

캡스톤디자인 I 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(Object Detection) 모델은 사전에 학습된 데이터에 의존하여 새로운 유형의 객체가 등장하면 탐지 성능이 저하되는 한계를 가지고 있다. 본 프로젝트는 Open-Vocabulary Object Detection(OVD) 모델을 활용하여 제로샷(Zero-shot) 방식으로 새로운 군용 객체를 탐지할 수 있도록 하는 기술을 개발하는 것을 목표로 한다.</p> <p>군사 데이터는 보안상의 이유로 대규모 수집이 어렵다. 따라서, 이미지 합성 기법을 활용하여 현실적인 군용 객체 데이터를 생성하고, 다양한 환경에서도 강건한 탐지 성능을 유지할 수 있도록 모델을 최적화할 예정이다. 기존 객체 탐지 기술의 한계를 극복하고 실용적이고 확장 가능한 탐지 시스템을 구축하는 것을 목표로 한다.</p>				
프로젝트 내용	<p>본 프로젝트는 드론 영상에서 군용 객체 탐지를 위한 제로샷(Zero-shot) 방식의 Open-Vocabulary Object Detection(OVD) 탐지 기술을 개발하고, 데이터 부족 문제를 해결하기 위한 이미지 합성 기법을 적용하는 것에 초점을 맞춘다. 기존 객체 탐지 모델의 한계를 극복하기 위해 OVD 모델을 활용하여 자연어 설명만으로도 새로운 군용 객체를 탐지할 수 있도록 개선한다.</p> <p>드론 영상 데이터셋을 구축하고, 전처리 및 분석을 수행 후 최신 OVD 모델을 활용하여 군용 객체 탐지 성능을 높이기 위한 최적화 작업을 수행한다. 특히, 전차, 장갑차, 수송차의 세부 객체까지 탐지할 수 있도록 모델을 개선한다.</p> <p>모델의 성능은 탐지 정확도 성능의 정량적 지표를 활용하여 평가한다. 개발된 기술은 군사 작전뿐만 아니라 재난 대응, 국경 감시, 공공 안전 모니터링 등 다양한 분야에서의 활용 가능성을 검토하고, 연구 결과를 국제 학술지 및 컨퍼런스에 발표하여 학술적 기여를 목표로 한다. 이를 통해, OVD 기반 객체 탐지 기술의 실전 적용 가능성을 높이고, 원격 감지 및 감시 시스템의 발전에 기여할 것이다.</p>				
중심어(국문)	컴퓨터비전	개방형 어휘 객체 탐지	제로샷 탐지	이미지 합성	
Keywords (english)	Computer Vision	Open-Vocabulary Object Detection	Zero-shot Detection	Image Composition	
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>컴퓨터공학과의 캡스톤디자인 관리규정과 모든 지시사항을 준수하면서 본 캡스톤디자인을 성실히 수행하고자 아래와 같이 계획서를 제출합니다.</p> <p>2025 년 3월 5일</p> <p>책 임 자 : 박우진 </p> <p>희망 지도교수 : 장한얼</p>					

1. 캡스톤디자인의 배경 및 필요성

드론 기반 군사 감시 및 경찰 기술은 전장 및 국방 분야에서 필수적인 역할을 하고 있으며, 특히 전차, 장갑차, 수송차 등의 군용 객체 탐지는 정보 수집 및 전략적 의사 결정에 있어 중요한 기술로 자리 잡고 있다. 현재 국내·외에서 인공지능(AI) 기반 객체 탐지 기술이 활발하게 연구되고 있으며, 특히 YOLO[1], Faster R-CNN[2], DETR[3] 등 다양한 딥러닝 기반 객체 탐지 모델이 군사 및 보안 분야에서 활용되고 있다. 하지만 이러한 모델들은 사전에 학습된 객체만 탐지할 수 있는 한계를 가지며, 새로운 객체가 등장하면 탐지 성능이 저하될 가능성이 크다. 이에 따라, YOLO-World[4]와 같은 Open-Vocabulary Object Detection(OVD) 기술이 최근 주목받고 있으며, 이는 자연어 기반 탐지를 활용하여 학습되지 않은 객체도 탐지할 수 있도록 설계된 최신 객체 탐지 기법이다. 하지만 현재 OVD 기술이 군용 객체 탐지에 최적화 되어 연구된 사례는 부족하며, 실제 드론 영상 환경에서 신뢰할 수 있는 탐지 성능을 확보하는 데 어려움이 존재한다.

본 프로젝트는 OVD 모델을 최적화하여 드론 영상에서 군용 객체를 탐지할 수 있는 기술을 개발하는 것을 목표로 한다. 기존 객체 탐지 모델이 가지는 한계를 극복하고, 제로샷(Zero-shot) 방식으로 새로운 군용 객체를 탐지할 수 있도록 개선함으로써 실전 군사 작전에서의 적용 가능성을 높이하고자 한다. 또한, 군사 데이터가 부족한 문제를 해결하기 위해 이미지 합성 기법을 활용하여 현실적인 군용 객체 데이터를 생성하고, 이를 학습에 활용하는 방식으로 수행한다.

2. 캡스톤디자인 목표 및 비전

본 프로젝트는 기존 객체 탐지 모델의 한계를 극복하고, 자연어 기반 제로샷 탐지를 실현하는 OVD 기술을 군용 객체 탐지에 적용하는 도전적인 연구이다. 기존의 탐지 모델은 사전 학습된 데이터에 의존하여 새로운 군용 차량이 도입되면 재훈련이 필요하나, 본 연구에서는 자연어 설명을 활용하여 학습되지 않은 객체도 탐지할 수 있도록 하는 기술을 개발한다.

또한, 군사 데이터는 보안상의 이유로 대규모 수집이 어렵기 때문에 이미지 합성 기법을 활용하여 데이터 부족 문제를 해결한다. 최근 딥러닝 기반 이미지 합성 기술이 발전하면서, 현실적인 합성 데이터를 생성할 수 있는 가능성이 높아지고 있으며, 이를 군용 객체 탐지 모델에 효과적으로 적용하는 것이 본 연구의 창의적인 도전 과제 중 하나이다.

최종적으로, 본 연구가 성공적으로 수행될 경우, 군사 작전에서 실시간 객체 탐지 및 정보 수집을 가능하게 하며, 국방뿐만 아니라 보안 감시, 재난 대응 등 다양한 분야에서 활용할 수 있는 새로운 탐지 기술을 제시할 수 있을 것으로 기대된다.

3. 캡스톤디자인 내용

본 프로젝트는 드론 영상에서 군용 객체 탐지를 위한 OVD 모델을 개발하고, 데이터 부족 문제를 해결하기 위한 이미지 합성 기법을 적용하는 것을 목표로 한다. 이를 위해 다음과 같은 주요 기능을 포함한다.

주요 기능

항공영상 제로샷 군용 객체 탐지	- OVD 모델을 활용하여 항공영상 기반 학습하지 않은 전차, 장갑차, 수송차 등의 군용 객체 탐지 수행
데이터 구축 및 분석	- Image Composition[5] 모델을 활용한 데이터셋 구축 - opencv-python 라이브러리를 활용하여 데이터 시각화 - Image Harmonization[6] 기법을 활용한 데이터 전처리
환경 적응성	- 주·야간, 날씨 변화 등 다양한 환경에서도 탐지(궁극적 목표)

비기능적 요구사항

성능	- 기존 모델 대비 탐지 정확도를 높인 모델을 개발
유지보수성	- 모델의 확장 가능성을 고려하여 지속적 개선이 용이한 구조로 개발
결과	- OVD 모델을 이용해 제로샷 방식의 군용 객체 탐지 모델 개발 - 연구 결과를 토대로 논문 작성

4. 캡스톤디자인 추진전략 및 방법

1) 추진전략

- (1) 캡스톤디자인에 대한 이해
 - OVD 기술의 개념 및 기존 객체 탐지 모델과의 차이 분석
 - 드론 영상에서의 군용 객체 탐지의 중요성과 도전 과제 파악
 - 이미지 합성 기법 개념 및 기존 이미지 합성 모델 분석
- (2) 캡스톤디자인 경험
 - 이전 프로젝트에서 습득한 YOLO, Faster R-CNN 등의 객체 탐지 기술 활용
 - 멀티모달 학습 경험을 바탕으로 CLIP 기반 OVD 모델 최적화
- (3) 검증된 멘토 활용
 - 데이터 가공 서비스를 제공하는 회사인 데이터메이커의 나현우 연구원을 멘토로 섭외
- (4) 프로젝트 관리체계 수립
 - 역할 및 책임을 명확히 하여 프로젝트 수행을 체계적으로 관리
 - 연구 진행 상황을 주기적으로 점검하여 문제 해결 및 최적화 수행

2) 수행방법 및 추진절차

- (1) 데이터 수집 및 전처리
 - 드론 영상 데이터셋 구축 및 분석
 - 데이터 부족 문제 해결을 위해 이미지 합성 기법 활용
 - 현실적 데이터 생성을 위해 Image harmonization 기법 활용
- (2) OVD 모델 선정 및 최적화
 - 최신 OVD 모델 비교 분석을 통해 최적 모델 선정
 - 군용 객체 탐지에 적합하도록 모델을 최적화 및 성능 개선
 - 예) K2, K200, T80 BMP3, 수송차 등의 세부 객체 탐지 기능 개선
- (3) 모델 학습 및 검증
 - 다양한 드론 영상 데이터셋을 활용하여 모델 학습 진행
 - 탐지 정확도(mAP_50, mAP_75, mAP_s, mAP_m, mAP_l) 등의 정량적 지표를 활용하여 모델 평가
- (4) 실전 적용 가능성 검토
 - 군사 작전뿐만 아니라 재난 대응, 국경 감시 등의 분야에서 활용 가능성 평가
 - 연구 결과를 국제 학술지 및 컨퍼런스에 발표하여 학술적 기여

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

사용 프레임워크



5. 참고문헌

- [1] Redmon, J., & Farhadi, A. (2016). "You Only Look Once: Unified, Real-Time Object Detection." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 779–788.
- [2] Ren, S., He, K., Girshick, R., & Sun, J. (2015). "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks." Advances in Neural Information Processing Systems (NeurIPS), vol. 28.
- [3] Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A., & Zagoruyko, S. (2020). "End-to-End Object Detection with Transformers." Proceedings of the European Conference on Computer Vision (ECCV), pp. 213–229.
- [4] Tianheng Cheng, Lin Song, Yixiao Ge, Wenyu Liu, Xing-gang Wang, and Ying Shan. Yolo-world: Real-time open-vocabulary object detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 16901–16911, 2024
- [5] Yizhi Song, Zhifei Zhang, Zhe Lin, Scott Cohen, Brian Price, Jianming Zhang, Soo Ye Kim, and Daniel Aliaga. Objectstitch: Generative object compositing. In CVPR, 2023.
- [6] Linfeng Tan, Jiangtong Li, Li Niu, Liqing Zhang. Deep Image Harmonization in Dual Color Spaces. MM '23: Proceedings of the 31st ACM International Conference on Multimedia, pp. 2159 – 2167