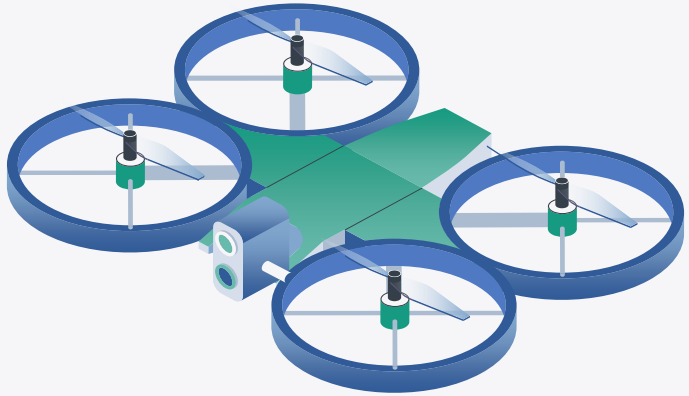
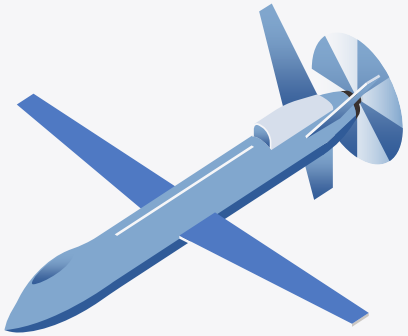


캡스톤디자인1 계획발표

항공영상 기반의 제로샷 객체 탐지 연구
(Zero-shot Object Detection Research Based on Aerial Imagery)



지도 교수: 장한얼 교수님

팀 명 : AerD(Aero Detector)

팀 원 : 컴퓨터공학과 20201735 박우진
컴퓨터공학과 20222019 김다빈

국립한밭대학교

CONTENTS

01

팀원 소개

01 팀원 소개

박우진

논문 및 자료조사,
모델 코드 작성 및 실험,
데이터셋 구축

김다빈

논문 및 자료조사,
모델 코드 작성 및 실험,
데이터셋 구축

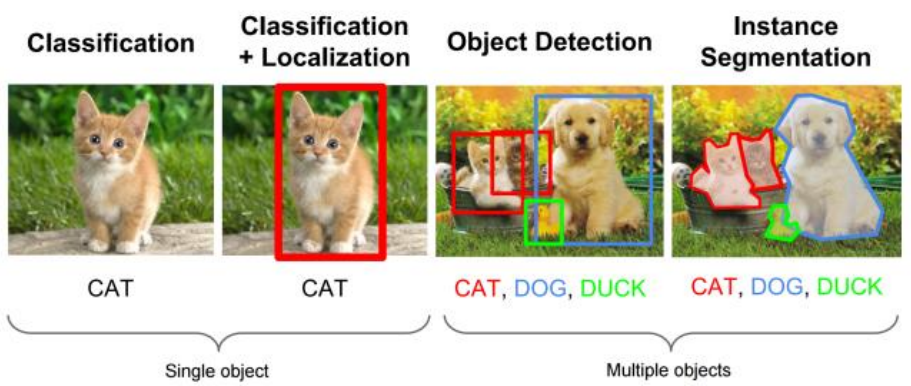
02

연구 배경 및 필요성

02 연구 배경 및 필요성

Object Detection이란?

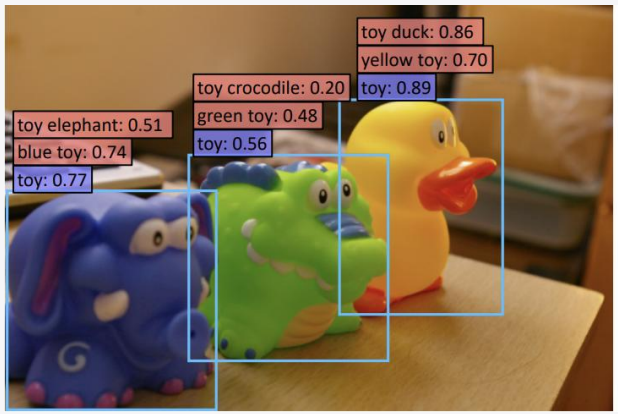
- **이미지나 영상에서 특정 객체의 위치(Bounding box)와 종류(Class)를 탐지하는 컴퓨터 비전 기술**
  - 객체의 위치와 클래스 식별
  - 바운딩 박스 출력



02 연구 배경 및 필요성

Open-Vocabulary Object Detection(OVD)란?

- **모델이 학습하지 않은 새로운 객체(Novel category)도 탐지할 수 있는 객체 탐지 기법**
  - 사전 정의된 클래스에 의존하지 않음
  - 텍스트 기반 객체 탐지
  - 제로샷 탐지



: Novel categories

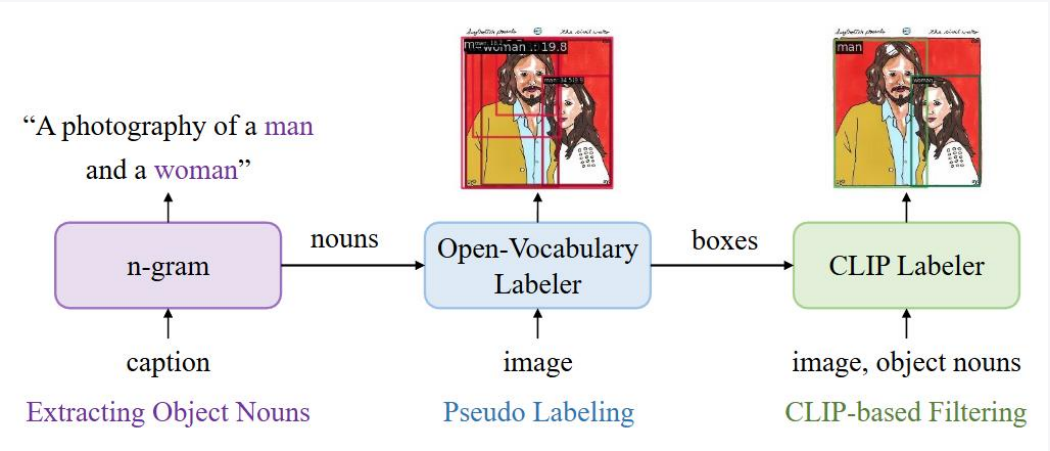
: Base categories

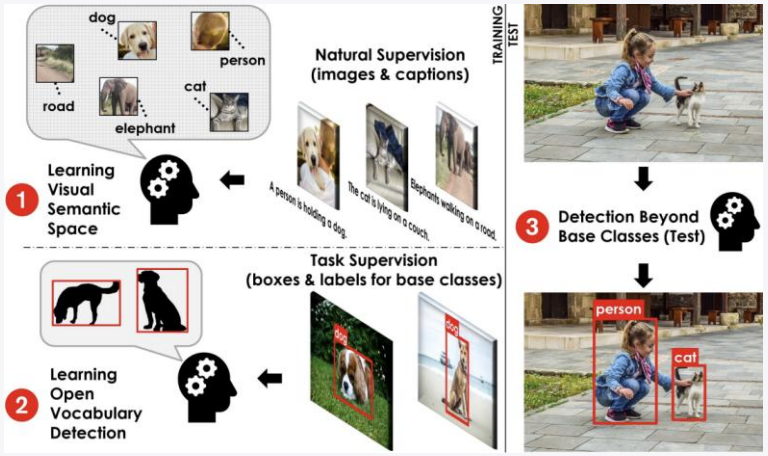
02 연구 배경 및 필요성

Open-Vocabulary Object Detection의 필요성

- **Object Detection 모델은 학습한 객체만 탐지 가능**
  → 추가 학습 필요

- **Open-Vocabulary Object Detection 모델은 학습하지 않은 객체도 탐지 가능**
  → 자연어 설명





02 연구 배경 및 필요성

Open-Vocabulary Object Detection의 필요성

- **자연어를 기반으로 새로운 객체를 탐지할 수 있음**
  → 다양한 도메인에 적응 가능

- **기존 탐지 모델보다 유연성이 뛰어나고, 빠른 도입이 가능**





03

기존 연구

03 기존 연구

YOLO-World

- YOLO backbone과 CLIP 임베딩을 활용하여 범용적인 객체 인식을 수행하는 것이 목표

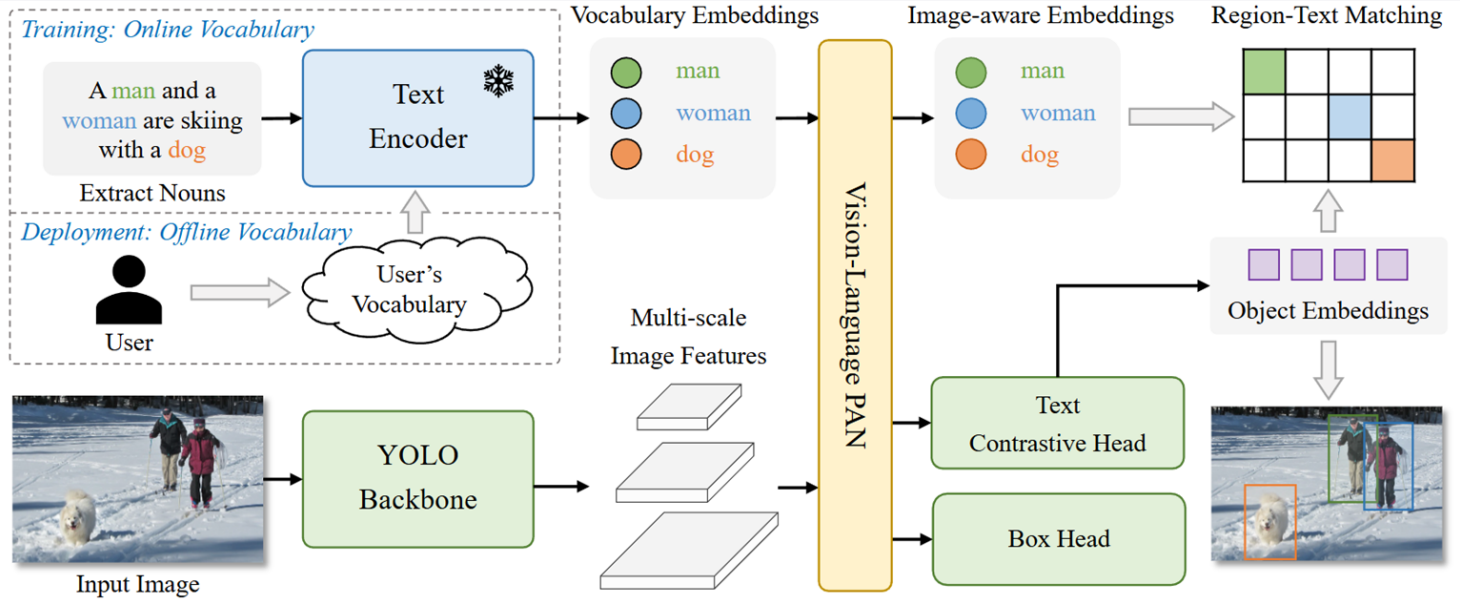


Figure 3. **Overall Architecture of YOLO-World.** Compared to traditional YOLO detectors, YOLO-World as an open-vocabulary detector adopts text as input. The *Text Encoder* first encodes the input text input text embeddings. Then the *Image Encoder* encodes the input image into multi-scale image features and the proposed *RepVL-PAN* exploits the multi-level cross-modality fusion for both image and text features. Finally, YOLO-World predicts the regressed bounding boxes and the object embeddings for matching the categories or nouns that appeared in the input text.

03 기존 연구

ObjectStitch

- Image guidance를 사용하여 생성된 합성 이미지에서 원래 객체의 정체성과 외형을 보존하는 것이 목표

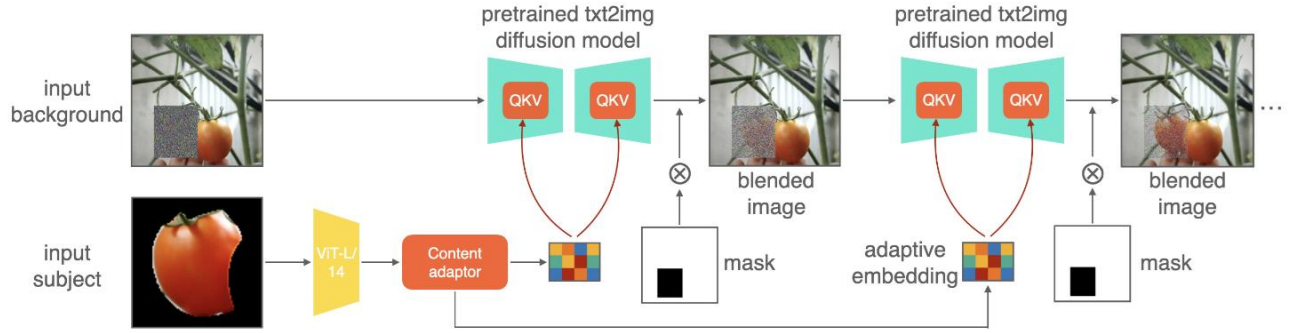


Figure 2. System pipeline. Our framework consists of a content adaptor and a generator (a pretrained text-to-image diffusion model). The input image I_o is fed into a ViT and the adaptor which produces a descriptive embedding. At the same time the background image I_{bg} is taken as input by the diffusion model. At each iteration during the denoising stage, we apply the mask M on the generated image I_{out}, so that the generator only denoises the masked area $I_{out} \otimes M$.

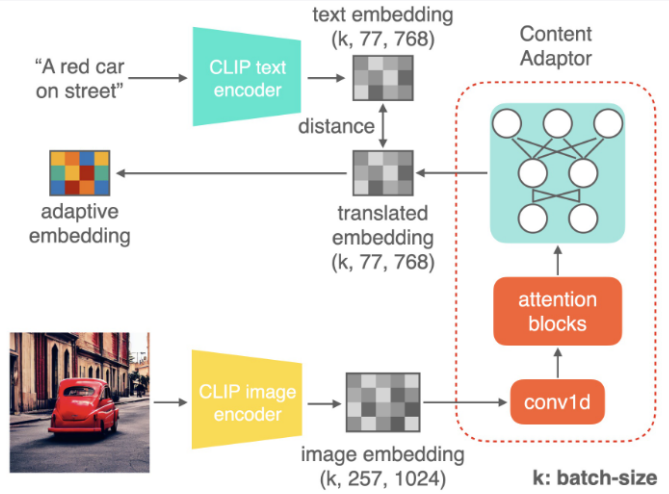


Figure 3. Structure of the Content Adaptor. In the first stage, it is trained on a large dataset of image-caption pairs to learn multi-modal sequential embeddings containing high-level semantics. In the second stage, it is fine-tuned under the diffusion framework to learn to encode identity features in *adaptive embedding*.

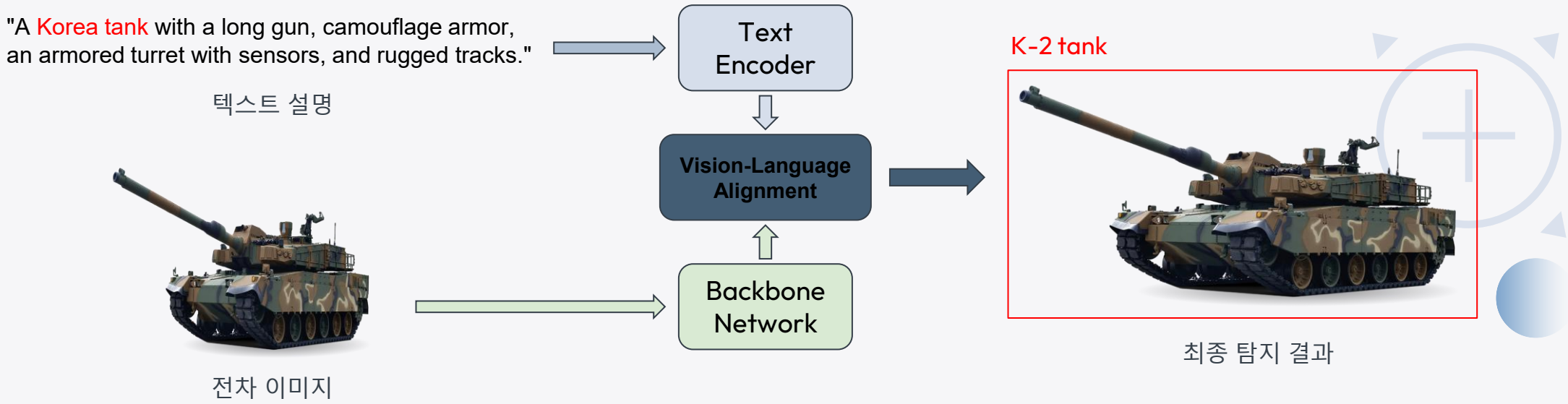
04

연구 목표 및 계획

04 연구 목표 및 계획

연구 목표

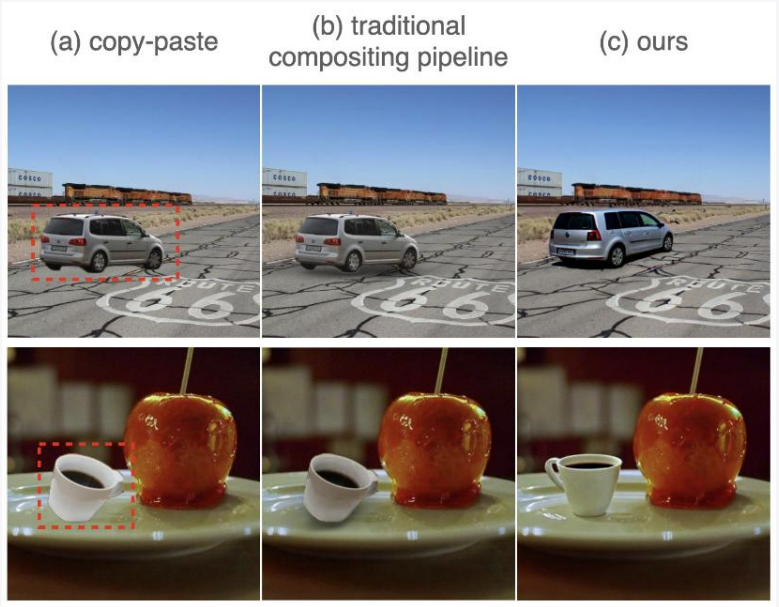
- **군용 차량 탐지에 OVD 기술 적용**
  - **OVD는 자연어 설명 활용, 새로운 객체 탐지 가능**
    → 군사 작전에서 제로샷 탐지를 활용한 정찰 및 상황 인식 강화
  - **현재 OVD 기술이 군용 객체 탐지에 최적화되어 연구된 사례는 부족**
    → 데이터셋 부족, 일반적 OVD 모델의 학습 도메인 문제

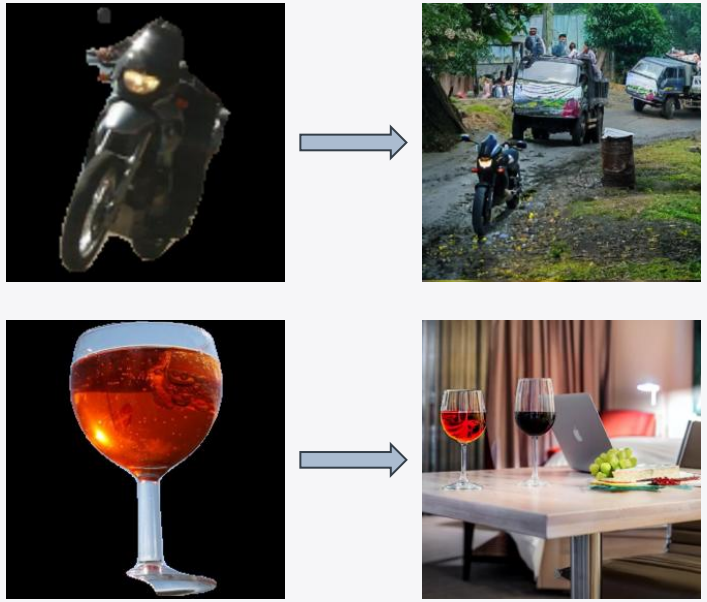


04 연구 목표 및 계획

연구 목표

- **이미지 합성 기법 사용**
  - 군사 데이터는 보안 문제로 대규모 수집 어려움
  - 이미지 합성 기법을 활용하여 현실적인 합성 데이터를 생성 및 적용





04 연구 목표 및 계획

연구 계획

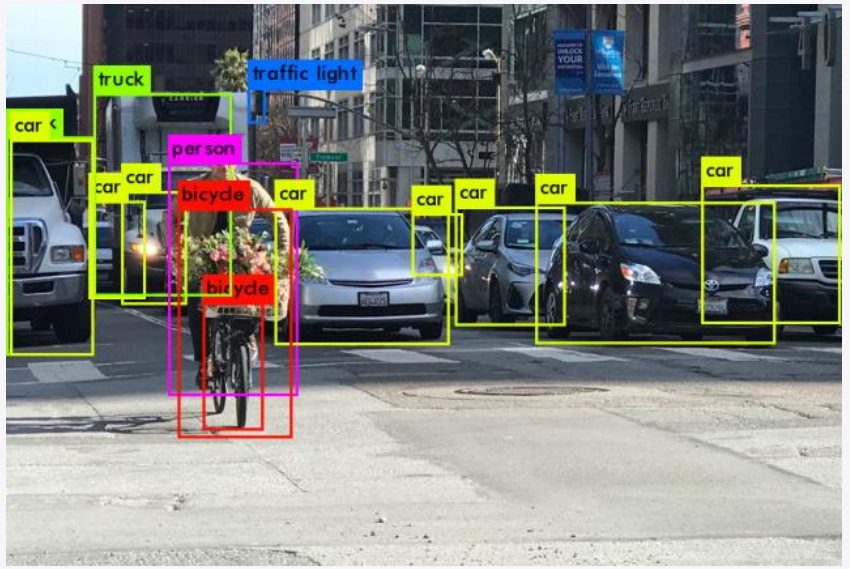
데이터 수집 및 전처리	• 드론 영상 데이터셋 구축 및 분석 • 데이터 부족 문제 해결을 위해 image composition 기법 활용 • 현실적 데이터 생성을 위해 image harmonization 기법 활용
OVD 모델 선정 및 최적화	• 최신 OVD 모델 비교 분석을 통해 최적 모델 선정 • 군용 객체 탐지에 적합하도록 모델을 최적화 및 성능 개선 • K2, K200, T80, BMP3, 수송차 등의 세부 객체 탐지 기능 개선
모델 학습 및 검증	• 다양한 드론 영상 데이터셋을 활용하여 모델 학습 진행 • 탐지 정확도(mAP_50, mAP_75, mAP_s, mAP_m, mAP_l) 등의 정량적 지표를 활용하여 모델 평가
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

05

연구 활용성

05 연구 활용성

- **OVD의 장점**
  - 자연어 설명을 활용하여 학습하지 않은 객체도 탐지 가능
- **OVD의 활용 가능성**
  - 군사 분야 → 새로운 군용 차량을 신속하게 탐지하여 적군의 변화 감시 가능
  - 자율주행 분야 → 새로운 장애물이나 위험 요소 탐지 가능



05 연구 활용성

연구 기대효과

성능

- 기존 모델 대비 탐지 정확도를 높인 모델을 개발
- OVD 모델을 활용하여 항공영상 기반 학습하지 않은 전차, 장갑차, 수송차 등의 군용 객체 탐지 수행

결과

- OVD 모델을 이용해 제로샷 방식의 군용 객체 탐지 모델 개발
- 연구 결과를 토대로 논문 작성
- 새로운 군용 차량을 제로샷 탐지 기법으로 탐지하여 군사 전략적으로 중요한 기술 발전에 기여

06 참고 문헌

- **Open-Vocabulary Object Detection**
  - 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
  - Amir Zareian, Shih-Fu Chang, Dongdong Yu, and Xiu Shen Wei. Learning open-vocabulary object detection via vision and language knowledge distillation. *In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 5186–5195, 2021.
  - Rohit Girdhar, Alireza Fathi, Zeynep Akata, and Ian Misra. Open-vocabulary object detection using captions. *In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 14393–14403, 2023.
- **Image Composition**
  - 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.
- **Image Harmonization**
  - 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

THANK YOU

컴퓨터공학과 20201735 박우진
컴퓨터공학과 20222019 김다빈