

## 개인\_정진원\_T2206

### 1. 프로젝트 목표

카메라로 비춰진 사람 얼굴 이미지만으로 이 사람이 마스크를 쓰고 있는지, 쓰지 않았는지, 정확히 쓴 것이 맞는지 자동으로 가려내며 성별과 나이 또한 예측하는 시스템을 개발하고, 협업과 딥러닝 모델 개발에 필요한 실력을 쌓는다.

### 2. 학습목표를 달성을 위해 무엇을 어떻게 했는가?

#### A. 개인 학습

- 주어진 시간 동안 모델의 성능을 높이기 위해 최대한 많은 시도를 하려고 했고 이를 베이스라인 코드에 적용시켰습니다. 또한 진행했던 시도들에 대한 기록을 통해 효율적으로 팀원들에게 정보를 전달하였습니다.

#### B. 공동 학습

- 피어세션에 적극적으로 참여하며, 알고 있는 것이나 새로 알게 된 것에 대해 공유하였고, 팀원들의 말을 경청하며, 팀원들의 작업을 보고 많이 배우려고 했습니다.

### 3. 나는 어떤 방식으로 모델을 개선했는가?

#### A. Baseline 성능을 높이기 위한 다양한 실험 진행

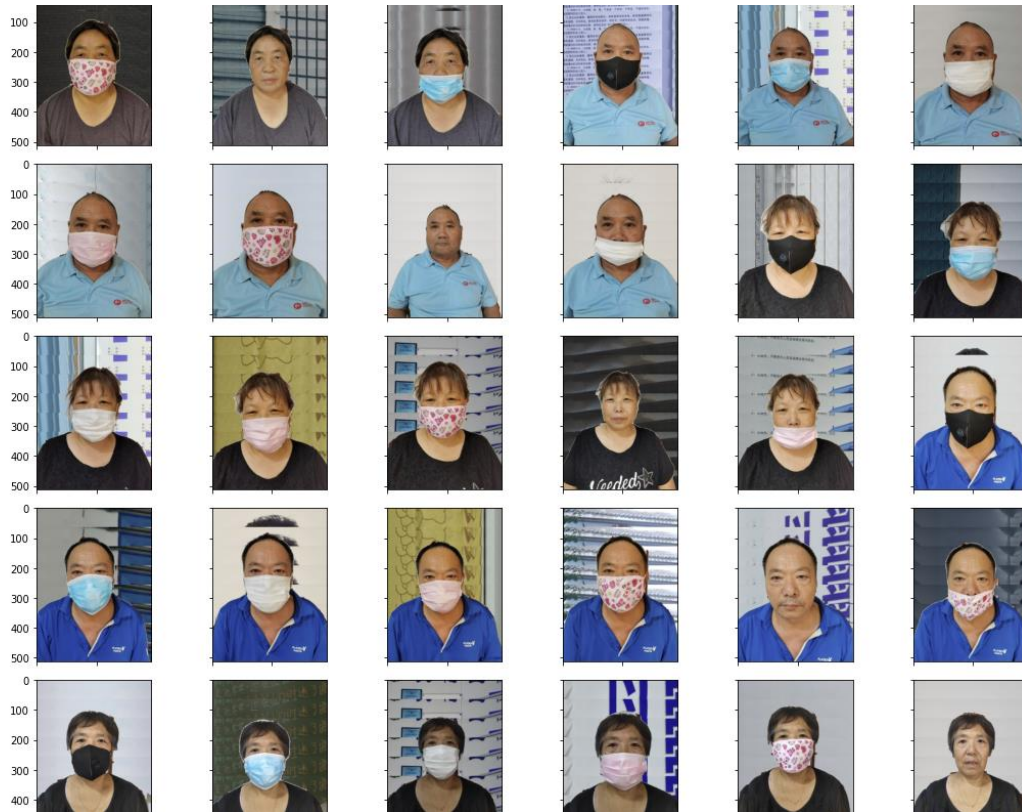
- ResNet-18을 비롯한 VGG-16, EfficientNet-b2, EfficientNet-b3, EfficientNet-b4로 실험을 진행하여 다른 모델들과 ResNet 사이의 큰 성능 차이가 없다는 것을 확인했습니다. 다양한 하이퍼 파라미터와 augmentation도 실험해봤습니다.

#### B. Dataset의 문제를 해결하기 위한 시도

- 데이터를 분석한 결과 유사한 나이대의 사람들이 유사한 배경에 있는 것을 확인하였고, 이는 모델이 배경을 바탕으로 판단하는 원치 않는 결과를 낼 수 있다고 판단했습니다.



데이터 양이 부족한 것도 문제였기 때문에, 배경을 random으로 합성한 이미지를 생성하여 데이터셋에 추가했습니다.



1 배경을 임의로 합성한 이미지

- Class Imbalance를 해결하기 위해 생성한 데이터와 기존의 데이터를 oversampling하여 분포를 비슷하게 맞추고, GaussianBlur, RandomRotation, RandomAffine, RandomHorizontalFlip을 적용하여 같은 이미지로 인식하지 못하게 하려고 했습니다. 또한 데이터의 증강을 바탕으로 모델이 너무 빠르게 train set에 overfitting 되는 문제들을 해결하고자 했습니다.
- 딥러닝적 관점으로 분류 에러에 근거한 loss에 가중치를 주어 문제가 있는 loss에 더 집중하는 방식으로 불균형한 클래스 문제를 해결하는 Focal Loss 적용했습니다



2 다양한 loss function 실험

### C. 성능을 높이기 위한 시도

- LB score가 높은 5개의 결과를 hard voting 방식으로 앙상블하여 최종 리더보드 score 0.765로 6위를 달성했습니다.

#### 4. 달성한 지점과 얻은 깨달음

좋은 팀원들이 있어 팀원들의 결과를 앙상블하여 LB에서 성과를 거둘 수 있었습니다. 협업을 하며 깃을 체계적으로 사용하는 방법을 배웠고 (PR, Issue 활용 등) 동일한 baseline으로 작업하며 남들이 알아보기 쉽고, 내 환경이 아닌 다른 환경에서도 내가 구현한 기능을 사용하기 쉽게 코드를 작성하는 연습을 할 수 있는 좋은 경험이었습니다. 프로젝트를 python script 형식으로 만들어 argument 형태로 parameter를 주고, 이것이 logging 되게 했던 것도 다양한 시도를 빠르게 해보고 결과를 공유하는데 큰 도움이 됐던 것 같습니다. 또한 데이터를 분석하며 문제점을 찾고, 이를 어떤 방식으로 해결해야 하는지, 어떤 시도들을 해볼 수 있는지 배울 수 있었습니다.

#### 5. 마주한 한계와 아쉬웠던 점

LB 제출 횟수가 정해져 있어서 새롭게 시도한 것들이 정말 효과가 있는지 확인하기 어려웠습니다. 그리고 Validation Score로 확인을 하면 사실 새롭게 적용한 방법들이 크게 효과가 있다는 것을 체감하기 어려웠습니다. 성능이 잘 나온 팀원이 한 방식으로 baseline을 잡고자 하였는데, 같은 방법으로 훈련을 진행하더라도 성능을 재현하는 것이 어려웠습니다.

#### 6. 다음 P-Stage에 새롭게 시도해볼 것

먼저 팀적으로 통합된 baseline을 잡는 것이 중요하다고 느꼈습니다. 이 통합된 baseline을 이용하여 각자 실험을 분업하여 진행하고, 성능이 얼마나 올라갔는지 내려갔는지 확인하는 것이 좋을 것 같습니다. 또한 '해봤더니 잘됐다'가 아닌, 원인을 찾고 이에 대한 해결 방안을 찾는 방법으로 대회를 진행해야 할 것 같습니다. 이번 대회에서는 제가 했던 시도들이 효과가 없는 경우가 많았는데, 다음에는 이번 대회의 경험을 살려서 조금 더 정밀한 분석을 하고, 발전된 방법을 이용하여 성능 개선에 큰 도움을 주고 싶습니다.