



# 8/26(목) 회의록

|           |                        |
|-----------|------------------------|
| 🕒 작성일시    | @2021년 8월 20일 오전 9:49  |
| 👤 작성자     | <span>하람</span> 이하람    |
| 👥 참석자     |                        |
| 🕒 최종 편집일시 | @2021년 8월 27일 오전 12:52 |
| 📌 회의 유형   | 일일 회의                  |

## 👉 학습 내용 공유

### 1. 대회 내용 토론 🙋

- 하람)
  - efficientnet-b3 = 0.6471(f1-score)
  - efficientnet-b3 + ImbalancedDatasetSampler + weighted loss = 0.6817(f1-score)
  - from torchsampler import ImbalancedDatasetSampler 사용

```
def get_labels(self):  
    return self.labeled_df['label']  
  
train_data = TrainDataset(trainimage_dir, train, transform)  
valid_data = TrainDataset(trainimage_dir, valid, transform)  
  
batch_size = 32  
train_loader = DataLoader(train_data, batch_size=batch_size, num_workers=4, sampler=ImbalancedDatasetSampler(train_data))  
valid_loader = DataLoader(valid_data, batch_size=batch_size, num_workers=4, shuffle=True)
```

- weighted loss

```
weight = torch.tensor([1., 1., 2.] * 6).to(device)  
loss = F.cross_entropy(scores, targets, weight=weight)
```

- 마스크, 성별, 나이 모델을 나눠서 학습하시는 분들은 외부 데이터셋 사용해보시는 건 어떨지?

- 마스크 : <https://www.kaggle.com/andrewmvd/face-mask-detection>

- 나이 : <https://www.kaggle.com/frabbisw/facial-age>

- 보성) LR Scheduler 추천: [OneCycleLr...](#)

보성) 해볼 것 : vit (pretrained, deeeep) , 데이터 셋 가지고 놀기

## 2. 건의사항 🙄

**wandb 써봐요! :**

슬랙 스레드에 이메일 적어주세요...! 나중에 초대드릴게요...!

## 3. 멘토링

함수는 메인에서 사용하자.

각 파일 내부에서 변수를 지정해주고 하는건 그리 좋은 생각이 아닙니다.

아마 컨스턴트는 어트리뷰트로 잡으면 될듯.

클래스 임밸런스를 어떻게 잡아야할까?

1. 데이터 셋 자체를 오버샘플링, 언더샘플링,
2. 학습할때 특정 클래스에 웨이트를 줄 수 있다.  
옵티마이저 옵션에 클래스 가중 주는게 있습니다.  
ex.) 9번 클래스 데이터가 적다. 9번 클래스에 대한 로스에 더 크게 곱한다던가....?  
분포를 먼저 구하고, 분포를 정규화하던가 해서- 데이터 수가 적은 순으로 해도 괜찮은 듯.
3. 모든 클래스를 따로 해버려도 되긴 하다.
4. 그룹별로 묶어버려도 된다
5. 그 클래스가 점수에 얼마 영향을 미치는지 생각해봐라.
6. `nn. cross_entropy(outputs, labels, weight=weight)`