

# Hand Bone Image Segmentation



1. Introduction
2. Collaboration Tools
3. EDA
4. Hypothesis & Experiment
5. Result
6. Review

# 4. Hypothesis & Experiment

4.1 최종 모델 도식화

4.2 학습 전략

4.3 전처리 및 증강

4.4 모델링

4.5 추론 전략 및 양상들

4.6 후처리

# Introduction

## 1 Introduction



# 1 Introduction

## Media in category "X-rays of the hands"

The following 146 files are in this category, out of 146 total.



[01 16 X-ray of Hand \(cropped\).jpg](#)  
469 × 469; 110 KB



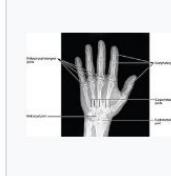
[01 16 X-ray of Hand.jpg](#)  
750 × 881; 207 KB



[1896 x-rays of Physikalischer Verein, hand of man.tif](#)  
1,896 × 2,574; 6.53 MB



[1896 x-rays of Physikalischer Verein, hand of woman.tif](#)  
1,867 × 2,541; 6.31 MB



[814 Radiograph of Hand.jpg](#)  
968 × 639; 186 KB



[American quarterly of roentgenology \(1909\) \(14570808208\).jpg](#)  
2,646 × 4,136; 645 KB



[American quarterly of roentgenology \(1911\) \(14756932095\).jpg](#)  
2,068 × 1,610; 1.44 MB



[American X-ray journal \(1897\) \(14570496517\).jpg](#)  
893 × 1,297; 147 KB



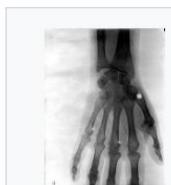
[Bender gruppe-02.jpg](#)  
1,066 × 542; 210 KB



[Bones of a hand. Radiograph, 1900-1904. Wellcome L0026317.jpg](#)  
1,350 × 1,386; 487 KB



[Bones of the hand of Sergeant Mather, possibly of 2nd Middle Wellcome L0026316.jpg](#)  
1,184 × 1,489; 542 KB



[Bones of the hand of W. Scott-Moncrieff, showing the effect Wellcome L0026315.jpg](#)  
1,184 × 1,548; 604 KB



[Bouchard nodes Armenian.png](#)  
600 × 359; 288 KB

[Buckwalter X-Ray Hand.jpg](#)  
1,413 × 1,129; 153 KB

# 1 Introduction

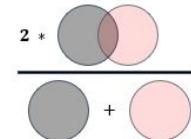
- 크게 손가락 / 손등 / 팔로 구성되며, 총 29개의 뼈 종류가 존재



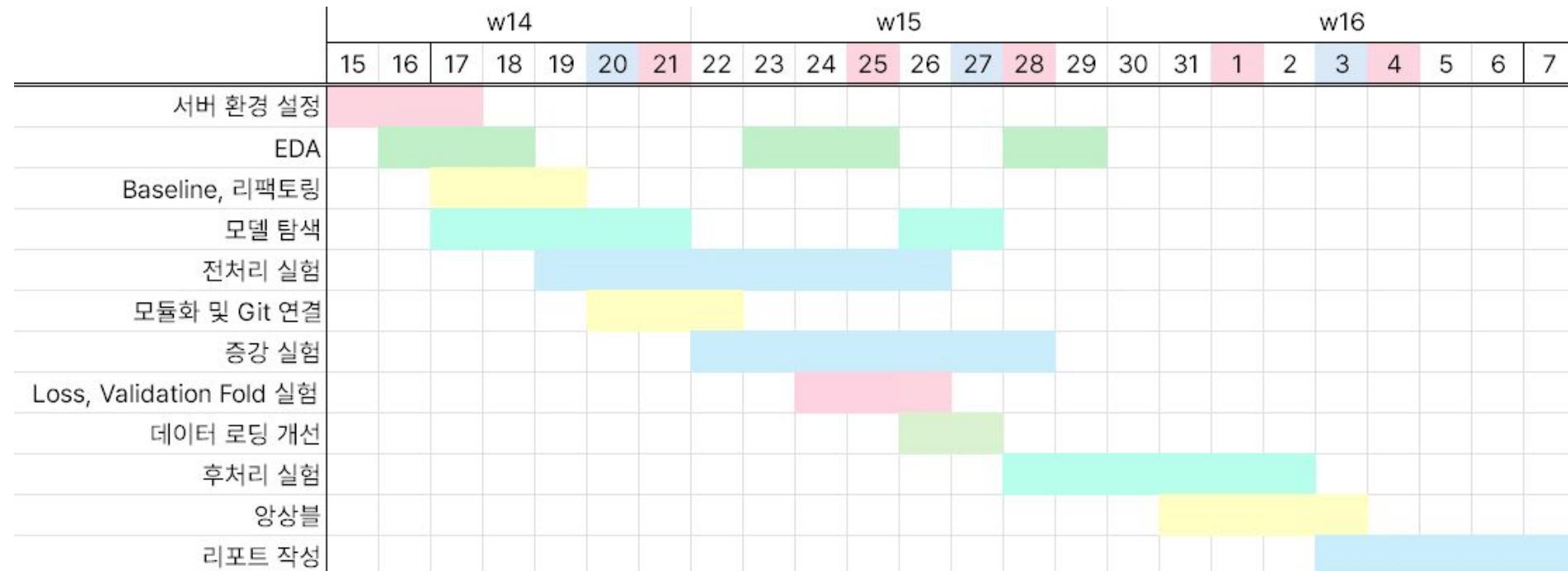
[ 'f1', 'f2', 'f3', 'f4', 'f5', 'f6', 'f7', 'f8', 'f9', 'f10',  
'f11', 'f12', 'f13', 'f14', 'f15', 'f16', 'f17', 'f18', 'f19',  
'Trapezium', 'Trapezoid', 'Capitate', 'Hamate', 'Scaphoid',  
'Lunate', 'Triquetrum', 'Pisiform',  
'Radius', 'Ulna' ]



$$Dice = \frac{2 * |A \cap B|}{|A| + |B|}$$



# 1 Introduction



# Collaboration Tools

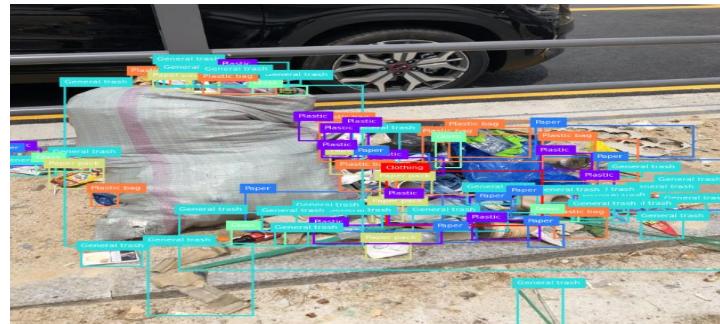
## 2 Collaboration Tools



이전에 진행했던 Object Detection 회고

1. 역할을 나누자
  2. 깃허브를 활용해보자
  3. 실험 관리를 더 철저히 하자

\*특히 벤인통제



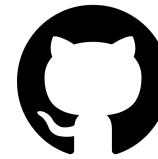
## 2 Collaboration Tools

Notion



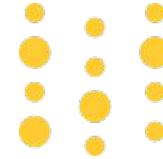
전반적인 프로젝트  
진행사항 관리  
및 기록

Github



저장소  
파일 관리

WandB



실험 기록  
및 관리

Slack



소통  
회의록

# 2 Collaboration Tools

Notion



전반적인 프로젝트  
진행사항 관리  
및 기록

TO DO

Project name	Status	Owner	Dates	Completion	...
EDA	In Progress	루미 두	2025/12/16 → 2025/12/17	100.00%	
모델화	In Progress	위정호	2025/12/16		
model baseline	In Progress	장동호 / 학생	2025/12/17 → 2025/12/18	97.73%	
전처리, 중간 가설 및 실험	In Progress	루미 두	2025/12/18 → 2025/12/20	86.96%	

Score

Model	제작명	Person
ensemble	WIH_096_ensemble_sliding_tta	위정호
ensemble	WIH_098_ensemble_sliding_GaussianOptThr	위정호
ensemble	WIH_055_ensemble_03	위정호
ensemble	WIH_099_ensemble_sliding_GaussianOptThr	위정호
ensemble	WIH_049_ens_Smooths_global_0.61-0.29-0.09	위정호
HRNet_w48_exclude_sliding	WIH_037_hnet_w48_1024_focal_dice_sw	위정호
HRNet_w48_exclude_sliding	WIH_038_hnet_w48_1024_focal_dice_sw_denoise	위정호
ensemble	WIH_046_ens_Smooths_class_matrix_3	위정호
ensemble	WIH_067_ensemble_sliding_tta_Hflip	위정호
ensemble	WIH_045_ens_Smooths_class_matrix_2	위정호
class-wise ensemble	CSB_019_cwe_WIH_038_CS8_015	cho

Tasks

Project	Status	Assignee	Val Score	Img_size	Loss	# epoch
EDA	Done	cho				
EDA 2 - 이전 가수	Done	cho				
이상치 - 문자열 외	Done	cho				

COUNT 3 COMPLETE 3/3

Task name	Status	Val Score	Assignee	Img_size	Loss	# epoch				
fcn_resnet50	Done		cho	S12	BCE	82				
HRNet_OCR	Done				BCE	DICE	73			
UserNet_swin_l	Done				BCE	DICE	73			
rn_UNetv2	Done				BCE	DICE	73			
UNet_efficientnet-b4	Done			S12	BCE	DICE	12			
UNet+_efficientnet-b4	In Progress	J Kim		S12	1024	2048	BCE	DICE	7.9	50
segFormer_B3(imagenet)	Done	J Kim		S12			BCE	DICE	55	15
DeepLabV3+	Done			S12			BCE	DICE	55	50
UNet_efficientnet-b0	Done			S12			BCE	DICE	55	50
UNet_efficientnet-b2	Done			S12			BCE	DICE	55	80
UNet_efficientnet-b2	Done			S12			BCE	Focal	55	60
UNet_efficientnet-b2	Done			S12			BCE	DICE	92	20
UNet_efficientnet-b2	Done			S12			BCE	DICE	82	25

서버 세팅 5

Task name	Status	Val Score	Assignee
베이스라인 동작 확인	Done		
Git 연결	Done		
tmux 사용 (오프라인으로 실행하기)	Done		
TTA	In Progress		

+ New task

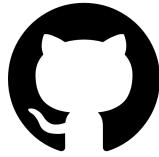
COUNT 5 COMPLETE 4/5

전처리, 중간 가설 및 실험 15

Task name	Status	Val Score	Assignee
One hand training (right-hand)	Done		루 루미 두
손 뻐 정렬	Failed		루 루미 두
Center Crop	Done		루 루미 두
전처리 비교 BASE 성능	Done		위정호
multi-label	Done		J Kim
손목뼈-수근골...	Archived		cho
색(톤, 명암)	Done		J Kim
중간 비교 BASE 성능	Done		위정호
multi-channel feature engineering	Done		J Kim
Validation set 설정	Done		위정호
LOSS 별 성능	Done		kim kyoungmu
후처리	Done		cho
Sliding Window	In Progress		루 루미 두
Validation set으로 학습	Failed		kim kyoungmu
class-wise ensemble	In Progress		cho

# 2 Collaboration Tools

Github



저장소  
파일 관리

The screenshot shows a GitHub desktop application interface. On the left, a file tree lists various project components: bi\_map, checkpoints/Base\_UNet, dataset, eda, inference, model, tools, .gitignore, README.md, and config.py. To the right of the file tree is a detailed commit history. The commits are color-coded by author and include:

- Add LBH\_010\_N (blue dot)
- feat: add overlap analysis (pink dot)
- feat: add binary map data (pink dot)
- feat: add binary map creation (pink dot)
- Update inference scripts and data (green dot)
- feat: add multi-GPU support (pink dot)
- Fix: NameError in inference (pink dot)
- Refactor: Robust Ensemble (pink dot)
- del (blue dot)
- Merge branch 'subin' of git (green dot)
- add .gitignore (pink dot)
- Merge pull request #6 from subin (green dot)
- Remove utility scripts (cheat sheet) (orange dot)

At the bottom of the commit history, there are sections for 'SOURCE CONTROL' (with a '3' badge), 'REPOSITORIES' (listing 'work' and 'ss' repositories), and 'CHANGES' (listing the same two repositories). A blue badge at the bottom center indicates 14 unread notifications.

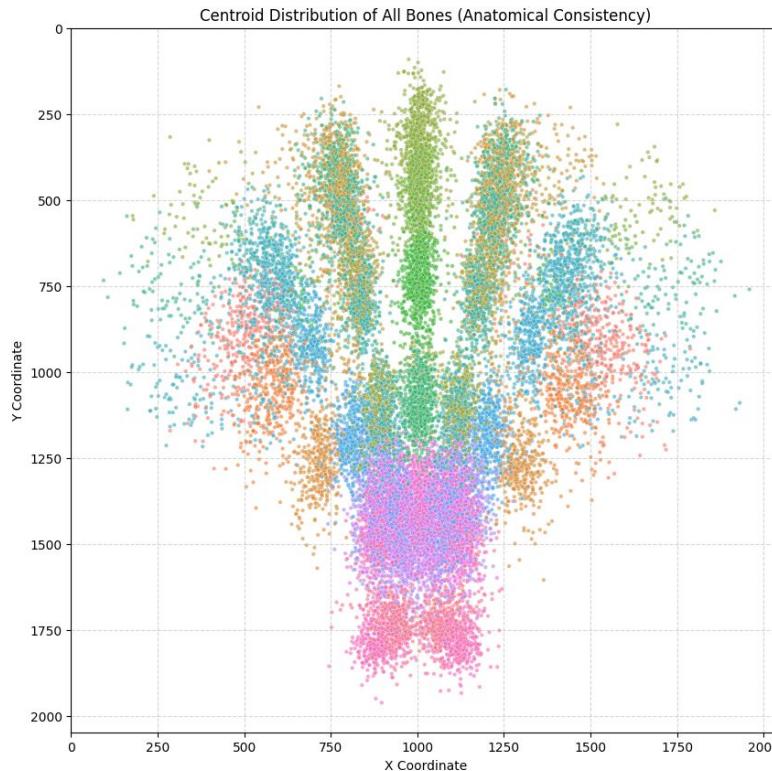
# EDA

### 3 EDA



1. 손 뼈 : 정형화된 구조
2. 흑백 이미지 : 0-255
3. 2048\*2048px
4. 한 사람이 양손 촬영
  - a. 쌍끼리 유사
5. 손가락, 팔 뼈들은 길쭉
6. 손목 뼈들은 동글

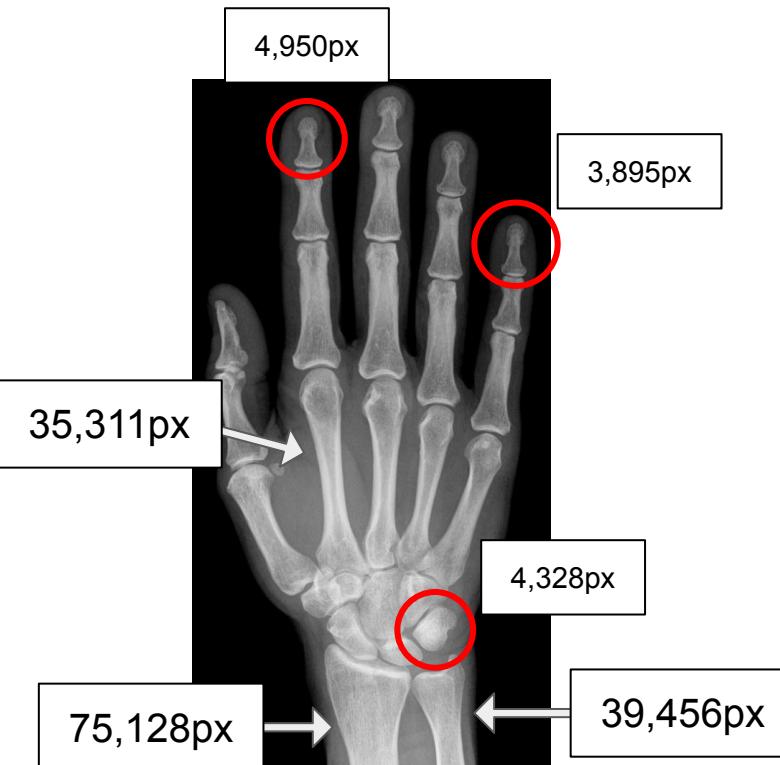
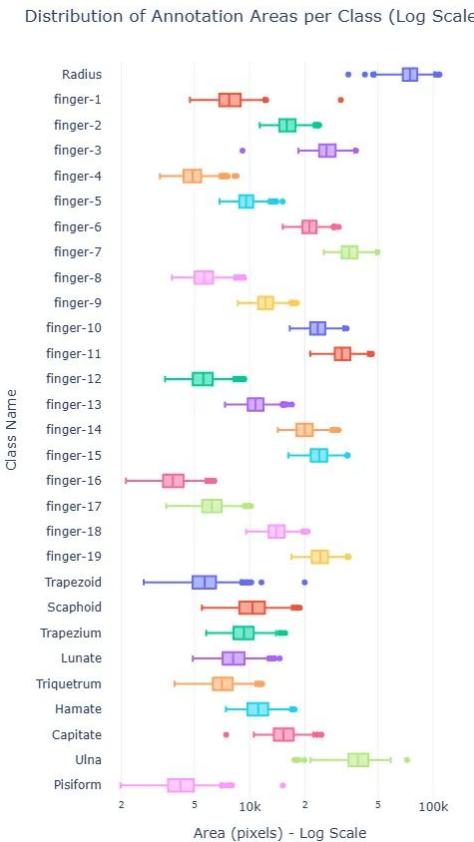
### 3 EDA



면적 중심에 점을 찍고 그 분포를 봤을 때,  
손목을 꺾어서 촬영한 이미지가 있음을 파악

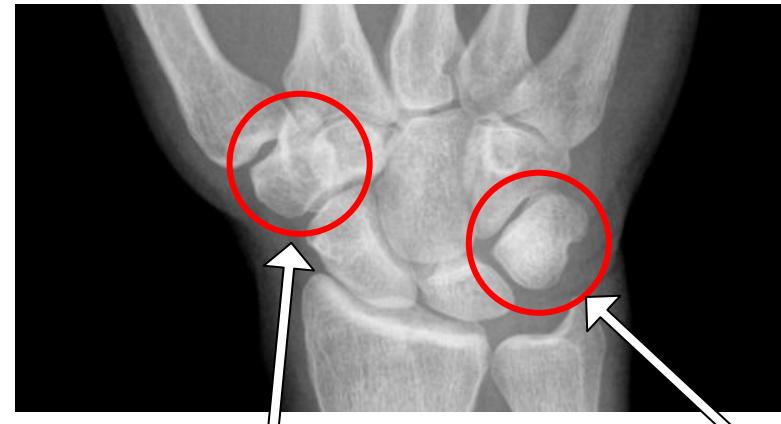
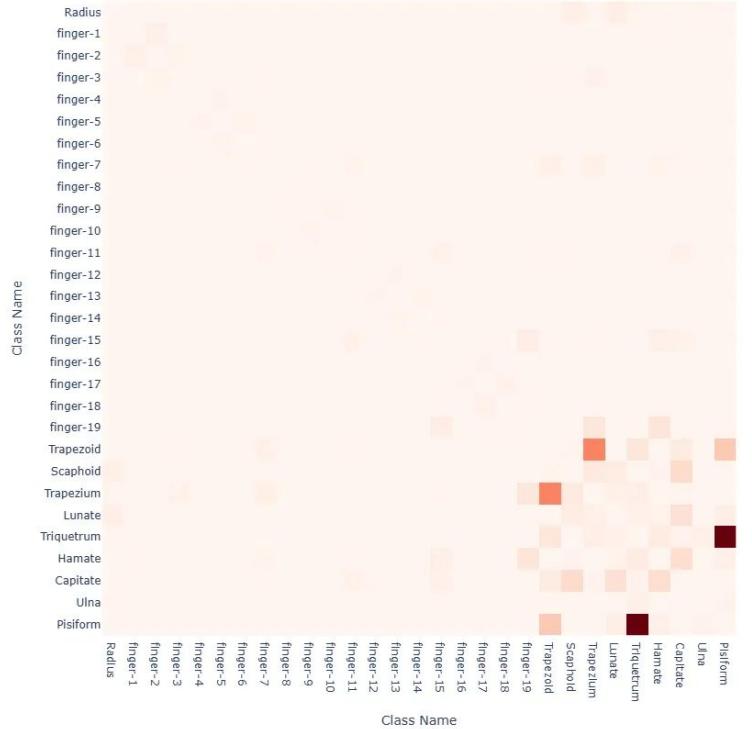
같은 픽셀 좌표에서  
한 이미지는 finger-1가 위치  
다른 이미지는 finger-4가 위치에 있을 수  
있음

### 3 EDA



# 3 EDA

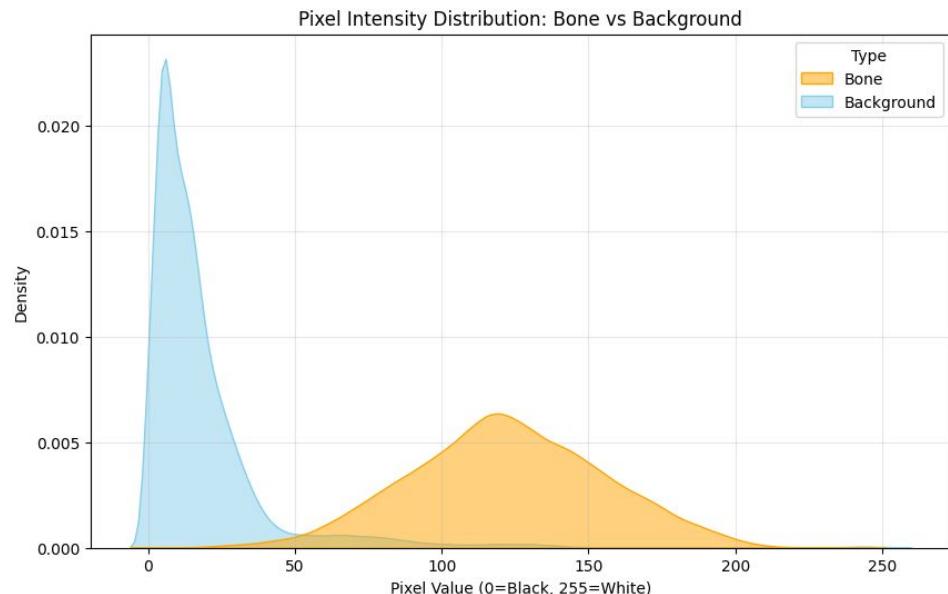
Average IoU Heatmap (Severity of Overlap)



Trapezium  
& Trapezoid

Triquetrum  
& Pisiform

### 3 EDA



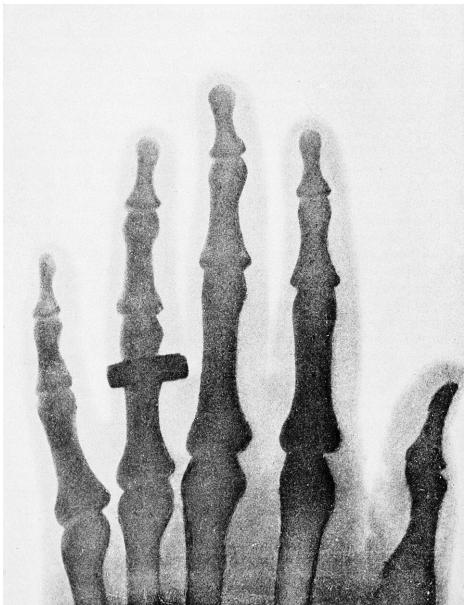
뼈 - 배경 픽셀의 화소값 비교

분포의 구분이 뚜렷한 편이나  
화소값이 비슷한 일정 구간이 존재

### 3 EDA

장신구

- Annotation에 포함

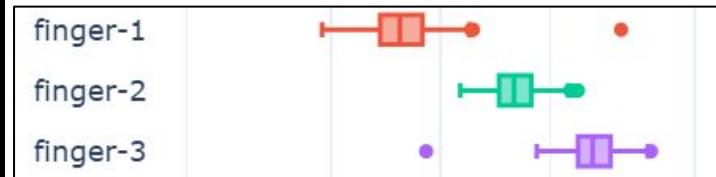


네일아트

- 화소값 높게 나옴



수상하게 크거나 작은 이상치



== Top 10 Most Severely Overlapping Pairs (High IoU)			
	Class A	Class B	Frequency Avg IoU
64	Triquetrum	Pisiform	799 0.448332
45	Trapezoid	Trapezium	800 0.188486
<b>48</b>	<b>Trapezoid</b>	<b>Pisiform</b>	<b>1 0.090350</b>
59	Solidus	Scaphoid	799 0.058616

구조상 겹쳐있지 않을 Class 쌍

\* 실제 데이터가 아닌 발표 이해를 위해 임의로 선정, 가공된 사진입니다

### 3 EDA

## 요약

- 픽셀 좌표 기준 클래스 분포 차이
  - 각도의 다양성을 부여하는 증강 필요
- 클래스 간 면적 불균형
  - Multi-Scale 학습 필요
- 픽셀 단위 Multi-label 중첩
  - Sigmoid 기반 학습 필요
- 모호한 경계
  - 대비 강화
- 모델 학습에 혼동을 줄 수 있는 이상치
  - 선별적 정제 필요

# Hypothesis & Experiment

4.1 최종 모델 도식화

4.2 학습 전략

4.3 전처리 및 증강

4.4 모델링

4.5 추론 전략 및 앙상블

4.6 후처리

## 4 Hypothesis & Experiment - 4.1 최종 모델 도식화

### Sliding Window

- Window 1024, Stride 512 → 3\*3 Patch



### Preprocessing

- Data Cleansing
- CLAHE
- SSR
- Normalization

HRNet

nnUNet

DeepLabV3

### Inference

- Gaussian Blending
- TTA

Final Segmentation Output  
Public : 0.9758  
Private : 0.9767 (+0.0009)

Soft Voting Ensemble  
- Uniform :  $1/N$   
가중치

## 4 Hypothesis & Experiment - 4.2 학습 전략

### 데이터 병목 개선 및 로딩 최적화

- 베이스라인 기준 1 epoch 당 3분 정도 소요
- 이미지 전처리 과정이 CPU 상에서 이뤄지면서 병목 발생 확인
  - GPU 사용률 0~100% 변동이 반복
- NVIDIA DALI 라이브러리 적용
  - 1 epoch 당 3분에서 20초로 단축
  - 더 많은 모델을 빠르게 학습시켜 다양한 실험 진행

## 4 Hypothesis & Experiment - 4.2 학습 전략

### Loss

- Segmentation에서 사용되는 Loss 함수들 탐색 및 실험
  - 실험 모델 : UNet-EfficientNetb2 (Baseline)
  - Dice가 가장 높지만, Dice + BCE, Dice + Focal과 큰 차이 없음

Loss	Mean Dice
Dice	<b>0.9638</b>
Dice + BCE	0.9634
Dice + Focal	0.9630
Dice + Lovasz	0.9255
BCE	0.9148

## 4 Hypothesis & Experiment - 4.2 학습 전략

### Validation Fold

- Validation Score와 LB Score 점수 간의 차이를 최소화하기 위해 Validation set 선정 과정을 거침
- 5 Fold Cross Validation을 수행하고 각 Fold별 결과를 모두 제출하여 Validation Score와 LB Score의 차이가 가장 적은 Fold를 Validation set으로 선정

Fold	Mean 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	<b>0.9543</b>	<b>0.9423</b>	<b>0.0120</b>

## 4 Hypothesis & Experiment - 4.2 학습 전략

### Resolution & OOM

- 이미지 크기 512 → 1024 변경 시 성능 향상 확인
- 2048도 괜찮을까? → OOM 발생 → 쪼개자!

Model	Resolution	LB Score
DeepLabV3+	512	0.9426
<b>DeepLabV3+</b>	<b>1024</b>	<b>0.9681</b>
Unet-EfficientNet-B2	512	0.9490
<b>Unet-EfficientNet-B2</b>	<b>1024</b>	<b>0.9618</b>
nnUnet	512	0.9563
<b>nnUnet</b>	<b>1024</b>	<b>0.9689</b>

## 4 Hypothesis & Experiment - 4.2 학습 전략

### Sliding Window

- 원본 이미지를 고정 크기 패치로 분할해 각 패치를 입력으로 활용
  - 정보 손실 없이 학습에 사용



## 4 Hypothesis & Experiment - 4.2 학습 전략

### Sliding Window

- 실험 결과

model	Window Size	Stride	No Sliding LB Score	LB Score
DeepLabv3	1024	512	0.9681	0.9726
DeepLabv3	1024	1024	0.9681	0.9702
UNet-EfficientNet-b2	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

## 4 Hypothesis & Experiment - 4.3 전처리 및 증강

### 이상치 제거

- Annotation에 이상이 있었던 4건의 데이터를 학습에서 제외
  - 1건 - 반지가 함께 라벨링된 경우
  - 3건 - 라벨링 오류
- Annotation에 이상은 없으나, 네일아트, 보철물 등 화소값에 영향을 주는 데이터들은 모델이 어려운 데이터도 학습할 수 있도록 처리하지 않음

## 4 Hypothesis & Experiment - 4.3 전처리 및 증강

### CLAHE Contrast Limited Adaptive Histogram Equalization

- X-ray 이미지에서 나타나는 뼈의 모호한 경계선을 명확히 하기 위해 대비를 제한적으로 강화

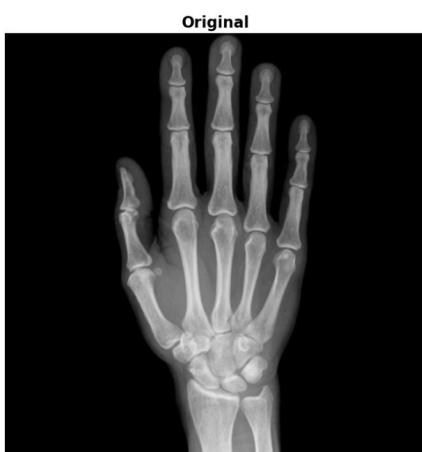


-	Base	CLAHE 4.0	CLAHE 2.0
<b>Val Mean Dice</b>	0.9512	0.9488	0.9567
<b>LB Public Score</b>	0.9364	0.93	0.9416

## 4 Hypothesis & Experiment - 4.3 전처리 및 증강

### SSR Shift, Scale, Rotate

- 촬영 시 발생할 수 있는 미세한 위치 및 각도 변화에 대한 강건성 확보
- 이미지가 잘리는 손실을 방지하고자 SSR 강도를 약하게 설정



-	Base (512)	SSR (512)	Base (1024)	SSR (1024)
Val Mean Dice	0.9569	0.9450	0.9656	<b>0.9677</b>
LB Public Score	0.9420	0.9342	0.9618	<b>0.9643</b>

## 4 Hypothesis & Experiment - 4.3 전처리 및 증강

### Normalize

- EDA에서 확인한 뼈와 배경 간 픽셀 분포 특성을 바탕으로 두 가지 정규화를 비교 및 실험
  - ImageNet 통계치 기반 정규화
  - 이미지별 Percentile Clipping + Z-score

방법	설명	Dice Score
ImageNet 통계치 기반 정규화	mean=[0.485], std=[0.229] 적용	<b>0.9364</b>
이미지별 Percentile Clipping + Z-score	0.5~99.5% 클리핑 후 이미지별 평균/표준편차로 정규화	0.9337

## 4 Hypothesis & Experiment - 4.4 모델링

### 모델링

- nnU-Net
  - CNN 기반 Encoder-Decoder 구조, 고해상도 의료 영상에서 검증된 프레임워크
- HRNet
  - CNN 기반다중해상도 병렬구조로, 고해상도 Feature Map을 전 과정에서 유지
- DeepLabV3+
  - CNN Encoder + ASPP + Decoder 구조, 멀티스케일 특징 추출을 수행
- SegFormer
  - Transformer Encoder + MLP Decoder 구조, Global Context
- UNet + MiT
  - Transformer Encoder(MiT-b3) + CNN Decoder(UNet) 구조의 하이브리드 아키텍처

## 4 Hypothesis & Experiment - 4.5 추론 전략 및 앙상블

### Gaussian Blending

- Sliding Window 추론 과정에서 발생하는 패치 간 경계 불연속성(Edge Artifacts)을 최소화
- 단순 평균 대비 가우시안 가중 평균 적용시 성능 향상을 보임

## 4 Hypothesis & Experiment - 4.5 추론 전략 및 앙상블

### TTA

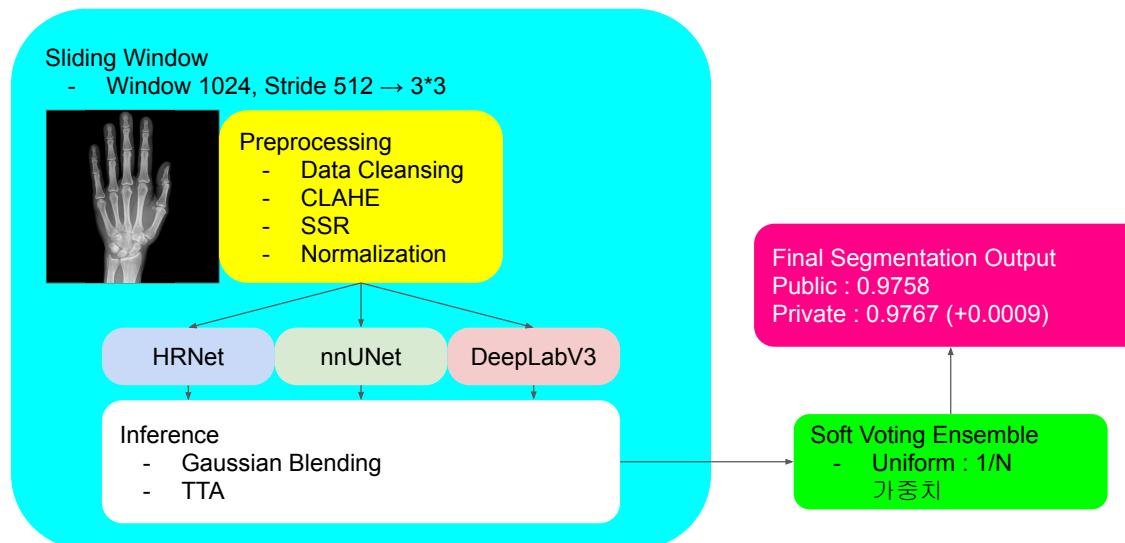
- 각 이미지의 크기, 형태에 따라 모델이 특정 뼈를 잘 예측할 수 있을 것이라 기대
- 멀티 스케일로 된 이미지를 추론하고, 평균 내어 최종 예측을 생성  
(img\_size = [0.8, 1.0, 1.5])

model	TTA	LB Score
DeepLabV3+	-	0.9681
<b>DeepLabV3+</b>	<b>TTA</b>	<b>0.9695(+0.0014)</b>
Segformerb3	-	0.9602
<b>Segformerb3</b>	<b>TTA</b>	<b>0.9624(+0.0022)</b>

## 4 Hypothesis & Experiment - 4.5 추론 전략 및 앙상블

### 앙상블 Soft Voting

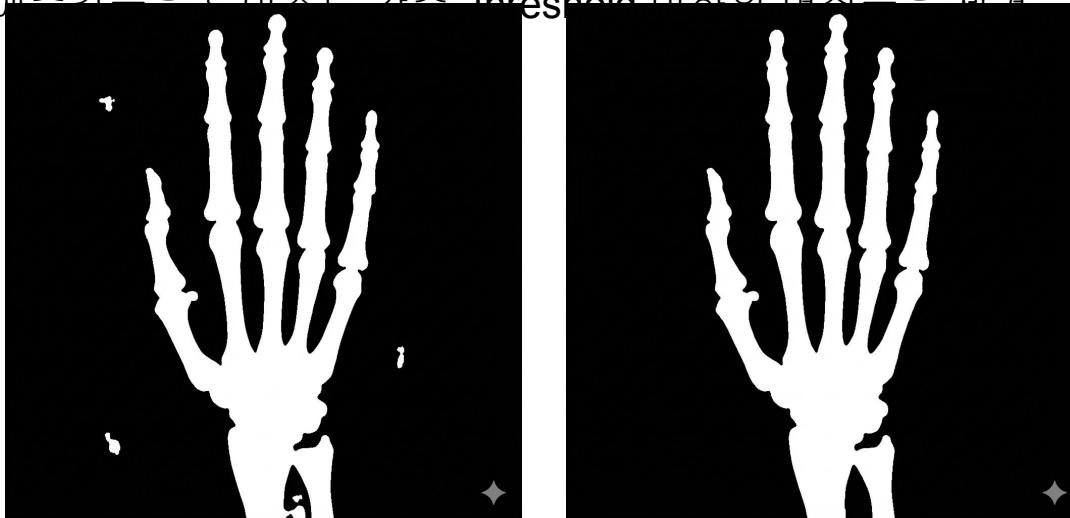
- 모델의 출력을 확률값 그대로 사용하여 평균을 내는 방식
- 모든 모델에 동일한 가중치( $1/N$ )를 부여하는 Uniform 방식 사용



## 4 Hypothesis & Experiment - 4.6 후처리

### 소형 객체 제거

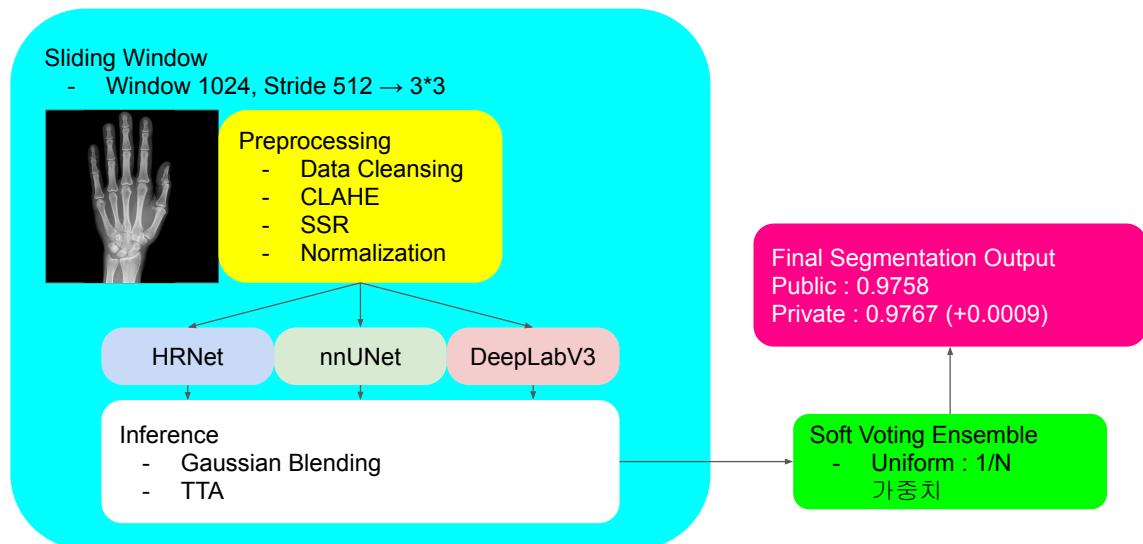
- 추론 결과를 시각화하여 확인했을 때 미세하게 작은 픽셀들이 뒤어있음을 발견
- 너무 작은 예측 가드온이 이므로 가즈 threshold 이하의 면적은 제거



\* 실제 데이터가 아닌 발표 이해를 위해 임의로 선정, 가공된 사진입니다

# 4 Hypothesis & Experiment +

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



# Result

## 5 Result

### 최종 결과

- 0.9758 → 0.9767
- 0.0009 상승

My Rank 1	CV-01		0.9758	137	5d
My Rank 4	CV-01		0.9767	137	5d

# Review

# 6 Review

## 대회 리뷰

- 잘한 점
  - EDA 기반 실험 설계로 전처리·증강·모델·후처리를 체계적으로 비교
  - Sliding Window, TTA, Ensemble 등 실전적인 추론 전략 적용
  - Notion·GitHub 활용으로 협업 및 실험 관리 효율 개선
- 아쉬운 점
  - 라이브러리 의존으로 아키텍처 수준의 근본 원인 분석 부족
  - 일부 실험이 가설 검증보다 경험적 해석에 머무름
  - 초기 라벨 품질 개선에 충분한 시간 투입 부족

감사합니다