연구인력 현장지원 실적보고서(상세)

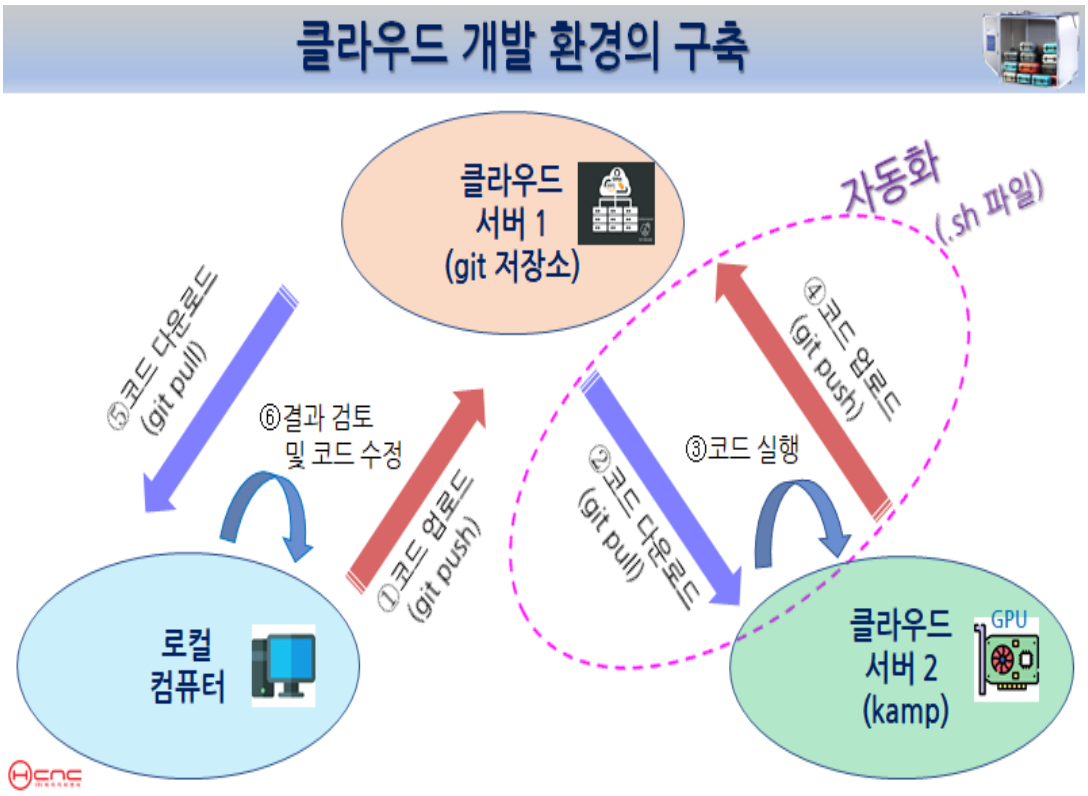
지원 내용 및 결과

1.수행업무 1: Git을 활용한 로컬컴퓨터 - 클라우드서버1(Github) - 클라우드 서버2(Kamp) 간의 클라우드 개발환경 구축

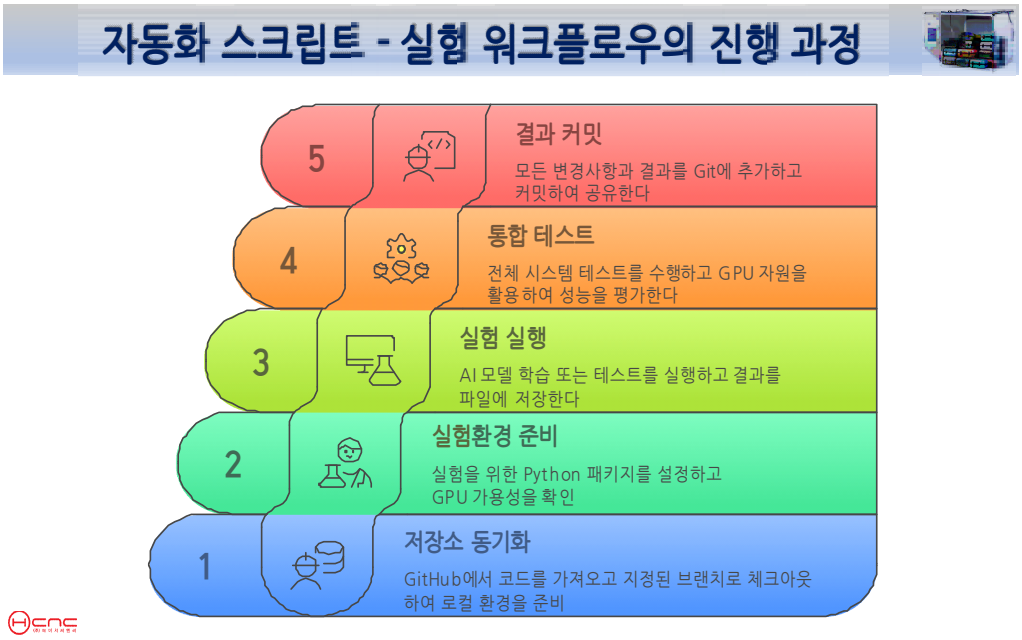
- 로컬 컴퓨터 - 클라우드 서버1(Github) 간의 git 동작 구축
- 클라우드 서버1(Github) - 클라우드 서버2(Kamp) 간의 git 동작 구축
- 클라우드 서버 측의 시뮬레이션 수행절차를 자동화하는 쉘 스크립트 구현
- 로컬 컴퓨터 - 클라우드 서버1(Github) - 클라우드 서버2(Kamp)에 걸친 개발 전주기 워크플로우의 구축

□ **결과:**

- 클라우드 개발환경의 동작을 설명하는 개념도



- 클라우드 개발환경에서 자동화된 실험 워크플로우의 진행 과정

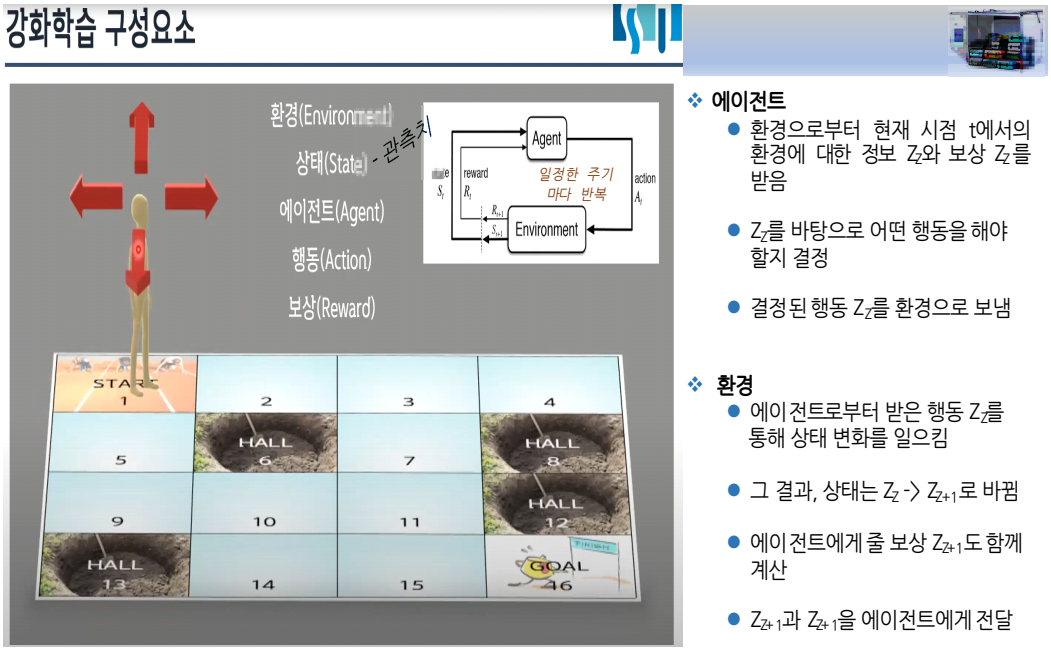


2. 수행업무 2:

ULD 적재 최적화를 위한 AI 솔루션에 사용되는 강화학습의 액터-크리틱 /PPO 알고리듬의 동작을 스터디/분석 완료

□ **결과:**

- 강화학습의 구성요소

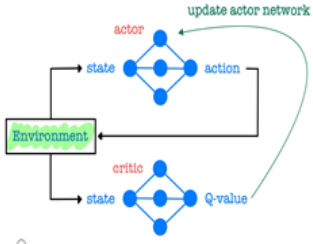


● 액터-크리틱 방법의 개요

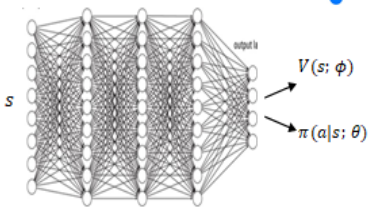
액터-평론가 방법

액터-평론가 방법은 두 개의 네트웍으로 구성 :

- 평론가 네트웍은 가치함수 $V(s;\phi)$ or $Q(s,a;\phi)$ 에 대한 파라미터 ϕ 를 업데이트함
  액터 네트웍은 정책 그래디엔트 $\nabla_\theta J(\theta)$를 사용하여 정책 $\pi_\theta(a|s)$ 에 대한 파라미터 θ를 업데이트함



- 정책과 더불어 추가로 가치함수를 학습하는 것은 분산을 줄이기 위해 기준선 강화에서처럼 가치함수를 알면 정책 업데이트에 도움이 될 수 있으므로 유용함



- 실제로는 평론가와 액터가 네트웍을 공유하되, 각자의 가중치 파라미터 ϕ 와 θ 를 사용함

● Proximal Policy Optimization(PPO)의 개념

Proximal Policy Optimization (PPO) - 개념

- 정책 업데이트에서 너무 큰 단계를 밟으면?

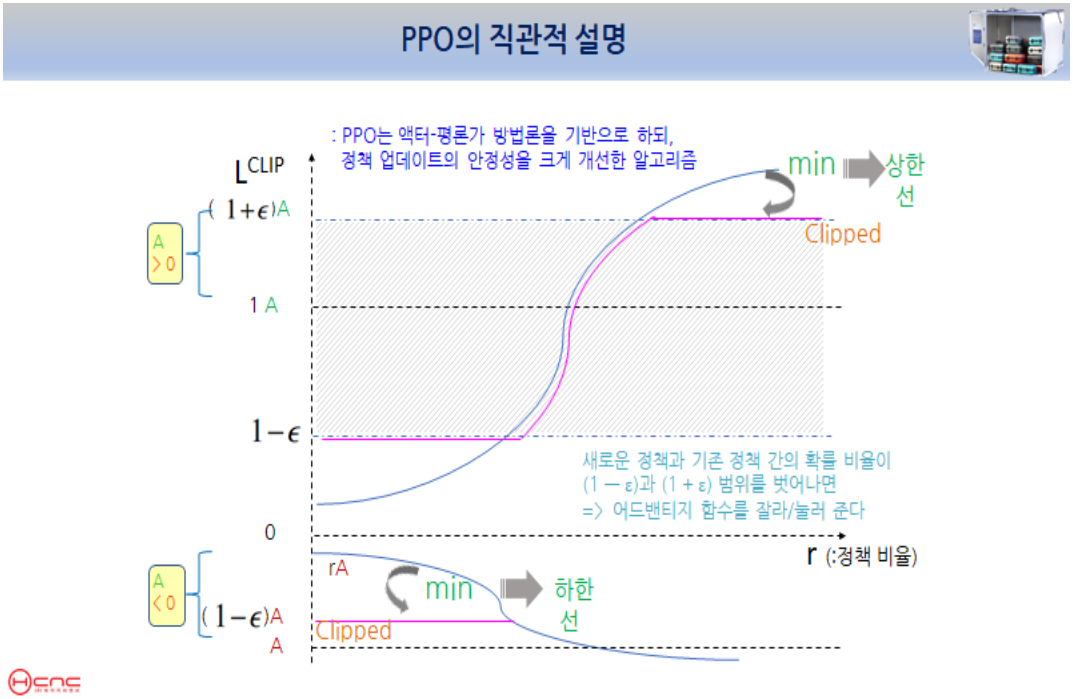
"절벽에서 떨어지는"(나쁜 정책을 얻는) 결과를 초래하여

회복하는 데 오랜 시간이 걸리거나 전혀 불가능할 수도 ..

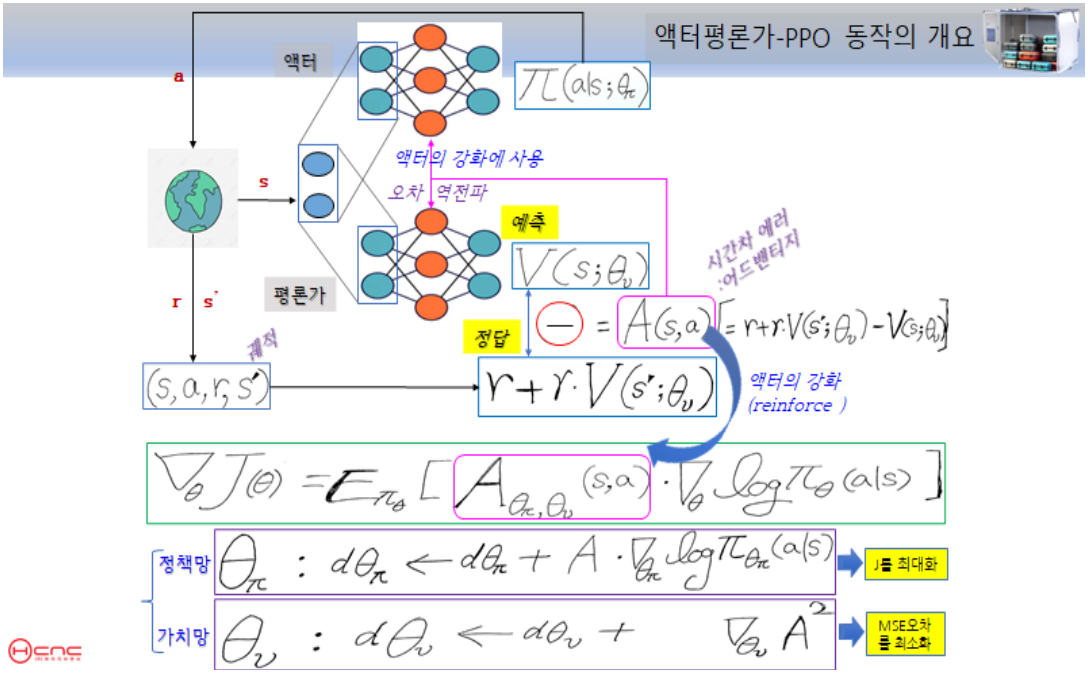
*두 가지 주요 **최적화** 방법*
- 경사 하강법과 같은 라인 써치
신뢰 영역 방법

- 각 학습 에포크에서 정책이 변경되는 정도를 제한하여 정책의 학습 안정성을 개선하자
  (정책 업데이트가 너무 커지는 것을 방지하고자 함)
- 현재 정책이 이전 정책과 비교하여 얼마나 바뀌었는지
  현재 정책과 이전 정책 간의 비율을 계산하여 측정
- 구조는 고전적인 액터-평론가 프레임워크를 따름

- PPO의 직관적 설명



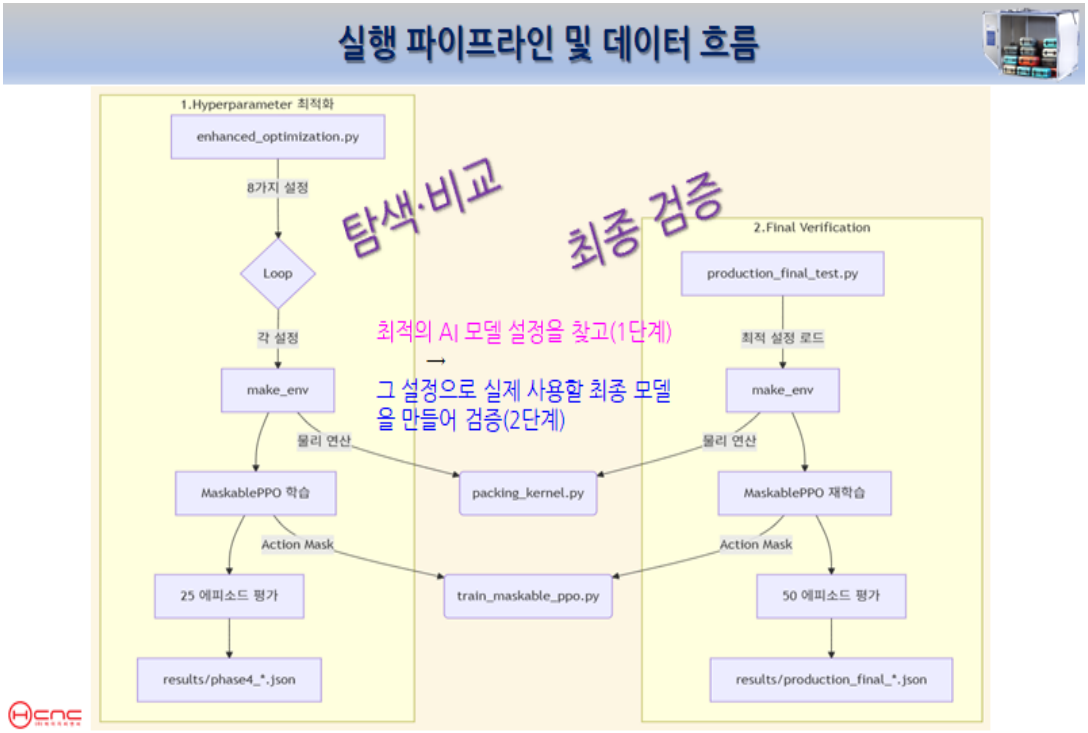
- 액터크리틱-PPO 동작의 개요



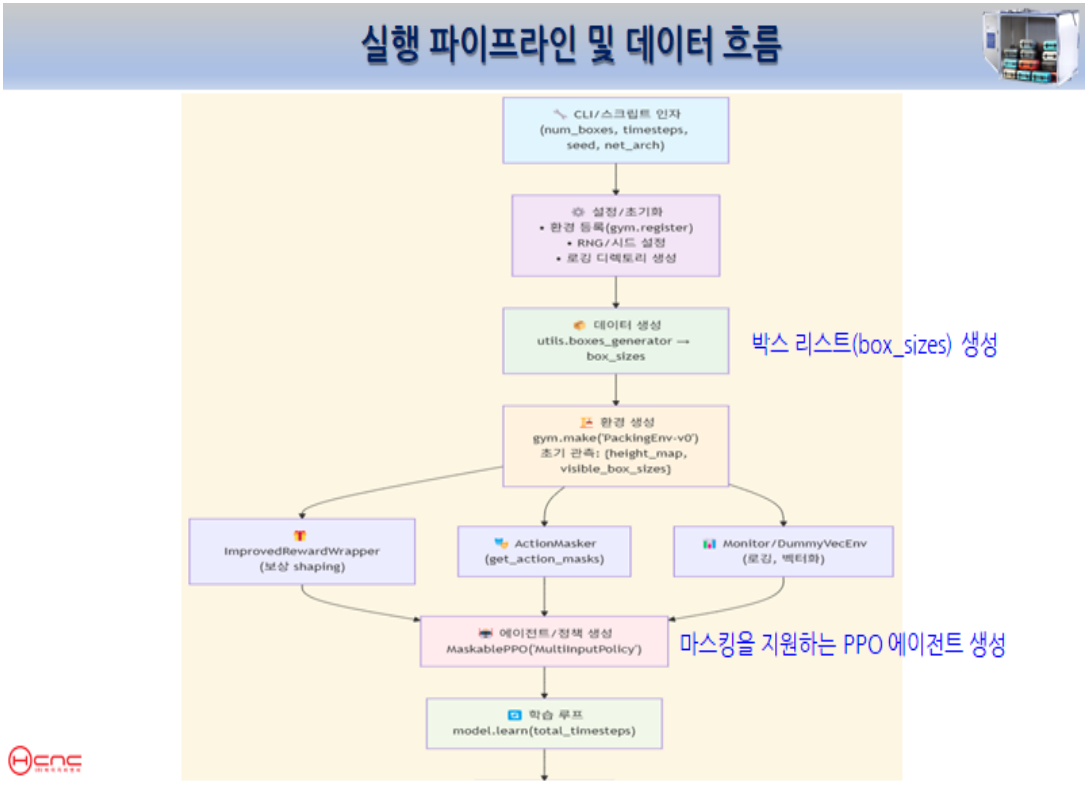
3. 수행업무 3: 최상위 실행 스크립트의 실행 파이프라인 및 코드베이스 전체의 실행 파이프라인 다이어그램 작성

□ 결과:

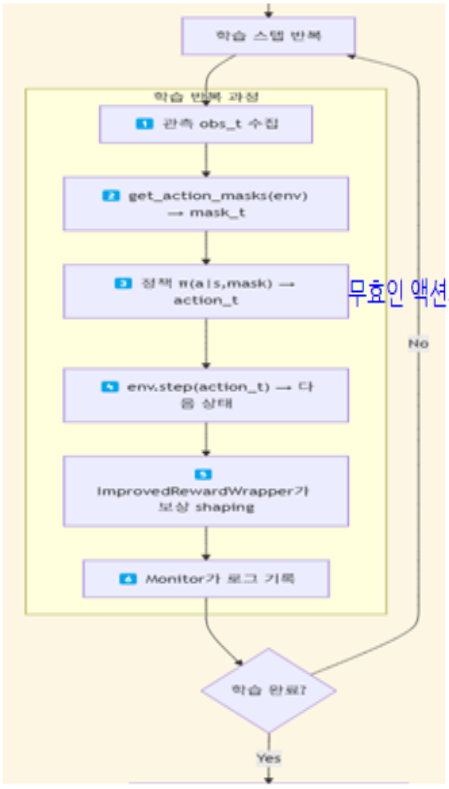
● 최상위 실행 스크립트의 실행 파이프라인 작성



● 코드베이스 전체의 실행 파이프라인 다이어그램 작성



실행 파이프라인 및 데이터 흐름



무효인 액션의 확률 0 처리

실행 파이프라인 및 데이터 흐름

각 trial이 위의 "설정→데이터→환경→학습→평가" 파이프라인을 반복 수행

4. 수행업무 4: 액터-크리틱/PPO 알고리듬의 성능 고도화를 위해 기본적으로 액션 매스킹, 정교한 보상 설계(보상 쉐이핑) 전략을 채용함

- 액션 매스킹(Action Masking): 불가능한 좌표/박스의 조합을 사전에 제거함으로써 탐색 공간을 축소하는 효과가 있음
- 보상 쉐이핑(Reward Shaping): 에이전트가 더 빠르고 효율적으로 학습하도록 돕기 위해 기존의 보상(reward) 시스템에 추가적인 보상(예: 효율/안정성 보너스 , 지연/실패 페널티 등)을 설계하여 제공하는 기술. 특히, 최종 목표 달성까지 오랜 시간이 걸리거나, 목표를 달성했을 때만 보상이 주어지는 희소한(sparse) 보상 환경에서 학습 속도를 높이는 데 효과적임

□ **결과:**

- 액션 공간

```
box_index = action // (self.container.size[0] * self.container.size[1])
res = action % (self.container.size[0] * self.container.size[1])
position = np.array([res // self.container.size[0], res % self.container.size[0]])
return box_index, position.astype(np.int32)
```

- 액션 매스킹

```
act_mask = np.zeros((self.num_visible_boxes, self.container.size[0] * self.container.size[1]), dtype=np.int8)
for index in range(len(self.unpacked_visible_boxes)):
    acm = self.container.action_mask(box=self.unpacked_visible_boxes[index], check_area=100)
    act_mask[index] = np.reshape(acm, (self.container.size[0] * self.container.size[1],))
return [x == 1 for x in act_mask.flatten()]
```

- 보상 쉐이핑

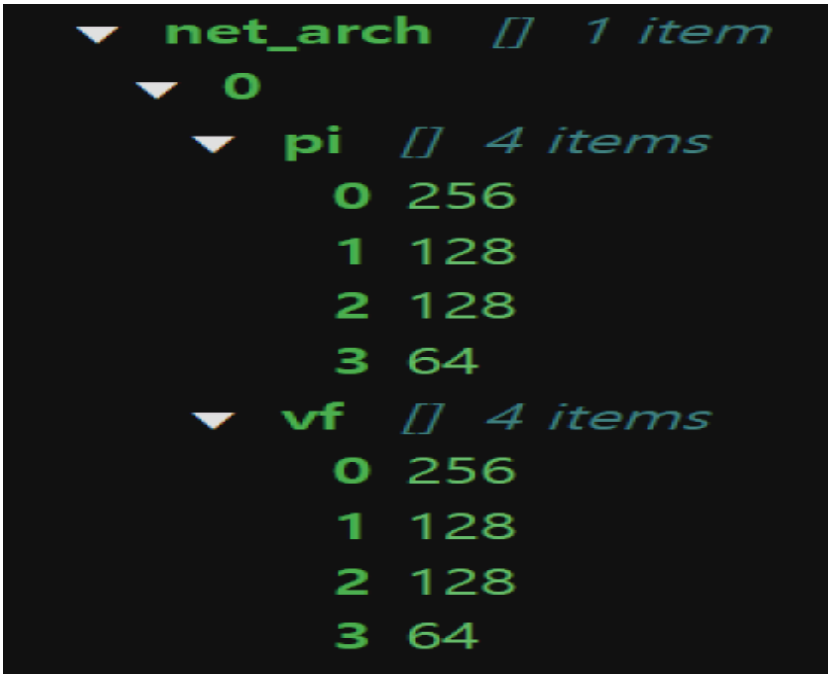
```
if reward_type == "terminal_step":
    reward = packed_volume / container_volume
elif reward_type == "interm_step":
    reward = packed_volume / ((max_x-min_x)*(max_y-min_y)*(max_z-min_z))
```

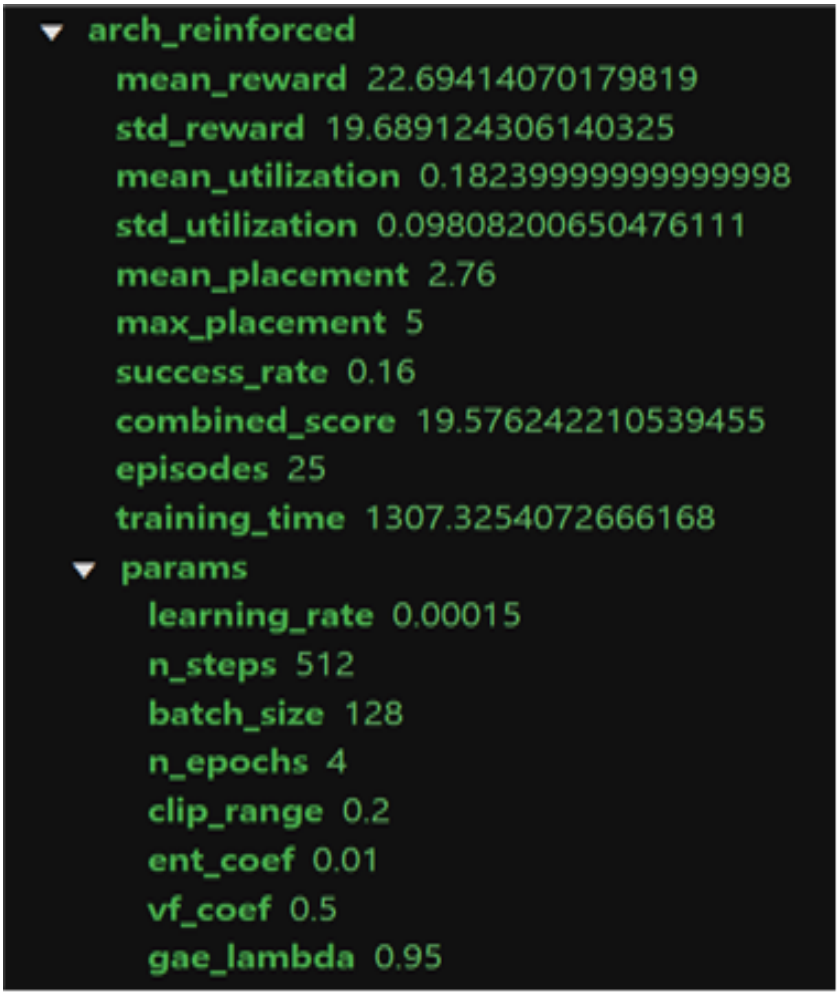
5. 수행업무 5: 추가적으로, Optuna 및 W&B 최적화 도구의 발굴/도입에 의한 네트워크 구조 및 Hyperparameter의 동시 최적화를 자동화함으로써 PPO 학습 동작의 최적화 및 성능의 고도화 달성 완료

- Optuna 및 W&B 최적화 도구를 신규로 발굴하여 본 최적화 솔루션에 도입
- 1단계-정밀최적화(enhanced_optimization.py): 전략별 다중 실험에 의한 탐색/비교를 통해 → 최적의 조합을 도출
- 2단계-최종검증(production_final_test.py): 1단계에서 도출된 설정으로 실제 사용할 최종 모델을 만들어 50 에피소드에 걸친 반복 평가로 재현성을 확인

□ **결과:**

- 네트워크 구조의 최적화



	● Hyperparameter의 최적화 
지원성과/ 기대효과	❍ 지원에 따른 예상/기대효과에 대해 기술 - 본 과제에서 확보된 Optuna 및 W&B라는 최적화 도구를 일반적인 타 과제 딥러닝의 경우에도 도입/적용함으로써 AI 모델 성능의 최적화 시간을 기존에 비해 대폭 절감하여 생산성의 혁신을 기대할 수 있음 - SOTA논문의 트랜스포머 모델 대신에 경량 MLP 기반으로 비용 절감과 동시에 20배 이상 빠른 반복 실험이 가능할 정도의 속도 및 운영의 용이성 확보 - 액션 매스킹·보상설계 고도화로 학습 안정화 및 학습 효율성을 150% 개선 ❍ 항공물류 분야의 ULD 적재 최적화를 위한 AI 솔루션의 기술경쟁력 확보 및 이를 통한 기업가치의 제고와 현재 보류 중인 개발계획의 향후 재개시에 솔루션의 고도화, 상용화를 위한 프로토타입으로써 활용 가능 ❍ 개발된 적재 최적화 AI 솔루션을 기반으로 향후 유사 물류분야인 육상/해상 물류 분야의 창고 관리/컨테이너 적재 및 화물운송 경로 관리를

	위한 최적화 AI 솔루션으로 사업분야의 확장도 기대 가능
현장지원 소감	❍ 파견 기업이 일반적인 경우와 달리 제조현장의 데이터를 보유하지 못한 데다가, GPU 등의 기본 AI 인프라도 보유하고 있지 않아 파견 초기에 애로를 많이 겪었음 ❍ 결국, 이를 극복하기 위해 상술한 클라우드 개발환경의 자체 개발로 갈 수 밖에 없어 에너지의 소모가 극심했는데, 경영진은 이러한 사정에 아예 무관심한 점이 아쉬웠음 ❍ 개발한 AI 기반 ULD 적재 최적화 솔루션이 트럼프 관세 대란으로 인한 수주 불발 사태로 항공화물 업계 굴지의 업체인 다이후쿠 사의 자동화 솔루션 상용화에 기여할 기회가 사라진 점이 아쉬운 점으로 남음