

시뮬레이션 강화학습 실험 기록

# MicroDuck 보행 학습과 교정

2026.11.04

이한민 · hmlee@kimm.re.kr

# 1. 전체 진행 흐름

KIMM 한국기계연구원
KOREA INSTITUTE OF MACHINERY & MATERIALS

**재현 → 관찰 → 진단 → 교정의 근거를 보존**

## 01
### 재현
고정 버전
·GUI·headless 학습

## 02
### 관찰
숫자 평가와 보행 영상

## 03
### 진단
관절·접촉·정규화 기록

## 04
### 교정
조건별 비교와 재검증

# 2. 재현 범위



## 현재 대상은 평지 정책 · 실물·험지 성능은 별도

### 수행 범위

- 환경 설치·초기 GUI 확인
- 4,096개 환경 평지 학습
- 체크포인트 평가·보행 진단

### 남은 범위

- 자연스러운 보행 품질 확보
- 험지·외란 조건 평가
- ONNX 검증·실물은 별도

# 3. 실행 환경

KIMM 한국기계연구원 KOREA INSTITUTE OF MACHINERY & MATERIALS

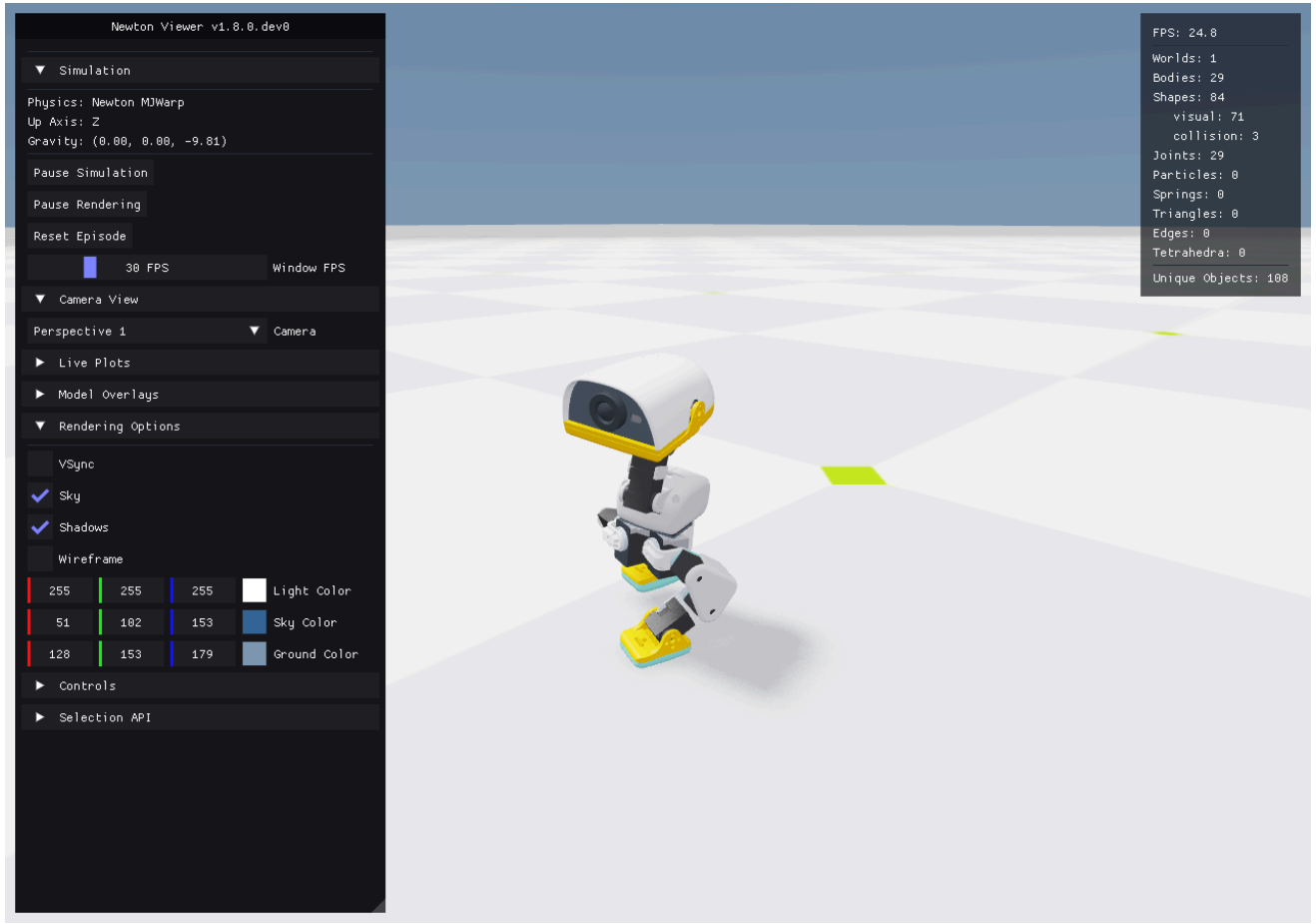
## 버전·물리·모터 모델을 고정하여 비교

| 항목 | 실제 사용 조건 |
| --- | --- |
| **GPU / 메모리** | RTX 5080 16GB / 시스템 RAM 64GB |
| **소프트웨어** | Isaac Lab 3.0.0 · PyTorch 2.11 cu130 |
| **물리 / 모터** | Newton MJWarp · BAM · backlash |
| **학습 가상환경** | .venv-preview — 이름과 달리 본 학습에도 사용 |

# 4. 화면과 물리 엔진

KIMM 한국기계연구원 KOREA INSTITUTE OF MACHINERY & MATERIALS

## Newton 뷰어로 학습과 같은 물리 조건 확인



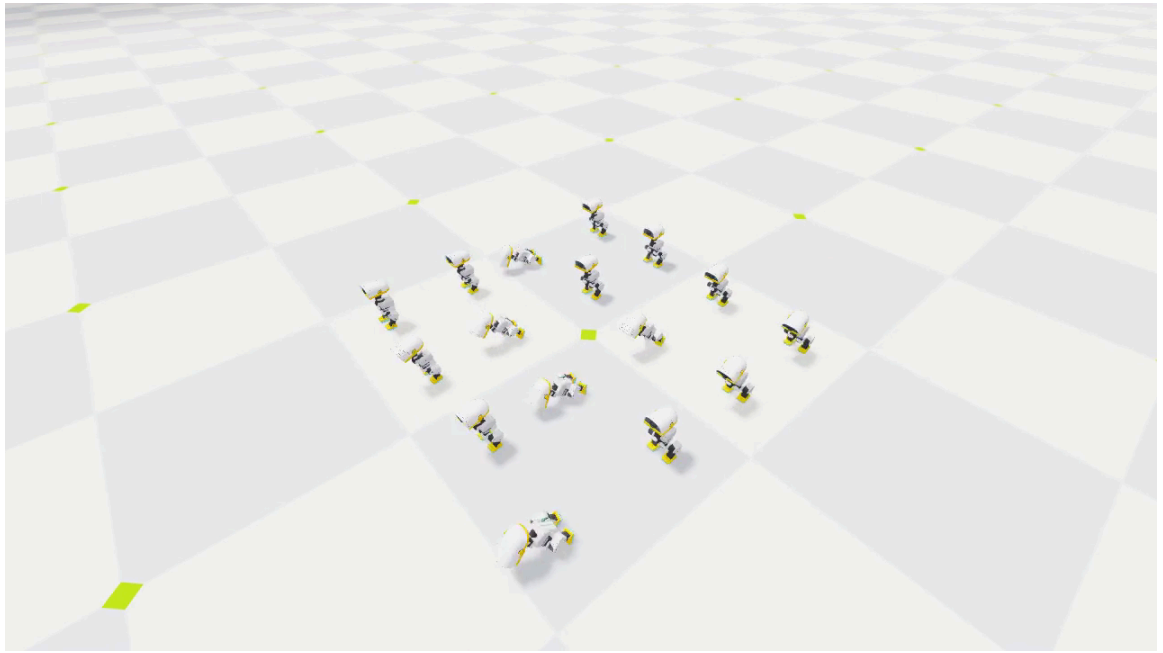

- Newton: 학습한 모델의 보행 확인
- Isaac Sim Kit: 시각화 대안
- ONNX: PyTorch 재생에 불필요

# 5. 병렬 환경의 학습 전후

KIMM 한국기계연구원
KOREA INSTITUTE OF MACHINERY & MATERIALS

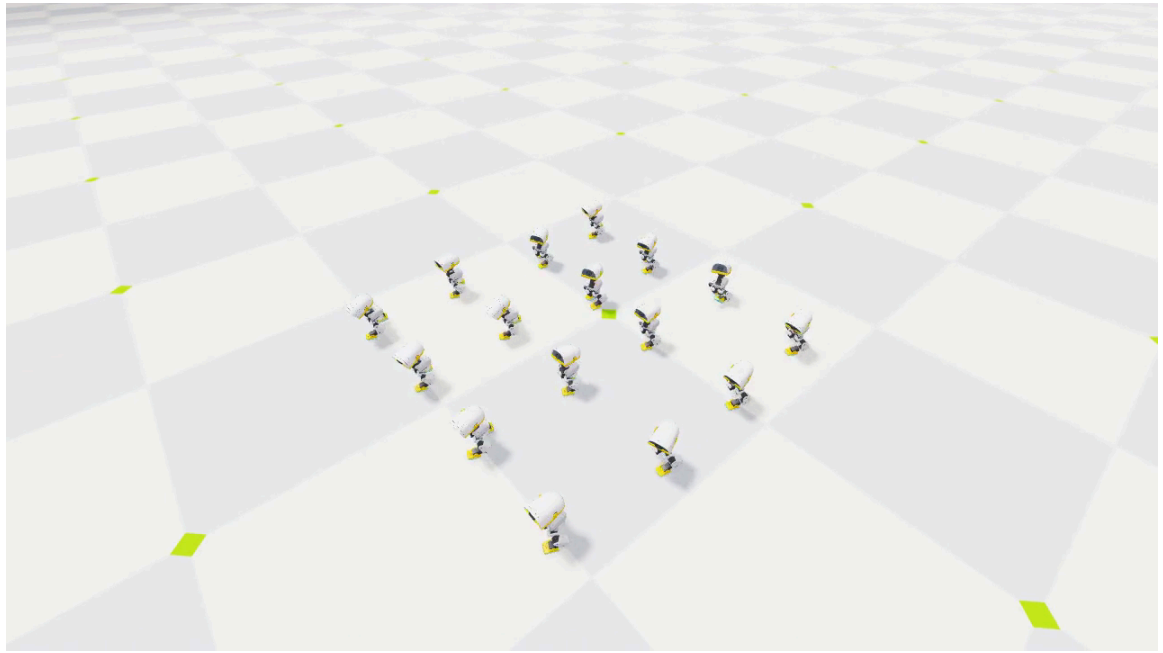
## 같은 정책을 공유하는 여러 환경에서 보행이 형성됨

### 초기 · 19회



넘어짐 발생 · 환경별 재시작

### 학습 후 · 45,000회



전진 행동 형성 · 비대칭 잔존

저장 정책 비교 · 각 16대 · 같은 조건에서 전진 명령

# 6. 정책의 입출력

KIMM 한국기계연구원
KOREA INSTITUTE OF MACHINERY & MATERIALS

## 목표 명령과 관절 행동 출력을 구별

**01**

### 입력 61개

센서·직전 행동·목표 명령

**02**

### 정책 MLP

512 → 256 → 128

**03**

### 출력 14개

관절 기준 자세 대비 목표 변화

# 7. 입력과 관절의 구성

KIMM 한국기계연구원 KOREA INSTITUTE OF MACHINERY & MATERIALS

## 상태·이력48개 + 명령13개 → 관절14개 제어

| 구분 | 구성 |
|---|---|
| **센서·이력 48개** | IMU6 + 관절각14 + 관절속도14 + 직전행동14 |
| **명령 13개** | 이동·회전3 + 머리4 + 몸통6 |
| **양쪽 다리 각5개** | hip_yaw · hip_roll · hip_pitch · knee · ankle |
| **목·머리4개** | neck_pitch · head_pitch · head_yaw · head_roll |

# 8. 머리·몸통 명령의 역할



## 머리는 추종 학습 · 몸통은 입력만 유지

### 머리 4개

- 목 굽힘·머리 숙임·좌우 회전·기울임
- 기준자세 대비 각도 변화 [rad]
- 추종 + 지속 편향 보상 사용
- 최대 변화: ±1.1 / ±1.1 / ±1.4 / ±0.31 rad

### 몸통 6개

- 위치 x/y/z + roll/pitch/yaw
- 위치 ±5mm · 회전 ±0.05rad
- 입력은 존재 · 자세 추종 보상 꺼짐

# 9. PPO 데이터 흐름

KIMM 한국기계연구원
KOREA INSTITUTE OF MACHINERY & MATERIALS

**98,304개 전이를 5회 사용 · 총 20번 갱신**

01

## 경험 수집

4,096 환경 × 24 제어 스텝

02

## 모델 갱신

4 미니배치 × 5 epochs

03

## 다음 수집

갱신된 정책으로 다시 행동

# 10. PPO 갱신의 제한

KIMM 한국기계연구원 KOREA INSTITUTE OF MACHINERY & MATERIALS

**확률비 제한은 관절 목표의 물리 한계를 대신하지 않음**

| 핵심 계산 | 의미 |
|---|---|
| $\rho = \pi_{새정책}(a\|s) / \pi_{수집정책}(a\|s)$ | 수집 때보다 같은 행동을 얼마나 더 선택하는가 |
| $L = E[\min(\rho\hat{A}, clip(\rho,0.8,1.2)\hat{A})]$ | 이득 $\hat{A}$를 반영하되 과도한 정책 변화를 억제 |
| 배치 반복 사용: 원본5회 / 교정3회 | 수집 비용을 줄이면서 여러 번 경사 갱신 |
| 별도 필요한 제약 | 모터 목표 한계·행동 이력 범위·유한값 검사 |

# 11. 제어 주기



## 물리 4회 계산마다 새 정책 행동 1회

### 물리 계산

**200 Hz**

5ms 간격으로 접촉·관절 상태 계산

### 정책 제어

**50 Hz**

20ms마다 새 목표 · decimation=4

# 12. 보상 설계

KIMM 한국기계연구원
KOREA INSTITUTE OF MACHINERY & MATERIALS

## 실제 로봇의 움직임에 목표와 감점을 연결

### 머리 · 명령 각도 추종

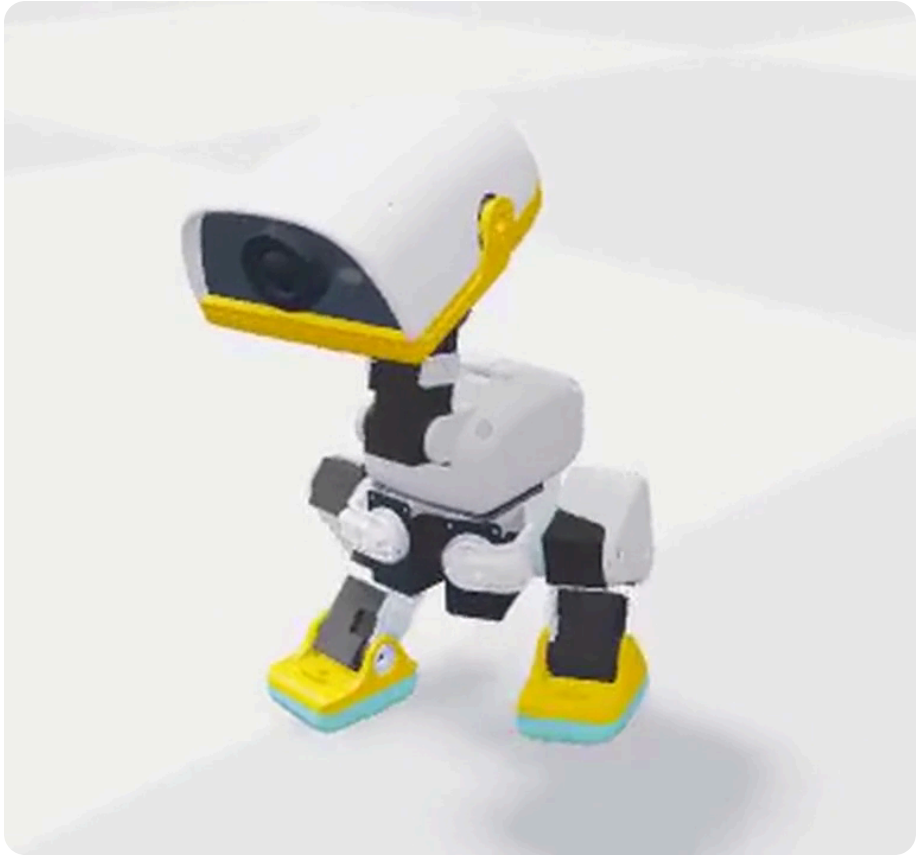
목·머리 4축 추종 +2

지속되는 각도 편향 −3

### 이동 · 속도 명령 추종

전진·회전 추종 +4 / +4

목표 속도와의 오차에 보상



### 몸통·관절 · 안정성

직립 +2 · 몸통 흔들림 −0.05

행동 급변·관절 한계 감점

### 발 · 지지와 스윙

공중시간 +3 · 미끄러짐 −0.1

발 높이·자기 충돌 감점

**15개 항목의 가중합 × 제어 주기 0.02초**

# 13. 랜덤화 적용 시점

KIMM 한국기계연구원 KOREA INSTITUTE OF MACHINERY & MATERIALS

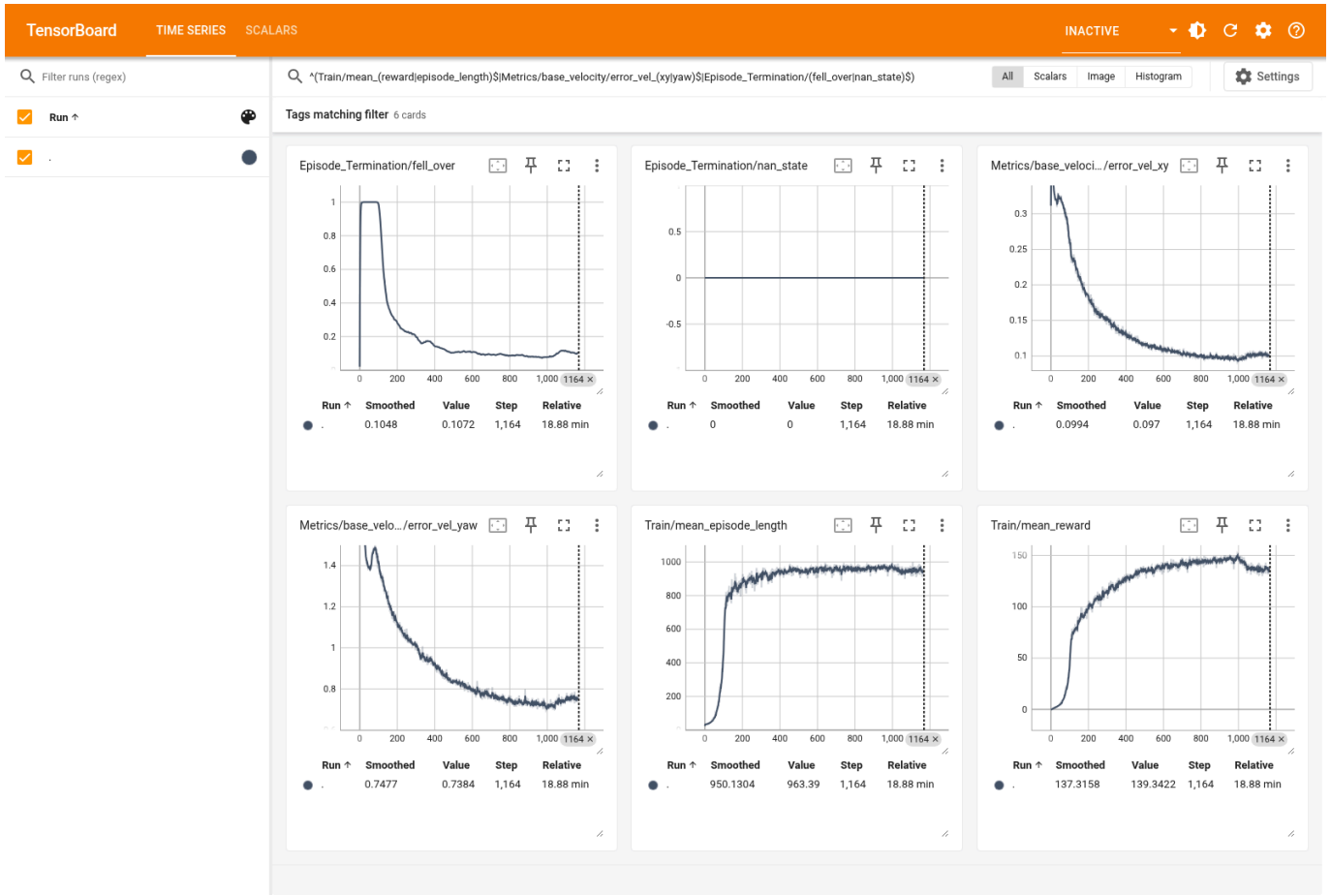
## 모든 항목이 에피소드마다 바뀌지는 않음

| 시점 | 주요 항목 |
|---|---|
| **환경 생성** | 마찰·질량·엔코더 편차·IMU 장착·공급전압 |
| **에피소드 시작** | 초기 상태·무게중심·관절 마찰·추가 관성 |
| **도중 3~6초** | x/y 속도에 ±0.3m/s 범위 변화 추가 |
| **관측 시** | 센서 잡음·지연 |

# 14. 학습 그래프의 의미

KIMM 한국기계연구원 KOREA INSTITUTE OF MACHINERY & MATERIALS

## 보상·오차·종료 사유를 함께 확인

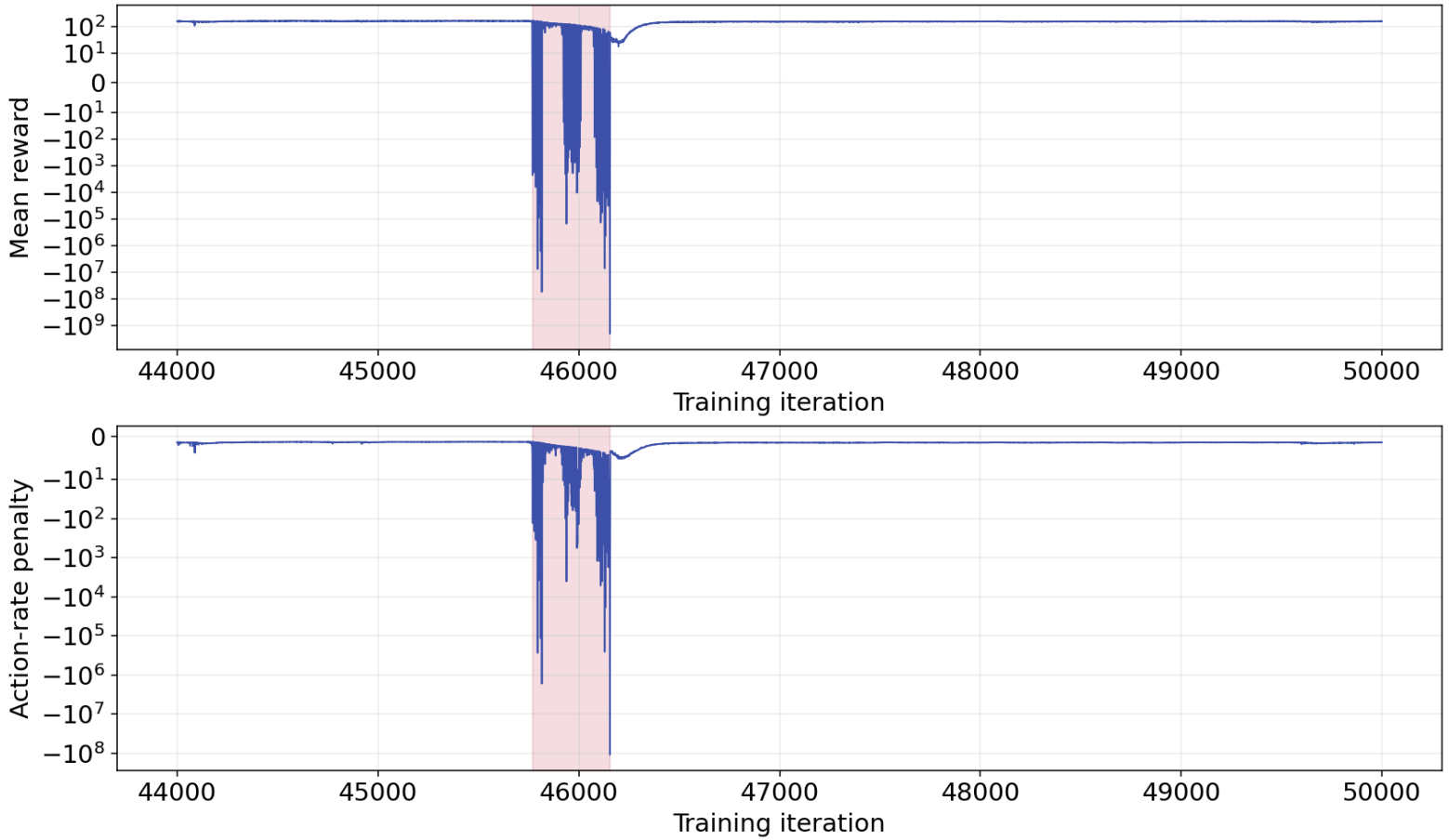



- 보상: 높을수록 목표에 유리
- 추종 오차: 낮을수록 좋음
- 생존과 자연스러움은 별개

# 15. 46,000회 부근 급락

KIMM 한국기계연구원 KOREA INSTITUTE OF MACHINERY & MATERIALS

## 큰 행동 변화 감점과 가치 손실 이상치 확인

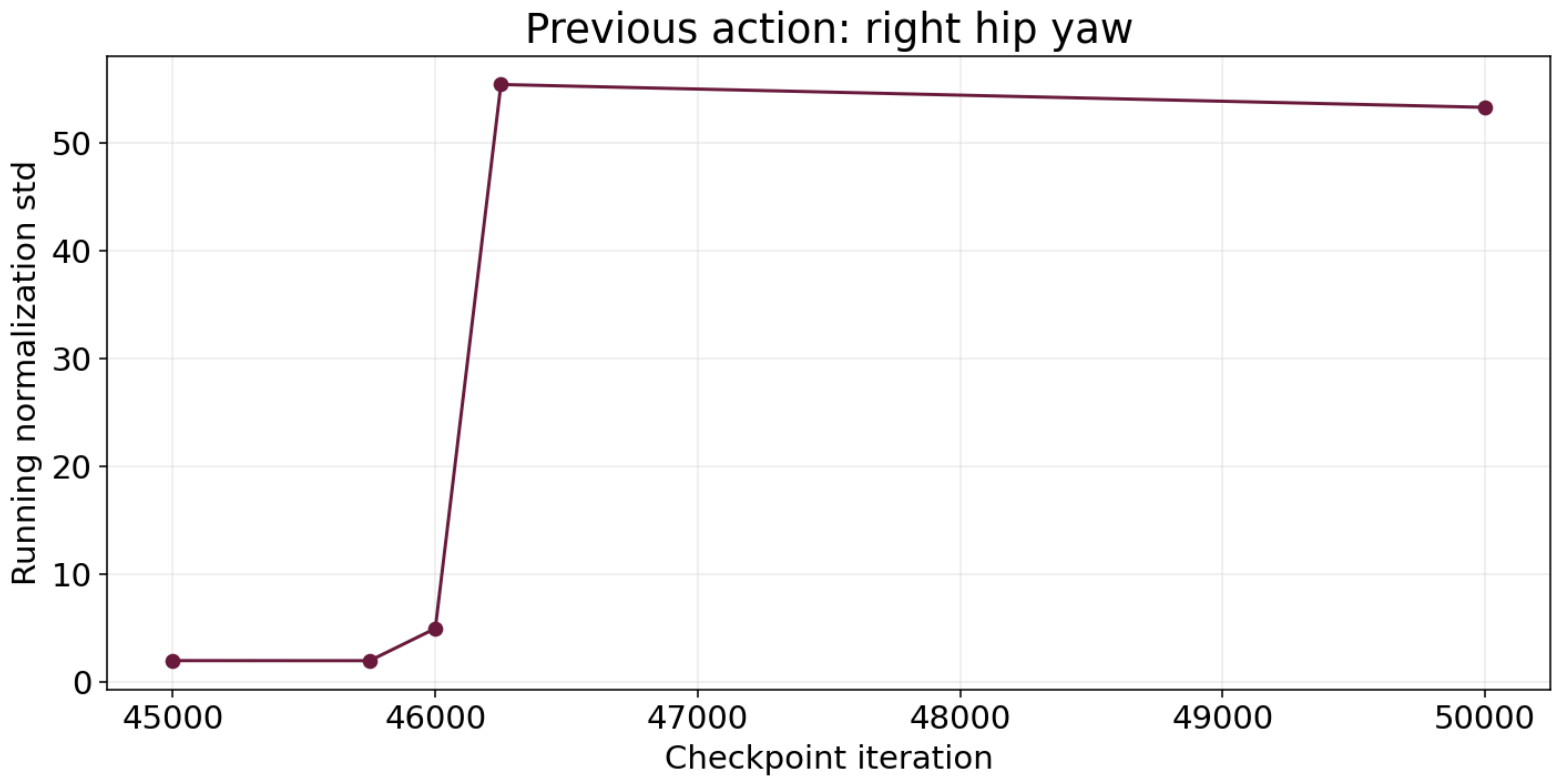



- 약45,769~46,156회 이상 구간
- 마지막 커리큘럼 변경은7,000회
- 최초 유발 원인은 미확정

# 16. 이상치의 후유증

KIMM 한국기계연구원 KOREA INSTITUTE OF MACHINERY & MATERIALS

## 입력 정규화 통계에도 영향이 남음

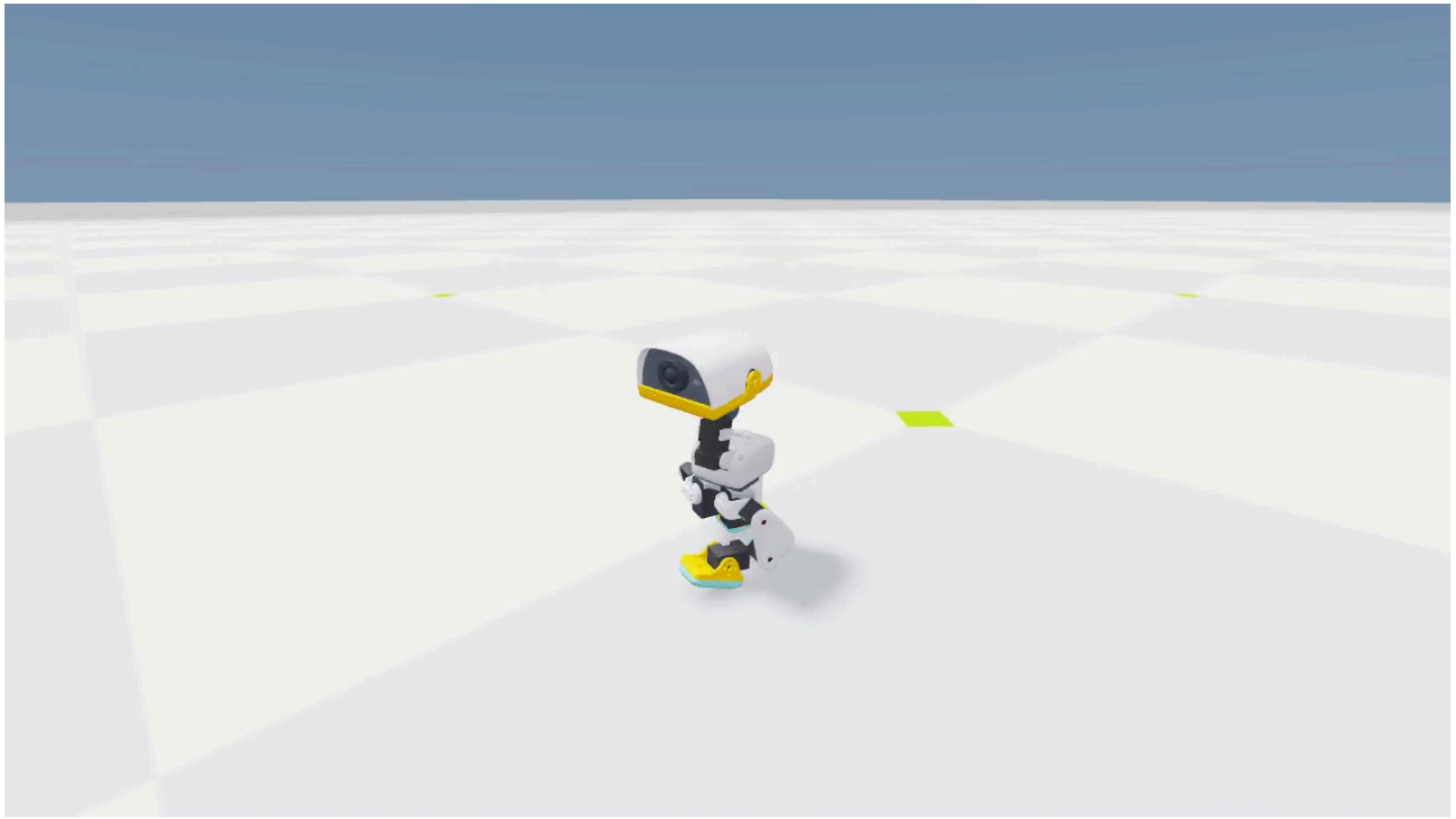



- 직전 오른쪽 hip _ yaw 행동
- 표준편차 약2 → 55 → 53
- 최종 모델이 항상 최선은 아님

# 17. 사용자가 발견한 보행 문제

KIMM 한국기계연구원
KOREA INSTITUTE OF MACHINERY & MATERIALS

## 넘어지지 않아도 절뚝거리는 보행은 품질 미달

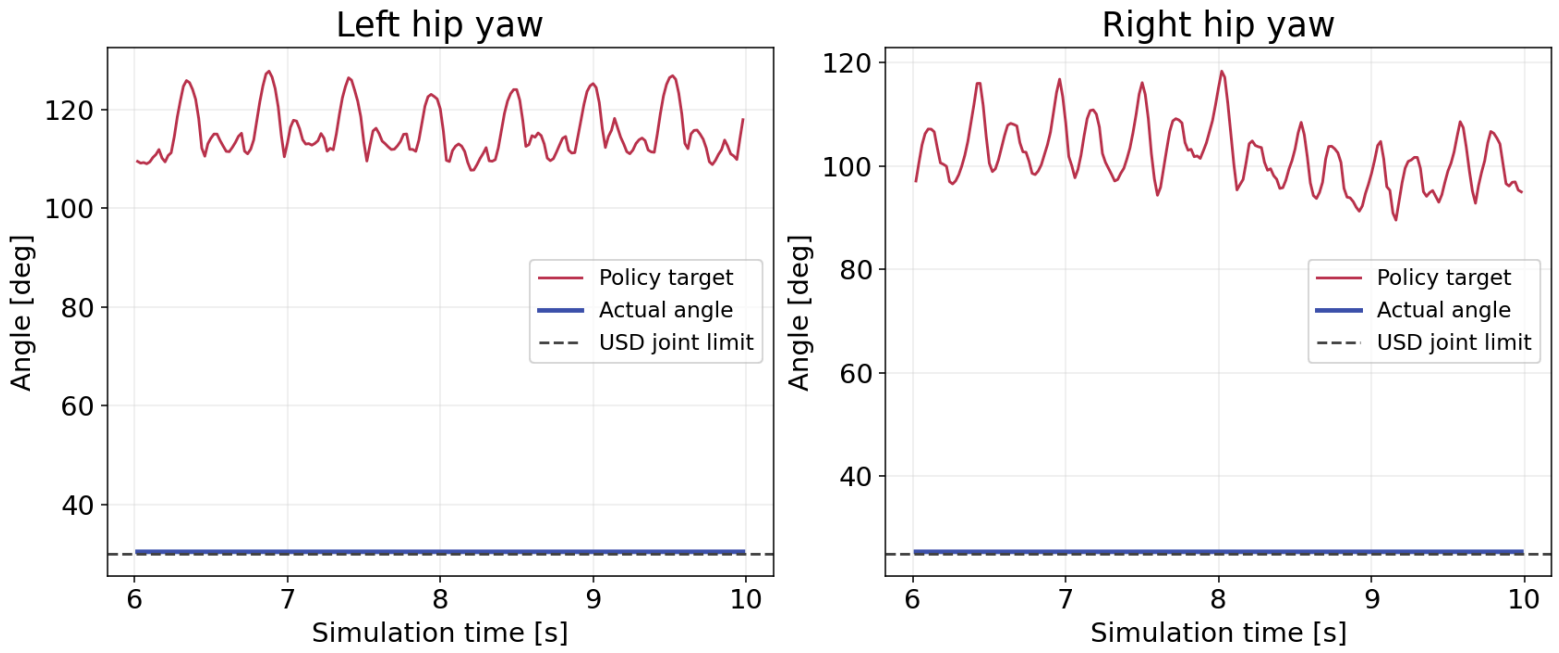


- 45,000회 체크포인트
- 정지 → 전진 → 회전 → 정지
- 사용자 관찰: 한쪽 다리 고정처럼 보임

# 18. 관절 목표와 실제 각도

KIMM 한국기계연구원 KOREA INSTITUTE OF MACHINERY & MATERIALS

## 양쪽 hip _ yaw가 끝단에 밀린 상태 확인

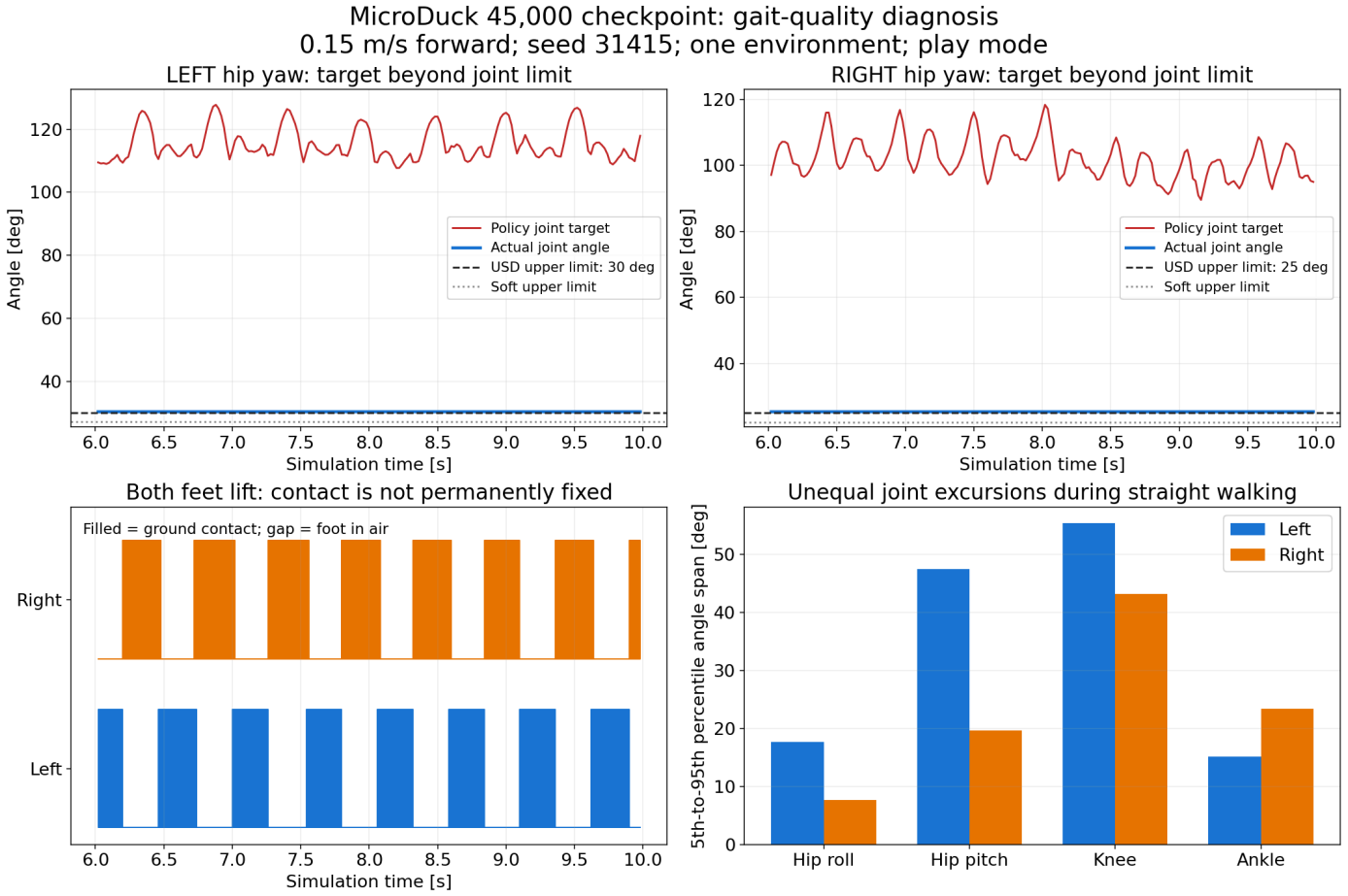



- 허용 상한: 좌30° / 우25°
- 목표 평균: 약115° / 101°
- 실제 각도는 끝단 부근 고정

# 19. 발 접촉과 좌우 사용

KIMM 한국기계연구원 KOREA INSTITUTE OF MACHINERY & MATERIALS

## 양발은 번갈아 들지만 관절 가동 범위가 비대칭




- 발 접촉 비율 약49~53%
- hip_pitch 가동 범위 약2.4배 차이
- 최종 모델에서도 같은 경향

# 20. 평가 기준의 보완

KIMM 한국기계연구원 KOREA INSTITUTE OF MACHINERY & MATERIALS

## 속도·생존율에 관절·보행 품질 기준 추가

| 기존 기준 | 추가 기준 |
| --- | --- |
| **생존율 ≥90%** | 목표 각도의 soft-limit 초과0% |
| **XY RMSE ≤0.15m/s** | hip _yaw 전류포화 시간 비율 <1% |
| **yaw RMSE ≤0.35rad/s** | 직진 hip _pitch 좌우 가동범위 비 ≤1.5 |
| **정지·전진·회전** | 여러 seed + 영상으로 절뚝거림 확인 |

# 21. 정지 정책의 오판 방지



**오차가 작아도 실제로 전진하고 양발을 써야 함**

## 속도 기준 보완

- 명령: 0.15m/s 전진
- 평균 전진속도 ≥ 명령의80%
- 완전 정지를 통과시키지 않음

## 관절 기준 보완

- 양쪽 hip_pitch 범위 각각≥0.15rad
- 그다음 좌우 범위비≤1.5 확인
- 둘 다 고정된 상태도 탈락

# 22. 교정 실험 순서

KIMM 한국기계연구원 KOREA INSTITUTE OF MACHINERY & MATERIALS

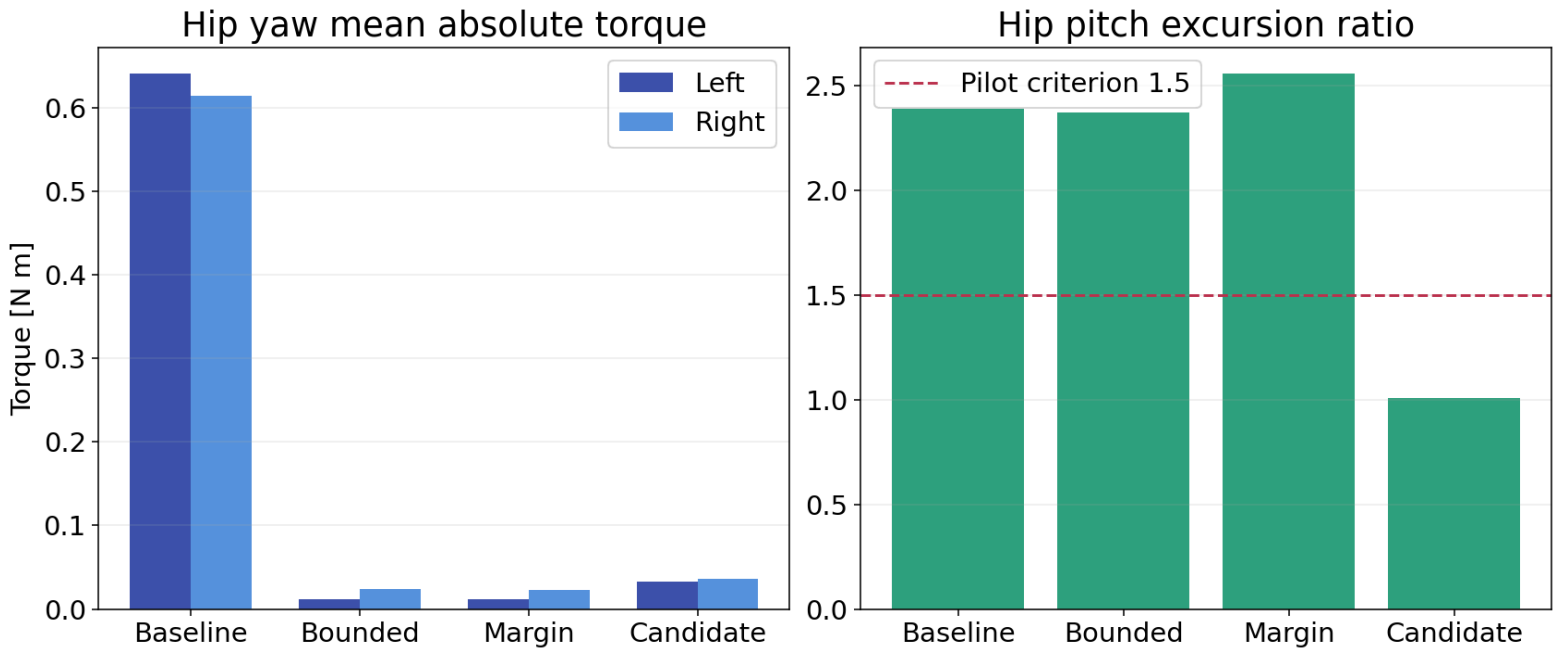
## 조건을 분리하고 부분 개선과 실패도 기록

| 실험 | 변경 | 판단 |
|---|---|---|
| E0 | 원본 정책·영상·진단 보존 | 기준선 |
| E1 | 모터 목표를soft-limit안으로 제한 | 토크 감소·비대칭 잔존 |
| E2 | 목표 범위에0.05rad 여유 추가 | 한계 초과 제거·비대칭 잔존 |
| E3 | 범위 제한+자세 보상으로 교정 학습 | 출력 guard에 의해 중단 |
| E3b | 환경 전달·이력 행동 ±3 제한 | 별도 교정 학습·재평가 |

# 23. 목표 제한의 효과

KIMM 한국기계연구원 KOREA INSTITUTE OF MACHINERY & MATERIALS

## 과도한 모터 토크는 줄었지만 비대칭은 남음




- E1:
  평균토크0.64→0.01~0.02N·m
- E2: 실제soft-limit초과0%
- 단순 제한만으로 품질 완료 아님

# 24. 교정 학습의 조건

KIMM 한국기계연구원 KOREA INSTITUTE OF MACHINERY & MATERIALS

## 원본 복원 후 작은 학습률로 짧게 비교

| 항목 | 설정 |
| --- | --- |
| 초기 상태 | model45000의정책·가치·정규화·optimizer복원 |
| 학습량 | 1,024환경 × 24step × 추가500회 |
| 최적화 | fixed LR 1e-4 · 4미니배치 · 3epochs |
| 보상 추가 | 목표 초과량 -0.5 · hip_yaw 중심편차 -0.5 |
| 환경 | 성숙 curriculum고정 · 원본DR/잡음/밀기유지 |

# 25. 실패를 조기에 포착



## 모터 목표 제한만으로 원시 행동 되먹임은 막지 못함

### E3에서 확인

- 약12회 갱신 후 guard중단
- 직전 행동 최대28.54
- 다음 정책 출력 최대55.71

### E3b 보완

- 환경 전달 행동 ±3 제한
- 관절별 최종 목표 제한 유지
- 같은감시기준·부모모델로재시도

# 26. 행동에서 모터 목표까지

KIMM 한국기계연구원 KOREA INSTITUTE OF MACHINERY & MATERIALS

## 행동 이력과 최종 관절 목표에 각각 범위 제한

| 단계 | 교정 후 계산 |
| --- | --- |
| **환경 전달 행동** | a = clip(정책 출력, −3, +3) |
| **요청 목표 [rad]** | q요청 = q기준 + a − 엔코더 편차 |
| **최종 모터 목표 [rad]** | q목표 = clip(q요청, soft하한+0.05, soft상한−0.05) |
| **실행 검증** | 3환경 × 14관절 · 극단 입력 ±50 · 계산 오차0 |

# 27. 좌우 대칭 증강



## 거울에 비친 상태·행동도 경험으로 학습

### 추가한 학습 경험

- 왼다리 ↔ 오른다리 교환
- 회전축·측면 성분의 부호 반전
- 반전 두 번이면 원본 복원

### E4 비교 조건

- E3b 500회 모델에서 시작
- 나머지 보상·제한·DR 유지
- 대칭 증강만 추가 · 1,000회

# 28. 대칭 증강의 결과

KIMM 한국기계연구원 KOREA INSTITUTE OF MACHINERY & MATERIALS

## 가동범위 차이는 감소했으나 기준에 미달

| 항목 | E3b → E4 결과 |
| --- | --- |
| 추가 학습 | 대칭 증강 1,000회 |
| 좌우 hip_pitch 범위비 | 2.08 → 1.92 · 목표 ≤1.5 |
| 회전 추종 RMSE | 0.296 → 0.346rad/s · 상한0.35 |
| 판정 | 부분 개선 · 직접 대칭 오차 학습으로 다음 실험 |

# 29. 대칭 대응 오차의 학습



## 반전 상태의 예측과 원래 예측의 반전을 맞춤

### E4: 경험을 추가

- 원래 경험 + 좌우 반전 경험
- PPO가 두 경험에서 학습
- 정책 대응 관계는 간접 유도

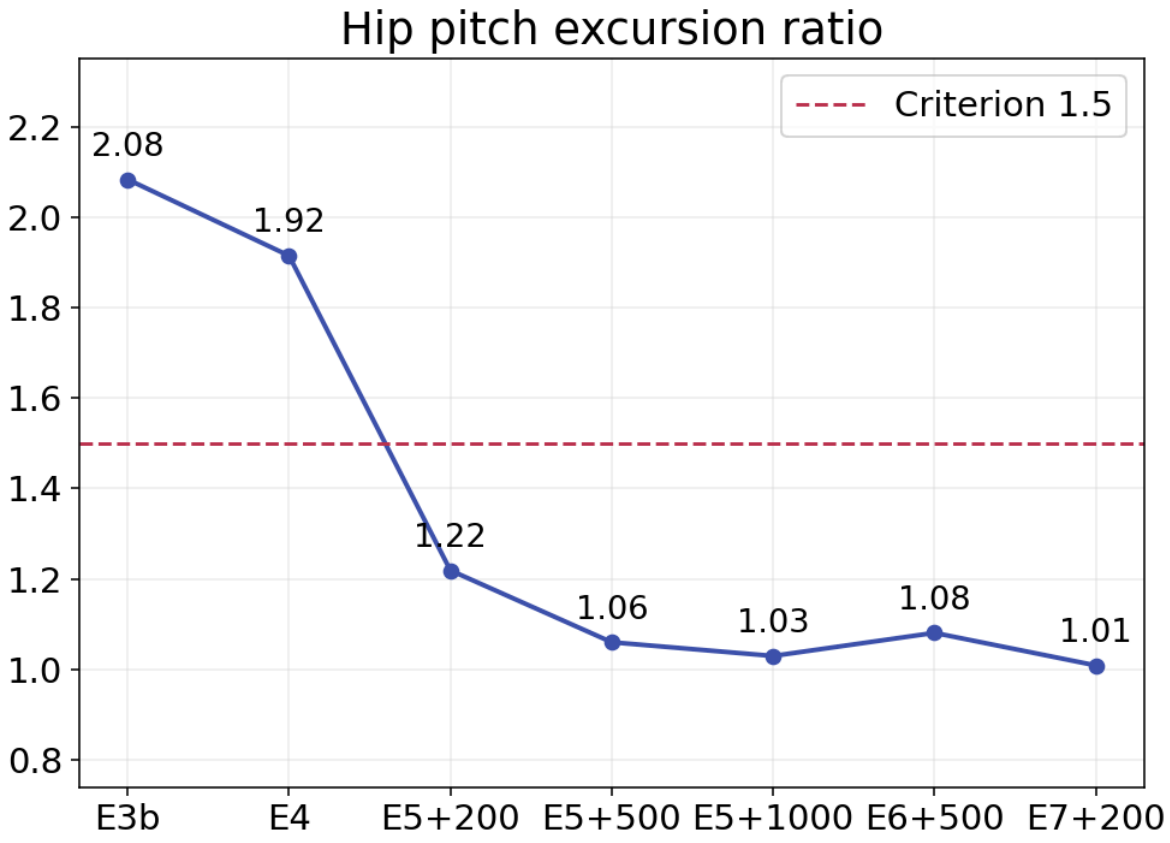
### E5: 오차도 직접 줄임

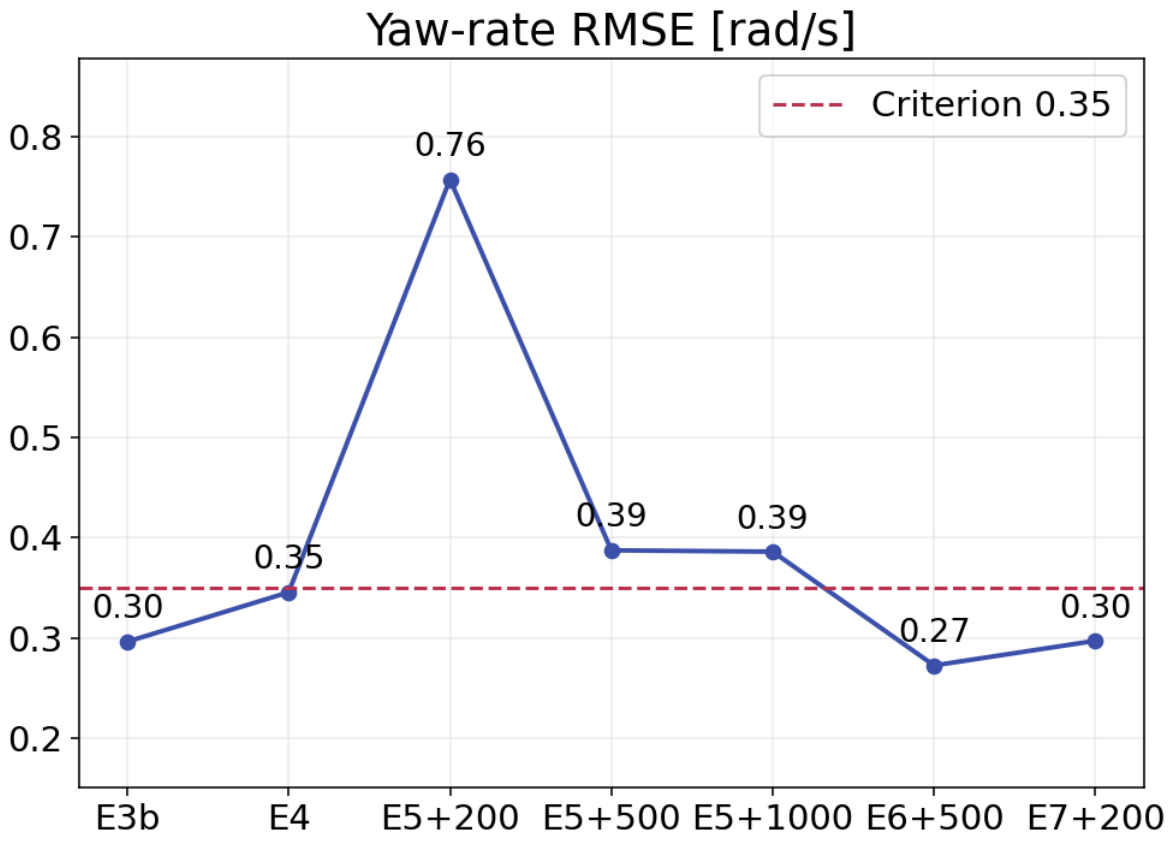
- π(Ms) 와 Mπ(s) 비교
- 제곱오차를 PPO 손실에 추가
- 계수1.0 · 나머지 조건 유지

# 30. 대칭과 안정성의 교환

KIMM 한국기계연구원 KOREA INSTITUTE OF MACHINERY & MATERIALS

## 좌우 차이 감소와 회전 흔들림을 함께 평가



# 31. 대칭 교정의 검증 결과

KIMM 한국기계연구원 KOREA INSTITUTE OF MACHINERY & MATERIALS

## 좌우 비대칭은 해소 · 회전 추종은 추가 보완

| 3개 seed × 32환경 | E5 최종 후보 |
|---|---|
| 정지·직진·회전 생존율 | 각 조건100% · 20초 평가 |
| 직진 hip_pitch 범위비 | 1.021~1.030 · 기준≤1.5 |
| 직진 yaw RMSE | 0.372~0.386rad/s · 기준0.35 초과 |
| 제자리 회전 평균속도 | 0.319~0.361rad/s · 목표0.5 / 기준≥0.4 |

# 32. 회전 추종의 보완



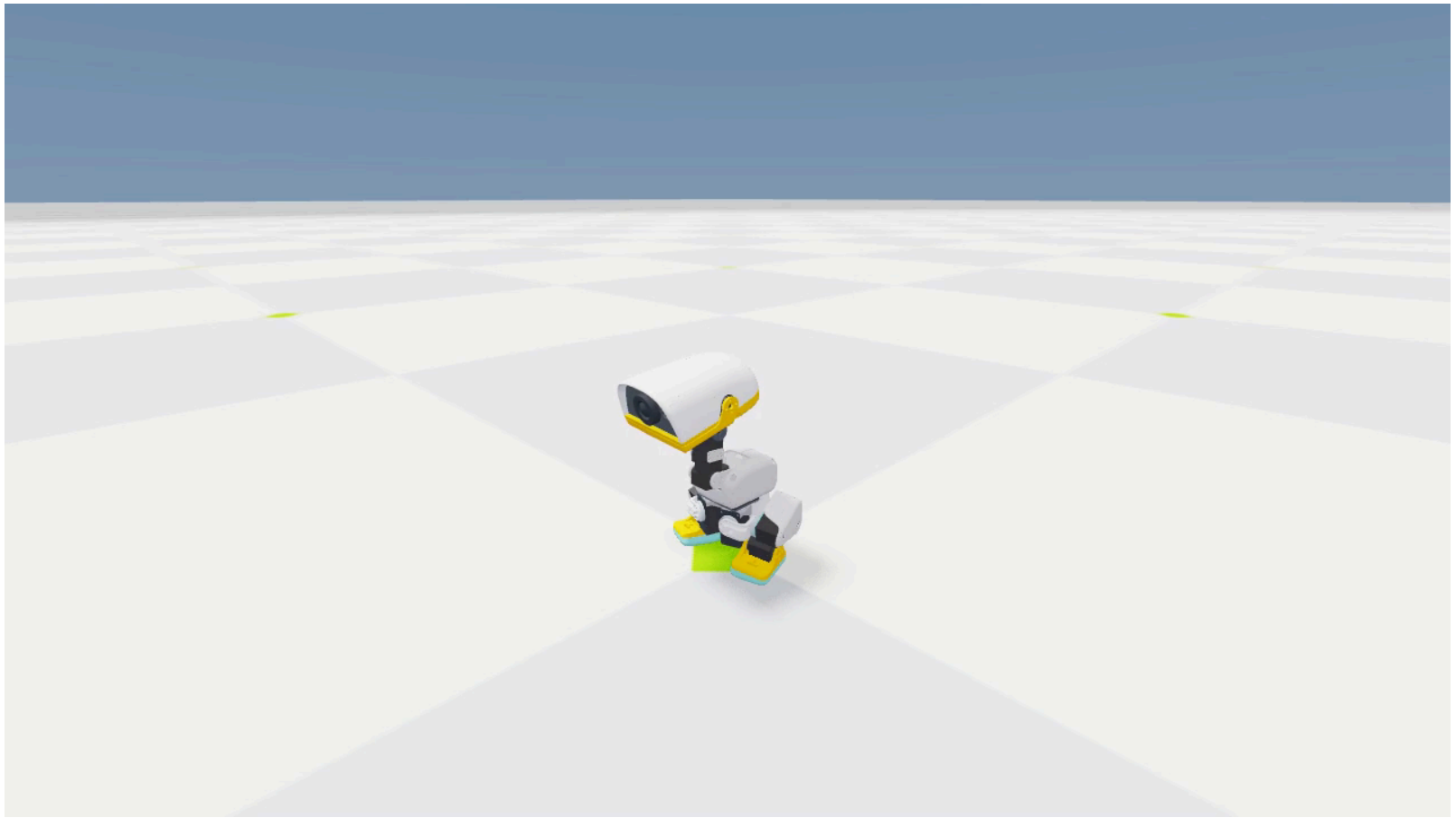
## 큰 회전 오차에도 감점이 증가하도록 제곱항 추가

| 항목 | E6 변경 |
| --- | --- |
| 기존 회전 보상 | $4\exp(-e^2/0.35^2)$ · 큰 오차에서0에 수렴 |
| 추가 감점 | $-2e^2 \cdot e$ = 실제 세계z각속도 − 목표 회전속도 |
| 직진 / 제자리 회전 | 목표0 / 0.5rad/s를 같은 오차식으로 처리 |
| 비교 조건 | E5 model47500 +500회 · 그 외 설정 유지 |

# 33. 고정 명령 평가의 사각

KIMM 한국기계연구원 KOREA INSTITUTE OF MACHINERY & MATERIALS

## 9조건 통과 모델도 정지 후 출발에 실패



- 4초 정지 후 전진 명령
- 목표0.15m/s → 실제≈0
- 명령 입력 정상 전달 확인

# 34. 명령 전환 경험의 보완



## 보상식을 유지하고 출발 상황의 학습 빈도를 높임

### 집중 환경 512개

- 머리·몸통 명령0
- 정지35% / 전진0.15m/s40%
- 회전±0.5rad/s25%

### 기존 환경 512개

- 원본 속도·머리 명령 분포
- 보상18개·물리·랜덤화 유지
- 최대1,000회 계획 · 검증 후 선택

# 35. 부분 개선 영상

KIMM 한국기계연구원
KOREA INSTITUTE OF MACHINERY & MATERIALS

## E3b는 한계·포화를 해소했지만 비대칭 기준 미달



- model45500 · 추가500회
- 전진 가동범위 비 2.08
- 비교 목표 ≤1.5에 미달

# 36. 실험 진행 기록

KIMM 한국기계연구원 KOREA INSTITUTE OF MACHINERY & MATERIALS

## 실패·부분 개선·검증 통과를 한 장에 정리

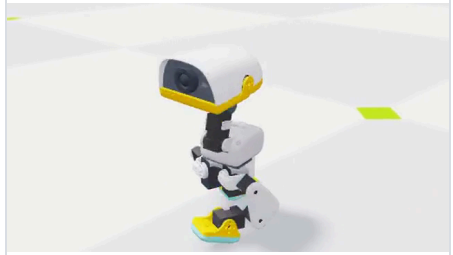
| 실험 | 주요 변경 | 추가 학습 | 결과 |
|---|---|---|---|
| **E3 → E3b** | 목표 제한 + 원시 행동 제한 | 중단 → 500회 | 한계·포화 해소 |
| **E4** | 좌우 대칭 경험 증강 | 1,000회 | 가동범위 비 1.92 |
| **E5** | 대칭 보조손실 추가 | 1,000회 | 가동범위 비 1.03 |
| **E6** | 회전 오차 제곱 감점 | 500회 | 회전 개선 · 출발 실패 |
| **E7** | 정지·출발·회전 경험 보강 | 200회 모델 채택 | 고정·전환 평가 통과 |

# 37. 보행 교정의 전체 히스토리

KIMM 한국기계연구원 KOREA INSTITUTE OF MACHINERY & MATERIALS

**한계 의존 → 대칭 개선 → 출발 실패 → 전환 보완**

## 01 원본



**45,000회 학습**

끝단 의존·절뚝거림

**가동비 2.39**

## 02 목표 제한

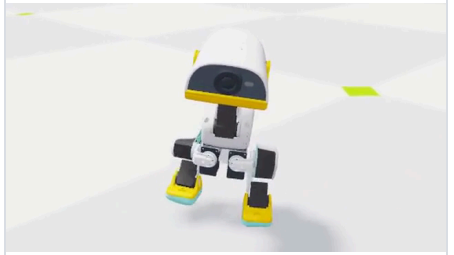


**목표·행동 범위 제한**

포화 해소·비대칭 잔존

**가동비 2.08**

## 03 대칭 학습



**대칭 경험 + 보조손실**

좌우 가동 차이 감소

**가동비 ≈1.03**

## 04 회전 보완



**회전 오차 감점 추가**

회전 개선·출발 실패

**정지 뒤 멈춤**

## 05 전환 학습



**정지·출발 경험 보강**

출발·회전·재출발 통과

**가동비 ≈1.01**

# 38. 교정 후보의 측정 결과



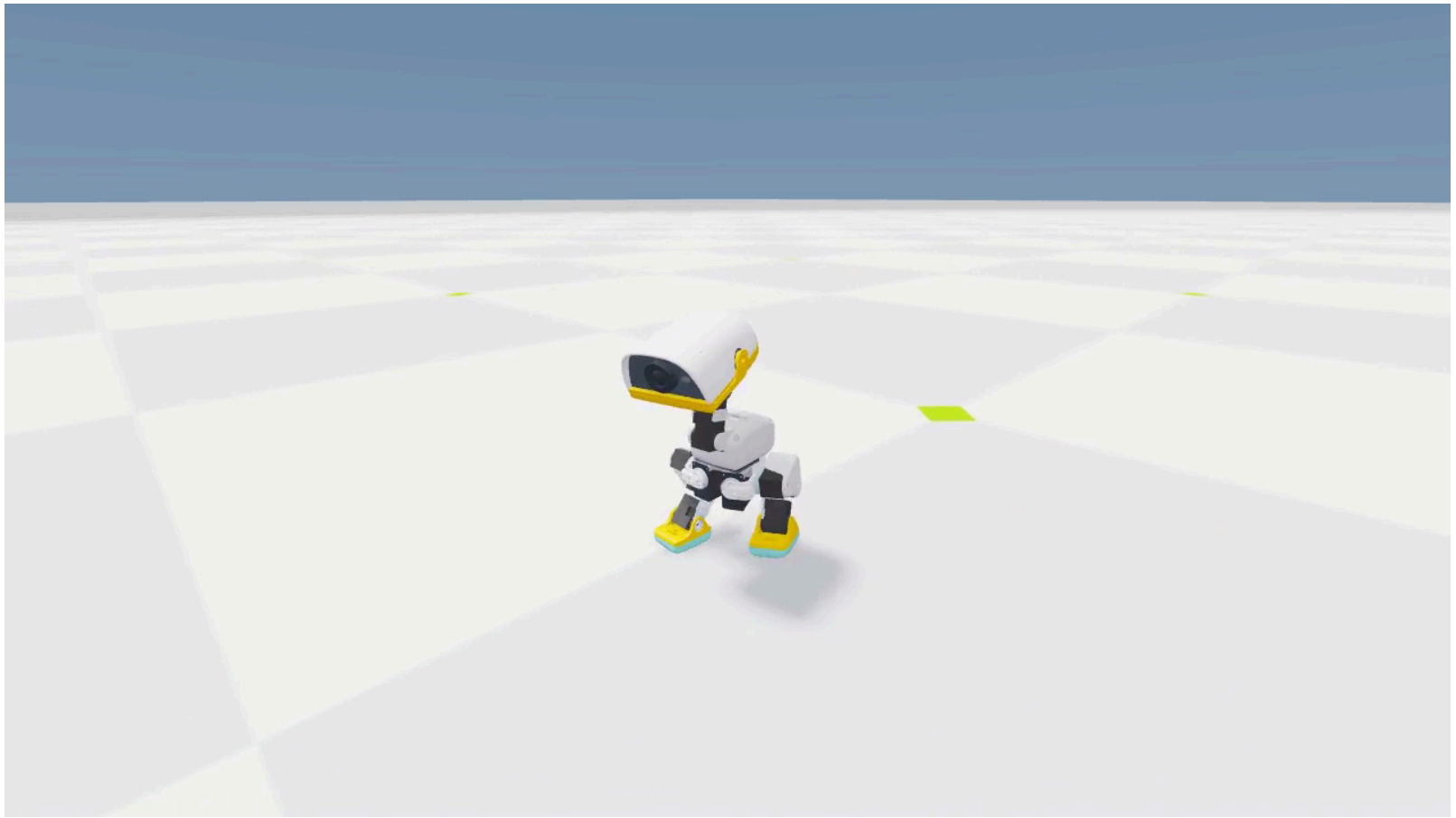
## 개선된 항목과 남은 한계를 함께 판단

| 항목 | 측정값 |
| --- | --- |
| 첫에피소드생존율 | 100.0% |
| XY / yaw RMSE | 0.0789m/s / 0.2971rad/s |
| hip_pitch좌우가동범위비 | 1.01 |
| 평가 조건 | 32환경 · seed31415 · bounded-posture-yaw |

# 39. 교정 후보 영상

KIMM 한국기계연구원
KOREA INSTITUTE OF MACHINERY & MATERIALS

## 동일한 명령 순서로 원본과 비교

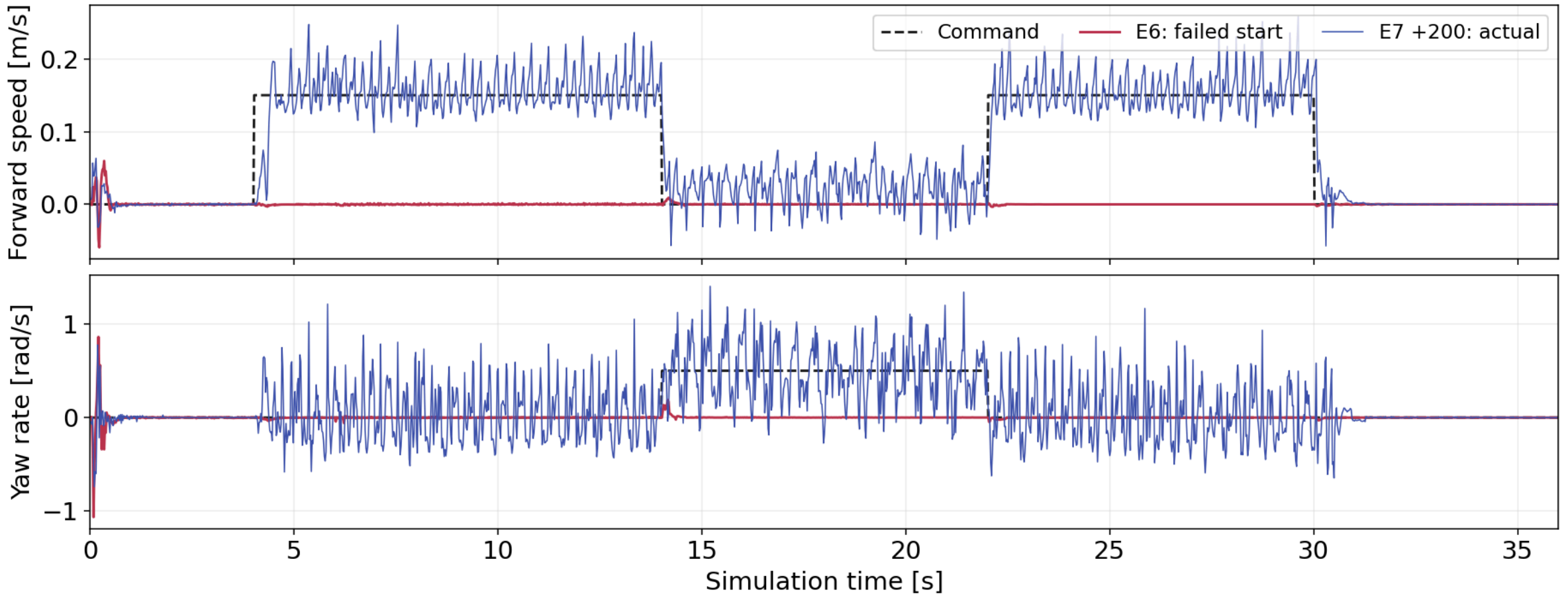


- 같은 정지·전진·회전 순서
- 한계 의존·좌우 자세 관찰
- 영상과 수치 평가를 함께 판정

# 40. 명령 전환의 교정 결과

KIMM 한국기계연구원 KOREA INSTITUTE OF MACHINERY & MATERIALS

## 정지 뒤 출발·회전·재출발·정지까지 확인



# 41. 독립 조건 검증

KIMM 한국기계연구원 KOREA INSTITUTE OF MACHINERY & MATERIALS

## 3개 seed × 3개 명령 · 각32환경에서 평가

| 명령 | 최저 생존율 | yaw RMSE | 실제 속도 | 통과 |
|---|---|---|---|---|
| **정지** | 100% | 0.007~0.008 | -0.000~0.000 | 3/3 |
| **전진0.15m/s** | 100% | 0.294~0.297 | 0.151~0.153 | 3/3 |
| **회전0.5rad/s** | 100% | 0.267~0.274 | 0.497~0.512 | 3/3 |

# 42. 여러 환경의 전환 검증

KIMM 한국기계연구원 KOREA INSTITUTE OF MACHINERY & MATERIALS

## 평균 추종과 환경별 출발 성공을 함께 확인

| seed | 1차 / 2차 전진 | 회전 속도 | 이동 기준 충족 | 판정 |
|---|---|---|---|---|
| **31415** | 0.153 / 0.153 | 0.504 | 96.9% | 통과 |
| **27182** | 0.151 / 0.151 | 0.498 | 93.8% | 통과 |
| **16180** | 0.152 / 0.152 | 0.510 | 100.0% | 통과 |

# 43. 이후 진행 분기

KIMM 한국기계연구원 KOREA INSTITUTE OF MACHINERY & MATERIALS

## 보행 품질을 확보한 뒤 험지·강인성으로 확장

**01**

### 평지 품질

관절끝단·좌우가동·정지전환

**02**

### 일반화

미사용seed·속도·회전·외란

**03**

### 확장

험지·ONNX비교·실물별도

# 44. 다른 로봇에 적용할 원칙

KIMM 한국기계연구원 KOREA INSTITUTE OF MACHINERY & MATERIALS

## 보상·실행 제한·경험 분포·평가를 분리 설계

| 설계 영역 | 이번 실험에서 얻은 근거 |
| --- | --- |
| **목표 보상** | 속도 추종 + 자세·접촉 안정성부터 구성 |
| **실행 제한** | 관절 목표 한계는 보상 감점과 별도로 적용 |
| **실패 대응** | 비대칭→대칭 학습 / 회전 오차→오차 감점 |
| **경험·검증** | 출발 실패→전환 경험 / 영상→누락된 평가 발견 |

# 45. 재현·인계 자료

KIMM 한국기계연구원 KOREA INSTITUTE OF MACHINERY & MATERIALS

## 설정·모델·원자료·판단을 함께 보존

| 자료 | 프로젝트 내 위치 |
| --- | --- |
| 슬라이드 원본 | presentations/microduck-learning/index.html |
| 전체 진행·결정 | docs/reproduction-plan.md |
| 실험 구현·실행 | scripts/gait_recovery.py · train_gait_recovery.py |
| 평가 원자료 | artifacts/gait-diagnosis/ · gait-recovery/ |
| 실험 설정·부모SHA·모델 | experiments/gait-recovery/<실험>/ |

# 46. 보상 상세 — 이동·안정성

KIMM 한국기계연구원 KOREA INSTITUTE OF MACHINERY & MATERIALS

## 가중치뿐 아니라 함수·단위·조건을 함께 해석

| 항목 | 최종 가중치 / 의미 |
| --- | --- |
| **이동 / 회전 추종** | +4 / +4 · exp(-오차²/σ²), σ=0.15m/s / 0.35rad/s |
| **직립 / 기준자세** | +2 / +1 · 기준자세는정지때만 |
| **몸통각속도 / 각운동량** | -0.05 / -0.02 · 제곱합 |
| **관절한계 / 행동변화** | -1 / -1 · 실제soft-limit초과 / 원시행동차이² |

# 47. 보상 상세 — 발·머리

KIMM 한국기계연구원 KOREA INSTITUTE OF MACHINERY & MATERIALS

## 발 들기 보상이 보행 품질 전체를 대표하지 않음

| 항목 | 최종 가중치 / 의미 |
|---|---|
| 발공중시간 / 발높이 | +3 / -2 · 0.125~0.3s구간 / 목표높이오차 × 속도 |
| 스윙최고높이 / 미끄러짐 | -1.12890625 / -0.1 · 착지시높이오차 / 접촉발속도² |
| 자기충돌 | -1 · 발사이접촉 |
| 머리자세 / 지속편향 | +2 / -3 · 목표각추종 / 1초EMA편향 |

# 48. 출처와 검증 수준



## 직접 확인·설계 해석·미검증을 분리

### 주요 출처

- Hugging Face 재현 튜토리얼
- Isaac Lab 공식 보행 문서
- 후속 좌우 대칭 증강 PR

### 자료 해석

- 확인: 원자료·코드·실행결과
- 해석: 현상을 설명하는 가설
- 미검증: 실물·장시간강인성

# 49. 머리 자유도와 학습 효율



## 동일 초기 가중치에서 머리 제어 허용·관절 고정 비교

### 공통 조건

- 1024환경 · 학습 seed 42
- 보상·명령·랜덤화·PPO 동일
- 머리 자세 추종·편향 보상 제거
- 관절 공통 감점은 유지

### 바꾸는 조건

- 머리 제어 허용: 4축을 정책이 제어
- 머리 관절 고정: 기본 자세 유지
- 질량·관성·백래시 구조 보존

# 50. 학습 시간과 성공률의 구분

KIMM 한국기계연구원 KOREA INSTITUTE OF MACHINERY & MATERIALS

## 경험 효율·계산 속도·평가 성공을 따로 기록

| 질문 | 비교 지표 |
| --- | --- |
| 얼마나 빨리 배우나 | 동일 보행 기준 최초 통과까지의 경험 수 |
| 물리 계산은 느려지나 | 수집·최적화 시간 / 초당 상태 전이 수 |
| 얼마나 자주 성공하나 | 미사용 평가 seed의 생존·추종 성공 비율 |
| 안정적으로 학습되나 | 3회 연속 평가 통과 / 여러 학습 seed 반복 |

# 51. 기존 정책에 머리 고정 적용

KIMM 한국기계연구원 KOREA INSTITUTE OF MACHINERY & MATERIALS

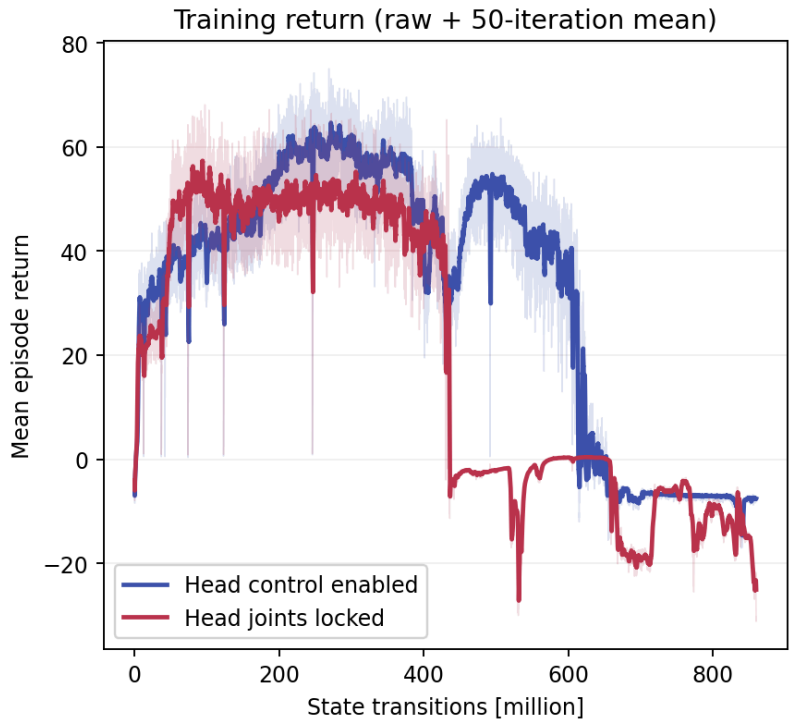
## 고정 직후에도 보행 가능 · 새 학습 결과와 구분

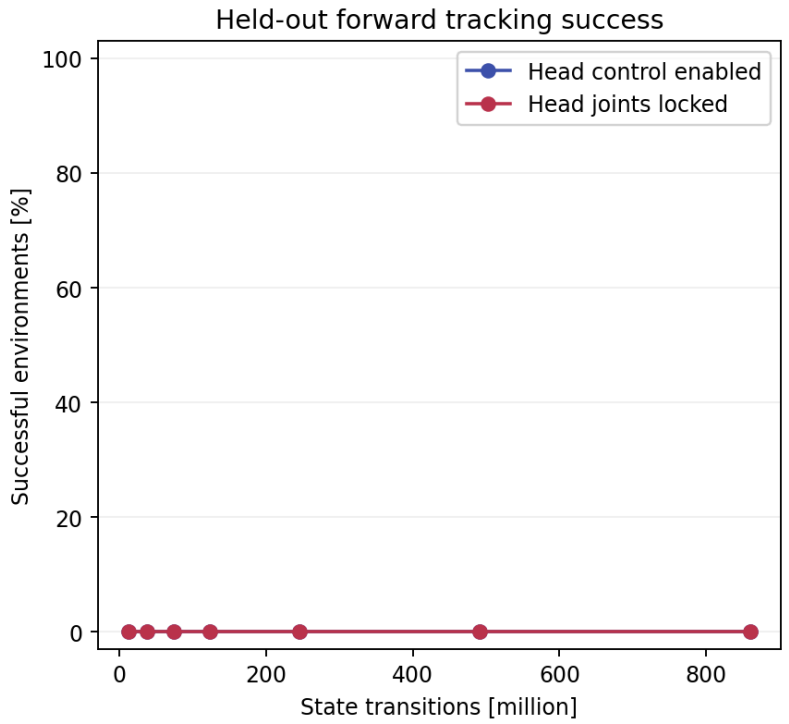
| 측정 항목 | 머리 제어 허용 | 머리 관절 고정 |
|---|---|---|
| 20초 생존율 | 100% | 100% |
| 전진 평균 [m/s] | 0.153 | 0.160 |
| yaw 오차 RMS [rad/s] | 0.300 | 0.318 |
| 몸통 roll 속도 RMS | 0.936 | 0.951 |
| 몸통 pitch 속도 RMS | 0.301 | 0.345 |

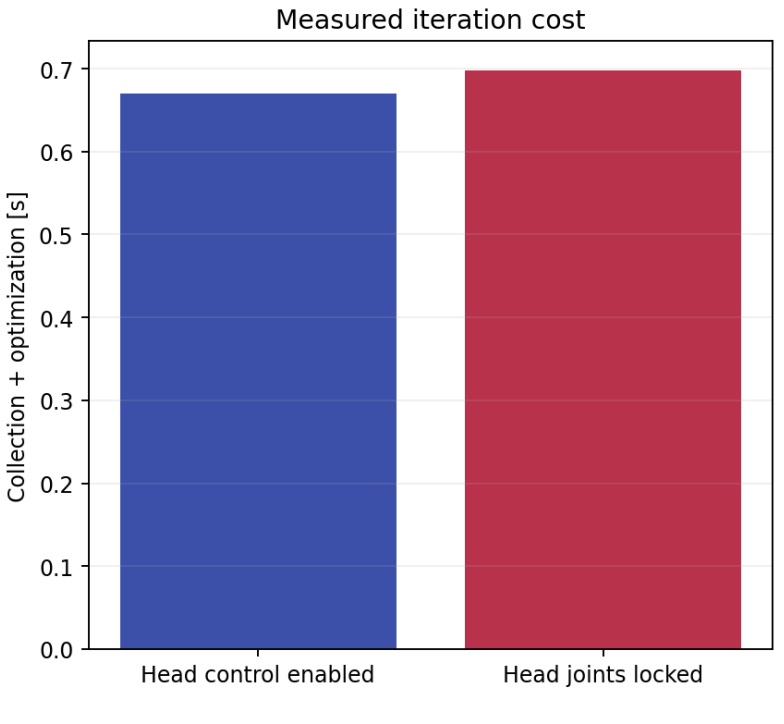
# 52. 머리 자유도 비교 학습 곡선

KIMM 한국기계연구원 KOREA INSTITUTE OF MACHINERY & MATERIALS

## 학습 보상·독립 평가·반복 계산비용을 함께 확인



# 53. 처음부터 학습한 비교 결과

KIMM 한국기계연구원 KOREA INSTITUTE OF MACHINERY & MATERIALS

## 같은 경험 예산의 중간 평가 · 최종 결론 보류

| 조건 | 누적 반복 | 학습 분 | 생존율 | 전체 기준 |
|---|---|---|---|---|
| 머리 제어 허용 | 35,000 | 391.8 | 0.0% | 미달 |
| 머리 관절 고정 | 35,000 | 398.3 | 0.0% | 미달 |

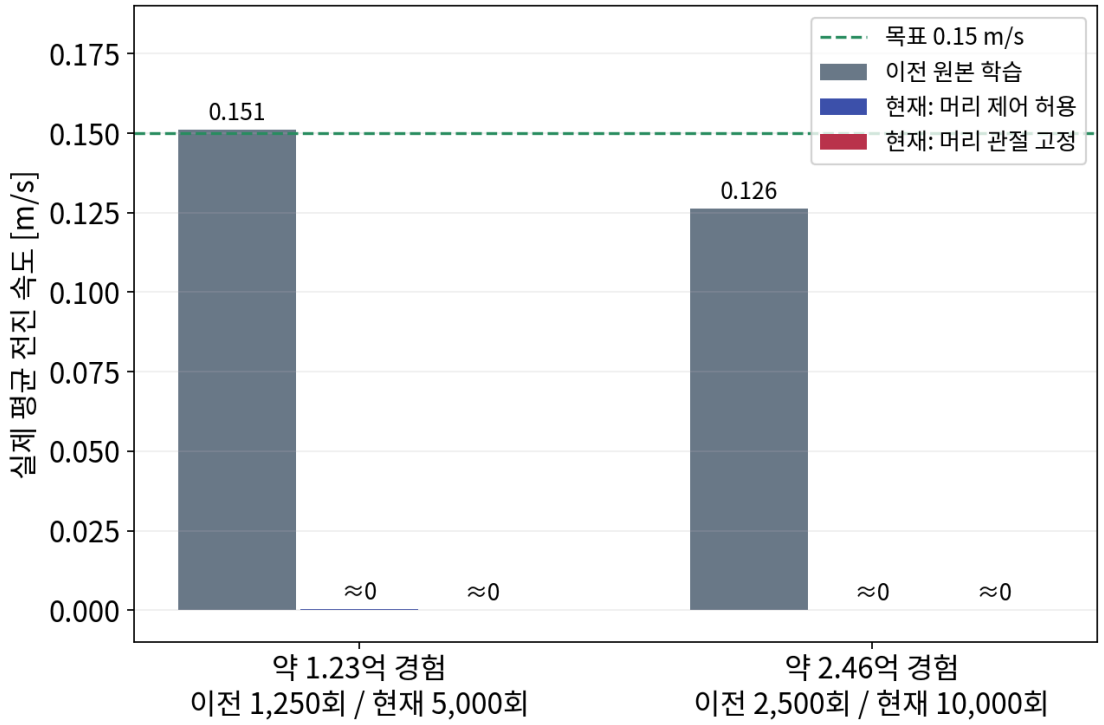
# 54. 이전 학습과 같은 경험량으로 비교

KIMM 한국기계연구원 KOREA INSTITUTE OF MACHINERY & MATERIALS

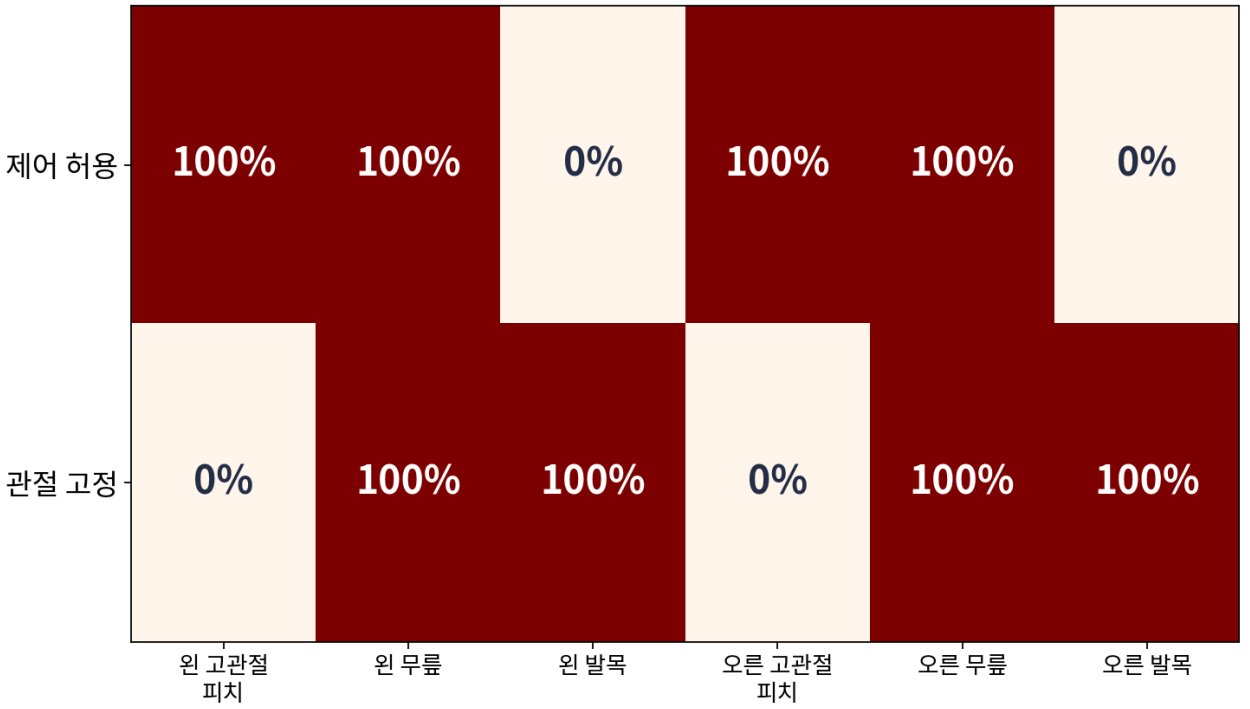
## 환경 수 차이를 보정해도 현재 두 조건은 정체

### 정지 상태 유지와 보행 학습을 구별해야 함

경험량을 맞춰도 현재 학습은 정체




현재 1만 회: 다리 원시 출력의 ±3 초과 비율




동일 평가: 32환경 · 20초 · 전진 0.15 m/s · seed 31415 | 출력 포화는 평가의 초기 2초 제외

# 55. 정체 원인: 출력 포화와 학습 설정



## 실제 제어가 변하는지 확인한 뒤 머리 효과 비교

### 직접 확인

- 주요 다리 출력: 평가 내내 ±3 초과
- 제한 후 적용 행동: 변화량 0
- 원시 탐색 표준편차: 약13~21

### 원인 후보 · 대조실험 필요

- 감점은 제한 후 값 · 탐색은 원시 분포
- 커리큘럼: 경험 기준 4배 빠른 강화
- 미세조정용 교정 항목을 처음부터 적용



**MicroDuck** 평지 보행