

Palette.AI

Classification of Mask wearing status, Gender and Age

8조: 김지수 김혜수 배지연 이승현 임문경 진명훈



Mask

☒ Wear ☐ Incorrect ☐ Wear

Gender

☐ Male ☒ Female

Age

☒ <30 ☐ >=30 & <60 ☐ >=60

2021.08.23 – 2021.09.02

Public 2등, Private 3등 Solution (38조 中)



We are Palette.AI !!

01

Introduction

02

Baseline

03

Improvement

04

Conclusion





- 어떤 Task를 풀었는가?
- 학습 데이터는?
- Exploratory Data Analysis

우리가 어떤 문제를 풀었는지,
그리고 문제에 대해 소개해드릴게요!



01. Introduction



01

우리가 누구?

Introduction



🎨 AI Palette

“색칠을 하다 보면 그림이 완성되어 있을 거예요”

⚙️ Members

김지수	김혜수	배지연	이승현	임문경	진명훈
					
Github	Github	Github	Github	Github	Github

⚙️ Contribution

김지수	Modeling • Feature Engineering • Age-specific model • EfficientNet Master • Out of fold
김혜수	Dataset curation • Construct Pipeline • Mental Care • Data license verification
배지연	Dataset generation • Dataset curation • Mask synthesis • Hyperparameter tuning
이승현	Team Management • Dataset preprocessing • Modeling • Make task-specific loss
임문경	EDA, Modeling • Visualizing • Search augmentation technique • MLops
진명훈	Modeling • Active Learning • Mentoring • Huggingface pipeline • Handling imbalance problem

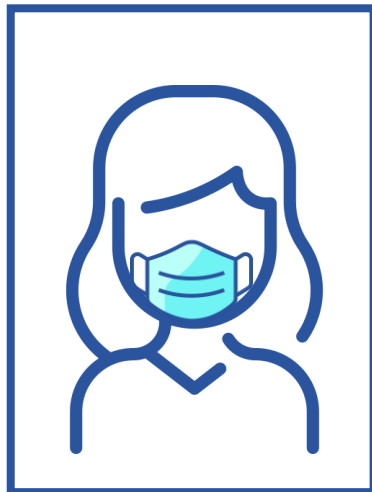


마스크 착용 상태 분류!

카메라로 촬영한 사람 얼굴 이미지의 마스크 착용 여부를 판단하는 Task

정면의 얼굴이 찍힌 사진

384 X 512



Mask

☒ Wear ☐ Incorrect ☐ Wear

Gender

☐ Male ☒ Female

Age

☒ <30 ☐ >=30 & <60 ☐ >=60

마스크 착용 / 오착용 / 미착용

남성 / 여성

30대 미만 / 30대 이상 60대 미만 / 60대 이상



01

학습 데이터는?

Introduction

Boostcamp에서 제공한 Asian Mask Dataset!

- 전체 사람 명 수: 4,500 명
- 한 사람 당 사진의 개수: 7 장 (마스크 착용 5장, 코스크 or 텍스크 1장, 미착용 1장)
- 이미지 크기: (384 x 512)
- 전체 데이터 중 60%를 학습 데이터로, 20%는 public, 20%는 private
 - Train: 18,900
 - Test1: 6,300 (@Public)
 - Test2: 6,300 (@Private)
- 평가 지표: Macro F1 Score



Class 1	Mask	Gender	Age
0	Wear	Male	< 30
1	Wear	Male	>= 30 and < 60
2	Wear	Male	>= 60
3	Wear	Female	< 30
4	Wear	Female	>= 30 and < 60
5	Wear	Female	>= 60
6	Incorrect	Male	< 30
7	Incorrect	Male	>= 30 and < 60
8	Incorrect	Male	>= 60
9	Incorrect	Female	< 30
10	Incorrect	Female	>= 30 and < 60
11	Incorrect	Female	>= 60
12	Not Wear	Male	< 30
13	Not Wear	Male	>= 30 and < 60
14	Not Wear	Male	>= 60
15	Not Wear	Female	< 30
16	Not Wear	Female	>= 30 and < 60
17	Not Wear	Female	>= 60

Mask

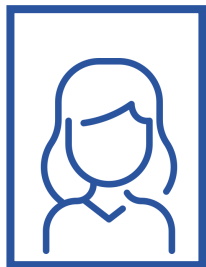
☒ Wear
 ☐ Incorrect
 ☐ Wear

Gender

☐ Male
 ☒ Female

Age

☒ <30
 ☐ >=30 & <60
 ☐ >=60



Normal



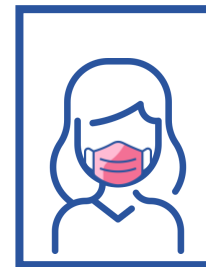
Mask1



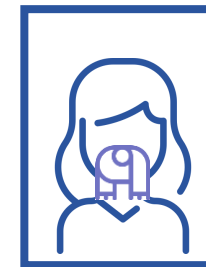
Mask2



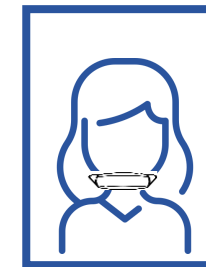
Mask3



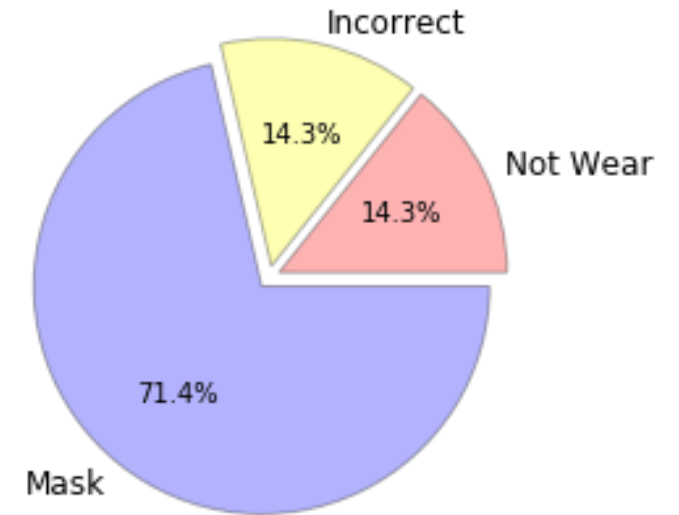
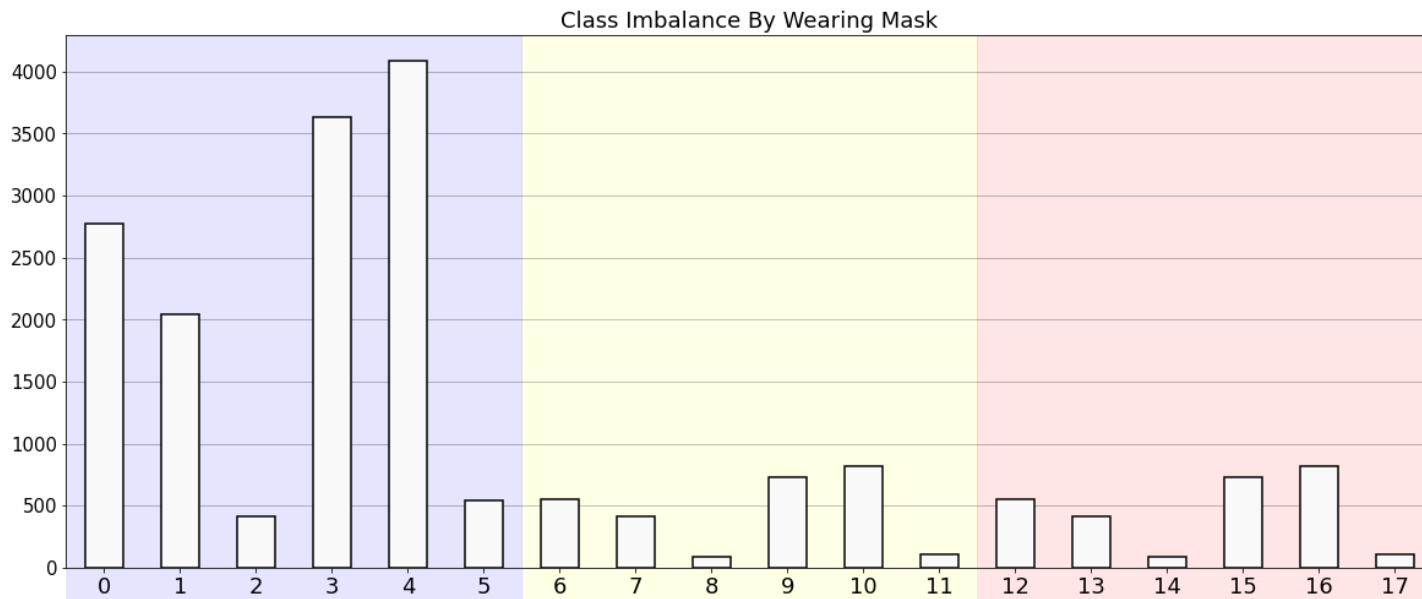
Mask4



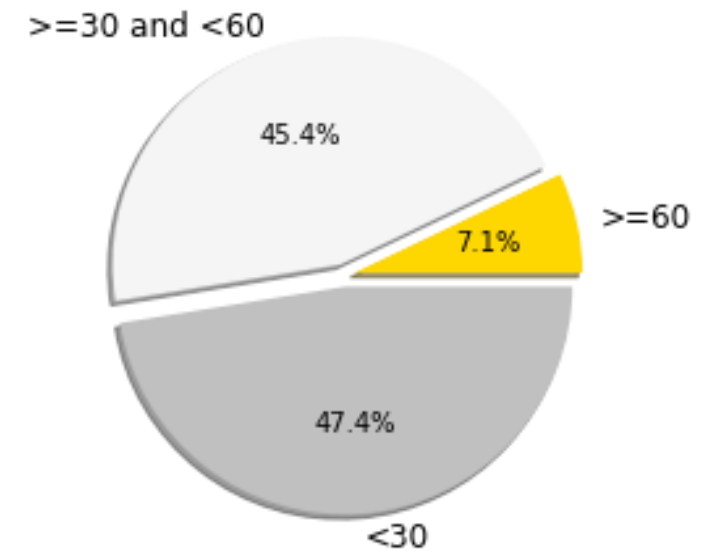
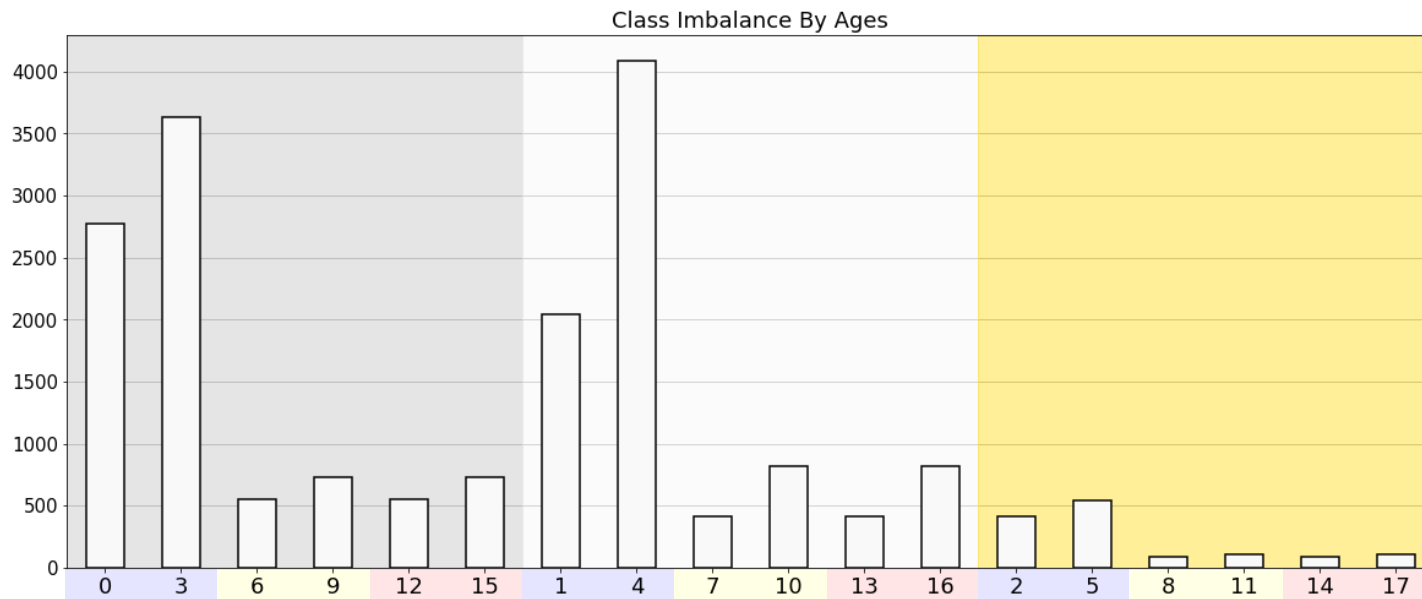
Mask5



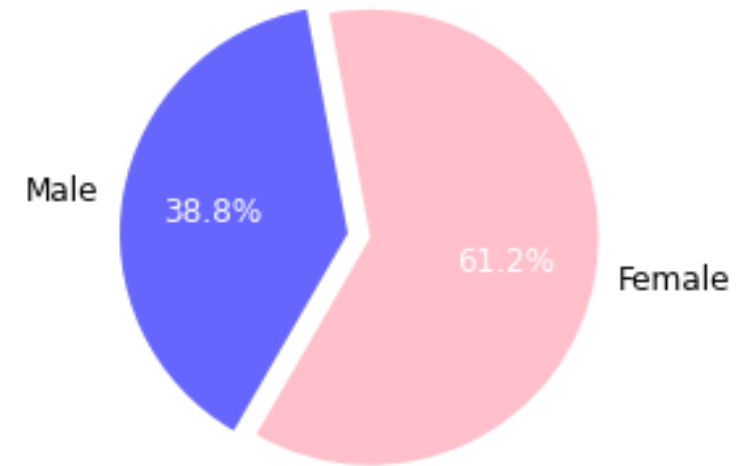
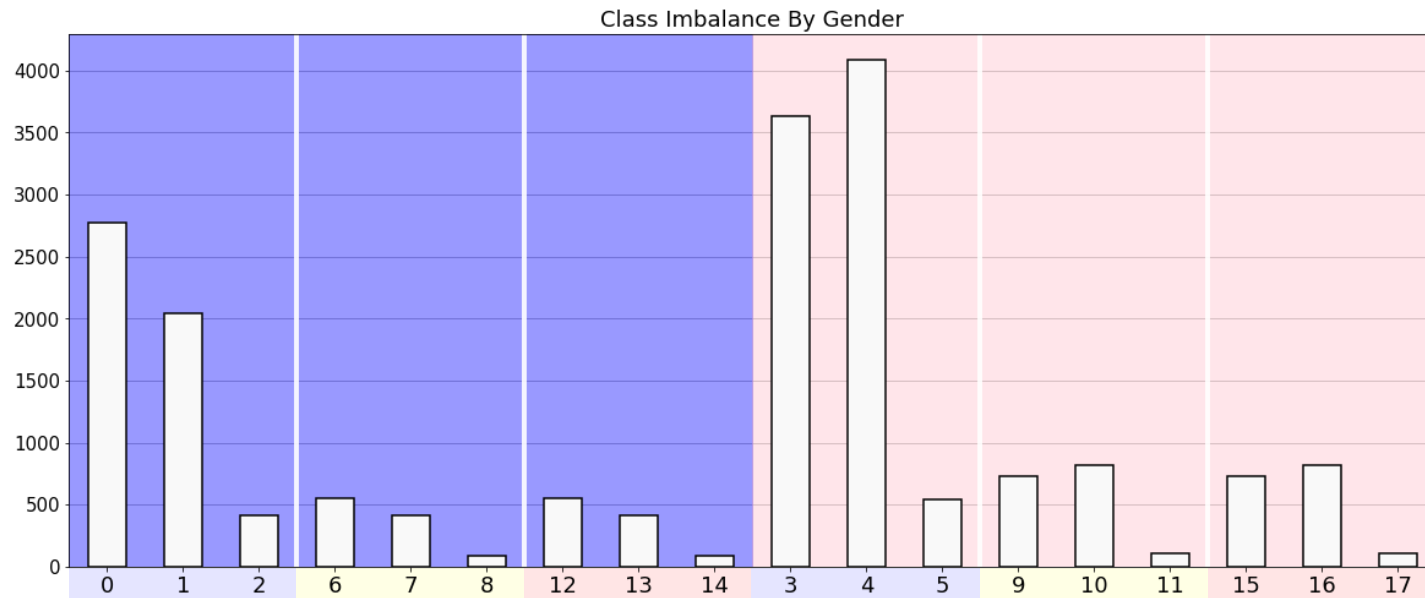
Incorrect



Mask Class Imbalance 문제가 심각해 보입니다.



Age에 따른 불균형 문제는 더 심각해 보여요



Gender에도 불균형이... 잘 예측할 수 있을까요?



Baseline

- 베이스 라인: 모델
- 베이스 라인: 실험 세팅
- 베이스 라인: Best Models
- 베이스 라인: Insights

1주차 (08.23 ~ 08.27)에 구축한
저희의 Baseline에 대해 소개드릴게요!
그리고 저희가 얻은 Insight도...!



02. Baseline



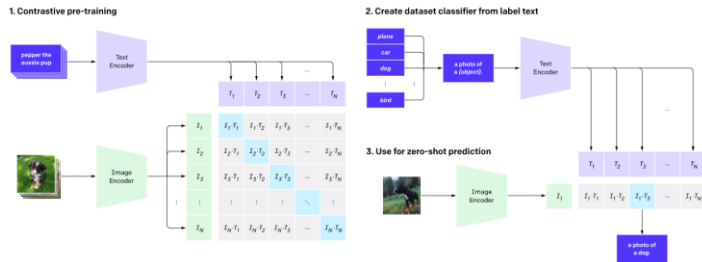
02

베이스 라인: 모델

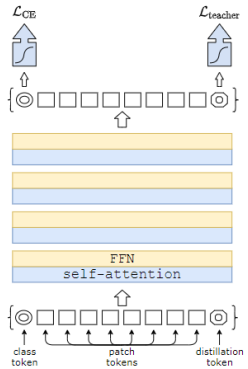
Baseline



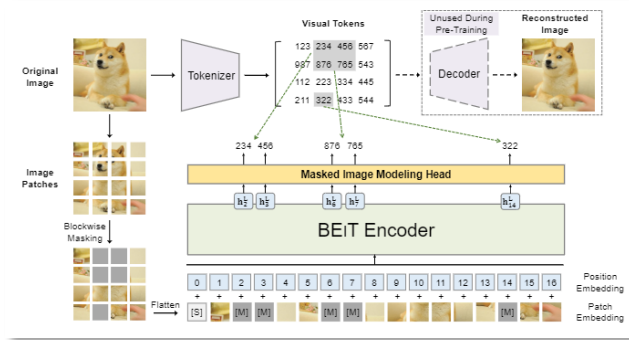
CLIP (OpenAI, '21)



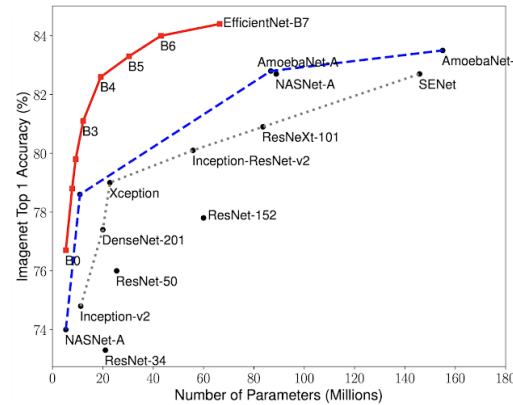
DeiT (FAIR, '21)



BEiT (MS, '21)



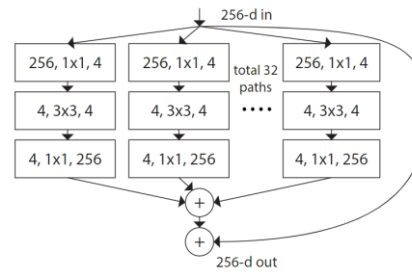
EfficientNet (Google Brain, '19)



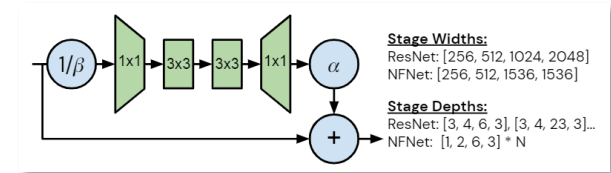
ResNet50 (MS, '15)



ResNeXt (US San Diego, FAIR, '17)



NFNet (DeepMind, '21)



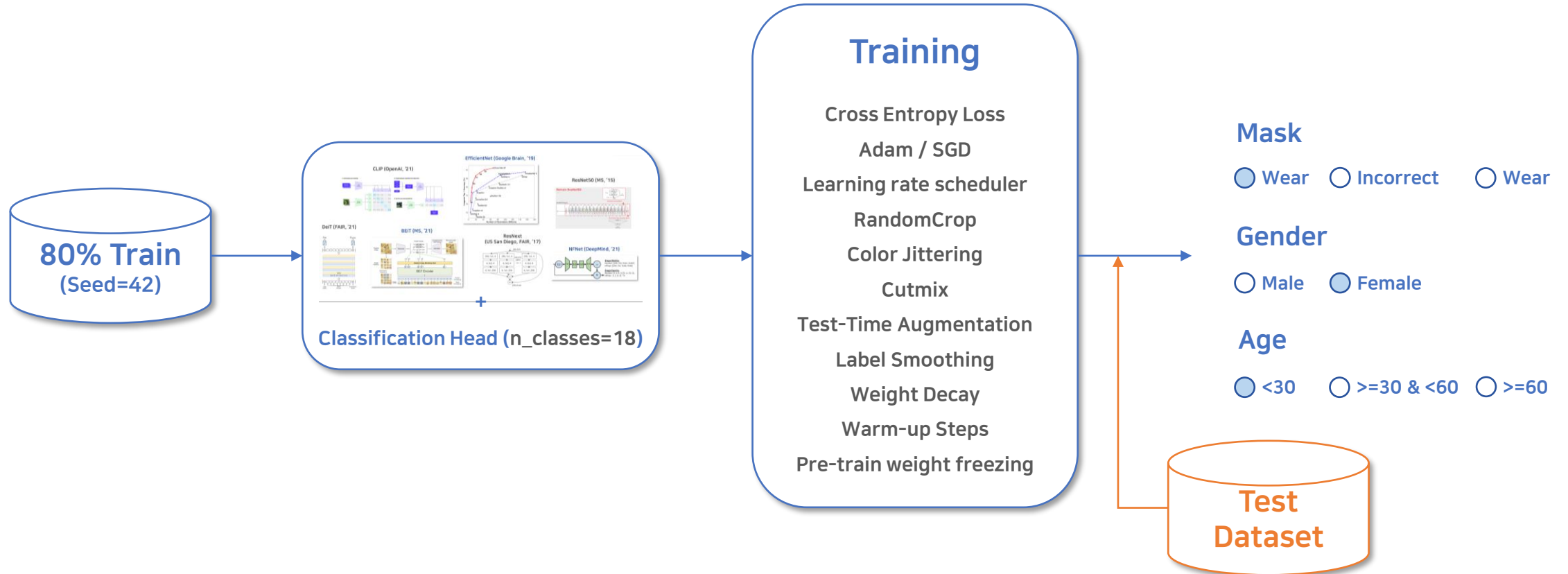
초반엔 어떤 모델이 잘하는지 알기 위해 다양하게 실험했어요!



02

Baseline

베이스 라인: 실험 세팅



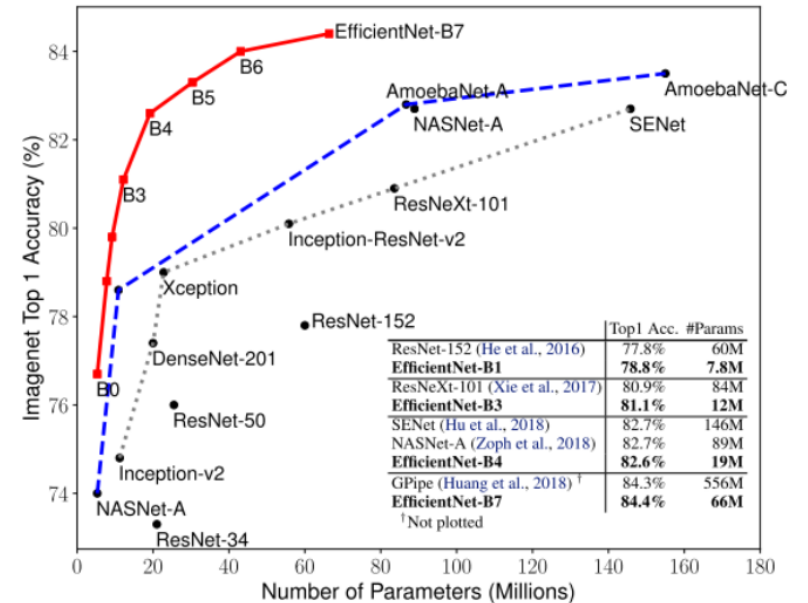
18개의 클래스를 분류하는 모델을 다양한 조건에서 실험!



제일 잘하는 모델 3개에 대해 Study를 진행했어요!

EfficientNet (Google Brain, '19)

- <https://arxiv.org/pdf/1905.11946.pdf>
- Depth(d), Width(w), Resolution(r)을 조절하여 성능을 향상시킨 모델
- 기존의 모델 (ResNet, MobileNet)에서는 Depth scaling, Width scaling만을 조절하여 성능을 향상
- EfficientNet은 Resolution scaling을 통해 성능 향상이 잘 된다는 것을 증명
- EfficientNet은 Depth scaling, Width scaling, Resolution scaling을 동시에 조절했을 때 가장 최적의 성능을 찾아내 학습을 진행
- Image Classification Task에 좋은 성능을 보이는 EfficientNet으로 Pre-Training





02

베이스 라인: Best Models

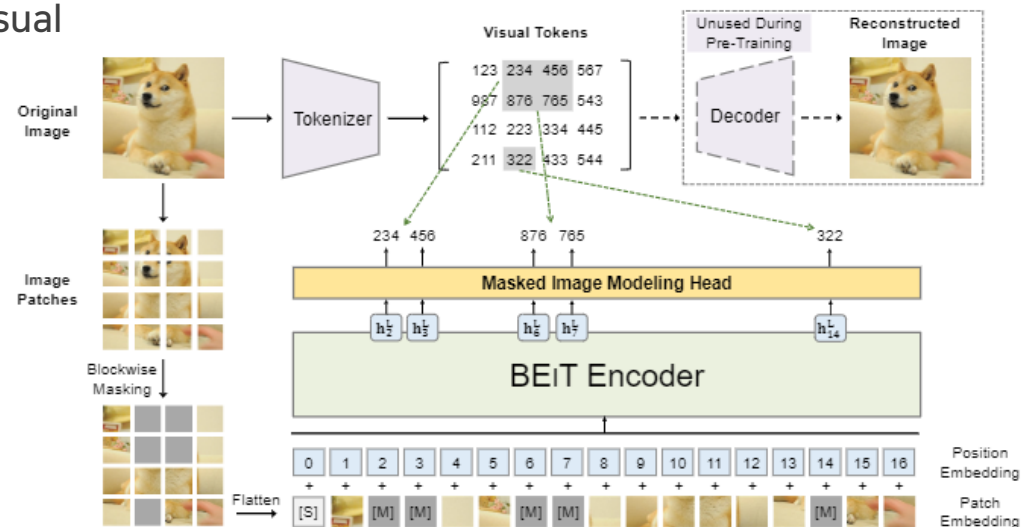
Baseline



제일 잘하는 모델 3개에 대해 Study를 진행했어요!

BEiT (MicroSoft, '21)

- <https://arxiv.org/pdf/2106.08254.pdf>
- ViT를 BERT의 **Masked Language Modeling**에서 영감받은 **Macked Image Modeling**으로 Pre-Training한 모델
- DALL-E의 **discrete-VAE**로 image를 tokenize하여 pre-training label 구축
- 224*224 image를 resolution이 16*16인 196개의 patch으로 쪼갬
- 196개의 patch embedding 중 40% 정도를 **blockwise masking** 실시
- BEiT Backbone model에 **Language Model Head**를 붙여서 masking token과 visual token 간의 Cross Entropy Loss로 사전 학습 실시 (총 8,092개의 vocab을 가짐)
- 데이터가 많이 필요한 ViT의 단점을 개선!

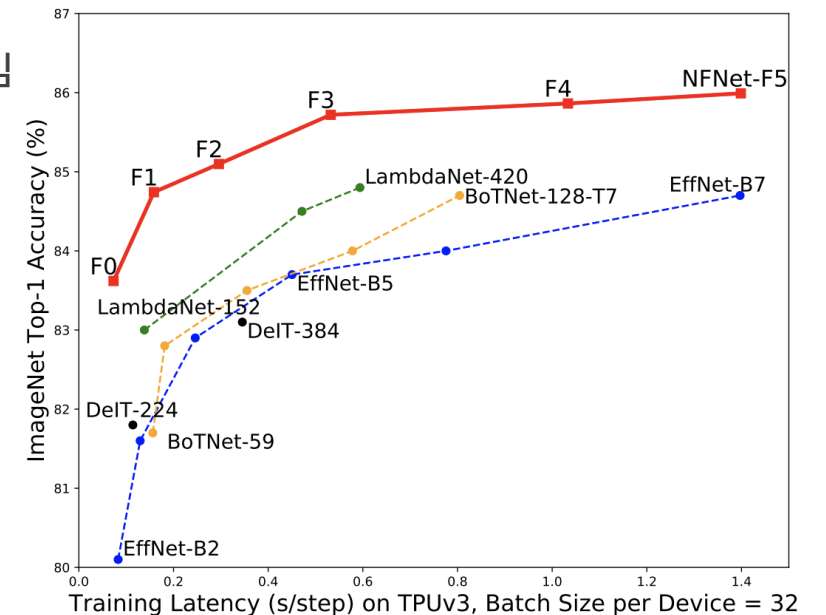




제일 잘하는 모델 3개에 대해 Study를 진행했어요!

NFNet (DeepMind, '21)

- <https://arxiv.org/pdf/2102.06171.pdf>
- BatchNorm을 하지 않고 AGC, Dropout, SAM 등을 이용한 모델
- Full name: Normalizer Free Network
- Batch Normalization은 아래와 같은 단점이 존재
 - 분산 학습에서 구현 Error가 잦음
 - Expensive Computational Cost
- Adaptive Gradient Clipping (AGC)를 제안, 모델이 Large Batch에도 안정적인 성능을 보임
- 가중치에 gradient가 변했는지를 보고 학습의 안정성을 평가
- 이 후 Gradient Clipping 실시
- 최종적으로 단일 모델로 Best Performance를 보여준 모델!





02

베이스 라인: Insights

Baseline



각 모델이 Ensemble하지 않고도 잘했어요! (최고 73.8점 + 개인 리더보드 9위 / 255명)

하지만 이 모델로 마스크 착용 여부 판별을 정확하게 판단하기엔 일반화도 그렇고 설명력이 부족한 것 같아요...

우리의 모델은 어떤 Sample에 대해 틀렸을까요?

1. 60대 이상 분들을 30대 이상 60대 미만으로 잡는 경우가 빈번했어요
2. 특정 복장(검은 옷 + 빨간 외투)을 입은 성별에 대해 구별하지 못했어요
(같은 복장일 경우 해당 옷을 입은 남성분들을 여성으로 판별하는 경우가 엄청 많았어요!)
3. Incorrect Mask(턱스크 or 코스크) 판별을 힘들어 했어요
(심지어 Not Wear과 class 수가 비슷했음에도 불구하고요!)

아하 역시 EDA에서 모델이 헛갈릴 것이라 생각했던 부분들이었네요!

이 부분을 어떻게 잡느냐가 관건으로 보여요!

문제점은 확인했고 코드도 구축했으니 성능 향상에 도전!



Improvement

- 학습 개선: ① Re-labeling
- 학습 개선: ② Handling Class Imbalance
- 학습 개선: ③ Age 모델을 따로 구축
- 학습 개선: ④ Data Augmentation

2주차 (08.30 ~ 09.02)에 진행한
일반화를 고려한 성능 개선 방안 소개!



03. Improvement



03

학습 개선: ① Re-labeling

Improvement

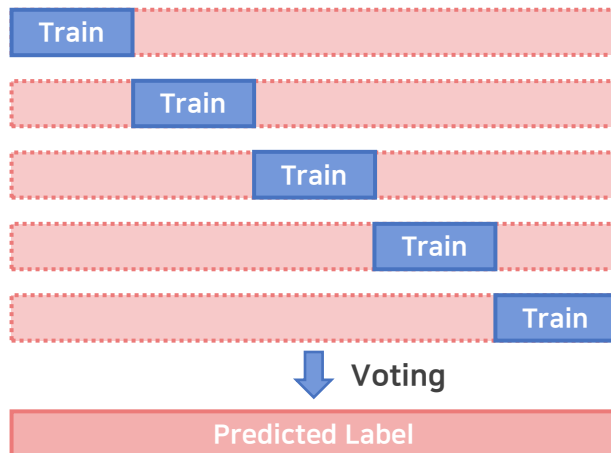


우선 **Active Learning**에서 Idea를 얻어 **Noise Labeling**을 교정하는 작업을 거쳤어요!

인간이 모든 데이터를 보는 것은 불가능하기 때문에, 아래와 같은 과정으로 **Cleaning** 작업을 수행했어요

- Train Dataset을 5개의 Fold로 나눈다
- BEiT 모델을 5개 만들어서 각 Fold(20%)의 데이터만을 학습시킨다
- 각 모델은 학습하지 않은 ouf-of-fold(80%)를 예측한다
- 결과로 생긴 sample 당 4개의 new label을 Voting한다 (2:2일 경우 틀린 label로 간주한다)
- 원래의 label과 비교하여 틀린 label을 체크한다
- 틀린 label을 인간이 직접 data를 살펴보고 해당 데이터에 맞는 label을 부여한다

위 과정을 통해 총 18,900장의 Image에서 138장의 후보를 찾았고 그 중 틀린 69장의 이미지의 Label을 교정했습니다!



VS

Golden Truth

	path	mis_label	label
106	../input/data/train/images/000020_female_Asian...	10	16
107	../input/data/train/images/000020_female_Asian...	16	10
5348	../input/data/train/images/001498-1_male_Asian...	3	0
5349	../input/data/train/images/001498-1_male_Asian...	15	12
5350	../input/data/train/images/001498-1_male_Asian...	9	6
...
17635	../input/data/train/images/006504_female_Asian...	12	15
17636	../input/data/train/images/006504_female_Asian...	0	3
17637	../input/data/train/images/006504_female_Asian...	0	3
17638	../input/data/train/images/006504_female_Asian...	6	9
17639	../input/data/train/images/006504_female_Asian...	0	3

69 rows × 3 columns



03

학습 개선: ② Handling Class Imbalance



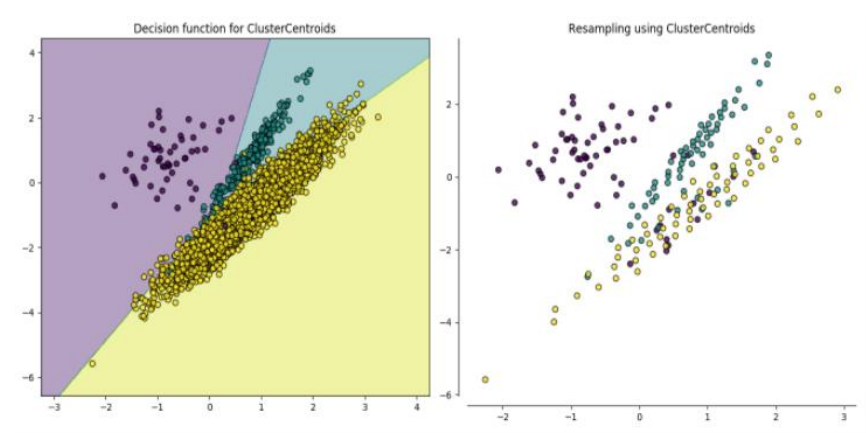
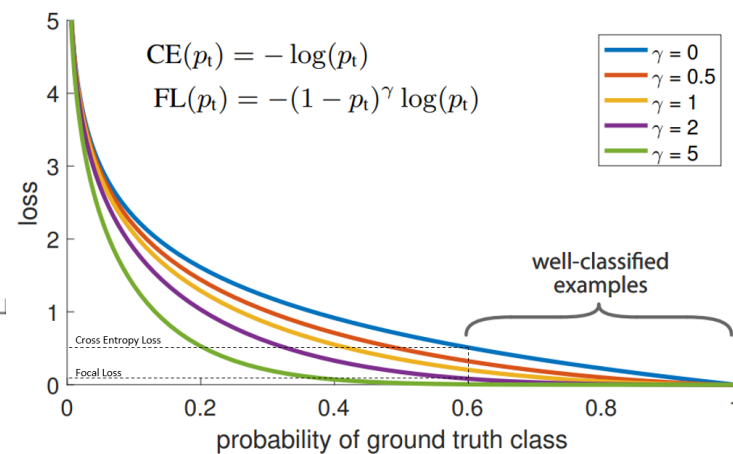
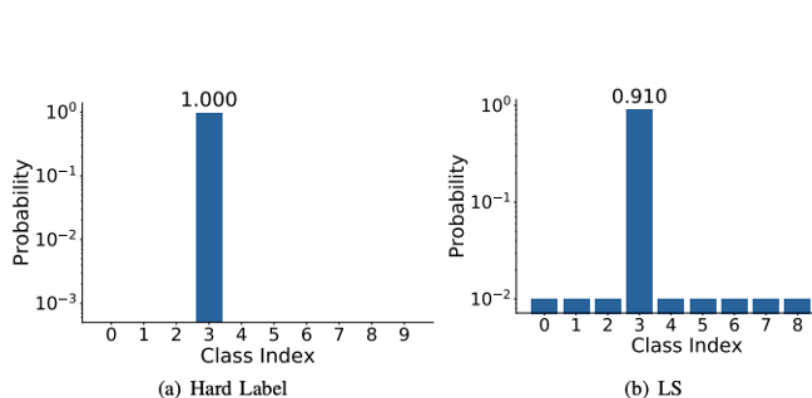
Improvement

다음으론 Class Imbalance 문제가 심각했기에 이를 해결하기 위한 다양한 방법들을 적용했어요!

- Label Smoothing
- Weighted Cross Entropy Loss
- Focal Loss
- Weighted Sampling

Accuracy는 떨어졌으나 Class Imbalance 를 이전 모델보다 훨씬 잘 잡아냈어요! F1 Score가 소폭 상승했습니다 ☺

가중 샘플링 기법의 경우엔 1% 가량 큰 성능 향상을 거뒀어요!



<https://dining-developer.tistory.com/27>

<https://www.tutorialexample.com/implement-focal-loss-for-multi-label-classification-in-tensorflow/>
<https://towardsdatascience.com/label-smoothing-make-your-model-less-over-confident-b12ea6f81a9a>



② 번으로 Class Imbalance 문제가 전부 해결되진 않았어요! 3가지의 Imbalance 중 60대 분들을 구분 짓는 일이 너무 힘들었죠...

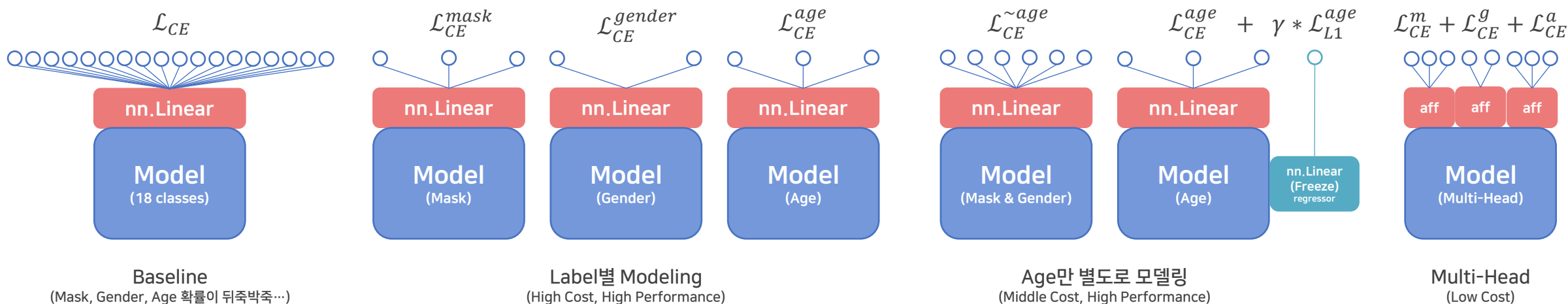
이에 대해 문제를 다른 시선으로 봤어요!

- 18개의 Class를 구분하는 **Multi-Class Problem**이 아닌
- 각각 3개(**Mask**), 2개(**Gender**), 3개(**Age**)의 Class를 가지는 **Multi-Label Problem**으로 해석했죠!

그러나 위 방법은 모델을 3개나 구축해야 하기 때문에 용량 및 inference time이 늘리는 만큼 증가하는 단점이 있어요...

때문에 아래 두 방식을 추가해서 Multi-Head 문제를 구현했습니다

- ✓ **Model 3개**를 사용하여 각각 Label을 예측 (**High Cost & High Performance**)
- ✓ **Mask & Gender**를 동시에 예측하는 **모델 1개**와 **Age만을 예측**하는 모델 구축 (**Middle Cost & High Performance**)
Age 모델의 경우, 60대를 더욱 잘 예측시키기 위해 Regression Task도 같이 풀었습니다!
- ✓ Backbone은 그대로 두고 **nn.Linear**를 3개 두고 해당 head만 update (**Low Cost**)





03

학습 개선: ④ Data Augmentation



Improvement

Age만을 예측하는 모델로 0.7% 정도 성능 개선이 있었지만 incorrect, 60대 분들, gender 예측 개선이 좀 더 필요해요!
이를 위해 아래 세 데이터 셋에 label을 추가하여 위 3가지 class imbalance 문제를 풀고자 했습니다
시간 관계 상 Face Mask는 특정 모델에만 먹여줬으며 Data Augmentation을 수행했을 때 큰 성능 향상을 거뒀습니다.

- **Mega Age**: train + test 약 8,000건
 - <https://www.kaggle.com/baopmessi/megaage>
 - 파일의 나이 정보를 토대로 60대 이상만 추출
 - MaskTheFace SW로 마스크 합성
 - Pseudo Labeling으로 성별 합성
 - 대회 규정의 Commercial Free에 의해 사용하지 X
- **All Age Face**: Original Images 13,322건
 - <https://github.com/JingchunCheng/All-Age-Faces-Dataset>
 - 2~80세까지 다양한 연령의 얼굴 사진
 - MaskTheFace SW로 마스크 합성
- **Face Mask**: 250,000 Images, 4 types of mask worn, 28,000 unique faces
 - <https://www.kaggle.com/tapakah68/medical-masks-p4>
 - Incorrect * 2, Mask * 1, Gender * 1

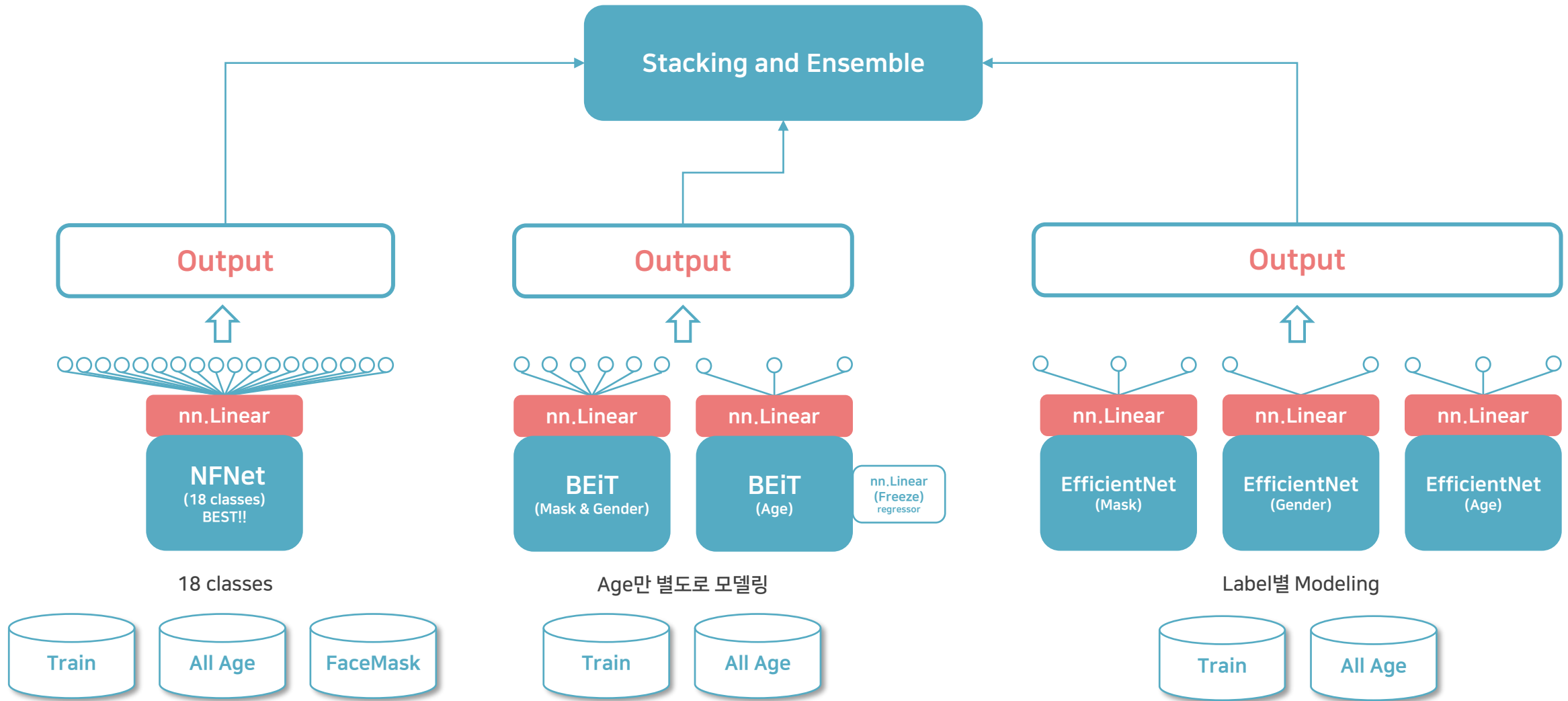


Conclusion

- Our Solution
- 최종 결과는?
- 아쉬운 점 + 중요했던 점

04. Conclusion





+ Label Smoothing + Focal Loss + Weighted Sampling + ...











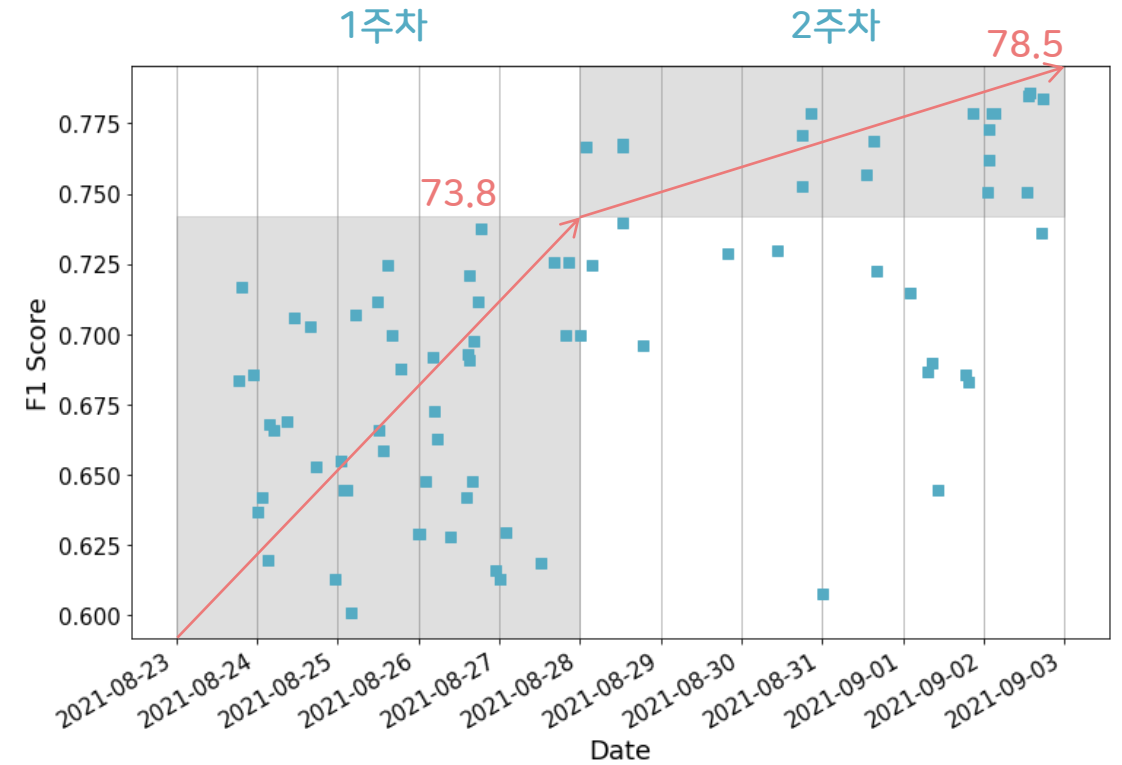
04

최종 결과는?

Conclusion



			Macro F1	Accuracy	
Public	1	이미지분류_38조		0.786	81.762
	2	이미지분류_8조		0.784	81.619
	3	이미지분류_16조		0.781	81.921
	4	이미지분류_28조		0.779	82.095
Private	1	이미지분류_16조		0.776	82.238
	2	이미지분류_33조		0.771	81.524
	3	이미지분류_8조		0.771	81.254
	4	이미지분류_6조		0.769	81.492



전략적으로 접근하여 성능 향상 및 일반화 도모!



아쉬웠던 점

- Ray Tune에 대해 내부적으로 스터디를 진행했으나 적용하지 못한 점
- MC-Dropout, FixMatch 등 다른 일반화 기법에 대한 고민이 적었던 점
- Pseudo Labeling 이외에 Semi-Supervised Learning으로 접근했다면?
- BEiT에 Further Pre-Training을 적용하고 데이터를 더 태워줬을 때의 성능은?

프로젝트를 진행하며 느낀 점과 중요했던 점

- Model Architecture는 생각보다 중요하지 않았음
- 데이터에 대한 이해와 어떤 문제를 지니고 있는지를 파악하는 것 (Insight 발견과 문제 풀이를 위한 방법들)
- Test Dataset과 Validation Dataset 간의 Alignment
- WanbB 등의 Tool로 실험 관리
- 팀원들 간의 Communication과 전략적인 시간 활용

한 달간 너무 뜻깊은 시간을 보냈습니다!

Thank you!

