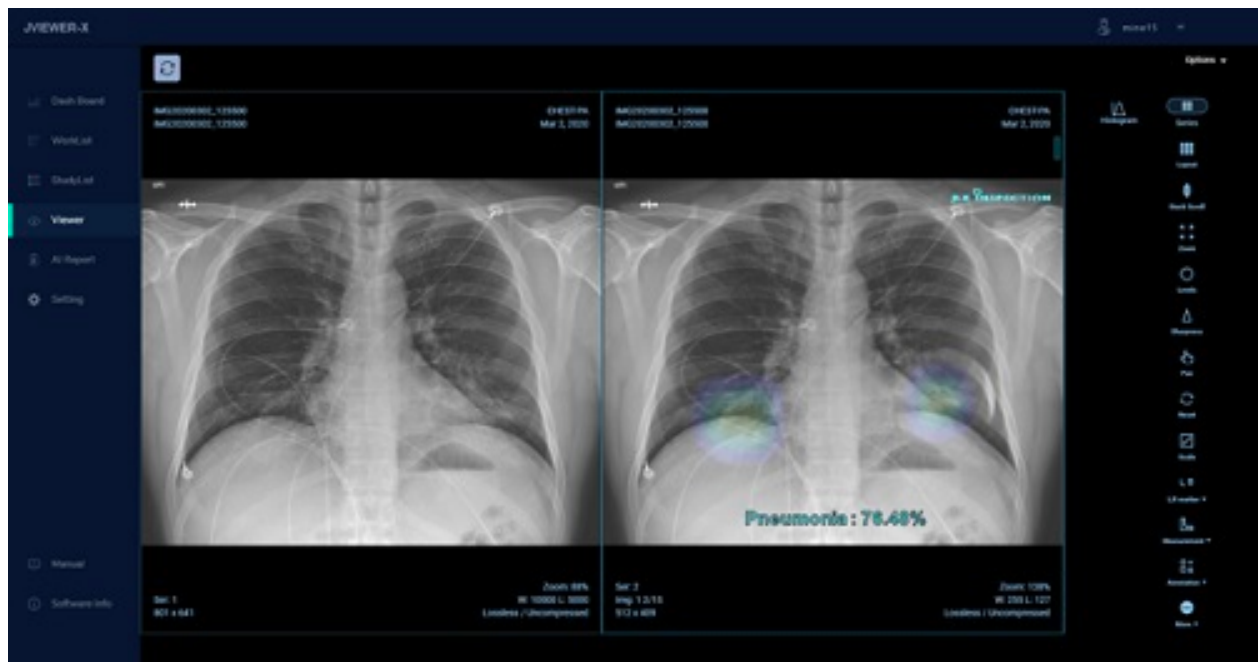
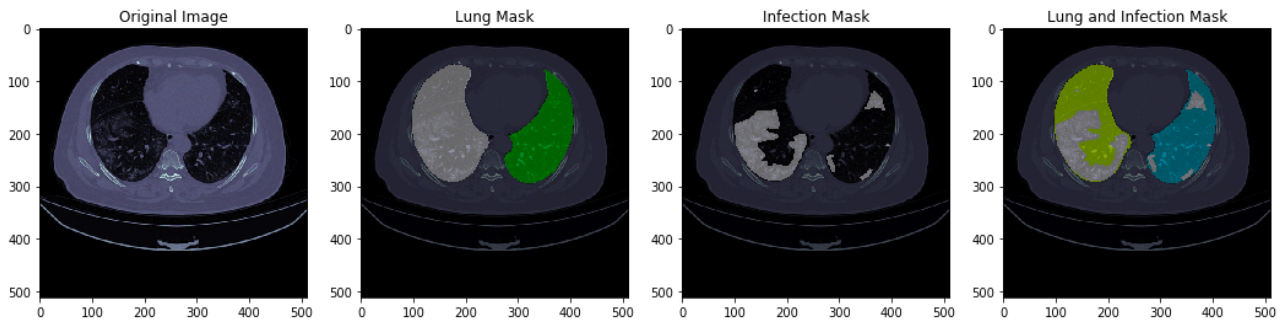


IOI Project #1

COVID-19 Diagnosis

20181512 김현준

1. Dataset Description



해당 Dataset은 정상인들의 흉부 X-Ray 촬영본, COVID-19 (코로나 바이러스 감염증) 환자들의 흉부 X-Ray 촬영본들로 이루어져 있고, 이외에도 COVID-19 환자들의 흉부 CT 촬영본 Dataset도 함께 활용할 수 있다.

초기엔 X-Ray 촬영 및 CT 촬영만으로는 COVID-19 진단에 어려움을 겪은 바 있었으나, 최근에는 WHO 공식 발표에 의거, 펜데믹 사태에 이르면서 감염 사례가 많아지면서 Dataset이 늘어나며 트레이닝 리소스가 늘어났다.

2. Kernel Description



Covid-19 Detection from Lung X-rays

Python notebook using data from [multiple data sources](#) · 1,973 views · 1mo ago

해당 Kernel은 위에 설명한 Dataset을 활용하여(Kaggle에 업로드 되어있는 여러 종류의 Chest X-Ray Dataset, CT Dataset 을 사용) 특정 사람의 흉부 X-Ray 사진 및 CT 촬영본을 가지고 코로나 바이러스에 감염이 되었는지 판별할 수 있는 머신러닝 모델을 만들었다. (Keras, TensorFlow를 사용했다.)

```
import glob
import matplotlib.pyplot as plt
import matplotlib.image as mpimg
%matplotlib inline

normal_images = []
for img_path in glob.glob(DATASET_DIR + '/normal/*'):
    normal_images.append(mpimg.imread(img_path))

fig = plt.figure()
fig.suptitle('normal')
plt.imshow(normal_images[0], cmap='gray')

covid_images = []
for img_path in glob.glob(DATASET_DIR + '/covid/*'):
    covid_images.append(mpimg.imread(img_path))

fig = plt.figure()
fig.suptitle('covid')
plt.imshow(covid_images[0], cmap='gray')
```

```
train_datagen = ImageDataGenerator(rescale=1./255,
    shear_range=0.2,
    zoom_range=0.2,
    horizontal_flip=True,
    validation_split=0.3)

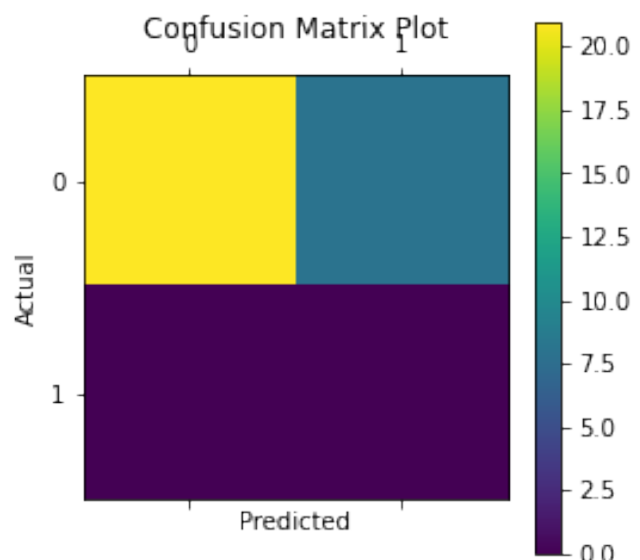
train_generator = train_datagen.flow_from_directory(
    DATASET_DIR,
    target_size=(IMG_H, IMG_W),
    batch_size=BATCH_SIZE,
    class_mode='binary',
    subset='training')

validation_generator = train_datagen.flow_from_directory(
    DATASET_DIR,
    target_size=(IMG_H, IMG_W),
    batch_size=BATCH_SIZE,
    class_mode='binary',
    shuffle=False,
    subset='validation')

history = model.fit_generator(
    train_generator,
    steps_per_epoch = train_generator.samples // BATCH_SIZE,
    validation_data = validation_generator,
    validation_steps = validation_generator.samples // BATCH_SIZE,
    epochs = EPOCHS)
```

위와 같이 Dataset을 Label에 맞게 입력하고 모델을 만들어 직접 Training 및 테스트를 했을 때, 다음과 같은 Confusion Matrix Plot을 보인다. 표를 간단히 해석하자면, 해당 머신러닝 / 딥러닝 모델이 '이 사진은 COVID-19 환자의 폐 사진이다'라고 판별했는데, 실제로 코로나 바이러스 감염증 환자인 경우가 약 21건, 그리고 '이 사진은 COVID-19 환자의 폐 사진이 아니다'라고 판별했는데 실제로는 바이러스 감염자인 경우는 약 8건에 해당된다고 볼 수 있다. (29명의 COVID-19 환자 X-Ray로 테스트한 결과)

Predicted	0	All
Actual		
0	21	21
1	8	8
All	29	29



2. How to utilize this data

COVID-19를 진단하는 기법 중 가장 널리 쓰이는 방법은 PCR(중합효소 연쇄 반응) 검사 키트를 이용하는 것이다. 해당 키트를 사용했을 때 만약 코로나 바이러스의 유전 물질이 있으면 (코로나 바이러스를 보유하고 있으면) 그 유전 물질에 대응하는 프라이머를 제작하여 중합효소 연쇄 반응을 통해 바이러스만이 가지고 있는 유전 물질을 증폭시키는데, 만약 증폭이 일어나면 검사 키트 상으로 식별되어 COVID-19 환자라고 진단할 수 있는 기법이다.

그러나 PCR 검사 기법을 활용한 키트를 이용하는 것은, 시간적으로 부담이 될 수 있는 기법이다. 지금처럼 전세계가 팬데믹으로 고통받고 있을 때, 사람들은 자신도 모르게 갑자기 나오는 기침에 불안감을 호소하기 마련이다. 따라서 검사 키트를 통해 검사를 하더라도, 검사 결과가 너무 느리게 나오는 것은 치명적인 단점으로 다가올 수 있다.

그렇다면 다른 기법을 생각해볼 수 있다. 가장 쉽게 생각하기론 흉부 X-Ray 촬영이 있다. 호흡기관에 악영향을 미치는 코로나 바이러스의 특성 상, 흉부 X-Ray를 촬영했을 때에 정상인의 흉부 사진과 다른 결과를 보이는 것은 당연한 이치이다. 하지만 전문가가 아닌 이상 쉽게 COVID-19를 진단하기엔 어려움이 있다. 왜냐하면 X-Ray 결과만 얼핏보면 폐렴 등 일반 질환들과 다른 점을 찾기 어렵기 때문이다.

하지만, WHO가 공식으로 '팬데믹' 사태를 공표할 정도로 전세계적으로 코로나 바이러스의 확산이 매우 빠르게 이루어져왔다. 따라서, 흑시모를 COVID-19 진단을 위해 병원을 찾는 사람이 많아졌다. 이렇게 병원에서도 코로나 바이러스 감염증 진단에 참고하기 위해 촬영한 X-Ray, CT 데이터들이 축적되면서 원래는 진단하기 힘들었지만, 분명히 폐렴과 같은 일반 질환들과는 다른 부분이 발견되기 시작하고 또 감염된 경우의 특이한 점도 발견이 가능해졌다. 즉, ML/DL 모델을 개발하는 것이 가능해졌다는 소리다.

위에서 분석한 Kernel은 많지 않은 Dataset으로 충분히 높은 Accuracy를 보이고 있다. 이 말은 지금까지 각 나라의 병원에 수집된 수많은 Dataset만 있다면, 이를 통해 수많은 Training을 거치게 되어 값싸고 빠른 X-Ray 검사만으로도 코로나 바이러스 감염증 진단이 쉽고 빨라진다는 소리다.

만약 이 기술이 상용화된다면, X-Ray 촬영 장비가 있는 보건소, 동네 작은 병원에서도 방사선 전문가들조차 필요없이 COVID-19 진단이 가능해진다는 소리이며, 이로 인해 진단 키트 품귀현상을 극복할 수 있고 또한 공포에 떨고 있을 국민들의 입장에서 값싸고 간편하고 빠르게 자신이 코로나 바이러스에 감염이 됐는지 안됐는지 알 수 있어 정신적으로 매우 안정될 것이다.

안그래도 국내 진단 키트 제작회사들도 난항을 겪고 있는 상황이다. 시약 원재료 값이 5배 가량 폭등하면서 생산 단가가 높아지고, 또 진단 키트 수요는 나날이 급증하고 있어 가격은 내릴 수 밖에 없고, 또 자동화 생산에도 한계가 있어 생산량이 전면적으로 대폭 감소하고 있다고 한다. 위의 활용 방안이 상용화가 되면 진단 키트 생산에 억지로 많은 공을 들일 필요가 없어질 뿐만 아니라 기존 진단 키트의 치명적 단점들에 더이상 연연하지 않아도 된다.