**딥러닝 OCR활용 차량개인정보 삭제**

**1. 프로젝트 개요**

중고차거래, 사고차량 처리, 폐차처리 등 차량사진을 이용하여 업무 처리를 하는 비즈니스 영역이 있는데, 업무를 마친 후에는 번호판, 사람얼굴, 전화번호, 차량등록증 내 주소 등 개인식별정보를 삭제해야만 함. 이러한 업무를 일일이 사람이 수작업 진행하면 인력이 많이 필요하므로 이를 AI로 자동화하면 편리할 것임

**2. 프로젝트 목적**

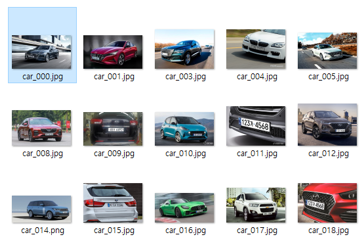
Yolov3 모델과 Open OCR API를 이용해서 차량사진에서 개인식별정보를 인식하고, 특히 차량등록증인 경우 OCR로 해당 내용을 점검하여 개인식별 정보를 삭제 처리하는 예제 프로그램을 개발하는 것을 목표로 함

**3. 프로젝트 내용**

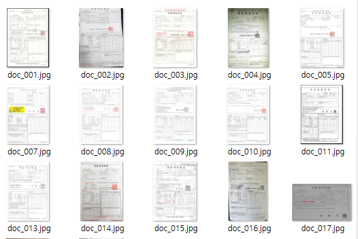
1) 데이터 수집

- 학습데이터는 관련 회사에서 제공한 데이터를 이용하였으며, 회사기밀상 해당 데이터는 학습모델에만 반영하고 본 프로젝트에서는 제공하지 않는 것으로 함

- 인터넷에서 조회하여 수집한 데이터는 제공 가능함



[그림 1] 차량사진 (번호판 인식용)



[그림 2] 차량등록증 사진 (주소정보 인식용)

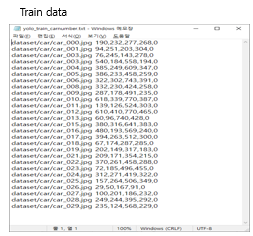
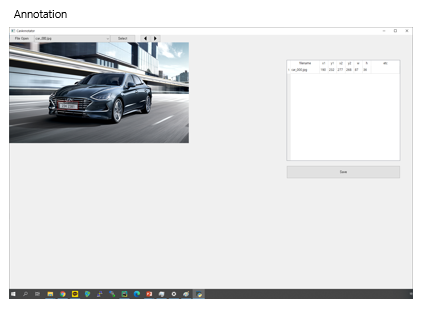
- 별도의 annotation tool로 수집된 데이터에서 인식할 영역을 찾아 yolo모델 학습용 데이터를 구성함

2) Yolov3 Training

- Detection model은 Keras2.2.4 기반 yolov3 모델을 사용함

- Pre-trained 모델을 사용하여 학습의 효율성을 높임

- 기본 모델코드에 모델경로, annotation정보(학습데이터), class정보, yolo\_anchors.txt 경로를 설정하고 학습 진행

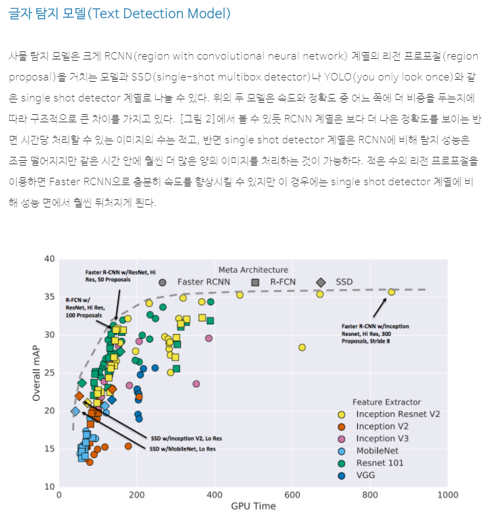


[그림 3] annotation tool과 생성된 yolov3 학습데이터

3) OCR 원리

- OCR은 글자 탐지 모델(Text Detection Model)과 글자 인식 모델(Text Recognition Model)로 이루어짐

- 글자 탐지 모델은 글자로 추정되는 영역을 찾는 것을 말하며, 글자 인식에 특화된 일종의 Detection모델로 볼 수 있음



[그림 4] Object Detection Model성능 비교

- 글자 인식 모델은 글자 영역에서 실제 어떤 글자가 씌어있는지 인식하는 것으로 RNN/LSTM계열 모델과 Attention 개념을 적용하여 정확도를 높이고 있음

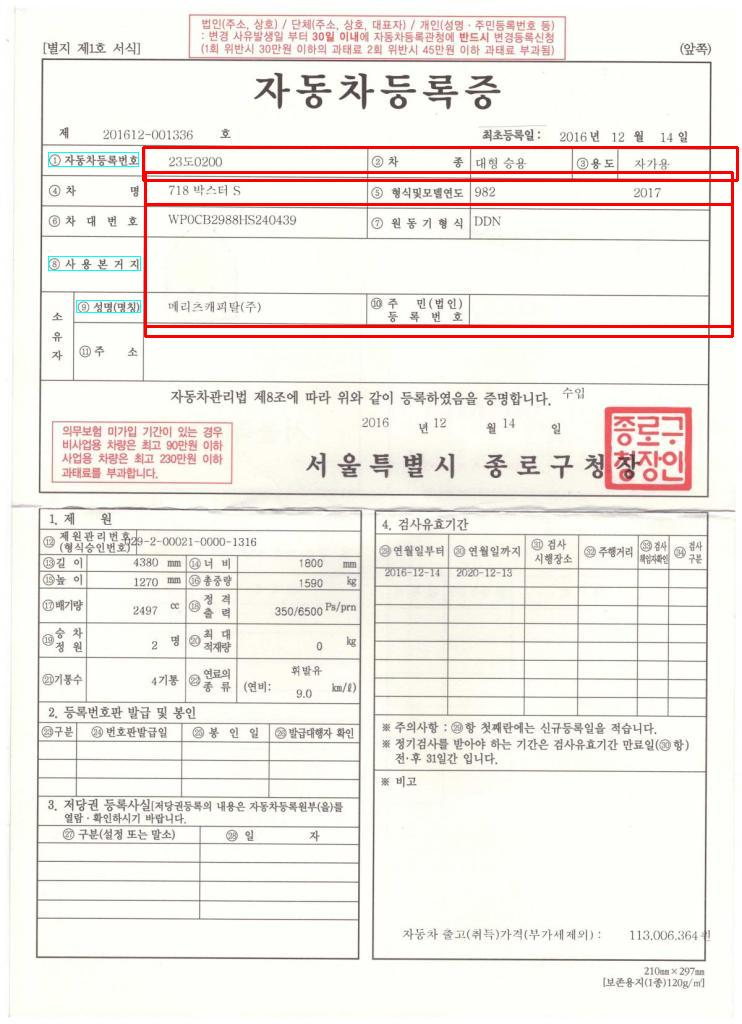


[그림 5] 글자 인식 모델 구조

- 본 프로젝트에서 OCR모델을 처음부터 새로 만들기는 어렵고, open-API중에 무료로 제공되는 Kakao-OCR을 이용하여 인식하고자 함 (성능은 NAVER-OCR이 더 뛰어나지만 NAVER는 유료임)

4) **학습결과**

개인식별정보에 해당하는 차량번호판과, 차량등록증 내 등록번호, 주소, 성명 등을 인식하여 삭제 가능함



[그림 6] 차량번호판과 차량등록증내 개인식별정보 인식 결과

**4. 주요 기술**

- 언어 : Python

- 모델 : Yolov3, Kakao-OCR api

- 프레임워크 : Keras2.2.4

**5. 주요 소스코드**

- Yolov3 학습 부분



- 종합테스트 부분

