

# 팀 프로젝트 지도 활동 보고서

교과목명	AI및데이터분석의 기초		
강사	한상호 교수		
활동 일자	2020년 09월 19일 (토요일)		
학과	성명	연락처	서명
스마트시스템SW학과	정학제	010-3895-1306	
스마트시스템SW학과	정재윤	010-9008-4847	
스마트시스템SW학과	정해문	010-7176-4289	
스마트시스템SW학과	조재홍	010-4062-9523	
활동 내용	<p>(미팅시 팀원들 사진 붙이기 - 선택사항)</p> <p>- 미팅 결과물에 대한 URL:  <a href="https://github.com/Hott-J/AI-BigData-Basic">https://github.com/Hott-J/AI-BigData-Basic</a>            팀장 github 계정 아래 project repo를 하나 만들고 (repo 이름은 project 취지에 맞게 작명), 코드 파일들, PPT 파일 등을 commit 하고 해당 파일 접근을 위한 URL을 여기에 작성하기</p> <p>- 2페이지에 작성했습니다. 데이터 수집에 관해 조사하고 의견을 나누었습니다.</p>		

## < 정학제 팀장 >

마스크 3구분을 위한 데이터 수집

모델을 만들기 위한 데이터를 어떻게 수집할것인가?

1. Teachable Machine 홈페이지에서 직접 본인 얼굴을 촬영한 뒤  
모델 생성해보기

<https://ai-creator.tistory.com/76>

2. 구글링해서 마스크쓴 사진, 안쓴 사진, 턱스크 사진 찾아오기

3. microsoft search api 사용해보기(그냥 이렇게있다 참고만...)

<https://docs.microsoft.com/en-us/graph/search-concept-overview>

만약 본인 얼굴로만 모델을 생성했을 경우, 다른 사람에 적용해도  
제대로 판단할 수 있을까...?

어떤 방법으로 데이터를 수집하는게 가장 좋고, overfitting(과적합)이  
안날지 생각해보고 찾아보고 간단히 자기 생각 정리해주시면 됩니다.

과적합 관련 : <https://wikidocs.net/61374>

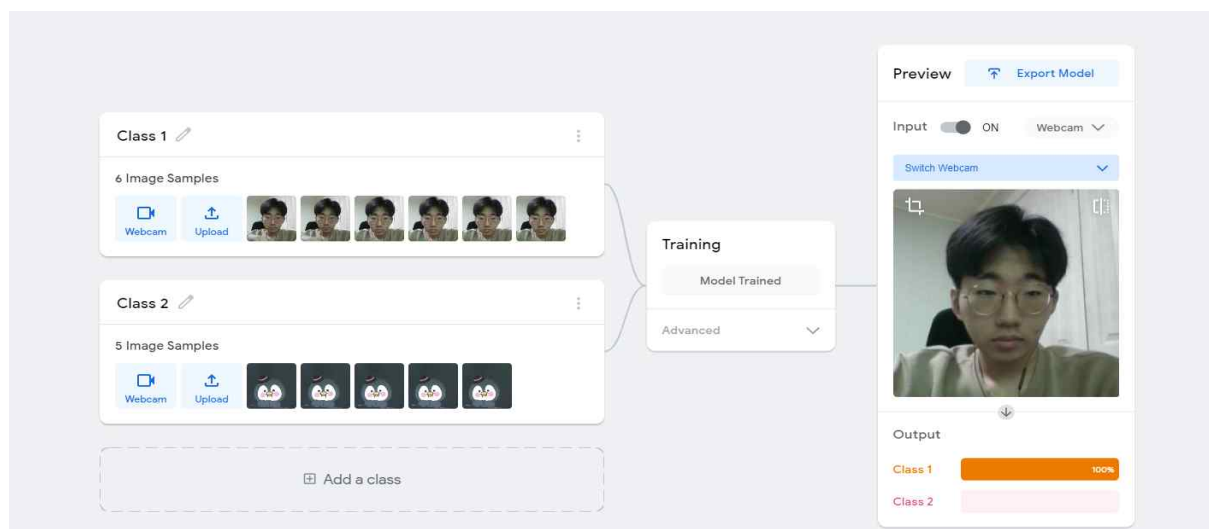
~2020-11.29 정도까지

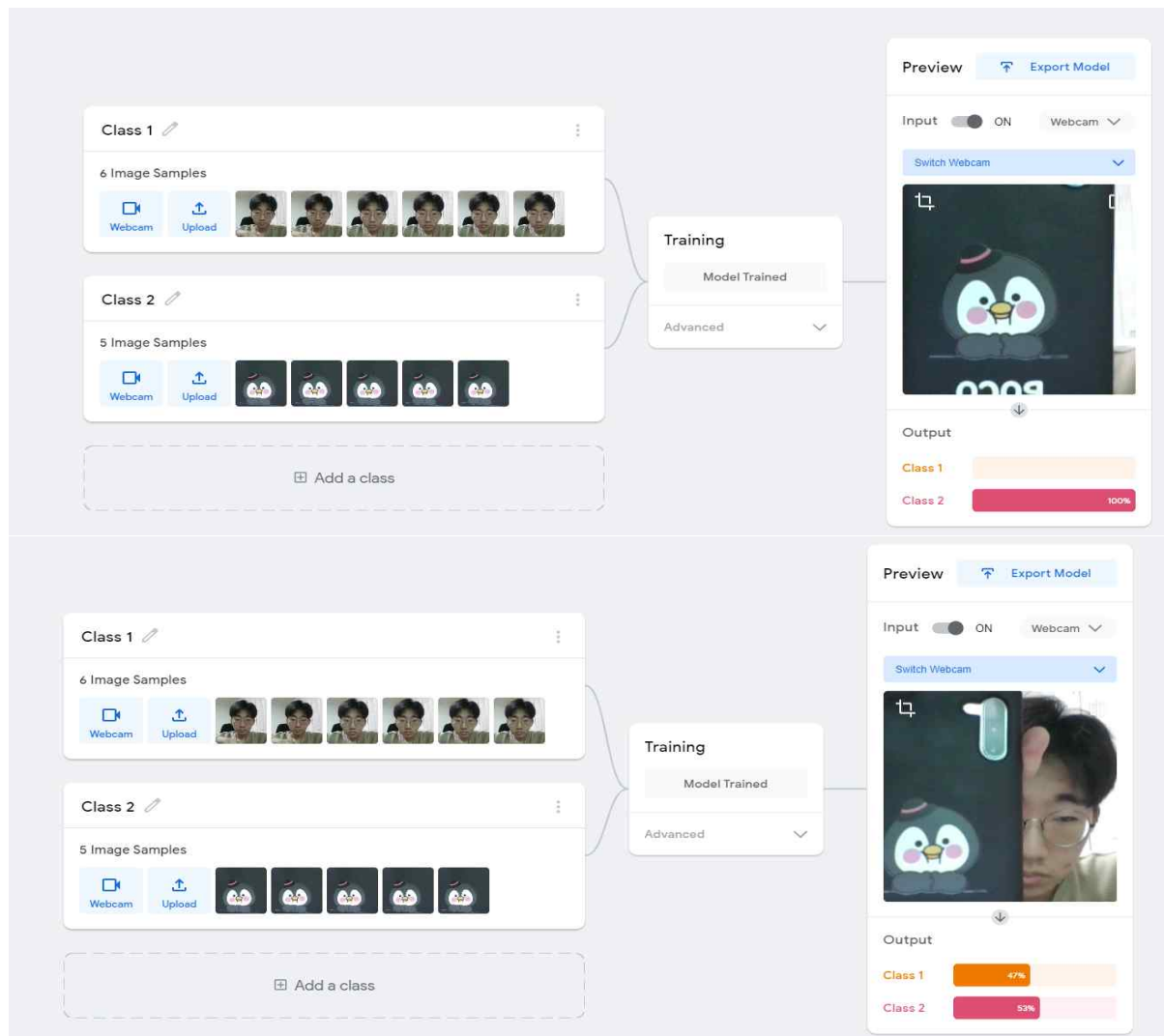
위와 같은 내용으로 과제 제시하였습니다. 팀원들이 정리해준 내용으로, 보고서 작성하였습니다.

## < 조재홍 팀원 >

Teachable Machine 홈페이지에서 직접 자신의 얼굴을 촬영하여 모델을 생성하였다.

비교 대상은 핸드폰 케이스의 캐릭터 이미지를 사용하였다.





위의 사진과 같이 정확도의 결과가 드러났다.

(마스크를 쓴 사진 / 마스크를 쓰지 않은 사진 / 마스크로 입만 가리거나 코와 입 모두 가리지 않은 사진) 이 3장의 사진을 구분하기 위한 방안 생각해보기

첫째로, 마스크를 쓴 상태와 쓰지 않은 상태에 큰 차이점은 입이 보이는지, 코가 보이는 지로 구분할 수 있다. 입을 가렸는지 썼는지는 입술의 색과 주변 피부색을 비교하여 입의 위치를 알 수 있다. 또한 일부 카메라 애플리케이션을 보면 자동 보정을 해주고, 입술에 색을 더해주는 기능도 포함되어 있다. 이는 마스크를 쓴 상태에서 마스크 위로 색을 더해줄 때도 있기에 입술의 색을 비교하여 색을 더해주는 기능이 아니라 얼굴형을 보고 입의 위치를 가늠하여 더해준다고 생각한다. 따라서 마스크를 썼을 때, 입을 가렸는지 판별하는 여부는 위의 2가지를 고려하면 될 것 이라고 생각한다.

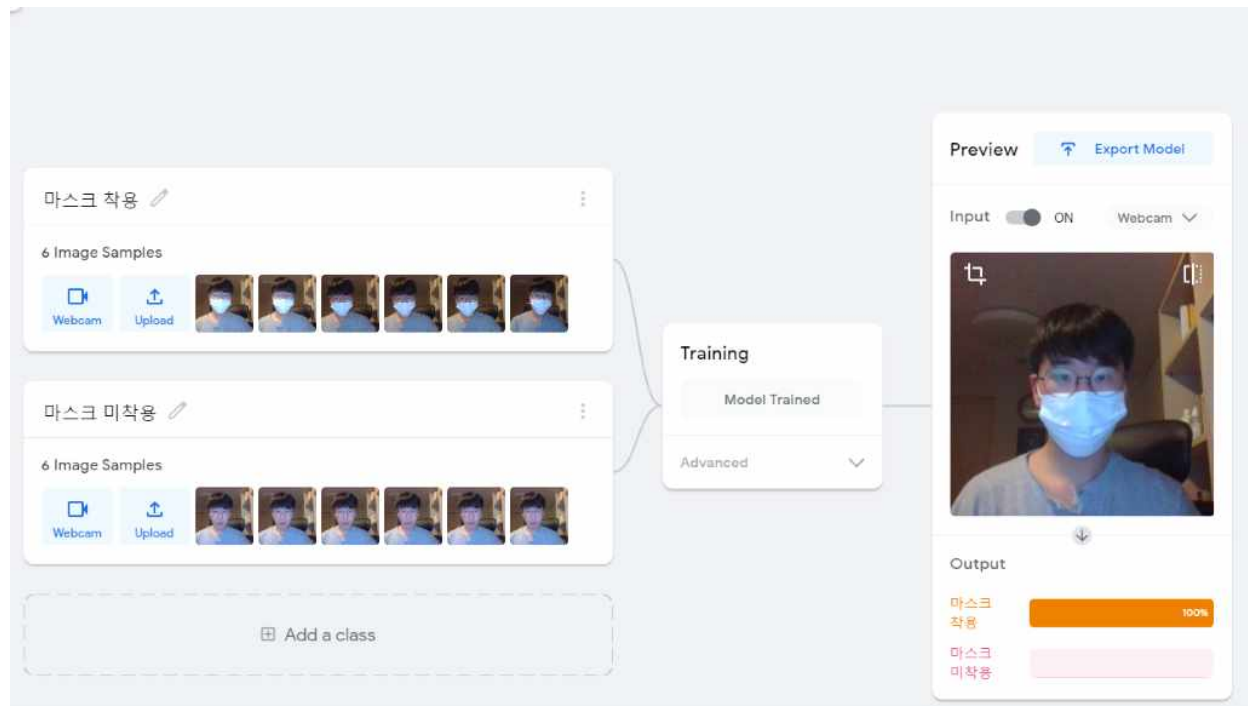
둘째로, 입은 가렸으나 코는 가리지 않은 상태이다. 코는 입술과 다르게 피부색과 색으로 구분되어지지 않기 때문에 한가지 생각한 것으로는 명암으로 코의 위치를 구분하는 것이다. 사진을 찍거나 영상을 촬영할 때, 빛이 존재하기 때문에 얼굴의 정면으로 빛이 드는 환경이 아니라면 코로 인해 얼굴에 그림자가 질 것이다. 그러므로 마스크를 썼을 때는 그림자가 없어야 하고, 그렇지 않을 때는 그림자가 존재한다는 것이다. 그러나 이 부분에서는 검정색 마스크를 썼을 때에는 그림자의 여부를 판별하는 것이 어렵기 때문에 그럴 때에는 볼에 피부색을 비교할 필요가 있다고 생각한다.

결론적으로 입술의 색, 코의 그림자가 생기는데 여부, 볼이나 주변의 피부색, 얼굴형에 대한 코와 입의 위치를 예측하는 것을 통해서 마스크를 썼는지에 대한 여부를 알아볼 수 있을 것이다.

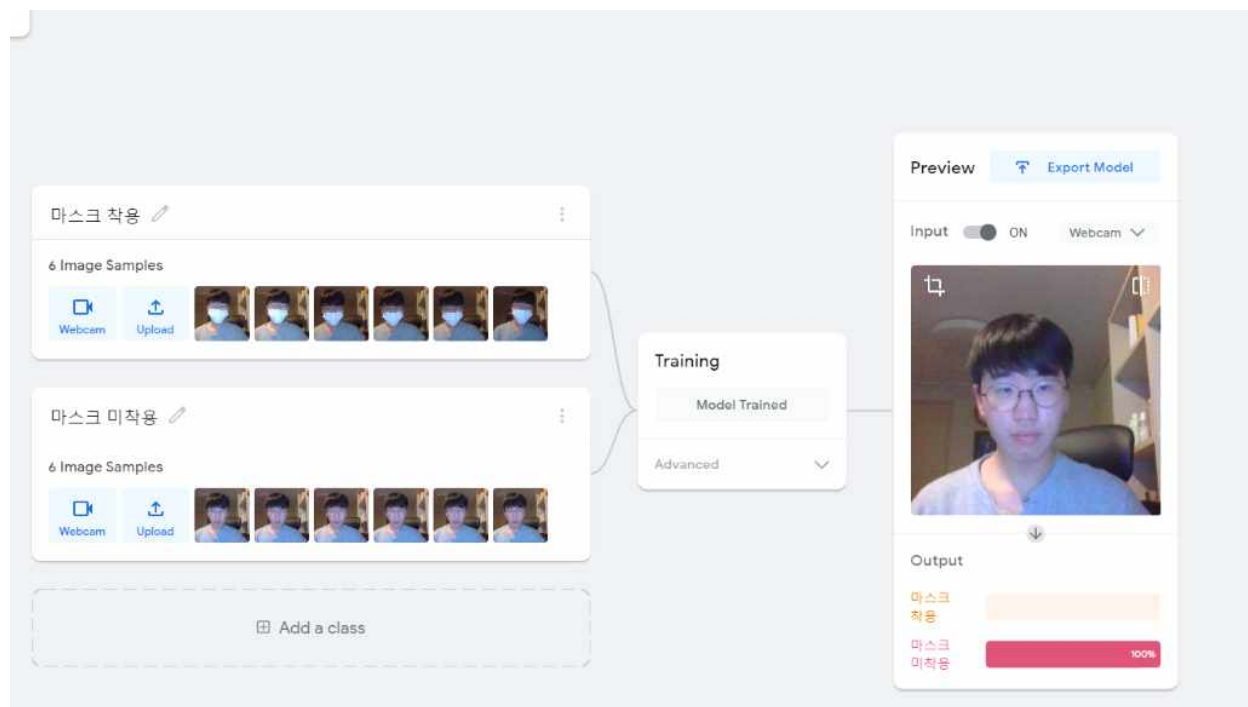
## < 정해문 팀원 >

Teachable machine 을 사용하여 마스크 여부를 확인

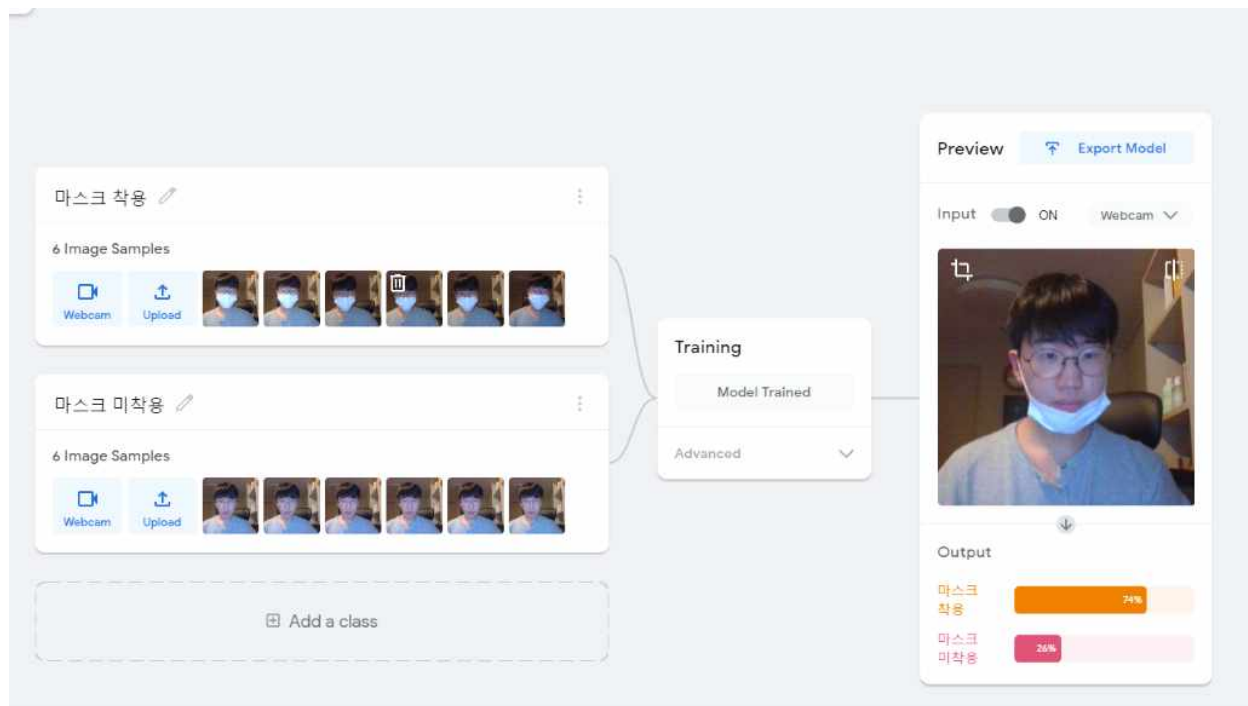
### 1. 마스크 착용



### 2. 마스크 미착용



### 3. 마스크 반정도만 착용



마스크 착용하거나 착용하지 않았을 때는 정확하게 식별이 가능한 모습을 볼수 있다.(동일 인물인 경우)

#### 과적합

machine learning 에서 학습 데이터를 과하게 학습(overfitting) 하는 것을 뜻함  
학습시킨 얼굴로 모델 생성후 다른 사람에게 적용할 때 제대로 판단하기 위해서는  
더 많은 데이터 수집이 필요하다.

데이터 수집 방법은 웹 사이트에 있는 자료들을 긁어오기 위해 웹 크롤링을 할 수 도 있고, 자신이 운영하는 서비스에서 유저들의 행동 데이터를 수집하기위해 로그를 남길 수 도 있다.

위 인물의 학습데이터의 특징을 보면 마스크는 흰색, 안경 착용, 배경은 좀 어둡다.

이에 위 모델에 다른 인물을 적용해 제대로 판단 하기 위해선 마스크의 색깔 등 더 많은 요소를 추가하여 데이터를 학습해야 할 것이다.

복잡한 모델이 간단한 모델보다 과적화될 가능성이 높는데, 이때 복잡한 모델을 좀 더 간단하게 하는 방법으로 가중치 규제가 있다. 학습 중에 가중치가 큰 것에 대해서 일종의 패널티를 부과해 과적합의 위험을 줄이는 방법이다.

신경망 모델이 복잡해질 때 가중치 규제만으로는 어려운데 이때 드롭아웃 기법을 사용한다.

드롭아웃 기법은 뉴런의 연결을 임의로 삭제하는 것으로, 훈련할 때 임의의 뉴런을 골라 삭제하여 신호를 전달하지 않게 한다.

#### 마스크 착용 사진

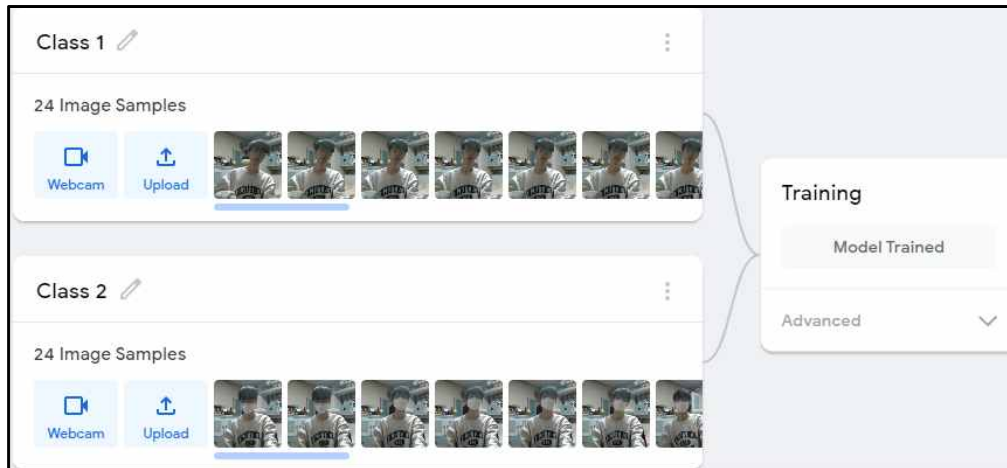
- 구글링

마스크 턱에 걸친 사진

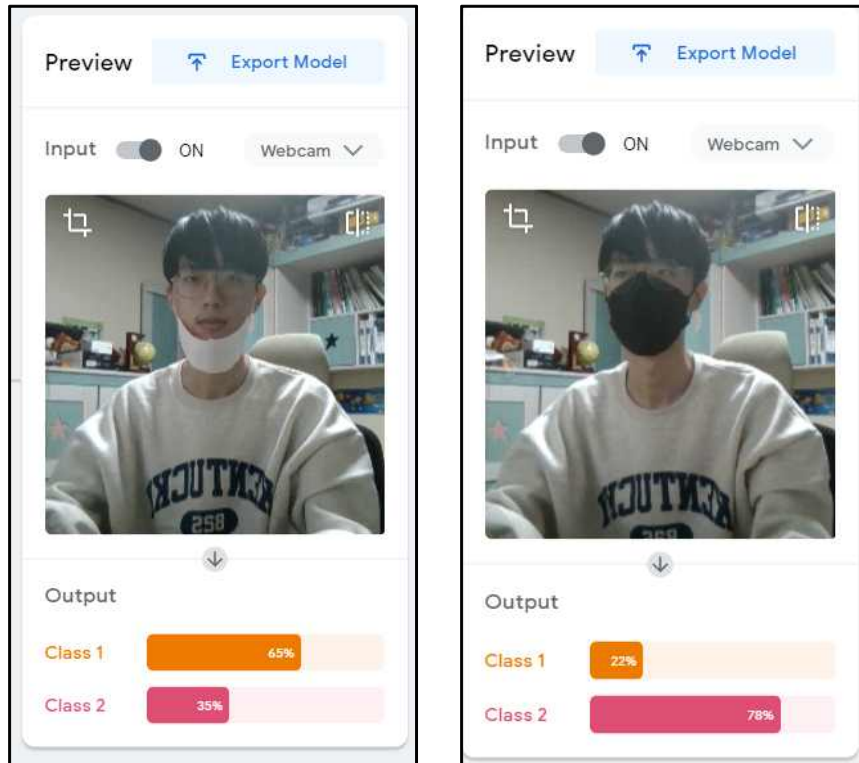
-구글링

< 정재윤 팀원 >

Teachable Machine 홈페이지에서 직접 그냥 얼굴과 하얀 마스크를 쓴 이미지를 학습해서 모델을 생성해보았다.



이처럼 학습 이미지와 같은 테스트 이미지는 잘 인식하지만, 턱에 마스크를 쓰거나 검은색 마스크를 쓰면 정밀도가 떨어지는 것을 알 수 있다.



다음으로 데이터를 이용한 모델 학습 방법을 생각해 보자.

우선 내 얼굴로만 모델을 생성해본다고 치자. 이미지 머신러닝의 가장 중요한 점은 특징(feature) 추출이라고 생각한다. 내 얼굴로만 모델을 생성했을 경우 너무 학습이 내 얼굴에만 맞추어지게 될 것이다. 그 결과 다른 사람에게 이 모델을 적용한다면 정확도가 내려가고, 심하면 작동하지 않는 테스트셋에 치중된 오버피팅이 발생할 것이다.

그렇기에 학습을 위한 데이터를 수집할 때에는 편향되지 않도록 얼굴과 마스크라는 것의 특징을 학습시키기 위해 얼굴형, 인종, 마스크 색깔 등등 다양하고 방대한 테스트 이미지들을 수집하고 활용하여 테스트셋을 생성하는 것이 오버피팅을 방지하고 모델을 정확하게 학습하게 할 수 있을 것이다.

이후 방대한 데이터를 이용해 마스크 쓴 사진, 안쓴 사진, 마스크로 일부분만 가린 사진을 구분해야 한다면, 학습할때 우선 마스크를 제대로 쓴 사람의 경우 얼굴에 존재하는 입과 코의 특징이 존재하지 않는다는 점을 이용해서 학습할 수 있을 것이다. 하지만 턱스크와 같은 경우에는 조금 까다로울것 같다. 얼굴색과 마스크의 색이 다른 경우에는 턱의 특징이 다른 점을 이용하여 인식하기 쉬울 것 같으나, 비슷한 경우에는 조금 인식하기 힘들것 같다.

팀프로젝트 지도 수행과 관련하여 상기와 같이 활동하였음을 확인합니다.

2020 년 11월 21 일

담당 교수 : 한 상 호 (인)