

Project 4 Wrap-up Report

Level2 Semantic Segmentation

CV03 뻔뻔(FunFun)

1. 프로젝트 개요

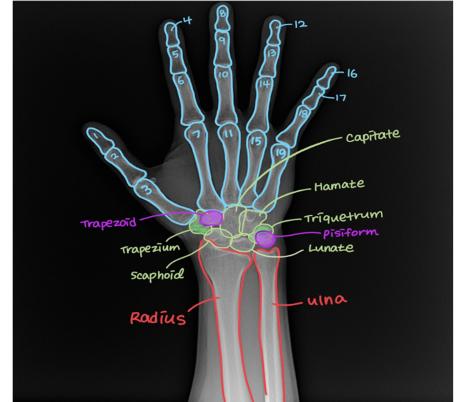
1.1 프로젝트 소개

Hand Bone Image Segmentation

뼈는 우리 몸의 구조와 기능에 중요한 영향을 미친다. Bone Segmentation을 수행해 정확한 뼈 분할 기능을 발전시키고자 한다.

1.2 프로젝트 상세

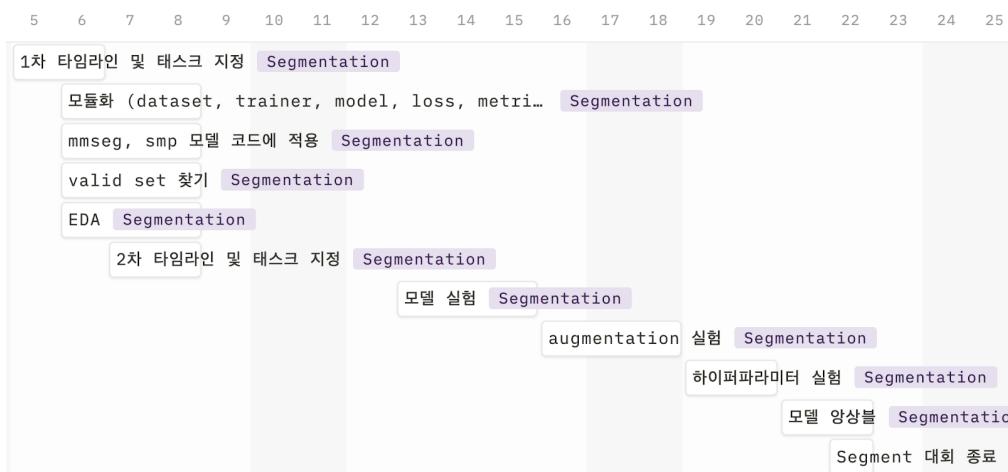
- ▶ Dataset: Train(800장), Test(288장), 크기(2048, 2048, 3)
- ▶ Class: 29개의 뼈
- ▶ Annotation: polygon point 형식으로 제공
- ▶ Evaluation Metric: Dice Coefficient
- ▶ Framework: PyTorch, smp, mmseg
- ▶ Cooperation Tool: Github, Notion, Wandb, Slack, Zoom, Gather



2. 프로젝트 팀 구성 및 역할

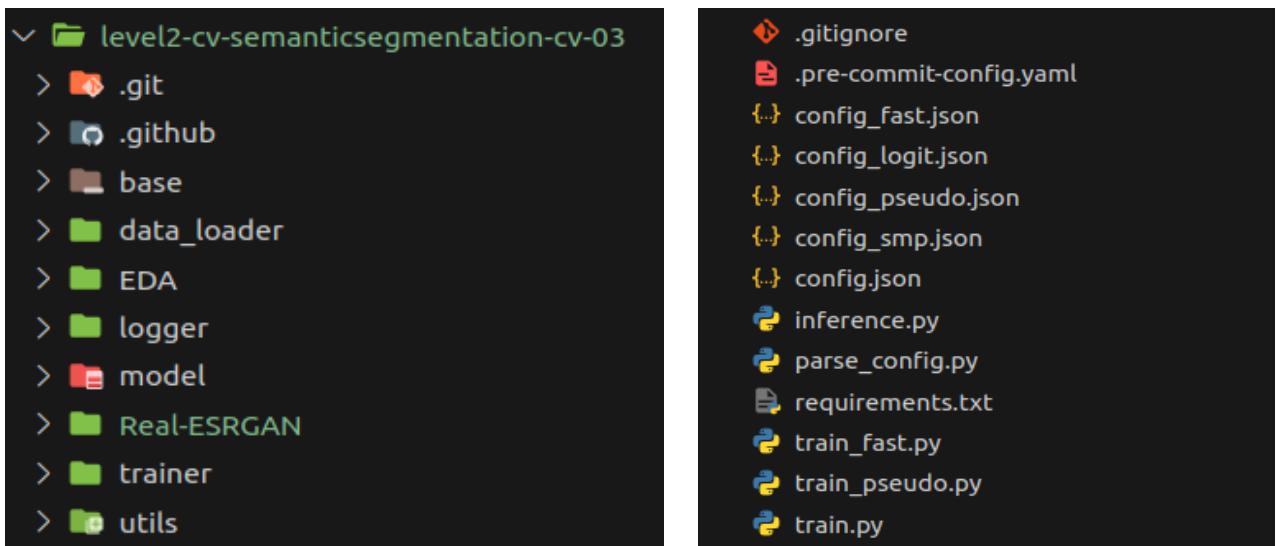
강현서	EDA, Augmentation 실험, Sweep 실험
김정택	프로젝트 모듈화, SuperResolution 실험, Soft Voting Multi processing으로 코드 재구현
박진영	EDA, Augmentation 실험, GradCAM 구현, Hard Voting 구현
선경은	프로젝트 모듈화, Unet/Unet++/Optimizer 실험, Pseudo Labeling/Soft Voting 구현, Dataset 압축
이선우	프로젝트 모듈화, Dataset 압축, Loss 실험, Augmentation 실험, DeeplabV3Plus 실험
최현우	Resize 실험, smp 모델 실험, mmseg 모델 구현, Class-Specific Train, Validation 실험

3. 프로젝트 일정



4. 프로젝트 수행 과정

4.1 Project 모듈화



4.2 EDA

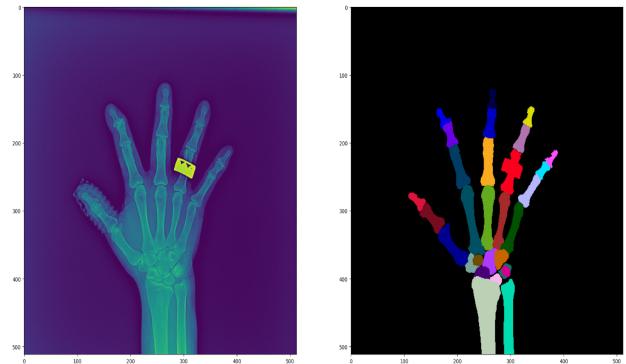
Noise

▶ 실험 가설

반지 착용 케이스 1장을 제거하여 데이터에 노이즈를 줄이면 성능이 향상될 것이다.

▶ 실험 결과

반지를 착용한 뼈에 대한 성능은 증가하였으나, 전체 mean dice 값은 하락하였다. 데이터 제거로 인해 다른 뼈에 대한 정보를 누락하는 셈이 되었기 때문에 전체 성능이 약간 하락한 것으로 보였다. 따라서 해당 데이터를 제거하지 않고 사용하였다.



Method	Model	Val Finger-14 Dice (반지 낀 손가락)	Val Dice
origin	U-Net - resnet34	0.9825	0.9638
No Ring	U-Net - resnet34	0.9838	0.9629

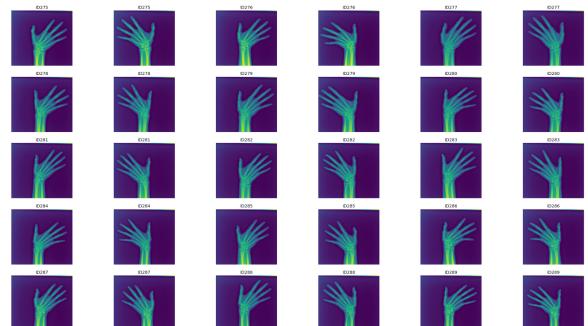
꺾인 손목

▶ 데이터 특징

Train 46 쌍 / Test 83 쌍의 꺾인 손목 이미지가 존재했다.

▶ 조치

Group K Stratified Fold를 사용하여 동일한 비율의 꺾인 손목 사진을 Train과 Valid Data Set에 사용하였다.



겹치는 뼈

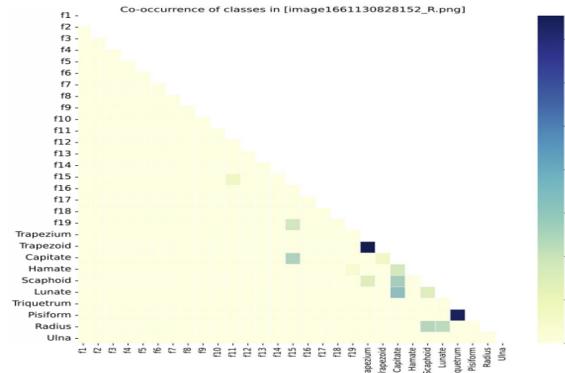
▶ 데이터 특징

Trapezium & Trapezoid 손등 뼈가 자주 겹침

Triquetrum & Pisiform 손등 뼈가 자주 겹침

▶ Class별 Dice Coefficient 실험 결과

겹치는 뼈 가운데 Pisiform과 Trapezoid가 점수가 전체 클래스 대비 가장 낮았으며, 높은 클래스 점수 대비 약 17%, 10% 낮은 것으로 나타났다. 따라서 겹친 뼈 가운데 Pisiform과 Trapezoid가 본 프로젝트에서 해결해야 할 가장 큰 문제일 것이라 예측하였다.



Model	Pisiform	Trapezoid	Capitate	Hamate	Lunate	Radius	Scaphoid	Trapezium	Triquetrum	Ulna	Finger (mean)	Val Dice
U-Net - ResNet34	0.8225	0.9081	0.9598	0.9567	0.9643	0.9918	0.9741	0.9513	0.9557	0.9899	0.9758	0.9637
U-Net - EfficientNet-B7	0.9088	0.9227	0.9691	0.9464	0.9542	0.9901	0.966	0.9436	0.928	0.9879	0.9816	0.9736

(fold 1 기준)

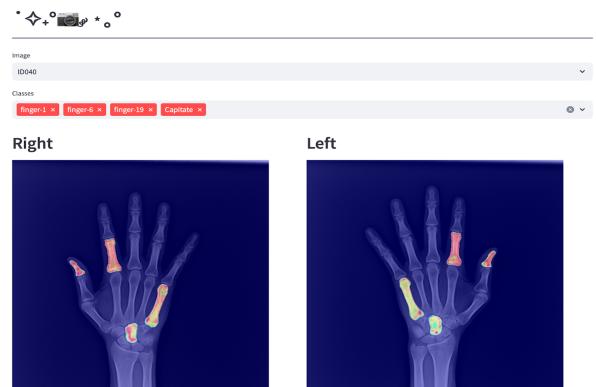
GradCAM

▶ 사용 이유

모델이 test dataset에 대해서 어떻게 바라보고 있는지 확인하기 위해 사용하였다. 각 클래스에 대해 어떤 부분에 주목하고 있는지 확인하고, 그 부분이 잘못되지는 않았는지 혹은 너무 낮은 확률로 확신을 갖고 있진 않은지 확인할 수 있었다.

▶ 사용 방법

Streamlit을 이용하여 이미지 한장을 불러온 뒤, 클래스를 선택해 모델이 얼만큼 확신을 가지는지 GradCAM 결과를 시각화하였다.



4.3 Dataset Preparing

Train Dataset

▶ 문제 상황

DataLoader에서 병목이 생기는 현상을 확인하였다. 특히 poly-seg 형식의 label을 mask로 변환하는 CPU 작업량이 많아 학습 시간이 오래 소요되었다. 미리 mask 형식으로 변환 후 저장하여 사용하려 했지만 전체 용량이 90GB 이상 차지하였다.

▶ 해결 방법1

Memmap 라이브러리를 이용해 이미지를 미리 더 빨리 로딩할 수 있는 형식으로 저장하였고, 그 결과 10에폭 기준 약 3분의 학습 시간 감소가 가능했다.

▶ 해결 방법2

이진 배열인 mask를 byte에서 bit 단위로 저장하였고, 그 결과 용량은 약 8분의 1로(90GB→11GB), 학습 시간은 약 30% 감소하였다.

Test Dataset

▶ 문제 상황

각 픽셀의 확률값으로 양상들을 진행하기 위해서는 float type의 mask를 사용해야 하지만 한 모델당 전체 mask 용량이 100GB 이상을 차지하였다.

▶ 해결 방법

float64 type에 10000을 곱하여 uint16 type으로 변경한 후, numpy의 savez_compressed로 압축하여 저장하였고, 그 결과 용량은 약 1000분의 1로(100GB→100MB) 감소하였다.

4.4 Modeling

FCN

▶ 기존 CNN의 완전 연결층을 합성곱으로 대체한 구조이므로 픽셀 단위 예측을 위한 기본 모델로 선정하였다.

U-Net

▶ 이미지의 세밀한 구조를 포착하면서도 컨텍스트 정보를 유지할 수 있는 대칭적인 구조를 가지고 있어 의료 영상에 적합하다고 판단하였다.

U-Net++

▶ U-Net의 확장 버전으로 기본 구조에 중첩된 U-Net 구조를 추가하여 더 세밀한 특징 추출을 가능하게 한다. 특히 경계가 모호하거나 복잡한 구조를 가지는 의료 영상에 적합하다고 판단하였다.

DeepLabV3Plus

▶ U-Net 계열의 모델과 함께 Segmentation에서 가장 평이 좋은 모델이다.

▶ 양상들을 통해 추가적인 성능 향상을 도모하고 싶었다.

Model	Backbone	Val Dice	LB Dice
FCN	resnet50	0.9646	0.9611
U-Net	efficientnet-b5	0.9678	0.9663
U-Net++	efficientnet-b7	0.9741	0.9713
DeepLabV3Plus	efficientnet-b7	0.9725	0.9703

(fold 1 기준)

4.5 Augmentation

Resize

▶ 실험 가설

원본 이미지의 resolution이 (2048, 2048)이므로 낮은 resolution으로 resize가 될 때 정보 손실이 있을 것이다.

▶ 실험 내용

실험에 진행한 모델은 smp.FPN 모델이며 세부 파라미터는 resnet34, imagenet을 사용하였다.

실험한 resolution은 512, 1024, 2048이며 이미지 사이즈에 따른 batch size를 조절하였다.

▶ 실험 결과

이미지 사이즈가 클 수록 좋은 성능을 보였다.

Resolution	Train Batch	Valid Batch	Val Dice	LB Dice
(512 x 512)	8	8	0.9452	0.9425
(1024 x 1024)	8	4	0.9638	0.9616
(2048 x 2048)	2	1	0.9715	0.9688

Super Resolution

▶ 실험 가설

Resize 실험 결과 이미지 사이즈가 클 수록 좋은 성능을 보인 것으로 확인하였다. 그러나 단순히 뼈를 확대하는 것은 화질 자체를 개선시키지 못하기 때문에 Super Resolution을 적용하여 화질을 개선시킨 뒤, 실험해보면 성능이 더 좋아질 것으로 기대하였다.

▶ 실험 내용

Real-ESRGAN을 사용하여 화질을 2배 개선한 뒤, 2048x2048 크기로 resize후 저장하여 오프라인 증강 실험을 진행하였다.

▶ 실험 결과

Super Resolution 사용하여 육안으로 확인한 결과, 작은 뼈에 대해 화질 차이가 매우 두드러졌다. 그러나 Base 모델과 비교 실험한 결과 Dice 값의 차이는 거의 존재하지 않았다.



origin v.s. 2x

Method	Model	Val Dice
Base	U-Net - EfficientNet-B7	0.9738
Super Resolution (x2)	U-Net - EfficientNet-B7	0.9739

Flip, Rotate

▶ 실험 가설

Horizontal Flip, Rotate를 적용하면 성능 향상이 있을 것이다.

▶ 실험 내용

Rotate의 경우 의료 도메인에서는 보통 -5 ~ 5도 사이의 증강이 일반적이라고 하여 limit을 5로 설정하였다. 두 기법에 대한 확률 파라미터를 조정하며 성능을 비교하였다.

▶ 실험 결과

증강을 적용하지 않은 baseline 실험을 기준으로 valid set에 대한 Mean dice, Pisiform dice, Trapezoid dice 값을 비교하였다. 실험 결과 baseline이 전체적으로 가장 우수한 성능을 보였다. Pisiform과 Trapezoid dice 값을 성능 지표로 활용한 이유는 두 뼈가 겹쳐있어 모델이 가장 낮은 성능을 나타내는 영역이기 때문이다.

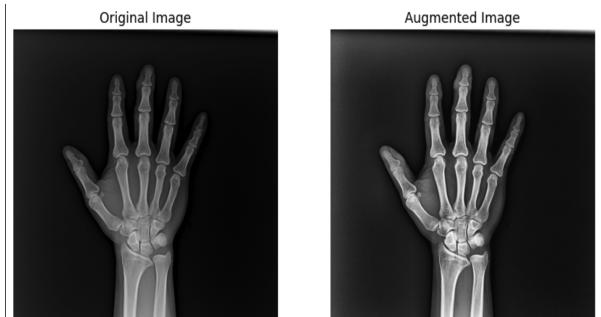
Hflip	Rotate	Epoch	Val Dice	Pisiform Dice	Trapezoid Dice
-	-	50	0.9674	0.8942	0.9116

0.1	0.1	50	0.9561	0.8791	0.8704
0.3	0.3	50	0.9586	0.8862	0.9119
0.5	-	100	0.9558	0.8897	0.8762
-	0.3	50	0.9631	0.8	0.9076

CLAHE

▶ 실험 가설

이미지 대비를 향상시키는 CLAHE 기법을 사용해 흰 뼈와 검은 배경의 차이를 극대화하여 뼈 특징을 보다 뚜렷하게 감지하고 분류할 수 있도록 돋는다.



▶ 실험 내용

efficient net을 encoding으로 사용한 smp.Unet 모델에 CLAHE를 0.1과 0.5의 확률로 설정하여 성능을 비교하였다.

▶ 실험 결과

CLAHE	Epoch	Val Dice	Pisiform Dice	Trapezoid Dice
-	63	0.9677	0.8903	0.914
0.1	63	0.9655	0.8902	0.8825
0.5	63	0.9632	0.8975	0.8886

4.6 Optimizer

Adam

▶ 기본 optimizer로 사용하였다.

AdamW

▶ weight decay를 최적화 과정과 분리하여 가중치 감소가 모멘트 추정치에 영향을 주지 않도록 하여 일반화 성능을 더 향상시키는 데에 도움이 될 것이라 생각했다.

Optimizer	LR	Weight decay	Val Dice	LB Dice
Adam	1e-3	1e-5	0.9678	0.9663
AdamW	1e-3	1e-5	0.9721	0.9681

(U-Net_efficientnet-b5)

4.7 Loss

▶ 실험 가설

Loss 값에 따라 Valid Dice 값이 큰 차이가 나는 것을 확인해 최적의 loss값을 찾고자 하였다.

▶ 실험 내용

epoch은 100을 기준으로 잡고 돌렸으나, 대부분 early-stopping으로 50 epoch 내외로 끝났다.

▶ 실험 결과

전체적으로 단일 Loss보다 여러 Loss function을 섞은 Combo Loss가 효과가 좋았다. 최종적으로 Jaccard+Dice Loss를 사용하기로 결정했다.

Loss	Val Dice
BCE Logit Loss	0.9592
Dice Loss	0.6646
Jaccard Loss	0.9674
BCE Logit + Dice	0.9650
Jaccard + BCE Logit + Dice	0.9688
Jaccard + Dice	0.9691

4.8 Ensemble

Hard Voting

▶ 실험 가설

여러 모델의 결과를 비교하여 한 픽셀 당 가장 많이 예측된 클래스를 사용하면 한 모델을 사용하는 것보다 더 높은 성능을 기대할 수 있을 것이다.

▶ 실험 내용

첫 번째로 진행한 폴드 양상블은 한 모델만 예측한 클래스는 제거하고, 두 개 이상의 모델이 예측한 클래스만 가져가는 방식으로 양상블을 하였고. 두 번째론, 예측 성능이 낮은 2 class만 따로 양상블하여 Unet++(efficient-b7) 모델에 합치는 방식으로 진행했다.

▶ 실험 내용

Method	Model	Threshold	LB Dice
단일모델	Unet++(efficientnet-b7)	-	0.9713
폴드 양상블	Unet++(efficientnet-b7)	0.2	0.9721
2 class 양상블	FCN(resnet34), DeepLabV3Plus(efficient-b5), Unet++(efficient-b7)	0.6	0.9725

Soft Voting

▶ 실험 가설

각 픽셀의 확률값으로 양상블하는 것이 각 모델의 특징을 더 잘 반영하고 모호한 픽셀에 대해 효과적일 것이다.

▶ 실험 내용

dice 점수 차이 또는 모델 특징을 고려하여 가중치를 다르게 주고 양상블을 진행하였다.

▶ 실험 결과

단일 폴드 최고 dice 점수보다 5개 폴드 양상블시 점수가 상승하였고, 다른 모델과 양상블시 더 좋은 효과를 보였다.

Method	Model	Weights	LB Dice
단일모델	Unet++(efficientnet-b7)	[1]	0.9713

풀드 양상블	Unet++(efficientnet-b7)	[2, 1, 1, 1, 1]	0.9724
모델 양상블	DeepLabV3Plus, Unet++(efficientnet-b7)	[1.5, 2, 1, 1, 1, 1]	0.9726

(threshold 0.5)

5. 프로젝트 결과

public 5th	5	CV_03조		0.9728
private 5th	5	CV_03조		0.9739

6. 팀 회고

잘했던 점

- ▶ 모듈화된 학습 환경을 공유하여 여러가지 모델을 통한 학습을 진행할 수 있었다.
- ▶ 데이터셋을 모두 관찰하여 제외할 데이터를 골라낼 수 있었다.
- ▶ Parameter 실험을 통해 모델 일반화 성능을 향상시킬 수 있었다.

아쉬웠던 점

- ▶ TTA를 적용해보지 못했습니다.
- ▶ Validation set의 correlation을 제대로 확인해보지 못했다.
- ▶ 다양한 모델 실험을 진행하지 못했습니다. 특히 YOLOv8을 사용한 Segmentation을 못 해본게 아쉽다.
- ▶ 높은 이미지 해상도로 인해 학습을 진행하는데 소요되는 시간이 길었다.
- ▶ 특정 클래스에 대한 성능저하를 빠르게 파악했지만 효과적인 대처를 하지 못했다.

프로젝트를 통해 배운 점

- ▶ 데이터셋과 데이터 로더의 구조에 따라 학습 시간에 큰 영향을 미칠 수 있다는 것을 배울 수 있었다.
- ▶ 실험의 가설이 결과와 맞지 않을 때는 더 다양한 방법으로 분석을 시도하고 실험해봐야 하는 것을 배웠다.

7. 개인 회고

강현서

▶ 나는 내 학습목표 달성을 위해 무엇을 어떻게 했는가?

Nohup를 이용해 비는 시간 없이 최대한 많은 실험을 할 수 있도록 했고, 팀원들과 Wandb를 사용해 train 상황을 공유하여 효율적인 실험이 가능했다.

▶ 나는 어떤 방식으로 모델을 개선했는가?

효과적인 Augmentation 전략을 수립하기 위해 우리의 Hand Bone Segmentation 대회와 그 성격이 유사한 데이터를 사용한 대회를 탐색해보았다. 결과적으로 COVID-19 진단 관련 폐 X-ray 이미지를 활용한 Kaggle 대회를 참고하였다. 우리 대회와 유사하다고 판단한 부분은 다음과 같다.

1. 상하 방향이 고정되어 있다: 흉부 X-ray 사진은 사람이 서 있는 형태로 촬영되므로 상하 방향이 고정된다. 우리의 데이터셋도 상하 방향이 고정되어 있다.

2. 폐가 뼈 등과 겹치는 부분이 있다: 우리 대회에서 가장 중요한 성능 요인은 뼈가 많이 겹쳐 있는 손등 뼈 부분에서 Segmentation을 잘 해내는 것이었는데, 이 데이터셋에서도 겹침 부분이 있었다.

이 대회를 통해 얻은 인사이트를 공유해놓은 글들을 조사해본 결과, Reflection 부분에서 y-axis reflection, Rotation (의료 도메인에서는 -5~5도 사용이 일반적), Translation, Mixup(겹치는 부분에 대한 탐지 성능에 도움) 등의 Augmentation이 효과적이라는 것을 알게 되었다. 이 과정을 통해 도메인 지식의 중요성을 느꼈다.

▶ 전과 비교하여 새롭게 시도한 변화는 무엇이고, 어떤 효과가 있었는가?

이전에는 주로 직관적인 가설을 세우고 이를 실험적으로 검증하는 방식을 주로 사용했다. 그러나 이번에는 다른 접근법을 시도해보았다. 비슷한 도메인의 Kaggle 대회를 조사한 후, 그 대회에서 유효성이 입증된 Augmentation 방법을 집중적으로 적용해보았다. 이 과정에서 팀원들에게 시도해볼 가치가 있는 Augmentation 기법들을 빠르게 제시할 수 있어 업무 진행의 효율성을 높일 수 있었다.

▶ 마주한 한계는 무엇이며, 아쉬웠던 점은 무엇인가?

조사를 통해 진행한 Augmentation이 기대와는 달리 baseline보다 낮아진 성능을 보였다. 마스터 클래스에서 타 조의 발표를 참고하면, 그분들도 비슷한 Augmentation을 통해 성능 향상을 확인했다고 하시는데, 어떤 부분에서 이런 차이가 날 수 있는지 궁금하며 아쉬움이 남았다. Validation set을 잘못 찾은 것인지, TTA를 같이 적용하지 않아서 그런 것인지 궁금하다. Public score과 Correlation이 좋은 Valid set을 찾는 것의 중요성을 느꼈다. 대회 마지막 부분에 Wandb에서 성능이 오른 것을 확인하고 팀원들과 합의 하에 제출한 파일이 리더보드 상에서 더 낮은 성능을 보일 때, 제출 횟수가 얼마 남지 않아 아쉬움이 있었다.

▶ 한계/교훈을 바탕으로 다음 프로젝트에서 시도해볼 것은 무엇인가?

Public score과 Correlation이 좋은 Valid set을 찾을 것이다. 이 부분이 간단함에도 그 후 하는 모든 실험에 영향을 끼쳐 실험 결과의 정당성에도 중요하기 때문이다. 따라서 리더보드 점수를 함께 확인하며 많은 Valid set을 시도해보는 과정을 거칠 것이다.

김정택

▶ 나는 내 학습목표 달성을 위해 무엇을 어떻게 했는가?

프로젝트를 직접 모듈화 시키면서, Pytorch Project가 어떻게 구성되는지 이해하고 직접 코드를 구현하는게 이번 대회 학습 목표였습니다.

▶ 나는 어떤 방식으로 모델을 개선했는가?

EDA를 바탕으로 최대한 효과적인 Augmentation 실험을 진행하고자 했습니다.

▶ 전과 비교하여 새롭게 시도한 변화는 무엇이고, 어떤 효과가 있었는가?

Super Resolution에 대한 코드를 직접 가져와 작성해볼 수 있었고, 다른 깃헙 프로젝트의 코드를 사용하는데 조금 더 익숙해진 것 같습니다.

▶ 마주한 한계는 무엇이며, 아쉬웠던 점은 무엇인가?

이전 Object Detection 대회 때와 달리 이번 대회는 체계적으로 대회를 진행하지 못해 아쉬웠습니다.

만약 보다 많은 모델 실험을 했더라면, 앙상블 때 성능 개선을 더 많이 기대해볼 수 있었을텐데 그러지 못했던게 아쉬웠습니다.

▶ 한계/교훈을 바탕으로 다음 프로젝트에서 시도해볼 것은 무엇인가?

다음에 따로 대회에 참여하게 된다면, 그때는 대회를 위해 필요한 항목들에 대해 사전에 정리해놓고

이에 대해 타임라인을 지정하고, 보다 체계적으로 대회를 준비해볼 수 있으면 좋겠습니다.

박진영

▶ 나는 내 학습목표 달성을 위해 무엇을 어떻게 했는가?

이번 대회의 학습목표는 ‘코드를 많이 보고 구현해보자’였다. 그동안 실험을 맡아 진행하면서 기존 코드를 수정하거나 새롭게 구현했던 적이 몇번 있었는데 그럴 때마다 하나씩 실수하거나 논리적으로 맞지 않는 등 여러움을 겪었다. 이런 점들을 개선하기 위해선 코드를 많이 보고 만져봐야 된다는 생각이 들어 이번 대회에서는 모델 코드 뿐만 아니라 여러 실험을 위한 코드들을 많이 짜고 개선하려고 노력했다. 이런 노력 덕분에 GradCAM 코드를 구현하고 더 나아가 Streamlit으로 시각화 해 팀원 모두가 편하게 사용할 수 있도록 하였고, 양상블 코드도 구현하여 모델 성능 향상을 도모하였다.

▶ 나는 어떤 방식으로 모델을 개선했는가?

EDA를 통해 데이터셋을 살펴본 결과 train이나 test set이 매우 흡사했다. 따라서 이미지 특징을 변화시키는 증강 기법보다 이미지 밝기, 대비 등을 변화시키는 기법을 사용하는 게 더욱 효과적일 거라는 생각이 들었다. 따라서 대비를 주는 CLAHE 기법을 사용해 데이터 증강을 적용시켰다. 이어, 모델 간 양상블을 통해 모델 성능을 올릴 수 있도록 Hard Voting 양상블 기법을 적용할 수 있도록 코드를 구현하였다. 픽셀 클래스를 하나씩 비교해나가며 많은 모델이 예측한 클래스를 선택하는 방식을 통해 모델 성능을 올릴 수 있었다.

▶ 전과 비교하여 새롭게 시도한 변화는 무엇이고, 어떤 효과가 있었는가?

전에는 모델 자체의 성능을 올리기 위한 실험 위주로 진행했다면 이번 대회에선 그 외의 방법들을 통해 모델 성능을 높이려고 노력했다. 덕분에 그동안 어렵다고 생각했던 내용들을 이해할 수 있었고 코드를 이해하는 능력을 많이 발전시킬 수 있었다.

▶ 마주한 한계는 무엇이며, 아쉬웠던 점은 무엇인가?

그동안 했던 대회 내용을 바탕으로 대회를 진행한 것, ‘전에 데이터 증강 기법 적용했고 양상블 적용했으니까 이번에도 적용해보자’라는 생각으로 실험을 진행했던 것 같다. 이전에 했던 내용들과 배웠던 내용들에 대해 지금 대회를 진행한 내용을 살펴보며 새로운 인사이트를 얻어 실험을 진행하고 싶었는데 이런 점이 부족해 많이 아쉬웠다.

▶ 한계/교훈을 바탕으로 다음 프로젝트에서 시도해볼 것은 무엇인가?

위에서 작성한 한계를 개선하기 위해선 지금 진행되는 내용에 대해서 정확히 아는 것이 중요하다는 생각이 들었다. 최종 프로젝트에서 모델링 파트를 맡았으니 다음엔 내가 진행한 내용 이외에도 다른 팀원이 진행한 내용에 대해 정확히 파악하고 이를 바탕으로 새로운 실험을 어떤 걸 하면 좋을지 생각하려 한다. 모델 실험 과정을 그때 그때 정리하고 모르는 부분은 물어보며 이해하며 최대한 모든 내용을 파악할 수 있도록 할 것이다.

선경은

▶ 나는 내 학습목표 달성을 위해 무엇을 어떻게 했는가?

이번 프로젝트는 단순 실험보다 구현을 많이 해보기 위해 프로젝트 모듈화 작업을 제안하였다. 전체 프로젝트 구조를 정리하여 팀원들과 파트를 나눠 작업하였고 github으로 많은 코드 리뷰를 통해 개선해 나갔다. 최종적으로 config 파일만 수정하면 모든 코드가 문제 없이 실행되도록 완성하였고 다양한 실험도 쉽게 진행할 수 있었다.

▶ 나는 어떤 방식으로 모델을 개선했는가?

모델의 구조를 FCN, Unet, Unet++, DeepLabV3 다양하게 사용해보았고, 백본은 resnet50, efficientnet-b0,b5,b7, mobilenet로 다양하게 변경해가며 실험했다. 그 결과 가장 좋은 성능의 모델을 찾을 수 있었다.

▶ 전과 비교하여 새롭게 시도한 변화는 무엇이고, 어떤 효과가 있었는가?

데이터셋의 원본을 사용하여 바로 학습하는 것이 아닌 전처리를 미리 진행하여 새로 저장한 후 불러와서 사용하였다. 이 과정에서 단순히 저장하면 용량의 문제가 있었는데 이를 해결하기 위해 여러 압축 방법을 고민해서 적용해 보았다. 단순히 파일 용량만을 압축하는 것이 아닌 데이터 구조와 정보를 이해하고 데이터의 정보는 사라지지 않되 저장 공간을 줄이는 방법을 사용하였다.

▶ 마주한 한계는 무엇이며, 아쉬웠던 점은 무엇인가?

데이터 증강 실험이 효과가 거의 나타나지 않았다. 다른 팀들을 보니 증강으로 성능을 올린 것을 확인했다. 증강을 조금 더 분석해보고 우리가 어떤 적용을 잘못 하고 있는지를 파악했어야 하는데 이 부분이 부족했던 것 같다. 또한 프로젝트 처음에

겹치는 뼈에 대한 시각화를 했고 멀티 레이블 성능을 올리는 데 분석용으로 사용하려고 했는데 진행을 하다보니 이를 활용하지 못하고 단순히 모델 실험만 했던 것이 아쉽다.

▶ 한계/교훈을 바탕으로 다음 프로젝트에서 시도해볼 것은 무엇인가?

다음 프로젝트에서 모델링 파트를 맡았기 때문에 조금 더 실험 계획서를 자세하게 써보고 다양하게 분석해보는 시간을 많이 가져야할 것 같다. 또한 문제 상황과 해결 아이디어, 해결한 방법등을 작성해가며 진행해야겠다.

이선우

▶ 나는 내 학습목표 달성을 위해 무엇을 어떻게 했는가?

베이스라인 코드가 jupyter notebook 하나였기 때문에 직접 모듈화를 통해서 동일한 실험 환경을 만들 수 있도록 하는 것에 힘을 쏟았다. 실험을 진행할 때는 WanDB와 Artifact를 통해서 조금 더 빠른 공유가 가능하도록 하였다.

▶ 나는 어떤 방식으로 모델을 개선했는가?

모델을 고정한 후, 여러 backbone을 시도해보면서 성능을 확인했다. 또한, 여러 Loss function을 비교해가면서 성능 향상에 어떤 영향이 있는지 실험해보았다. 최종적으로는 양상불을 통해서 점수 향상을 노렸다.

▶ 전과 비교하여 새롭게 시도한 변화는 무엇이고, 어떤 효과가 있었는가?

데이터 로딩이 너무 오래 걸려서 로딩 병목에 대한 개선을 하기 위해 노력했다. 실제로 여러 방법을 통해 로딩 속도를 개선할 수 있어서 만족하고 있다. 또한, 실험 계획서를 미리 작성해서 내가 어떠한 실험을 하고 있는지 공지함과 동시에 실험에 대한 결과도 주기적으로 업데이트해주었다.

▶ 마주한 한계는 무엇이며, 아쉬웠던 점은 무엇인가?

전체적으로 실험보다 다른 내용들에 할애한 시간이 많았다는 느낌이 든다. 1-stage와 2-stage의 양상불 효과가 좋다는 것을 인지했었음에도 불구하고 YOLO를 이용한 Segmentation ensemble을 깜빡하고 시도해보지 않았다는 것이 아쉬웠다.

▶ 한계/교훈을 바탕으로 다음 프로젝트에서 시도해볼 것은 무엇인가?

전체적으로 프로젝트의 모듈화에 많이 신경을 쓰고 MLOps를 통한 훈련 과정의 자동화를 시도해 조금 더 퀄리티 높은 결과물을 만들어내고 싶다.

최현우

▶ 나는 내 학습목표 달성을 위해 무엇을 어떻게 했는가?

초반 학습 담당으로 베이스라인코드를 활용하여 여러 모델간의 성능차이를 비교하였습니다., num worker와 resize 옵션을 조정하며 데이터 로딩과 전처리 속도의 영향을 테스트했고, 학습 속도 테스트를 통해 가장 효율적인 학습 환경을 설정하려고 노력했습니다. 또한, streamlit을 활용해 학습 결과를 시각화하며 직관적으로 모델의 성능을 평가할 수 있는 환경을 마련했습니다. validation 테스트와 class29 + class2 테스트를 통해 모델의 정확도를 검증하고, random seed test를 실시하여 실험의 재현성을 확보했습니다.

▶ 나는 어떤 방식으로 모델을 개선했는가?

모델의 성능을 개선하기 위해, 데이터 전처리 단계에서 resize를 통해 성능 개선의 가능성을 확인하여 resize 크기별 score 향상을 실험계획서로 작성하여 공유하였습니다. 그리고 예측 성능이 상대적으로 낮았던 두 개의 클래스에 초점을 맞추어 이들에 대한 학습을 강화했습니다. 이 두 클래스에 대한 집중적인 학습 후, 결과를 양상불 방식으로 통합함으로써 모델의 성능을 개선하였습니다.

더 나아가, 팀원들이 각자 학습한 모델들에 대한 validation 과정을 거쳐, 가장 효과적인 양상불 조합을 도출해내기 위해 노력하였습니다.

▶ 전과 비교하여 새롭게 시도한 변화는 무엇이고, 어떤 효과가 있었는가?

이전에는 단일 조건으로 모델 학습을 진행했으나, 이번 프로젝트에서는 다양한 조건(데이터 사이즈 조정, 클래스 조합, random seed 조정)을 실험하여 모델의 성능과 안정성을 동시에 고려했습니다. 특히, streamlit을 이용한 시각화는 모델의 학습 과정과 결과를 직관적으로 이해할 수 있게 하여, 어떤 부분이 개선되어야 할지 빠르게 파악할 수 있는 효과가 있었습니다

▶ 마주한 한계는 무엇이며, 아쉬웠던 점은 무엇인가?

한계로는 시작부터 높았던 score로 인해 제대로된 효과가 있었는지 구분하기 어려웠습니다.

대회 초반부터 0.95이상의 score가 유지되어 왔기 때문에 실험 내용에 대해서 정확하게 향상이 있음을 파악하는데 어려움을 겪었고, TTA를 적용을 시켜보지 못한것이 아쉬웠습니다

▶ 한계/교훈을 바탕으로 다음 프로젝트에서 시도해볼 것은 무엇인가?

다음 프로젝트에서는 다양하게 시도해볼수있는 방식은 빠르게 실험하여 정리를 하며 진행하고 싶습니다. 또한 이번 대회와 같이 기본 score가 높은 상황에서 더욱 더 세밀한 실험 환경 구성과 실험 계획서 작성은 진행하여 실험 과정에서의 효율성을 극대화하여 더 좋은 분석과 결과를 도출 할 수 있을 것으로 생각됩니다.