2021

자율주행 객체의 강화학습 수행체계 연구

Research: DRL for Autonomous Object

GNTP 경남테크노파크
개방형혁신네트워크
R&D 지원 사업

R&D 기초연구 수행
동준상.넥스트플랫폼
naebon1@gmail.com

v210625

자율주행 객체의 강화학습 수행체계 연구

Research: DRL for Autonomous Object

GNTP 경남테크노파크
개방형혁신네트워크
R&D 지원 사업

R&D 기초연구 수행
동준상.넥스트플랫폼
naebon1@gmail.com

v210625

자율주행 객체의 강화학습 수행체계 연구

심층강화학습 실험사례 분석 (FOI)

DRL 심층강화학습 개요

- DRL, Deep Reinforcement Learning
- 기존 머신러닝 기법과의 차이점: 명시적인 특성 추출(feature extraction) 없이 에이전트의 환경 적응 반복 시행으로 학습
- 연구 사례: 스웨덴국방연구소(FOI)와 일렉트로닉아츠(EA)의 CGF 강화학습 연구

DRL 심층강화학습의 장점

명시적인 프로그래밍 대비, DRL 심층강화학습 기법의 장점

- **효율성 제고:** 에이전트가 취해야할 동작을 프로그래밍 기법으로 입력할 수 있지만, 에이전트가 고려해야 할 변수의 수가 급증하면서 명령 입력 및 수행 효율성 급감
- **현실성 반영:** 환경 요소에 도메인 전문가의 전문성, 현장의 규칙을 적용하고, 에이전트가 이에 적응하도록 하여 현장에 존재하는 현실감(맥락) 부여 가능
- **복잡성 반영:** 딥러닝과 강화학습 융합 통해 프로그래머 또는 전문가가 단독으로 전달하기 어려운 수준의 복잡성을 시행착오 기법으로 탐색 및 해소해 나감

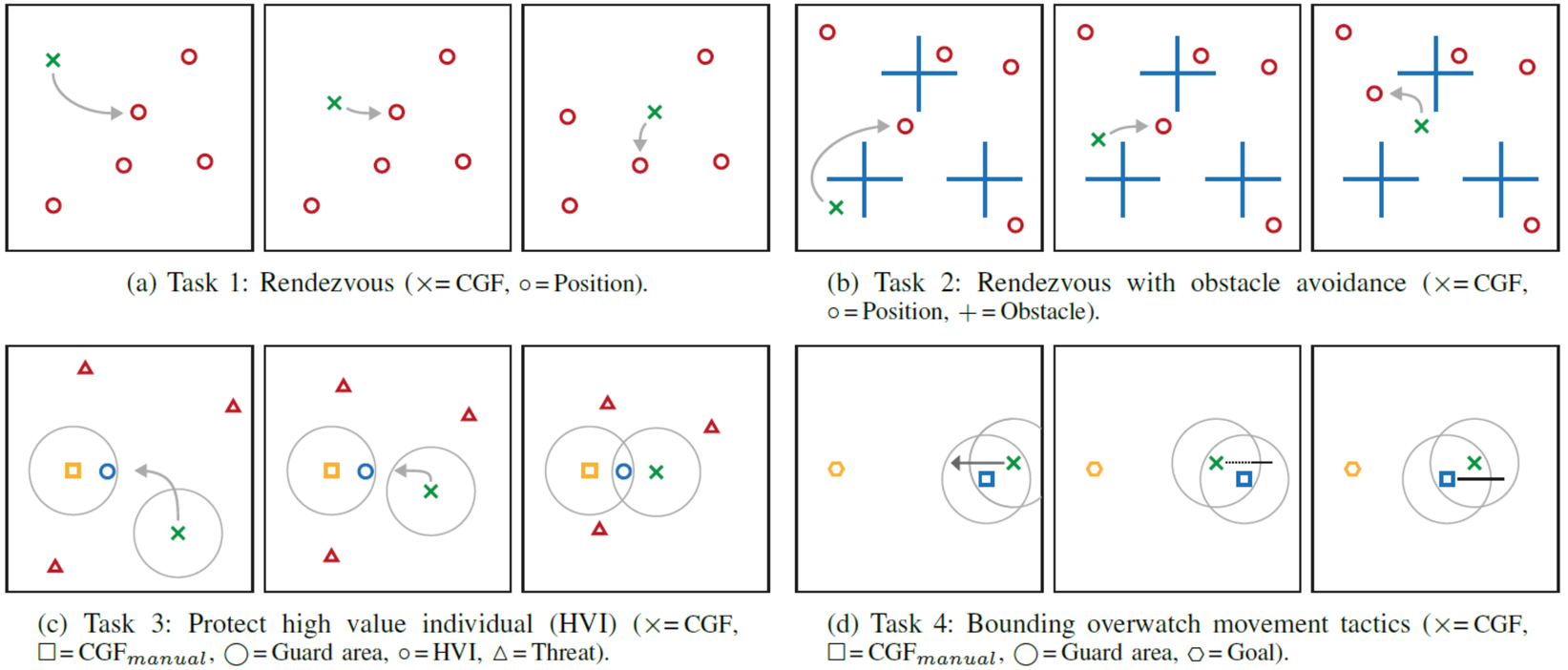
미션 및 DRL 모델

미션: 다양한 DRL 모델을 이용하여 물류 환경에서의 협업, 상호작용, 업무조정 등이 가능하도록 하고, 물류 에이전트의 목표지점 설정, 파렛트 픽업, 목표지점 이동 구현

- **DQL** (Deep Q-Learning)
- **A3C-FF** (Asynchronous Advantage Actor-Critic with Feed Forward)
- **A3C-LSTM** (Asynchronous Advantage Actor-Critic with LSTM)

강화학습의 목표 설정

1. 목표 지점 이동
2. 장애물 랑데뷰
3. 요인 보호 이동
4. 영역 겹침 이동



(a) Task 1: Rendezvous (×= CGF, ○= Position).

(b) Task 2: Rendezvous with obstacle avoidance (×= CGF, ○= Position, + = Obstacle).

(c) Task 3: Protect high value individual (HVI) (×= CGF, □= CGF_{manual}, ◯= Guard area, ○= HVI, △= Threat).

(d) Task 4: Bounding overwatch movement tactics (×= CGF, □= CGF_{manual}, ◯= Guard area, ○= Goal).

Fig. 1: Learning tasks used to evaluate DRL in the context of CGF behaviors for ground combat simulation.

강화학습 실험 환경설정

- 신경망: CNN,
  입력층 25600개 뉴론

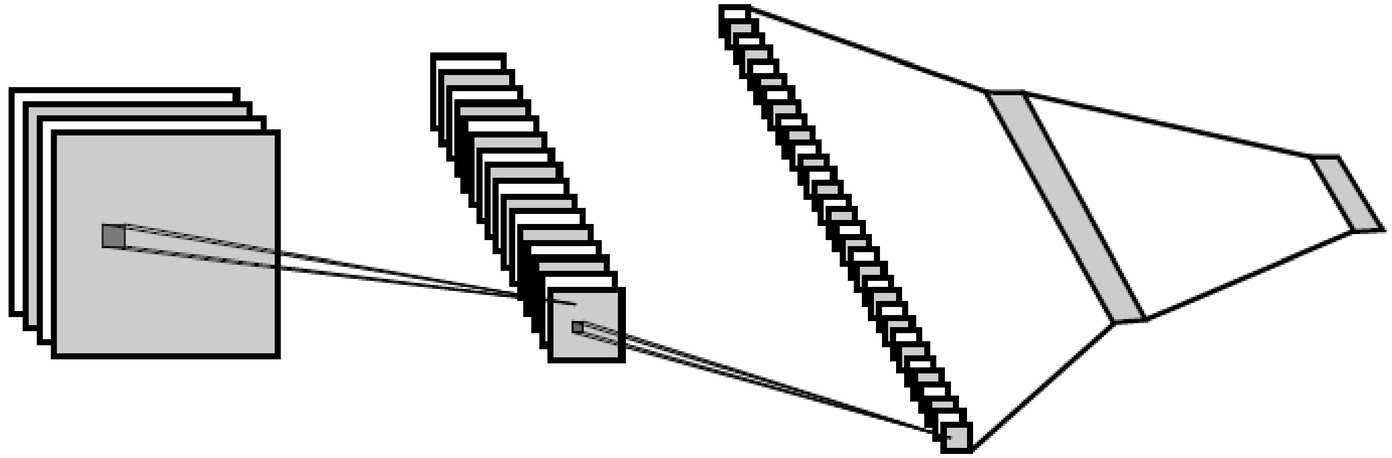


Fig. 2: Network structure used in the DQL and A3C-FF algorithms.

강화학습 실험 환경설정

- 하이퍼파라미터

DQL		
Variable	**Hyperparameter**	**Value**
C	Target network update freq.	10 000
ϵ_{start}	Initial exploration rate	1.0
ϵ_{end}	Final exploration rate	0.1
$t_{annealing}$	Annealing period	1 000 000
$\mathcal{D}_{size}$	Replay memory size	1 000 000

A3C-FF		
Variable	**Hyperparameter**	**Value**
$\mathcal{P}$	Parallel CGFs	16

A3C-LSTM		
Variable	**Hyperparameter**	**Value**
$\mathcal{P}$	Parallel CGFs	16
t_{max}	LSTM sequence length	5

강화학습 실험 환경설정

- 보상함수

1. 랑데뷰

$$r = \begin{cases} 1 & d(O_i, CGF) < d_{min} \\ -1 & \text{moving outside the environment} \\ 0 & \text{otherwise,} \end{cases}$$

2. 장애물 랑데뷰

$$r = \begin{cases} 1 & d(O_i, CGF) < d_{min} \\ -1 & \text{moving outside the environment} \\ 0 & \text{otherwise,} \end{cases}$$

3. 요인 보호 이동

$$r = \begin{cases} r_d & d(threat_i, CGF) < 10 \wedge HVI \text{ is guarded} \\ 0.1 & d(threat_i, CGF) < 10 \wedge HVI \text{ is unguarded} \\ -1 & d(threat_i, HVI) < 5 \\ -1 & \text{moving outside the environment} \\ 0 & \text{otherwise,} \end{cases}$$

where

$$r_d = \frac{d(CGF_{manual}, CGF)}{d_{max}}$$

4. 영역 겹침 이동

$$r = \begin{cases} r_d & \text{if } action = a_{overwatch} \wedge CGF_{manual} \text{ is guarded} \\ 1 & \text{if } d(goal, CGF_{manual}) < 10 \\ 0 & \text{otherwise,} \end{cases}$$

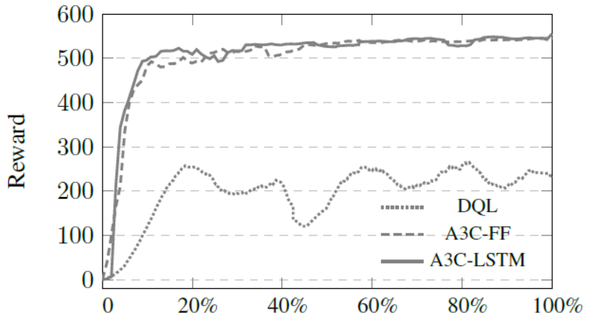
where

$$r_d = \left(\frac{d_{adv}}{d_{max}}\right)^2$$

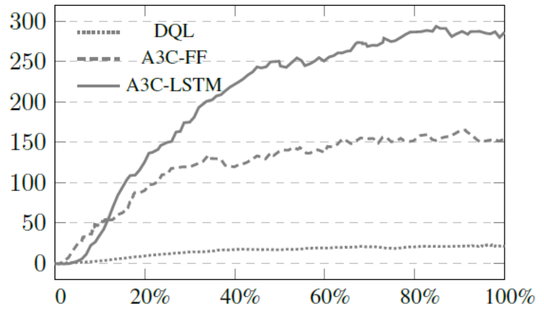
강화학습 실험결과 요약: 미션별 비교

성과의 측정 방식:
단위시간동안 보상점수의 누적액

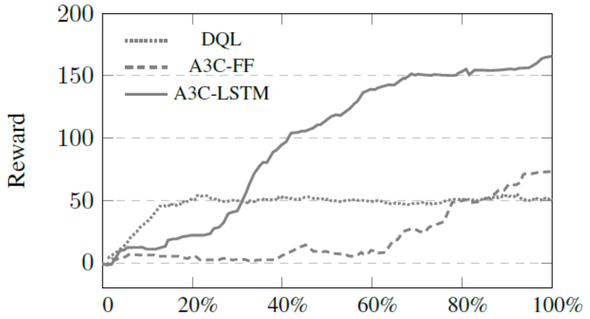
- 랑데뷰: A3C 우수
- 장애물 랑데뷰: A3C-LSTM 우수
- 작업자 보호 이동: A3C-LSTM 우수
- 영역 겹침 이동: A3C 우수



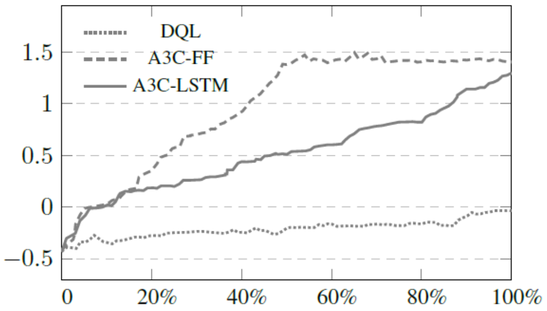
(a) Experiment 1: Rendezvous



(b) Experiment 2: Rendezvous with obstacle avoidance



(c) Experiment 3: Protect HVI

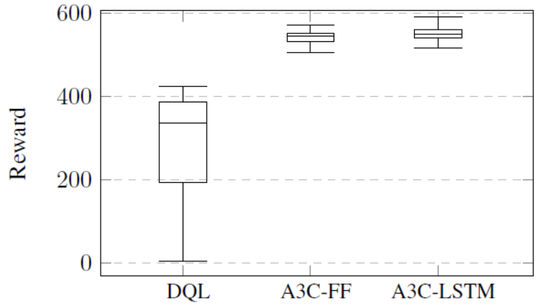


(d) Experiment 4: Bounding overwatch

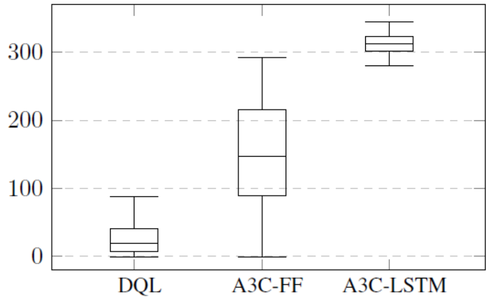
강화학습 실험결과 요약: 모델별 비교

학습모델의 일반화 수준 평가

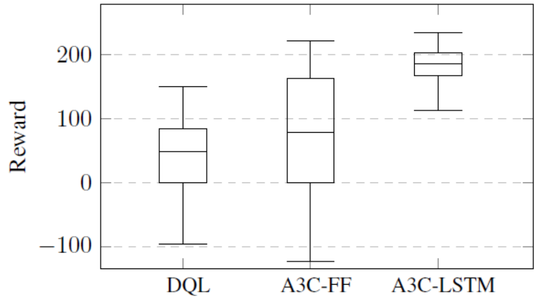
- DQL: 전반적으로 저조
- A3C-FF: A3C의 베이스라인
- A3C-LSTM: 전반적으로 우수



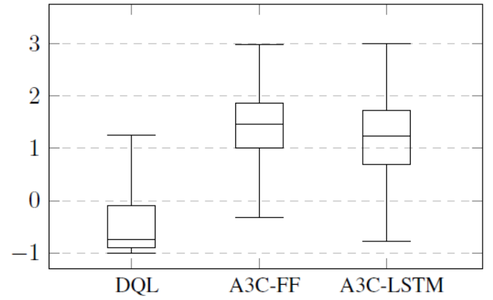
(a) Experiment 1: Rendezvous



(b) Experiment 2: Rendezvous with obstacle avoidance



(c) Experiment 3: Protect HVI



(d) Experiment 4: Bounding overwatch

강화학습 실험 환경설정: 물류 에이전트에 적용

- 강화학습 기반 최적 이동방식 습득, 최적 경로 탐색
- 유니티 ML Agent 활용, 시뮬레이션 방식으로 학습 진행

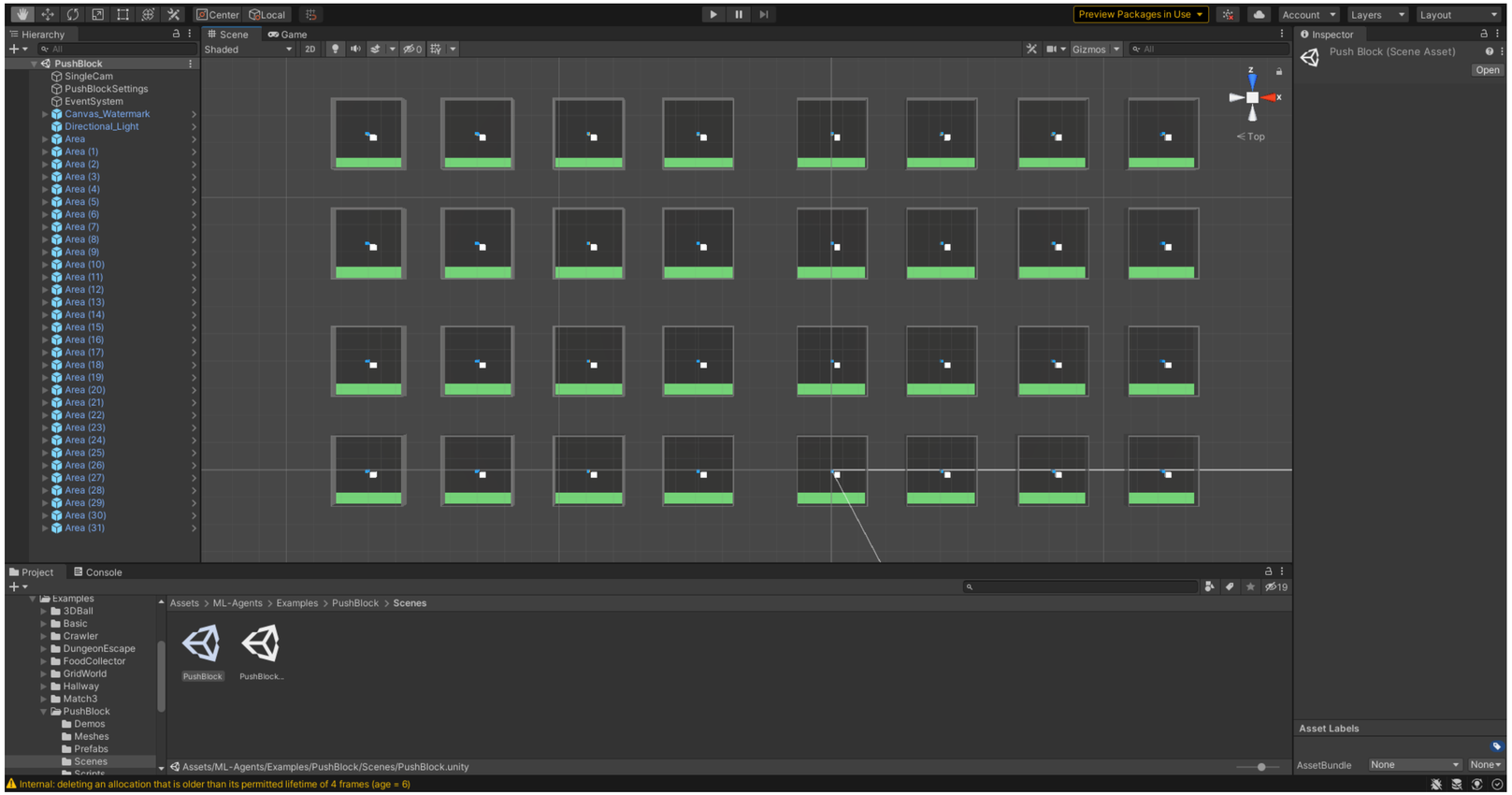
- **가상 운송 기기의 학습 결과를 실제 운송 기기에 적용할 수 있는 방법 연구**
- 1. 가상 운송 기기의 학습 구조 저장 (NN), 다른 기기에 이식 〉〉〉 금번 연구개발 목표
- 2. 가상 운송 기기의 구동부, 전환부 데이터 출력, 수집 필요 〉〉〉 차기 연구개발 목표


DARPA
SUBTERRANEAN CHALLENGE
EXPLORER
CNH
EPSON
Honeywell
DARPA

자율주행 객체의 강화학습 수행체계 연구

Unity ML Agent 환경구성 및 시뮬레이션 체계

유니티 ML Agent 기반 강화학습 수행 체계

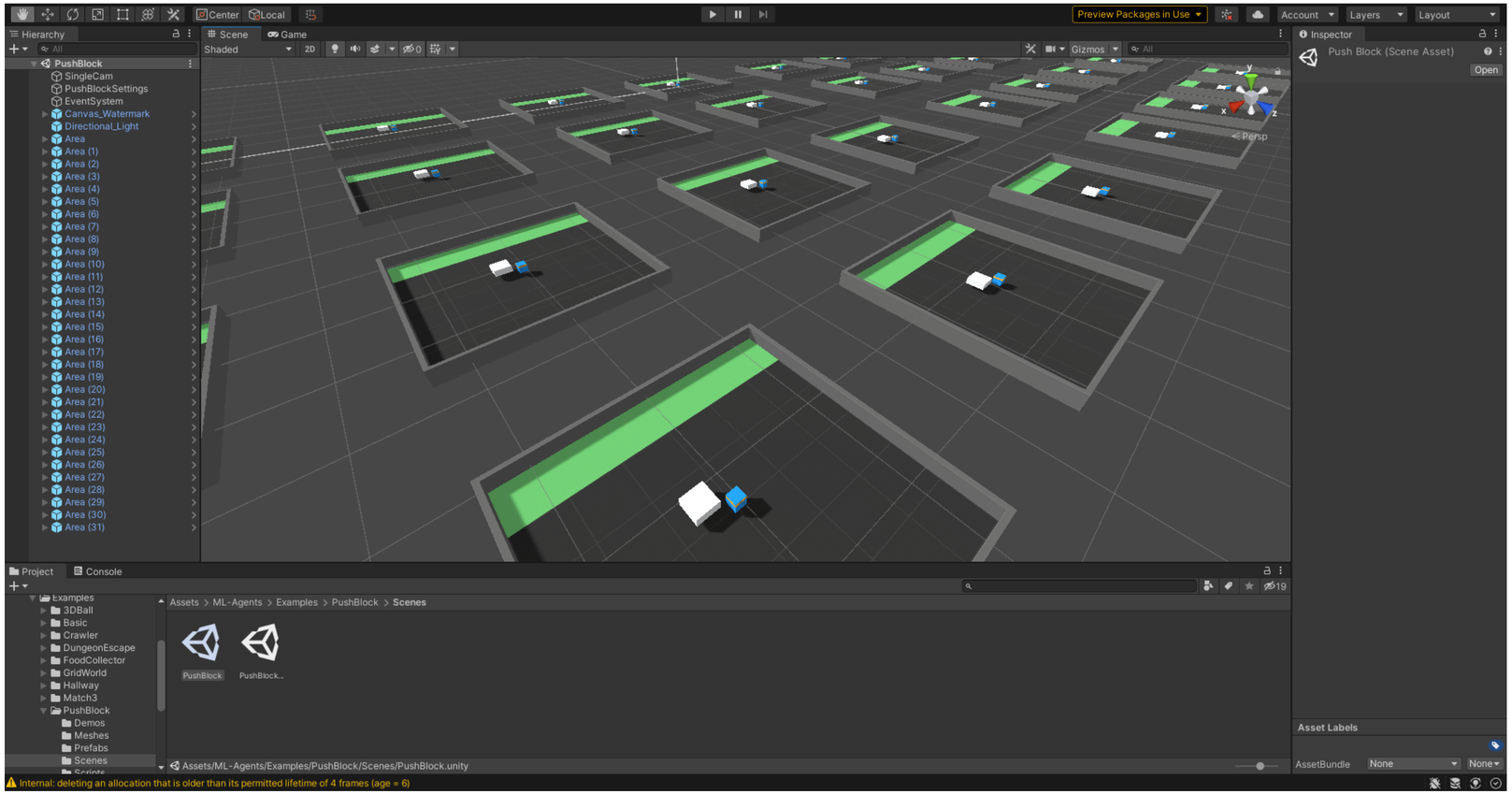


- 게임엔진이자 물리엔진인 유니티 기반 강화학습 수행
- 유니티에서 배포한 ML Agent 패키지 설치 필요

유니티 ML Agent 기반 강화학습 수행 체계

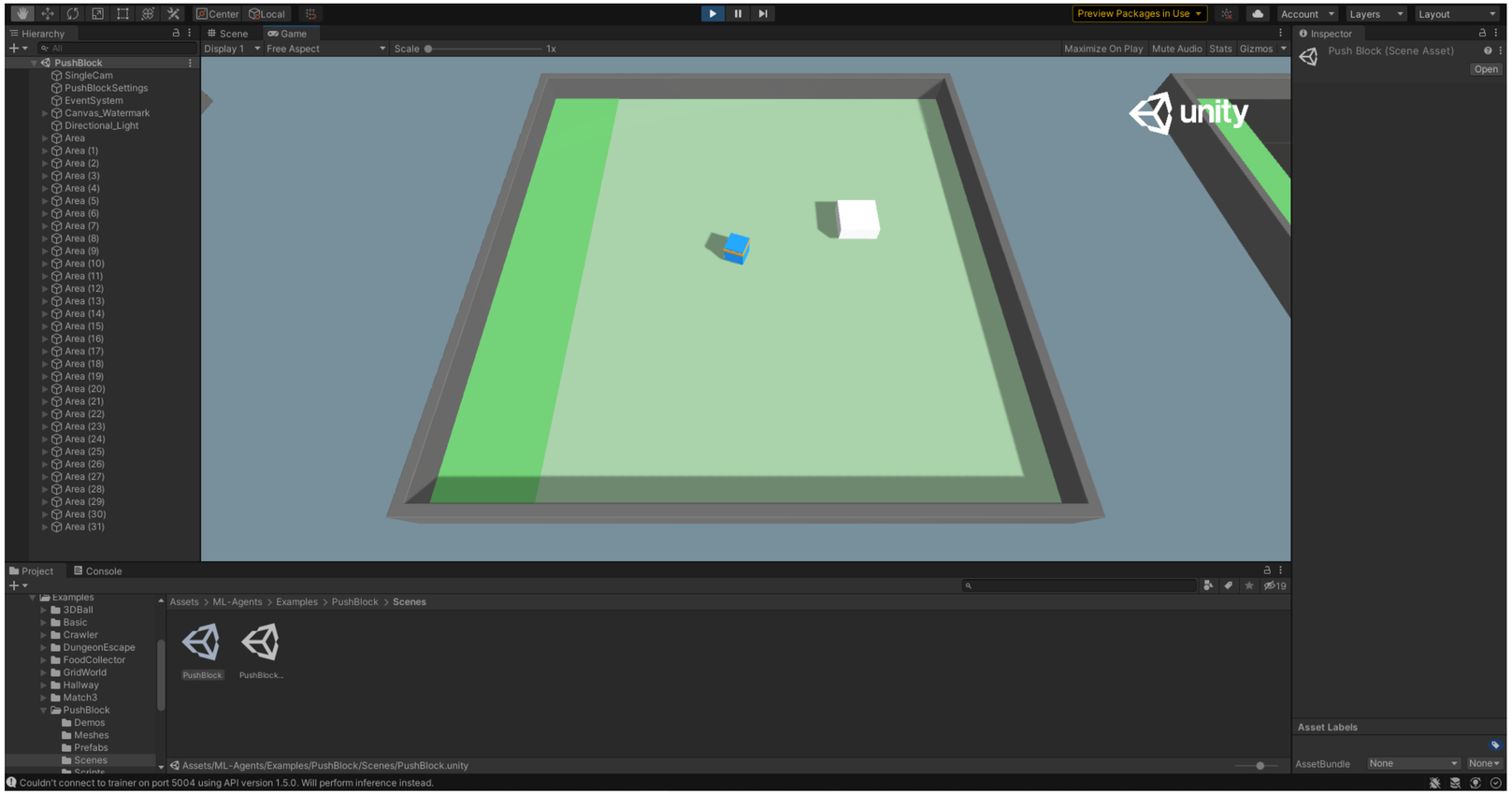
- ML Agent 내에서 강화학습 환경 선택 가능
- Push Block 프로젝트 선택

유니티 ML Agent 기반 강화학습 수행 체계



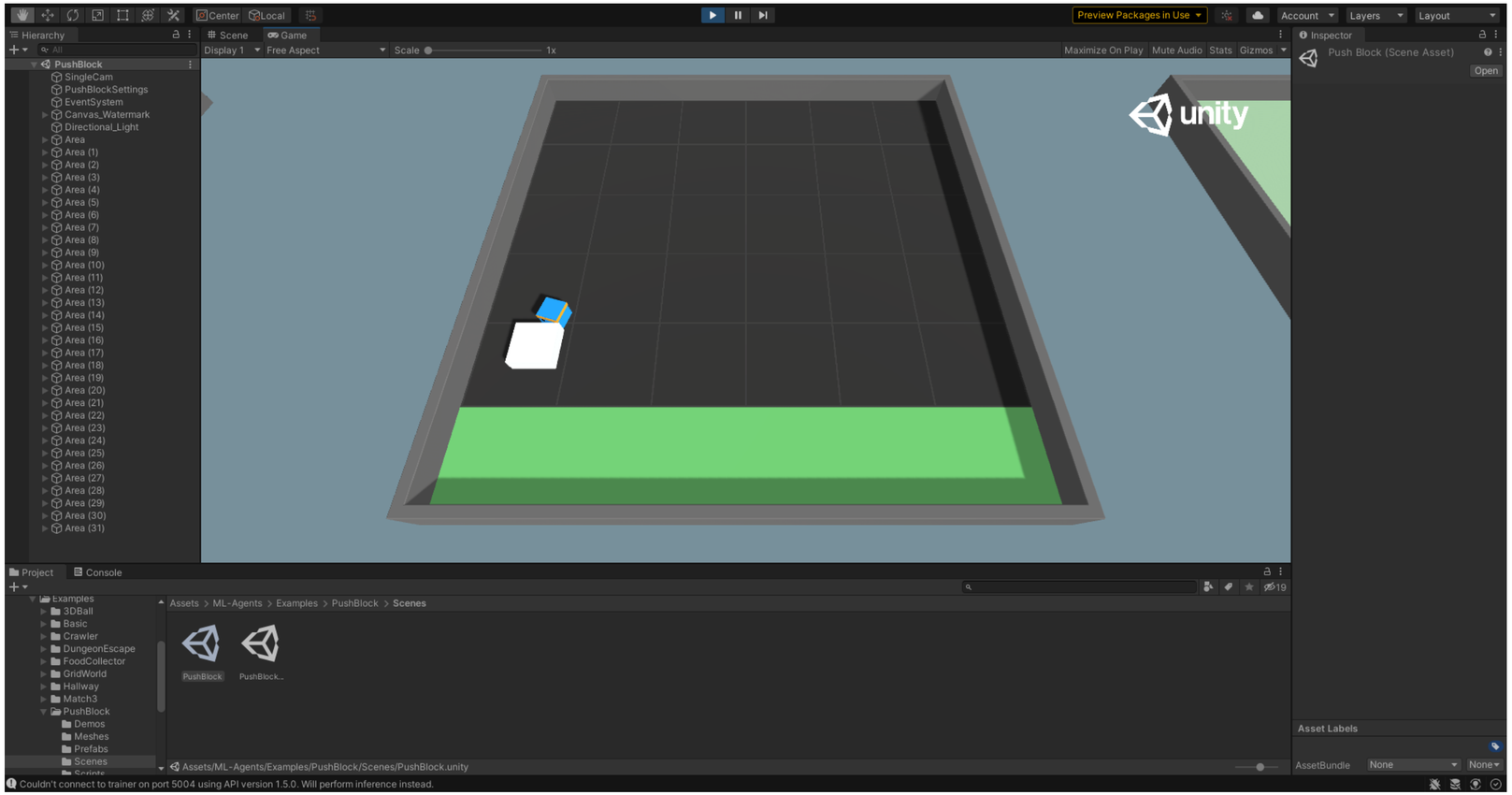
- RL 시뮬레이션은 최소 수만 회에서 보통 수백만 회 반복 시행됨
- 경험한 제약 조건을 반영하여 최소한의 리소스를 지닌 다수의 에이전트를 학습시킴

유니티 ML Agent 기반 강화학습 수행 체계



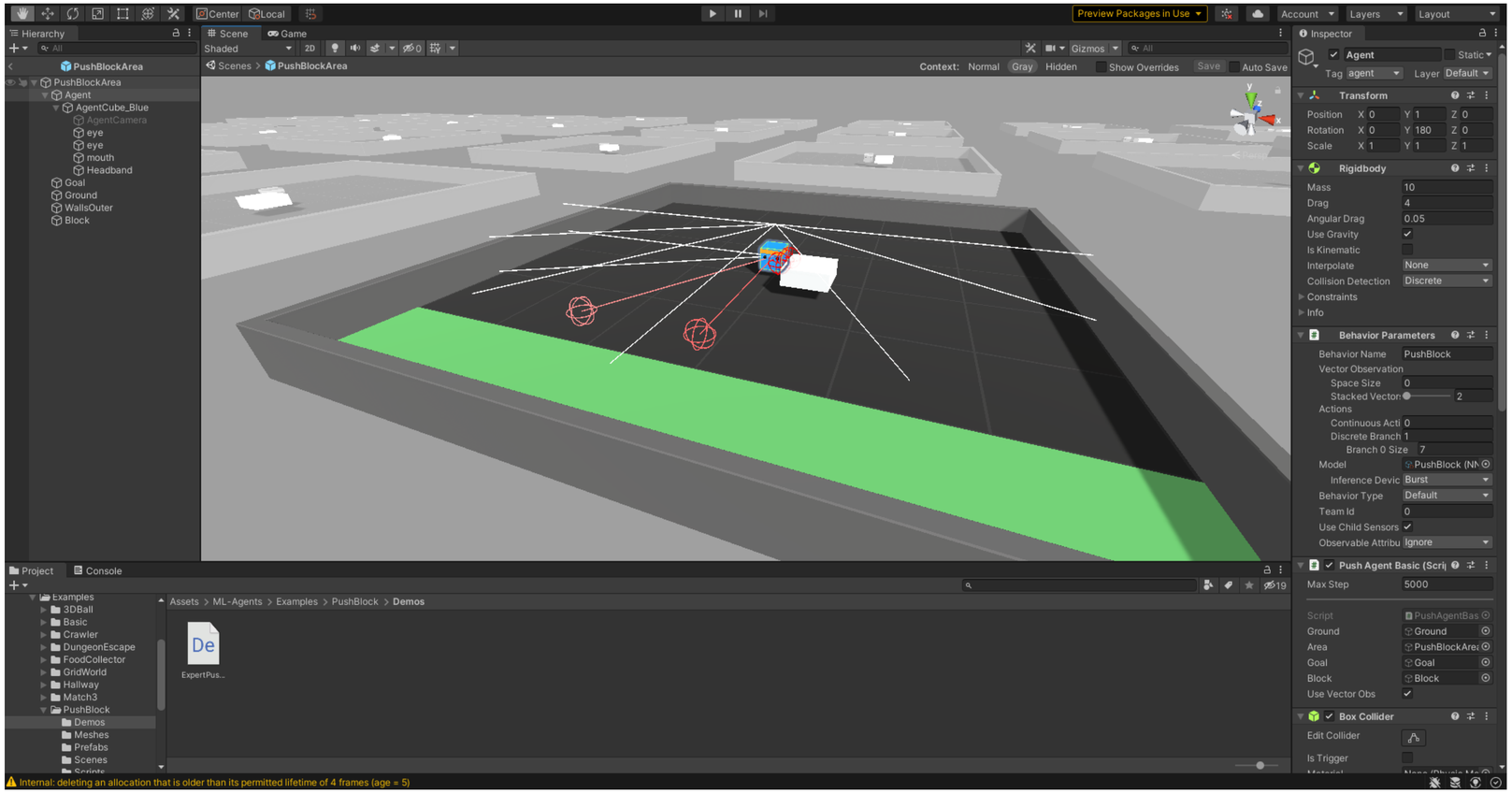
- 강화학습을 거친 에이전트의 영역 배치 작업 시뮬레이션 화면
- 흰색 박스: 배송 객체
- 파란색 박스: 물류 운송 객체
- 연두색 영역: 배송 목적지

유니티 ML Agent 기반 강화학습 수행 체계



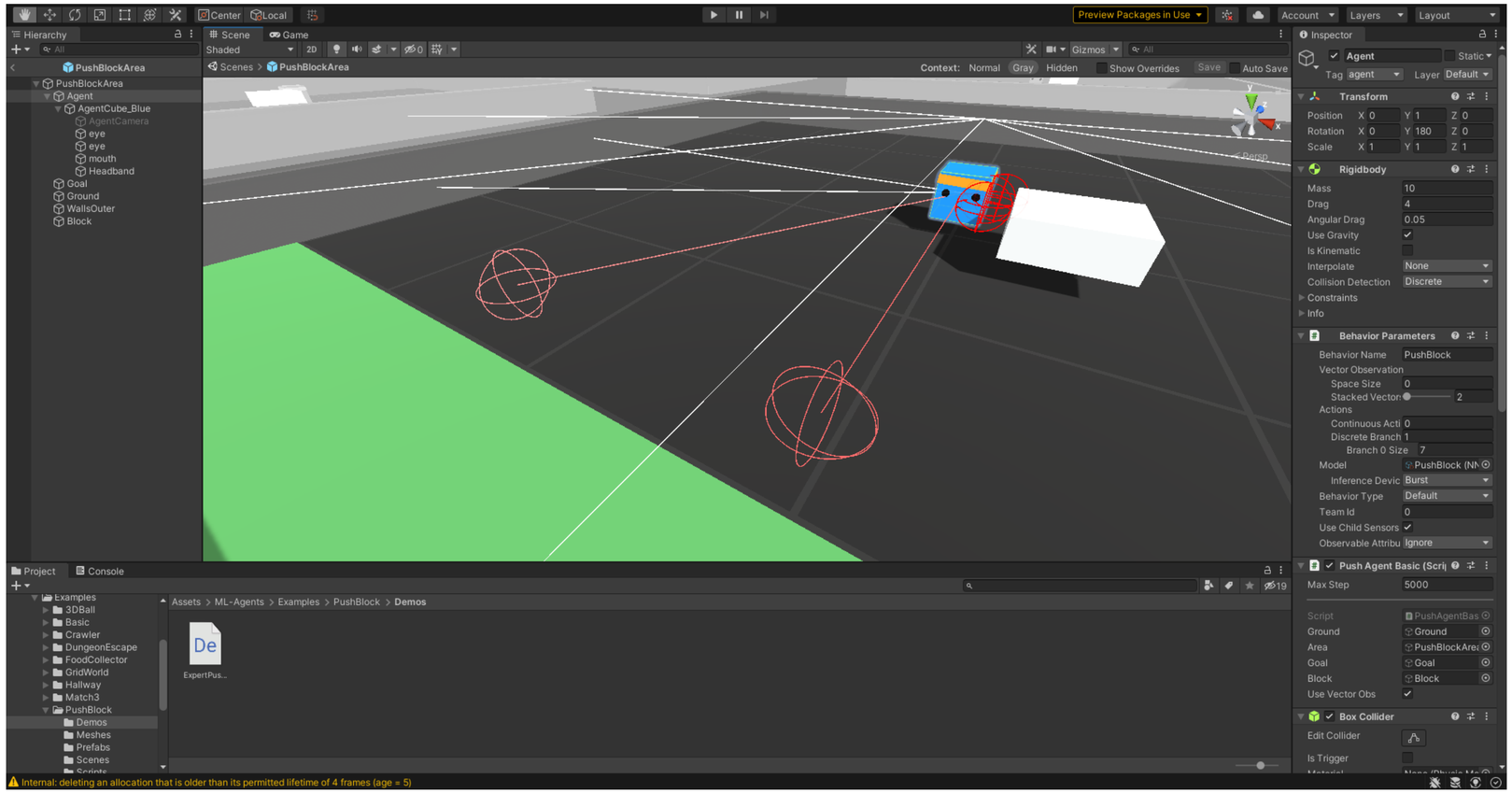
- 금번 강화학습에서 에이전트는 목표 영역에 배송 객체를 밀어넣었을 때 보상을 획득

유니티 ML Agent 기반 강화학습 수행 체계

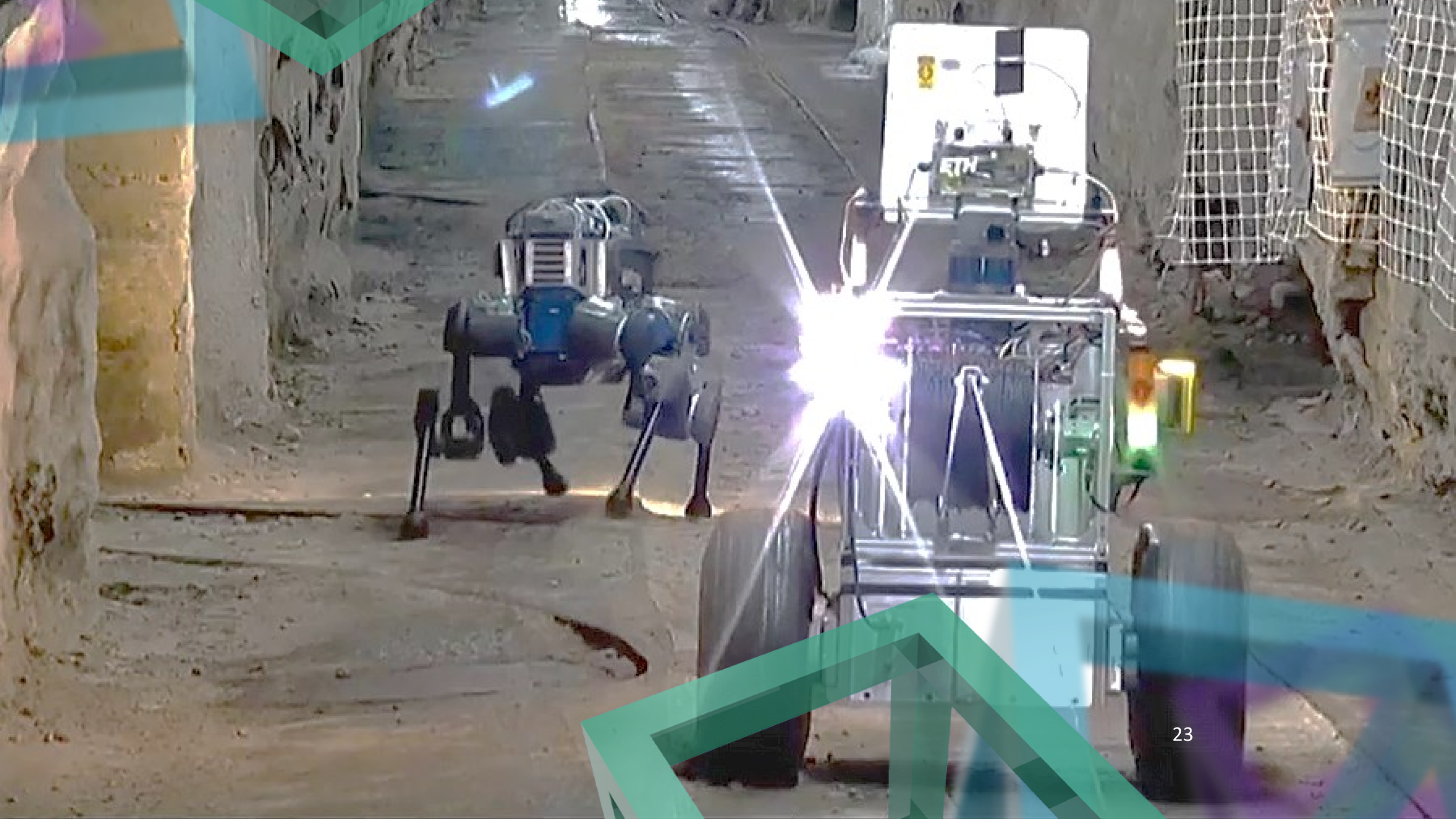


- 에이전트에 탑재된 센서 및 카메라의 구성 확인
- 2개의 카메라와 10개의 센서를 통해 외부 환경 데이터 수집, 반응

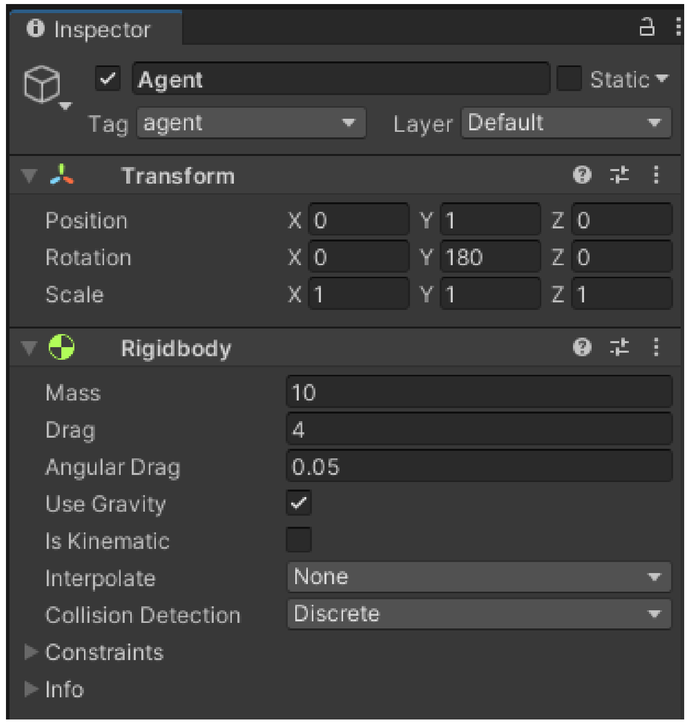
유니티 ML Agent 기반 강화학습 수행 체계



- 강화학습의 결과물: 학습된 구조 (Learned-Structure)
- 다른 가상 객체 또는 실제 운송 장비에 이식 및 탑재하는 것을 목표

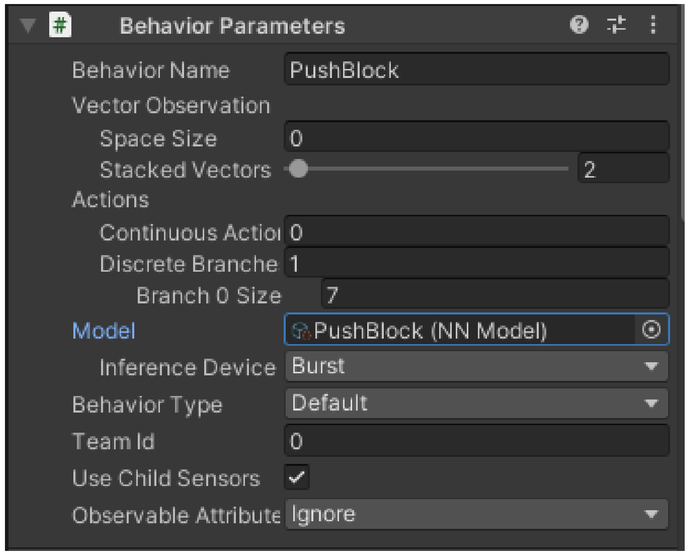


유니티 ML Agent 기반 강화학습 수행 체계



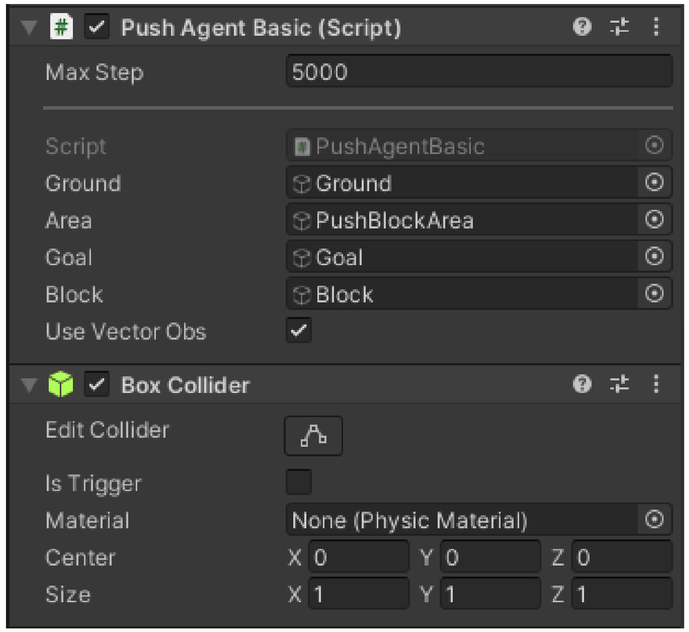
- ML Agent의 반복시행 작업에서 도출되는 정보의 종류
- 전후, 좌우, 상하 등 배송 작업과 관련된 데이터 생성
- 회전값은 물론, 수학적 변환 작업을 통해 배송 에이전트의 속도, 바퀴당 회전수 등도 측정 가능

유니티 ML Agent 기반 강화학습 수행 체계



- 사전 훈련된 강화학습 경험치는 행동 파라미터를 통해 입력 가능
- 이는 훈련한 에이전트가 아닌, 다른 에이전트로 경험치를 전달할 수 있음을 의미
- 전달 호환성은 유니티 및 파이토치 커뮤니티의 프리트레인드 모델 저장방식인 NN Model 사용

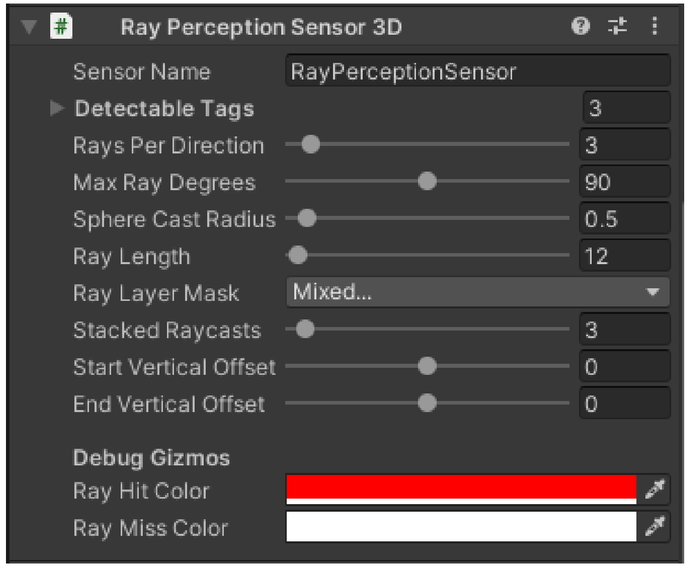
유니티 ML Agent 기반 강화학습 수행 체계



- 강화학습을 현실/가상 공간에서 실험하기 어려운 이유
- 1. 수백만 회의 반복 시행이 가능한 에이전트 설계 난이
- 2. 에이전트가 상호작용할 수 있는 환경 구성 난이

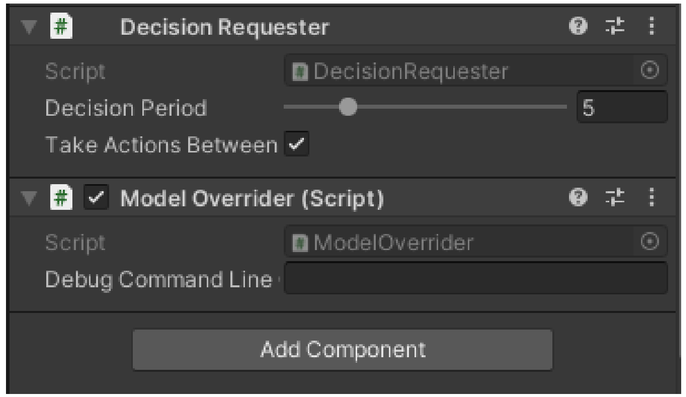
- 유니티 ML Agent는 디지털 트윈스 구현을 위한 환경 및 에이전트를 제공한다는 데 의의가 있음

유니티 ML Agent 기반 강화학습 수행 체계



- 자율주행 에이전트에 장착할 카메라 및 센서 설정
- 기본 장비 외 다양한 감지 장치를 추가 가능
- 학습과정에서 다양한 데이터 수집 가능

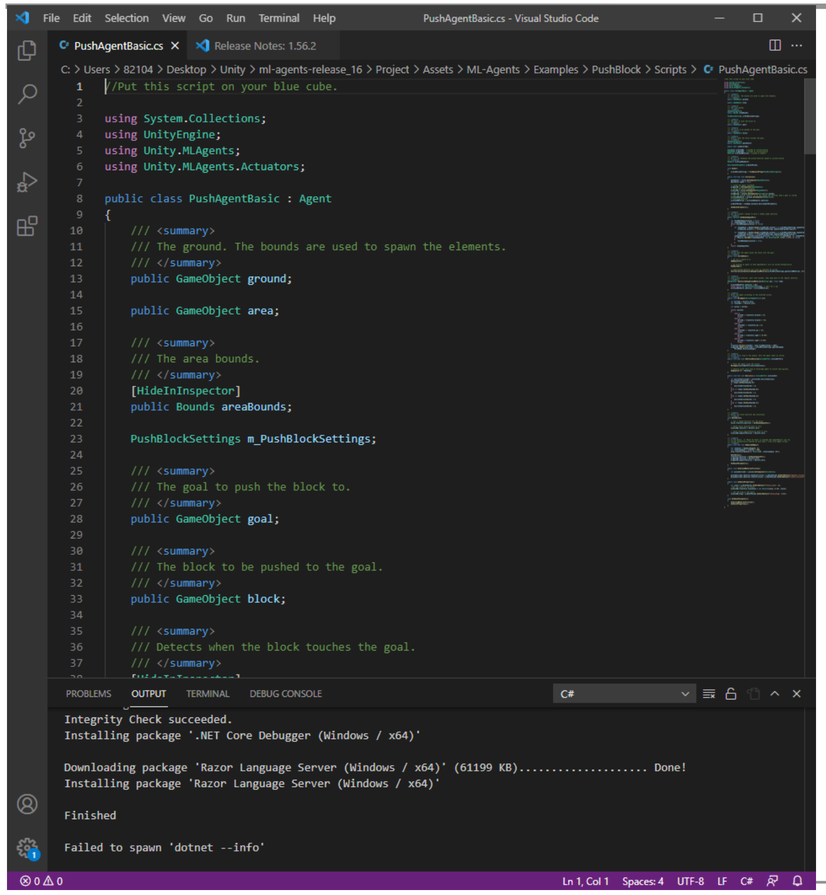
유니티 ML Agent 기반 강화학습 수행 체계



- 강화학습의 또 다른 어려운 영역인 의사결정 체계의 전환
- 다수의 동작이 연결된 에이전트의 경우, 어떤 상황에서 어떤 동작을 할 것인지 결정할 수 있어야 함

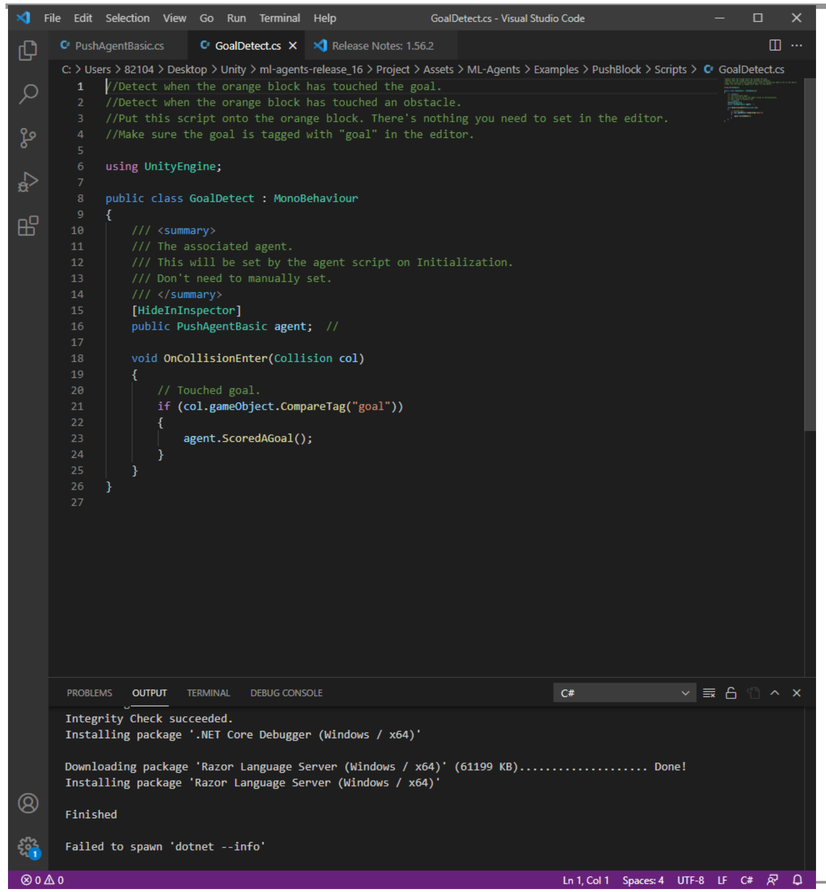
- ML Agent에서는 Decision Requester로 구현

유니티 ML Agent 기반 강화학습 수행 체계



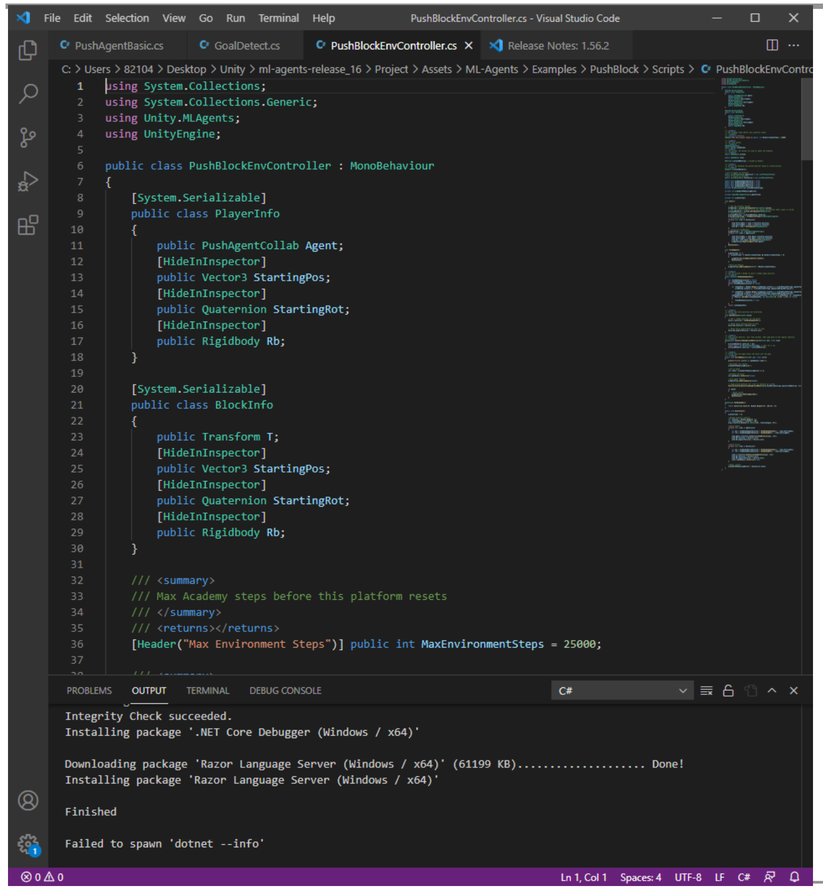
- 에이전트의 동작 및 보상과 처벌을 정의한 PushAgent 코드

유니티 ML Agent 기반 강화학습 수행 체계

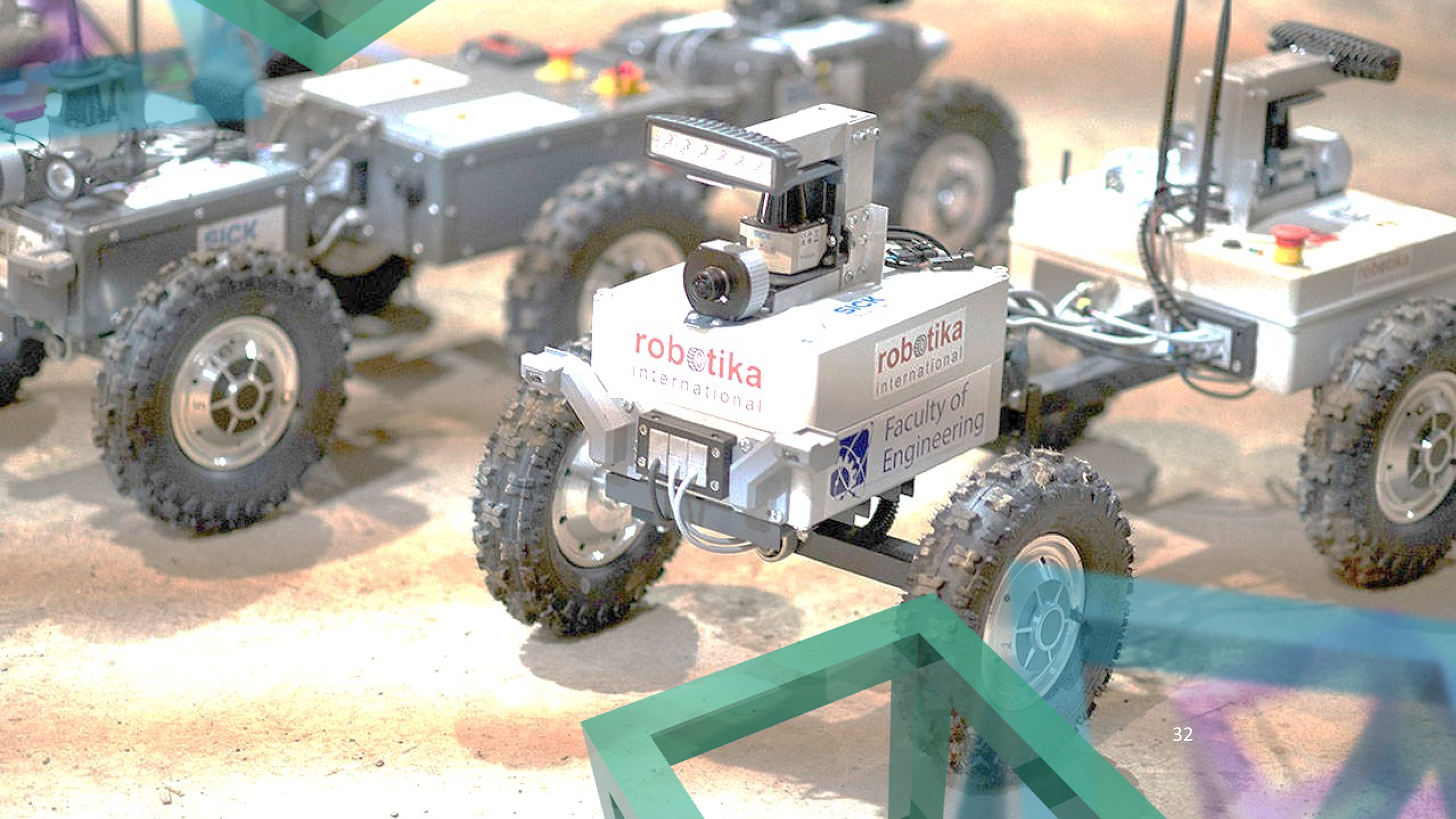


- 환경 요소인 타겟 영역을 찾기 위한 GoalDetect 소스 코드

유니티 ML Agent 기반 강화학습 수행 체계



- 환경과 에이전트의 상호작용 및 강화학습의 전반적인 체계를 정의한 PushBlockEnvController 소스 코드


SICK
robotika
international
robotika
international
Faculty of
Engineering

자율주행 객체의 강화학습 수행체계 연구

참고자료

Reference | 논문 및 보고서

- Varila, M.; Seppanen, M.; Suomala, P. (2007). Detailed cost modelling: a case study in warehouse logistics
- Dukic, G.; Oluic, C. (2007). Order-picking methods: improving order-picking efficiency
- Giaglis, G. M.; Minis, I.; Tatarakis, A.; Zeimpekis, V. (2004). Minimizing logistics risk through real-time vehicle routing
- Erkan, T. E.; Can, G. F. (2014). Selecting the best warehouse data collecting system by using AHP and FAHP methods

Reference | 논문 및 보고서

- Laporte, G., 1992, “The vehicle routing problem: An overview of exact and approximate algorithms”
- Larsen, A., 2000, “The dynamic Vehicle Routing Problem”
- Möhring, R. H., et al., 2008, “Dynamic Routing of Automated Guided Vehicles in Real-time”
- Psaraftis, H.N., 1988 “Dynamic vehicle routing problems”

Reference | 논문 및 보고서

- Baglivo, L., Biasi, N., Biral, F. : Autonomous pallet localization and picking for industrial forklifts: a robust range and look method. (2011)
- Brust, C.A., Sickert, S., Simon, M., Rodner, E., Denzler, J.: Convolutional patch networks with spatial prior for road detection. (2015)
- Chen, G., Peng, R., Wang, Z., Zhao, W.: Pallet recognition and localization method for vision guided forklift. (2012)
- Cucchiara, R., Piccardi, M., Prati, A.: Focus based feature extraction for pallets recognition. (2000)

자율주행 객체의 강화학습 수행체계 연구

Research: DRL for Autonomous Object

GNTP 경남테크노파크 개방형혁신네트워크 R&D 지원 사업

끝 감사합니다.

R&D 기초연구 수행
동준상.넥스트플랫폼
naebon1@gmail.com

v210625