거나 대체하는 방향으로 확장할 가능성을 염두에 두고 있다.

향후 연구에서는 외력 추정 정확도를 개선하고 순응 성능도 개선하기 위한 연구를 수행하고, 일반화를 위해 다중 접촉 상황 같은 다양한 환경에서의 학습과 순응 거동을 암묵적으로 학습하는 방안, 실환경에서의 정량적 순응 성능 평가를 진행할 계획이다.

References

- [1] J. Hwangbo et al., "Learning agile and dynamic motor skills for legged robots," *Science robotics*, vol. 4, no. 26, Art. no. eaau 5872, Jan., 2019, DOI: 10.1126/scirobotics.aau5872.
- [2] N. Rudin, D. Hoeller, P. Reist, and M. Hutter, "Learning to walk in minutes using massively parallel deep reinforcement learning," *Conference on robot learning (CoRL)*, London, UK, pp. 91-100, Jan., 2022, [Online], <https://proceedings.mlr.press/v164/rudin22a.html>.
- [3] N. Hogan, "Impedance Control: An Approach to Manipulation, Part II—Implementation," *Journal of Dynamic Systems, Measurement, and Control*, vol. 107, no. 1, pp. 8-16, Mar., 1985, DOI: 10.1115/1.3140713.
- [4] A. Q. Keemink, H. van der Kooij, and A. H. A. Stienen, "Admittance control for physical human-robot interaction," *The International Journal of Robotics Research*, vol. 37, no. 11, pp. 1421-1444, Sep., 2018, DOI: 10.1177/0278364918768950.
- [5] J. Dong, J. Xu, Q. Zhou, and S. Hu, "Physical human-robot interaction force control method based on adaptive variable impedance," *Journal of the Franklin Institute*, vol. 357, no. 12, pp. 7864-7878, Aug., 2020, DOI: 10.1016/j.jfranklin.2020.06.007.
- [6] G. Park, M.-J. Kim, and J. Park, "Center of Mass Compliance Control of Humanoid Using Disturbance Observer," *The Journal of Korea Robotics Society*, vol. 17, no. 3, pp. 339-346, Aug., 2022, DOI: 10.7746/jkros.2022.17.3.339.

- [7] H.-M. Joe, "Locomotion Control of a Humanoid Robot using a Biped Walking Pattern based on Physical Interaction," *Journal of Institute of Control, Robotics and Systems*, vol. 27, no. 9, pp. 668-675, Sep., 2021, DOI: 10.5302/J.ICROS.2021.21.0075.
- [8] P. Zhi, P. Li, J. Yin, B. Jia, and S. Huang, "Learning unified force and position control for legged loco-manipulation," *arXiv:2505.20829*, 2025, DOI: 10.48550/arXiv.2505.20829.
- [9] B. Xu, H. Weng, Q. Lu, Y. Gao, and H. Xu, "Facet: Force-adaptive control via impedance reference tracking for legged robots," *arXiv:2505.06883*, 2025, DOI: 10.48550/arXiv.2505.06883.
- [10] Q. Lu, Y. Feng, B. Shi, M. Pisen, Z. Bao, and C. K. Liu, "GentleHumanoid: Learning Upper-body Compliance for Contact-rich Human and Object Interaction," *arXiv:2511.04679*, 2025, DOI: 10.48550/arXiv.2511.04679.
- [11] Y. Zhang et al., "Falcon: Learning force-adaptive humanoid loco-manipulation," *arXiv:2505.06776*, 2025, DOI: 10.48550/arXiv.2505.06776.
- [12] T. Leng, Y. Wang, T. Zheng, C. Luo, and M. Zhao, "Preference-Conditioned Multi-Objective RL for Integrated Command Tracking and Force Compliance in Humanoid Locomotion," *arXiv:2510.10851*, 2025, DOI: 10.48550/arXiv.2510.10851.
- [13] W. Lin, L. Qian, S. Wang, J. Qiang, P. Sun, and X. Luo, "RL-Force: Reinforcement Learning with Force Estimation for Humanoid Locomotion Subject to Continuous External Disturbances," *International Conference on Intelligent Robotics and Applications (ICIRA)*, Okayama, Japan, pp. 52-64, Aug., 2025, DOI: 10.1007/978-981-95-2095-4_5.
- [14] G. B. Margolis, M. Wang, N. Fey, and P. Agrawal, "Softmimic: Learning compliant whole-body control from examples," *arXiv:2510.17792*, 2025, DOI: 10.48550/arXiv.2510.17792.



박연수

2025 경희대학교 기계공학과(학사)
2025~현재 경희대학교 기계공학과(석사)

관심분야: Humanoids, Reinforcement learning



이창주

2019 경희대학교 기계공학과(학사)
2021 경희대학교 기계공학과(석사)
2025 경희대학교 기계공학과(박사 수료)
2025~현재 Safetics 연구원

관심분야: Robotic safety, Human-robot collaboration, Robotic manipulation, Integrated robot system development, Mobile manipulator control, Imitation learning for humanoid robots



김건우

2019 경희대학교 기계공학과(학사)
2021 경희대학교 기계공학과(석사)
2020~현재 Safetics 연구원

관심분야: Humanoids, Collaborative robots, Robot collision safety



임성수

1990 서울대학교 기계설계학과(공학사)
1992 서울대학교 대학원 기계설계학과(공학석사)
2000 Georgia Insti. of Tech. 기계공학과(공학박사)
2003~현재 경희대학교 교수

관심분야: Robotic Manipulation, Vibration, Safety, Human-Robot Collaboration



박성관

2018 서울과학기술대학교 기계공학과(학사)
2020 서울과학기술대학교 기계공학과(석사)
2021~현재 경희대학교 기계공학과(박사)

관심분야: Reinforcement learning, Human-robot interaction, Large language models