

2025 2학기

VQA 스터디 중간 공유회

김태경 / 박찬욱

목차



1. 스터디 프로젝트 소개



2. 데이터셋 구축 완료

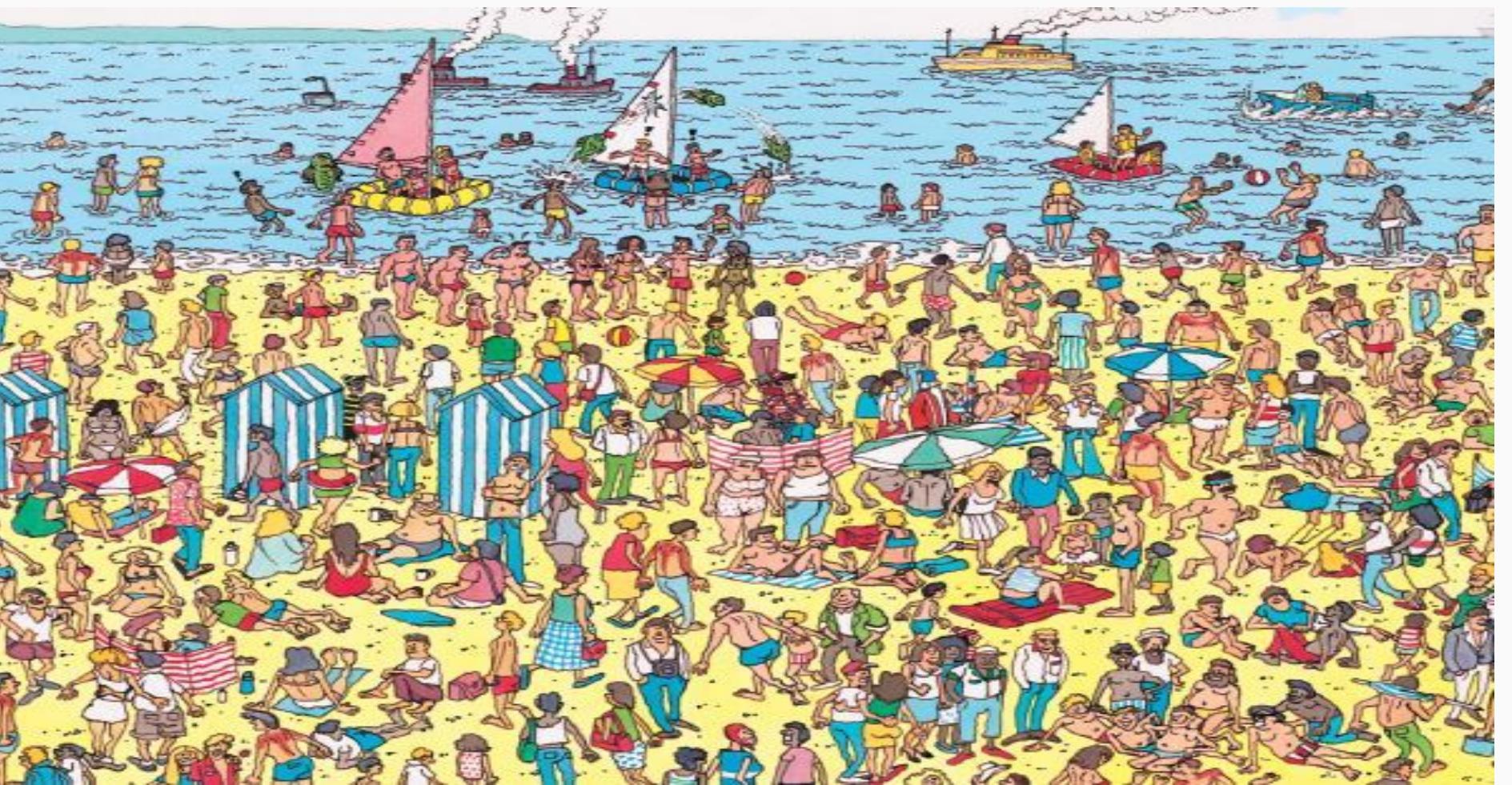


3. 모델 학습중



4. 향후 활동 안내

2학기 프로젝트

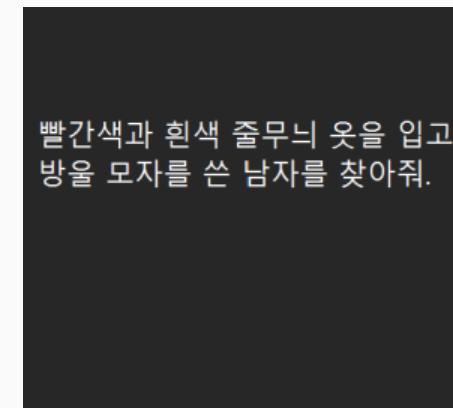


숨은그림을 찾고, 이에 대해 설명까지 해주는
만능 숨은그림찾기 VQA 모델

데이터셋 구성 완료

크게 데이터 수집 단계와, 데이터 어노테이션 단계로 구분

DATA COLLECTION

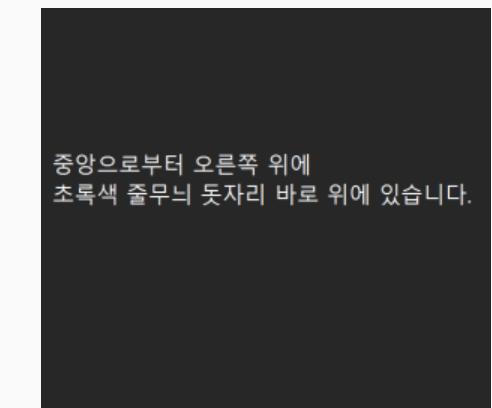


"월리를 찾아라" 책 구입 후 스캔을 통해 고화질의 이미지 데이터 확보 예정

이미지와 함께, 이미지에서 무엇을 찾을지에 대한 질문을 생성형 AI를 통해

각 이미지별로 30개씩 생성 예정

DATA ANNOTATION



수집한 이미지와 질문을 사용 가능한 하나의 데이터로 가공하는 과정

바운딩박스로 질문에 대한 해답의 위치를 표시

해답의 위치에 대해, 근처를 자세히 묘사한 답변을 제시

최종적으로 수집하고 가공한 데이터를 전부 하나의 JSON 형식으로 구성

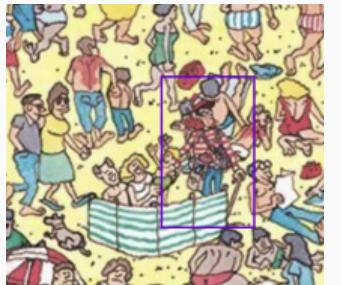
데이터셋 변경사항



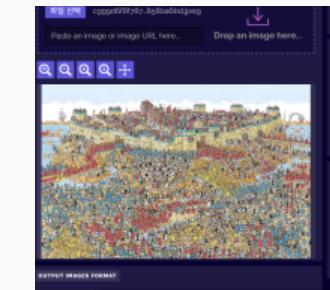
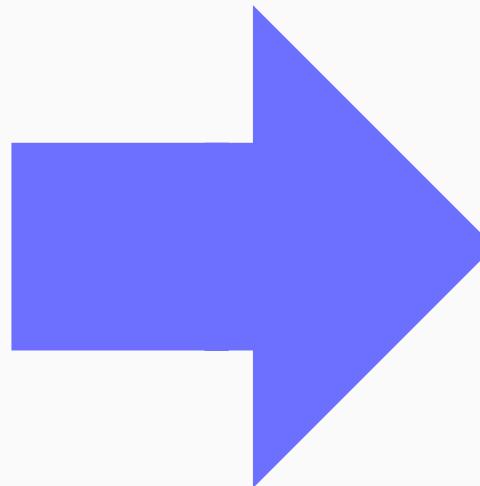
- 전체 이미지 입력
- 월리를 포함한 이미지만 입력



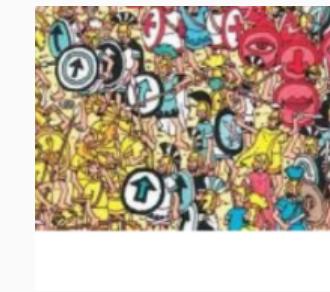
- 낮은 이미지 화질을 개선



- 데이터 불균형 문제
- 월리가 포함된 데이터와 포함되지 않은 데이터 간 차이가 많음



- 하나의 큰 이미지를 여러 정사각형으로 쪼개기
- 월리가 없는 이미지도 포함, 월리를 검출하지 못하는 경우도 만들기



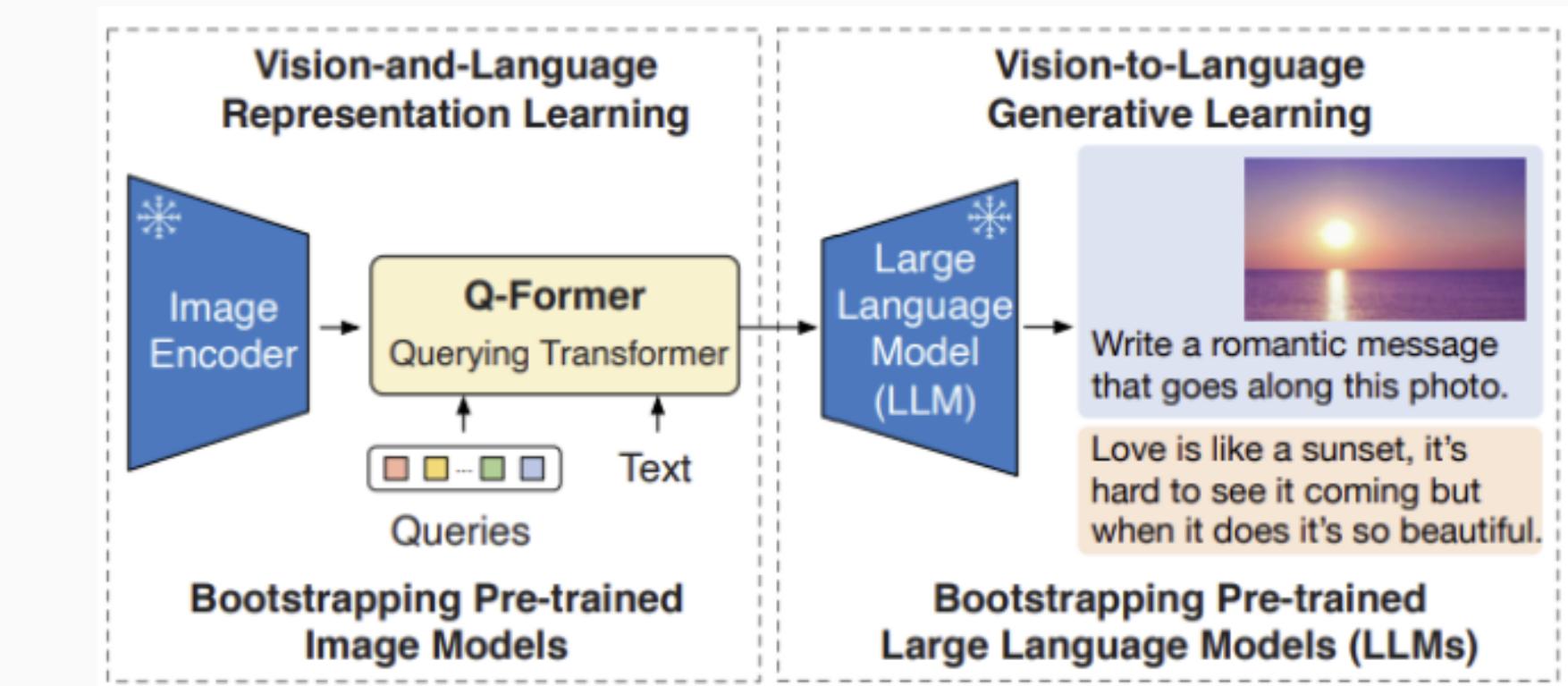
- 이미지 화질 증가 시 기존 이미지가 망가짐
- 저차원의 이미지가 다루기 더 쉬우므로 그대로 사용



- Data Augmentation 과정에서 월리 포함 클래스에 회전 / crop / flip 등을 가해 데이터 증강

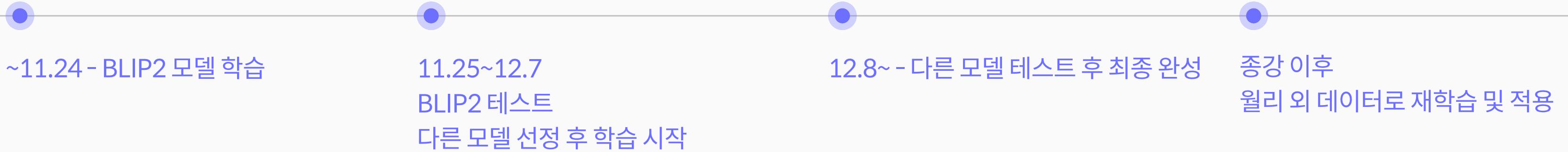
모델 구성

- 전체 학습 VS 파인 투닝
데이터가 여전히 많지 않았기에 파인튜닝 선택



- 모델 설정
지난번 대회의 연장선으로 BLIP2 기반 하나, 최신 모델 기반 하나 사용

향후 계획 설명

- 
- ~11.24 - BLIP2 모델 학습
 - 11.25~12.7
BLIP2 테스트
다른 모델 선정 후 학습 시작
 - 12.8~ - 다른 모델 테스트 후 최종 완성
 - 종강 이후
월리 외 데이터로 재학습 및 적용

Q&A