**Weekly report**

날짜 : **2022-11-10**

연구원 : **최윤석**

* **이전 수행계획**
* **차량 이미지 분류:**
  + 모델의 성능 향상을 위해 epoch와 다른 요소를 변경하여 학습 진행
* **수행결과**
* **차량 이미지 분류 :** 
  + SSD 모델 학습 및 성능측정
    - 기존 2만개의 epoch로 모델을 학습했을 때 mAP 0.77 수준으로 성능이 만족스럽지 않아서 epoch를 20만개로 늘려서 학습을 진행
    - Figure 1을 보면 mAP가 0.963으로 많이 향상되었고, 20개중 12개의 카테고리에 대해 0.97 이상의 average\_precision이 나옴.

텍스트이(가) 표시된 사진

자동 생성된 설명

Figure . 각 카테고리별 average\_precision과 mAP의 값

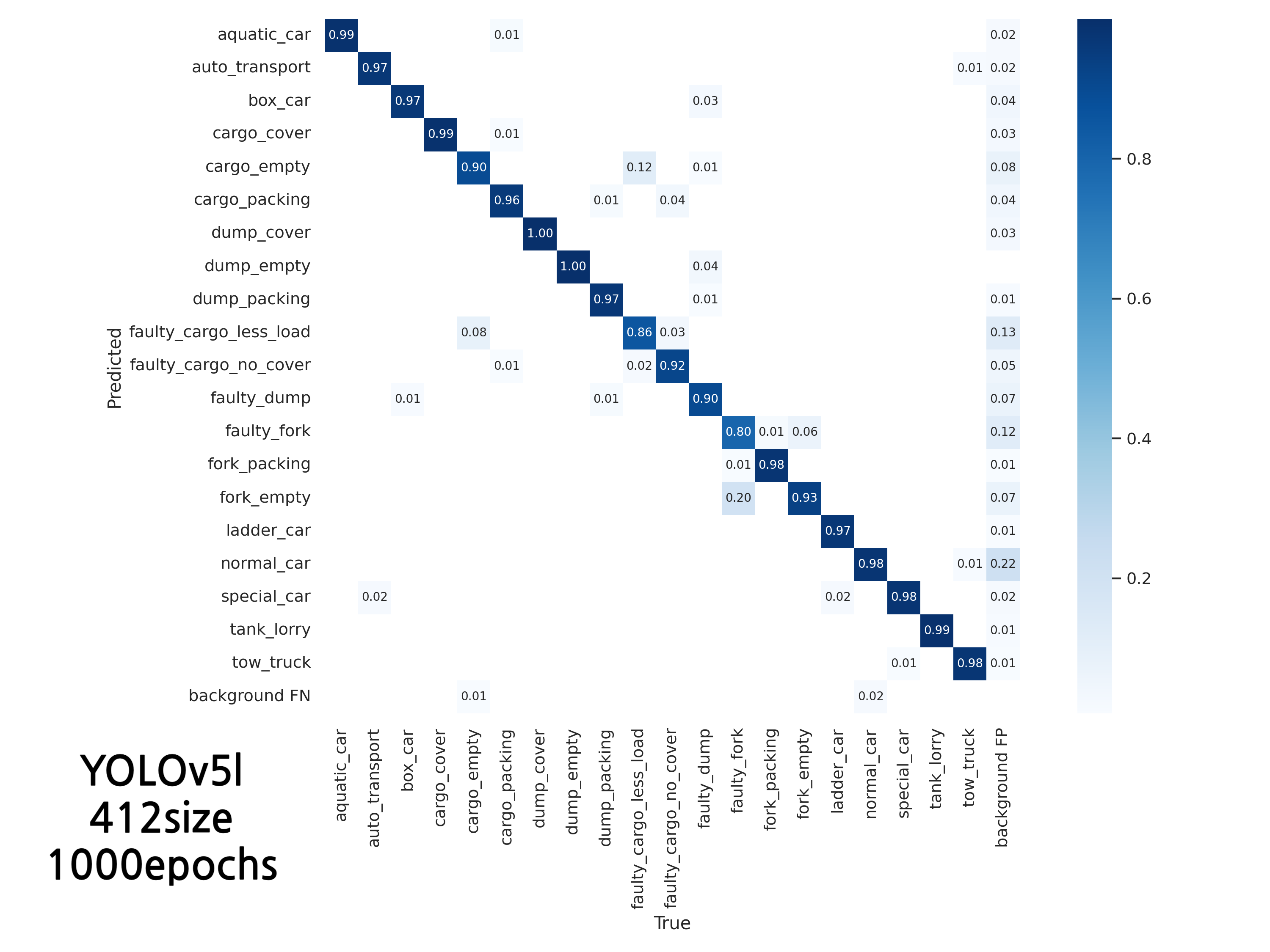


Figure . 김민중 연구원의 YOLO 모델 confusion matrix

* Average\_precision이 안 좋게 나온 라벨(2, 4, 5, 9, 10, 11, 12, 14)이 김민중 연구원의 confusion matrix에서 성능이 안 좋게 나온 라벨(4, 9, 10, 11, 12, 14)과 비슷함.

테이블이(가) 표시된 사진

자동 생성된 설명

Figure . SSD 모델에서 Averag\_precision이 안 좋게 나온 라벨의 국문명을 노란색으로 표시한 표

텍스트이(가) 표시된 사진

자동 생성된 설명

Figure 4. 라벨 19번 – epoch 2만번으로 학습한 모델이 견인차를 검출한 모습

텍스트, 상점이(가) 표시된 사진

자동 생성된 설명

Figure 5. 라벨 19번 – epoch 20만번으로 학습한 모델이 견인차를 검출한 모습

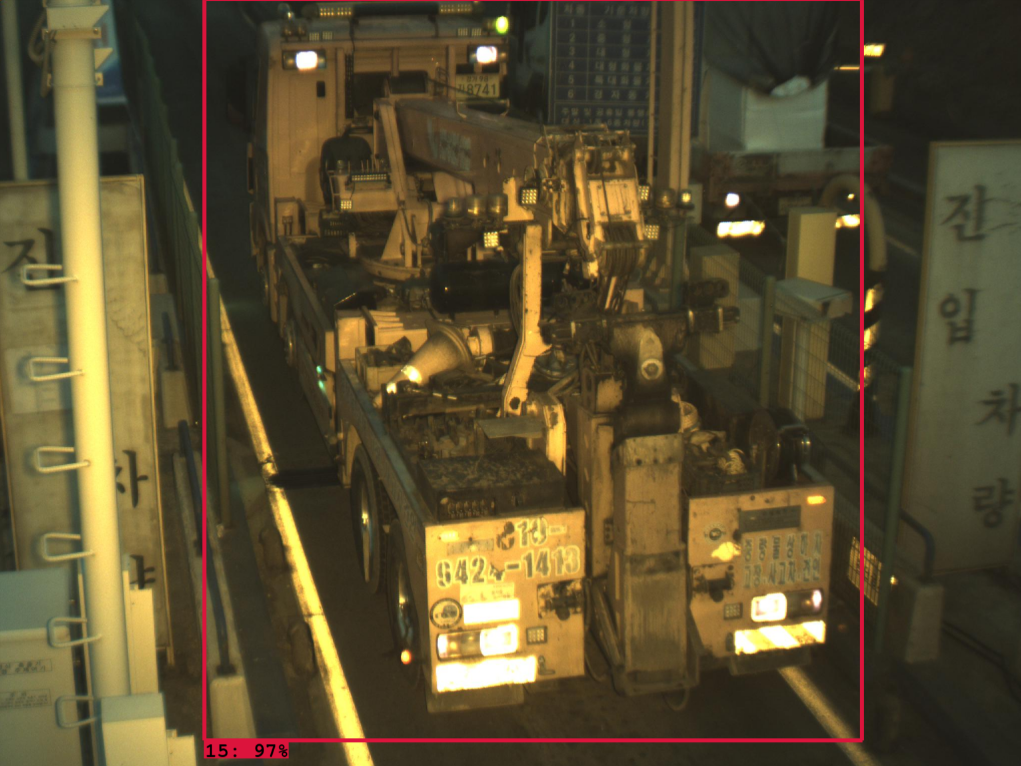


Figure 6. 라벨 17 - epoch 2만번으로 학습한 모델이 라벨 17을 라벨 15(사다리차)로 잘못 검출한 모습

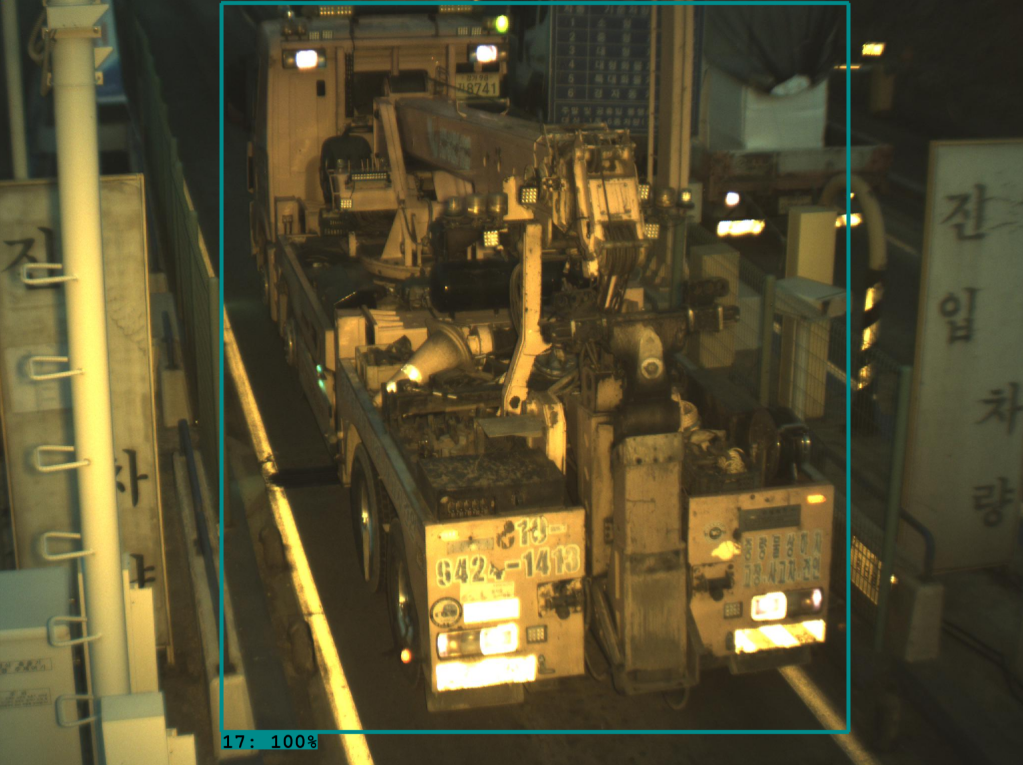
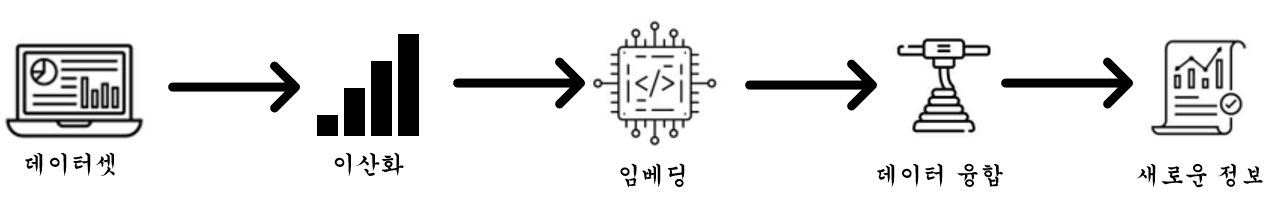


Figure 7. 라벨 17 - epoch 20만번으로 학습한 모델이 특수차(펌프카)를 검출한 모습

* + - 20만 epoch로 학습된 모델로 객체 검출을 진행. 각 라벨당 2개씩 뽑은 테스트 데이터 40개에 진행한 결과 bounding box를 제대로 잡아내는 것을 확인
    - 2만번 학습한 모델과 20만번 학습한 모델을 비교해보니 20만번 학습한 모델이 라벨을 더 정확히 분류하고 Bounding Box도 더 정교하게 잡는 것을 확인하였음.

Figure . 발생하는 오류 메시지

* **관계형 테이블 임베딩** :
  + SSD 오류 해결전까지 휴지
* **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
* **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