ETRI 연구인력 현장지원 실적보고서

과제명 : 강화학습 기반 3D Bin Packing 최적화 시스템 개발(ULD 적재 자동화)
지원기간 : 2024.07 ~ 2025.09
지원기관 : 한국전자통신연구원(ETRI)
현장기업 : (기입)
지원연구원 : 박정우
본 보고서는 Code동작 기술서 및 실행 산출물(results/, models/, logs/)을 근거로 작성

1. 지원 배경 및 목적

- 항공 ULD 적재의 수작업 의존 문제 개선 및 자동화 PoC 구축
- 강화학습(RL) 기반 3D Bin Packing으로 컨테이너 공간 활용률 극대화
- 연구용 Transformer 대신 **경량 MLP + MaskablePPO**로 실용 성능/비용 최적화

2. 지원 내용

2.1 시스템/모듈 구조 설계

- **환경**(`src/packing_env.py`): 관측(Height Map, visible_box_sizes), 액션(Discrete: 위치×가시박스)
- **커널**(`src/packing_kernel.py`): 배치/충돌/지지율/높이맵 갱신 로직
- **학습**(`src/train_maskable_ppo.py`): `make_env()`, `get_action_masks()`, `ImprovedRewardWrapper`
- **유틸**(`src/utils.py`): `boxes_generator()` 등 기하 유틸

2.2 핵심 알고리즘 구현

- **Action Masking**: 불가능 좌표/박스 조합 사전 제거 → 탐색 공간 축소
- **Reward Shaping**: 활용률 개선·배치 성공·효율/안정성 보너스 등 단계별 보상
- **결합 점수**: 0.3×mean_reward + 0.7×(mean_utilization×100)

2.3 안정성/운영 이슈 해결

- **999 스텝 무한대기**: 안전 콜백(평가 타임아웃/주기 제어)로 100% 해결
- **CUDA/의존성**: CPU-only 실행 옵션, 의존성 고정
- **Import/Type 오류**: `BoxCreator` 제거, `activation_fn=nn.ReLU` 통합

2.4 커리큘럼 학습(안전형)

- 최근 에피소드 성과 기반으로 박스 수 난이도 점진 증가
- 연속 성공(patience) 조건 충족 시 단계 상승, 불안정 시 유지

2.5 HPO 파이프라인

- **정밀 최적화**(`enhanced_optimization.py`): 전략별 다중 실험 → 최적 조합 도출
- **최종 검증**(`production_final_test.py`): 50 에피소드 반복평가로 재현성 확인

3. 지원 실적 (정량/정성)

3.1 정량 성과

지표	성과	근거
Phase4 Best Combined	19.576	results/phase4_enhanced_all_0804a.json (arch_reinforced)
최종 평균 보상 (No Curri)	6.8344	results/ultimate_results0717_noCurri.txt
최종 평균 보상 (Curri)	5.8725	results/ultimate_results0717_Curri.txt
학습 시간 (No Curri)	172.8초	results/ultimate_results0717_noCurri.txt
학습 시간 (Curri)	137.7초	results/ultimate_results0717_Curri.txt

3.2 정성 성과

- 경량 MLP 기반으로 빠른 반복실험 및 운영 용이성 확보
- 실시간 모니터링/대시보드/GIF로 가시성 강화
- 커리큘럼·Masking·보상설계로 학습 안정화 및 수렴 가속

4. 문제점 및 개선사항

- 단기 실험에서 커리큘럼 기본 설정이 보상 평균을 소폭 저하 → **난이도 상승 조건/속도 재튜닝** 필요
- 장시간 학습 시 성능 분산 존재 → 평가 주기 세분화, 조기중단 규칙 강화
- 대형 박스 시나리오의 초기 결정 치명성 → 초기 탐색 강화, 온정책 보상 미세조정

5. 성과 및 기대효과

- **실용화**: 일반 GPU 1장 기준 단시간 재학습/배포 가능
- **비용 절감**: Transformer 미사용으로 인프라 비용 절감
- **현장 이득**: 인력 운영 효율화, 적재 품질/속도 향상

6. 향후 계획

- **커리큘럼 재튜닝**: 박스 수 증가 단계의 임계/윈도우/보상 조건 개선
- **속도 최적화**: 60초 내 학습 목표(롤아웃/로딩 병렬화)
- **모델 고도화**: MLP + Self-Attention 하이브리드 탐색

7. 첨부 및 근거자료

- `results/`: phase4_enhanced_all_0804a.json, ultimate_results0717_*.txt, *.png
- `models/`: 학습 결과 zip
- `logs/`: 모니터링 CSV/텐서보드 로그
- `docs/`: Code동작 기술서.md 외 문서

참고: Combined Score = 0.3×mean_reward + 0.7×(mean_utilization×100)

작성일 : 2025-09-03
작성자 : 박정우