

캡스톤디자인 II 중간보고서

프로젝트명 : 항공영상 기반의 제로샷 객체 탐지 연구
캡스톤 디자인 II, 중간보고서

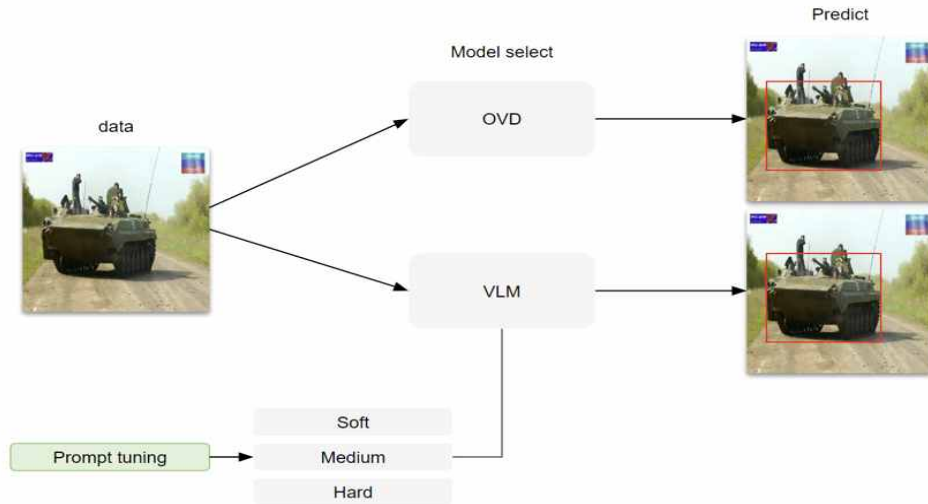
Version 1.0

개발 팀원 명(팀리더): 박우진
김다빈

대표 연락처: 010-2360-2799
e-mail: 20201735@edu.hanbat.ac.kr

캡스톤 디자인 II 중간보고서 내용

1. 요구사항 정의서에 명시된 기능에 대하여 현재까지 진척된 결과 및 그 내용을 기술하시오.
 - 시각-언어 모델(VLM)을 기반으로 군용 객체 탐지 성능 향상을 위한 프롬프트 튜닝 실험을 수행하였다.



| 모델 | 이미지1 | 이미지2 |
|----------------|------|------|
| YOLO-World | | |
| Grounding DINO | | |
| Florence-2 | | |
| Qwen_S | | |
| Qwen_M | | |
| Qwen_H | | |

| 모델 | 이미지1 | 이미지2 |
|--------|------|------|
| Qwen_S | | |
| Qwen_M | | |
| Qwen_H | | |

- 데이터 부족 문제를 보완하기 위한 합성 기법을 활용해 군용 차량 이미지를 추가하였다.



- OVD 기반 모델들을 활용하여 군용 객체 탐지 성능을 실험적으로 비교하였으며, 그 결과는 아래와 같다.

| model | class | mAP(0.5~0.95) |
|----------------|-----------------|---------------|
| YOLO-World-S | military object | 0.016 |
| YOLO-World-L | military object | 0.299 |
| Grounding-DINO | military object | 0.745 |



- VLM을 대상으로 다양한 프롬프트를 적용하여 군용 객체 제로샷 탐지 성능을 평가, 비교하였다.

| model | method | class | mAP(0.5~0.95) | all mAP(0.5~0.95) |
|------------|-------------------------------|----------------|---------------|----------------------|
| Qwen2.5-VL | 외형적 정보를 기반으로 반환 형식대로 탐지 | T-80 | 0.121 | 0.264 |
| | | K-2 | 0.149 | |
| | | BMP-3 | 0.223 | |
| | | K-200 | 0.178 | |
| | | Military Truck | 0.384 | |

| model | method | class | mAP(0.5~0.95) | all mAP(0.5~0.95) |
|------------|--------------------------------------|----------------|---------------|----------------------|
| Qwen2.5-VL | 드론뷰 명시 + 외형적으로 구별되는 특징만을 선별 | T-80 | 0.155 | 0.368 |
| | | K-2 | 0.342 | |
| | | BMP-3 | 0.253 | |
| | | K-200 | 0.277 | |
| | | Military Truck | 0.556 | |



사용 모델

- Tianheng Cheng, et al. “YOLO-World: Real-Time Open-Vocabulary Object Detection” CVPR 2024.
- S. Liu, et al. “Grounding DINO: Marrying DINO with Grounded Pre-Training for Open-Set Object Detection” ECCV 2024.
- Shuai Bai, et al. “Qwen2.5-VL Technical Report” arXiv 2025.

2. 프로젝트 수행을 위해 적용된 추진전략, 수행 방법의 결과를 작성하고, 만일 적용과정에서 문제점이 도출되었다면 그 문제를 분석하고 해결방안을 기술하시오.
- 단일 OVD에서 클래스명, 서술형 프롬프트만으로는 성능 개선 폭이 제한되고 저해상도에서 세부가 소실되었다. 이에 VLM을 결합한 Two-Stage 구조로 전환하고, 객체 외형 특징을 반영한 프롬프트를 설계하였다.
 - 탐지는 OVD가, 분류는 지식이 풍부한 VLM이 맡도록 역할을 분리했다. 1단계에서는 Grounding DINO가 localization 역할을 수행하여 안정적으로 박스를 획득하고, 2단계에서는 해당 ROI를 Qwen2.5-VL에 전달해 classification 역할을 수행하며 세부 클래스를 식별한다.

프로젝트명 : 항공영상 기반의 제로샷 객체 탐지 연구

소프트웨어 요구사항 정의서

Version 1.0

개발 팀원 명(팀리더):박우진
김다빈

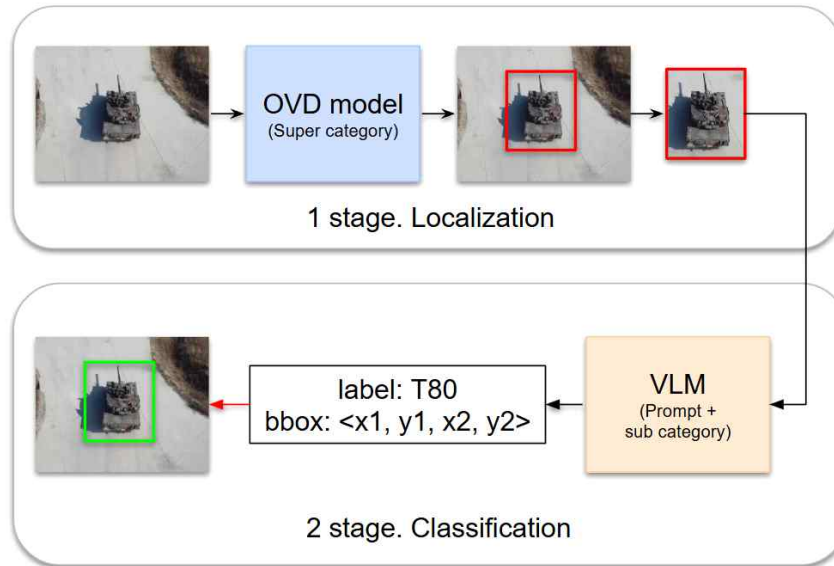
대표 연락처: 010-2360-2799
e-mail: 20201735@edu.hanbat.ac.kr

목 차

| | |
|-------------------------|----|
| 1. 시스템 개요 | 8 |
| 2. 시스템 장비 구성 요구사항 | 8 |
| 3. 기능 요구사항 | 9 |
| 4. 성능 요구사항 | 9 |
| 5. 인터페이스 요구사항 | 9 |
| 6. 데이터 요구사항 | 10 |
| 7. 테스트 요구사항 | 10 |
| 8. 보안 요구사항 | 10 |
| 9. 품질 요구사항 | 10 |
| 10. 제약 사항 | 11 |
| 11. 프로젝트 관리 요구사항 | 11 |

1. 시스템 개요

- Zero-shot 성능을 향상시키기 위한 Two-Stage 파이프라인 구조



2. 시스템 장비 구성요구사항

| 요구사항 고유번호 | | ECR-001 | | |
|------------|-------|--|------|-----|
| 요구사항 명칭 | | 장비 요구사항 | | |
| 요구사항 분류 | | 시스템 장비구성 요구사항 | 응락수준 | 필 수 |
| 요구사항 상세 설명 | 정의 | 모델 추론 장비 | | |
| | 세부 내용 | <ul style="list-style-type: none"> - 장비 품목 : GPU (NVIDIA A6000) - 장비 수량 : 2개 - 장비 기능 : 모델의 연산 속도를 높인다. - 장비 성능 및 특징 : 개당 VRAM 48GB | | |

3. 기능 요구사항

| 요구사항 고유번호 | | SFR-001 | | |
|------------|-------|---|------|-----|
| 요구사항 명칭 | | AI 모델 개발 | | |
| 요구사항 분류 | | 기능 | 응락수준 | 필 수 |
| 요구사항 상세 설명 | 정의 | OVD로 ROI를 검출한 뒤 VLM으로 세부 분류하는 end-to-end 파이프라인 개발 | | |
| | 세부 내용 | <ul style="list-style-type: none"> - 1단계(OVD): OVD로 군용 객체(super-class) 탐지 - 2단계(VLM): VLM에 ROI 전달, 세부 클래스로 분류 - ROI 마진·해상도·threshold를 설정값으로 관리 | | |

4. 성능 요구사항

| 요구사항 고유번호 | | PER-001 | | |
|------------|-------|-------------------------------------|------|-----|
| 요구사항 명칭 | | 처리 속도 및 시간 | | |
| 요구사항 분류 | | 성능 요구사항 | 응락수준 | 필 수 |
| 요구사항 상세 설명 | 정의 | 처리 속도 및 시간 | | |
| | 세부 내용 | - 이미지를 입력으로 받아 모델이 답변을 추론하는 시간을 의미함 | | |

5. 인터페이스 요구사항

| 요구사항 고유번호 | | SIR-001 | | |
|------------|-------|---|------|-----|
| 요구사항 명칭 | | 대화형 컴퓨팅환경 | | |
| 요구사항 분류 | | 사용자 인터페이스 | 응락수준 | 필 수 |
| 요구사항 상세 설명 | 정의 | 두 단계 제로샷 탐지를 위한 대화형 컴퓨팅환경 구현 | | |
| | 세부 내용 | <ul style="list-style-type: none"> - 사용자가 OVD(탐지)·VLM(분류) 실험/검증을 쉽게 수행할 수 있도록 Jupyter(ipynb) 기반 인터페이스 제공 - detector(OVD) · classifier(VLM) · evaluator 모듈별 라이브러리로 구성하고, 탐지/분류/평가를 인터페이스에서 명확히 구분하여 표시 | | |

6. 데이터 요구사항

| | | | |
|------------|---|------|-----|
| 요구사항 고유번호 | DAR-001 | | |
| 요구사항 명칭 | 드론뷰 군용 객체 데이터셋 | | |
| 요구사항 분류 | 데이터 | 응락수준 | 필 수 |
| 요구사항 상세 설명 | - 드론 시점 이미지 수집 후 coco형식 바운딩박스 어노테이션 라벨링 후 JPG 파일과 JSON 파일로 저장 | | |

7. 테스트 요구사항

| | | | |
|------------|--|------|-----|
| 요구사항 고유번호 | TER-001 | | |
| 요구사항 명칭 | 성능 테스트 | | |
| 요구사항 분류 | 테스트 | 응락수준 | 필 수 |
| 요구사항 상세 설명 | - mAP를 사용하여 드론뷰 데이터셋에 대한 OVD, VLM, Two-Stage에 따른 성능을 비교하여 평가 | | |

8. 보안 요구사항

| | | | |
|------------|---|------|-----|
| 요구사항 고유번호 | SER-001 | | |
| 요구사항 명칭 | 보안지침 준수 | | |
| 요구사항 분류 | 보안 | 응락수준 | 필 수 |
| 요구사항 상세 설명 | - 군용 객체 데이터는 출처·라이선스 준수하고, 보안 및 저작권 관련 데이터 이용정책에 따라 개발이 수행되어야 함 | | |

9. 품질 요구사항

| | | | |
|------------|----------|--|-----|
| 요구사항 고유번호 | QUR-001 | | |
| 요구사항 명칭 | 데이터 품질관리 | | |
| 요구사항 분류 | 품질 | 응락수준 | 필 수 |
| 요구사항 상세 설명 | 정의 | 품질관리(기술 관점) | |
| | 세부 내용 | <ul style="list-style-type: none"> - 데이터는 드론 시점으로 구성하고 포맷(JPG)을 통일해야함. - 데이터는 중복·손상·비정상 메타데이터가 없어야 함 - 바운딩박스/라벨 결측치가 없어야 하며, 서브클래스 명칭은 클래스 맵 평표 와 일치해야 함. | |

10. 제약 사항

| 요구사항 고유번호 | COR-001 | | |
|------------|---|------|-----|
| 요구사항 명칭 | 시스템 개발과 설계 및 구현 제약사항 | | |
| 요구사항 분류 | 제약사항 | 응락수준 | 필 수 |
| 요구사항 상세 설명 | <ul style="list-style-type: none"> - 현재 보유하여 활용 가능한 H/W, S/W를 최대한 활용함 - 대부분의 인공지능 모델 개발에 사용되는 Python(언어), PyTorch(프레임워크)를 사용함 | | |

11. 프로젝트 관리 요구사항

| 요구사항 고유번호 | PMR-001 | | |
|------------|--|------|-----|
| 요구사항 명칭 | 프로젝트 관리 | | |
| 요구사항 분류 | 프로젝트 관리 | 응락수준 | 필 수 |
| 요구사항 상세 설명 | <ul style="list-style-type: none"> - 세부 작업 분할 구조 : <ol style="list-style-type: none"> 1. 분석 <ul style="list-style-type: none"> - Two-Stage 제로샷 탐지의 범위, 지표, 성능 목표 정의 - 제로샷 탐지를 위한 기존의 방법론 조사 2. 데이터 수집 및 전처리 <ul style="list-style-type: none"> - 드론뷰 데이터셋 수집 - COCO 포맷으로 변환 후 중복/손상된 이미지 제거 3. 시스템 설계 <ul style="list-style-type: none"> - OVD 모델과 VLM을 사용하여 제로샷 탐지 성능을 향상 시키는 모델 설계 4. 실험 <ul style="list-style-type: none"> - mAP를 사용하여 드론뷰 데이터셋에 대한 OVD, VLM, Two-stage에 대한 비교 실험을 진행하여 성능 측정 - 프로젝트 수행조직에 대한 구성, 역할 <ul style="list-style-type: none"> 김다빈 : 논문 및 자료 조사, 데이터셋 전처리, 모델 코드 작성 박우진 : 논문 및 자료 조사, 데이터셋 전처리, 모델 코드 작성 | | |