

프로젝트 수행 내용

프로젝트	분야	수행 기간
시각장애인의 온라인쇼핑을 위한 이미지 내 텍스트 인식 모델 개발	딥러닝, OCR	2024.03.11 ~ 2024.04.05 (4주)
사용 모델	사용 데이터	업무
Transformer 기반의 TrOCR	AI 허브의 건강기능식품 이미지 데이터셋	2인 프로젝트 60% 담당 (데이터 전처리, 모델 학습, 결과 분석)

목표

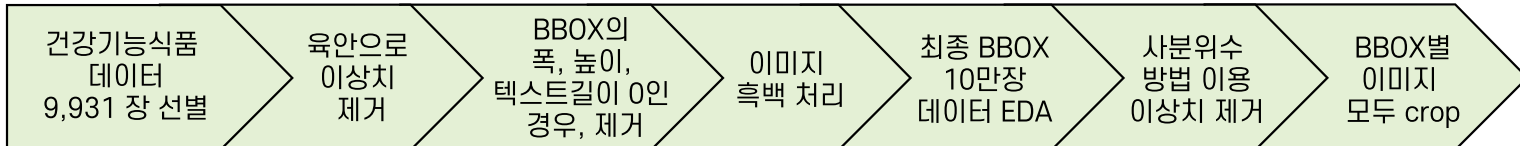
- 모델 Finetuning
- 제한된 시간과 자원에서 최대한의 결과내기
- 명확한 근거 있는 선택하기
- MLFlow로 모델 학습 관리 및 기록하기

배운 점 및 결과

- 모델의 최초 학습 dataset과 동일한 형태로 custom dataset을 구축하는 것이 성능 향상에 도움이 됨.
- Transformer 기반 모델과 MLFlow 연결의 난이도가 높아 실험 관리 실패하여 추후 재시도해볼 계획.
- 프로젝트 시작 전, 논문을 기반으로 모델 및 구성요소 (토큰라이저 등)을 확실하게 이해할 필요성 리마인드.

데이터 전처리

Crop 이미지 10만장 및 Annotation 파일 구축

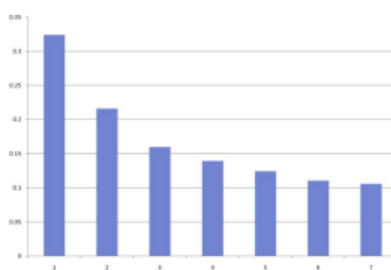


모델 학습

- 모델 선택 조건
 - 오픈소스 모델
 - 단일 모듈 구성 유무
 - 적절한 파인튜닝 난이도
 - 파인튜닝 관련 레퍼런스 유무
- Parameter 정보
 - 학습 이미지 수 : 40800장
 - 사용 모델 : Trocr-Small-printed
 - batch size : 4
 - epoch : 7
 - optimizer : AdamW
 - tokenizer : RobertaTokenizerFast

학습 결과

- 평가지표 : CER(character error rate, 잘못된 문자의 개수 / 총 문자의 개수)



[Validation CER per epoch]

- 데이터 42007장 Inference 결과, **82.7%** 텍스트 일치.

결과 분석

- 불일치 데이터 유형 분석
 - 휘어진 텍스트
 - 전처리 안된 잘못 잘린 이미지
- 텍스트의 음각 양각은 성능에 영향 미치지 않음.
- 동일한 텍스트의 학습 이미지 수가 많을수록 해당 텍스트에 대해 우수한 추론 성능을 보임.



[74개의 학습데이터가 있던 단어 '어린이의']