# VinBigData
# Chest X-ray Abnormalities Detection

김기영, 권형주, 손원용, 오상수, 이준호

2021.02.16

01 개발 환경

02 진행 일정

03 자료 조사

04 Yolo, DB

05 향후 일정

# 개발 환경

kaggle

PyTorch

git + GitHub

Trello

프로젝트

# 진행 일정

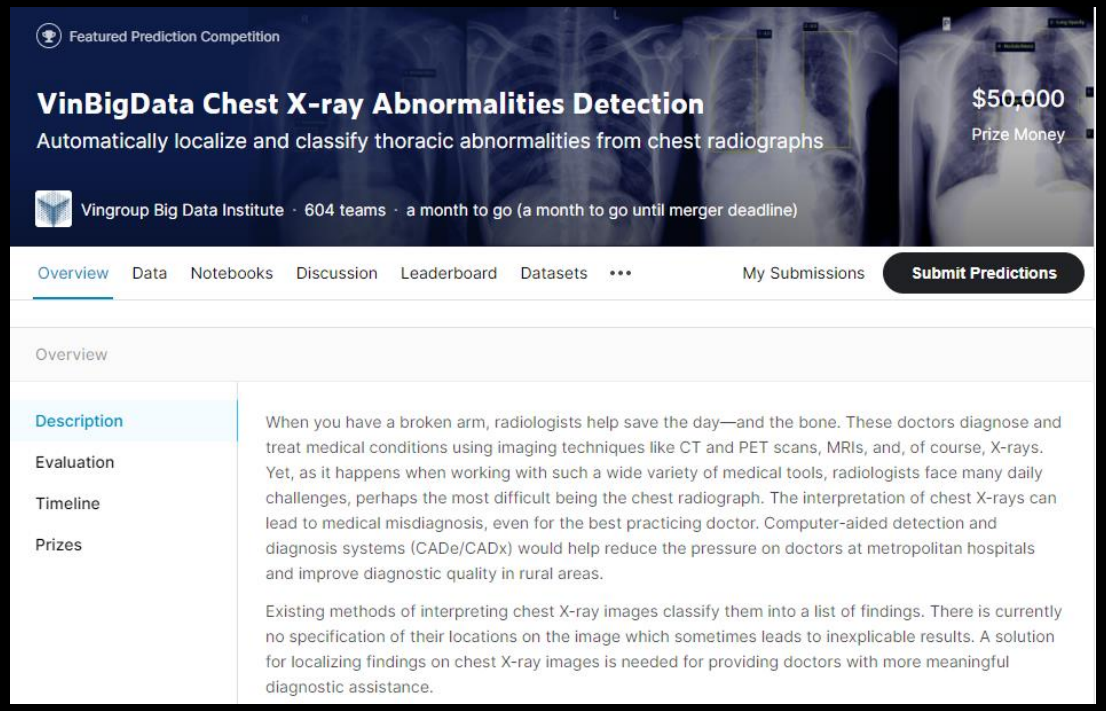
## 1주차 - 21년 2/8(月) ~ 12(金)

※ 2월 11일 ~ 2월 12일 설날

| | |
|---|---|
| 자료 조사 | Y-Label 조사 및 설명<br>Kaggle Overview,<br>data 해석 |
| | yolov5 모델<br>탐색, 코딩, 테스트 |
| | 데이터분석<br>DETR 모델 탐색 |
| | DICOM 파일 조사(Dicom viewer)<br>파일 변환(to PNG, JPEG)<br>MongoDB 업로드, 다운로드 |
| DB | MongoDB 파일 업로드 성능 측정(작업 시간)<br>MongoDB 데이터 업로드, 추가 다운로드 |

설날 연휴

# 자료 조사
## - 캐글 요약




**전문의도 해석하기 힘든 흉부 X-ray 진단의 품질 개선에 도움을 주기 위한 목적**
**Train : 15000 / Test : 3000**
**성능평가지표 : mAP at IOU > 0.4**

# 자료 조사
## - 캐글 요약

### Submission File

Images in the test set may contain more than one object. For each object in a given test image, you must predict a class ID, `confidence` score, and bounding box in format `xmin ymin xmax ymax`. If you predict that there are NO objects in a given image, you should predict `14 1.0 0 0 1 1`, where `14` is the class ID for "No finding", 1.0 is the confidence, and `0 0 1 1` is a one-pixel bounding box.

The submission file should contain a header and have the following format:

```
ID,TARGET
004f33259ee4aef671c2b95d54e4be68,14 1 0 0 1 1
004f33259ee4aef671c2b95d54e4be69,11 0.5 100 100 200 200 13 0.7 10 10 20 20
etc.
```

환자 고유번호 (정보 보안)

□ : 예측(1개 이상, 없을 경우 0 0 1 1)

□ : Box 좌표 표시

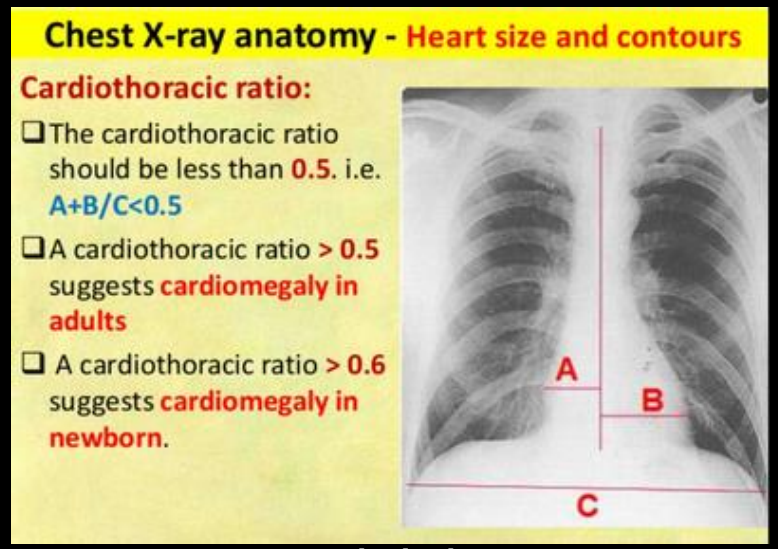
총 15개의 Class

# 자료 조사
## - 클래스 분석

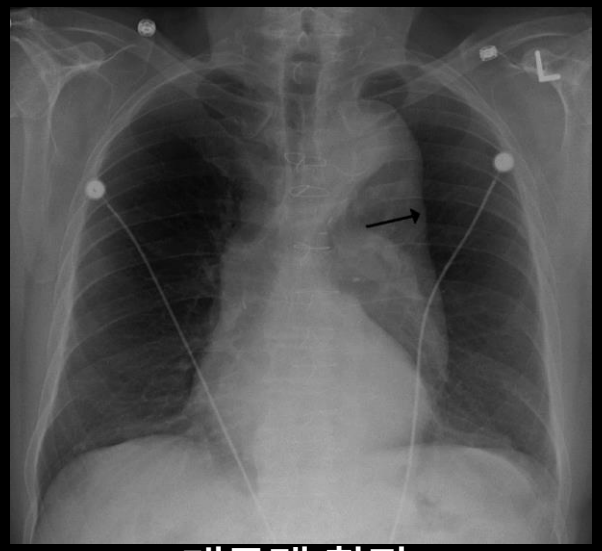
Cardiomegaly 심비대
Aortic enlargement 대동맥 확장
Pleural thickening 늑막 비후
ILD 간질성 폐질환
Nodule/Mass 종양
Pulmonary fibrosis 폐 섬유화
Lung Opacity 폐 불투명도

무기폐 Ateletasis
기타 병변 Other lesion
침투 Infiltration
흉막 삼출 Pleural effusion
석회화 Calcification
간경화 Consolidation
기흉 Pneumothorax

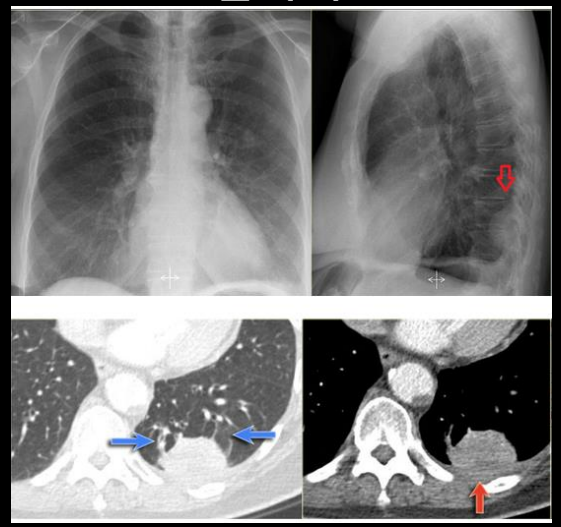
# 특징 분석




심비대




대동맥 확장

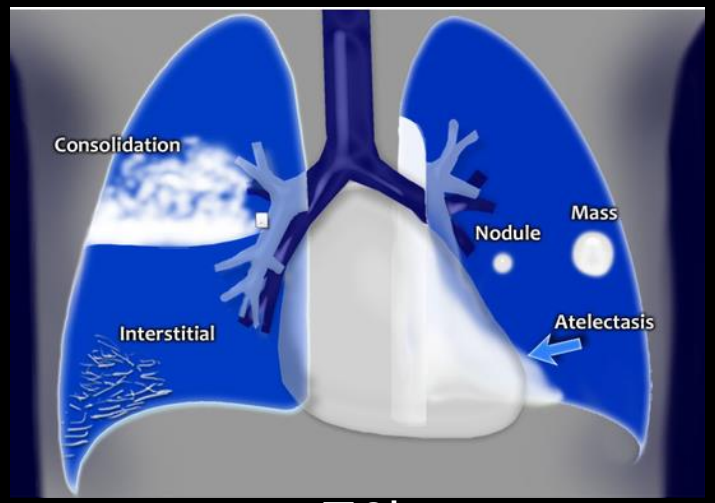


늑막 비후

간질성 폐질환

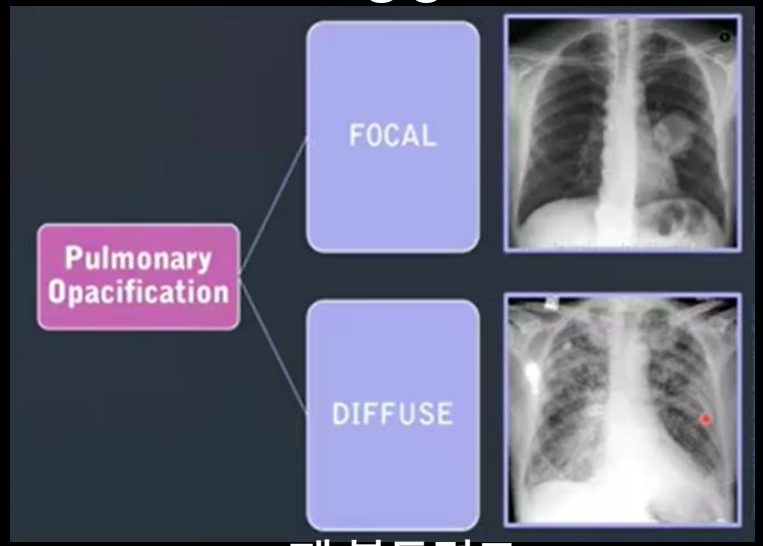
## 자료 조사
### - 클래스 분석

# 특징 분석

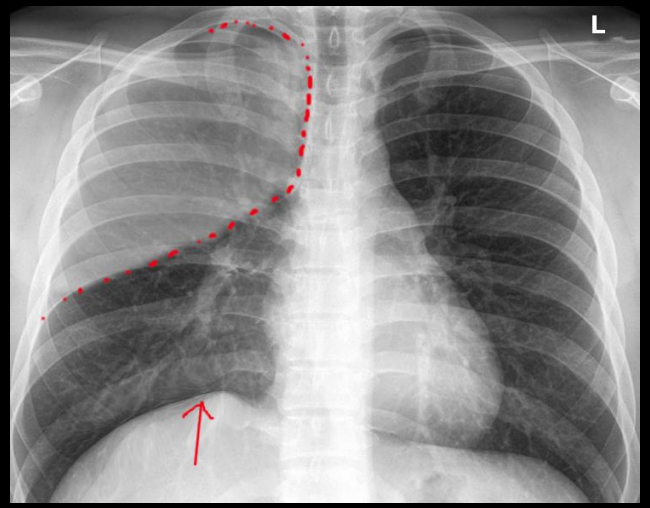



종양

섬유화



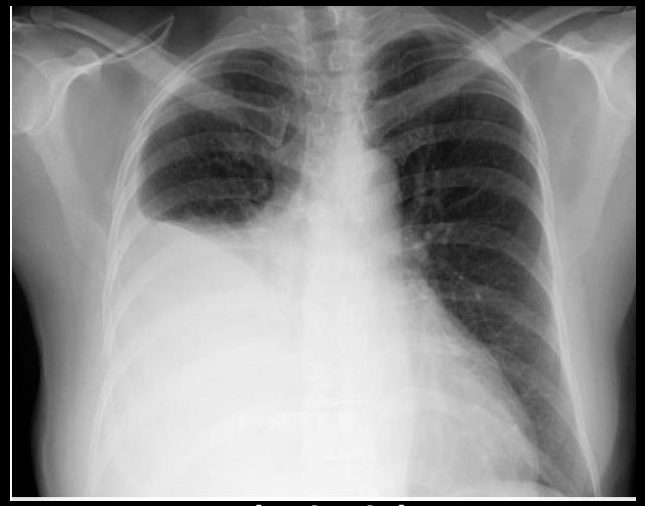

폐 불투명도



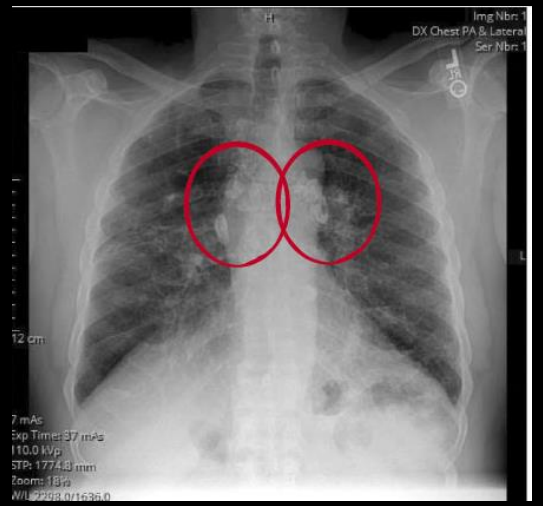

무기폐

## 자료 조사
- 클래스 분석

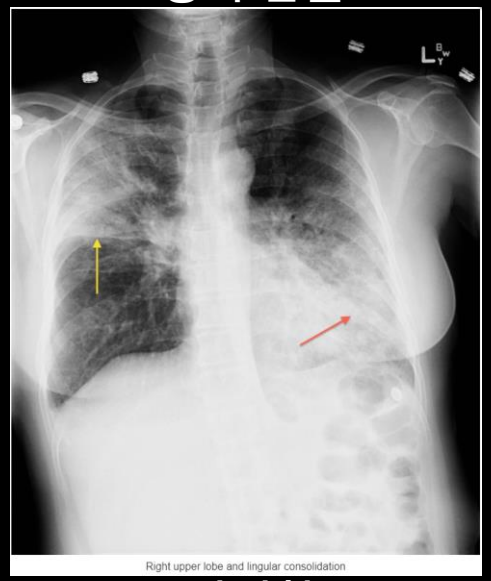
# 특징 분석



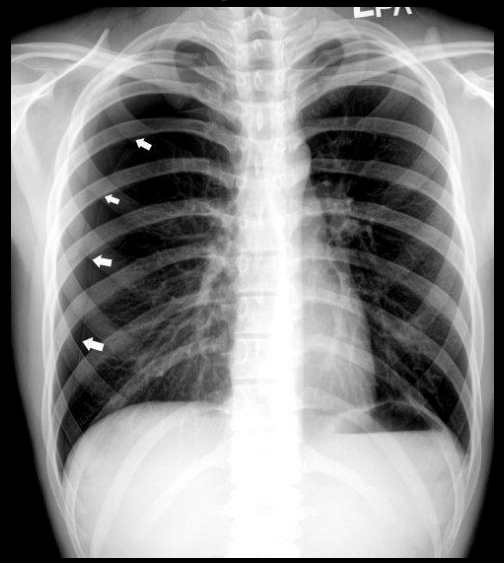
흉막 삼출




석회화




간경화



기흉

## 자료 조사
### - 클래스 분석

# 문제점

## 기타 병변 Other lesion

기타 발견 가능성 있는 질환

## 침투 Infiltration

ILD 간질성 폐질환
Consolidation 간경화
Lung Opacity 폐 불투명도

- 질병 구분의 기준 모호
- 한 질병이 다른 질병의 상위 범주에 포함

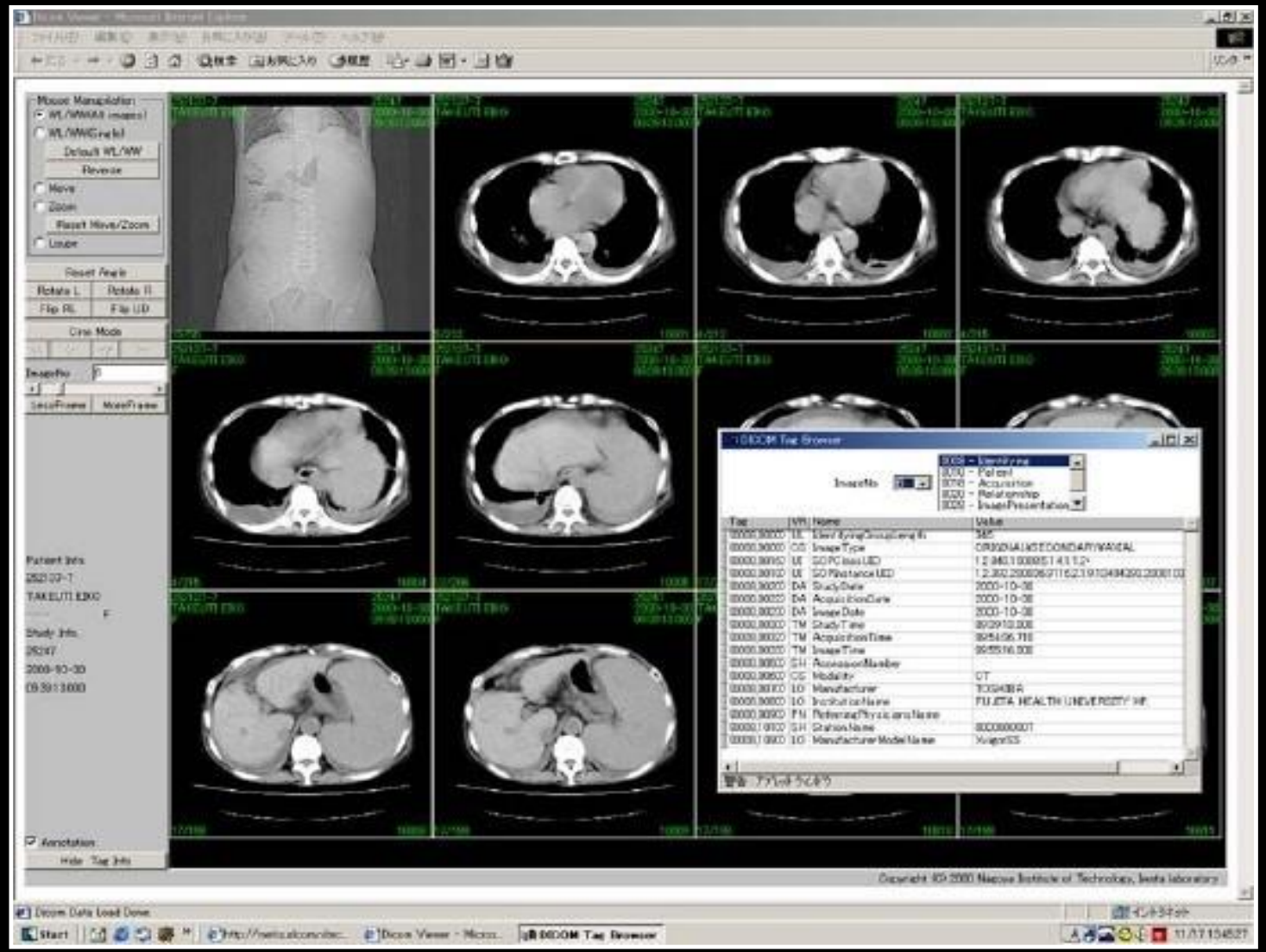
- X-ray 에서 형태 흡사

Pleural thickening 늑막 비후

Pleural effusion 흉막 삼출

Ateletasis 무기폐

## 자료 조사
- 클래스 분석

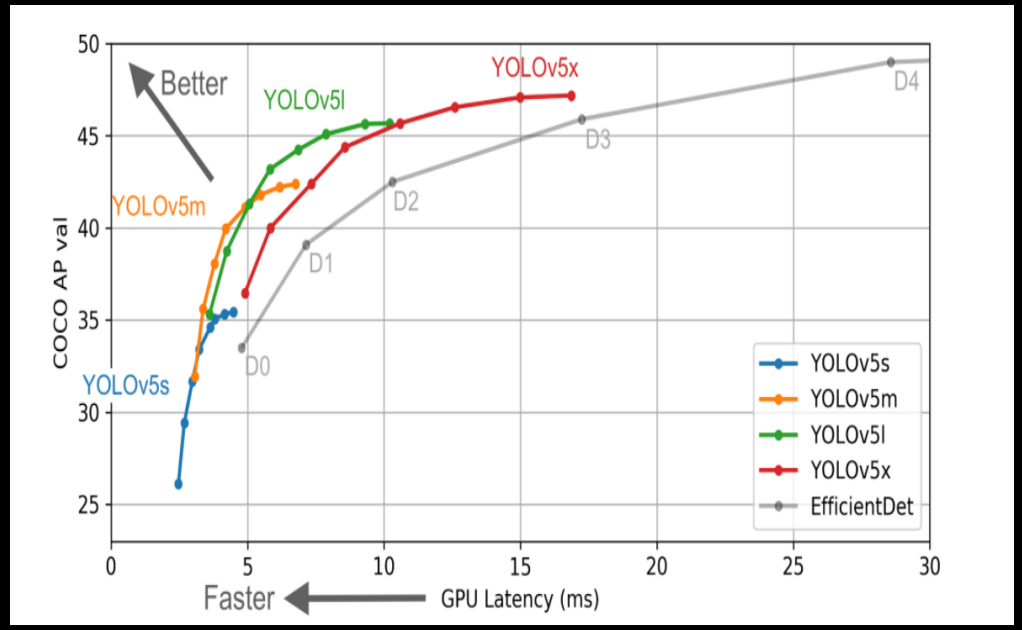


MongoDB

# MongoDB

| 파일명 | 개수 | 용량 (MB) | 업로드 / 다운로드 시간(초) |
|---|---|---|---|
| DICOM | 30 | 333 | Done<br>Upload time : 111.02231812477112<br>Done<br>Download time : 86.94876837730408 |
| PNG | 30 | 76.6 | Done<br>Uplode time : 41.680994272232056<br>Done<br>DownLode time : 40.191808462142944 |
| JPG | 30 | 13.9 | Done<br>Uplode time : 11.518765211105347<br>Done<br>DownLode time : 2.778932809829712 |

# 선정 이유




빠른 속도로 좋은 성능에 도달

**Yolov5**

# 선정 이유


다양한 Object Detection에 적합함

**Yolov5**

# Modeling

Train Data : 3515
Validation Data : 879
Epochs : 30
mAP/IoU = .5 : 0.31

1차 테스트 검증 (설정 내용)

**Yolov5**

# 1차 제출선정

| | | | | |
|---|---|---|---|---|
| 589 | ivanwang | 0.044 | 1 | 19d |
| 590 | ilyes | 0.040 | 2 | 1mo |
| 591 | Sylvain LEJAMBLE | 0.035 | 1 | 14d |
| 592 | Rookie | 0.030 | 1 | 1mo |
| 593 | Xuling | 0.028 | 1 | 12d |
| 594 | JunHo Lee | 0.023 | 1 | 3h |
| | Your First Entry ↑ Welcome to the leaderboard! | | | |
| 595 | AnhPham79 | 0.021 | 1 | 20d |
| 596 | Howie Wu | 0.019 | 3 | 2d |
| 597 | gbdmms | 0.019 | 10 | 1mo |
| 598 | Nadir Tariverdiyev | 0.018 | 1 | 1mo |
| 599 | Michal Kucirka | 0.000 | 1 | 20h |

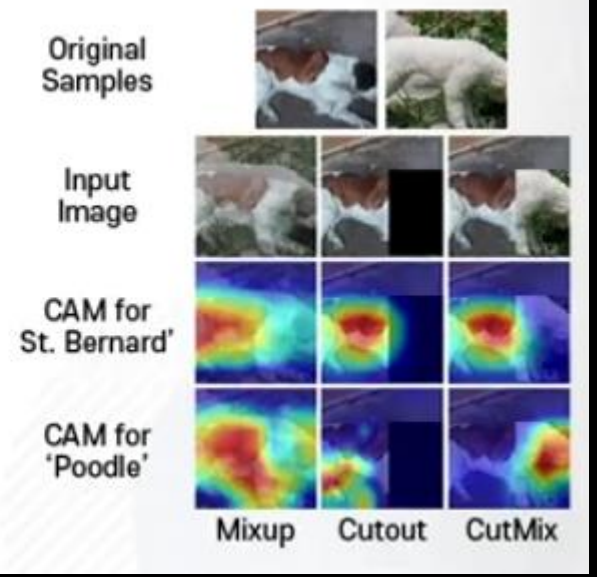
1차 SubMission : 0.023

Yolov5

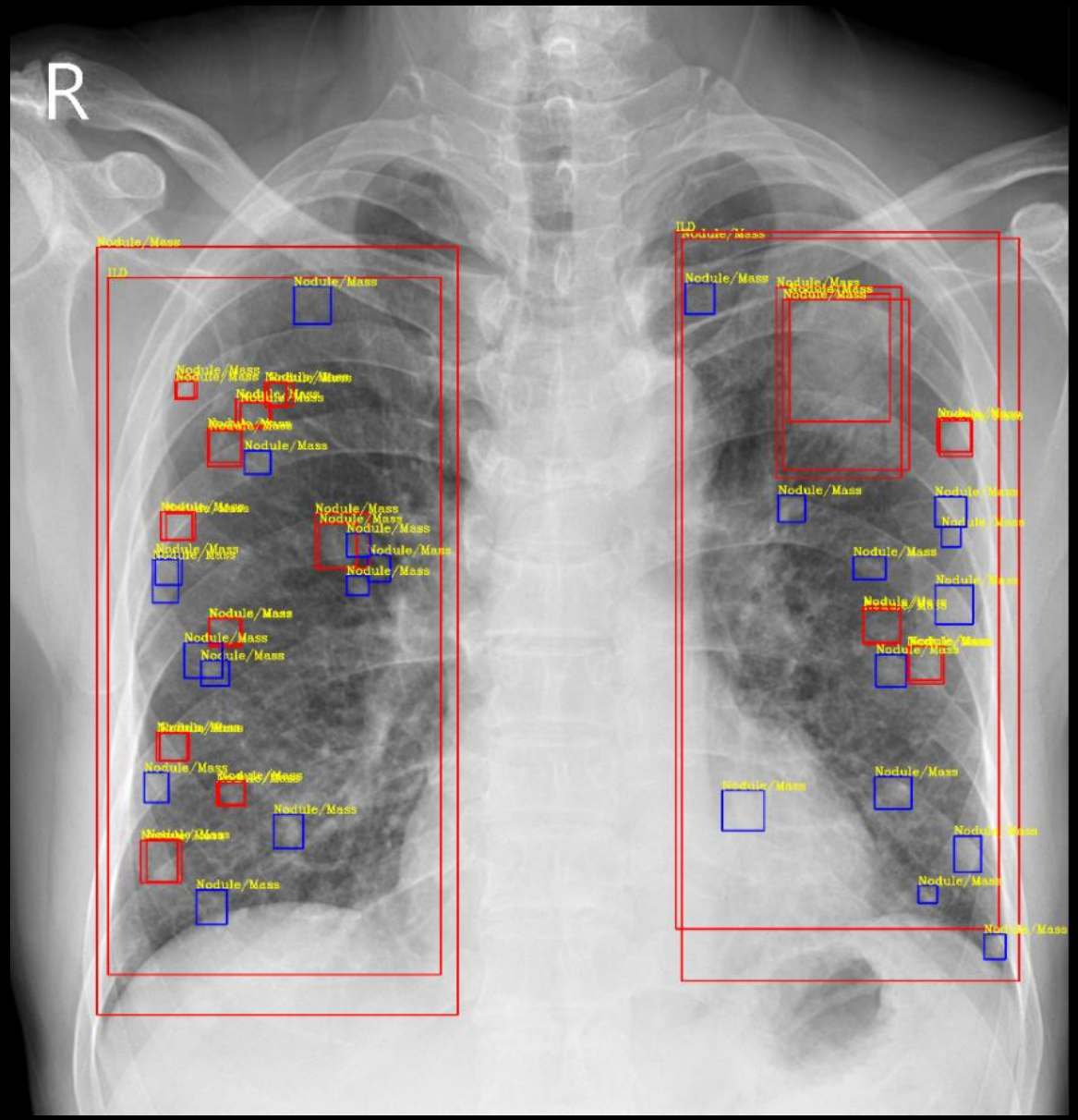
# 진행 상황 상세 공유 (추가)

1) ICCV2019 - "CutMix: Regularization Strategy to Train Strong Classifiers with Localizable Features"
- 서로 다른 클래스에 해당되는 이미지가 2개 이상 있을 때 이미지들을 랜덤하게 섞어서 새로운 이미지 생성 및 활용
- 기대효과 : Feature 비중이 낮은 부분으로 Object 탐지 및 분류

| ResNet-50 | Mixup[47] | Cutout[3] | CutMix |
|---|---|---|---|
|  |  |  |  |
| Dog 1.0 | Dog 0.5<br>Cat 0.5 | Dog 1.0 | Dog 0.6<br>Cat 0.4 |
| 76.3<br>(+0.0) | 77.4<br>(+1.1) | 77.1<br>(+0.8) | 78.6<br>(+2.3) |
| 46.3<br>(+0.0) | 45.8<br>(-0.5) | 46.7<br>(+0.4) | 47.3<br>(+1.0) |
| 75.6<br>(+0.0) | 73.9<br>(-1.7) | 75.1<br>(-0.5) | 76.7<br>(+1.1) |




# 진행 상황 상세 공유 (추가)




2) 증상별 상관관계 분석
- IOU 0.4 기준 4,352개의 이미지에 증상 중복

< 겹치는 증상 요약 (Top 3) >
1. 대동맥 확장 + 심비대 3,847
2. 흉막 삼출 + 늑막 비후 479
3. 늑막 비후 + 폐 섬유화 310

등 대략 60개 케이스에 대해 증상 중복

# 향후 일정

| | 2주차 - 21년 2/15(月) ~ 19(金)<br>※ 2월 11 ~ 2월 12일 설날 | 3주차 - 21년 2/22(月) ~ 26(金)<br>※ 2월 11 ~ 2월 12일 설날 |
|---|---|---|
| 자료 조사 | yolov5 Submission<br>CutMix 분석<br>DETR 모델 탐색 | 새로운 모델 탐색<br>기존 모델 튜닝 |
| DB | DICOM 영상처리 방법 학습<br>MongoDB 대용량 파일<br>업로드, 다운로드 | 이미지 전처리 |

# 참고 자료

## Y-Label


1) https://m.blog.naver.com/PostView.nhn?blogId=cloveryou1&logNo=221457514606&referrerCode=0&searchKeyword=Cardiomegaly
2) https://radiopaedia.org/articles/pulmonary-infiltrates-1
3) https://twitter.com/radiologistpage
4) https://radiologyassistant.nl/
5) Youtube – KIM'S CLINIC, hammadshams, Ayman Taji


## yolov5


1) https://github.com/ultralytics/yolov5
2) https://www.kaggle.com/awsaf49/vinbigdata-cxr-ad-yolov5-14-class-train
3) https://www.kaggle.com/xhlulu/vinbigdata-process-and-resize-to-png-1024x1024


## MongoDB


1) https://dicom.innolitics.com/ciods
2) https://www.mongodb.com/blog/post/building-mongodb-applications-binary-files-using-gridfs-part-1
3) https://pymongo.readthedocs.io/en/stable/api/gridfs/

# THANK YOU

https://trello.com/b/bUIob804/kaggle
https://github.com/orgs/X-Ray-Project/teams/x-ray