

접촉력 추정 기반 강화학습을 활용한 휴머노이드 이족 보행의 순응적 착지 프레임워크

Contact Force Estimation-Guided Reinforcement Learning Framework for Compliant Humanoid Bipedal Walking

조 용 현¹ · 이 관 우¹ · 박 경 재¹ · 임 서 현² · 신 재 용¹ · 최 현 준¹ · 박 상 혁¹ · 박 재 흥³ · 허 성 문[†]
Yonghyun Jo¹, Kwanwoo Lee¹, Gyeongjae Park¹, Seohyeon Lim², Jaeyong Shin¹,
Hyunjun Choi¹, Sanghyuk Park¹, Jaeheung Park³, SungMoon Hur[†]

Abstract: This paper presents a contact force-aware reinforcement learning framework for compliant humanoid bipedal walking. The framework estimates foot contact forces from proprioceptive history and incorporates them into policy learning. A history encoder is trained with auxiliary losses to infer the current and short-horizon future contact forces, while a virtual spring reward encourages compliant landing by inducing knee flexion when the estimated force exceeds a reference level. Through this design, the policy learns to absorb impact forces during both nominal walking and disturbance recovery, without relying on mechanical compliance or force/torque sensors. To transfer the learned policy to a four-bar-linkage humanoid, we build an actuator model combining motor parameter identification with an actuator network to reduce the sim-to-real gap. The framework reduces the maximum foot contact force in both nominal walking and disturbance recovery, and the policy transfers to the real robot zero-shot for stable compliant walking and push recovery. Experimental videos can be found at https://www.youtube.com/watch?v=uU9eGgM5_8w

Keywords: Humanoid Robots, Reinforcement Learning, Contact Force

1. 서 론

휴머노이드 로봇은 인간의 생활 및 작업 환경에서 별도의 환경 개조 없이 다양한 작업을 수행할 수 있는 플랫폼으로서 주목

을 받고 있다. 가사 보조, 물류, 재난 대응 등으로 휴머노이드 로봇의 활용 범위를 넓히기 위해서는, 작업 수행 과정을 안정적으로 보조할 수 있는 이족 보행 능력이 핵심 요구사항으로 고려된다.

강화학습은 대규모 병렬 시뮬레이션^[1]을 통해 수천 개의 환경을 동시에 활용한 효율적인 정책 학습을 가능하게 하여, 최근 이족 보행 기술의 비약적인 발전에 크게 기여하고 있다. 기존 궤적을 활용하거나^[2], 혹은 보상 함수 설계를 통해 보행 패턴을 유도^[3]하는 등의 방식을 통해 다양한 환경 및 상황에서 안정적으로 보행할 수 있음이 입증되고 있다.

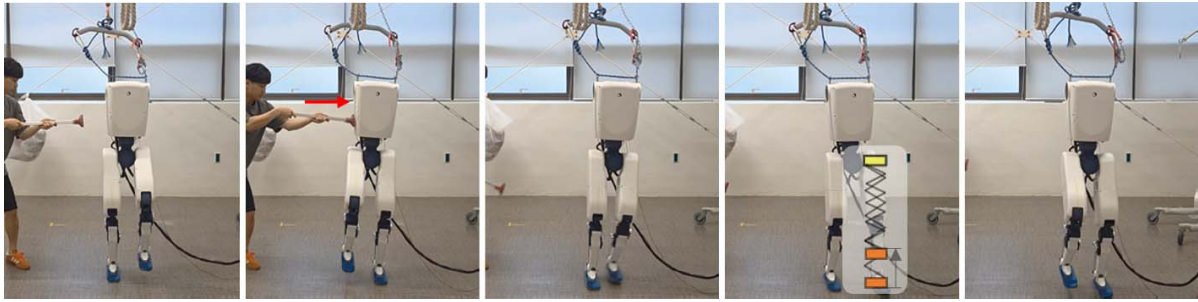
그러나 이러한 연구들은 주로 목표 속도 추종과 지형 적응에 초점을 두고 있어, 실제 로봇 보행에서 자세 안정성 저하와 하드웨어 부담을 유발할 수 있는 착지 충격력 완화 문제는 상대적으로 충분히 고려되지 않았다. 시뮬레이션에서는 과도한 접촉력이 물리적 제약 없이 허용되므로, 명시적인 접촉력 완화 방법이 고려되지 않을 경우 충격력이 크게 발생하는 정책이 학습될 수 있다. 이러한 정책을 실제 로봇에 배포하면 발과 지면의 충

Received : Jun. 22. 2026; Revised : Aug. 13. 2026; Accepted : Aug. 14. 2026

※ This work was supported by the Technology Innovation Program (RS-2025-25450033, Development of AI Foundation Model-Based Humanoid Robots Specialized for Home and Healthcare Environments) funded by the Ministry of Trade, Industry & Resources (MOTIR, Korea).

1. Student, Graduate School of Convergence Science and Technology, Seoul National University, Seoul, Korea (pieneee, kwlee365, rudwo1301, jasonshin0537, joony1121, sang0823@snu.ac.kr)
2. Student, Department of Mechanical Engineering, Yonsei University, Seoul, Korea (joan0219@yonsei.ac.kr)
3. Professor, Graduate School of Convergence Science and Technology, ASRI, RICS, Seoul National University, Seoul, Korea; Advanced Institutes of Convergence Technology, Seoul, Korea (park73@snu.ac.kr)

† Co-CEO, Corresponding author: Blue-robin Incorporation, Seoul, Korea (sm_hur@bluerobin.co.kr)



[Fig. 1] Push recovery sequence of the P73 Walker. From initial posture (1), an external push displaces the robot (2-3), which responds with a corrective step. At foot contact (4), the virtual spring mechanism induces compliant knee flexion to absorb the impact. The robot recovers balance (5) without force sensors or mechanical compliance

돌이 자세 안정성을 저하시키고^[4], 구동기 부하 및 하드웨어 마모를 가속시킬 수 있다^[5].

과거 휴머노이드 로봇의 보행 연구에서는 발과 지면 사이의 접촉 순응성을 조절하기 위해 힘/토크 센서 피드백 기반 제어 방법이 널리 사용되어 왔다^[6,7]. 그러나 힘/토크 센서의 사용은 시스템 비용, 다리 구조의 무게, 그리고 하드웨어 복잡도를 증가시키므로, 비용 효율성과 경량화를 중시하는 최근 휴머노이드 로봇 설계에서는 그 적용이 제한적일 수 있다. 이러한 배경 아래, 접촉력을 직접 측정하지 못하는 상황에서도 보행 시 발생하는 발과 지면 사이의 충격을 간접적으로 완화하기 위한 다양한 방법론들이 제안되어 왔다. Denoising World Model Learning (DWL)^[8]은 5차 다항식으로 스윙 발의 궤적을 사전 설계하여 착지 시 발 속도를 줄이는 방식을 사용하고, QuietWalk^[9]은 발 접촉 속도 패널티를 통해 착지 충격을 간접적으로 완화한다. 이러한 방법들은 접촉이 일어나기 전에 충격의 크기를 미리 추정하지 못한다. 따라서 외란 대응 중 발이 계획보다 빠르게 지면에 도달하거나 지면 위 장애물에 의해 조기 접촉이 발생하면, 접촉 직전 발 속도를 충분히 감속하지 못해 착지 충격이 그대로 전달될 수 있다.

본 논문에서는 힘/토크 센서 없이 보행 중 발과 지면 사이의 접촉력을 추정하고, 이를 정책 학습에 직접 활용함으로써 착지 충격을 완화하는 접촉력 인지 강화학습 프레임워크를 제안한다.

본 논문의 핵심 기여는 다음과 같다:

1. 힘/토크 센서 없이 고유수용성 정보만으로 보행 중 발과 지면 사이의 현재 및 미래 접촉력을 추정하는 방법을 제안한다. 고유수용성 이력에 내재된 접촉력 상승의 초기 신호를 포착하여, 접촉력이 최댓값에 도달하기 전에 정책이 활용할 수 있는 정보를 제공한다.
2. 추정된 접촉력에 비례하여 작동하는 가상 스프링 보상을 설계하여, 착지 전부터 선제적으로 순응적 동작을 유도한다. 이를 통해 정상 보행과 외란 상황 모두에서 최대 접촉력과 하중 증가율을 완화한다.

3. 외란 환경에서의 접촉력 비교 실험을 통해 최대 접촉력과 하중 증가율을 효과적으로 완화를 검증한다. 또한, 잠재 표현 제거 분석(latent ablation)을 통해 추정된 접촉력 정보가 보행 정책에 유의미하게 활용됨을 확인한다.
4. 제안 프레임워크를 사절 링크 구조의 휴머노이드 로봇에 적용하기 위해, 모터 파라미터 추정^[10]과 구동기 네트워크^[11]를 결합한 구동기 모델을 구축한다. 이를 통해 실제 로봇의 구동기 특성을 시뮬레이션에 반영하고, 시뮬레이션에서 학습된 정책의 실제 로봇 전이를 용이하게 한다. 학습된 정책을 추가 학습 없이 실제 로봇에 이식하여 [Fig. 1]과 같은 순응적 착지를 확인하고, 역산한 지면 반력으로 접촉력 완화 효과를 검증한다.

본 논문의 구성은 다음과 같다. 2장에서는 관련 연구를 정리하고, 3장에서는 접촉력 추정 기반 강화학습 보행 제어 프레임워크와 사절 링크 구조로 구동되는 휴머노이드 로봇을 위한 시뮬레이터 내 구동기 모델링 방법을 소개한다. 4장에서는 실험 환경과 정량적 검증 결과를 제시하며, 5장에서는 결론과 향후 연구 방향을 논의한다.

2. 관련 연구

2.1 강화학습 기반 이족 보행

대규모 병렬 시뮬레이션^[1]의 등장 이후, 강화학습 기반 보행 제어는 사족 로봇^[11,12]과 이족보행 로봇 모두에서 비약적으로 발전하였다. 이족 보행에 대한 강화학습 접근은 크게 두 계열로 나뉜다.

참조 궤적 기반 접근(reference-based)은 모션 캡처 데이터 또는 사전 설계된 기준 궤적을 활용하여 보행 패턴을 학습한다^[2]. Cheng 등^[13]은 대규모 인간 모션 캡처 데이터를 활용하여 자연스러운 전신 동작을 학습한 후 실제 휴머노이드에 이식하였다. 참조 궤적 기반을 사용하지 않는(reference-free) 접근은 기준 궤적 없이 보상 설계만으로 보행 패턴을 유도한다. Siekmann 등^[3]은 주기적 보상 합성을 통해 다양한 이족 보행 동작을 학습

하고, 시뮬레이션에서 학습된 정책을 실제 로봇으로 전이할 수 있음을 보였다. 이후 파쿠르 수준의 고난도 지형 보행^[14]이나 지각 정보를 결합한 전신 활용 보행^[15]까지 이족 보행의 적용 범위가 빠르게 확대되고 있다. 본 연구는 이 기준 궤적을 사용하지 않는 접근 방식에 속한다.

최근에는 단순히 보행의 성공률을 넘어, 보행 과정에서 에너지 효율, 착지 소음, 충격력 등의 보행 품질에 대한 관심이 대두되고 있다^[9,16]. 그러나 이러한 시도들은 주로 사족 로봇을 대상으로 이루어졌으며, 이족 보행 로봇에서 착지 충격력 완화와 같은 보행 품질 개선을 직접적으로 다룬 연구는 아직 제한적이다.

2.2 보행 로봇의 접촉력 제어

보행 로봇의 접촉력 제어는 전통적으로 임피던스 제어^[17], 작업 공간 기반 접촉 제어^[18], 지면 반력 기반 동역학 제어^[6,7], 이차 프로그래밍 기반 전신 제어^[19,20] 등 모델 기반 방법을 중심으로 연구되어 왔다. 이러한 방법들은 보행시 발생하는 지면과 발 사이의 접촉 충격 완화와 균형 유지에 효과적이지만, 정확한 동역학 모델과 힘/토크 센서 피드백에 의존하므로 모델링 오차가 존재하거나 센서 피드백을 사용할 수 없는 경우 적용이 제한된다.

최근에는 강화학습 기반 보행 연구에서 착지 충격을 완화하려는 시도가 이루어지고 있다. DWL^[8]은 5차 다항식으로 스윙 구간(swing phase)에서의 발 높이 및 속도 프로파일을 사전 설계하여 착지 직전 발 속도를 감소시켰다. QuietWalk^[9]는 발 접촉 속도 페널티와 가변 비례-미분 이득으로 보행 소음을 감소시켰다. QuietPaw^[10]는 접촉력과 충격 속도 기반의 소음 비용을 정의하고, 소음 제약 수준에 따라 소음과 보행 성능의 균형을 조절하는 사족 보행 정책을 제안하였다. 그러나 이러한 연구들은 접촉 전 발의 운동을 제어하는 간접적 방식이거나 사족 로봇의 소음 감소에 초점을 두고 있어, 이족 보행 로봇이 외란 대응과 같이 예기치 않은 조기 접촉 상황을 직면했을 때 지면과 발 사이의 충격을 선제적으로 완화하기 어렵다는 한계를 공유한다.

강화학습 기반 힘-임피던스 제어 역시 활발히 연구되어 왔다. Portela 등^[21]은 매니퓰레이터에서 힘/토크 센서 없이 직접 힘 제어를 학습하는 강화학습 방법을 제안하였다. FACET^[22]은 임피던스 기준 모델 추종을 통해 제어 가능한 순응성을 달성하였으며, FALCON^[23]은 힘 커리큘럼 기반의 이중 에이전트 구조를 통해 휴머노이드 로봇의 힘 적응형 이동 조작을 실현하였다. 이들 연구는 로봇의 중심부 또는 말단 장치에 가해지는 외력에 순응하는 action을 학습하는 데 초점을 두고 있어, 보행 중 스윙 발의 착지에서 반복적으로 발생하는 충격 접촉력을 추정하고 완화하는 문제는 다루지 않는다. 본 연구는 발과 지면 사이의 수직 접촉력을 추정 대상으로 하여, 착지 충격을 선제적으로 흡수하는 순응적 보행 정책을 학습한다.

2.3 시뮬레이션 학습 정책의 실제 로봇 전이

시뮬레이션에서 학습된 보행 정책을 실제 로봇에 배포할 때, 중심부 속도, 지형, 접촉 상태 등 시뮬레이션에서만 접근 가능한 특권 정보(privileged information)는 실제 환경에서 직접 관측할 수 없다. 이를 해결하기 위해 고유수용성 이력으로부터 특권 정보를 추정하는 다양한 접근이 제안되어 왔다. Rapid Motor Adaptation (RMA)^[24]는 특권 정보를 인코딩한 잠재 벡터(latent vector)를, 고유수용성 이력으로부터 회귀(regress)하도록 학습하는 2단계 적응 방식을 사용한다. DWL^[8]은 디코더(decoder)가 잠재 벡터로부터 전체 상태(full state)를 복원하도록 학습시켜 인코더에 환경 정보를 내재시킨다. 본 연구는 단계별 학습과 디코더 구조를 활용하지 않고 인코더의 잠재 벡터를 시뮬레이터의 물리량(중심부 속도, 지면과 발 사이의 접촉력)에 직접 회귀시키는 보조 손실로 학습한다. 잠재 벡터에 접촉력 추정 영역을 할당하여 보행 정책이 접촉력 정보를 명시적으로 활용할 수 있도록 한다. 추가로, 시뮬레이션과 실제 로봇 간 동역학 차이를 줄이기 위해, 모터 파라미터 추정^[10]과 구동기 네트워크^[11]를 결합하여 실제 로봇의 구동기 특성을 시뮬레이션에 반영한다.

3. 방 법

본 장에서는 제안하는 강화학습 프레임워크를 학습 설계 순서에 따라 기술한다. 먼저 관측·action 공간과 정책 구조를 정의하고, 학습 목적 함수와 이를 통한 접촉력 추정 메커니즘을 설명한다. 이어서 추정된 접촉력을 활용하는 가상 스프링 보상과 전체 보상 함수 설계를 제시하고, 마지막으로 학습된 정책을 실제 로봇으로 전이기 위한 real-to-sim 구동기 모델링을 기술한다.

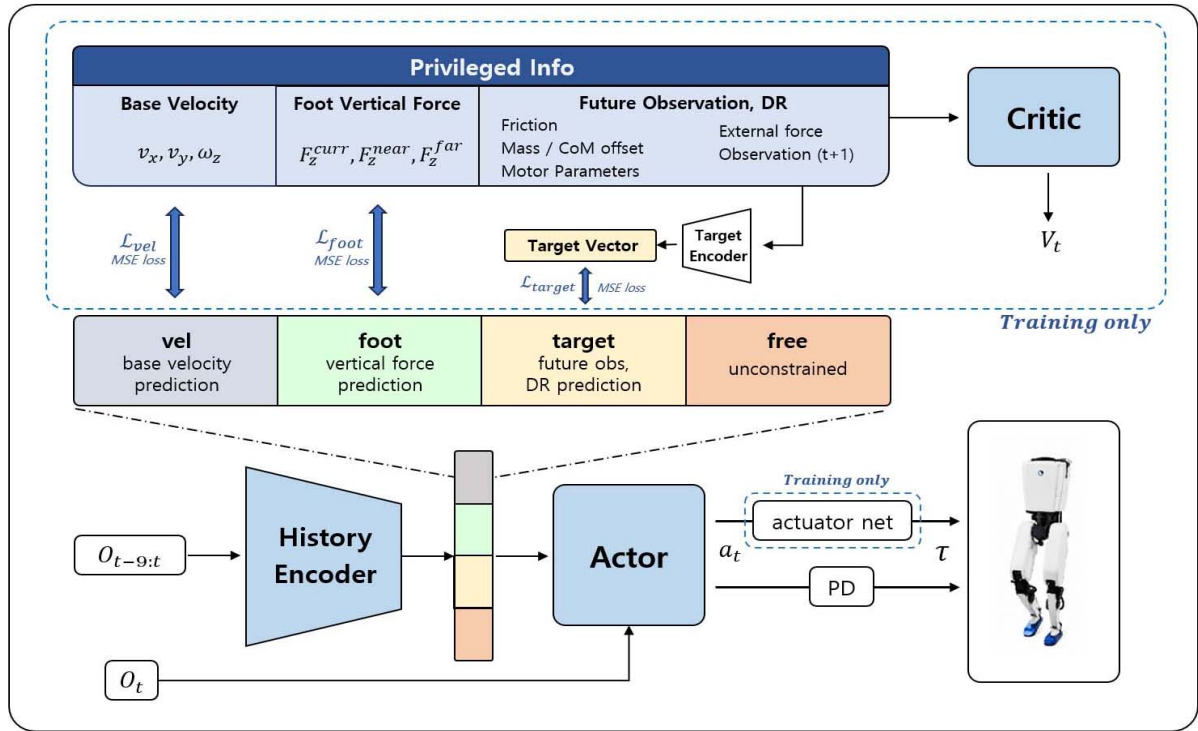
3.1 관측, action 및 정책 구조

본 연구는 휴머노이드 보행 제어를 partially observable Markov decision process(POMDP)로 정식화하며, Proximal Policy Optimization(PPO)^[25]와 asymmetric actor-critic^[26]을 사용하여 학습한다. 전체 네트워크 구조를 [Fig. 2]에 나타내었다.

단일 프레임 관측 $o_t \in \mathbb{R}^{47}$ 은 다음과 같이 구성된다:

$$o_t = [\omega_t, g_t, v_t^{cmd}, \sin(\phi_t), \cos(\phi_t), q_t, \dot{q}_t, a_{(t-1)}] \quad (1)$$

여기서 $\omega_t \in \mathbb{R}^3$ 은 중심부 좌표계 기준 각속도, $g_t \in \mathbb{R}^3$ 은 중력 벡터의 중심부 좌표계 투영, $v_t^{cmd} \in \mathbb{R}^3$ 은 속도 명령, $\sin(\phi_t) \in$



[Fig. 2] Overall architecture of the proposed framework. Training path: the history encoder extracts a latent vector from proprioceptive history, which is split into supervised regions (velocity, foot force, target) and a free region. The critic uses privileged information for value estimation. Deployment path: only the history encoder and actor are deployed, requiring proprioception only

$\in \mathbb{R}, \cos(\phi_t) \in \mathbb{R}$ 는 보행 위상(gait phase), $q_t \in \mathbb{R}^{12}$ 은 관절 위치, $\dot{q}_t \in \mathbb{R}^{12}$ 은 관절 속도, $a_{t-1} \in \mathbb{R}^{12}$ 은 이전 action이다. 위상 변수 ϕ_t 는 주기 동작에서 현재 위상을 나타내기 위한 것으로, 시간에 따라 0에서 2π 까지 증가하고, 2π 에 도달하면 다시 0이 된다. 10 프레임의 관측 이력을 쌓아 $o_{t-9:t} \in \mathbb{R}^{470}$ 을 정책 입력으로 사용한다.

action $a_t \in \mathbb{R}^{12}$ 는 12개 하체 관절의 위치 오프셋이다. 기본 자세 q_0 에 대해 $q_{des} = q_0 + s_a \cdot a_t$ 로 목표 관절 위치를 계산하며(s_a 는 action scaling factor), PD 제어로 토크 명령을 생성한다:

$$\tau_{cmd} = K_p(q_{des} - q) - K_d\dot{q} \quad (2)$$

정책은 이력 인코더(history encoder), actor, critic의 세 모듈로 구성된다([Fig. 2]). 이력 인코더는 관측 이력으로부터 잠재 벡터 $z_t \in \mathbb{R}^{64}$ 를 추출한다. Actor는 잠재 벡터 z_t 와 현재 관측 o_t 를 결합하여 action을 출력한다. Critic은 특권 정보를 포함한 관측을 입력으로 상태 가치를 추정하며, 시뮬레이션에서만 사용된다.

3.2 학습 목적 함수 및 접촉력 추정

이력 인코더가 추출하는 잠재 벡터 z_t 의 각 영역은 서로 다른 물리량을 추정하도록 보조 손실 함수(auxiliary loss)를 통해 학습된다.

$$z_t = [z_{(t,0:2)} \ z_{(t,3:8)} \ z_{(t,9:38)} \ z_{(t,39:63)}] \quad (3)$$

z_t 의 첫 3칸에 해당하는 항목인 $z_{t,0:2} \in \mathbb{R}^3$ 는 중심부의 x, y 축 선속도와 z 축 각속도, $z_{t,3:8} \in \mathbb{R}^6$ 는 양발의 현재 및 미래 수직 접촉력, $z_{t,9:38} \in \mathbb{R}^{30}$ 는 타깃 인코더(target encoder)가 다음 시점 관측과 특권 정보를 압축한 표현(target encoding)을 추종 및 저장하도록 할당된 공간이다. $z_{t,39:63} \in \mathbb{R}^{25}$ 는 PPO gradient만으로 학습되는 자유 영역(free region)으로, 보조 손실 영역이 포착하지 못하는 환경 정보를 정책이 자율적으로 인코딩할 수 있도록 확보했다^[27].

전체 학습 손실은 다음과 같다:

$$L = L_{PPO} + \lambda_v L_{value} + \lambda_e L_{entropy} + L_{aux} + \lambda_m L_{mirror} \quad (4)$$

여기서 λ_v 는 가치 손실 계수(value loss coefficient), λ_e 는 엔트로피 계수(entropy coefficient), λ_m 은 대칭 손실 계수(mirror loss coefficient)이다. L_{aux} 는 잠재 벡터의 각 영역을 참값(ground truth)으로 학습시키기 위한 보조 손실이다.

$$L_{aux} = \underbrace{\| \mathbf{z}_{t,0:2} - \mathbf{v}_{gt} \|^2}_{L_{vel}} + \underbrace{\| \mathbf{z}_{t,3:8} - \mathbf{F}_{gt} \|^2}_{L_{foot}} + \underbrace{\| \mathbf{z}_{t,9:38} - \bar{\mathbf{T}} \|^2}_{L_{target}} \quad (5)$$

$\mathbf{v}_{gt} = [v_x, v_y, \omega_z] \in \mathbb{R}^3$ 은 로봇 중심부 좌표계에서 표현된 x , y 축 방향 선속도와 z 축 방향 각속도의 시뮬레이션 참값이다. $\mathbf{F}_{gt} = [F_L, F_R] \in \mathbb{R}^6$, $F_i = [F_{z,now}^i, F_{z,near}^i, F_{z,far}^i]$, $i = L, R$ 은 양발의 수직 접촉력에 대한 시뮬레이션 참값으로, $F_{z,now}^i$ 는 현재 수직 접촉력, $F_{z,near}^i$ 는 $t+1$ 부터 $t+5$ 까지(20~100 ms 후)의 평균, $F_{z,far}^i$ 는 $t+6$ 부터 $t+10$ 까지(120~200 ms 후)의 평균이다. $\bar{\mathbf{T}} = f_{\theta}(o_{t+1})$ 는 타깃 인코더가 다음 시점($t+1$) 관측 o_{t+1} —미래 고유수용성 관측과 특권 정보(motor strength/armature, push force 등)를 함께 포함—을 30차원으로 압축한 표현이다. 타깃 인코더의 출력은 고정된 회귀 목표로 사용되며, 이력 인코더 학습 시 타깃 인코더로는 gradient가 전파되지 않는다. 좌우 대칭성을 활용한 대칭성 손실(mirror loss)^[28]을 추가로 적용한다.

이러한 보조 학습을 통해 인코더는 노이즈가 포함된 고유수용성 이력만으로 현재 접촉력과 미래 접촉력 변화를 추정하게 된다. 특히 $\mathbf{z}_{t,3:8}$ 에 부여된 L_{foot} 는 양발의 현재 수직 접촉력과 미래 접촉력(20~100 ms, 120~200 ms 미래 구간 평균)을 추정하도록 강제하므로, 인코더는 고유수용성 이력에 내재된 접촉력 상승 초기 신호를 포착하여, 접촉력이 최댓값에 도달하기 전에 정책이 선제적으로 대응할 수 있는 정보를 제공한다. 실제 배포 시에도 고유수용성 관측만으로 동작하며, 별도의 힘 센서가 필요하지 않다.

3.3 가상 스프링 보상

본 논문의 핵심 기여는 3.2에서 추정한 현재 및 미래 접촉력을 가상 스프링 보상과 결합하여 선제적으로 접촉력 최댓값을 억제하는 것이다. 추정된 접촉력을 기반으로, 각 발에 가상 스프링을 부착한 것처럼 동작하도록 유도하는 보상을 통해 순응적 착지 동작을 능동적으로 학습하게 한다.

먼저, 순응적 착지 동작을 위한 기준점을 중심부 기준 발 i 의 상대 높이 $h_{d,i}$ 로 다음과 같이 캡처한다:

$$h_{d,i} = h_{foot,i} - h_{base} \quad (6)$$

여기서 $h_{foot,i}$ 와 h_{base} 는 각각 발 i 와 중심부의 높이를 나타낸다. 이때 각 발의 접촉 상태를 매 스텝 추적하여, 스윙(비접촉)에서 지지(접촉)로 전환되는 순간을 초기 접촉으로 감지한다. 본 논문에서는 수직 접촉력이 임계값인 5 N을 초과하면 해당 발이 지면과 접촉 중인 것으로 판정한다. 이 기준 위치는 해당 발이 다시 스윙 구간을 거쳐 새로운 초기 접촉이 발생할 때까지 유지된다.

다음, 기준 접촉력을 다음과 같이 설정한다:

$$F_{z,0} = \alpha \cdot m \cdot g \quad (7)$$

여기서 $F_{z,0}$ 은 기준 접촉력, m 은 로봇의 전체 질량, g 는 중력 가속도, $\alpha (< 1.0)$ 는 기준력 계수이다. 이 기준력은 발마다 독립적으로 적용되는 고정 임계값으로, 현재 지면에 접촉 중인 발의 수에 따라 변하지 않는다. 이는 지지 상태와 관계없이, 개별 발에 가해지는 접촉력이 임계값을 초과하면 중심부 높이를 낮추어 충격을 흡수하도록 하기 위함이다. 지지 형태와 무관하게 각 발의 하드웨어 보호 기준을 일관되게 유지함으로써, 예기치 않은 과부하 상황에서도 구동기와 기구부에 전달되는 충격을 억제한다. 본 연구에서는 정상 보행 중에도 순응 동작이 지속적 으로 유도되도록 α 를 1.0보다 낮게 설정한다. 이를 통해 기준력이 단일 지지 시 정상 하중보다 낮아져, 보행 전반에 걸쳐 충격 흡수 동작이 유도된다.

현재 접촉력이 기준력을 초과하면, 초과분에 비례하여 가상 스프링의 변위가 발생한다:

$$h_{target,i} = h_{d,i} + \frac{\max(0, F_{z,i} - F_{z,0})}{K} \quad (8)$$

여기서 $h_{target,i}$ 는 발 i 의 목표 상대 높이, $F_{z,i}$ 는 현재 수직 접촉력, K 는 가상 스프링 강성이다. $h_{d,i}$ 는 음수(발이 중심부 아래)이므로, 양의 변위항 $\max(0, F_{z,i} - F_{z,0})/K$ 가 더해지면 목표 상대 높이가 $h_{d,i}$ 보다 0에 가까워진다. 이는 중심부가 발 쪽으로 더 가까워져야 함을 의미하며, 정책은 이를 달성하기 위해 무릎을 굽혀 충격을 흡수하는 동작을 학습한다. $F_{z,i} \leq F_{z,0}$ 이면 변위가 0이 되는 단방향 스프링으로, 기준력 이하의 정상 접촉에서는 불필요한 굴곡을 유도하지 않는다.

최종 보상 항은 실제 상대 높이와 목표 높이의 L2 오차로 정의된다:

$$r_{spring} = w_{spring} \sum_{i \in L, R} (h_i - h_{target,i})^2 \quad (9)$$

여기서 h_i 는 발 i 의 현재 중심부 대비 상대 높이이며, 가중치

w_{spring} 을 음수로 두어 오차가 클수록 보상이 감소하는 벌점 항으로 설계하였다. 이 보상은 보행 중 지면과 접촉 중인 발에만 적용된다.

본 논문에서 제안하는 가상 스프링 보상은 추정된 현재 접촉력에 기반해 계산되지만, 이력 인코더가 미래 접촉력을 동시에 예측함으로써 정책은 접촉력 상승을 사전에 인지할 수 있다. 이를 통해 실제 접촉력이 최댓값에 도달하기 이전에 무릎을 굽히는 선제적 착지 전략을 학습하며, 접촉력 변화량에 따라 유연 동작의 정도가 조절된다. 이는 외란 대응 과정이 별도로 학습되는 것이 아닌, 정상 보행에서 학습된 순응성이 외란 상황에도 일반화됨을 의미한다. 접촉력이 임계값을 초과할 때 벌점을 주는 방식^[8]도 과도한 접촉력을 회피하는 동작을 습득할 수 있으나, 순응 동작을 직접 명시하지 않으므로 정책이 그러한 전략을 스스로 탐색해야 하는 부담이 있다.

3.4 보상 함수 설계

전체 보상 함수는 속도 추종, 접촉력 제어, 보행 패턴, 정규화의 네 범주로 구성되며, 각 항목의 함수와 가중치는 [Table 1]에 정리하였다.

명령 선속도 및 각속도 추종에는 보행 로봇 강화학습에서 표준적으로 사용되는 exponential-kernel 보상^[1]을 적용한다. 접촉력 제어 항은 3.3절의 가상 스프링 보상(식 (9))을 중심으로 구성된다. 여기에 접촉력 제한 벌점, 지지 구간 중 발 미끄러짐 억제 벌점, 발바닥이 지면과 수평을 유지하도록 유도하는 벌점을 추가한다. 이때 발 미끄러짐 억제 벌점은 접촉 상태 지표 c_i 를 이용하여 지지 구간에서만 활성화된다. 보행 패턴 항은 보행 주기 $T_{period} = 0.8$ s와 양발 지지 비율 0.20을 갖는 보행 위상 ϕ_i 을 기반으로 스윙 및 지지 구간을 정의한다. 이를 바탕으로 접촉 스케줄 및 체공 시간 보상을 적용하여 주기적인 보행 패턴을 유도한다. 스윙 구간에서는 DWL^[8]의 5차 다항식 높이 궤적을 기준으로, 실제 발 높이가 기준 궤적보다 낮은 경우에만 벌점을 부여하는 단방향 발 높이 궤적 추종을 적용한다. 마지막으로 정규화 항은 자세 안정성, 제어 입력 변화 억제, 관절 토크 제한, 조기 종료 벌점으로 구성된다.

3.5 실측 데이터 기반 시뮬레이션 구동기 모델링

시뮬레이션과 실제 로봇 간의 구동기 동역학 차이는 시뮬레이션 학습 정책을 실제 로봇으로 전이할 때 발생하는 핵심 병목 중 하나이다. 따라서, 본 연구에서는 실제 로봇에서 수집한 데이터를 이용하여, 시뮬레이션 구동기가 실제 구동기와 유사한 동적 응답을 재현하도록 하는 실측 데이터 기반 구동기 모델링 방법을 사용한다^[11].

[Table 1] Reward terms summary. Penalties use negative weights. m : robot mass, ϕ : gait phase, c_i : foot contact

Reward	Function	Weight
<i>Velocity Tracking</i>		
Linear vel	$\exp(-\ v_{xy} - v_{xy}^{cmd}\ ^2/\sigma^2)$	3.5
Angular vel	$\exp(-(\omega_z - \omega_z^{cmd})^2/\sigma^2)$	2.0
<i>Contact-Force Control</i>		
Virtual spring (Ours)	Eq. (9), $K=4000, \alpha=0.9$	-2.0
Force limit	$\sum_i (1 - e^{-e_i^2/\sigma_e^2}),$ $e_i = \max(0, \ \mathbf{F}_i\ - 1.2mg)$	-15
Feet slide	$\sum_i \ \mathbf{v}_{foot,i}^{xy}\ , c_i$	-5.0
Ground parallel	$\sum_i \ (\hat{\mathbf{g}}_i^{foot})_{xy}\ ^2, c_i$	-15
<i>Gait & Swing</i>		
Contact schedule	$\sum_i \mathbb{1}[c_i = \hat{c}_i(\phi)]$	5.0
Air time	$\sum_i t_{air,i}^{\min} \mathbb{1}[t_{air}^{\max} < T_0]$	4.0
Swing clearance	$\sum_i \max(h_i^{ref} - h_i, 0) \mathbb{1}[swing_i]$	-80
<i>Regularization</i>		
Base height	$\exp(-(h - h_d)^2/\sigma_h^2)$	3.0
Flat orientation	$\ \mathbf{g}^{xy}\ ^2$	-30
Vertical vel.	v_z^2	-0.3
Action rate	$\ \mathbf{a}_i - \mathbf{a}_{i-1}\ ^2$	-0.1
Joint torques	$\ \boldsymbol{\tau}\ ^2$	-10^{-5}
Joint accel.	$\ \dot{\mathbf{q}}\ ^2$	-5×10^{-7}
Termination	$\mathbb{1}[\text{terminated}]$	-200

먼저, 병렬 사절 링크 구조와 모터 구동기에서 발생하는 비선형성을 시뮬레이션상에서 재현하기 위해 LSTM 기반 구동기 네트워크를 학습하였다. 본 논문에서 사용된 이족 보행 로봇은 무릎과 발목에 병렬 사절 링크 구조를 포함하고 있으며, 이로 인해 모터 공간과 관절 공간 사이에 각도 의존적인 비선형 토크 전달 관계가 발생한다. 직렬 관절 모델로 근사된 시뮬레이터에서는 이러한 사절 링크 전달 특성과 모터의 비이상적 응답을 해석적으로 모두 반영하기 어렵다. 따라서 Hwangbo 등^[11]의 구동기 네트워크 접근을 참고하여, 관절 위치 오차와 관절 속도 이력을 입력으로 받고 실제 로봇에서 측정된 토크 응답을 출력으로 예측하는 관절별 독립 LSTM 네트워크를 학습하였다. 이를 통해 모터 마찰, 통신 지연, 감속기 비선형성, 사절 링크의 각도 의존적 토크 전달 효과가 내재화된 관절 공간 구동기 모델을 시뮬레이터에 적용하였다.

추가로, 학습된 구동기 네트워크는 실제 토크 응답을 데이터 기반으로 근사하는 모델이므로, 전기자 관성, 점성 감쇠, 쿨롱

마찰과 같은 물리 매개변수를 명시적으로 분리하여 제공하지 않는다. 따라서 학습된 구동기 네트워크를 시뮬레이터의 구동기 모델로 탑재한 이후, PACE^[10]의 방법론을 참고하여 실제 로봇의 chirp 궤적 추종 응답을 가장 잘 재현하는 전기자 관성, 점성 감쇠, 쿨롱 마찰값을 추가로 식별하였다. 최종적으로, 데이터 기반 구동기 네트워크와 식별된 물리 매개변수를 함께 사용함으로써 병렬 사절 링크 및 모터 구동계의 비선형성을 시뮬레이션에서 보다 정확하게 모델링하였다.

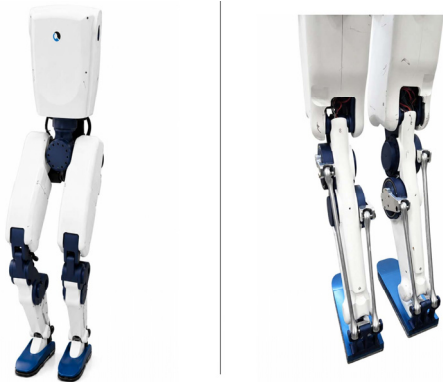
4. 실험

본 장에서는 제안하는 방법의 유효성을 검증한다. 먼저 실험 환경과 비교 정책을 소개하고, 이력 인코더가 고유 수용성 이력만을 이용하여 현재 및 가까운 미래 발과 지면 사이의 접촉력을 추정할 수 있음을 정량적으로 평가한다. 이후 외란 환경에서 추정된 접촉력과 가상 스프링 보상이 실제 접촉력 역제로 이어짐을 검증하고, 잠재 표현 제거 분석을 통해 정책이 접촉력 추정 정보를 실제로 활용함을 확인한다. 마지막으로 실제 로봇으로의 이식 결과를 보고한다.

4.1 실험 환경

4.1.1 로봇 플랫폼

본 실험에 사용된 휴머노이드 로봇[Fig. 3]은 하체 12개 관절과 허리 1개 관절을 포함하여 총 13개의 구동 관절로 구성된 이족 보행 로봇이다. 학습 정책은 하체 12개 관절을 제어하며, 허리 관절은 초기 자세로 고정된다. 로봇의 전체 질량은 약 60 kg 이고, 키는 150 cm이다. 무릎과 발목 관절에는 사절 링크 기반의 병렬 기구가 적용되어있으며, 이로 인해 관절 각도에 따라 모터 토크와 관절 토크 사이의 관계가 비선형적으로 변화한다.



(a) Overall robot configuration (b) 4-bar linkage mechanism

[Fig. 3] Experimental humanoid robot. (a) Overall robot with 13 actuated joints (12 lower-body + 1 waist). (b) Four-bar linkage structure of the knee and ankle joints

4.1.2 시뮬레이션 환경

제안하는 정책을 학습하기 위한 시뮬레이션 환경은 IsaacLab^[29]을 사용하여 구축하였다. 정책 학습은 4,096개의 병렬 환경에서 수행되었으며, 정책은 PPO^[25]를 사용하여 50 Hz의 제어 주기로 학습하였다. 시뮬레이션에서 학습된 정책이 실제 로봇의 동역학 차이를 포괄할 수 있도록 도메인 무작위화를 적용하였으며, 무작위화한 매개변수 범위는 [Table 2]에 정리하였다. 다만 구동기 모델에는 실제 구동기 동역학을 학습한 LSTM 기반 구동기 네트워크가 적용되어 있으므로, 모터 출력 계수(motor strength)는 $\pm 5\%$ 범위에서만 보수적으로 무작위화하였다.

본 논문의 모든 실험은 해당 구동기 모델이 적용된 시뮬레이션 환경에서 수행하였다.

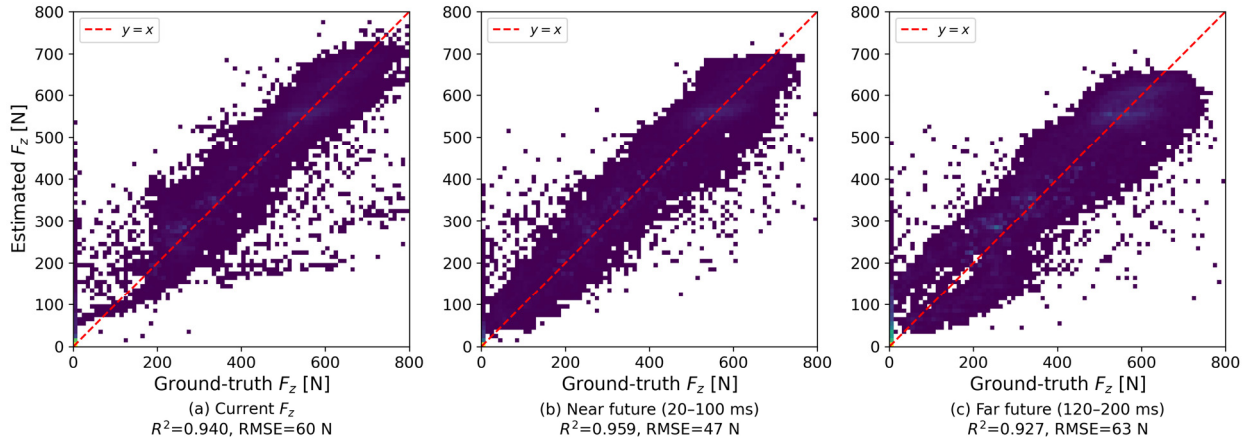
4.1.3 비교 정책

비교 대상 정책은 다음 네 가지로 구성한다. 네 정책은 동일한 네트워크 구조와 학습 설정을 공유한다. 속도 추종, 보행 패턴, 정규화 보상뿐 아니라 접촉력 제한 벌점, 발 미끄러짐 억제 벌점, 발바닥 수평 유지 벌점과 같은 공통 안정화 보상 또한 [Table 1]의 가중치를 그대로 유지한다. 정책 간 차이는 접촉력 추정 방식과 이에 의존하는 보상 항에 국한되며, 정책별 보상 구성을 [Table 3]에 정리하였다.

- **No-Est:** 접촉력 추정과 이에 의존하는 보상을 제거한 기준 정책. 공통 안정화 보상은 모두 유지한 채 가상 스프링 보상의 가중치만 0으로 설정
- **Vel-Track:** DWL^[8] 방식의 스윙 궤적 속도 추종 보상을 적용한 정책. 접촉력을 추정하지 않고, 스윙 궤적 추종으로 간접적으로 접촉력을 제어. 네 정책 중 유일하게 스윙 구간 보상이 5차 다항식 높이 및 속도 추종으로 대체됨.
- **CF-Est (curr):** 현재 접촉력만 추정하고 미래 접촉력 보조 손실을 제거한 비교 정책. 가상 스프링 보상은 동일하게 적용하되, 미래 예측의 인과적 기여를 분리

[Table 2] Domain Randomization Parameters

Parameter	Range	Unit
Ground friction	[0.2, 1.4]	-
Link mass scale	[0.8, 1.2]	-
Base mass offset	[-5, 10]	kg
Base COM offset	± 0.10	m
Motor strength	[0.95, 1.05]	-
Push force (x, y)	[-60, 60]	N
Push force (z)	[-30, 30]	N
Push duration	[0.5, 2.0]	s
Push interval	[5, 10]	s



[Fig. 4] Contact force estimation accuracy vs. ground truth, by prediction horizon, under the nominal condition (0.3 m/s, flat). Each panel is a 2D histogram of estimated vs. actual vertical foot force (red dashed: $y = x$). The estimator tracks ground truth closely at all horizons ($R^2 \geq 0.92$); accuracy decreases with prediction distance among the smoothed future targets (near $\rightarrow$ far). The remaining conditions are reported in [Table 4]

- **CF-Est (Ours):** 접촉력 추정과 가상 스프링 보상을 적용한 제안 정책. 잠재 벡터 $z_{t,3:8}$ 에서 현재 및 미래 접촉력을 추정하고, 가상 스프링 보상을 사용

4.2 접촉력 추정 정확도

본 절에서는 제안하는 프레임워크의 이력 인코더가 고유수용성 이력만을 이용하여 현재 시점의 발 지면 접촉력을 추정하

고, 근미래 및 원미래 시점의 접촉력을 예측할 수 있는지 검증한다. 이를 위해 64개의 병렬 환경에서 60초 동안 0.3 m/s의 평지 직진 보행을 수행하고, 잠재 벡터 $z_{t,3:8}$ 가 출력한 현재(t), 근미래($t+20-100ms$), 원미래($t+120-200ms$) 수직 접촉력을 시뮬레이터의 실측 접촉력과 비교한다. 검증이 단일 조건에 한정되지 않도록, 동일한 학습 결과를 재학습 없이 사용하여 보행 속도, 선회, 외란, 지면 형상을 달리한 네 조건에서도 같은 방식으로 평가한다.

[Table 4]와 [Fig. 4]에 접촉력 추종 결과를 제시한다. 접촉력 추정 정확도는 평균 제곱근 오차(root mean square error, RMSE)와 결정계수(R^2)로 평가한다. 기준 조건에서 현재·근미래·원미래 세 시점 모두 $R^2 \geq 0.92$ 를 달성하여, 잠재 벡터 $z_{t,3:8}$ 가 실측 접촉력의 변화 양상을 정확하게 반영함을 확인하였다. 미래 접촉력 추정 정확도는 예측 시점이 멀어질수록 감소하는데, 현재와 과거의 관측 이력만을 기반으로 추정되는 미래 접촉력은 시간적 거리가 멀어질수록 불확실성이 증가하기 때문이다

[Table 4] Contact-force estimation accuracy across evaluation conditions. Each cell reports the current (t), near future ($t+20-100$ ms), and far future ($t+120-200$ ms) horizons. The same policy is evaluated in every condition without retraining; near future is the most accurate horizon throughout (bold)

Component	No-Est	Vel-Track	CF-Est (curr)	CF-Est (Ours)
Contact-force estimation				
Current Fz	-	-	✓	✓
Future Fz	-	-	-	✓
Estimation-dependent reward				
Virtual spring	-	-	✓	✓
Swing-phase reward				
One-sided height prof.	✓	-	✓	✓
Quintic vel. tracking	-	✓	-	-
Common stabilization				
Force limit	✓	✓	✓	✓
Feet slide	✓	✓	✓	✓
Ground parallel	✓	✓	✓	✓
Velocity, gait, regularization				
All terms	✓	✓	✓	✓

Evaluation condition	RMSE [N]	R^2
	current / near / far	current / near / far
Nominal (0.3 m/s, flat)	59.9 / 47.3 / 63.3	0.940 / 0.959 / 0.927
Fast walking (0.6 m/s)	72.0 / 48.7 / 61.7	0.914 / 0.957 / 0.931
Turning (0.5 rad/s)	61.6 / 47.0 / 63.6	0.937 / 0.960 / 0.927
Push recovery (60 N)	110.3 / 77.5 / 103.6	0.816 / 0.889 / 0.802
Uneven ground (± 1 cm)	128.7 / 87.4 / 107.7	0.760 / 0.857 / 0.783

(near $R^2 = 0.96 \rightarrow$ far $R^2 = 0.93$). 현재 시점의 결정계수($R^2 = 0.94$)가 near 시점의 결정계수보다 약간 낮은 것은 현재 시점의 순간 접촉력이 착지 충격에 의한 고주파 피크(최대 약 5.5 kN)를 포함해 변동이 큰 반면 near와 far 시점의 학습 목표는 5-step 평균으로 평활화되어 상대적으로 추정이 용이하기 때문이다. 기준 조건의 접촉력 추정 RMSE는 47-63 N으로, 이는 정상 단일지지 하중(약 591 N)의 8-11%에 해당한다. 따라서 제안한 추정기가 접촉력을 충분히 정확하게 추정함을 확인할 수 있다. 이 결과는 3.3절에서 제안한 선제적 순응 착지가 신뢰도 높은 미래 접촉력 추정에 기반하고 있음을 정량적으로 뒷받침한다.

추정 성능이 단일 평가 조건에 국한되지 않음을 확인하기 위해, 동일한 정책을 재학습 없이 네 조건에서 추가로 평가하였다: 보행 속도를 두 배로 높은 고속 보행(0.6 m/s), 선회 명령을 부여한 방향 전환(0.5 rad/s), 보행 중 60 N 외란을 반복 인가한 외란 복구, 그리고 학습 중 경험하지 않은 비평탄 지면(± 1 cm)이다. 방향 전환과 고속 보행에서는 정확도가 기준 조건과 사실상 동일하였고(근미래 결정계수 0.960, 0.957), 외란 복구와 비평탄 지면에서는 현재 시점의 결정계수가 각각 0.816, 0.760으로 저하되었다. 다섯 조건 전부에서 근미래 시점이 세 시점 중 가장 정확하며, 조건이 가혹해질수록 현재 시점의 정확도가 가장 먼저 저하된다. 순간 접촉력은 착지 시점의 변동에 직접 노출되는 반면, 5-step 평균으로 평활화된 근미래 목표는 그 변동을 흡수하기 때문이다. 저하 폭 또한 근미래가 가장 작아(최대 0.10, 현재 시점은 0.18), 선제적 순응 착지가 의존하는 근미래 접촉력 예측이 평가 조건 변화에 대해 가장 강건한 채널임을 확인하였다.

4.3 외란 환경에서의 접촉력 비교

본 절에서는 외란 환경에서 추정된 접촉력이 가상 스프링 보상과 결합하여 발 지면 충격 완화를 유도할 수 있는지 검증한다. 실제 운용 시나리오를 모사하기 위해, 장시간 연속 보행 중 주기적으로 외란을 인가하고 발생하는 발과 지면 사이의 접촉력의 통계적 분포를 비교한다.

4.3.1 실험 설계

네 정책을 동일한 시뮬레이션 조건에서 평가한다. 평지에서 0.3 m/s 직진 명령을 부여하며, 외란 인가 프로토콜은 [Table 5]에 정리하였다. $\pm x$, $\pm y$ 네 방향과 30 N, 60 N 두 크기의 조합을 각 5회씩 반복하여, 모든 정책에 동일한 외란 시퀀스(총 40회, 10 s 간격, 1.0 s 지속)를 적용함으로써 외란 횟수와 조건을 통제한다. 다양한 방향과 크기의 외란을 연속 보행 중 인가하는 방식을 채택하여, 단일 외란 조건의 반복 시행으로는 평가하기 어려운 보행 전체의 접촉력 분포 특성과 과도 응답을 함께 분석한다.

[Table 5] Disturbance evaluation protocol, applied identically to all four policies

Parameter	Value
Terrain	Flat, uneven (± 1 cm)
Command velocity	0.3 m/s (forward)
Push direction	$\pm x, \pm y$
Push magnitude	30, 60 N
Push interval	10 s
Push duration	1.0 s
Repetitions per condition	5
Total pushes	40
Total walking time	400 s

또한 학습 중 경험하지 않은 지면 조건에서의 일반화를 확인하기 위해, 동일한 외란 프로토콜을 유지한 채 지면 형상만 변경한 조건을 함께 평가한다. 시뮬레이션 지면을 최대 높이 편차 ± 1 cm, 대표 폭 약 0.3 m의 비평탄 지형으로 교체하였으며, 네 정책은 모두 동일한 지면에서 평가하였다. 최대 높이 편차는 스윙 발 최고 높이의 8%에 해당하고, 지형의 대표 폭이 발 길이(0.2 m)와 비슷하므로 매 보행 주기마다 착지 시점이 달라진다.

4.3.2 평가 지표

접촉력 분포 분석을 위해 다음 지표를 사용한다:

- **표준편차 (Std):** 접촉력 응답의 일관성. 작을수록 스텝 간 접촉력 변동이 적다.
- **분포 백분위수 (P95, P99, Max):** 접촉력 분포의 상위 극단을 나타내며, 값이 클수록 드물지만 큰 충격이 발생함을 의미한다. 하드웨어 보호 관점의 핵심 지표이다.
- **하중 증가율 (loading rate) $\dot{F}_z = dF_z/dt$:** 접촉력의 시간 변화율. 충격성 하중(impulsive load)의 크기를 직접 반영한다.
- **외란당 최대 접촉력:** 각 외란 인가 시점부터 3 s 구간의 최대 수직 접촉력. 분포 전체의 최댓값이 단일 사건에 좌우되는 것과 달리, 외란 횟수만큼의 표본을 제공하므로 정책 간 차이를 통계적으로 검정할 수 있다.

4.3.3 결과 및 분석

[Table 6]과 [Fig. 5]에 네 정책의 접촉력 비교 결과를 제시한다.

표의 각 항목은 평탄/비평탄 지형 순이며, 본 절에서는 평탄 지형을 기준으로 분석한다. 비평탄 지형에 대한 결과는 4.3.4절에서 다룬다.

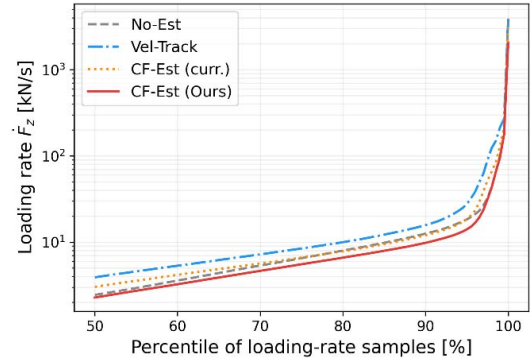
[Table 6] Contact force statistics on flat and uneven (± 1 cm) terrain. Each entry is flat / uneven. The disturbance protocol is identical in both conditions, and training was performed on flat terrain only. Best results in bold

Metric	No-Est	Vel-Track	CF-Est (curr)	CF-Est (Ours)
<i>Mean support [N]</i>				
Mean	591 / 591	591 / 591	591 / 591	591 / 591
<i>Distribution [N]</i>				
Std	133 / 150	153 / 164	141 / 161	110 / 130
P95	848 / 851	857 / 888	847 / 867	826 / 854
P99	1016 / 1020	962 / 1146	946 / 1055	901 / 1018
Max	5112 / 4907	4613 / 5029	4050 / 5416	3225 / 3647
<i>Per-disturbance peak [N]</i>				
Mean	2170 / 2379	2140 / 2155	1931 / 2267	1267 / 1535
<i>Loading rate $\dot{F}_z$ [kN/s]</i>				
P95	18 / 29	28 / 35	18 / 33	14 / 27
Max	3972 / 4630	3948 / 3007	3475 / 3691	2030 / 3034

접촉력 평균의 불변성: 측정된 수직 접촉력의 평균은 네 정책 모두 체중(약 60 kg, 591 N)의 0.05% 이내로 분포한다. 이는 정상 보행 시 평균 지면반력이 중력을 상쇄해야 하므로 기대되는 결과이며, 정책에 따른 유의미한 차이가 발생하기 어렵다. 따라서 접촉력 평균의 유사성은 접촉력 관련 보상의 도입이 정상 보행의 기본적인 동역학적 특성을 유지하면서 동작함을 시사한다.

접촉력 최대값 비교: 평지 조건에서 최대 접촉력은 No-Est 5,112 N, Vel-Track 4,613 N, CF-Est (curr) 4,050 N에 비해 CF-Est (Ours)는 3,225 N으로 No-Est 대비 약 36.9%, CF-Est (curr) 대비 약 20.4% 감소하였다. CF-Est (curr)는 현재 접촉력만 추정하여 No-Est 대비 20.8% 감소에 그쳐, 미래 접촉력 예측이 최대 접촉력 완화에 중요하다는 것을 보여준다. 접촉력의 표준편차 역시 CF-Est (Ours)가 110 N으로 No-Est (133 N), Vel-Track (153 N), CF-Est (curr) (141 N) 대비 가장 작아 접촉력 응답의 일관성이 가장 높다. 특히 Vel-Track은 No-Est보다 오히려 더 넓은 분포를 보이는데, 이는 속도 추종 보상이 스윙 궤적을 강하게 제약해 외란 후 순응적 착지 동작을 방해하기 때문으로 해석된다.

하중 증가율 억제: 평지 조건의 [Fig. 5]에서 CF-Est (Ours)의 하중 증가율 $\dot{F}_z$ 는 평가지표 전반에서 가장 낮게 나타났다(P95 기준 14 kN/s로 No-Est 18, Vel-Track 28, CF-Est (curr) 18 대비 각각 22%, 50%, 22% 감소). CF-Est (curr)는 No-Est와 유사한 수준에 머물러, 현재 접촉력만으로는 하중 증가율 억제에 한계



[Fig. 5] Loading rate $\dot{F}_z = dF_z/dt$ distribution under the disturbance protocol. Each curve is the quantile function of the per-step positive loading rate. CF-Est (Ours, red) is the lowest at every percentile, showing that the proposed policy builds up contact force the most gradually — the signature of a learned compliant landing

가 있음을 보인다. 평균 접촉력이 체중으로 동일한 상황에서 하중 증가율이 낮다는 것은, 동일한 충격량을 더 긴 시간에 걸쳐 완만하게 전달함을 의미하며, 이는 가상 스프링이 급격한 충격력의 변화를 억제하는 메커니즘과 일치한다. P95, 최대값 또한 CF-Est (Ours)가 가장 낮은 것으로 나타났다.

논의: CF-Est (Ours)가 과도한 접촉력을 억제하는 메커니즘은 가상 스프링 보상과 미래 접촉력 추정의 결합으로 설명된다. 이력 인코더의 잠재 벡터 $z_{t,3:8}$ 가 현재 접촉력과 함께 20-200 ms 후의 접촉력을 예측하므로, actor는 착지 전부터 무릎 굽힘을 유도하여 충격을 분산시킨다. CF-Est (curr)는 미래 예측 없이 현재 접촉력만 추정하므로, 최대 접촉력은 No-Est 대비 감소하나 하중 증가율(P95)은 동일한 수준에 머문다. 이는 선제적 순응 동작이 미래 예측에 의존함을 시사한다.

4.3.4 비평탄 지형에서의 일반화

학습은 평탄 지형에서만 수행되었으므로 비평탄 지형은 학습 분포 밖 조건이다. 비평탄 지형에서는 매 보행 주기마다 예상보다 이르거나 늦은 접촉이 발생하며, 이는 하중 증가율 P95가 평탄 지형 14 kN/s에서 비평탄 지형 27 kN/s로 증가한 데서 확인된다. 즉 지면 형상 변화가 의도한 대로 조기 접촉 조건을 만들어냈다.

이 조건에서 CF-Est (Ours)의 외란당 최대 접촉력은 $1,535 \pm 106$ N으로, No-Est($2,379 \pm 170$ N), Vel-Track($2,155 \pm 142$ N), CF-Est (curr)($2,267 \pm 182$ N) 대비 각각 35.5%, 28.7%, 32.3% 낮았다(평균 $\pm$ 표준오차, 모두 $p < 0.001$). 평탄 지형에서의 저감 폭이 각각 41.6%, 40.8%, 34.4%였음을 고려하면, 제안 방법의 접촉력 저감 효과는 학습 범위 밖 지면에서도 유지된다. 특히

CF-Est (curr) 대비 저감 폭은 두 지면에서 유사하게 유지되어, 미래 접촉력 예측의 기여가 지면 조건에 의존하지 않음을 보인다.

한편 네 정책의 P95(851-888 N)와 P99(1,018-1,146 N)는 비평탄 지형에서도 서로 유사하다. 즉 정책 간 차이는 정상 보행 구간이 아니라 외란 직후 접촉력이 급증하는 순간에 국한되며, 이는 제안 방법의 이득이 상시적인 컴플라이언스 증가가 아니라 접촉력 예측에 기반한 선제적 순응에서 비롯됨을 뒷받침한다.

4.4 잠재 표현 제거 분석

본 절에서는 제안하는 정책(CF-Est)이 발 지면 접촉력 추정 정보를 실제로 활용하는지 정량적으로 분석한다. 잠재 벡터 $z_t \in \mathbb{R}^{64}$ 의 특정 영역을 마스킹하고 action 변화량을 측정하는 제거(ablation) 분석^[30]을 수행한다.

4.4.1 마스킹 기반 제거 분석

잠재 벡터의 부분 영역을 0으로 마스킹한 후, 원래 action과의 L2 거리를 측정한다:

$$\Delta a_{\text{mask}} = \|\pi(o_t, z_t) - \pi(o_t, z_t^{\text{mask}})\|_2 \quad (10)$$

여기서 z_t^{mask} 은 정규화 좌표에서 부분 영역을 0으로 치환한 잠재 벡터이다. Δa_{mask} 이 클수록 해당 영역이 action 결정에 큰 영향을 미친다는 것을 의미한다.

분석 대상 영역은 다음과 같다:

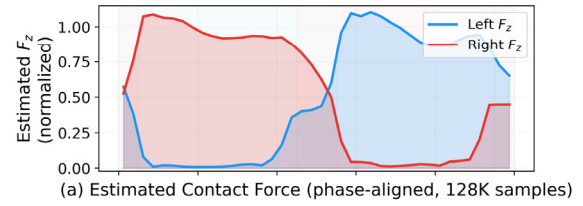
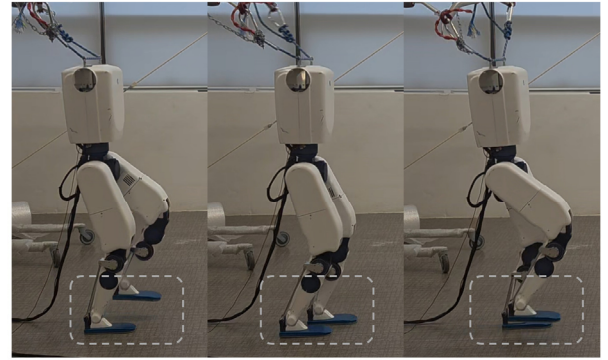
- $z_{t,0:2}$: Base velocity estimation
- $z_{t,3:8}$: Contact force estimation
- $z_{t,9:38}$: Future privileged-state encoding

64개 병렬 환경에서 0.3 m/s 직진 보행을 60초간 수행하여 데이터를 수집하고, 각 샘플의 보행 위상을 관측의 ($\sin \phi_t$, $\cos \phi_t$)로부터 복원하여 위상 정렬 평균을 산출한다.

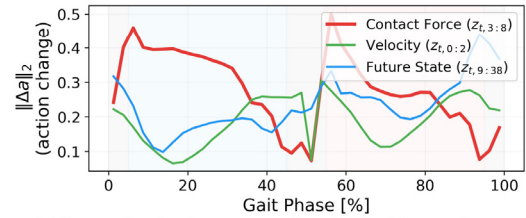
4.4.2 결과

[Fig. 6]에 보행 위상에 따른 잠재 벡터 영역별 action 기여도 변화를 제시한다.

위상별 접촉력 기여도 [Fig. 6(b)]에서 접촉력 영역의 기여도는 각 발이 지면에 닿는 착지(스윙에서 지지로의 전환) 순간에 최댓값을 보이며(위상 약 6%, 56%), 이는 추정 접촉력이 가파르게 상승하는 하중 구간과 일치한다. 이후 지지 구간 동안 접촉력의 크기가 높게 유지됨에도 기여도는 점차 감소하고, 양발 지지 구간(위상 약 50%, 95%)에서 최저로 떨어진다. 이는 정책



(a) Estimated Contact Force (phase-aligned, 128K samples)



(b) Latent Region Importance ($r = 0.58$ with max F_z , $p < 0.001$)

[Fig. 6] Latent ablation analysis across gait phases. (a) Phase-aligned estimated contact force shows clear left/right alternation. (b) Contact force region importance ($z_{t,3:8}$, red) peaks at each foot's touchdown (the swing-to-stance transition, phase $\approx 6\%$, 56%), aligning with the steep loading edge of the estimated F_z in (a), then decays through stance and drops to its minimum during double support

이 접촉력 추정 정보를 착지 충격이 발생하는 하중 순간에 선택적으로 가장 강하게 활용함을 보여주며, 충격 완화라는 본 연구의 목적과 부합한다.

4.5 실제 로봇 전이

학습된 CF-Est (Ours) 정책을 별도의 미세 조정이나 보정 없이 실제 휴머노이드 로봇[Fig. 3]에 zero-shot으로 이식하였다. 정책은 실제 로봇에서 안정적으로 보행하였으며, [Fig. 1]과 같이 외란이 인가되는 상황에서도 가상 스프링 보상으로 학습된 순응적 무릎 굽힘을 동반하여 균형을 회복하였다. 이는 시뮬레이션에서 학습된 접촉력 인지 순응 보행 정책이 추가 학습 없이 실제 하드웨어로 전이됨을 정성적으로 보여준다.

4.5.1 실제 로봇에서의 접촉력 정량 평가

실제 로봇에는 발바닥 힘/토크 센서가 장착되어 있지 않으므

[Table 7] Vertical ground reaction force on the real robot during 0.2 m/s forward walking, estimated via Eq. (11)

Metric	No-Est	CF-Est	Reduction
Mean [N]	588	591	-
Std [N]	217	90	-58.7%
P95 [N]	950	715	-24.7%
Loading rate P95 [kN/s]	87.1	59.5	-31.7%

로, 몸통의 수직 가속도로부터 전신 수직 지면 반력을 역산하여 착지 충격을 정량적으로 비교하였다. 로봇 전체를 하나의 강체로 보면 수직 방향 운동 방정식으로부터 지면 반력은 다음과 같이 주어진다.

$$F_z = m(g + a_z) \quad (11)$$

여기서 m 은 로봇 질량, g 는 중력 가속도, a_z 는 상태 추정기가 제공하는 몸통 수직 속도를 미분하여 얻은 수직 가속도이다. 평지에서 0.2 m/s 직진 명령을 인가한 구간을 대상으로 No-Est와 CF-Est 정책을 비교하였으며, 결과를 [Table 7]에 제시한다.

평균 지면 반력은 두 정책 모두 체중과 0.5% 이내로 일치하여 역산이 물리적으로 타당함을 확인하였다. 반면 분포의 상위 구간에서는 CF-Est가 P95 기준 24.7% 낮았고, 하중 증가율(P95)은 31.7% 낮았다. 접촉력 표준편차 또한 58.7% 감소하여 보행 주기 전반에서 접촉력 변동이 억제되었다. 평균은 유지된 채 분포의 상위 구간과 하중 증가율만 감소하는 양상은 시뮬레이션에서 관측된 경향과 일치한다.

다만 식 (11)은 로봇을 단일 강체로 근사하므로, 상태 추정기가 제공하는 몸통 가속도와 실제 질량중심 가속도의 차이, 즉 다리 스윙에 의한 항을 포함하지 못한다. 또한 역산은 양발의 합력만을 제공하여 양발 지지 구간에서 발별 착지 충격이 부분적으로 상쇄되며, 하중 증가율의 절대값은 차분 간격에 의존한다. 따라서 시뮬레이션의 절대값과 직접 비교하지 않으며, 본 결과는 실제 로봇에서의 경향에 대한 참고 자료로 제시한다.

5. 결 론

본 논문에서는 이족 보행 로봇의 순응적 보행을 위한 접촉력 인지 강화학습 프레임워크를 제안하였다. 제안하는 방법은 고 유수용성 이력으로부터 발과 지면 사이의 현재 및 미래 접촉력을 추정하고, 이를 가상 스프링 보상에 활용하여 선제적으로 순응적 동작을 유도한다. 접촉력을 별점화하거나 사전에 계획된 발 궤적에 의존하는 기존 방식과 달리, 순응 동작 자체를 직접 유도하여 정상 보행뿐 아니라 외란에 의한 예기치 않은 접촉에서도 접촉 충격을 효과적으로 완화할 수 있었다. 외란 환경 비

교 실험에서 제안하는 방법은 추정을 사용하지 않는 기준 정책 대비 최대 접촉력과 하중 증가율을 모두 가장 낮은 수준으로 완화하였으며, 잠재 표현 제거 분석을 통해 정책이 접촉력 추정 정보를 실제로 활용함을 확인하였다. 또한 학습 중 경험하지 않은 비평탄 지형에서도 외란당 최대 접촉력의 저감 효과가 유지되어, 제안 방법이 학습 범위 밖 지면 조건으로 일반화됨을 확인하였다.

또한 사절 링크 구조의 실제 휴머노이드 로봇으로 시뮬레이션 학습 정책을 전이하기 위해, 모터 파라미터 추정과 구동기 네트워크를 결합한 구동기 모델링 방법을 구축하였다. 이를 통해 학습된 접촉력 인지 정책을 추가 학습이나 보정 없이 실제 로봇에 적용하였으며, 안정적인 보행과 외란 대응이 가능함을 확인하였다. 몸통 수직 가속도로부터 역산한 전신 지면 반력에서도 접촉력 P95와 하중 증가율이 기준 정책 대비 각각 24.7%, 31.7% 낮아, 시뮬레이션에서 관측된 경향이 실제 로봇에서도 나타남을 확인하였다.

향후 연구에서는 발바닥 힘/토크 센서나 지면반력기(force plate)를 이용한 직접 계측으로 실제 로봇에서의 충격 완화 효과를 검증하고, height scan과 같은 외수용 정보를 결합하여 계단·경사와 같이 높이 변화가 큰 지형으로 접촉력 인지 보행을 확장하며, 나아가 달리기·점프와 같이 충격이 큰 동적 동작으로의 적용을 진행할 예정이다.

References

- [1] N. Rudin, D. Hoeller, P. Reist, and M. Hutter, "Learning to walk in minutes using massively parallel deep reinforcement learning," in *Conference on Robot Learning (CoRL)*, London, UK, pp. 91-100, 2021, [Online], <https://proceedings.mlr.press/v164/rudin22a.html>.
- [2] X. B. Peng, P. Abbeel, S. Levine, and M. van de Panne, "DeepMimic: Example-guided deep reinforcement learning of physics-based character skills," *ACM Transactions on Graphics*, vol. 37, no. 4, Art. no. 143, Jul., 2018, DOI: 10.1145/3197517.3201311.
- [3] J. Siekmann, Y. Godse, A. Fern, and J. Hurst, "Sim-to-real learning of all common bipedal gaits via periodic reward composition," in *IEEE International Conference on Robotics and Automation (ICRA)*, Xi'an, China, pp. 7309-7315, 2021, DOI: 10.1109/ICRA48506.2021.9561814.
- [4] J. Jo, G. Park, and Y. Oh, "Robust walking stabilization strategy of humanoid robots on uneven terrain via QP-based impedance/admittance control," *Robotics and Autonomous Systems*, vol. 154, Art. no. 104148, Aug., 2022, DOI: 10.1016/j.robot.2022.104148.
- [5] B. R. P. Singh and R. Featherstone, "Mechanical shock propagation reduction in robot legs," *IEEE Robotics and Automation Letters*, vol. 5, no. 2, pp. 1183-1190, Apr., 2020, DOI: 10.1109/LRA.2020.2966395.

- [6] S.-H. Hyon, J. G. Hale, and G. Cheng, "Full-body compliant human-humanoid interaction: Balancing in the presence of unknown external forces," *IEEE Transactions on Robotics*, vol. 23, no. 5, pp. 884-898, Oct., 2007, DOI: 10.1109/TRO.2007.904896.
- [7] B. J. Stephens and C. G. Atkeson, "Dynamic balance force control for compliant humanoid robots," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Taipei, Taiwan, pp. 1248-1255, 2010, DOI: 10.1109/IROS.2010.5648837.
- [8] X. Gu et al., "Advancing humanoid locomotion: Mastering challenging terrains with denoising world model learning," in *Robotics: Science and Systems (RSS)*, Delft, Netherlands, Jul., 2024, DOI: 10.15607/RSS.2024.XX.058.
- [9] R. Watanabe et al., "Learning quiet walking for a small home robot," in *IEEE International Conference on Robotics and Automation (ICRA)*, Atlanta, GA, USA, pp. 15285-15291, 2025, DOI: 10.1109/ICRA55743.2025.11128174.
- [10] F. Bjelonic, F. Tischhauser, and M. Hutter, "Towards bridging the gap: Systematic sim-to-real transfer for diverse legged robots," arXiv preprint *arXiv:2509.06342*, 2025, DOI: 10.48550/arXiv.2509.06342.
- [11] J. Hwangbo et al., "Learning agile and dynamic motor skills for legged robots," *Science Robotics*, vol. 4, no. 26, eaau5872, Jan., 2019, DOI: 10.1126/scirobotics.aau5872.
- [12] J. Lee, J. Hwangbo, L. Wellhausen, V. Koltun, and M. Hutter, "Learning quadrupedal locomotion over challenging terrain," *Science Robotics*, vol. 5, no. 47, Art. no. eabc5986, Oct., 2020, DOI: 10.1126/scirobotics.abc5986.
- [13] X. Cheng, Y. Ji, J. Chen, R. Yang, G. Yang, and X. Wang, "Expressive whole-body control for humanoid robots," in *Robotics: Science and Systems (RSS)*, Delft, Netherlands, Jul., 2024, DOI: 10.15607/RSS.2024.XX.107.
- [14] Z. Zhuang, S. Yao, and H. Zhao, "Humanoid parkour learning," in *Conference on Robot Learning (CoRL)*, Munich, Germany, pp. 1975-1991, 2024, [Online], <https://proceedings.mlr.press/v270/zhuang25a.html>.
- [15] J. Long et al., "Learning humanoid locomotion with perceptive internal model," in *IEEE International Conference on Robotics and Automation (ICRA)*, Atlanta, GA, USA, pp. 9997-10003, 2025, DOI: 10.1109/ICRA55743.2025.11128333.
- [16] Y. Zhang et al., "QuietPaw: Learning quadrupedal locomotion with versatile noise preference alignment," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Hangzhou, China, pp. 12561-12568, 2025, DOI: 10.1109/IROS60139.2025.11246267.
- [17] H.-O. Lim, S. A. Setiawan, and A. Takanishi, "Position-based impedance control of a biped humanoid robot," *Advanced Robotics*, vol. 18, no. 4, pp. 415-435, Jan., 2004, DOI: 10.1163/156855304773822491.
- [18] J. Park and O. Khatib, "Contact consistent control framework for humanoid robots," in *IEEE International Conference on Robotics and Automation (ICRA)*, Orlando, FL, USA, pp. 1963-1969, 2006, DOI: 10.1109/ROBOT.2006.1641993.
- [19] T. Koolen et al., "Design of a momentum-based control framework and application to the humanoid robot Atlas," *International Journal of Humanoid Robotics*, vol. 13, no. 1, Art. no. 1650007, Mar., 2016, DOI: 10.1142/S0219843616500079.
- [20] A. Herzog, L. Righetti, F. Grimmering, P. Pastor, and S. Schaal, "Balancing experiments on a torque-controlled humanoid with hierarchical inverse dynamics," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Chicago, IL, USA, pp. 981-988, 2014, DOI: 10.1109/IROS.2014.6942678.
- [21] T. Portela, G. B. Margolis, Y. Ji, and P. Agrawal, "Learning force control for legged manipulation," in *IEEE International Conference on Robotics and Automation (ICRA)*, Yokohama, Japan, pp. 15366-15372, 2024, DOI: 10.1109/ICRA57147.2024.10611066.
- [22] B. Xu, H. Weng, Q. Lu, Y. Gao, and H. Xu, "FACET: Force-adaptive control via impedance reference tracking for legged robots," in *Conference on Robot Learning (CoRL)*, Seoul, Korea, pp. 705-720, 2025, [Online], <https://proceedings.mlr.press/v305/xu25c.html>.
- [23] Y. Zhang et al., "FALCON: Learning force-adaptive humanoid loco-manipulation," arXiv preprint *arXiv:2505.06776*, 2025, DOI: 10.48550/arXiv.2505.06776.
- [24] A. Kumar, Z. Fu, D. Pathak, and J. Malik, "RMA: Rapid motor adaptation for legged robots," in *Robotics: Science and Systems (RSS)*, Jul., 2021, DOI: 10.15607/RSS.2021.XVII.011.
- [25] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal policy optimization algorithms," arXiv preprint *arXiv:1707.06347*, 2017, DOI: 10.48550/arXiv.1707.06347.
- [26] L. Pinto, M. Andrychowicz, P. Welinder, W. Zaremba, and P. Abbeel, "Asymmetric actor critic for image-based robot learning," in *Robotics: Science and Systems (RSS)*, Pittsburgh, PA, USA, Jun., 2018, DOI: 10.15607/RSS.2018.XIV.008.
- [27] Z. Wang, W. Wei, R. Yu, J. Wu, and Q. Zhu, "Toward understanding key estimation in learning robust humanoid locomotion," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Abu Dhabi, United Arab Emirates, pp. 11232-11239, 2024, DOI: 10.1109/IROS58592.2024.10801648.
- [28] F. Abdolhosseini, H. Y. Ling, Z. Xie, X. B. Peng, and M. van de Panne, "On learning symmetric locomotion," in *Motion, Interaction and Games (MIG)*, Newcastle upon Tyne, UK, Art. no. 19, Oct., 2019, DOI: 10.1145/3359566.3360070.
- [29] M. Mittal et al., "Orbit: A unified simulation framework for interactive robot learning environments," *IEEE Robotics and Automation Letters*, vol. 8, no. 6, pp. 3740-3747, Jun., 2023, DOI: 10.1109/LRA.2023.3270034.
- [30] J. Li, W. Monroe, and D. Jurafsky, "Understanding neural networks through representation erasure," arXiv preprint *arXiv:1612.08220*, 2016, DOI: 10.48550/arXiv.1612.08220.



조 용 현

2023 서강대학교 기계공학과(학사)
2025~현재 서울대학교 지능정보융합학과
석사, 박사 통합과정

관심분야: Reinforcement Learning, Whole body control



최 현 준

2025 서울대학교 건축학부(학사)
2025~현재 서울대학교 지능정보융합학과
석사, 박사 통합과정

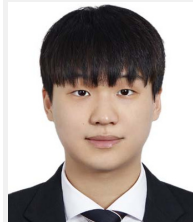
관심분야: Reinforcement learning, legged locomotion control



이 관 우

2022 인하대학교 기계공학과(학사)
2022~현재 서울대학교 지능정보융합학과
석사, 박사 통합과정

관심분야: Bipedal walking of humanoid robot



박 상 혁

2024 성균관대학교 기계공학부(학사)
2024~현재 서울대학교 지능정보융합학과
석사, 박사 통합과정

관심분야: Bipedal Locomotion, Environmental Perception



박 경 재

2021 서울대학교 기계항공공학부(학사)
2021~현재 서울대학교 지능정보융합학과
석사과정

관심분야: Bipedal walking of humanoid robot



박 재 흥

1995 서울대학교 항공우주공학과(학사)
1999 서울대학교 항공우주공학과(석사)
2006 Stanford University Aero/Astro(박사)
2009~현재 서울대학교 융합과학기술대학원
교수

관심분야: Robot-environment Interaction, Contact Force Control,
Multi Contact Control, Whole-body Dynamic Control,
Bio-mechanics



임 서 현

2023 연세대학교 기계공학과 학사과정

관심분야: Whole body control, bipedal locomotion



허 성 문

2010 한양대학교 전자정보시스템학과(공학사)
2012 과학기술연합대학원대학교 HCI 및 로
봇응용공학(공학석사)
2026 서울대학교 융합과학기술대학원(공학박사)
2022~2025 블루로빈 주식회사 이사
2025~현재 블루로빈 주식회사 공동대표

관심분야: Manipulation, Haptics, Rehabilitation Robot



신 재 용

2021 연세대학교 기계공학과(학사)
2021~현재 서울대학교 지능정보융합학과
석사, 박사 통합과정

관심분야: Humanoid, Bipedal Walking, Reinforcement Learning