**Weekly report**

날짜 : **2022-10-13**

연구원 : **최윤석**

* **수행결과**
* **차량 이미지 분류 :** 
  + SSD 모델 학습
    - 20개의 카테고리별로 10개씩 뽑아 만든 데이터셋으로 학습 – 이미지 개수 축소
    - 2개의 카테고리에서 100개씩 뽑아 만든 데이터셋으로 학습 – 카테고리 개수 축소
  + 두 모델 다 축소하기 전 기존 모델(약 1.5만개의 이미지로 학습하여 생성)과 동일하게 학습이 진행되지 않는 에러가 발생하여 해결중
* **관계형 테이블 임베딩** :
  + 기존 윤종찬 연구원의 실험에 히트맵 방식을 적용해 임베딩 성능의 베이스를 파악
    - 272개의 데이터를 272\*272 히트맵에 나타내니 시각적으로 구분하기 어려워 히트맵에서 테이블 벡터의 코사인 유사도가 0.975 이상인 것만 추출

텍스트, 명판이(가) 표시된 사진

자동 생성된 설명

Figure . 두 테이블 벡터의 코사인 유사도가 0.975 이상인 테이블의 이름과 유사도 값을 추출

텍스트이(가) 표시된 사진

자동 생성된 설명

Figure . Figure 1에서 높은 유사도를 가진 두 테이블

* + - 윤종찬 연구원의 임베딩 벡터에서 코사인 유사도가 높은 경우 실제로 테이블들이 유의미한 관련이 있는 것을 확인
* **수행계획**
* **차량 이미지 분류** : 학습 도중 발생하는 오류 해결 및 모델 학습
* **관계형 테이블 임베딩** :
  + 수치형 데이터를 이산화한 것을 범주형 데이터로 바꾸는 임베딩방식을 보완하는 방법을 찾고 실험하여, 기존의 윤종찬 연구원의 연구와 성능 비교
* **기타사항**
  + **데이터 융합 프로세스**





혼합형 데이터셋 예시