

Hand Bone X-ray Image Segmentation

김경모, 김지은, 위정호, 이봉학, 조수빈

Naver Boostcamp AI Tech 8th Cohort, CV 01 Team

South Korea

요약

본 프로젝트는 손 뼈 X-ray 이미지에서 29개의 뼈 구조를 분할하는 Semantic Segmentation 과제를 다룬다. 고해상도 (2048×2048) 이미지에서 다수의 작은 뼈가 중첩된 상태로 존재하는 Multi-label Segmentation 문제의 특성을 분석하고, 이에 적합한 전처리, 모델링, 후처리 전략을 제안한다. CLAHE 기반 대비 강화, Sliding Window 추론, 다중 모델 앙상블을 통해 Dice Coefficient 0.9767을 달성하였다.

1 프로젝트 개요

뼈는 인체의 구조와 기능에 중요한 영향을 미치기 때문에, 정확한 뼈 분할은 의료 진단 및 치료 계획을 개발하는 데 필수적이다. Bone Segmentation은 인공지능 분야에서 중요한 응용 분야 중 하나로, 특히 딥러닝 기술을 이용한 뼈 Segmentation은 많은 연구가 이루어지고 있으며, 질병 진단, 수술 계획, 의료 장비 제작, 의료 교육 등 다양한 목적으로 활용될 수 있다.

1.1 Task 정의

본 프로젝트는 손 뼈 X-ray 이미지에서 29개의 뼈 구조를 분할하는 Semantic Segmentation 과제를 다룬다. 모델은 입력 이미지에 대해 각 클래스별 확률맵을 생성하는 멀티채널 예측을 수행한다. 이후 각 픽셀에 threshold 이상의 값을 갖는 클래스를 할당하여 최종 Segmentation 결과를 도출하며, 예측 결과는 RLE (Run-Length Encoding) 형식으로 변환하여 제출한다.

본 과제의 핵심 난점은 2048×2048 고해상도 이미지에서 다수의 작은 뼈가 서로 중첩된 상태로 존재한다는 점이다. 이로 인해 단일 픽셀에 복수의 클래스가 공존하는 Multi-label Segmentation 문제로 접근해야 한다.

1.2 평가 기준

본 프로젝트에서는 Semantic Segmentation 분야에서 널리 사용되는 Dice Coefficient를 평가 지표로 사용한다. Dice Coefficient는 예측 마스크와 정답 마스크 간의 중첩 정도를 정량화하는 지표로, 다음과 같이 정의된다.

$$Dice = \frac{2|GT \cap Pred|}{|GT| + |Pred|} \quad (1)$$

평가 시 29개의 뼈 클래스 각각에 대해 Dice 값을 산출한 후, 이를 평균하여 최종 점수를 도출한다. 이러한 방식은 클래스 간 크기 차이가 큰 데이터셋 특성을 고려하여, 특정 대형 클래스에 편향되지 않은 균형 잡힌 분할 성능을 평가하기 위함이다.

1.3 팀 구성 및 역할

이름	역할
김경모	Loss 실험, Validation set train 실험, 양상블
김지은	Multi-label 학습 전략 수립, 정규화 실험, Multi-channel featuring, Sliding window, 후처리 (Gaussian Weighted Averaging), 양상블 (Hard Voting)
위정호	코드 모듈화, 리팩토링, Validation Fold 실험, Dali, HRNet, 후처리 (CRF, morphology), Sliding window, 양상블 (Class-wise, Soft Voting)
이봉학	전처리 (Flip & One Hand Training, Center Crop), Sliding window, NNUNet 실험
조수빈	EDA, 전처리 (CLAHE, SSR, 이상치 제거), UNet+MiT, 후처리 (소형 객체 제거), 양상블 (Class-wise)

1.4 프로젝트 일정

리더보드 운영 기간: 2025.12.17(수) 10:00 – 2026.01.06(화) 19:00



그림 1: 프로젝트 일정 간트차트

1.5 개발환경

- CUDA Version : 12.2
- GPU : Tesla V100-SXM2-32GB 3대
- Nvidia-DALI : GPU 가속 데이터 파이프라인으로 CPU 전처리 병목 완화
- 주요 라이브러리 : PyTorch, Albumentations, Segmentation Models PyTorch(smp)
- 협업 툴 : Github, Notion, WandB, Slack

2 데이터 분석

2.1 데이터 구성 및 특징

2.1.1 Image

이미지는 2048×2048 해상도로 구성되며, Train 800장, Test 288장이 제공된다. 한 사람의 양손을 촬영하였으므로, Train은 400명, Test는 144명의 데이터로 구성된다.

2.1.2 Annotation

총 29개의 클래스에 대한 다각형(Polygon) Annotation 정보가 JSON 형식으로 제공된다. 클래스는 손가락(f1-f19, 19개), 손목(Trapezium, Trapezoid, Capitate, Hamate, Scaphoid, Lunate, Triquetrum, Pisiform, 8개), 팔(Radius, Ulna, 2개)로 구성된다.

2.1.3 Meta Data

촬영자의 ID, 성별, 키, 체중 정보가 제공된다.



그림 2: 데이터셋 구성: (좌) 원본 X-ray 이미지, (우) 29개 클래스 segmentation mask

2.2 EDA(Exploratory Data Analysis)

2.2.1 클래스 면적 분포

클래스 간 면적 불균형이 심하게 나타난다. 표 1에서 볼 수 있듯이, 가장 작은 클래스(finger-16)와 가장 큰 클래스(Radius) 간에 약 19배의 면적 차이가 존재한다. 특히 손목에 해당하는 8개 클래스는 면적이 작고 밀집되어 있어, 단일 해상도 Resize 기반 학습 시 정보 손실이 발생할 가능성이 높다.

표 1: 클래스별 평균 면적 (Top 5)

Smallest		Largest	
Class	Area (px)	Class	Area (px)
finger-16	3,895	Radius	75,128
Pisiform	4,327	Ulna	39,456
finger-4	4,950	finger-7	35,311
finger-12	5,668	finger-11	32,278
finger-8	5,743	finger-3	26,666

Distribution of Annotation Areas per Class (Log Scale)

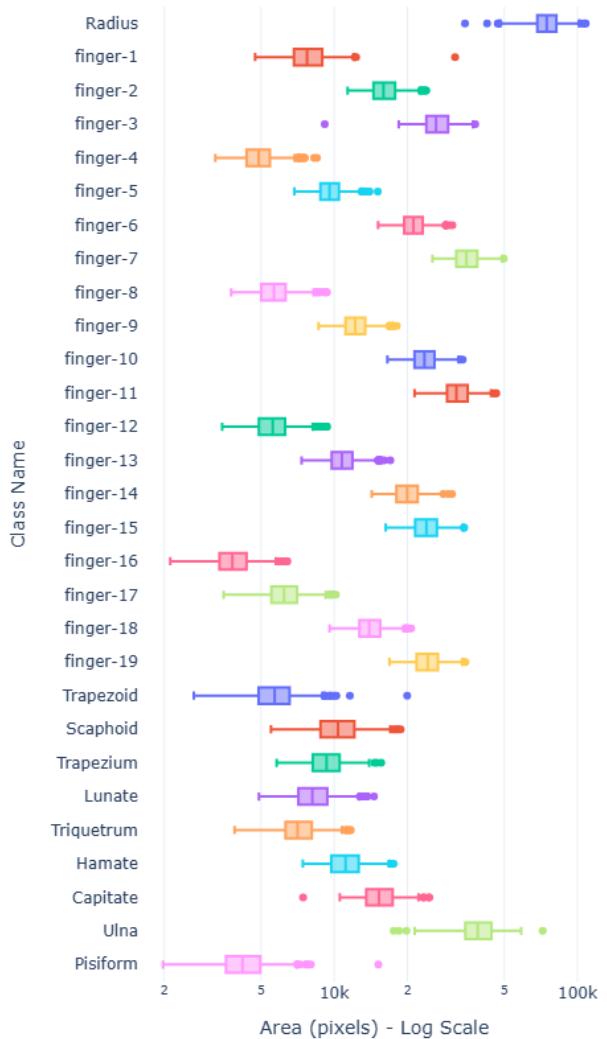


그림 3: 클래스별 annotation 면적 분포 (Log Scale)

2.2.2 픽셀 분포

뼈와 배경 간의 픽셀 분포는 뚜렷하나, 일부 영역에서 중첩이 확인된다. 배경에 포함된 근육, 살 등과 저밝기 뼈 영역 간의 혼동 가능성성이 존재한다.

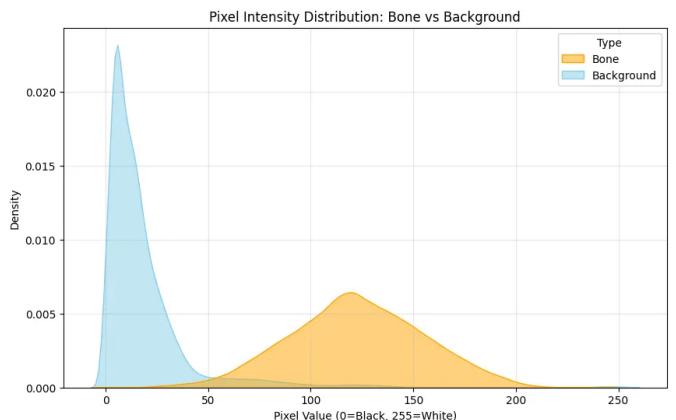


그림 4: 뼈와 배경 픽셀 분포

2.2.3 뼈 위치 분포

이미지 중심을 기준으로 양손의 뼈 annotation 중심 좌표를 시각화하였다. 손등, 팔, 중앙 손가락은 중심부에 밀집되어 있으나, 손을 꺾은 채 촬영된 경우 좌우로 분포가 흩어지는 경향을 보인다. 전체적으로 해부학적 구조에 따른 일관된 위치 패턴이 나타나므로, 무작위 위치 변형 적용 시 구조적 왜곡에 유의해야 한다.

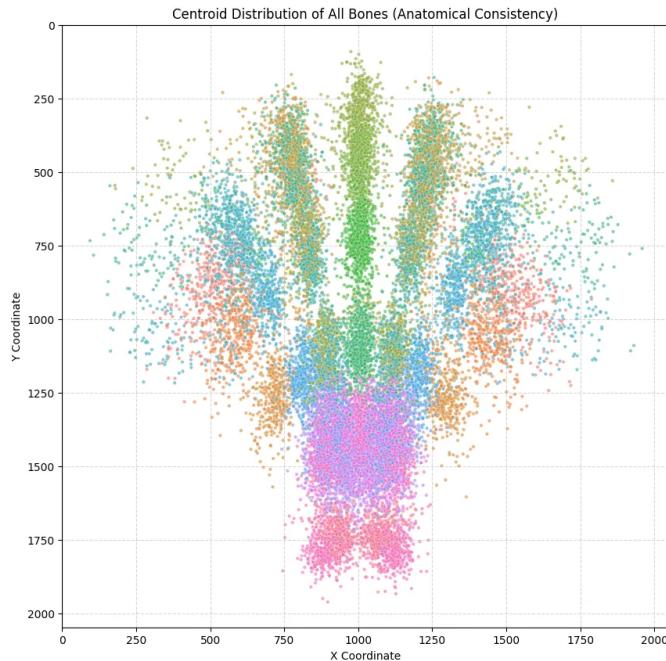


그림 5: 클래스별 annotation 중심 좌표 분포

2.2.4 공간적 중첩

X-ray 이미지는 3차원 인체 구조가 2차원으로 투영되므로, 뼈 간 중첩이 필연적으로 발생한다. 이러한 중첩 영역은 모델이 명확한 경계를 학습하기 어렵게 만드는 요인이다. 학습 데이터 800장 전체에서 중첩이 확인되며, 전체 픽셀 중 약 0.41%가 2개 이상의 클래스 라벨을 동시에 보유한다. 중첩은 주로 손목 부위의 작은 뼈들에 집중되어 있다.

표 2: 주요 중첩 클래스 쌍

Class A	Class B	Frequency	Avg IoU
Triquetrum	Pisiform	799	0.4483
Trapezoid	Trapezium	800	0.1885

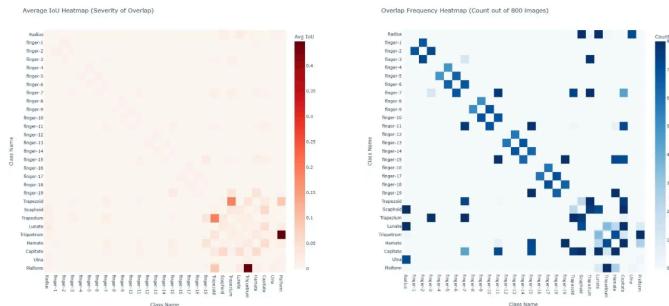


그림 6: 클래스 간 중첩 분석: (좌) 평균 IoU, (우) 중첩 비도

2.2.5 이상치 탐지

데이터셋에서 표3과 같은 다양한 이상 케이스와 라벨링 오류가 발견되었다.

표 3: 발견된 이상치 유형

유형	내용
반지 착용	finger-14 annotation에 반지 포함
네일아트	X-ray 상 밝게 나타나 뼈와 혼동 가능
손가락 휘어짐	일부 이미지에서 발견
라벨링 오류	전체 위치 반대, 손목 클래스 반대, 부골 라벨링

2.2.6 종자골

finger-3(엄지 손허리뼈) 옆에 위치한 동그란 형태의 뼈이다. 해부학적으로는 별도의 뼈이나, 해부학적으로는 별도의 뼈이나, 데이터셋 대다수는 이를 finger-3에 포함하여 라벨링하였다. 단, 800장 중 61장 (7.7%)은 종자골이 제외되어 있어 라벨링 일관성 문제가 존재한다.



그림 7: 종자골 X-ray 이미지 [1]

2.2.7 Meta Data

촬영자는 여성 436명, 남성 364명으로 구성되며, 20-30대가 주를 이루고 50대 이상은 소수이다. 신장과 체중이 클수록 손 크기가 비례하여 증가하며, 남성이 여성보다 약 1.2배 크게 나타난다. 본 데이터셋은 젊은 연령대 중심으로 구성되어 골밀도에 따른 화소값 변동이 크지 않으나, 고연령 데이터가 포함될 경우 이에 대한 처리가 필요할 것으로 예상된다.

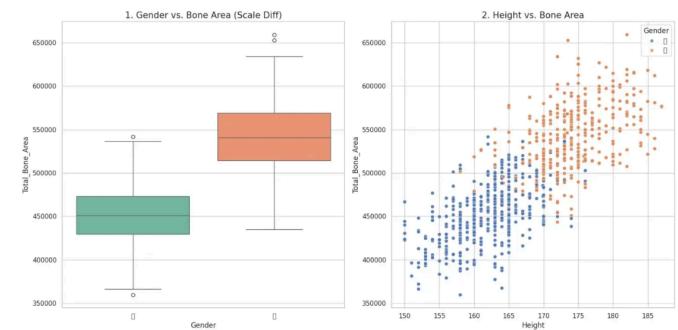


그림 8: 메타데이터 분석: (좌) 성별에 따른 빠 면적, (우) 신장과 빠 면적 관계

2.3 EDA 요약

EDA를 통해 도출된 전략적 방향은 다음과 같다.

- 클래스 간 면적 불균형 (최대 19배) → Multi-scale 학습 및 Sliding Window
- 고해상도 · 모호한 경계 → CLAHE 대비 강화
- 이상치 및 라벨링 오류 → 선별적 데이터 정제

3 실험

3.1 Base Model

기준 모델로 UNet-EfficientNet-B2를 선정하였다. EfficientNet-B2는 파라미터 대비 표현력이 우수하고, UNet Decoder는 의료 영상에서 검증된 구조이다. 이후 실험은 이 모델을 기준으로 변인 통제를 고려하여 진행하였다.

표 4: Base Model 설정

항목	설정
Architecture	UNet + EfficientNet-B2
Image Size	512 × 512
Epoch	25
Loss	Dice Loss
Optimizer	AdamW
Dice Score	0.9364

3.2 전처리 및 증강

3.2.1 Normalize

ImageNet 통계치 기반 정규화와 이미지별 정규화 (Percentile Clipping + Z-score)를 비교하였다. 이미지별 정규화가 촬영 조건 차이를 보정할 수 있어 유리할 것으로 예상하였으나, ImageNet 기반 정규화가 0.0027 높은 성능을 보였다. 이는 사전학습된 인코더가 ImageNet 통계치에 최적화되어 있어, 이를 벗어난 분포에서 특징 추출 성능이 저하되기 때문으로 분석된다.

표 5: 정규화 방식 비교

방법	Dice Score
ImageNet 통계치 기반 정규화	0.9364
이미지별 Percentile Clipping + Z-score	0.9337

3.2.2 CLAHE

X-ray 이미지에서 뼈의 모호한 경계선을 명확히 하기 위해 대비를 제한적으로 강화하였다. clip_limit=4.0 적용 시 오히려 성능이 하락하였으나, 2.0으로 낮추자 0.0052 향상되었다. 강도가 높을 경우 배경의 Pixel Value 도 함께 상승하여 모델이 혼동을 겪은 것으로 분석된다.

표 6: CLAHE clip_limit에 따른 성능

	Base	CLAHE 4.0	CLAHE 2.0
Val Dice	0.9512	0.9488	0.9567
LB Score	0.9364	0.9300	0.9416

3.2.3 Hand Crop

2048×2048 이미지에서 손 영역만 크롭하여 학습을 시도하였다. 배경 축소로 학습 효율 향상을 기대하였으나, Crop + Resize는 위치 정보 손실로 성능이 크게 하락하였으며, Padding도 개선 효과가 없었다.

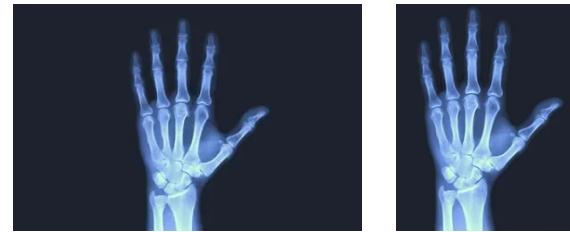


그림 9: Hand Crop 적용 예시: (좌) 원본, (우) 크롭 후

표 7: Hand Crop 방식에 따른 성능

	Base	Crop+Resize	Crop+Padding
Val Dice	0.9677	0.8434	0.9364

3.2.4 Horizontal Flip

좌우 방향을 오른손으로 통일하여 패턴의 일관성을 확보하여 성능을 향상시키고자 좌우 방향을 오른손으로 통일하는 실험을 수행하였다. finger-1 기준으로 이미지 중심선 왼쪽에 위치할 경우 왼손으로 판정하고, Flip을 적용하여 모든 이미지를 오른손 형태로 통일한 뒤 학습 및 추론을 진행하였다. 그러나 오히려 성능이 하락하였다.

확률적 좌우 반전 (HorizontalFlip, p=0.5)을 통한 데이터 다양성 확보도 시도하였으나 마찬가지로 성능이 하락하였다. Horizontal Flip은 일반적인 증강 기법이나, 본 데이터셋은 양손이 하나의 이미지에 동시에 포함된 구조이다. Flip 적용 시 좌우 손의 해부학적 위치 관계가 왜곡되어 공간적 문맥 (spatial prior) 학습이 불안정해지며, 데이터 다양성 증가보다 구조적 일관성 저하가 더 크게 작용한 것으로 분석된다.

표 8: Horizontal Flip 방식에 따른 성능

	Base	Right Hand Flip	HorizontalFlip
Val Dice	0.9677	0.9353	0.9653

3.2.5 SSR (Shift, Scale, Rotate)

촬영 시 발생할 수 있는 미세한 위치 및 각도 변화에 대한 강건성을 확보하기 위해 적용하였다. SSR 적용 후 빈 배경을 검은색으로 채우도록 설정

하였으며, 영역별 대비를 조정하는 CLAHE가 이에 영향을 받지 않도록 CLAHE 후 SSR을 적용하는 순서로 진행하였다.

일부 이미지는 손끝이 테두리에 근접해 있어 SSR 강도를 약하게 설정하였다. 증강 적용 시 클래스 전체가 사라지지 않을 정도로 강도를 조정하였으며, 해상도 1024 기준 Shift limit = 0.025, Scale limit = 0.025, Rotate limit = 10으로 설정하였다.

512 해상도에서는 오히려 성능이 하락하였다. 원본 2048 이미지를 512로 압축하면서 발생한 픽셀 정보 손실과 Rotate 보정 시 추가 정보 손실이 원인으로 분석된다. 이후 해상도를 1024로 높여 실험을 진행한 결과 성능이 향상되었다.

표 9: SSR 적용 및 해상도에 따른 성능

	Base (512)	SSR (512)	Base (1024)	SSR (1024)
Val Dice	0.9569	0.9450	0.9656	0.9677
LB Score	0.9420	0.9342	0.9618	0.9643

3.2.6 Multi-channel Featuring

X-ray 이미지는 기본적으로 단일 채널(Grayscale)이나, ImageNet으로 사전학습된 인코더는 3채널 RGB 입력을 기대한다. 단순히 동일한 그레이스케일을 3채널로 복제하는 대신, 각 채널에 서로 다른 특징 정보를 담아 모델에 더 풍부한 입력을 제공하고자 하였다.

채널 구성은 표 10와 같다.

표 10: Multi-channel 구성

채널	설명	목적
channel 0	CLAHE (clip_limit=2.0)	대비 강화된 원본 정보
channel 1	Sobel Edge Detection	뼈 경계 정보 강조
channel 2	Morphological Gradient	형태학적 경계 정보

그러나 Baseline 대비 성능이 하락하였다. 주요 원인으로는 첫째, ImageNet으로 사전학습된 인코더가 자연 영상에 최적화되어 있어 인공적 특징 맵 분포에 적응하지 못한 점, 둘째, CLAHE 채널에 이미 경계 정보가 포함되어 추가 채널이 정보 중복과 노이즈만 유발한 점, 셋째, 채널별 상이한 값 범위에 대한 정규화가 최적이 아니었던 점을 들 수 있다.

표 11: Multi-channel Featuring 실험 결과

방법	Dice Score	비고
Baseline (3ch Grayscale + CLAHE)	0.9618	-
Multi-channel 실험1	0.9585	-0.0033
Multi-channel 실험2	0.9523	-0.0095

실험1: 원본에 dilate/erode 연산 후 차분으로 Morphological Gradient 계산

실험2: CLAHE 적용 이미지에 cv2.morphologyEx(MORPH_GRADIENT) 직접 적용

3.2.7 이상치 제거

EDA(2.2.5 이상치 탐지)에서 확인된 annotation 오류 4건(반지 포함 1건, 라벨링 오류 3건)을 학습 데이터에서 제거하였다. 네일아트, 흰 손가락, 보철물 등은 annotation에 큰 이상이 없어 다양성을 위해 유지하였다.

표 12: 이상치 제거에 따른 성능

	제거 전	제거 후
Val Dice	0.9674	0.9702
LB Score	0.9673	0.9685

3.3 학습 전략

3.3.1 Sliding Window

본 데이터는 2048×2048 고해상도 이미지이나, 메모리 제약으로 축소 해상도에서 학습을 수행하였다. 단일 Resize 기반 학습은 소형 뼈(Finger tip, Pisiform) 및 경계 정보 손실을 유발할 수 있어, 원본 해상도를 보존하기 위해 Sliding Window를 적용하였다.

입력 이미지를 1024×1024 패치로 분할하여 각 패치에 대해 독립적으로 학습 및 추론을 수행하였다. 추론 시 패치 간 중첩 영역은 예측값을 평균화하여 경계 부근의 불연속성을 완화하였다. Stride 512 적용 시 패치 간 50% 중첩이 발생하여 경계 영역의 예측 안정성이 향상되었다.

대부분의 모델에서 Sliding Window 적용 시 성능이 향상되었으며, 특히 손목 영역의 소형 뼈에서 False Positive가 감소하였다. 다만 SegFormer는 오히려 성능이 하락하였는데, 이는 Transformer의 Global Context Modeling 특성상 전체 이미지 맥락이 필요한데, 패치 단위 입력으로 인해 장거리 의존성 학습이 제한되었기 때문으로 분석된다.

표 13: Sliding Window 적용에 따른 모델별 성능

Model	Window	Stride	Baseline	LB Score
DeepLabv3	1024	512	0.9681	0.9726
DeepLabv3	1024	1024	0.9681	0.9702
UNet-EfficientNet	1024	1024	0.9490	0.9641
SegFormer-b4	1024	512	0.9672	0.9319
nnUNet	1024	512	0.9689	0.9741
HRNet	1024	512	0.9724	0.9752

3.3.2 Label Strategy

학습 데이터의 Ground Truth는 뼈 간 중첩을 허용하는 Multi-label 형식이다. 평가 지표인 Dice를 극대화하기 위해 중첩 정보 처리에 대한 전략적 의사결정이 필요하였다.

Pisiform과 같이 크기가 작은 뼈는 Dice의 분모가 작아 FP에 매우 민감하다. 이에 중첩을 허용하는 Sigmoid(독립 확률)와 중첩을 배제하는 Argmax(상호 배타적 분류) 방식을 비교하였다. Argmax는 FP를 원천 차단할 수 있어 유리할 것으로 예상되었으나, 실제로는 Sigmoid 방식이 압도

적으로 우수한 성능을 보였다. 가장 많이 겹치는 Triquetrum-Pisiform 쌍의 경우, Argmax 적용 시 중첩된 실제 정답(TP)이 잘려나가면서 Dice 점수가 0.91에서 0.60으로 급락하였다. 이를 토대로 Multi-label(Sigmoid) 기반의 학습 전략을 확정하였다.

표 14: Label Strategy에 따른 성능 비교

방식	Dice Score
Sigmoid (Multi-label)	0.9621
Argmax (Multi-class)	0.8122

3.3.3 Loss

평가 지표가 Dice이므로 Dice Loss를 기본으로 하고, BCE, Focal, Lovasz 와의 조합을 비교하였다.

실험 결과, Validation Dice는 Dice Loss 단독 사용 시, LB Score는 Dice + BCE 조합 시 가장 높았다. 다만 Dice, Dice + BCE, Dice + Focal 간 성능 차이가 크지 않아, 이후 실험에서는 각 모델 특성에 맞게 세 Loss 중 하나를 선택하여 사용하였다.

표 15: Loss 함수에 따른 성능 비교

Loss	Val Dice	LB Score
Dice	0.9638	0.9491
BCE	0.9148	-
Dice + BCE	0.9634	0.9493
Dice + Focal	0.9630	0.9476
Dice + Lovasz	0.9255	0.9170

3.3.4 Validation Fold

Validation Score와 LB Score 간의 차이를 최소화하기 위해 Validation set 선정 과정을 거쳤다. k=5 Fold Cross Validation을 수행하고 각 Fold 별 결과를 모두 제출하여, Validation Score와 LB Score의 차이가 가장 작은 Fold를 선정하였다.

표 16: Fold별 성능 비교

Fold	Val Dice	LB Score	diff
0	0.9476	0.9323	0.0153
1	0.9479	0.9327	0.0152
2	0.9583	0.9420	0.0163
3	0.9507	0.9365	0.0142
4	0.9543	0.9423	0.0120

Fold 4가 가장 높은 성능과 가장 작은 차이를 보여 해당 Fold를 Validation set으로 사용하였다.

3.3.5 Validation Train

본 프로젝트에서 Validation set은 전체 데이터의 20%로 설정하였다. Validation set이 160장으로 비교적 큰 데이터라고 판단하여, 학습을 완료한

모델에 Validation set을 추가로 학습시켜 모든 데이터를 활용하고자 하였다.

표 17: Validation Train 실험 결과

	LB Score
Base	0.9652
추가 학습	0.9656

추가 학습을 통해 소폭의 성능 향상을 확인하였다.

3.4 모델

단일 아키텍처에 의존하기보다, 서로 다른 inductive bias를 가진 모델들을 후보군으로 설정하고 비교 실험을 진행하였다.

nnU-Net v2. CNN 기반 Encoder-Decoder 구조로, 데이터 특성 기반 자동 파이프라인 최적화를 수행한다. 고해상도 의료 영상에서 검증된 프레임워크로, 수작업 설계 대비 상한선 성능(Upper Bound)을 확인하고자 선정하였다.

SegFormer. Transformer Encoder + MLP Decoder 구조로, Global Context Modeling을 수행한다. 손가락 끝-손목 뼈 간 장거리 관계 파악 및 작은 뼈 검출 성능 보완을 기대하여 선정하였다.

DeepLabV3+. CNN Encoder + ASPP + Decoder 구조로, 멀티스케일 특징 추출을 수행한다. 클래스 스케일 불균형 대응 및 대형/소형 뼈 동시 학습 성능 확인을 위해 선정하였다.

UNet + MiT. Transformer Encoder (MiT-b3) + CNN Decoder (UNet) 구조의 하이브리드 아키텍처이다. 정형화된 손 뼈 구조에서 Transformer의 맥락적 추론 능력을 기대하여 선정하였다.

HRNet. CNN 기반 다중 해상도 병렬 구조로, 고해상도 Feature Map을 전 과정에서 유지한다. 픽셀 정밀도가 중요한 의료 영상에서 Fine-grained segmentation 성능 확인을 위해 선정하였다.

3.5 추론 전략 및 후처리

3.5.1 TTA (Test Time Augmentation)

멀티 스케일(0.8, 1.0, 1.5)로 이미지를 추론한 후 예측값을 평균하여 최종 예측을 생성하였다. 두 모델 모두 TTA 적용 시 성능이 향상되었다.

표 18: TTA 적용에 따른 성능

Model	TTA	LB Score
DeepLabV3+	-	0.9681
DeepLabV3+	O	0.9695 (+0.0014)
SegFormer-b3	-	0.9602
SegFormer-b3	O	0.9624 (+0.0022)

3.5.2 Class-specific Threshold

Multi-label(Sigmoid) 학습에서 모든 클래스에 동일한 임계값(0.5)을 적용하면 데이터 크기 불균형과 지표 민감도를 반영하지 못한다는 한계가 있다. 이에 클래스별 특성에 따라 최적 임계값을 탐색하였다.

작고 중첩이 많은 클래스(Pisiform, Trapezoid 등)는 임계값을 0.4로 낮추어 Recall을 확보하였고, 경계가 모호한 클래스(finger-16, Triquetrum)는 0.6으로 높여 Precision을 확보하였다. 크고 분리가 명확한 클래스는 기본값 0.5를 유지하였다.

표 19: Class-specific Threshold 설정

Thr.	Classes	특성	개선
0.4	finger-3,4,5,14,15,18, Pisiform, Trapezoid, Trapezium, Hamate, Capitate	작음, 중첩	+0.0026
0.5	finger-1,2,6,7,8,9,10,11,12,13,17,19, Scaphoid, Lunate, Radius, Ulna	크고, 분리	-
0.6	finger-16, Triquetrum	모호한 경계	+0.0012

3.5.3 소형 객체 제거

추론 결과에서 미세한 노이즈 픽셀을 제거하기 위해 cv2.connectedComponentsWithStats를 적용하였다. 종자골 객체가 함께 제거되지 않도록 threshold를 900px로 설정하였으나, 성능 변화는 없었다.

3.5.4 가우시안 스무딩

슬라이딩 윈도우 추론 시 패치 간 중첩 영역의 예측값을 병합하는 방법을 비교하였다. 단순 평균(Simple Average) 대비 가우시안 가중 평균 적용 시 패치 경계의 불연속성이 완화되어 Validation Dice가 0.8441에서 0.8789로 향상되었다.

3.5.5 CRF(Conditional Random Field)

CNN 기반 Segmentation 모델은 픽셀 단위로 독립적인 예측을 수행하므로, 경계가 뭉개지거나 불필요한 노이즈가 발생할 수 있다. X-ray 이미지 특성상 뼈와 배경의 명암 차이가 뚜렷하므로, 픽셀 강도(Intensity) 정보를 활용하여 경계면을 정교하게 보정하고자 CRF를 적용하였다.

기존 DenseCRF의 높은 연산 비용을 해결하기 위해 ConvCRF를 도입하였다. Message Passing을 Convolution 연산으로 근사하여 GPU 병렬 처리가 가능하도록 최적화하였다.

그러나 양상을 확률 맵에 적용 시 오히려 Dice Score가 감소하였다. 양상을(Soft Voting)로 이미 정제된 확률 맵에 CRF를 추가 적용할 경우, 픽셀 강도 기반의 강제적인 경계 재조정이 소형 뼈나 관절 간격 같은 미세 구조를 뭉개뜨려 정확도가 저하된 것으로 분석된다.

3.6 Ensemble

3.6.1 Class-wise

각 클래스별로 가장 우수한 모델의 예측을 선택하거나 더 높은 가중치를 부여하는 전략이다. 최고 성능 모델을 Base로 하여 클래스별로 더 우수한 모델의 RLE로 대체하는 방식과, Grid Search와 Nelder-Mead Simplex 알고리즘으로 클래스별 가중치 매트릭스(29, N_models)를 최적화하는 방식을 시도하였다. 그러나 두 방식 모두 검증 데이터에 과적합되어 성능이 하락하였다.

3.6.2 Hard Voting

모델들의 예측 결과를 다수결 원칙으로 합치는 방식이다. 각 모델이 출력한 이진 마스크를 합산하여 임계값(N//2) 이상 득표한 픽셀만 양성으로 판단한다. 특정 모델에서 발생하는 False Positive를 효과적으로 제거하고 Precision을 높이는 데 기여한다.

표 20: Hard Voting 양상을 결과

모델	Dice Score
HRNet_w48	0.9724
DeepLabV3	0.9702
Segformer_b4	0.9672
Ensemble	0.9742 (+0.0018)

3.6.3 Soft Voting

모델의 출력을 확률값 그대로 사용하여 평균을 내는 방식으로, 정보 손실을 최소화한다. 모든 모델에 동일한 가중치(1/N)를 부여하는 Uniform 방식을 베이스라인으로 사용하였다. 모델별 최적 가중치를 탐색하는 Optimized 방식도 시도하였으나, 검증 데이터에 과적합되어 성능이 하락하였다.

3.7 최종 모델

최종 제출은 두 가지 모델을 선정하였다.

모델 1. HRNet, nnUNet, DeepLabV3를 Soft Voting(Uniform)으로 양상을한 구조이다. 모든 모델에 동일한 전처리(Normalization, CLAHE)와 증강(SSR)을 적용하였으며, Sliding Window(Window 1024, Stride 512)를 통해 고해상도 정보를 보존하였다. 추론 시 소형 객체 제거(De-noise)를 적용하였다.

모델 2. 모델 1과 동일한 양상을 구조에 소형 객체 제거 대신 TTA, 가우시안 스무딩을 적용한 버전이다.

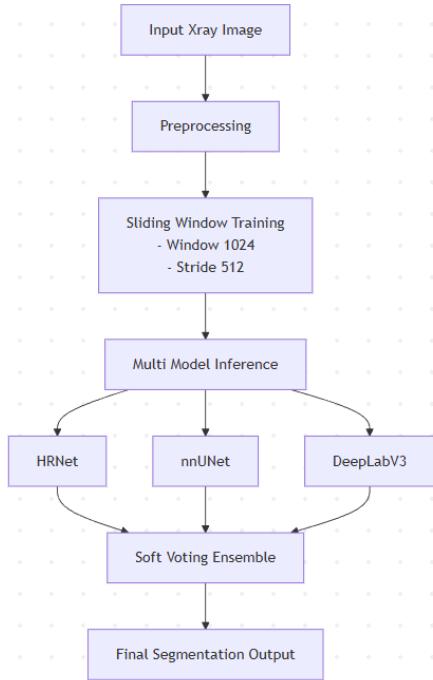


그림 10: 최종 모델 파이프라인

4 실험 결과

최종 선정된 두 모델의 Leaderboard 성능은 다음과 같다.

표 21: 최종 모델 성능

	Public	Private
모델 1 (Denoise)	0.9757	0.9766
모델 2 (TTA + Gaussian)	0.9758	0.9767

두 모델 모두 유사한 성능을 보였으며, Public과 Private 간 점수 차이가 거의 없어 일반화 성능이 안정적임을 확인하였다. 최종 Private 기준 모델 2가 0.9767로 가장 높은 성능을 기록하였다.

5 결론

본 연구는 손 뼈 X-ray 이미지에서 29개의 뼈 구조를 분할하는 Semantic Segmentation 과제를 수행하였다. 고해상도 이미지에서 다수의 작은 뼈가 중첩된 상태로 존재하는 Multi-label Segmentation 문제의 특성을 분석하고, 이에 적합한 전처리, 모델링, 후처리 전략을 제안하였다.

EDA를 통해 클래스 간 면적 불균형, 픽셀 단위 Multi-label 중첩, 고해상도·모호한 경계 등의 특성을 파악하였고, 이를 기반으로 CLAHE 기반 대비 강화, SSR 증강, Sliding Window 추론, 다중 모델 양상률 등의 전략을 수립하였다.

최종적으로 HRNet, nnU-Net, DeepLabV3 세 모델을 Sliding Window 와 Soft Voting으로 양상률하여 Dice Coefficient 0.9758을 달성하였다.

이는 체계적인 EDA와 변인 통제를 통한 실험 관리, 그리고 다양한 모델의 상호 보완적 양상률이 효과적이었음을 보여준다.

6 적용 기법 정리

본 프로젝트에서 시도한 전체 기법과 최종 모델 적용 여부는 표 22에 정리하였다.

표 22: 전체 실험 기법 요약

구분	기법	적용
전처리 및 증강	Normalize	✓
	CLAHE	✓
	Hand Crop	
	Horizontal Flip	
	SSR	✓
	Multi-channel	
학습 전략	이상치 제거	✓
	Sliding Window	✓
	Label Strategy	✓
	Loss (Dice+BCE)	✓
	Validation Fold	✓
추론 전략 및 후처리	Validation Train	
	TTA	✓
	Class-specific Thr.	
	소형 객체 제거	✓
양상률	가우시안 스무딩	✓
	CRF	
	Class-wise	
양상률	Hard Voting	
	Soft Voting	✓

7 논의

7.1 주요 성과

- 체계적인 실험 관리와 변인 통제: GitHub, WandB, Notion, Slack을 활용하여 실험 설정과 결과를 기록하고, 단일 변인 변경 원칙을 유지함으로써 각 기법의 성능 기여도를 명확히 분석할 수 있었다.
- 파이프라인 모듈화: 데이터 로딩, 전처리, 학습, 추론 단계를 모듈화하여 다양한 실험을 빠르게 반복할 수 있었으며, 후반부 Sliding Window 및 TTA 결합 실험에도 유연하게 대응할 수 있었다.
- 학습 병목 개선: NVIDIA DALI를 활용하여 이미지 전처리 병목을 완화함으로써 학습 시간을 단축하고, 더 많은 실험을 수행할 수 있는 환경을 구축하였다.

7.2 핵심 인사이트

- 문제 정의의 중요성: 본 데이터는 단일 픽셀에 다수 클래스가 공존하는 구조를 가지며, 이를 Multi-class가 아닌 Multi-label Segmentation 문제로 정의한 것이 이후 모든 모델 및 후처리 전략의 핵심 전제가 되었다.

- **데이터 이해에 따른 성능 상한:** 손 뼈의 해부학적 구조, 클래스 간 면적 불균형, 중첩 패턴을 EDA 단계에서 충분히 분석한 것이 이후 실험 단계에서의 전처리, 모델 구조, Sliding Window 추론 전략으로 자연스럽게 이어졌다.
- **메타데이터의 역할:** 성별, 키, 체중 등의 메타데이터는 뼈 크기 분포와 상관관계를 보였으나, 본 문제에서는 이미지 전처리 및 스케일 보정으로 대부분 흡수 가능함을 확인하였다.

7.3 한계점

- **라이브러리 의존성 증가:** nnU-Net, segmentation models pytorch (smp) 등 고수준 라이브러리를 활용하면서 빠른 실험은 가능했으나, 일부 모델은 내부 동작을 충분히 이해하지 못한 채 사용하게 되는 한계가 있었다.
- **도메인 지식 부족:** 의료 영상(손 뼈 X-ray)에 대한 해부학적·의학적 배경 지식이 충분하지 않아, 일부 클래스 간 관계나 오류 패턴을 해석하는 데 제약이 있었다.

- **협업 규칙의 일관성 부족:** GitHub 커밋 메시지 및 브랜치 관리 규칙이 규정되지 않아, 실험 이력 추적에 일부 불편함이 발생하였다.

7.4 향후 개선 방향

- **모델 내부 구조에 대한 이해 강화:** 단순히 성능 비교에 그치지 않고, Encoder–Decoder 구조 및 Attention 메커니즘을 보다 깊이 이해하고 직접 구현하는 방향으로 확장할 필요가 있다.
- **도메인 지식 기반 전략 설계:** 다음 프로젝트를 진행할 경우, 해당 도메인과 관련된 사전 조사 단계를 EDA와 함께 진행하면서 좀 더 체계적이고 확실한 근거에 기반한 전략 수립을 세우고자 한다.
- **실험 관리 고도화:** 대조군 (Baseline)과 변경 요소를 명확히 기록하는 실험 관리 체계를 강화하여, 실험 결과 해석의 신뢰도를 높이고 재현성을 개선할 예정이다.

참고 문헌

- [1] OpenStax. Anatomy and Physiology. <https://openstax.org/books/anatomy-and-physiology/pages/1-7-medical-imaging>

개인 회고

▼ 김경모

1. 프로젝트 목표 및 역할

- 개인 학습 목표:
 - Segmentation 이해 : 따로 공부해보지 않았기 때문에 프로젝트를 하면서 공부하고 이해하는 것을 목표로 했다.
 - 협업툴 사용 : 이전 프로젝트에서 팀원들과 원활한 교류가 잘 안되었다고 생각해서 적극적으로 협업툴을 사용하고자 했다.
 - 알게 된 내용 정리 : 그동안 강의나 따로 찾아본 내용들을 따로 정리하지 않아서, 빨리 잊어버리는 거 같아 기록을 남기는 것이 중요하다고 생각했다.
- 주요 기여 역할
 - Loss 함수 테스트, Validation set으로 학습

2. 목표 달성을 위한 과정 (나의 핵심 활동)

- 기술적 시도 및 성과:
 - 이번 Hand bone segmentation task에 맞는 loss function을 찾기 위해 다양한 loss에 대한 성능을 평가함. 결과적으로 Dice를 사용하는 것이 가장 성능이 좋았지만, 모델에 따른 차이가 있기 때문에 상황에 맞게 변경하는 경우도 있었음.
 - Validation set도 약 160장 정도의 개수이기 때문에 fine-tuning 된 모델에 해당 Validation set으로 추가 학습하면 성능이 더 오를 것으로 기대하여 실험을 진행함. val set + train의 20%를 사용하여 학습했지만, 성능이 떨어졌음. 하지만 Validation set을 학습에 넣는 건 괜찮은 아이디어인 것 같아서, 적절한 하이퍼파라미터를 찾은 모델에 validation 없이 모든 데이터를 사용하는 방법을 통해 약간의 성능 개선을 이루었음.

3. 한계점 및 아쉬운 점

- 실험을 다양하게 진행했고, 그 과정에서 다양한 경험을 할 수 있었지만, 직접 진행한 실험을 통해 큰 성능 개선이 되는 것을 경험하지 못함
- 시간 분배의 문제로 강의에 집중하지 못함

4. 프로젝트를 통해 얻은 교훈 및 다음 목표

팀원과 협업하며 점차 높은 리더보드 스코어를 달성하는 과정에서 저번 대회의 문제점이 소통과 협업툴의 부재라는 것을 깨달았다. 대회를 진행하면서 자연스럽게 Segmentation이라는 분야를 공부하게 되었고, 다양한 방식의 모델을 알게 되는 기회가 되었다. 하지만 시간 분배를 잘 못한 탓인지 제공하는 강의에는 조금 소홀해지게 되었고 초기 목표였던 내용 정리도 후반부터는 흐지부지된 것 같아 반성하게 되었다. 다음 프로젝트부터는 정리를 꾸준히 하고, 블로그를 시작하는 것을 목표로 할 예정이다.

▼ 김지은

1. 프로젝트 목표 및 역할

- 개인 학습 목표:
 - 라이브러리 숙련 : **SMP(Segmentation Models Pytorch), MMSegmentation** 등 다양한 라이브러리를 활용해보고, Encoder-Decoder 구조를 자유롭게 커스터마이징하여 실험할 수 있는 능력 기르기.
 - 모델 탐색 및 선정 : UNet 계열과 Transformer계열 모델 비교, 탐색하여 속도와 정확도 확인.
 - 데이터 전처리 및 증강 전략 수립 : EDA 결과(뼈의 중첩, 모호한 경계)를 바탕으로, 모델 학습에 부정적 영향을 미치는 요소를 해결하기 위한 **Sliding Window, CLAHE, Rotate** 등의 구체적인 전처리 전략을 수립.
 - Fine-tuning 및 성능 최적화 : 모델 출력값을 단순히 Thresholding 하는 것 을 넘어, 데이터 분포에 기반한 후처리 최적화를 수행.
- 주요 기여 역할
 - **Data & Strategy:** Multi-label 학습 전략 수립, 정규화(Normalization) 실험, Multi-channel featuring 적용.
 - **Modeling:** SegFormer 기반 모델 실험 및 성능 검증.
 - **Inference & Post-processing:** 고해상도 대응을 위한 **Sliding Window** 기법 적용, **양상블(Hard Voting)** 구현

2. 목표 달성을 위한 과정 (나의 핵심 활동)

- 기술적 시도 및 성과:
 - 빠른 속도를 위해 Segformer를 베이스라인으로 선정하고, Multi-channel Featuring 실험.

- EDA → 모델학습 순환적 루프를 통해 '작은 뼈'보다는 '뼈 중첩'이 병목임을 파악하고, Sliding window, Overlap penalty loss 실험.
- 하드보팅 양상을 기법 적용

3. 한계점 및 아쉬운 점

- 라이브러리를 통해 최신 모델을 다뤘지만 아키텍처에 대한 깊이 있는 이해가 부족하여 특정 클래스(중첩 뼈)의 성능이 낮은 이유에 대해 모델의 receptive field 문제인지, gradient 업데이트의 문제인지 근본적인 해답을 얻지 못하여 아쉬웠다. 프로젝트가 끝나면 라이브러리 의존도를 낮추고 논문 구현 단계까지 학습하여 데이터 특성에 맞춰 아키텍처를 수정할 수 있는 역량을 기르고자 한다.

4. 프로젝트를 통해 얻은 교훈 및 다음 목표

- 지난 프로젝트에 비해 협업툴을 적극적으로 활용하여 팀원 간 실험 결과 공유와 코드 버전 관리가 원활하게 이루어졌다. Notion을 통한 EDA 결과 문서화, GitHub를 통한 코드 관리, 그리고 실험 로그의 체계적인 기록을 통해 중복 실험을 방지하고 효율적인 역할 분담이 가능했다.

▼ 위정호

1. 프로젝트 목표 및 역할

- 개인 학습 목표:
 - 팀원 간의 효율적인 협업을 위한 코드 모듈화(Modularization) 및 실험 표준화(Standardization)
 - 다양한 모델(HRNet 등)과 후처리 기법(CRF, Morphology)을 실험하며 최적의 조합을 찾아내는 Inference Optimization 경험 축적
- 주요 기여 역할
 - 실험 환경 표준화 및 리팩토링: 코드를 모듈화하고 리팩토링하여, 팀원들이 동일한 조건에서 객관적으로 실험 결과를 비교할 수 있도록 변인 통제(Control Variables) 환경 조성
 - Inference Pipeline 고도화: DALI 기반 데이터 로딩 가속화, 양상을(Soft Voting 최적화) 및 후처리 파이프라인 구축 및 검증

2. 목표 달성을 위한 과정 (나의 핵심 활동)

- 엄밀한 변인 통제 및 비교 분석: Validation Fold 구성, 전처리 방식 통일 등 실험의 객관성을 확보하기 위한 기반 작업을 수행하여 팀 전체의 의사결정 신뢰도를 높임

- DALI & HRNet 도입: 대용량 데이터 처리 속도를 높이기 위해 DALI를 적용하고, Segmentation에 특화된 High-Resolution Network(HRNet)를 실험하여 베이스라인 성능을 끌어올림
- Automated Soft Voting Optimization: Hand X-ray 데이터의 구조적 유사성을 활용, 검증 셋에 대해 Grid Search와 Nelder-Mead 알고리즘을 결합한 최적 가중치 탐색을 수행하여 단순 평균 대비 성능 향상
- 후처리 기법의 심층 검증: 앙상블 결과에 대해 Morphology(Opening/Closing)와 ConvCRF 등 다양한 후처리를 시도하고, 정량/정성적 분석을 통해 최종 파이프라인에서의 채택 여부를 논리적으로 결정함

3. 한계점 및 아쉬운 점

- 모델 간 시너지 부족: HRNet 외에 더 다양한 이기종 모델(Transformer 기반 등)을 실험하여 앙상블의 다양성(Diversity)을 극대화하지 못한 점이 아쉬움. Transformer 기반 모델들은 Sliding window의 효과가 떨어져서 최종 성능이 잘 나오지 않아 최종 앙상블 모델로 활용하지 못했는데 2048 해상도를 활용할 수 있는 다른 추론 방식에 대해 생각해내지 못함
- 후처리 최적화 시간 부족: CRF 등의 후처리 기법이 앙상블 결과와 맞지 않을 때 확 인했으나, 단일 모델 단계에서 적용하거나 파라미터를 더욱 세밀하게 튜닝할 시간적 여유가 부족했음
- 변인 통제의 한계 : 모든 실험을 동일한 V100 GPU 환경에서 수행했음에도, CUDA 및 cuDNN 연산의 비결정성과 학습·검증 과정에서의 전역 RNG(Random Number Generator) 상태 공유로 인해 완전한 재현성 확보에는 한계가 있었다. 특히 Validation 단계에서의 난수 소모가 이후 학습 데이터 증강 및 배치 순서에 영향을 주어, 동일한 설정에서도 최종 성능이 달라지는 현상을 경험했다.

4. 프로젝트를 통해 얻은 교훈 및 다음 목표

- 향후에는 Docker 기반으로 실험 환경을 고정하고, RNG 격리와 검증 프로토콜을 명문화하여 보다 안정적인 실험 재현성을 확보하는 것을 목표로 한다.

▼ 이봉학

1. 프로젝트 목표 및 역할

- 개인 학습 목표:
 - Segmentation에 대한 기초 습득 및 이해
 - 깃허브 등 협업툴 사용하는 것에 익숙해지고 숙련도 올리기

- 주요 기여 역할

- Crop, Flip 등 전처리 테스트, 일부 모델 학습, Window Sliding 적용 및 테스트

2. 목표 달성을 위한 과정 (나의 핵심 활동)

- 모델 학습 보다는 데이터와 직접적인 이미지 처리쪽에 집중하여 손 정렬, 크롭, Flip을 통한 한손 학습 등을 적용했으나 모두 성능 개선에는 실패하면서 유의미한 결과를 내지 못함
- 초반 노션 정리를 통한 프로젝트 진행 기반 구축

3. 한계점 및 아쉬운 점

- 기본 베이스라인 자체가 성능이 매우 높게 나와서 어떤 방법을 적용했을 때 성능 개선이 미미했던 부분들이 많았고 데이터에서 얻은 인사이트를 기반으로 세운 가설들을 검증하는 일련의 과정보다는 해보고 좋으면 어떤 것 때문에 좋은 것 같다는 추측성으로 맞추는 것 같아 아쉬웠다
- 코드 작성 및 구현을 AI에게 일임, 직접 구현해보는 부분이 없었음. 데이터의 입출력 형태 뿐 아니라 어떤 기능, 동작이 어느 부분에서 동작하는지, 이미지를 처리하는 알고리즘과 OpenCV 코드들을 이해하고 쓰기보단 AI의 작업을 따라가기만 했다
- 깃허브를 쓰긴 했지만 코드 공유하는 정도만 사용되고 커밋 메세지나 이슈 등 좀 더 상황을 공유하거나 활용도가 적고 아쉬웠다. 커밋 메세지도 정해진 규칙이 없어서 여러 단어들로 혼용되었다.
- 각자 진행하면서 마주친 여러 문제점, 어려웠던 점이나 해결하지 못한 점들을 모아 보면 많을텐데 이런 부분에 대해 어떤 문제를 어떻게 해결이나 왜 그렇게 생각해서 시도했는지 등에 대한 문서화를 하고 같이 공유했으면 팀 전체에게 문제 상황을 더 잘 공유할 수 있을 것 같다
- 진행하면서 기록들이 많아지다보니 어디에 어떤 내용이 있었는지 찾고 공유하는 것도 딜레이가 조금씩 생기면서 인덱싱, 태그 등을 잘 해서 팀원 모두가 빠르게 필요한 정보를 찾을 수 있게끔 하면 더 좋을 것 같다

4. 프로젝트를 통해 얻은 교훈 및 다음 목표

- 초반에 협업툴과 프로젝트 룰, 규칙들을 잘 정리해두면 진행하는 동안 편의성이 많이 좋아짐을 느꼈다
- 도메인 지식도 부족하지만 딥러닝 자체가 블랙 박스다 보니 어떤 것에 근거하여 이게 좋을 것이다 보다는 이걸 해봤더니 성능이 좋았다 경험적인 부분이 강하다. 그런

만큼 모듈화와 실험을 바로바로 돌릴 수 있게 코드를 잘 정리하는 것이 정말 중요하다

▼ 조수빈

1. 프로젝트 목표 및 역할

- 개인 학습 목표 :
 - 손 뼈 X-ray, 픽셀 이미지에 대한 이해 : 프로젝트에서 다뤄야 하는 데이터에 대해 제대로 파악해두어야 실험 방향을 바르게 잡을 수 있고, 더 실험할 가치가 있는 가설을 수립할 수 있다고 생각한다.
 - Github 활용 : 협업에 있어 중요한 것은 정보 공유이며 그 정보에는 코드 역시 포함되어 있다. 동일한 코드를 사용하는 것, 즉 같은 실험 환경이 조성되어야 소통이 원활하게 이루어질 수 있고 성능 비교에 용이하다. 따라서 본 프로젝트에서는 Github를 적극 활용하는 것을 목표로 삼았다.
- 주요 기여 역할
 - 데이터 탐색, 전처리 : 데이터의 여러 분포에 대해 그래프를 추출하여 한 눈에 파악할 수 있게 하였다. 데이터를 가공하는 과정에서 모델이 어떤 입력값을 받게 되는지 확인할 수 있게 가공된 이미지를 시각화하였다. 입력에 넣었을 때 모델이 더 좋은 성능을 낼 수 있게 데이터를 가공하였다.

2. 목표 달성을 위한 과정 (나의 핵심 활동)

- 기술적 시도 및 성과 :
 - 이미지 증강을 위해 전처리 및 증강의 강도를 강하게 주었을 때의 정보 손실이 증강을 주어서 얻는 이득보다 더 크다는 것을 알 수 있었다. 픽셀 단위이기 때문에 사람이 육안으로 확인할 수 없는 미세한 차이가 성능에 영향을 준다는 것을 파악했다. Segmentation의 면적을 이루는 최소 단위들은 픽셀이라는 점을 충분히 이해하고 실험을 진행하여, 성능이 향상되는 정도의 설정값을 얻을 수 있었다.
 - 손의 해부학적 구조를 공부해서 구조적으로 붙어있지 않는 클래스들이 EDA에서 겹쳐져 있다는 결과를 이상치가 있다는 것으로 해석할 수 있었다. 익숙하지 않은 도메인, 데이터였기 때문에 익숙해지기 만들려고 주어진 이미지들을 계속 육안으로 확인하였는데, 덕분에 데이터의 어떤 부분이 문제인지 짐작하고 가설을 세워 개선하는 방향을 제시할 수 있었다고 생각한다.

3. 한계점 및 아쉬운 점

- 초기 EDA에서 더 심도 있는 분석을 진행하지 못한 것 같다. 실험이 본격적으로 진행되기 전에 염두해야 할 부분—이상치, 데이터 분포 등—to 더 자세히 파악하고 초기에 처리한 뒤 사항을 공유했다면 성능과 실험 관리 측면에서 훨씬 좋았을 것 같다.
- 주어진 학습 데이터는 800장으로 많지 않은 수량인데, 이상치가 발견된 데이터를 재라벨링 해서 되도록 학습에 쓰고 싶었으나 실행에 어려움이 있어 이를 처리하는 방법이 제거하는 것에 그쳐 아쉽다. 종자골의 경우도 라벨링 되어있지 않은 것들은 추가로 마스킹을 해 라벨링 일관성을 더 높여보고 싶었는데 아쉬움이 남는다.
- 가설에 대한 검증을 수치로 증명하는 것에 어려움이 있었다. 가령, CLAHE를 적용했을 때 대비 제한을 높게 주면 성능이 떨어지고, 대비를 낮추니 성능이 올랐었는데 이에 대한 결과 분석을 추측으로만 남겨두어 아쉽다. 더 확실하게 원인을 규명해보고 싶다.

4. 프로젝트를 통해 얻은 교훈 및 다음 목표

- 이번 프로젝트는 이전보다 더 효율적이고 체계적으로 진행되었다. 다양한 협업 툴 활용, 변인 통제, 역할 분담 등 부족하다 생각했던 부분을 개선하여 안정적인 협업이 가능했다.
- 코드를 수정하는 경우 Github에 기록이 남지만, 이를 하나하나 파악하고 전달하는 것에 번거로움이 있었다. 변경 사항을 기록하는 문서를 하나 만들고, 공지 게시판처럼 활용할 수 있다면 좋을 것 같다. 또 실험 관리를 기록하는 것에 있어, 팀원이 문서를 확인했을 때 필요한 정보를 쉽게 파악할 수 있도록 기록 양식을 어느 정도 구조화, 정형화 하는 것이 필요하다.