**Weekly report**

날짜 : **2022-11-17**

연구원 : **최윤석**

* **이전 수행계획**
* **차량 이미지 분류 :**
  + 모델의 epoch를 10만 – 20만 사이에서 조정하여 각각의 mAP를 측정
* **관계형 테이블 임베딩** :
  + 수치형 데이터를 이산화한 것을 범주형 데이터로 바꾸는 임베딩 방식을 보완하는 방법을 찾고 실험하여, 기존의 윤종찬 연구원의 연구와 **융합 가능도 척도**를 사용하여 성능 비교
* **수행결과**
* **차량 이미지 분류 :** 
  + SSD 모델 학습 및 성능측정
    - SSD에서 적절한 mAP 값을 얻기 위해 epoch 값을 조정하여 실험 진행
      * 8만 epoch와 10만 epoch로 진행

텍스트이(가) 표시된 사진

자동 생성된 설명

Figure . epoch 8만개 - 각 카테고리별 average\_precision과 mAP의 값

텍스트이(가) 표시된 사진

자동 생성된 설명



Figure 2. epoch 10만개 - 각 카테고리별 average\_precision과 mAP의 값

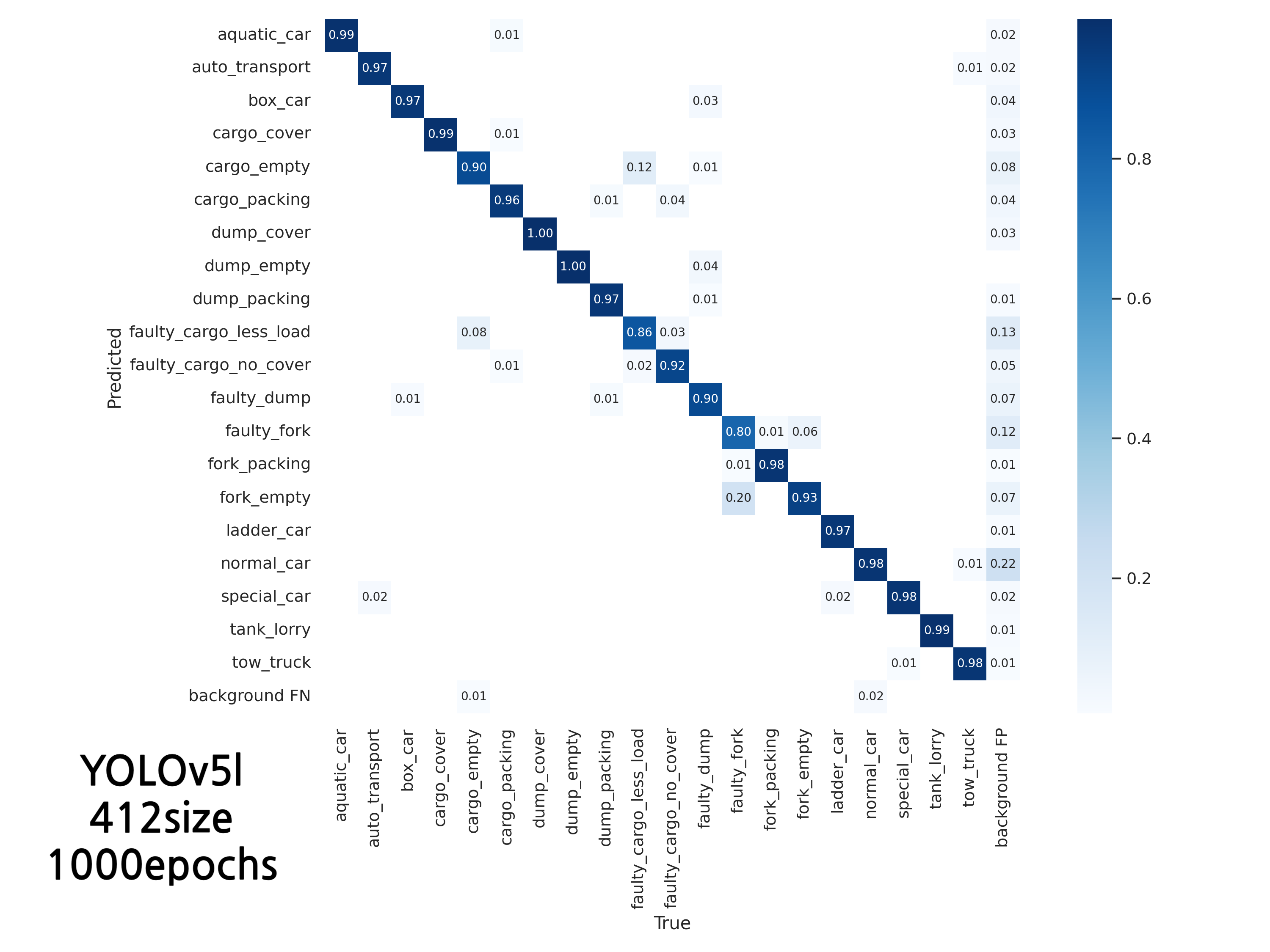


Figure 3. 김민중 연구원의 YOLO 모델 confusion matrix

* Epoch 10만개로 학습한 모델의 Average\_precision이 0.9보다 안 좋게 나온 라벨(4, 9, 12, 14)이 김민중 연구원의 confusion matrix에서 성능이 안 좋게 나온 라벨(4, 9, 12)과 비슷함.
* 김민중 연구원의 YOLO mAP는 0.96으로 SSD를 10만번 학습한 모델보다 약간 좋고, 사진에서 물체를 검출하는 속도도 YOLO가 장당 0.02초로 장당 1~2초 걸리는 SSD보다 우수한 것을 볼 수 있음.

Table 1. Train epoch 별 mAP

|  |  |  |  |
| --- | --- | --- | --- |
| epoch | 8만 | 10만 | 20만 |
| mAP | **0.903** | **0.954** | **0.963** |

* **관계형 테이블 임베딩** :
  + 관계형 테이블 임베딩 관련 호서대 ICT 보고서 작성
* **수행계획**
* **차량 이미지 분류** :
* **관계형 테이블 임베딩** :
  + 수치형 데이터를 이산화한 것을 범주형 데이터로 바꾸는 임베딩 방식을 보완하는 방법을 찾고 실험하여, 기존의 윤종찬 연구원의 연구와 **융합 가능도 척도**를 사용하여 성능 비교
  + 각 컬럼마다의 중요도를 고려해서 임베딩하는 방식과 수치형 컬럼에서 숫자의 크기가 크고 작음에 부여된 의미를 고려하는 임베딩 방식 연구
  + 혼합형 테이블 임베딩 관련 논문 공부
* **기타사항**
  + **데이터 융합 프로세스**



* **융합 가능도 척도**
  + 임베딩 거리 기준 값을 정하고 기준값보다 임베딩 거리가 가까우면 융합 가능한 데이터셋이라 판단함
  + 융합 가능한 데이터셋을 사람이 하나씩 확인하여 유의미한 융합이 되는 데이터셋을 찾음
  + 을 융합 가능도 척도라 정의함



혼합형 데이터셋 예시