

미니프로젝트 3차 조별 발표

AI 11반 42조



문제정의

- 조별로 이해한 프로젝트 목표를 간단히 작성해주세요
 - ✓ DataSet의 크기와 Model 크기 간 상관관계를 파악한다.
 - ✓ CNN Model과 Transformer Model 의 차이점을 이해한다.
 - ✓ Wandb를 통해 모델 학습 log를 공유한다.
 - ✓ 전이학습을 이해하고 결과를 도출한다.

과정 요약

- ✓ 어려운 테스트를 위해 테스트 데이터셋을 추가하였습니다.
- ✓ CutMix와 Elastic Transform을 통해 파손된 차량의 클래스를 추가하였습니다.
- ✓ 불균형 데이터셋을 처리하기 위해 Focal Loss 사용

데이터 보강

- 제공된 데이터셋의 협소한 규모(약 600장)
- 촬영 이미지가 파손 부위에 집중
- 딥러닝 모델에게 쉬운 Task로 판단 (이미지 참조)
- AI HUB를 통해 파손 강도가 약한 1200장

Confusion Matrix :

```
[[60  0]
 [ 0 37]]
```

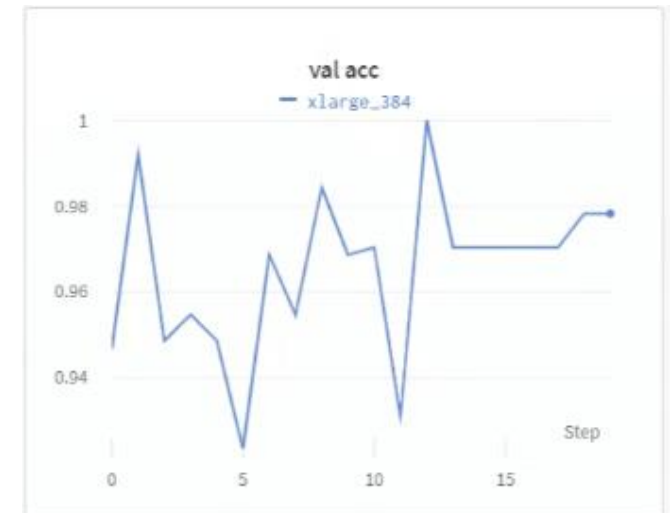
Accuracy : 1.0

Recall : 1.0

Precision : 1.0

F1 Score : 1.0

| | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| 0 | 1.00 | 1.00 | 1.00 | 60 |
| 1 | 1.00 | 1.00 | 1.00 | 37 |
| accuracy | | | 1.00 | 97 |
| macro avg | 1.00 | 1.00 | 1.00 | 97 |
| weighted avg | 1.00 | 1.00 | 1.00 | 97 |



>> 일반적인 경우를 고려해 데이터를 추가 보강



데이터 전처리

- 조에서 공유한 데이터 전처리 방법을 나열

- 히스토그램 평활화 (Histogram Equalization)



- 평활화 이후 파손 특성이 지나치게 강화되어 배제하였음.

- 정규화(Normalization)

- Image Augmentation

- Rotation
 - Shift
 - RGB Shift
 - Mixup
 - CutMix
 - Elastic Transform

Data Augmentation

- MixUp



- CutMix

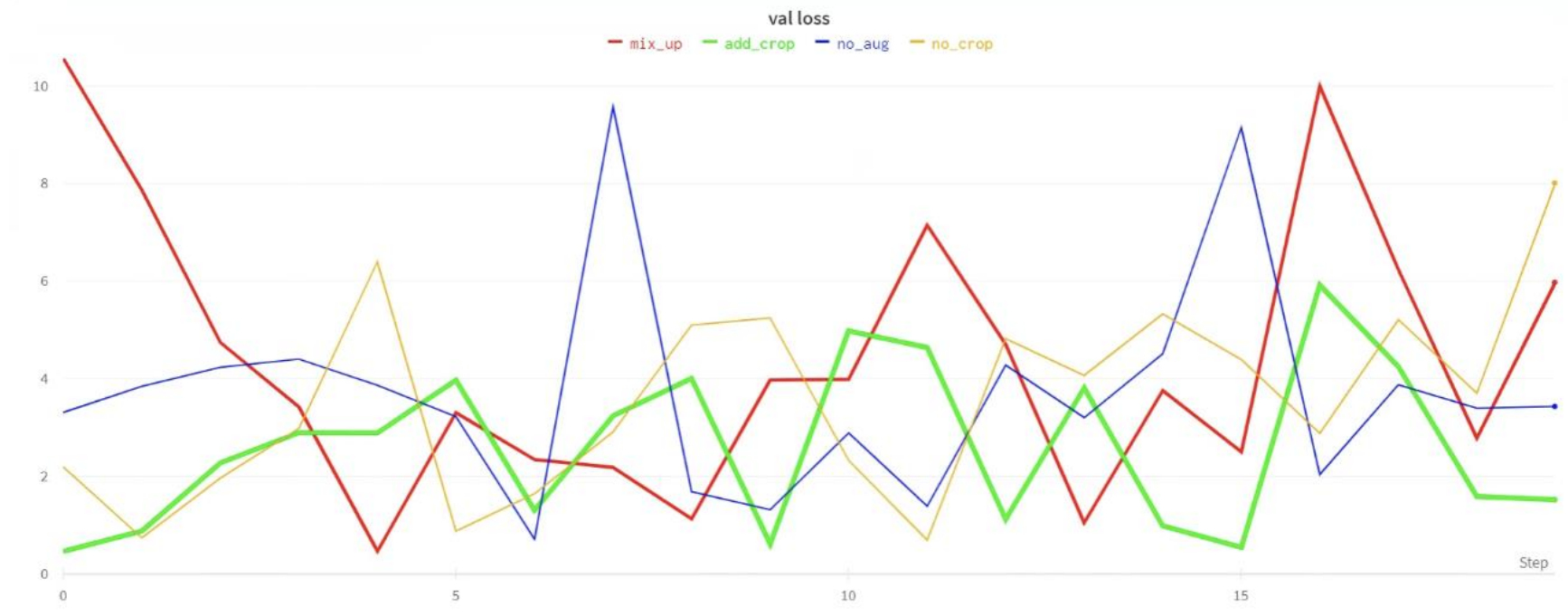


- Elastic Transform



Data Augmentation

- BASE MODEL : Resnet 34



- 새로 추가한 데이터의 파손된 부분이 작고, train의 파손된 부분이 크기 때문에 crop류의 augmentation이 성능이 좋다. 또한 mix_up은 오히려 파손되지 않은 데이터셋을 파손되게 느끼게 하기 때문에 성능이 떨어졌다.

모델링 (1) – CNN

✓ CNN

Layer의 깊이가 깊어질수록 깊은 특성값을 학습

즉, 데이터셋이 작은 지금 layer가 깊은 모델은 일반화되기가 어려워 좋은 성능을 내지 못한다.

Resnet34



Convnext



그외 CNN 모델링

✓ VGG

Vgg A를 tensorflow로 구현하여 acc를 측정하였으나 resnet보다 좋지 않았다.

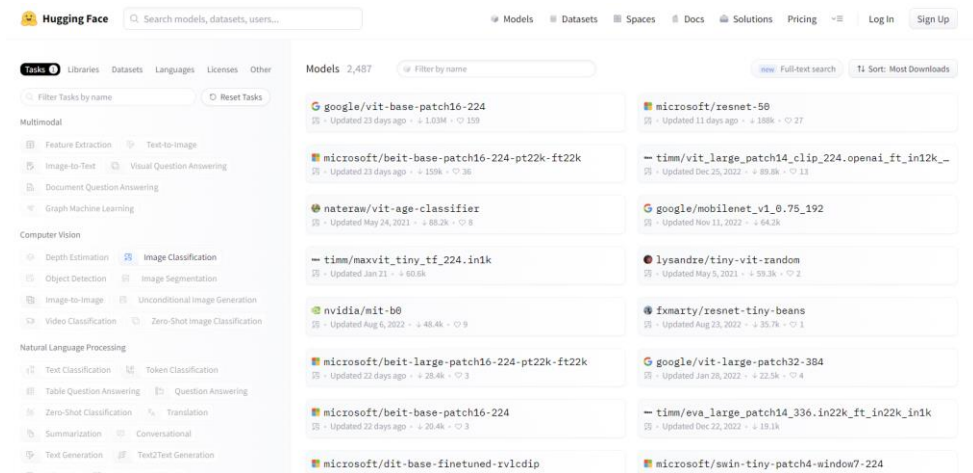
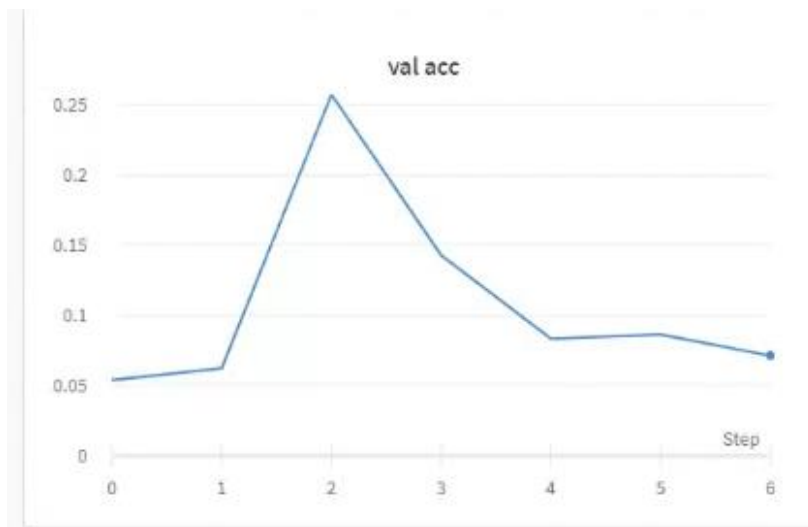
| ConvNet Configuration | | | | | |
|-----------------------------|------------------|------------------|------------------|------------------|------------------|
| A | A-LRN | B | C | D | E |
| 11 weight layers | 11 weight layers | 13 weight layers | 16 weight layers | 16 weight layers | 19 weight layers |
| input (224 × 224 RGB image) | | | | | |
| conv3-64 | conv3-64 | conv3-64 | conv3-64 | conv3-64 | conv3-64 |
| | LRN | conv3-64 | conv3-64 | conv3-64 | conv3-64 |
| maxpool | | | | | |
| conv3-128 | conv3-128 | conv3-128 | conv3-128 | conv3-128 | conv3-128 |
| | | conv3-128 | conv3-128 | conv3-128 | conv3-128 |
| maxpool | | | | | |
| conv3-256 | conv3-256 | conv3-256 | conv3-256 | conv3-256 | conv3-256 |
| conv3-256 | conv3-256 | conv3-256 | conv3-256 | conv3-256 | conv3-256 |
| | | | conv1-256 | conv3-256 | conv3-256 |
| | | | | conv3-256 | conv3-256 |
| | | | | conv3-256 | conv3-256 |
| maxpool | | | | | |
| conv3-512 | conv3-512 | conv3-512 | conv3-512 | conv3-512 | conv3-512 |
| conv3-512 | conv3-512 | conv3-512 | conv3-512 | conv3-512 | conv3-512 |
| | | | conv1-512 | conv3-512 | conv3-512 |
| | | | | conv3-512 | conv3-512 |
| | | | | conv3-512 | conv3-512 |
| maxpool | | | | | |
| conv3-512 | conv3-512 | conv3-512 | conv3-512 | conv3-512 | conv3-512 |
| conv3-512 | conv3-512 | conv3-512 | conv3-512 | conv3-512 | conv3-512 |
| | | | conv1-512 | conv3-512 | conv3-512 |
| | | | | conv3-512 | conv3-512 |
| | | | | conv3-512 | conv3-512 |
| maxpool | | | | | |
| FC-4096 | | | | | |
| FC-4096 | | | | | |
| FC-1000 | | | | | |
| soft-max | | | | | |



모델링 (2) – Transformer

• Huggingface

- 자연어 처리/AI 분야의 최신 기술을 사용할 수 있는 오픈소스 라이브러리와 플랫폼을 제공
- 상대적으로 쉽게 사전학습된 SOTA급 모델들을 사용 가능
- ViT(Vision Transformer)류의 모델인 **beit**를 사용



모델링 (2) – Transformer의 문제점

- Inductive Bias가 없는 모델

Translation Invariance

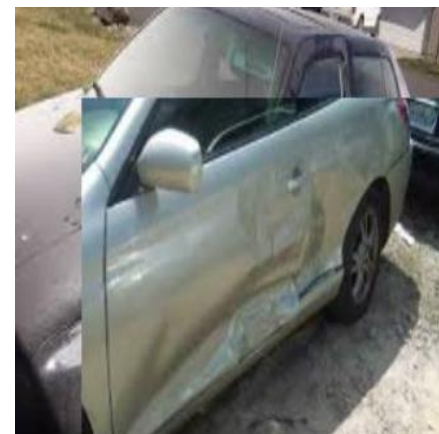


achieved during training. To save compute, we report few-shot linear accuracy instead of full fine-tuning accuracy. Figure 4 contains the results. Vision Transformers overfit more than ResNets with comparable computational cost on smaller datasets. For example, ViT-B/32 is slightly faster than ResNet50; it performs much worse on the 9M subset, but better on 90M+ subsets. The same is true for ResNet152x2 and ViT-L/16. This result reinforces the intuition that the convolutional inductive bias is useful for smaller datasets, but for larger ones, learning the relevant patterns directly from data is sufficient, even beneficial.

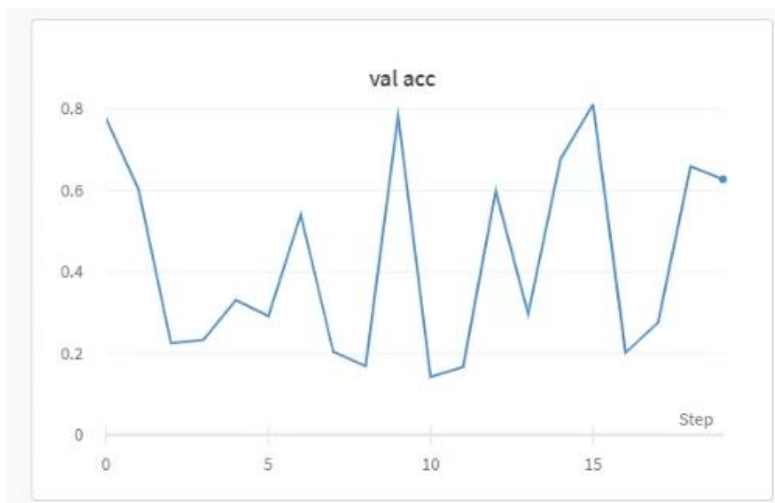
Overall the few-shot results on ImageNet (Figure 4) as well as the low-data results on VTAR

Cut Mix (상세)

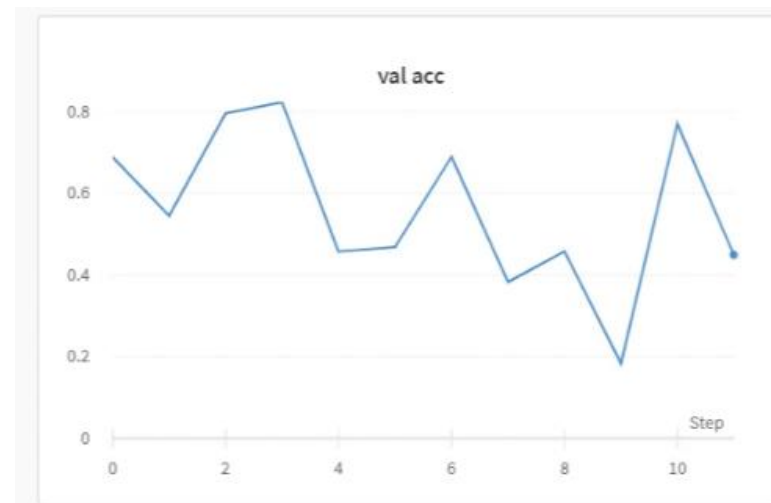
- 일반화로 인한 모델의 성능 상승
- Normal 데이터 간 산출물 또한 Abnormal로 취급



적용 전



적용 후



문제 - 클래스 불균형

Focal loss

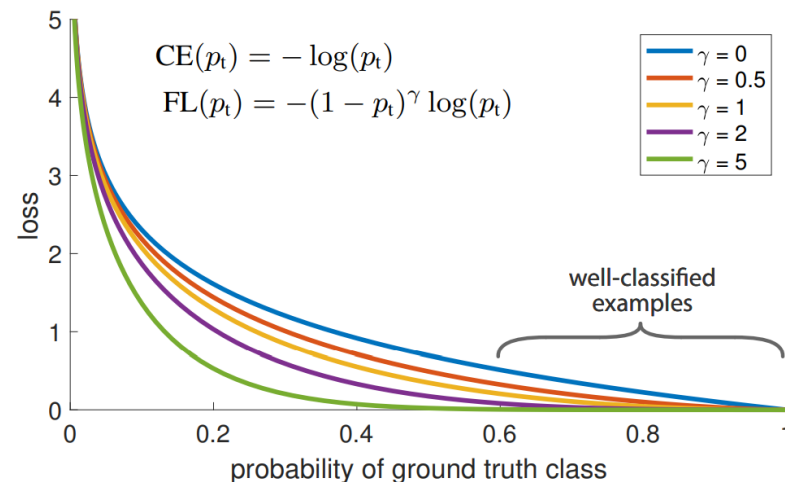
= class 분포가 불균형한 데이터셋의 분류 문제를 해결하기 위해 제안된 손실 함수

= Augmentation 과정에서 class 불균형 문제 발생
(elastic, mixup 등의 과정에서 abnormal class의 비중이 상대적으로 커지므로)

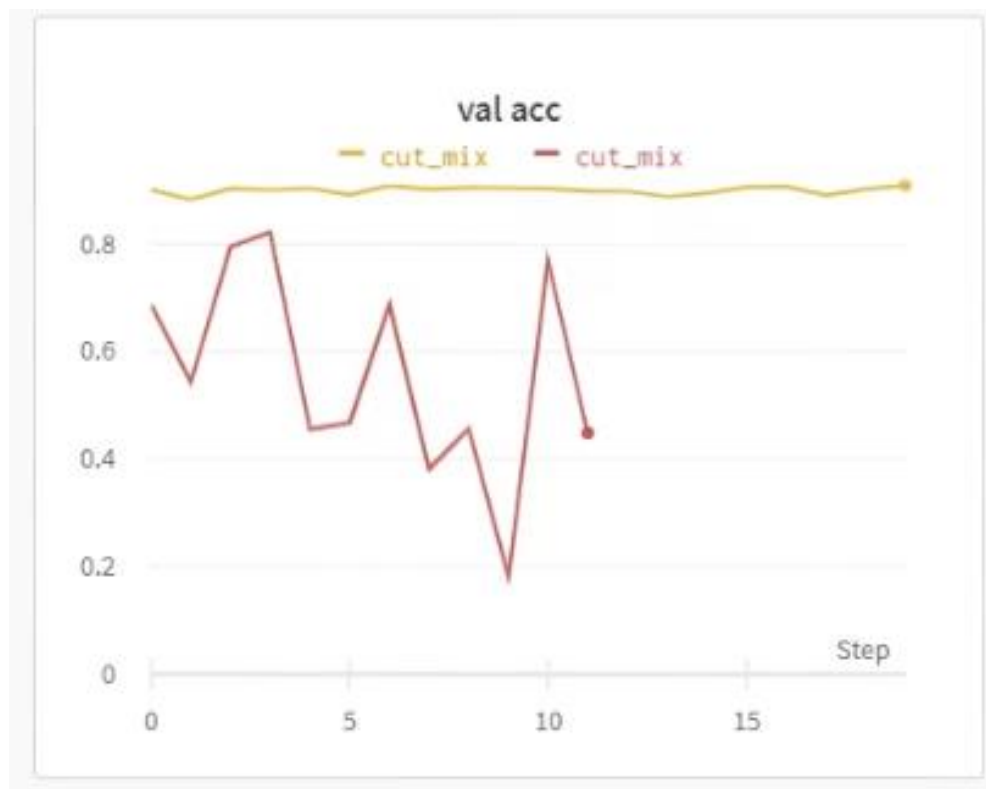
= 본 문제를 해결하기 위해 Focal loss 고안

Cross entropy loss의 경우,

*쉬운 예제와 어려운 예제 모두 동일 가중치를 부여,
 >> 훈련에 부정적 영향*



Focal Loss 사용



■ Focal loss 사용

■ 미사용

