

캡스톤디자인 II 계획서

제 목	국문	항공영상 기반의 제로샷 객체 탐지 연구				
	영문	Zero-shot Object Detection Research Based on Aerial Imagery				
프로젝트 목표 (500자 내외)	<p>기존 객체 탐지 모델은 사전 학습된 클래스에만 의존하여 새로운 객체를 탐지할 때 성능이 크게 저하되는 한계가 있다.</p> <p>본 프로젝트는 Open Vocabulary Object Detection(OVD) 모델과 Vision-Language Model(VLM)을 결합한 Two-stage 전략을 통해 제로샷(Zero-shot) 방식으로 군용 객체를 탐지하는 기술을 개발하는 것을 목표로 한다.</p> <p>이 접근을 통해 One-stage 탐지 기법의 제약을 극복하고, 향후 KCI급 학술지 2편 게재를 통해 연구 성과를 학술적으로 검증할 계획이다.</p>					
프로젝트 내용	<p>본 프로젝트는 다양한 각도에서 촬영된 드론 영상을 대상으로 제로샷 방식으로 군용 객체를 탐지하는 기술을 개발하는 데 중점을 둔다. 기존 One-stage 객체 탐지 모델은 새로운 유형의 군용 객체가 등장했을 때 탐지 성능이 크게 저하되는 문제가 있다. 이를 극복하기 위해 본 연구에서는 OVD와 VLM을 결합한 Two-stage 전략 기반의 탐지 방법을 제안한다. 첫 번째 단계에서는 OVD 모델을 활용하여 ‘military object’와 같은 상위 클래스에 대한 후보 영역을 정확히 검출하고, 두 번째 단계에서는 VLM을 이용하여 검출된 영역 내에서 K2전차, K200장갑차, T80전차, BM P3 장갑차, Military truck 등 서브클래스를 분류한다. 정량적 지표인 mAP를 통해 성능을 평가한다. 제안한 Two-stage 접근법은 새로운 군용 객체에 대해 빠르게 적응함으로써 객체 탐지 정확도를 향상시킬 것으로 기대된다.</p>					
기대효과 (500자 이내) (응용분야 및 활용범위)	<p>본 프로젝트에서 제안한 제로샷 군용 객체 탐지 기술은 사전 학습에 포함되지 않은 클래스를 학습 없이 탐지할 수 있으며, 실시간 추론 최적화를 통해 드론 영상, 위성 영상 등 다양한 원격 감시 시스템에 적용할 수 있다. 이를 통해 군사 정찰 및 국경 감시에서 신속한 위협 식별이 가능하며, 재난 대응, 공공 안전 모니터링, 교통 관리 등 민간 분야에서도 운영 비용 절감과 대응 속도 향상을 기대할 수 있다. 연구 결과는 KCI급 학술지 발표를 통해 원격 감시 분야의 연구 및 응용 범위 확장에 기여할 것이다.</p>					
중심어(국문)	컴퓨터비전	제로샷 탐지	개방형 어휘 객체 탐지	시각-언어 모델		
Keywords (english)	Computer Vision	Zero-shot Detection	Open-Vocabulary Object Detection	Vision-Language Model		
멘토	소속	데이터메이커	이름	나현우		
팀 구성원	학년 /반	학 번	이 름	연락처(전화번호/이메일)		
	4	20201735	박우진	010-2360-2799/20201735@edu.hanbat.ac.kr		
	4	20222019	김다빈	010-9855-2041/20222019@edu.hanbat.ac.kr		
<p>컴퓨터공학과 캡스톤디자인 관리규정과 모든 지시사항을 준수하면서 본 캡스톤디자인을 성실히 수행하고자 아래와 같이 계획서를 제출합니다.</p> <p>2025 년 6월 27일</p> <p>책 임 자 : 박우진 (인)</p> <p>지도교수 : 장한얼 (인)</p>						

캡스톤디자인 계획서(양식)

1. 캡스톤디자인의 배경 및 필요성

드론 영상은 다양한 고도와 각도, 복잡한 지형 환경을 대상으로 관측할 수 있어 군사 정찰·재난 대응·국경 감시 등 원격 감시 시스템의 핵심 데이터로 활용되고 있다. 최근 딥러닝 기반 객체 탐지(Object Detection) 연구는 광범위한 응용 가능성을 제시했지만, 라벨링된 학습 데이터에 의존한다는 한계가 있다. 특히 군용 차량과 같은 민감한 객체는 확보 가능한 실제 이미지가 극히 제한적일 뿐만 아니라, 전문가의 세밀한 라벨링이 요구되어 시간·비용 부담이 매우 크다.

기존 One-stage 모델은 빠른 추론이 가능하다는 장점이 있으나, 클래스 일반화 성능이 떨어질 수 있다는 단점이 있다. 특히, 학습 시 정의된 클래스 외의 새로운 객체가 등장할 경우 탐지 성능이 급격히 저하되어 실사용에 어려움이 있다. 이러한 한계를 극복하기 위해 Open-Vocabulary Object Detection(OVD)[1]과 Vision-Language Model(VLM)[2]을 결합한 Two-stage 전략이 주목받고 있다. Two-stage 방식은 상위 클래스 검출 과정과 세부 서브클래스 분류 과정을 분리하여 처리함으로써, 세부 클래스 간 미묘한 차이를 더 정확하게 구분할 수 있다는 장점이 있다.

최근 드론을 활용한 군 감시 및 정찰의 중요성이 커지고 있으나, 실전 환경에서의 다양한 군용 객체를 탐지하기에는 여전히 기술적 제약이 존재한다. 특히, 실제 데이터의 수집이 어렵고, 기존 모델은 학습된 클래스 외 객체에 대한 탐지 성능이 크게 저하되는 문제가 있다. 이러한 한계를 극복하기 위해서는 데이터 부족 문제를 완화하고, 학습되지 않은 새로운 객체도 탐지 가능한 제로샷 객체 탐지 기술의 개발이 필요하다.

2. 캡스톤디자인 목표 및 비전

본 캡스톤디자인은 드론 영상에서 제로샷(Zero-shot) 방식으로 군용 차량을 탐지하는 기술 개발에 중점을 둔다. 특히, 세부 클래스 간 미묘한 차이를 보다 정확하게 구분할 수 있는 OVD(Object-Vocabulary Detection) + VLM(Vision-Language Model) 기반 Two-stage 전략을 제안한다.

기존 One-stage 탐지 기법은 학습된 클래스 외 신종 객체에 대해 유연한 대처가 어렵고, 실제 작전 환경의 다양한 드론 영상 특성(고도·날씨 등)에 취약하다는 문제가 있다. 본 연구에서는 OVD모델을 활용한 상위 클래스 검출과 VLM을 활용한 세부 클래스 분류를 분리하여 수행함으로써, 새로운 군용 차량에 대한 자연어 지시만으로 객체를 탐지할 수 있는 파이프라인을 구현하고자 한다.

본 연구의 최종 목표는 다음과 같다. 첫째, 제안한 Two-stage 접근을 통해 baseline 대비 정량적 지표 mAP에서 의미 있는 성능 향상을 입증한다. 둘째, 연구 성과를 바탕으로 KCI급 학술지 2편을 게재하여 학술적으로 기여하는 것이다.

이 과정에서 다양한 OVD(YOLO-World[3], Grounding DINO[4] 등)모델과 VLM(Qwen 2.5 VL[5], Florence-2[6], Kimi-VL[7] 등)을 활용하여 각 모델의 성능을 비교 분석하여 성능이 우수한 모델을 선택하여 활용할 계획이다. 또한, 합성 데이터를 생성하여 훈련 데이터를 확장할 때 발생할 수 있는 편향(Bias)을 최소화하기 위해 다양한 데이터 증대 기법을 적용하고, 모델의 일반화 성능을 안정적으로 확보할 수 있도록 실제 데이터와 합성 데이터를 적절한 비율로 혼합하여 활용할 계획이다.

추가 연구를 통해 실제 응용 분야에서의 활용 가능성을 확장할 수 있을 것으로 기대된다.

3. 캡스톤디자인 내용

분류	내용	설명
기능적 요구사항	드론뷰 이미지 분석	<ul style="list-style-type: none"> - 다양한 고도·각도·조명·날씨 조건에서 수집된 드론 영상에 대한 전처리 및 분석 - 군용 객체의 크기, 형태, 배경 특성을 파악
	데이터 수집 및 전처리	<ul style="list-style-type: none"> - 실제 드론 영상 수집과 Stable Diffusion 기반 합성 이미지 생성 - 모델의 일반화 성능을 안정적으로 확보할 수 있는 증대법을 사용하여 합성 이미지 생성 - SAM을 이용한 전경 마스크 추출 및 면적 비율 기준 이상치 필터링을 통해 학습용 데이터 확보
	Two-stage 모델 설계	<ul style="list-style-type: none"> - OVD 모델을 활용한 상위 클래스(region proposal) 검출 및 VLM을 이용한 K2, K200, T80, BMP3, Military truck 서브클래스 분류 파이프라인 구현
	실제 효용성 검증	<ul style="list-style-type: none"> - 데이터메이커에서 제공하는 군용 차량 데이터셋[8]을 사용한 탐지 성능 평가 진행
비기능적 요구사항	성능	<ul style="list-style-type: none"> - 모델의 mAP 성능 비교 평가 진행
	보안	<ul style="list-style-type: none"> - 데이터셋은 Roboflow Universe에 일부 공개, 외부에서는 제한된 범위 내에서만 사용 가능
	확장성	<ul style="list-style-type: none"> - 개발한 모델이 다른 데이터셋에서 적용 가능한지 테스트하여 다양한 군용 객체에 대한 모델의 적용 범위 확장
	결과	<ul style="list-style-type: none"> - 제안한 제로샷 군용 객체 탐지 기술의 효과를 검증하여 KCI급 학술지 2편 게재 - GitHub 레포지토리에 (https://github.com/KimDaBin0704/AerD) 공개

4. 캡스톤디자인 추진전략 및 방법

본 연구는 제로샷 군용 객체 탐지 기술을 단계별로 구현·검증하기 위해 다음과 같은 전략과 방법론을 적용한다.

첫째, 데이터셋 구축 및 전처리 단계에서는 Roboflow Universe 웹사이트를 통해 수집한 K2, K200, T80, BMP3, Military truck의 실제 군용 차량 이미지를 기반으로 학습·검증·테스트 세 분할을 수행한다. 추가로 Stable Diffusion과 ControlNet을 활용한 합성 파이프라인을 통해 신규 환경(배경·조명·각도)에 맞춘 합성 이미지를 생성한다. 합성 이미지를 생성할 때 편향을 유발할 수 있기에 모델의 일반화 성능을 안정적으로 얻을 수 있는 데이터 증대법을 사용하고 학습 데이터에 추가한다.

둘째, Two-stage 모델 설계 및 학습 단계에서는 OVD 모델을 상위 클래스(“military object”) 검출용 모델로, VLM을 세부 서브클래스 분류용 모델로 활용하는 파이프라인을 구현할 예정이다.

데이터셋의 학습을 위해 NVIDIA RTX A5000 24GB GPU 2대를 사용할 예정이다. 팀 구성원 모두 전공과목인 “인공지능”, “컴퓨터비전”을 수강한 이력이 있으므로, 인공지능과 컴퓨터비전 관련 지식에 대한 이해가 있다. 개발 환경 및 협업 관리를 위해 전체 코드는 PyTorch 기반으로 개발될 예정이며, 최신 논문 기반 모델과 HuggingFace 기반 VLM 모듈을 통합하여 mAP를 측정한다. 이미지 전처리와 데이터 증강을 위해서 이미지 처리 라이브러리인 OpenCV를 활용한다. 딥러닝 기반 탐지 모델의 경우 최신 논문 기반 모델을 사용 가능한 MMDetection 프레임워크를 활용한다. 개발된 Two-stage 모델의 효용성을 데이터메이커에서 제공하는 군용 차량 데이터셋을 활용하여 검증한다. 모델의 성능은 기존 방법과의 비교 분석을 통해 정량, 정성적으로 평가한다.

관련 코드는 GitHub(<https://github.com/KimDaBin0704/AerD>)에 업로드하여 관리할 예정이다.

팀 구성원

	팀 구성	성명	역할
1	팀장	박우진	논문 및 자료조사, 모델 코드 작성 및 실험, 데이터셋 구축
2	팀원	김다빈	논문 및 자료조사, 모델 코드 작성 및 실험, 데이터셋 구축



5. 캡스톤디자인 결과의 활용방안

본 연구를 통해 개발된 제로샷 군용 객체 탐지 기술은 드론 영상 기반 군사 정찰 및 국경 감시에서 사전 학습되지 않은 신규 군용 차량을 즉시 식별함으로써 위협 감지와 대응 시간을 획기적으로 단축할 수 있다. 재난 현장에서는 피해 지역 내 차량 및 장비 분포 파악, 진입 경로 확보, 교통 통제 지원 등에 활용 가능하며, 더 나아가 산업 현장 원격 점검 시에도 알려지지 않은 객체의 식별 기능을 제공하여 현장 조사 비용과 인력 부담을 줄이는 것이 가능하다. 연구 결과는 KCI급 학술지 발표를 통해 원격 감시 및 컴퓨터 비전 분야의 학술적 기여와 기술적 발전에 중요한 역할을 할 것이다.

6. 참고문헌

- [1] Zareian, A., Dela Rosa, K., Hu, D. H., & Chang, S.-F. (2021). Open-Vocabulary Object Detection Using Captions. *CVPR 2021*.
- [2] Du, Y., Wei, F., Zhang, Z., Shi, M., Gao, Y., & Li, G. (2022). Learning to Prompt for Open-Vocabulary Object Detection with Vision-Language Model. *CVPR 2022*.
- [3] Cheng, T., Song, L., Ge, Y., Liu, W., Wang, X., & Shan, Y. (2024). YOLO-World: Real-Time Open-Vocabulary Object Detection. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2024*.
- [4] Ren, T., Shen, S., et al. (2024). Grounding-DINO: Grounded Open-Vocabulary Detection with Transformers. *CVPR 2024*.
- [5] Liu, et al. (2025). Qwen 2.5 VL Technical Report. *arXiv:2502.13923*.
- [6] Xiao, B., Wu, H., Xu, W., Dai, X., Hu, H., Lu, Y., Zeng, M., Liu, C., & Yuan, L. (2023). Florence-2: Advancing a Unified Representation for a Variety of Vision Tasks. *arXiv:2311.06242*.
- [7] Du, A., Yin, B., Xing, B., Qu, B., Wang, B., Chen, C., ... Sung, F., Wei, G., Lu, E., et al. (2025). Kimi-VL Technical Report. *arXiv:2504.07491*.
- [8] Roboflow Universe. (n.d.). Russian Military Annotated Dataset. Roboflow Universe. <https://universe.roboflow.com/capstoneproject/russian-military-annotated>

캡스톤디자인 II 계획발표 채점표

팀 구성원	학년/반	학 번	이 름				
제 목							
항목			점수				
			1	2	3	4	5
1. 프로젝트 주제의 필요성이나 중요성이 적절히 서술되었는가?							
2. 국내외 동향(문제 제기), 주요 기능(특징 포함) 및 범위가 적절히 서술되었는가?							
3. 기대효과(사회적, 기술적, 경제적 파급효과)가 적절히 서술되었는가?							
4. 추진 전략과 수행방법이 적절한가?							
5. 팀 구성과 역할 분담이 적절히 이루어졌는가?							
합계							
*수정 및 개선 의견							
<div style="text-align: center;">2013년 월 일</div> <div style="display: flex; justify-content: space-between; margin-top: 20px;"> <div>심사위원 :</div> <div>(인)</div> </div>							

※ 채점은 각 영역별 5점 만점을 기준으로 채점함.(상 5, 중 3, 하 1)

※ 계획서와 발표내용을 참고하여 채점표에 따라 평가함.