

# 심층강화학습 기반 4족보행 로봇 제어를 위한 Nvidia Isaac Gym 분석 및 사례 연구

## Analysis and Case Studies of Nvidia Isaac Gym for Deep Reinforcement Learning-Based Quadruped Robot Control

최요한(Yohan Choi), 지창훈(Chanhun Ji), 한연희(Youn-Hee Han)

Future Convergence Engineering  
Korea University of Technology and Education  
{yoweif, koir5660, yhhan}@koreatech.ac.kr

### 요약

본 논문은 심층강화학습(Deep Reinforcement Learning, DRL)을 기반으로 한 4족보행 로봇 제어와 관련된 기존 연구들을 분석하고, NVIDIA Isaac Gym 플랫폼을 중심으로 기술적 접근 방식과 사례 연구를 종합적으로 검토한다. NVIDIA Isaac Gym은 병렬 시뮬레이션과 GPU 가속을 통해 DRL 학습을 지원하며, 4족보행 로봇이 복잡한 환경에서 안정적이고 효율적으로 동작할 수 있도록 돕는다. 본 논문은 다양한 DRL 알고리즘의 특징과 성능을 비교하며, 이를 활용한 로봇 제어 사례를 분석하여 현재 기술의 한계와 향후 연구 방향을 논의한다.

**키워드:** 4족보행 로봇, 심층강화학습, NVIDIA Isaac Gym, 병렬 시뮬레이션, Sim-to-Real 전이, 로봇 제어

### Abstract

This paper provides a comprehensive analysis and survey of existing research on quadruped robot control using deep reinforcement learning, focusing on the NVIDIA Isaac Gym platform. NVIDIA Isaac Gym supports DRL training through parallel simulation and GPU acceleration, enabling quadruped robots to operate stably and efficiently in complex environments. The paper compares the characteristics and performance of various DRL algorithms and analyzes case studies to identify current technological limitations and future research directions.

**Keywords:** Quadruped Robots, Deep Reinforcement Learning, NVIDIA Isaac Gym, Parallel Simulation, Sim-to-Real Transfer, Robot Control

## 1. 서론

최근 로봇 공학에서 심층강화학습(Deep Reinforcement Learning, DRL)은 전통적인 제어 기법의 한계를 극복하고, 보다 복잡한 환경에서 자율성을 구현할 수 있는 새로운 가능성을 제시하고 있다. 특히, 4족보행 로

봇은 높은 이동성과 안정성을 바탕으로 다양한 환경에서의 활용이 가능하지만, 제어와 학습 측면에서 여전히 기술적 과제를 안고 있다. 균형 유지, 장애물 극복, 효율적 경로 탐색과 같은 기본적인 제어 문제는 복잡한 역학적 특성과 환경적 요인으로 인해 해결하기 어렵다. DRL은 데이터 기반 학습을 통해 이러한 문제를 해결할 수 있는 가능성을 제공하며, 로봇이 이전에 경험하지 못한 상황에도 적응할 수 있는 능력을 갖추도록 한다.

DRL 기반 4족보행 로봇 제어의 주요 도전 과제는 학습 시간의 과도한 소요, 시뮬레이션 환경과 실제 환경 간의 격차(Sim-to-Real Gap), 대규모 학습 환경 구축의 어려움 등으로 요약된다. 학습 시간 문제는 대규모 데이터 수집과 연산이 필요한 DRL의 특성상 필연적으로 발생하며, 이는 기존의 CPU 기반 시뮬레이션 환경에서 더욱 두드러진다. 또한, 시뮬레이션 환경에서 학습된 정책을 실제 로봇에 적용하려면, 실제 환경에서의 물리적 노이즈와 환경 변화에 대한 적응력이 필요하다. 이러한 격차는 로봇의 안정적인 작동을 저해하는 주요 원인으로 작용한다.

NVIDIA Isaac Gym은 이러한 한계를 해결하기 위해 설계된 GPU 기반 병렬 시뮬레이션 플랫폼이다. 이 플랫폼은 수천 개의 시뮬레이션 환경을 단일 GPU에서 병렬로 처리할 수 있어 데이터 처리 속도를 크게 향상시키며, DRL 학습의 효율성을 극대화한다. 또한, Isaac Gym은 물리적 파라미터의 런타임 랜덤화, GPU 가속 텐서 API와 같은 고급 기능을 제공하여, 로봇이 다양한 조건에서 학습할 수 있도록 지원한다. 이러한 특징은 DRL 기반 로봇 제어 학습에서 필수적인 요소로 작용하며, 특히 시뮬레이션-현실 격차를 줄이고 학습된 정책의 현실 적용 가능성을 높이는 데 기여한다.

본 논문은 NVIDIA Isaac Gym을 활용한 4족보행 로봇 제어와 관련된 기존 연구를 종합적으로 분석하고, DRL 기반 로봇 제어 기술의 현재 연구 상황을 논의한다. 연구의 주요 목적은 다음과 같다. 첫째, DRL 알고리즘(Proximal Policy Optimization, Soft Actor-Critic 등)의 특징과 활용 사례를 분석하

여, 4족보행 로봇 제어에서의 성능과 한계를 평가한다. 둘째, NVIDIA Isaac Gym이 제공하는 기술적 이점과 이를 활용한 로봇 제어 학습의 가능성을 평가한다. 셋째, 사례 연구를 통해 DRL 학습 결과가 실제 환경에서 성공적으로 적용될 수 있음을 입증하며, 현재 연구의 한계를 파악한다.

본 연구는 4족보행 로봇 제어 분야에서 DRL의 가능성과 한계를 이해하는 데 필요한 이론적 및 실용적 자료를 제공하며, NVIDIA Isaac Gym을 활용한 로봇 학습의 기술적 가능성과 실용성을 평가한다. 특히, 본 논문은 DRL을 활용한 4족보행 로봇 제어의 이론적 기초를 이해하고, 다양한 알고리즘과 학습 환경에서의 성능을 비교함으로써 연구자들이 새로운 연구를 설계하는 데 있어 유용한 참고 자료를 제공할 것이다. 이러한 연구 조사는 DRL 기반 로봇 제어의 현재 상태를 정확히 이해하고, 향후 연구와 응용 가능성을 확장하는 데 기여할 것으로 기대된다.

## 2. NVIDIA ISAAC GYM 소개

Table 1. URDF와 MJCF 특징 비교

항목	URDF	MJCF
형식	XML	XML
주 사용 목적	로봇의 기계적 구조 표현 및 시뮬레이션	동적 물리 시뮬레이션 및 DRL 연구
물리 속성 정의	제한적(기본적인 관성, 질량 등)	매우 상세(마찰, 충돌, 힘, 속도, 근육 등)
확장성	제한적	매우 유연, 복잡한 모델링 가능
주요 사용 사례	ROS, Gazebo 등 로봇 운영 시스템	MuJoCo 기반 DRL 및 물리 시뮬레이션

Nvidia Isaac Gym은 로봇 DRL 연구를 위해 설계된 GPU 기반 물리 시뮬레이션 플랫폼이다. 수천 개의 환경을 단일 GPU에서 병렬로 처리하며, 로봇 제어 및 학습과 관련된 강력한 기능을 제공한다. 이 플랫폼은 다음과 같은 주요 특징을 포함한다.

## 2.1. 주요 특징

NVIDIA Isaac Gym은 로봇 시뮬레이션 환경 설계와 학습 가속화를 위한 다양한 기능을 제공한다. 주요 특징은 다음과 같다.

- 다양한 모델링 파일 형식 지원: URDF (Unified Robot Description Format) 및 MJCF (MuJoCo XML File) 파일을 임포트할 수 있으며, 임포트된 3D 메시(Mesh)에 대해 자동으로 볼록 분해(Convex Decomposition)를 수행하여 물리 시뮬레이션을 용이하게 한다. URDF와 MJCF 특징 비교는 Table 1에 정리하였다.
- GPU 가속 텐서 API: 이것은 NVIDIA Isaac Gym의 핵심적인 기능 중 하나로, DRL 환경의 상태 평가와 액션 적용을 GPU에서 직접 수행할 수 있도록 한다. 이 API는 환경 데이터를 GPU 메모리에서 바로 처리하며, CPU와 GPU 간의 데이터 전송 병목현상을 제거하여 처리 속도를 크게 향상시킨다. 구체적으로, 이 API를 활용하면 각 환경의 상태(예: 로봇의 관절 위치, 속도, 센서 값)를 GPU 상에서 직접 계산하고, 이를 바탕으로 액션(예: 모터 토크, 힘)을 즉시 적용할 수 있다. 또한, 수천 개의 환경이 병렬로 실행되는 상황에서도 GPU의 텐서 연산 능력을 활용하여 높은 처리 효율을 유지할 수 있다. 이는 대규모 시뮬레이션에서 학습 속도를 비약적으로 증가시키며, 더욱 복잡한 환경에서도 실시간 학습이 가능하게 한다. 이와 같은 특성은 NVIDIA Isaac Gym이 고성능 DRL 연구에 적합한 플랫폼으로 자리 잡는 데 중요한 역할을 한다.
- 다양한 환경 센서 지원: NVIDIA Isaac Gym은 로봇 및 시뮬레이션 환경에서 발생하는 다양한 물리적 데이터를 실시간으로 측정하고 제공할 수 있도록 설계되었다. 이 기능은 로봇의 상태와 동작을 정확히 파악하고 제어하기 위해 필수적인 데이터를 다룬다. 구체적으로, NVIDIA Isaac Gym은 다음과 같은 센서를 통해 다양한 데이터를 지원한다:
  - ① 위치(Position): 로봇 링크 또는 관절의 현재 좌

표를 측정하여 로봇의 공간적 상태를 파악

- ② 속도(Velocity): 관절의 회전 속도나 링크의 선형 속도를 제공하여 동적 상태 추적
  - ③ 힘(Force): 특정 지점에서 발생하는 접촉력 또는 외부 환경과의 상호작용 데이터를 측정
  - ④ 토크(Torque): 로봇 관절에서 발생하는 회전력을 측정하여 동작의 세부 제어 가능
- 이러한 센서 데이터는 GPU에서 병렬로 처리되어 다수의 시뮬레이션 환경에서도 실시간으로 제공된다. 예를 들어, 로봇이 불균형한 표면을 걷는 시뮬레이션에서 힘과 토크 데이터를 활용하면, 로봇의 균형 유지와 충격 흡수를 효과적으로 학습할 수 있다. 이를 통해 DRL 환경에서 더욱 정교한 제어 정책을 학습하거나, 로봇의 물리적 성능을 분석 및 개선할 수 있다.
- 물리 파라미터의 런타임 랜덤화: NVIDIA Isaac Gym은 이 기능을 통해 시뮬레이션 환경의 다양성과 현실성을 크게 향상시킨다. 이 기능은 시뮬레이션 실행 중에 마찰 계수, 관성, 질량, 충돌 특성 등과 같은 물리적 파라미터를 동적으로 변동시킬 수 있도록 지원한다. 이를 통해 학습 중인 DRL 에이전트가 다양한 조건에서 적응력을 기를 수 있도록 환경을 조성한다. 특히, 도메인 랜덤화는 시뮬레이션과 실제 환경 간의 차이를 줄이는 데 유용하다. 시뮬레이션에서 다양한 물리적 조건을 경험하게 함으로써, 학습된 정책이 실제 환경에서도 안정적으로 작동할 가능성을 높인다. 예를 들어, 로봇의 바닥 마찰 계수를 랜덤화하여 다양한 표면에서의 로봇 이동성을 학습하거나, 링크의 질량과 관성 값을 변동시켜 제조 공정의 불확실성에 대비할 수 있다. NVIDIA Isaac Gym은 이러한 랜덤화를 GPU 기반으로 병렬 처리할 수 있어, 수천 개의 시뮬레이션 환경을 동시에 실행하며 학습 속도를 극대화한다. 이 기능은 특히 복잡한 로봇 제어 문제와 시뮬레이션-현실 격차(Sim-to-Real Gap)를 줄이기 위한 DRL 연구에 매우 유용하다.
  - 야코비안(Jacobian) 및 역기구학(Inverse

Kinematics) 지원: 야코비안은 로봇의 관절 공간(joint space)과 작업 공간(task space) 사이의 관계를 수학적으로 나타내며, 특정 작업 지점에서의 속도 및 힘 변환을 효율적으로 계산하는 데 활용된다. 이를 통해 작업 공간에서의 위치 및 방향 제어를 관절 공간으로 변환할 수 있다. 역기구학 기능은 로봇의 작업 공간에서 목표 위치를 설정했을 때, 이를 달성하기 위한 관절 값(joint configuration)을 계산하는 데 사용된다. NVIDIA Isaac Gym은 이러한 계산을 GPU에서 병렬 처리할 수 있도록 최적화하여, 수천 개의 환경에서 동시에 역기구학 문제를 해결할 수 있다. 이는 복잡한 로봇의 실시간 제어 및 DRL 환경 설계에 큰 도움을 준다. 예를 들어, 다관절 로봇의 그리퍼(Gripper)를 작업 공간에서 정확한 위치로 이동시키기 위한 제어 문제를 다룰 때, NVIDIA Isaac Gym의 야코비안 및 역기구학 지원 기능은 계산 속도와 정확도를 크게 향상시킨다. 이러한 기능은 특히 DRL 과정에서 필수적인 역할을 수행한다.

## 2.2. NVIDIA Isaac Gym 계층 구조

NVIDIA Isaac Gym의 시뮬레이션 환경은 Figure 1과 같이 계층적 구조를 기반으로 설계되어, 다양한 객체와 환경을 체계적으로 관리할 수 있도록 구성되어 있다. 이 계층 구조는 시뮬레이션의 확장성과 복잡성을 효율적으로 처리하는 데 중요한 역할을 한다.

시뮬레이션의 최상위 계층은 sim 인스턴스로, 전체 시뮬레이션 환경을 관리하는 역할을 한다. sim은 여러 독립적인 환경을 포함하며, 이러한 환경들은 각각 env 인스턴스로 표현된다. 각 env는 고유

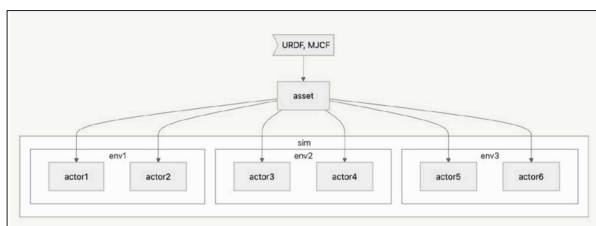


Figure 1. Isaac Gym의 계층 구조

한 시뮬레이션 공간을 제공하며, 환경 간의 상호작용 없이 독립적으로 작동할 수 있다. 이를 통해 여러 시뮬레이션 환경을 병렬로 실행하거나 서로 다른 실험 설정을 동시에 구현할 수 있다. 각 env 내부에는 actor 인스턴스가 포함된다. actor는 로봇, 장애물, 센서 등 물리적 상호작용을 수행하는 객체를 나타낸다. actor는 실제 동작이나 상호작용의 주체로서, DRL 에이전트가 제어하거나 관찰할 수 있는 대상이 된다. actor는 asset 인스턴스를 기반으로 생성된다. asset은 URDF 또는 MJCF 파일을 통해 정의되며, 로봇이나 환경 요소의 기본적인 설계와 물리적 속성을 포함한다. 이러한 구조는 asset이 로봇이나 장애물의 기본 청사진(class) 역할을 하고, actor가 이를 기반으로 생성된 구체적 인스턴스(instance) 역할을 한다는 점에서 “asset : actor = class : instance”의 관계로 요약할 수 있다. 결과적으로 sim은 전체 시뮬레이션을 관리하며, env는 각각의 실험 환경(예: 로봇이 장애물을 극복하는 공간)을 제공한다. 이 환경 내의 actor는 로봇 본체, 장애물, 센서와 같은 요소로 구성되며, 이러한 요소들은 asset 인스턴스를 기반으로 정의된 구조를 따른다. 이러한 계층적 설계는 다양한 시뮬레이션 구성 요소를 체계적으로 관리할 수 있게 하며, NVIDIA Isaac Gym을 DRL 시뮬레이션 환경의 강력한 도구로 자리 잡게 한다.

## 2.3. 최신 업데이트(Preview 4 Release)

2022년 2월에 출시된 NVIDIA Isaac Gym Preview 4 버전에서는 DRL 워크로드와 관련하여 플랫폼의 기능이 대폭 강화되었다. 이번 업데이트는 PhysX 물리 엔진과의 통합, 새로운 기능 추가, 그리고 사용자 설정의 확장으로 시뮬레이션의 정밀도와 효율성을 한 단계 끌어올렸다.

첫째, PhysX 통합을 통해 NVIDIA Isaac Gym과 Omniverse Isaac Sim 2022.1 간의 물리 엔진 구현이 정렬되었다. 이를 통해 Omniverse 플랫폼으로의 마이그레이션이 더욱 간소화되었으며, 다양한 물리적 시뮬레이션에서 일관된 결과를 얻을 수 있게

되었다. 둘째, SDF (Signed Distance Field) 충돌 지원이 새롭게 추가되었다. 이를 통해 나사와 볼트처럼 복잡한 기하학적 객체 간의 충돌을 보다 정밀하게 시뮬레이션할 수 있으며, 현실적인 물리 환경의 재현 가능성이 높아졌다. 이 기능은 정교한 기계 구조를 포함한 로봇 제어 및 설계에 유용하게 활용될 수 있다. 셋째, NVIDIA는 DRL 샘플을 추가로 제공하여 연구자들이 더욱 다양한 사례를 활용할 수 있도록 했다. 이러한 샘플들은 NVIDIA-Omniverse/IsaacGymEnvs 저장소에서 다운로드 가능하며, DRL 환경 설계 및 실험의 시작점을 제공한다. 넷째, 향상된 시뮬레이션 정확도를 통해 더 높은 품질의 시뮬레이션 결과를 제공한다. 자이로스코픽 힘이 기본적으로 활성화되어 로봇 동작의 물리적 정확성을 크게 개선하였다. 이는 특히 4족보행 로봇과 같은 복잡한 시스템에서 중요한 영향을 미친다. 다섯째, 세부 설정 가능성이 확대되었다. 사용자는 각 asset 및 형태별로 `rest_offset`과 `contact_offset`을 사용자 정의할 수 있어, 시뮬레이션의 세부적인 물리적 특성을 조정할 수 있다. 이는 고유한 시뮬레이션 환경을 구성하는 데 있어 높은 유연성을 제공한다. 여섯째, 구면 관절 파싱 지원이 URDF 임포터에 추가되어, 구형(ball) 관절을 포함한 로봇 모델의 구현이 가능해졌다. 이는 보다 복잡한 로봇 구조를 다룰 수 있도록 지원한다. 마지막으로, 버그 수정을 통해 플랫폼의 안정성이 향상되었다. 탄성 충돌 동작, 고정된 기반 액터 리셋, 마찰 랜덤화, 마찰 모드 설정 등과 관련된 다양한 버그가 해결되었다.

NVIDIA Isaac Gym Preview 4 버전의 이러한 업데이트는 4족보행 로봇과 같은 복잡한 로봇의 제어 정책 학습을 대규모로 수행할 수 있는 가능성을 더욱 확장시킨다. 특히, 병렬 시뮬레이션과 물리적 도메인 랜덤화 기능은 실제 환경을 효과적으로 모방하고, DRL 알고리즘의 일반화 성능을 향상시키는 데 크게 기여한다. 이를 통해 연구자와 개발자는 더욱 현실적인 조건에서 DRL 기반 로봇 제어 정책을 설계하고 검증할 수 있게 되었다.

### 3. 4족보행 로봇 제어

#### 3.1. 4족보행 로봇의 특성



Figure 2. 잘 알려진 4족보행 로봇들 (왼쪽부터 MIT Mini Cheeta [3], Stanford Doggo [4], Boston Dynamics Spot [5])

4족보행 로봇(Figure 2 참고)은 이동성과 안정성이 뛰어나며 다양한 환경에 적응할 수 있는 특성을 가진다. 이러한 로봇은 네 개의 다리를 사용하여 움직임을 제어하며, 비포장 도로나 장애물이 많은 환경에서도 높은 이동성을 유지할 수 있다. 특히, 복잡한 지형에서도 균형을 유지하면서 다양한 태스크를 수행할 수 있어 활용도가 높다. 이러한 특성으로 인해 재난 구조, 물류 이동, 탐사와 같은 실용적 응용 분야에서 주로 사용된다.

#### 3.2. 기존 제어 방법론

4족보행 로봇의 제어에는 전통적 제어 방법, 최적화 기반 제어, DRL 기반 제어와 같은 다양한 접근 방식이 존재한다 [6]. 각각의 방법은 고유한 강점과 한계를 가지며, 로봇이 특정 환경에서 효과적으로 동작할 수 있도록 다양한 제어 방식을 결합해 사용하는 경우도 많다.

우선, 전통적 제어 방법은 가장 오래된 제어 기법 중 하나로, PID 제어와 모델 기반 제어가 대표적이다. PID 제어는 비례(Proportional), 적분(Integral), 미분(Derivative) 제어 요소를 결합하여 로봇의 관절 위치와 속도를 정밀하게 제어한다. 이 방식은 설계와 구현이 간단하고 안정적이라는 장점이 있지만, 복잡한 동적 환경에서는 한계를 가진다. 반면, 모델 기반 제어는 로봇의 물리적 특성을 수학적으로 모델링하여 제어 알고리즘을 설계하는 방식이다. 이를 통해 로봇의 동적 특성을 잘 반영할 수 있지만, 모델링 오차와 계산 복잡성 때문에 실제 적용에서는 성능이 저하될 가능성이 있다.

다음으로, 최적화 기반 제어는 최적화 알고리즘을 활용하여 로봇의 동작을 계획하는 방식이다. 이 방식은 특정 작업, 예를 들어 장애물 회피나 이동 경로 최적화와 같은 문제를 해결하기 위해 에너지 소비를 최소화하거나 보폭과 궤적을 계산하여 불균형한 지형에서도 안정적인 이동을 가능하게 한다. 이 접근법은 정밀한 제어를 가능하게 하지만, 계산 자원이 많이 필요하며, 실시간 제어에서 효율성 문제가 발생할 수 있다.

마지막으로, DRL 기반 제어는 학습 데이터를 활용하여 로봇의 동작을 학습시키는 현대적인 접근 방식이다. 이 방법은 시뮬레이션 환경에서 로봇이 다양한 행동을 학습하도록 설계되며, 학습된 정책을 실제 환경에 적용(즉, Sim-to-Real 전이)하여 검증한다. Nvidia Isaac Gym을 활용한 연구에서는 강화학습을 통해 4족보행 로봇이 자신의 높이의 2배에 달하는 장애물을 점프하거나 좁은 공간을 통과하는 동작을 학습할 수 있음을 입증하였다. 이러한 방식은 전통적 제어 방식보다 적응성과 확장성이 뛰어나지만, 학습 데이터와 계산 자원 요구가 높다는 점에서 단점이 있다.

이처럼 다양한 제어 기법은 각기 다른 상황에 맞게 활용될 수 있으며, 특히 DRL 기반 제어는 복잡한 환경에서 로봇의 성능을 극대화할 수 있는 강력한 도구로 자리 잡고 있다.

### 3.3. 제어를 위한 핵심 문제

4족보행 로봇 제어는 다양한 응용 환경에서 안정적이고 효율적으로 동작하기 위해 여러 도전 과제를 해결해야 한다. 대표적인 도전 과제는 균형 유지, 장애물 극복, 그리고 효율적인 경로 탐색으로 나눌 수 있다.

첫 번째로, 균형 유지는 복잡한 지형이나 기울어진 경사면에서 로봇이 넘어지지 않고 안정적으로 이동할 수 있도록 보장하는 문제이다. 특히, 경사로와 같은 환경에서는 로봇의 무게 중심 이동과 발판 위치를 정밀하게 제어해야 한다. 이를 위해, 야코비안 행렬을 활용하여 로봇의 상태를 수학적으로 분석하고 균형을 유지하기 위한 최적의 움직임 계산을 하는 기법이 활용된다. 이러한 균형 유지 기술은 로봇이 다양한 환경에서 신뢰성 있는 동작을 수행하는 데 필수적이다.

두 번째로, 장애물 극복은 로봇이 물리적 제약을 극복하여 환경을 탐색하고 이동할 수 있도록 하는 문제이다. 장애물 점프, 낮은 구조물 아래를 기어가기, 좁은 통로를 통과하는 것과 같은 고난도의 동작은 세밀한 운동 제어를 필요로 한다. 이러한 문제를 해결하기 위해 강화학습이 주로 활용된다. 예를 들어, Robot Parkour 연구에서는 4족보행 로봇이 자신의 길이의 1.5배에 해당하는 간격을 뛰어넘는 기술을 학습하도록 설계하여, 실제 환경에서도 성공적으로 적용 가능함을 입증하였다. 이와 같은 기술은 강화학습의 적응성과 로봇의 운동 제어 능력을 극대화한다.

마지막으로, 효율적 경로 탐색은 로봇이 장애물과 복잡한 환경을 피해 목적지에 도달할 최적의 경로를 찾는 문제이다. 기존의 전통적 경로 탐색 알고리즘은 비교적 정형화된 환경에서 잘 작동하지만, 강화학습은 동적이고 예측 불가능한 환경에서도 적응 가능한 경로를 학습할 수 있는 장점을 제공한다. 예를 들어, DRL 알고리즘은 로봇이 움직이는 장애물이나 변화하는 환경 조건에 실시간으로 반응할 수 있도록 학습을 지원한다.

이러한 제어 문제와 도전 과제를 해결하기 위한 기술과 연구는 4족보행 로봇의 응용 가능성을 크게 확대하고 있다. 특히, DRL 기반 제어는 기존의 제어 방식으로 해결하기 어려운 문제를 새로운 방식으로 접근할 수 있게 하며, 로봇 제어 분야에서 혁신적인 가능성을 열어 주고 있다.

## 4. NVIDIA ISAAC GYM 활용 심층강화학습 설계

### 4.1. Markov Decision Process (MDP) 설계

NVIDIA Isaac Gym을 활용하여 4족보행 로봇의 DRL을 설계할 때, MDP (Markov Decision Process)는 핵심적인 프레임워크로 사용된다. MDP는 상태 공간(State Space), 행동 공간(Action Space), 보상 함수(Reward Function), 그리고 전이 확률(Transition Probability)로 구성되며, 로봇 제어 학습의 기초를

형성한다.

상태 공간은 로봇의 현재 상태를 설명하는 변수들의 집합이다. 이를 통해 로봇은 환경에서 자신의 위치와 조건을 이해하고 적절한 행동을 선택할 수 있다. 예를 들어, 관절의 각도와 속도, 관성 센서(IMU)에서 측정된 기울기 및 가속도, 발판 접촉 여부, 주변 환경 정보(예: 장애물 거리 및 위치) 등이 포함된다. Isaac Gym은 GPU 기반 텐서 API를 사용하여 이러한 데이터를 실시간으로 처리하며, 수천 개의 환경에서 병렬적으로 상태를 관리할 수 있도록 지원한다.

행동 공간은 로봇이 취할 수 있는 모든 가능한 행동의 집합을 정의한다. 4족보행 로봇의 경우, 각 관절에 적용할 토크나 속도와 같은 제어 명령이 행동 공간에 포함된다. 연속형 행동 공간이 주로 사용되며, 이는 각 관절의 값을 실수 범위로 표현하여 세밀한 제어가 가능하도록 한다. Isaac Gym은 이러한 연속형 행동 공간의 시뮬레이션을 효과적으로 처리한다.

보상 함수는 학습 목표를 설정하고 로봇이 특정 행동을 학습하도록 유도하는 역할을 한다. 이동 거리, 에너지 효율성, 균형 유지, 장애물 회피와 같은 요소들이 보상 설계에 반영된다. 예를 들어, 로봇이 전진하거나 장애물을 회피할 때 추가 보상을 부여하여 목표 행동을 강화할 수 있다. Isaac Gym은 이러한 보상을 병렬 처리로 계산하여 학습 효율성을 높인다.

전이 확률은 로봇이 특정 행동을 취했을 때 다음 상태로 전이될 가능성을 나타낸다. NVIDIA Isaac Gym은 PhysX 물리 엔진을 활용하여 로봇의 행동이 환경에 미치는 영향을 정밀하게 시뮬레이션하며, 로봇의 동작과 물리적 상호작용을 자연스럽게 모델링한다.

## 4.2. DRL 알고리즘 선택

NVIDIA Isaac Gym에서 구현된 DRL은 병렬 처리와 GPU 가속을 최대한 활용할 수 있는 학습 알고리즘을 선택하는 것이 중요하다. 4족보행 로봇 제어를 위한 DRL에서 주로 사용되는 대표적인 알고리즘은 Proximal Policy Optimization (PPO)과 Soft Actor-Critic (SAC)이다. 이 두 알고리즘은 각각의 특성과 장점을 통해 다양한 환경에서 로봇의 제어 정책

학습을 지원한다.

### • Proximal Policy Optimization (PPO)

Proximal Policy Optimization (PPO)는 DRL에서 널리 사용되는 정책 기반(policy-based) 알고리즘으로, 학습 안정성과 효율성을 동시에 달성하기 위해 설계되었다. 이 알고리즘은 기존의 정책 네트워크를 점진적으로 갱신하여 과도한 변화를 방지하는 데 중점을 둔다. 구체적으로, PPO는 기존 정책과 새롭게 갱신된 정책 사이의 차이를 제한하는 클리핑(clipping) 기법을 도입하여, 학습 과정에서 발생할 수 있는 정책의 불안정성을 최소화한다. 이러한 특성은 PPO를 다양한 환경에서 안정적으로 학습할 수 있는 알고리즘으로 자리 잡게 한다.

PPO의 또 다른 장점은 샘플 효율성이다. 이 알고리즘은 동일한 경험 데이터를 여러 번 재사용할 수 있어 학습 효율을 높이며, 데이터를 효과적으로 활용한다. 또한, 병렬 시뮬레이션 환경에서 다수의 에이전트가 동시에 학습할 수 있어 대규모 학습 작업에도 적합하다. 예를 들어, NVIDIA Isaac Gym과 같은 플랫폼에서는 수천 개의 시뮬레이션 환경에서 병렬 학습이 가능하며, PPO는 이러한 환경의 병렬 처리 성능을 극대화한다.

PPO는 특히 4족보행 로봇의 균형 유지, 장애물 회피와 같은 복잡한 동작을 학습하는 데 효과적이다. 예를 들어, 로봇이 불균형한 지형에서 안정적으로 이동하거나 예측 불가능한 장애물을 피할 때, PPO는 이러한 목표를 달성하기 위한 안정적이고 효율적인 학습을 제공한다. 이러한 특성으로 인해 PPO는 다양한 DRL 문제에서 표준 알고리즘으로 널리 사용되고 있다.

### • Soft Actor-Critic (SAC)

Soft Actor-Critic (SAC)는 연속형 행동 공간을 가진 환경에서 높은 샘플 효율성과 안정적인 학습을 제공하는 DRL 알고리즘이다. SAC는 정책 네트워크의 탐색 능력을 극대화하기 위해 엔트로피 보상(entropy regularization)을 추가한다. 이는 에이전트가 불확실한 환경에서 새로운 행동을 더 많이 탐색할 수 있도록 유도하며, 복잡한 환경에서의 학습 성능을 크게 향상시킨다.

SAC는 오프정책(Off-policy) 학습을 사용하여 학습 과정에서 수집된 경험 데이터를 반복적으로 활용한다. 이를 통해 샘플 효율성이 극대화되며, 기존 데이터를 재활용함으로써 학습 비용을 줄일 수 있다. 또한, SAC는 안정적인 정책 학습을 목표로 설계되었으며, 불확실한 환경에서도 신뢰할 수 있는 동작을 학습할 수 있는 능력을 갖추고 있다.

SAC는 특히 높은 정밀도를 요구하는 작업에 적합하다. 예를 들어, 4족보행 로봇이 높은 장애물을 뛰어넘거나 좁은 공간을 통과하는 것과 같은 고난도의 동작을 학습할 때 SAC는 그 유용성을 입증한다. 또한, 엔트로피 보상 덕분에 로봇은 보다 다양하고 창의적인 행동을 탐구할 수 있으며, 이는 새로운 환경에 적응하는데 필요한 적응성을 강화한다.

### 4.3. NVIDIA Isaac Gym 기반 DRL 학습 흐름

NVIDIA Isaac Gym 기반 DRL은 효율적인 병렬 처리와 GPU 가속을 활용하여 다음과 같은 단계로 진행된다. 먼저 MDP 정의에 따라 수천 개의 독립적인 시뮬레이션 환경이 GPU에서 병렬로 초기화된다. 각 환경은 고유의 초기 상태를 설정하며, 학습이 시작될 준비를 마친다.

학습 과정에서 각 환경의 상태는 GPU 텐서 API를 통해 실시간으로 수집된다. 수집된 상태 데이터는 정책 네트워크의 입력으로 사용되며, 네트워크는 최적의 행동을 예측한다. 예측된 행동은 각 환경에서 실행되며, NVIDIA Isaac Gym은 PhysX 엔진을 통해 로봇의 상태 변화를 정밀하게 시뮬레이션한다.

행동 수행 후에는 각 환경에서 발생한 보상이 GPU를 통해 병렬로 계산된다. 보상은 로봇의 전진 거리, 에너지 효율성, 균형 유지 등의 요소를 기준으로 평가되며, 각 환경의 상태는 이에 따라 업데이트된다. 이 데이터는 학습 알고리즘에 입력되어 정책 네트워크를 업데이트하는 데 사용된다.

이러한 학습 과정은 반복적으로 수행되며, 점차적으로 로봇의 정책이 개선된다. Isaac Gym의 병렬 처리와 GPU 가속은 학습 속도를 비약적으로 향상시키고, 대규모 환경에서의 학습을 가능하게 한다. 이

를 통해 4족보행 로봇 제어와 같은 복잡한 문제를 효과적으로 해결할 수 있다. NVIDIA Isaac Gym 기반 DRL 설계는 현실적인 시뮬레이션과 실제 제어 환경 간의 간극을 줄이며, DRL 연구에 강력한 도구를 제공한다.

## 5. 사례 연구

### 5.1. Rapid Motor Adaptation (RMA) 알고리즘 (RSS 2021) [7]

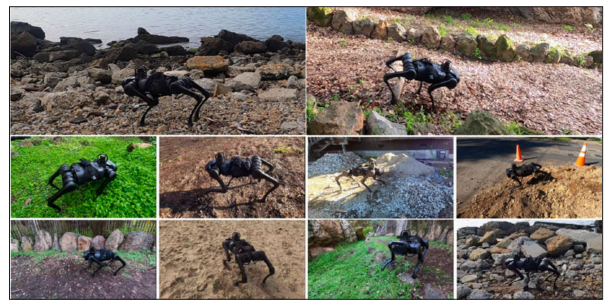


Figure 3. RMA: Rapid Motor Adaptation for Legged Robots (RSS 2021) [7]

#### • 연구 개요

본 연구는 4족보행 로봇이 다양한 환경 변화에 실시간으로 적응할 수 있도록 설계된 Rapid Motor Adaptation (RMA) 알고리즘을 제안한다. RMA는 추가적인 데이터 수집이나 파라미터 조정 없이 새로운 환경에서 바로 작동 가능하도록 설계되었으며, 기존의 DRL 기반 알고리즘과 달리 빠른 적응성을 강조한다. RMA는 두 가지 주요 구성 요소, 기본 정책(Base Policy)과 적응 모듈(Adaptation Module)로 이루어져 있다.

기본 정책은 로봇의 현재 상태와 환경 정보를 바탕으로 관할 동작을 제어하며, 적응 모듈은 로봇의 최근 상태-행동 데이터를 입력으로 받아 환경 정보를 실시간으로 추정한다. 이 두 모듈은 독립적으로 작동하면서도 상호 보완적인 역할을 수행하여 로봇이 예상치 못한 환경 변화에 신속히 대응할 수 있도록 한다. RMA는 시뮬레이션 환경에서 강화학습을 통해 훈련된 뒤, 실제 환경에서 테스트되었으며, 다양한 지형에



서 높은 성능과 안정성을 입증했다.

• 주요 기술

RMA는 두 가지 핵심 모듈, 즉 기본 정책과 적응 모듈로 구성된다. 기본 정책(Base Policy)은 로봇의 상태와 환경 정보를 바탕으로 로봇의 다음 행동을 결정한다. 상태 정보에는 로봇의 관절 위치와 속도, 지면 반발력, 로봇의 무게 중심 등이 포함되며, 환경 정보는 환경 특성을 표현한 외재 벡터(Extrinsics Vector)로 표현된다. 기본 정책은 강화학습을 통해 다양한 시뮬레이션 환경에서 훈련되며, 로봇이 환경 변화에 안정적으로 대응할 수 있는 정책을 학습한다.

적응 모듈(Adaptation Module)은 최근 상태-행동 데이터를 기반으로 환경 정보를 실시간으로 추정한다. 이 모듈은 환경의 물리적 특성(예: 마찰 계수, 경사도, 지면의 강도)을 학습하여 기본 정책에 적합한 환경 벡터를 제공한다. 적응 모듈은 시뮬레이션에서 지도학습(Supervised Learning)을 통해 훈련되며, 데이터의 시계열적 특성을 학습하기 위해 1D CNN 구조를 사용한다.

또한, RMA는 비동기 실행 구조를 채택하여 기본 정책은 100Hz의 높은 주기로 실행되고, 적응 모듈은 10Hz의 낮은 주기로 실행된다. 이러한 비동기 구조는 온보드 컴퓨팅 자원의 효율성을 극대화하며, 저성능 하드웨어에서도 효과적으로 작동할 수 있도록 한다.

• 연구 성과

RMA의 실험 결과는 알고리즘의 강력한 성능과 적응성을 입증했다. RMA는 메모리폼 매트리스, 기름진 표면, 경사로와 같은 다양한 환경에서 높은 성공률(80~100%)을 기록했다. 또한, A1 로봇이 12kg의 추가 하중(자체 무게의 100% 추가)을 안정적으로 운반하면서도 균형을 유지하는 데 성공하였다. 기존 컨트롤러 및 적응 모듈이 없는 RMA와 비교했을 때, 모든 테스트 조건에서 우수한 성능을 보였다. 적응 모듈을 비활성화한 경우 성공률과 이동 거리가 크게 감소했으며, 이는 적응 모듈이 환경 변화에 대응하는 데 필수적임을 나타낸다. 야외 실험에서도 RMA는 흙더미, 자갈, 경사로, 높은 초목이 있는 지형에서 안정적으로 이동하였다. 특히 계단과 경사로 같은 복잡한 환경에

서도 70~100%의 성공률을 기록하며, 기존의 컨트롤러 대비 월등한 성능을 보였다.

결론적으로, RMA는 환경 변화에 실시간으로 적응하며, 다양한 조건에서 높은 성능을 제공하는 알고리즘으로, 4족보행 로봇의 이동성과 적응성을 새로운 수준으로 끌어올렸다. 이 연구는 시뮬레이션과 현실 간의 격차를 줄이는 동시에, 로봇 이동 기술의 발전에 기여하는 중요한 성과를 제시한다.

5.2. GPU 기반 대규모 병렬 심층강화학습 (CoRL 2021) [8]

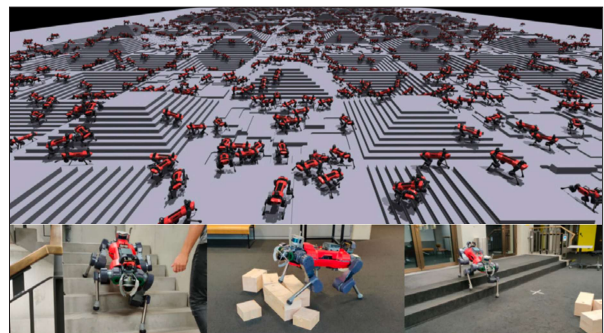


Figure 4. Learning to Walk in Minutes Using Massively Parallel Deep Reinforcement Learning (CoRL 2021) [8]

• 연구 개요

본 연구는 NVIDIA Isaac Gym을 기반으로 4족보행 로봇의 DRL을 가속화하고, Sim-to-Real 전이를 통해 실제 로봇에 성공적으로 적용하는 방법을 제안한다. 기존 DRL 학습은 데이터 수집과 정책 갱신 과정에서 많은 시간이 소요되며, 주로 CPU 기반 병렬 처리에 의존하여 속도와 효율성에 한계를 보였다. 본 연구는 GPU를 활용한 대규모 병렬 학습 환경을 통해 이러한 한계를 극복하고, 로봇 제어 정책 학습 시간을 기존 대비 수백 배 단축하는 데 성공하였다.

연구의 주요 목표는 병렬 학습을 극대화하여 학습 속도를 단축하고, 로봇이 다양한 지형에서 안정적으로 작동할 수 있는 정책을 학습하는 것이다. 또한, 학습된 정책을 실제 로봇 ANYmal C에 적용하여 Sim-to-Real 전이의 성능을 검증하였다. 이 접근법은 복잡한 로봇 학습 문제를 해결하기 위한 새로운 가능성을 제시하며, DRL 연구와 응용 분야에서 중요한 기여

를 한다.

• 주요 기술

본 연구에서는 대규모 병렬 학습을 위한 주요 기술로 NVIDIA Isaac Gym, 게임에서 영감을 받은 자동 커리큘럼, 그리고 PPO 알고리즘을 활용하였다. NVIDIA Isaac Gym 기반 시뮬레이션은 GPU에서 데이터 수집과 정책 업데이트를 병렬로 수행할 수 있도록 설계되었다. 수천 개의 로봇을 동시에 시뮬레이션하며, 모든 데이터 처리와 연산이 GPU 메모리 내에서 이루어진다. 이를 통해 CPU와의 데이터 전송 병목현상을 최소화하고, 시뮬레이션 속도와 효율성을 극대화하였다.

게임에서 영감을 받은 자동 커리큘럼은 학습 난이도를 동적으로 조정하여 로봇이 단계적으로 어려운 지형을 학습하도록 한다. 로봇은 평평한 지형에서 학습을 시작하여 점진적으로 경사로, 계단, 장애물 등 난이도가 높은 지형으로 이동한다. 성공적인 학습은 난이도를 증가시키고, 실패 시 낮은 난이도로 되돌아가 학습을 이어간다. 이를 통해 로봇은 각자의 성능에 맞는 난이도에서 효과적으로 학습할 수 있었다.

PPO 알고리즘은 대규모 병렬 환경에 최적화된 학습 알고리즘으로, 안정성과 효율성을 제공한다. 이 알고리즘은 수십만 개의 샘플을 병렬로 처리하며, 정책 갱신 과정에서 데이터 중복을 최소화하고 학습 성능을 극대화하였다. 또한, PPO는 단순한 설계와 강력한 성능으로 다양한 지형에서 로봇의 동작을 효과적으로 학습하도록 지원하였다.

• 연구 성과

본 연구는 GPU 기반 대규모 병렬 학습을 통해 DRL 학습 속도와 성능에서 뛰어난 성과를 달성하였다. 평평한 지형에서는 4분, 복잡한 지형에서는 20분 만에 학습을 완료하였으며, 이는 기존 연구 대비 수백 배의 학습 속도 향상을 의미한다. 학습된 정책은 다양한 지형(평지, 경사로, 계단, 불규칙 장애물)에서 높은 성공률을 보였다. 예를 들어, 로봇은 최대 25도 경사의 경사로를 오르내릴 수 있었으며, 최대 20cm 높이의 장애물을 극복하는 데 성공하였다. 이와 같은 성능은 기존 CPU 기반 시뮬레이션으로는 달성하기 어려웠던

결과이다.

Sim-to-Real 테스트에서는 학습된 정책을 실제 로봇에 성공적으로 적용하였다. 로봇은 계단과 경사로를 안정적으로 이동하였으며, 센서 노이즈와 불완전한 환경 정보에도 높은 정확도를 유지하였다. 도메인 랜덤화를 통해 시뮬레이션과 현실 간의 격차(Sim-to-Real Gap)를 줄임으로써, 실제 환경에서의 정책 성능을 더욱 강화하였다. 결론적으로, 본 연구는 DRL의 학습 속도를 획기적으로 단축하면서도 높은 성능을 유지하며, 4족보행 로봇 제어 문제에서 새로운 가능성을 열었다. 이 연구는 DRL의 응용 분야를 확장하고, 로봇 학습 분야에서 혁신적인 기여를 제공한다.

5.3. Robot Parkour Learning (CoRL 2023) [9]

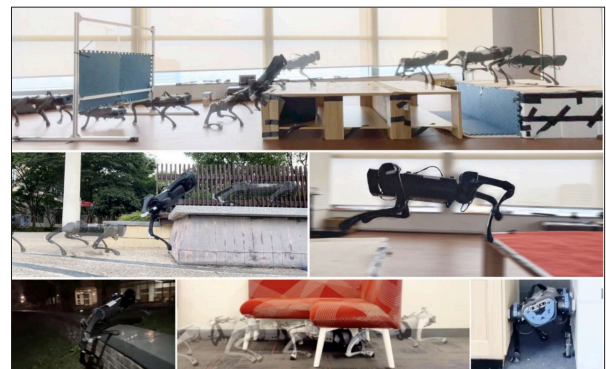


Figure 5. Robot Parkour Learning (CoRL 2023) [9]

• 연구 개요

본 연구는 저가형 4족보행 로봇이 다양한 장애물을 극복하고 복잡한 환경에서 이동할 수 있도록 설계된 로봇 파쿠르 학습 시스템을 제안한다. 로봇 파쿠르는 높이 뛰어넘기, 좁은 공간 통과, 낮은 구조물 아래 기어가기, 높은 장애물 오르기와 같은 다양한 기술을 요구하며, 이러한 기술을 하나의 통합된 정책으로 학습하고 적용하는 것은 큰 도전 과제이다. 기존 방식은 각 기술에 대한 별도 학습 또는 복잡한 보상 설계가 필요했지만, 본 연구는 간단한 보상 구조와 DRL 기반의 두 단계 학습 방법을 통해 이러한 문제를 해결한다.

연구의 주요 목표는 로봇이 실시간 비전 데이터를 활용해 다양한 파쿠르 기술을 학습하고, 시뮬레이션에서 훈련된 정책을 실제 로봇에 성공적으로 전이하는

것이다. 이를 통해 로봇은 다양한 환경에서 자율적으로 적절한 기술을 선택하여 장애물을 극복할 수 있다.

• 주요 기술

로봇 파쿠르 학습 시스템은 두 단계의 강화학습과 정책 증류(distillation) 과정을 통해 설계되었다. 첫 번째 단계는 RL 사전학습(Pre-training)으로, 부드러운 동적 제약 조건(Soft Dynamics Constraints)을 적용하여 학습 초기의 어려운 탐색 문제를 완화한다. 이 단계에서는 로봇이 장애물을 관통(penetrate)하도록 허용하여 지역 최소값(local minima)에 빠지지 않도록 유도하고, 점진적으로 실제 환경의 동적 제약 조건에 적응하도록 학습한다. 이 과정은 자동 커리큘럼을 통해 장애물 난이도를 동적으로 조정하여 로봇의 성능에 맞춘 학습 환경을 제공한다. 두 번째 단계는 RL 세부 튜닝(Fine-tuning)으로, 실제 동적 제약 조건(Hard Dynamics Constraints)을 적용하여 사전학습에서 얻은 기술을 더욱 정교하게 다듬는다. 이 단계에서는 장애물 관통이 허용되지 않으며, 모든 동작이 현실적인 물리 환경에서 수행된다.

정책 증류 과정에서는 각 기술별 학습 정책(예: 뛰어넘기, 기어가기 등)을 하나의 통합된 파쿠르 정책으로 결합한다. 이를 통해 로봇은 다양한 기술을 자율적으로 전환하며 사용할 수 있다. 이 과정에서는 DAgger 알고리즘을 사용하여 시뮬레이션에서 학습된 전문화된 기술 정책을 가르치는 방식으로 단일 정책을 학습한다.

• 연구 성과

본 연구는 시뮬레이션과 실제 환경에서 뛰어난 성과를 보여주었다. 시뮬레이션 결과에서는 로봇이 95% 이상의 성공률로 높은 장애물 오르기, 큰 간격 뛰어넘기, 좁은 공간 통과 등의 기술을 학습하였다. RL 사전학습에서 부드러운 동적 제약 조건을 사용한 접근 방식은 탐색 효율성과 성공률에서 기존 방식(RND, Oracles w/o Soft Dyn)보다 우수하였다. 실제 실험에서는 A1 로봇이 다양한 환경(예: 자갈, 경사로, 캠핑카 아래)에서 학습된 정책을 성공적으로 수행하였다. 로봇은 최대 0.4m 높이의 장애물을 80%의 성공률로 오르고, 최대 0.6m 너비의 간격을 80% 성

공률로 뛰어넘었다. 이 시스템은 시뮬레이션-현실 간 격차를 줄이고, 저가형 로봇에서 고성능 파쿠르 기술을 구현하는 데 성공하였다.

결론적으로, 본 연구는 간단한 보상 구조와 두 단계 학습 과정을 통해 저가형 로봇이 복잡한 파쿠르 기술을 학습하고 수행할 수 있음을 입증하였다. 이 접근법은 로봇 이동성과 자율성 분야의 새로운 가능성을 열어준다.

5.4. Extreme Parkour with Legged Robots (ICRA 2024) [10]



Figure 6. Extreme Parkour with Legged Robots (ICRA 2024) [10]

• 연구 개요

본 연구는 4족보행 로봇이 복잡한 장애물을 극복하고 고난도의 이동 기술을 수행할 수 있도록 설계된 Extreme Parkour 학습 시스템을 제안한다. 기존 로봇 파쿠르 연구는 각 기술을 개별적으로 학습하거나, 사전 설정된 장애물 코스에서만 작동하도록 설계되었다. 반면, 본 연구는 단일 신경망으로 로봇이 고유의 판단을 통해 다양한 장애물을 극복하고 환경에 적응할 수 있도록 설계되었다.

연구의 목표는 단일 신경망을 활용하여 로봇이 깊이 카메라(Depth Camera)로부터 입력 받은 이미지 데이터를 바탕으로 직접 제어 명령을 생성하는 것이다. 이를 통해 로봇은 높은 점프, 긴 간격을 뛰어넘기, 경사로 이동, 그리고 두 다리로 걷기를 포함한 고난도 동작을 학습하고 수행할 수 있었다.

• 주요 기술

Extreme Parkour 학습 시스템은 두 가지 핵심 기술인 Dual Distillation Method와 보상 설계 원칙을

중심으로 구축되었다: Dual Distillation Method는 두 단계 학습을 통해 로봇이 시뮬레이션에서 학습한 내용을 실제 환경에서 활용할 수 있도록 한다. 첫 번째 단계는 강화학습을 통해 스캔 데이터(scandots)와 환경 정보를 기반으로 로봇의 행동 정책을 학습하는 것이다. 두 번째 단계에서는 학습된 정책을 실제 환경에서 활용 가능하도록, 깊이 카메라에서 수집된 데이터만을 사용해 정책을 증류한다. 이를 통해 로봇은 사람의 사전 설정 없이도 장애물의 유형에 따라 이동 방향을 스스로 조정할 수 있게 된다.

보상 설계는 다양한 파쿠르 동작을 자동으로 학습할 수 있도록 단순하지만 효과적인 보상 구조를 기반으로 한다. 로봇은 목표 방향으로의 속도를 최대화하는 동시에, 장애물의 가장자리와 충돌하지 않도록 학습하며, 최적의 동작을 찾는다. 이 보상 구조는 로봇이 지형의 특성에 따라 행동을 조정할 수 있도록 유도한다.

#### • 연구 성과

본 연구는 시뮬레이션과 실제 환경에서 다음과 같은 뛰어난 성과를 보여주었다. 시뮬레이션 결과에서는 로봇이 자신의 높이의 2배에 달하는 장애물을 점프하거나, 자신의 길이의 2배에 해당하는 간격을 뛰어넘는 등 고난도 동작을 학습하였다. 또한, 로봇은 기울어진 경사로와 같은 복잡한 환경에서도 높은 성공률을 기록하였다. 실제 환경 테스트에서는 로봇이 깊이 카메라로 수집한 데이터만을 기반으로 높은 점프, 긴 간격 뛰어넘기, 경사로 이동, 두 다리로 서기를 성공적으로 수행하였다. 특히, 로봇은 새로운 장애물 코스에서도 높은 적응성을 보이며, 70~100%의 성공률로 과제를 완료하였다.

결론적으로, Extreme Parkour 학습 시스템은 단순한 보상 구조와 효율적인 학습 방법을 통해 4족보행 로봇이 다양한 파쿠르 기술을 학습하고 수행할 수 있음을 입증하였다. 이 연구는 로봇 이동성과 자율성 분야의 새로운 가능성을 열며, 복잡한 동작을 효율적으로 학습하고 적용할 수 있는 실용적인 프레임워크를 제공한다.

## 6. 결론

본 논문은 NVIDIA Isaac Gym 플랫폼을 중심으로 4족보행 로봇 제어에서 DRL을 활용한 연구를 분석하고, 사례 연구를 통해 기술적 접근 방식을 종합적으로 검토하였다. NVIDIA Isaac Gym은 병렬 시뮬레이션과 GPU 가속을 통해 DRL 학습의 효율성을 극대화하며, 다양한 환경에서 로봇이 안정적으로 동작할 수 있도록 돕는다. DRL 알고리즘인 PPO와 SAC 등을 활용한 학습 결과는 복잡한 환경에서도 로봇이 성공적으로 동작할 수 있음을 보여준다.

그러나 본 연구는 기존 기술을 분석하고 정리하는데 초점을 맞추었으며, 새로운 알고리즘이나 구조를 제안하지는 않는다. 사례 연구를 통해 기술적 가능성을 확인했지만, DRL 기반 로봇 제어에는 여전히 몇 가지 한계가 존재한다. 예를 들어, Sim-to-Real Gap 완화, 학습 시간 단축, 더 복잡한 환경에서의 적응력 향상 등은 추가 연구가 필요한 분야이다.

결론적으로, 본 논문은 4족보행 로봇 제어에서 DRL의 현재 상태를 이해하고, NVIDIA Isaac Gym의 잠재력을 평가하며, 향후 연구 방향을 제시하는 데 중점을 두었다. 우리는 본 논문이 DRL 기반 로봇 제어 연구를 더욱 발전시키기 위한 기초 자료로 활용될 수 있을 것으로 기대한다.

## 참고 문헌

- [1] NVIDIA Isaac Gym Benchmark Environments, <https://github.com/isaac-sim/IsaacGymEnvs>
- [2] NVIDIA Isaac Gym Tutorial, <https://forums.developer.nvidia.com/t/isaac-gym-tutorials/255290>
- [3] B. Katz, J. D. Carlo, and S. Kim, "Mini Cheetah: A Platform for Pushing The Limits of Dynamic Quadruped Control," in 2019 International Conference on Robotics and Automation (ICRA), pp. 6295-6301, 2019.

- [4] N. Kau, A. Schultz, N. Ferrante, and P. Slade, "Stanford Doggo: An Open-source, Quasi-direct-drive Quadruped," in 2019 International Conference on Robotics and Automation (ICRA), pp. 6309–6315, 2019.
- [5] E. Guizzo, "Boston Dynamics' Spot Robot Dog Goes on Sale," IEEE Spectrum, 2019.
- [6] Y. Fan, Z. Pei, C. Wang, M. Li, Z. Tang, and Q. Liu, "A Review of Quadruped Robots: Structure, Control, and Autonomous Motion," Advanced Intelligent Systems, March 2024.
- [7] A. Kumar, Z. Fu, D. Pathak, and J. Malik, "RMA: Rapid Motor Adaptation for Legged Robots," Robotics: Science and Systems (RSS), 2021
- [8] Nikita Rudin, David Hoeller, Philipp Reist, Marco Hutter, "Learning to Walk in Minutes Using Massively Parallel Deep Reinforcement Learning," The Conference on Robot Learning (CoRL), 2021.
- [9] Zhuang, Ziwen, et al. "Robot Parkour Learning," The Conference on Robot Learning (CoRL), 2023.
- [10] X. Cheng, K. Shi, A. Agarwal, and D Pathak, "Extreme Parkour with Legged Robots," IEEE International Conference on Robotics and Automation (ICRA), 2024.