

# Document Type Classification

4조

팀원

강승현 김창희 김하연 김형수 이소영 진성준

[www.fastcampus.co.kr](http://www.fastcampus.co.kr)

Copyright © FAST CAMPUS Corp. All Rights Reserved. 무단전재 및 재배포 금지

# Content

---

01. 팀원 소개

02. 대회 소개

03. Data Description

04. Modeling

05. 결과

06. 경진대회 진행 소감

01

# 팀원 소개

# 팀원 소개

강승현	김형수	김창희	김하연	이소영	진성준
팀장	조원	조원	조원	조원	조원



02

# 대회 소개

# 대회 개요

## Document Type Classification | 문서 타입 분류

Fast campus × upstage

문서 이미지 분류

- 17종 문서 이미지 분류

평가지표 : *Macro F1*





03

# Data Description

## Train Data

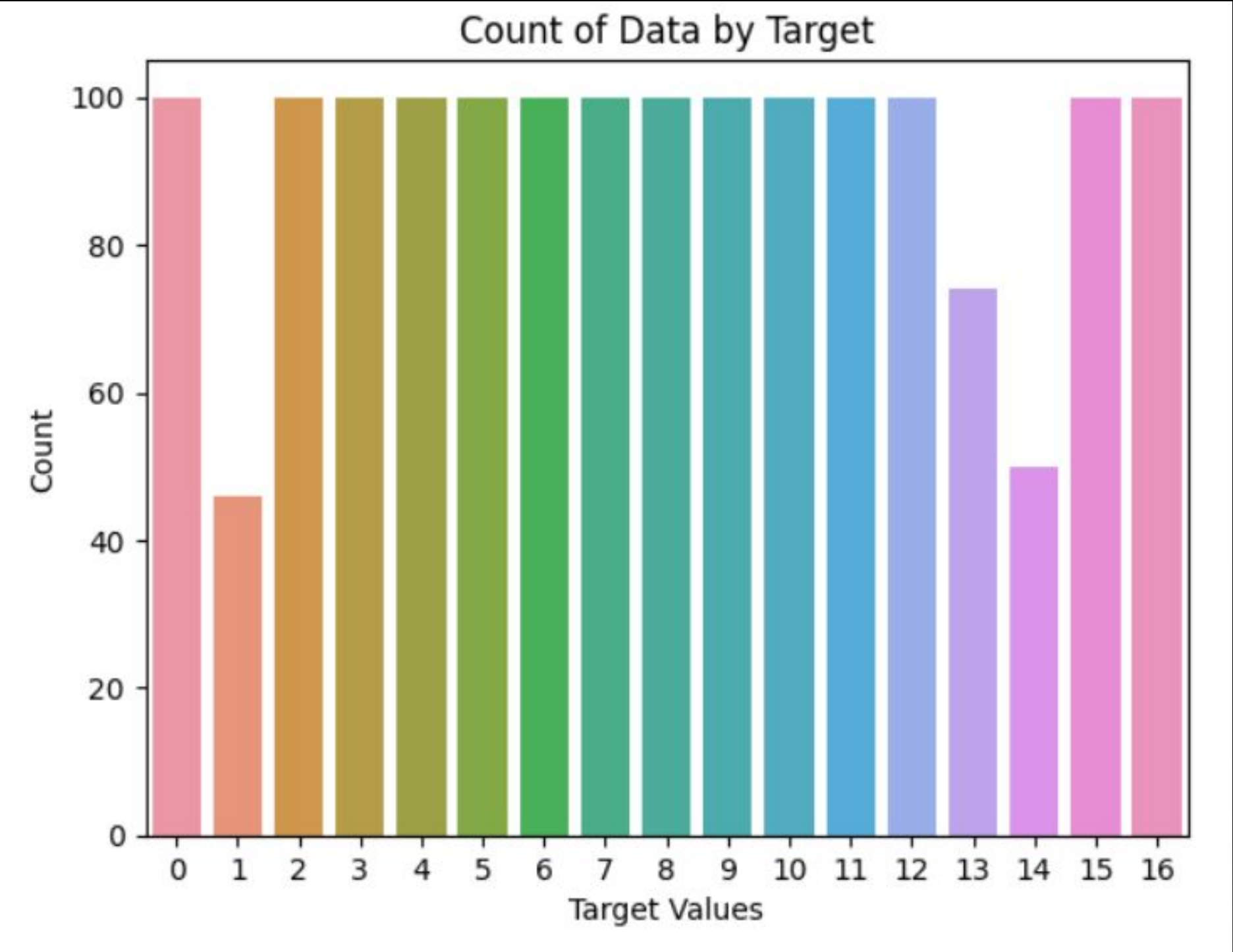
- Input
  - count : 1570개
  - 17 종
- Output
  - 주어진 이미지의 클래스

## Test Data

- count : 3140개



클래스 불균형



## 잘못 라벨링된 데이터

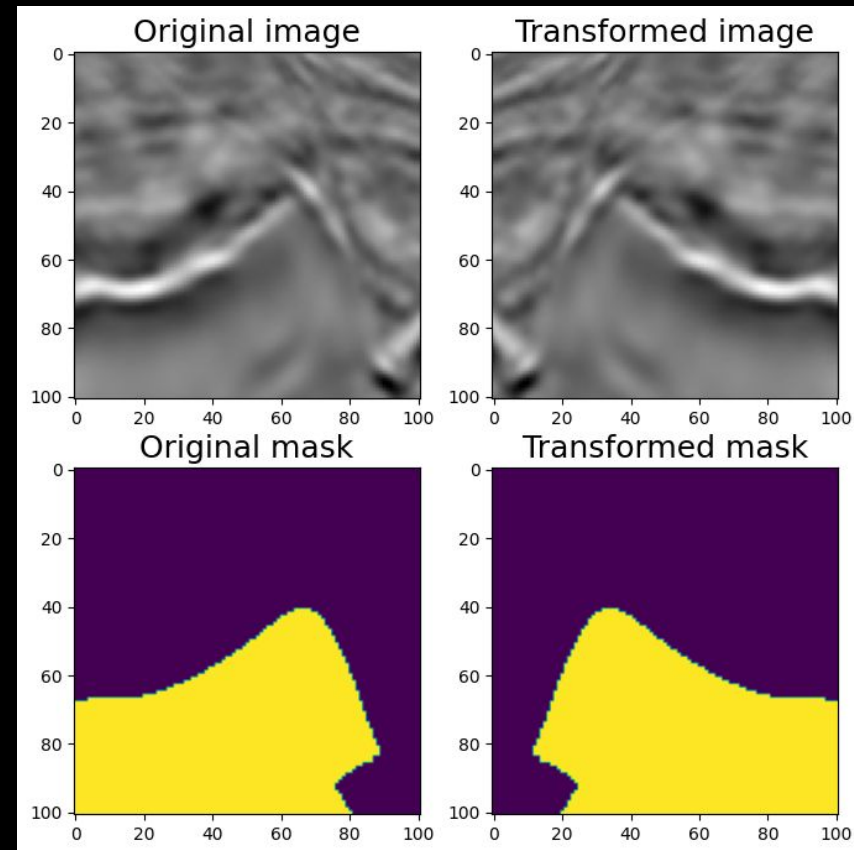
- 8개의 잘못 라벨링 된 데이터 수정

## 테스트셋의 이미지

- 비교적 선명하고 정방향인 훈련 데이터셋과 달리 원본사진에서 다양하게 변형된 사진들을 다수 확인할 수 있음
  - 회전
  - 좌우반전
  - 이동
  - 노이즈
  - 잉크 번짐
  - Mixup

# Data Processing - Albumentation

# A



HorizontalFlip



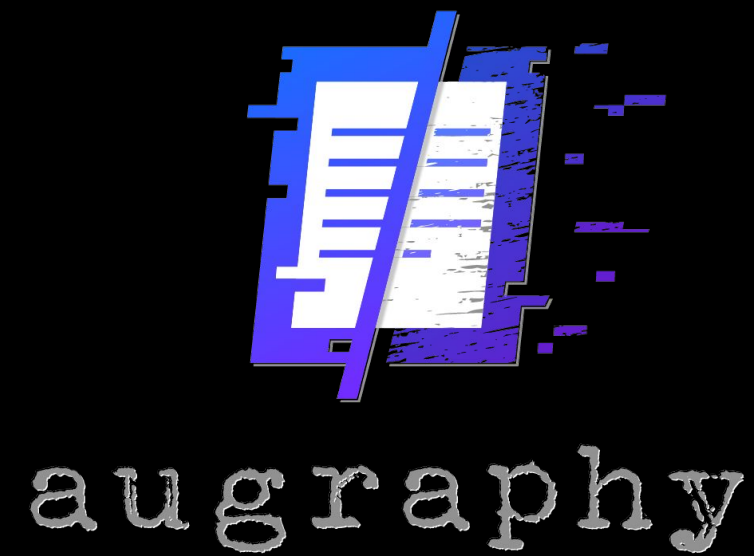
GaussianNoise



ShiftScaleRotation

- Transpose, ToGray, HorizontalFlip, Blur, GaussianNoise, ShiftScaleRotation,ISONoise 등을 사용





PatternGenerator - 패턴 노이즈 생성

Lorem ipsum dolor sit amet, consectetur adipiscing elit

Original

## Lorem Ipsum

dolor sit amet, consectetur  
adipiscing elit, sed do eiusmod  
tempor incididunt ut labore  
et dolore magna aliqua.

BrightnessTexturize

## Lorem Ipsum

dolor sit amet, consectetur  
adipiscing elit, sed do eiusmod  
tempor incididunt ut labore  
et dolore magna aliqua.

InkBleed

## Lorem Ipsum

dolor sit amet, consectetur  
adipiscing elit, sed do eiusmod  
tempor incididunt ut labore  
et dolore magna aliqua.



04

# Modeling

- **Model Selection**

- resnet34, resnet50, wide\_resnet101\_2, efficientnet\_b0, dit-large-finetuned-rvlcdip
- 학습 시간과 리더보드 결과를 고려했을 때 efficientnet\_b0 최종 선택

- **Modeling Process**

- 데이터 분할
  - K-Fold Cross-Validation 후 voting
  - 8 : 2 , 9 : 1 random split
  - 최종 선택 : K-Fold Cross-Validation 후 voting
- 학습률(LR) : 0.001 ~ 0.005
  - 최종 선택 : 0.001
- 손실 함수는 Cross-Entropy Loss 최적화 알고리즘은 Adam 사용



- **Model Selection**

- resnet\_34, deit\_base\_patch16\_224, vit, efficientnet\_b0, efficientnet\_b4 중 efficientnet\_b4가 가장 성능이 좋아 선택

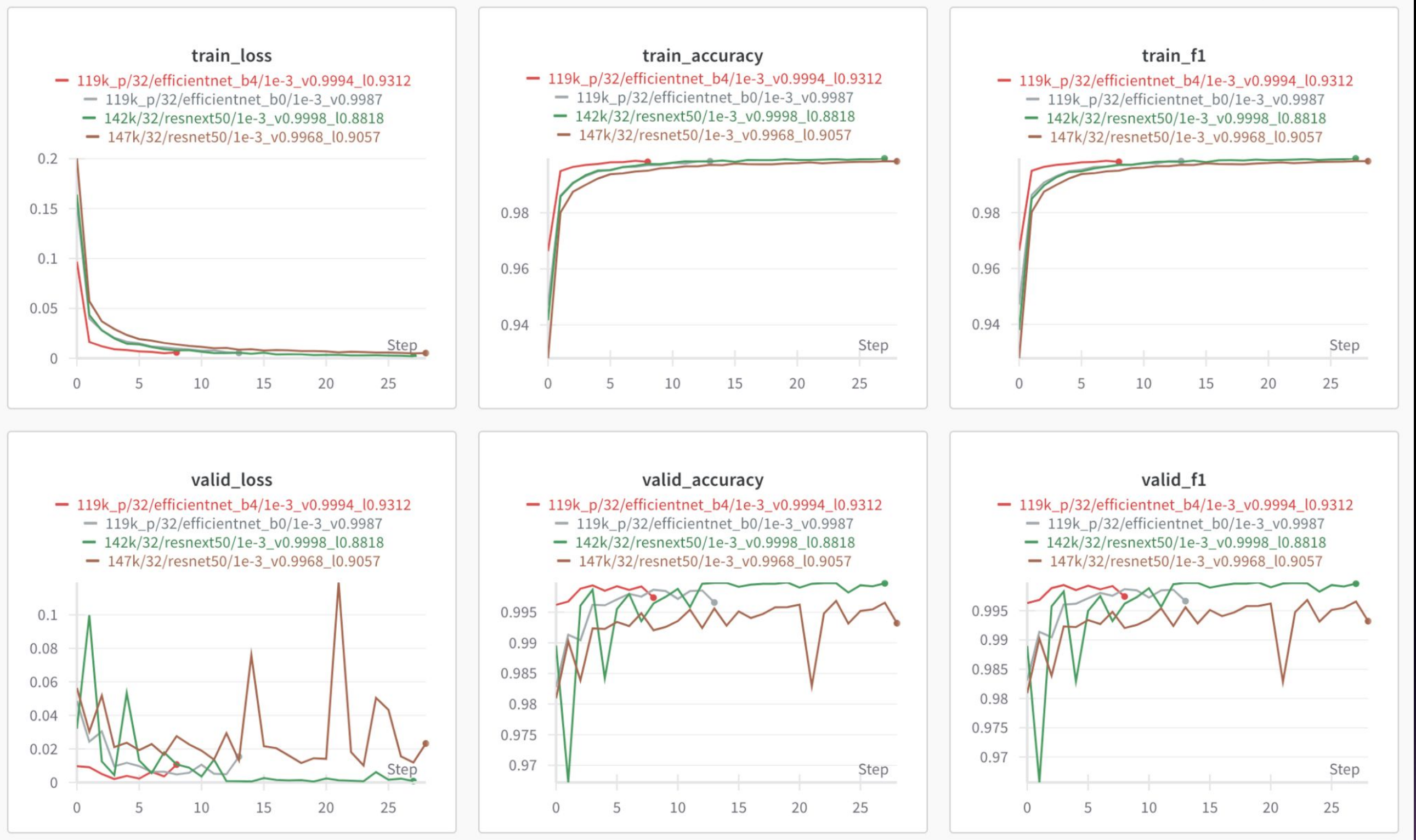
- **Modeling Process**

- 3, 4, 7, 14 번에 예측 성능이 떨어지는 것을 확인
  - 3, 4, 7, 14번에 클래스 가중치 적용
  - 3, 4, 7, 14번 데이터 추가적인 증강
- fine-tuning에서 lr과 batch\_size에는 양의 상관관계가 있다는 것을 확인
  - 최적의 조합인 batch\_size : 16, lr : 1e-4 사용
- 최종적으로 상위 3개의 예측값들을 hard-voting으로 앙상블 -> public f1 score 0.9560

# Model Select (이소영)

resnet50,  
resnext50,  
efficientnet\_b0,  
efficientnet\_b4  
pre-trained 모델로 실험

-> 성능이 가장 좋았던  
efficientnet\_b4으로 고정



# Modeling Process (이소영)

---

**Optimizer** - Adam

**Learning Rate** - 0.001 -> 0.0005

**Early Stopping**

- patience를 5로 설정하여 5 epoch 동안 validation score가 감소하지 않으면 조기 종료

**이미지 전처리**

- Resize 대신 문서의 가로 세로 비율이 유지되도록 Padding 적용 -> 성능 향상



예측 결과 시각화 결과 양식이 유사한 **3, 7, 14** 클래스에 대한 예측 성능이 떨어지는 것 확인

- **3, 7, 14 클래스에 대한 오버샘플링 가중치를 증가시킴**  
-> 성능 향상 (리더보드 F1 스코어 0.9444)
- **3, 7, 14 클래스 별도 학습**
  - efficientnet\_b5 pre-trained 모델로 해당 클래스만 따로 학습하여 기존 결과값 대체  
-> 기대한 만큼은 아니지만 스코어 향상 (리더보드 F1 스코어 0.9480)

## Test-Time Augmentation 적용

- inference 단계에서 평가 이미지에 online 방식으로 augmentation 적용
  - Flip(반전)
  - RandomRotate(90도 단위 랜덤 회전)
- N회 예측 수행
  - 20회 inference 후 soft-voting으로 앙상블 -> 스코어 향상 (리더보드 F1 스코어 0.9562)

## Modeling Process (이소영)

---

리더보드 Public 스코어 최상위 예측값 3개를 hard-voting으로 앙상블

-> Public F1 score: 0.9631



# Modeling Process (김 하연)

## Hybrid EfficientNet Swin-Transformer

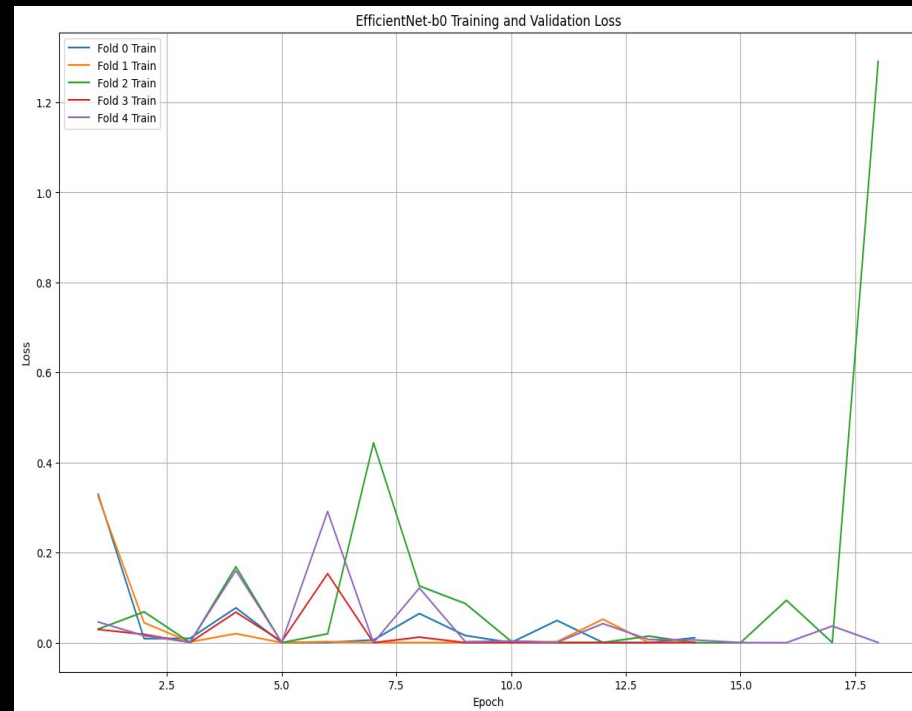
ImageNet이 사전 학습된 EfficientNet 인코더와 swin transformer block을 활용하는 Hybrid Swin Transformer 사용

Train Accuracy: 0.8893, Train F1 Score: 0.8546  
Validation Accuracy: 0.8604, Validation F1 Score: 0.8224

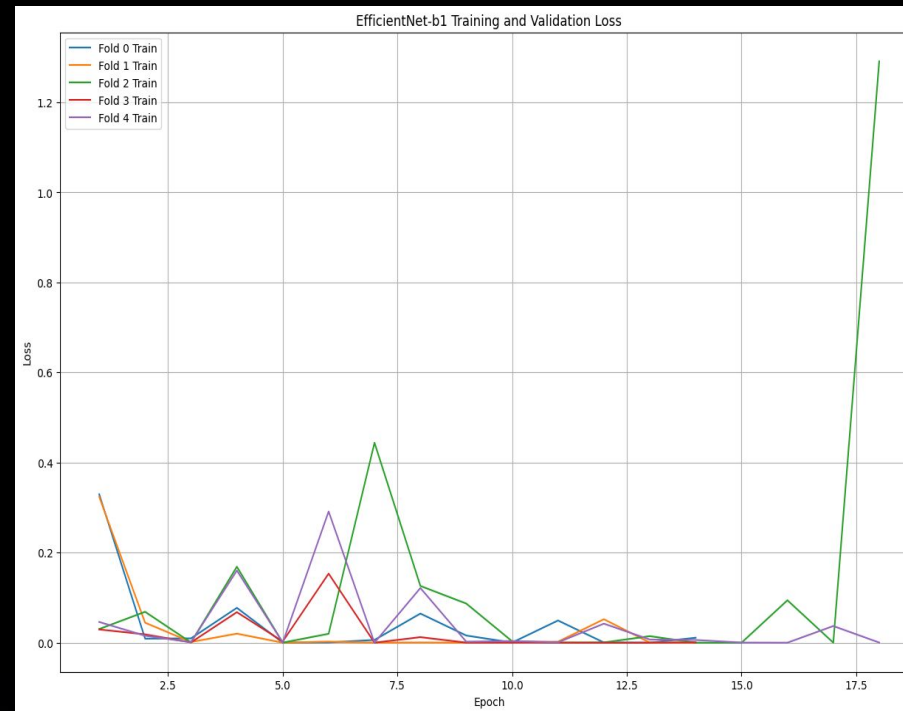
Layer (type)	Output Shape	Param #
efficientnet (Functional)	[(None, 24, 24, 1056), (None, 12, 12, 1280)]	53150388
conv_head (Sequential)	(None, 1280)	2560
patch_extract_5 (PatchExtract)	(None, 144, 4224)	0
patch_embedding_5 (PatchEmbedding)	(None, 144, 64)	279616
patch_merging_5 (PatchMerging)	((None, 36, 128), (None, 12, 12, 64))	32768
swin_blocks (Sequential)	(None, 144, 64)	33544
swin_head (Sequential)	(None, 128)	256
dense_41 (Dense)	(None, 17)	23953
=====		
Total params: 53,523,085		
Trainable params: 53,231,053		
Non-trainable params: 292,032		
=====		

# Modeling Process (김하연)

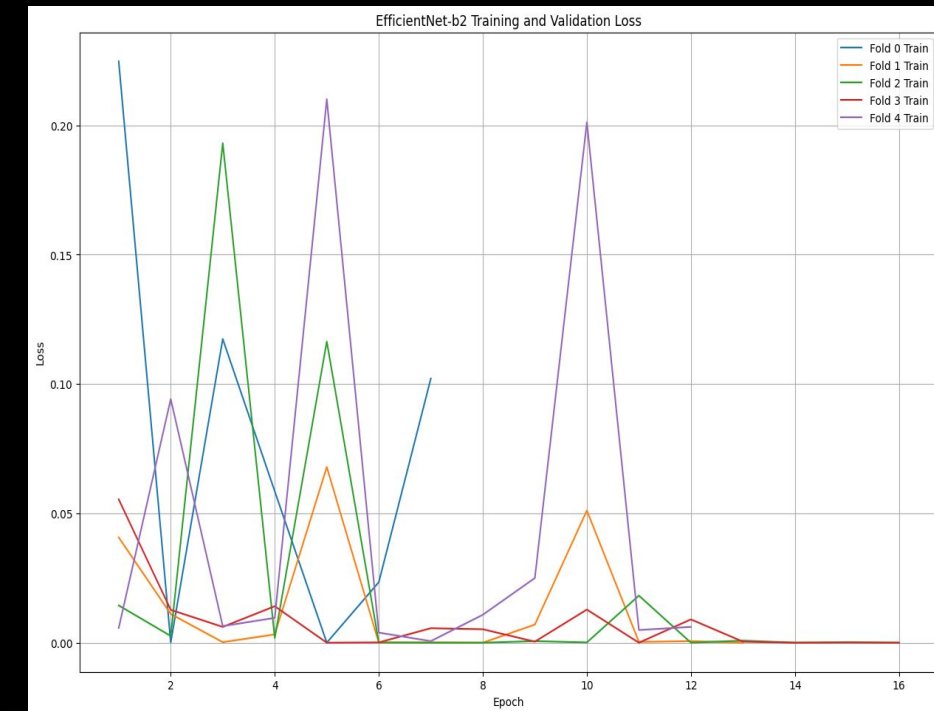
## EfficientNet Ensemble



EfficientNet-B0



EfficientNet-B1



EfficientNet-B2

EfficientNet B0, B1, B2는 모델 크기와 input image resolution에서 차이가 있으며 B0가 가장 작은 크기를 가지고 B2가 상대적으로 큰 크기를 가짐

각 모델 K-Fold 결과값 hard voting ensemble하여 최종 제출 결과 f1 score 0.9384

### 김용담 강사님

**Q:** Public score 0.94 에서 점수가 더 이상 오르지 않고, 특히 몇몇 클래스들에서 성능 떨어졌습니다. 혹시 추가로 시도해볼 수 있는 방법이 있을까요?

**A:** 점수를 높이기 위해서는 고퀄리티의 데이터가 필요하며, resnet 혹은 elasticnet 보다 더 큰 모델 사용(ViT LAION 2B)하면 성능 향상을 기대할 수 있다. 추가로 Contrastive Learning을 시도해보거나 loss function metric을 변경해 볼 수 있다.

### 민진홍 멘토님

**Q:** 현재까지 주어진 학습 데이터를 Albumentation의 여러 기법들을 활용하여 augmentation을 시도해보았습니다. 혹시 추가로 시도해 볼 수 있는 augmentation 기법들이 있을까요?

**A:** Augraphy의 augmentation 기법들을 사용해보는 것을 추천. 또 대회 마지막에 여러 모델들의 결과물을 앙상블 하는 것을 추천.

**Q:** 양식이 비슷한 클래스들 간의 예측 성능을 개선할 수 있는 방법이 있을까요?

**A:** 해당 클래스들만 따로 학습하는 방법이 도움이 될 것.



05

결과

# 최종 순위 및 평가지표 결과

Public F1 score: 0.9631(4/9)

Private F1 score: 0.9547 (4/9)

72	Finished	S	0.9631 → 0.9547	상세 보기	2024-02-19 18:38			
68	Finished	S	0.9603 → 0.9504	상세 보기	2024-02-19 16:18			

# 최종 순위 및 평가지표 결과

4	CV 4조	<div><div>성준</div><div>H</div><div></div><div></div><div>승현</div><div>S</div></div>	0.9631	69	9h
---	-------	---	--------	----	----



06

# 경진 대회 진행 소감

# 그룹 스터디 진행 소감

강승현	처음 경험해본 CV 경진대회여서 미숙한 점이 많았지만, 팀원들과 함께 진행하면서 다양한 아이디어와 접근 방식을 공유하고 해결해 나가는 과정에서 많은 것을 배울 수 있었습니다.
김형수	첫 CV 대회였던만큼 어려움도 많았지만, 다양한 augmentation 방법들을 공부해보고 시도해 볼 수 있었던 좋은 기회가 되었습니다.
김창희	강의에서 배운 내용을 복습하고, 알지 못했던 것들을 알게 되어 좋은 기회가 되었던 것 같습니다.

# 그룹 스터디 진행 소감

김하연	CV 모델링을 처음 해봤는데 다양한 기법들을 배울 수 있어서 좋았습니다.
이소영	CV 딥러닝 프로세스를 경험해보고 문제 해결을 위해 다양한 방법론을 시도해보면서 많은 것을 배울 수 있었습니다.
진성준	CV와 프로젝트 수행에 많은 부족함을 느꼈고, 타 조원들의 프로젝트 수행하는 것을 볼 기회가 되어 좋았습니다.



# Q&A

---

감사합니다.

—