

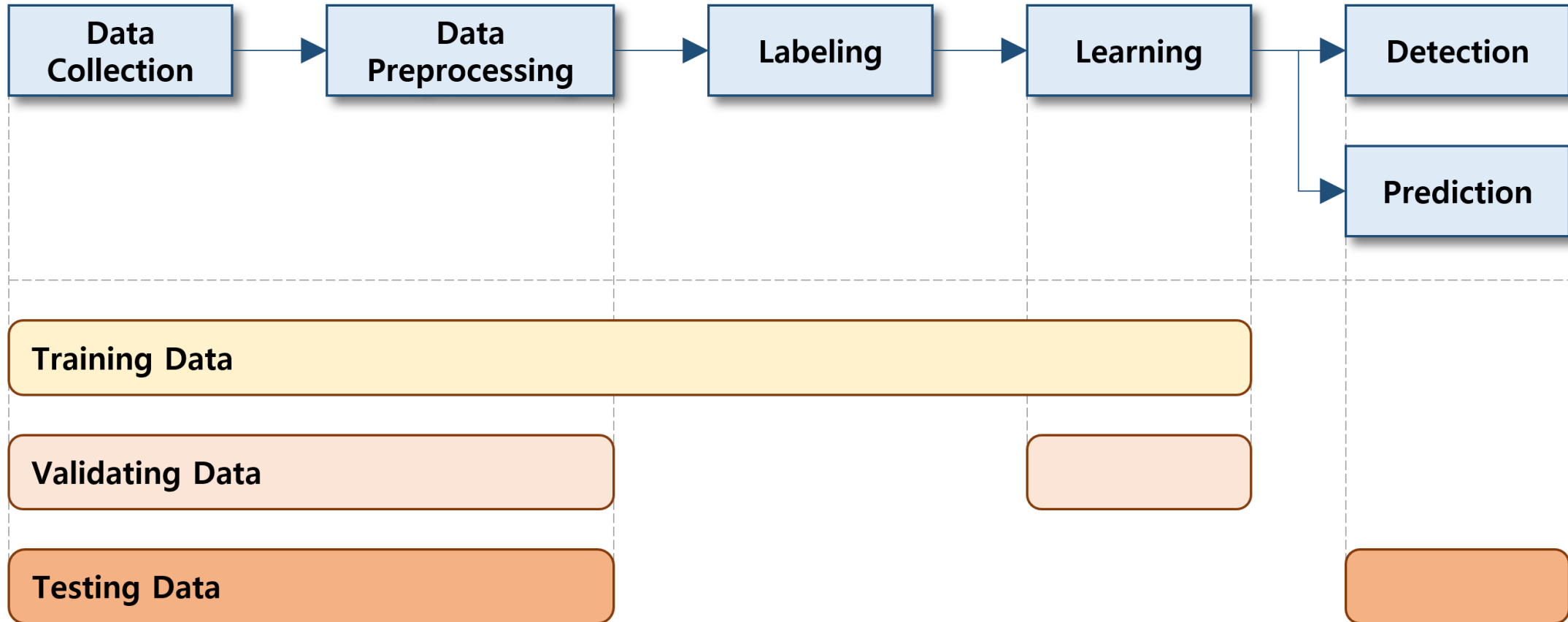
AI 고급 : 시각 처리 과정(정규)

데이터 레이블링



- 딥러닝에서 가장 중요한 것은 데이터(특히 학습을 위한 데이터)
- 같은 계열의 목적을 가진 모델이라면
 - 코드의 변형없이 데이터만 교체해서 서로 다른 분야에 적용할 수 있음
- CNN을 비롯한 영상인식, 분류 등을 목적으로 하는 딥러닝 모델은
 - 학습의 결과가 옳은지, 그른지 확인할 수 있도록 각각의 영역을 구분하고 표시를 해 둠
 - 이 작업 과정이 **레이블링** 작업 과정

이미지 인식 처리에서 데이터 셋의 적용 범위



정보가 없다

MNIST ?

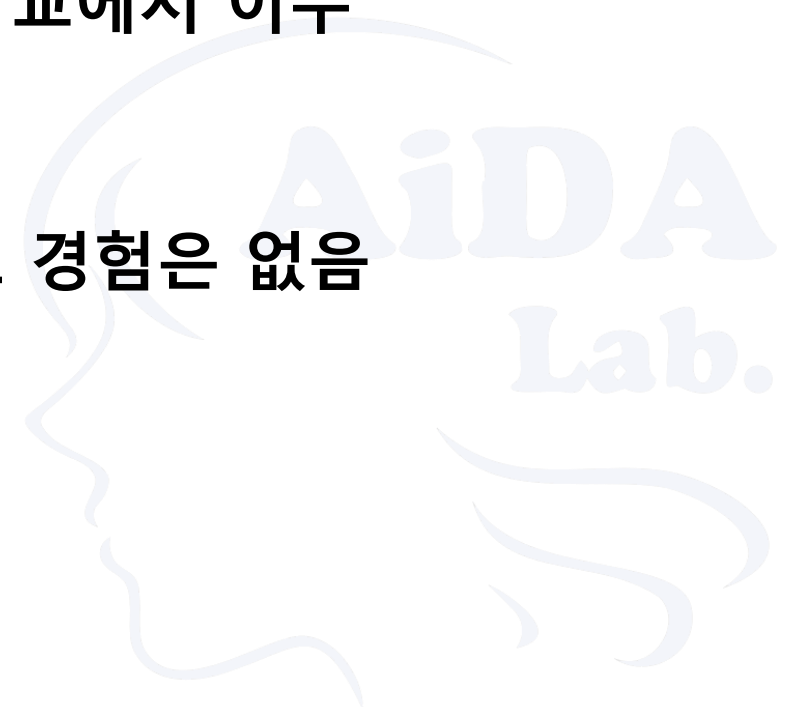
Dataset
Download

ImageNet ?

A
ib.

• 참여자

- 포항공대 2~3학년 인턴사원 3명
- AI / 프로그래밍에 대한 기본적인 교육, 실습은 학교에서 이수
- 데이터셋 개발 경험 없음
- 실습 프로젝트 경험은 있으나 실무 관련 프로젝트 경험은 없음



- 필요한 데이터

- 번호판을 장착한 자동차의 사진
- 자동차의 사진에서 번호판 영역에 적용된 레이블링 데이터
- 다양한 형태의 번호판 사진
- 기타 학습 조건들...



1차 시도: 이미지데이터 수집



- Google 검색을 통하여 약 50개의 차량 이미지 수집
- 전체 프로세스에 대한 과정을 처음 시도함
- 진행 방법이 올바른지에 대한 확인은 불가능한 상태



1차 시도: 이미지 레이블링 작업

1차

- 레이블링 도구: Labellmg

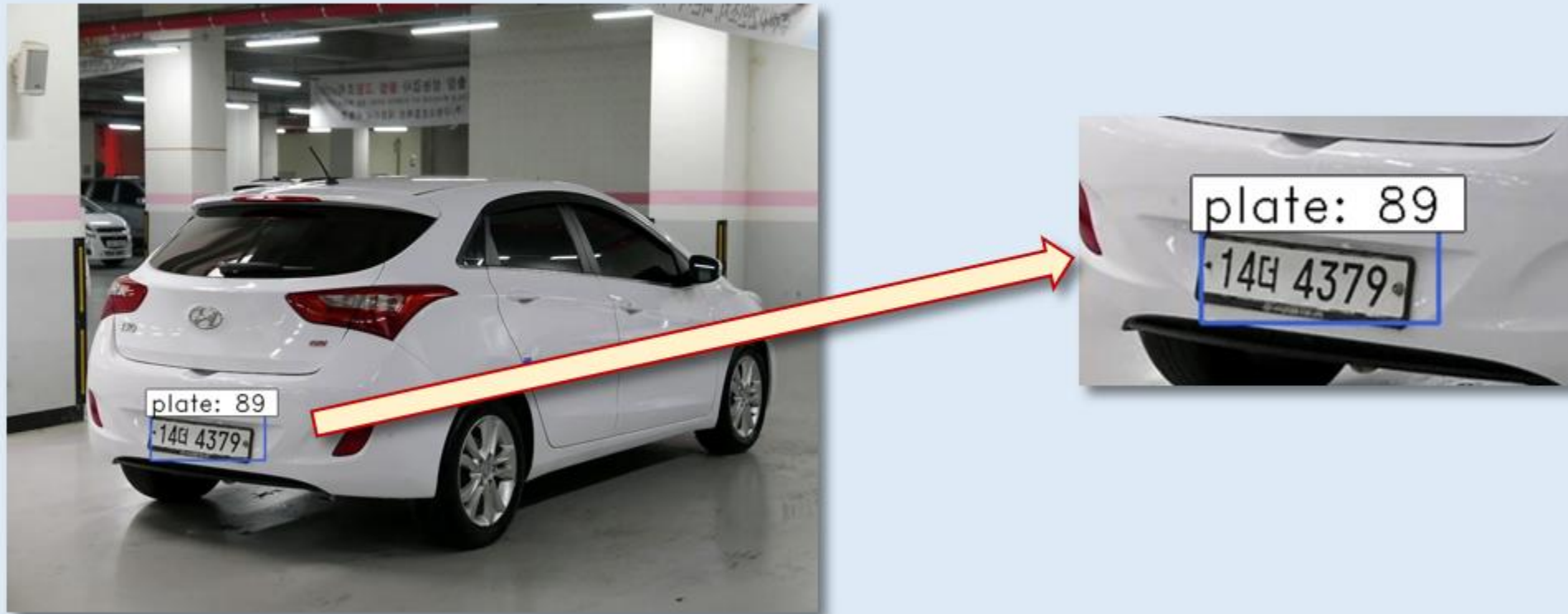


```
<annotation>
  <folder>CarNumber</folder>
  <filename>001.jpg</filename>
  <path>/training/001.jpg</path>
  <source>
    <database>Unknown</database>
  </source>
  <size>
    <width>980</width>
    <height>551</height>
    <depth>3</depth>
  </size>
  <segmented>0</segmented>
  <object>
    <name>plate</name>
    <pose>Unspecified</pose>
    <truncated>0</truncated>
    <difficult>0</difficult>
    <bndbox>
      <xmin>330</xmin>
      <ymin>321</ymin>
      <xmax>744</xmax>
      <ymax>386</ymax>
    </bndbox>
  </object>
</annotation>
```


1차 시도: Fast RCNN 적용 → Detection

1차

- Faster R-CNN 적용 테스트 성공



1차 시도: YOLO 적용 → Detection

1차

- YOLO 적용 테스트 실패



1차 시도: 실패 원인 파악



- 데이터 부족을 의심함
- Fast RCNN 모델에서 인식을 성공하였으므로 작업의 방향성은 맞는 것으로 간주함



2차 시도: 데이터 추가 + 레이블링 + 학습



- 중고차 매매 사이트를 통해 100개의 고화질 이미지 확보
- 새로 추가된 각 이미지에 대하여 레이블링 작업 수행
- 100개의 이미지를 이용하여 1,000번의 학습 수행



2차 시도: YOLO 적용 → Detection



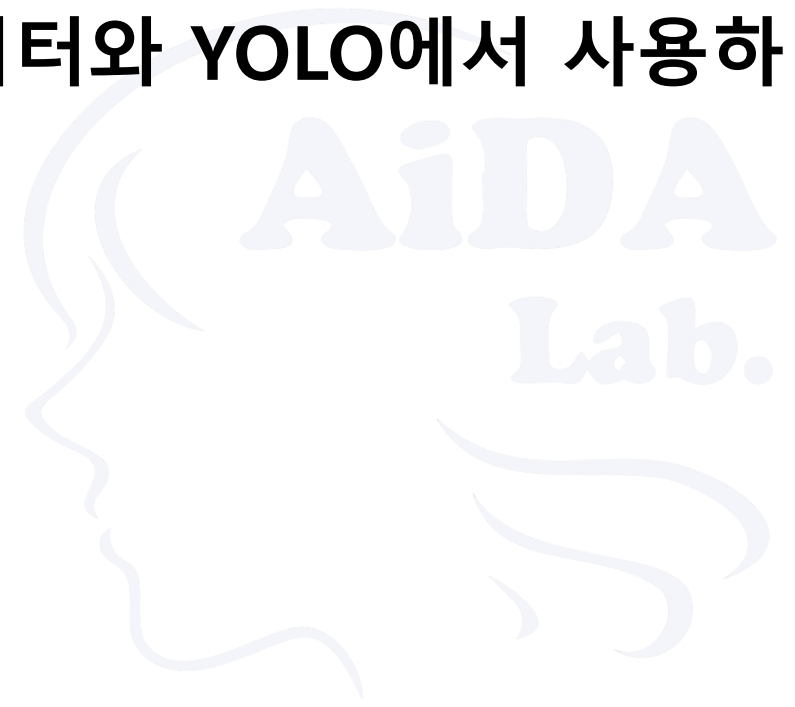
- 번호판 인식은 되지만 성능이 열악함
- 인식 영역이 조금씩, 또는 꽤 어긋나거나 잘못 그려지는 경우 발생



2차 시도: 실패 원인 파악



- 데이터 부족이 원인일 가능성 제시
- 또는 학습 횟수의 부족이 원인일 가능성 제시
- 기존 모델(CNN)에서 사용하는 레이블링 데이터와 YOLO에서 사용하는 데이터의 형태가 다름을 확인



2차 시도: 실패 원인 파악

2차

- YOLO 적용 테스트 실패

Labelling 용 XML

```
<annotation>
  <folder>CarNumber</folder>
  <filename>001.jpg</filename>
  <path>/training/001.jpg</path>
  <source>
    <database>Unknown</database>
  </source>
  <size>
    <width>980</width>
    <height>551</height>
    <depth>3</depth>
  </size>
  <segmented>0</segmented>
  <object>
    <name>plate</name>
    <pose>Unspecified</pose>
    <truncated>0</truncated>
    <difficult>0</difficult>
    <bndbox>
      <xmin>330</xmin>
      <ymin>321</ymin>
      <xmax>744</xmax>
      <ymax>386</ymax>
    </bndbox>
  </object>
</annotation>
```

Yolo 용 TXT

0 0.611328 0.700521 0.400391 0.640625

- 다수의 사이트에서 변경없이 사용할 수 있는 것으로 설명하거나 관련 설명 자체가 누락됨

2차 시도: 실패 원인 파악

2차

- YOLO 적용 테스트 실패

Labelimg 용 XML

```
</source>
<size>
  <width>980</width>
  <height>551</height>
  <depth>3</depth>
</size>
<segmented>0</segmented>
<object>
  <name>plate</name>
  <pose>Unspecified</pose>
  <truncated>0</truncated>
  <difficult>0</difficult>
  <bndbox>
    <xmin>330</xmin>
    <ymin>321</ymin>
    <xmax>744</xmax>
    <ymax>386</ymax>
  </bndbox>
</object>
</annotation>
```

Yolo 용 TXT

0 0.611328 0.700521 0.400391 0.640625

이미지 크기를 기준으로 하여
0~1 값으로 정규화, 파일 변환

3차 시도: 데이터 추가, 학습횟수 증가



- 1,675개의 이미지를 확보하여 2,000번의 학습 수행
- 기존의 이미지 레이블링 데이터 → YOLO 용 데이터로 변환



3차 시도: YOLO 적용 → Detection

2차

- YOLO 적용 테스트 실패

변환된 데이터는
제대로 적용되지 않음



3차 시도: 실패 원인 파악

- 기존 레이블링 데이터를 YOLO 형태로 변환한 것이 제대로 동작하지 않았을 가능성 제시
- 레이블링 작업을 처음부터 새로 하기로 결정
- 인터넷 검색을 통하여 YOLO 전용 레이블링 도구 검색

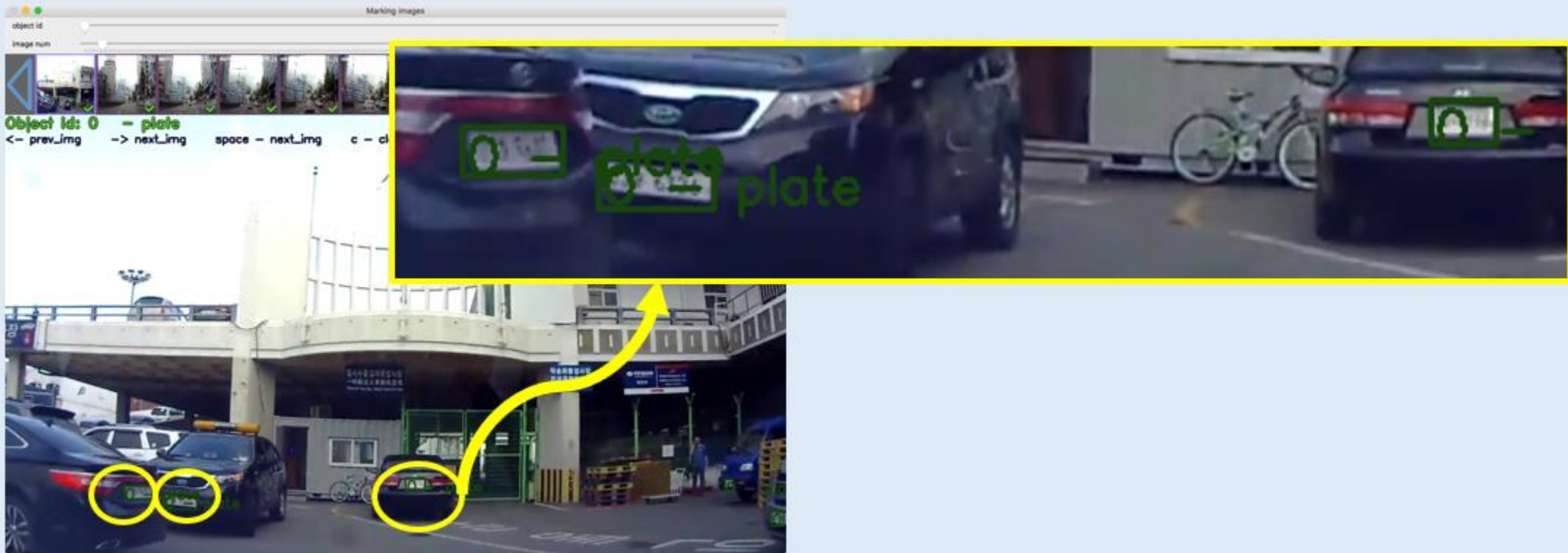


3차 시도: 레이블링 재작업

3차

- 레이블링 도구: Yolo-Mark

인터넷 검색을 통해 새로운 도구 발견



3차 시도: YOLO 적용 → Detection

3차

- YOLO 기반 차량번호판 인식 성공



3차 시도: YOLO 적용 → 동영상



실패



- 1,675개의 이미지를 이용하여 2,000번의 학습 수행
 - 임계 값: 80% → 동영상에 포함된 차량번호판의 약 10%정도 인식
 - 차량번호판이 아닌 유사 패턴 발견 시 → 차량번호판으로 오인식
 - 인식 성공률은 낮지만 대체로 안정적 (꾸준히 10%대...)
 - 정지영상(이미지)에 대한 인식 성공률은 약 90% → 안정적
 - **증가한 학습대상 이미지에 비하여 지나치게 적은 학습 횟수가 원인일 수 있다고 판단 → 학습 횟수를 늘리기로 결정**

4차 시도: 학습 횟수 증가

- 1,675개의 이미지를 이용하여 45,000번의 학습 수행
 - 임계값: 80% → 정지영상 인식 성공 (성공률 약 90%), 동영상 인식 실패
 - 임계값: 30% → 동영상에서 일부 차량번호판 인식 성공
 - 인식 기대값이 너무 낮게 계산됨 → 처음부터 인식을 시도하지 않음
- 너무 선명한 이미지만을 학습대상으로 사용한 것이 원인일 가능성 제시
 - 동영상에서 캡처한 이미지를 일부 학습에 적용하기로 결정

4차 시도: 데이터 추가 수집(동영상에서)

- 2,013개의 이미지를 이용하여 2,000번의 학습 수행
 - 동영상에서 캡처한 338장의 이미지 대상 → 레이블링 작업 추가 진행
 - 임계값: 50% → 인식 실패
 - 임계값: 30% →
 - 시도(4)보다 인식 수량 증가 확인
 - 중복 인식 증가
 - 유사패턴에 대한 오인식 증가
 - 어긋난 위치에 대한 인식 표시 증가

학습 횟수의 증가가 필요하다고 판단

4차 시도: 학습 횟수 증가

- 2,013개의 이미지를 이용하여 45,000번의 학습 수행
 - 임계값: 60% → 동영상에서도 웬만한 경우는 인식 성공(기대값 평균 65%~86%)
 - 차량번호판이 아닌 유사 패턴 발견 시 → 오인식 거의 없음
 - 거의 정확한 번호판의 위치 표시



정지영상에 대한 인식 결과



(92%)



(90%)



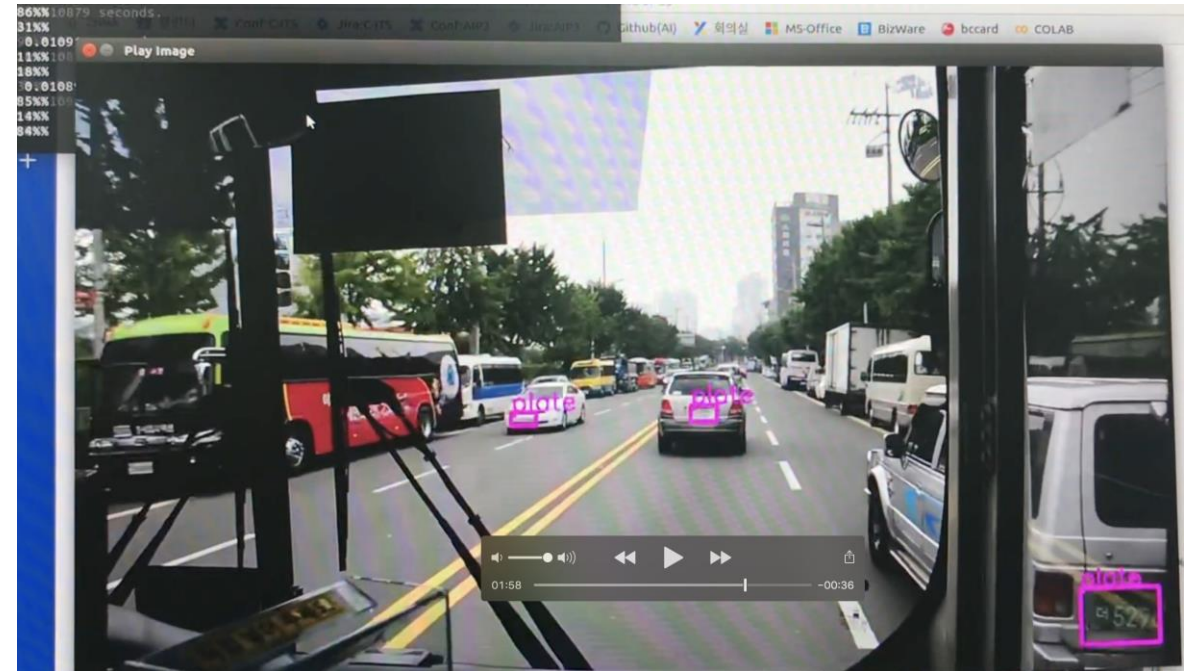
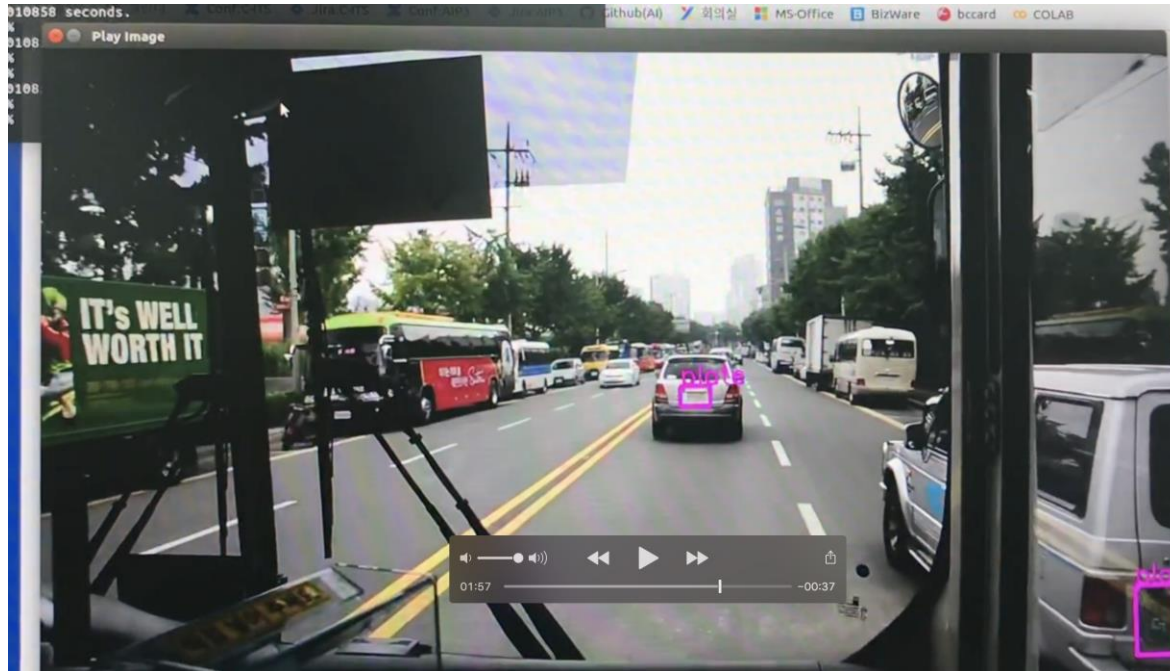
(89%, 81%)



(90%)

DA
Lab.

동영상에 대한 인식 결과



- 일부만 표시된 차량번호판도 인식 성공

- Labeling 영역의 지정에 대한 규칙 필요

- 인식하고자 하는 대상과의 거리
- 시선의 초점으로부터 허용되는 좌우 범위(각도)
- 시선의 초점으로부터 허용되는 상하 범위(각도)
- 인식 대상에 대한 선명도(흐릿함)의 허용 범위
- 인식 대상에 적용되는 밝기의 허용 범위
- 인식 대상의 크기는 다양하게 적용 가능하도록 이미지 선택

•정확한 수치 규정일 필요는 없음
•대략적인 느낌을 통해 적용

결론: Labeling 작업에 대한 사전 고려사항

- 예 1



결론: Labeling 작업에 대한 사전 고려사항

• 예 2



결론: Labeling 작업에 대한 사전 고려사항

• 예 3

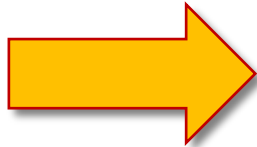


결론: Labeling 작업에 대한 사전 고려사항

• 예 4



이런 이미
지만 학습
하면



이런 크기의 대상
은 인식하지 못하
는 경우 증가

- 영상의 전처리 범위에 대한 규칙 필요
 - 입력 영상의 주변 환경 (실내, 실외, 맑음, 흐림 등)을 기준으로 입력 영상에 대한 표준 밝기
 - 입력 영상의 해상도를 기준으로 하는 이미지의 표준 크기
 - 그 외, 특정 환경에서의 인식이 필요한 경우, 해당 환경에 맞는 밝기, 크기 등의 사전 규칙 결정

- 컴퓨터 비전 데이터에 대한 모든 것 (superb-ai.com)



4년이 지났는데도 레이블링 작업흐름이 아직 "해결"되지 않았습니다. 레이블링, QA, 최종 QA, 자동 레이블링, 오류 발견, 다양성 마사지, 레이블링 문서 + 버전 지정, PPL 교육, 에스컬레이션, 데이터 정리, 처리량 & 품질 통계, 평가 세트 + 분류 & 부스팅, ...

- 이미지나 영상 내의 객체를 레이블링 하는 것이 레이블링 작업의 전부라고 단순히 생각하기 쉽다.
- 그러나 안드레아가 나열한 항목들을 보면 레이블링 워크플로우(labeling workflows)는 그보다 더 많은 작업을 수반하고 있음을 알 수 있다

- 레이블링 워크플로우를 구성하는 작업들

- 데이터에 대한 레이블링 유형 기준 마련
- 레이블링 작업 결과물에 대한 품질관리(QA), 피드백 수집 과정 구축
- 레이블링 작업자의 훈련과 퍼포먼스 측정 등을 포함하는 인력관리
- 레이블링 과정 중 이슈 발생시 커뮤니케이션 비용 관리
- 라벨링을 통해 구축된 데이터셋의 버전관리 등

→ 나열된 것은 레이블링 작업 중 예상되는 워크플로우 범위 중 일부일 뿐...

- 이미지, 영상 데이터의 레이블링 작업은 가장 덜 중요한 것으로 인식
- 그러나 컴퓨터 비전 문제를 푸는 데 있어서 가장 선도적인 테슬라팀이 여전히 난항을 겪고 있는 부분이 라벨링 워크플로우라고 지적
- 상용화 단계에서 컴퓨터 비전 문제를 해결하기 위해 우리가 주목해야 할 부분은, **사실은 데이터**였다.

- **사람은 이미지를 보지만 컴퓨터는 숫자를 본다**

- **저명한 컴퓨터 비전 전문가, UC버클리 교수 Jitendra Malik(지텐드라 말릭)**

- 사람이 이미지를 인식하는 과정이 무의식적으로 혹은 잠재 의식적으로 진행되기 때문에 흔히 컴퓨터가 비전을 처리하는 방식도 매우 쉬울 거라고 착각할 수 있다고 지적
- 실제 사람의 경우에도, 대뇌피질의 상당 부분은 이미지 프로세싱 처리에 집중
- 컴퓨터가 이미지를 처리하는 방식은 인간과 다름
- 인간이 매우 시각적이고 직관적인 방식으로 이미지를 인식하는 반면, 컴퓨터는 이미지의 모든 부분을 개별 픽셀로 환산하여 숫자로 인식
- 이미지를 인식할 때마다 처리해야 하는 데이터 양이 많다는 것을 의미
 - 복잡한 시각적 작업을 수행하기 위한 계산 및 데이터 리소스 또한 방대해짐을 의미

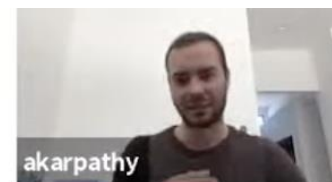
- 사람보다 더 많은 데이터를 필요로 하는 컴퓨터 비전 시스템
 - 현재 컴퓨터 비전 분야를 이끌어 가는 가장 고도화된 기법은 지도학습
 - 지도학습은 '정답'이라고 여겨지는 피쳐를 레이블링한 데이터를 기반으로 학습을 이행
 - 학습에 사용된 데이터를 기반으로 모델의 성능이 결정되기 때문에 모델이 인지하려는 모든 다양한 상황에 대해 데이터로 준비해서 학습시켜야 하는 것

- **CVPR 21** (Conference on Computer Vision and Pattern Recognition)
 - 컴퓨터 비전 분야에 있어서 세계적으로 저명한 학회
 - 안드레아 카파시의 발표 내용
 - 기계가 이해하기 어려운 상황 혹은 자연스럽게 얻을 수 있는,
 - 데이터 수는 적지만 실생활에서 충분히 일어나는 상황 등을 221가지로 추려 나열한 후,
 - 해당 케이스의 데이터를 집중적으로 확보 → 엣지 케이스(Edge Case)
 - 확보된 엣지 케이스를 4개월 동안 7차례 반복 테스트 수행 중

Triggers

Developed and maintained 221 triggers. E.g.:

- radar vision mismatch
- bounding box jitter
- detection flicker
- detection in Main camera but not Narrow camera
- driver didn't break but tracker thinks CIPV is rapidly decelerating
- break lights are detected as on but acceleration is positive
- rarely high/low velocity or acceleration
- CIPV cuts in / cuts out
- CIPV has high lateral velocity
- bounding-box derived depth disagrees with network-predicted depth
- rarely sloping road surface (hillcrest or dip)
- rarely sharp turning road surface
- driver breaks sharply on the highway
- stop an go traffic
- Main or Narrow or both cameras appear to be blinded
- driver enters/exits tunnel
- objects on the roof (e.g. canoes)
- driver breaks harshly and there is a VRU cloy to us but there is no intersection
- motorcycle on the highway at night
- vehicle type of CIPV flickers between network signal classes
- ...



zoom

- 엣지 케이스를 넣어 학습을 반복하며 모델의 성능을 높이는 것은 데이터를 통해 모델의 성능을 높이는 전형적인 Data-centric(데이터 중심) 방법론
- 비전 데이터를 다루는 문제를 풀고 있는 기업에서는
 - 모델 성능이 일정 수준에 도달하면
 - 이렇게 데이터를 중심으로 하는 반복사이클을 지속할 수 있는 견고한 머신러닝 데이터 플랫폼에 대한 필요성을 절감할 수 밖에 없음

수업시간에 질문하지 못한 내용은
아래에서 댓글로 올려주세요.

<https://aidalab.tistory.com/148>