솔루션 설명

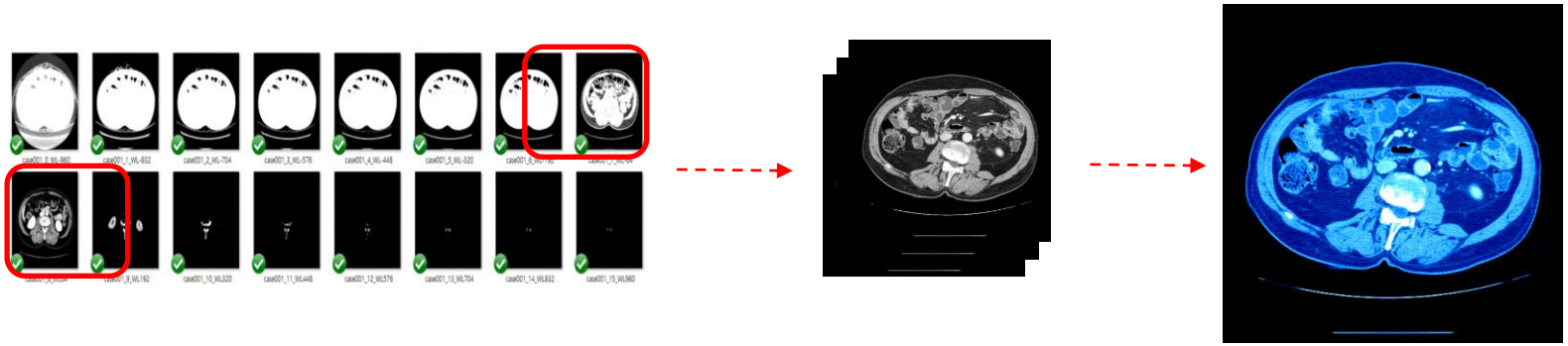
팀: HallymMMC

한림대학교: 허종욱, 신유승, 심보석
KAIST: 안원혁, 강지현

Body morphometry for sarcopenia
MOAI 2020 Body Morphometry AI Segmentation Online Challenge

데이터 전처리 및 데이터셋 구성

- **DICOM 원 데이터의 픽셀 분포를 확인 (좌)**
  - -256~256 사이에 주로 의미있는 값이 있는것으로 확인
- **WL: -128, 0, 128, WW: 256을 이용하여 .png 3장 추출**
  - 3장의 추출 이미지는 color image의 각 R/G/B 채널로 이동 후 저장 (우)



  - **(규정에 따라) TEST셋의 Psuedo-label 사용하지 않음**
  - 외부 데이터 사용하지 않음

학습 방법

■ 이용 모델 및 기타사항

- AlbUNet: U-Net with ResNet encoder
  - RTX 2070S 2대 사용, Pretrained 이용
- 이미지 사이즈는 1024로 학습 후 제출 전 512로 축소
  - 데이터는 10-fold cross validation

■ 실험 설정: 2종류 이용

- 1번: albunet1024fold10: 강한 image augmentation
- 2번: albunet1024fold10albJW: 보통수준의 image augmentation

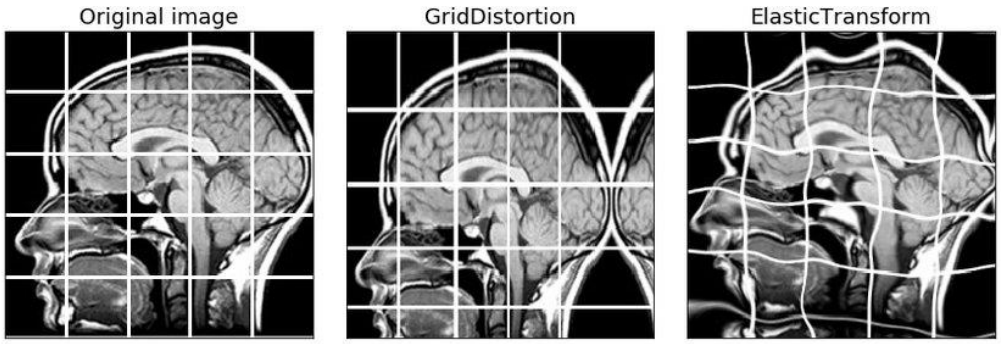
■ 학습 단계:

- 1단계: 비교적 큰 LR = 0.0001 (50epoch)
- 2단계:
  - 1단계의 best 모델 선택
  - 작은 LR = 0.00001 (50epoch, early stopping = 12)

데이터 augmentation

- **train_transforms_complex_1024.json (강한 데이터 증강)**
- **train_transforms_hjw_1024.json (보통 데이터 증강)**

- **사용한 데이터 transform 목록**
  - HorizontalFlip
  - RandomContrast, RandomGamma, RandomBrightness
  - ElasticTransform, GridDistortion, OpticalDistortion
  - ShiftScaleRotate, Resize, Normalize



평가 및 submission 생성

Inference and ensemble

- 2단계의 각 fold별 상위 3개의 결과를 평균(총 10종류의 fold * 3)
- 학습 종류(2종)별 앙상블
- Test time augmentation: Rotation, Scaling 사용 (최종 결과엔 미적용)
  - 총 2 * 10 * 3 = 60개의 weights를 사용

앙상블 후 후처리

RLE 및 CSV 파일 작성 후 제출

앙상블 후 후처리 세부 사항

- **MIN_CONTOUR_AREA 수치로 작은 영역 필터링**
  - 작은 영역 선택 후, 다른 클래스 영역으로 segment 이동 (=1300)
- **4개의 class중 인접한 영역의 확률로 segment 이동 설정**
  - step1) Small segmen에 팽창 연산을 적용 후, 타 채널과 중첩되는 부분의 확률 값을 모두 합하여 score로 저장.
  - step2) score가 최대로 나온 class index로 segment를 이동

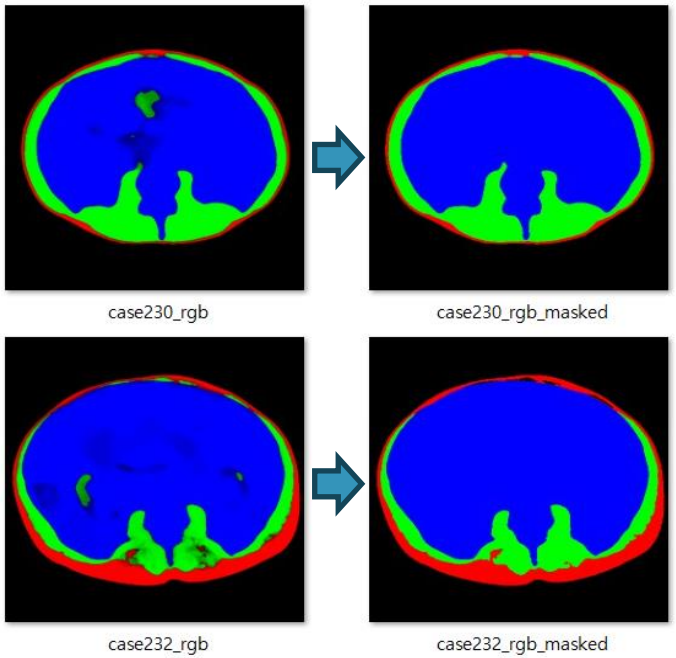
```
# 작은영역 boolean mask
mask_small_dilate = (mask_small_dilate>0)

# 큰 영역의 확률맵
prob_large_contour = mask * mask_larges_

# 확률이 제일 높은 영역의 채널 번호를 가져온다.
score = []
for c in range(0, num_class):
    score.append(((prob_large_contour[c, :, :] * mask_small_dilate).sum()))

# 최고 점수가 나온 채널 획득
idx = np.argwhere(score == np.amax(score))

# 높은 영역의 채널로 small contour를 편입시킨다.
mask_larges[idx[0], :, :] += mask_small
```



HallymMMC 팀 및 solution 설명

- **맴버 (5명):**
  - 한림대학교 소프트웨어융합대학: 허종욱, 신유승, 심보석
  - KAIST 전산학부: 안원혁, 강지현
- **최고 leaderboard score**
  - **0.96277**

- **코드 및 weight 다운로드 링크**
  - **git:** https://github.com/juhou/body-morp-segmentation
  - **케글 notebook:** https://www.kaggle.com/jongukhou/bodymorpseg-hou-master
  - **데이터셋:** https://www.kaggle.com/jongukhou/bodymorpsegmentationmaster
  - **데이터 weights:**
    - https://www.dropbox.com/sh/tspl5h5myx3423s/AACS5RqV59jGtGow6AiEjQoca?dl=0

■감사합니다.

(appendix) 코드 구성 및 폴더구조 아웃라인

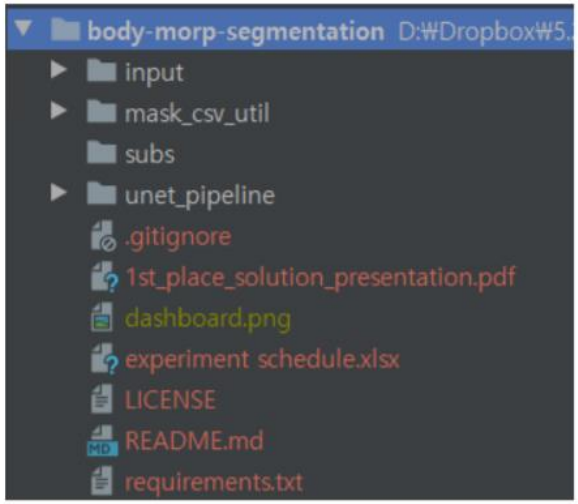
input: 학습/테스트용 데이터

mask_csv_util: 초기 raw 데이터 변환용 코드

subs : 제출용 csv와 prediction 저장 폴더

unet_pipline: 실험 전체 스크립트

나머지 파일들은 참고자료

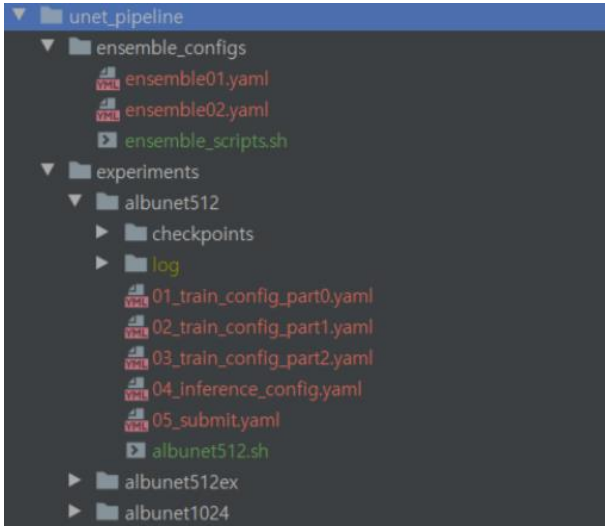


- 모든 데이터는 fold 인덱스를 미리 부여했음
- /unet_pipeline/folds/ *.csv 파일을 통해 관리됨.
- csv 부여하는 코드는 mask_csv_util 에 있음

(appendix) unet_pipline (1)

■ ./ensemble_configs:

앙상블을 위한 yaml 설정파일과 예제 스크립트



■ ./experiments:

모델/세팅 별 학습 스크립트 모음집

■ ./*/checkpoints: fold별 best 모델 저장

■ yaml 파일을 수정해서 실험 설정한 후에 *.sh에 있는 스크립트 실행

(appendix) unet_pipline (2)

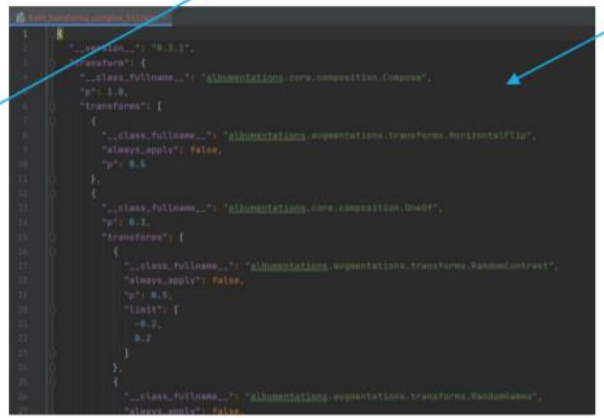
unet_pipline (2)

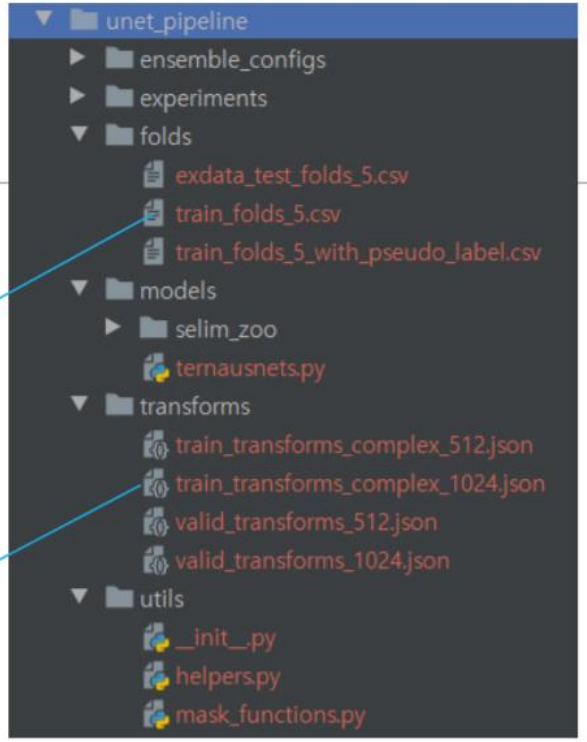
./folds: 학습실험 데이터 fold 관리용 csv

./models: 학습 가능한 모델 (일단 teranusnets만 준비됨)

./transforms: Albumentation 설정 파일

./utils: 각종 util (수정 필요없음)





(appendix) 실험 전체 결과

No.	Experiment name	Public 리더보드	Validation	설명
1	albunet512	0.95931	0.99006	원사이즈 데이터
2	albunet512PL	-	0.989998	슈도레이블(1번 pred) + 후처리 필요
3	albunet512fold10	-	0	
4	albunet512AL_JW	-	0	
5	albunet1024	0.9595	0.985848	
6	albunet1024fold10	0.96007	0.985742	
7	albunet1024fold10albJW	0.96056	0.986304	JW albumentation, 1024사이즈, 10fold
아래부터는 MIN)CONTOUR_AREA 1300 적용한것				
8	albunet1024fold10albJW_2	-	0.986428	JW albumentation, 학습 50epoch 더
9	net1024fold10albJWexpandedL	-	0	앨범JW, 경계제거DB
10	4fold10albJWexpandedLabel_fl	0.94794	0	앨범JW, 경계제거DB, tta scale 추가
11	albunet1024fold10albJW2	학습중	0	앨범JW2(중간복잡), 경계제거DB
12	albunet1024fold10albJW3	학습중	0	앨범JW3(scale강화), 오리지날DB, tta scale
13	albunet1024fold10alb_wh2			원혁 2

No.	Ensemble	Public LB	앙상블내용	설명
1	ensemble567	0.9618	5+6+7	worst를 좀 없애줌
2	ensemble_567_with209	0.96249	6+7	MIN_CONTOUR_AREA 1200
3	ensemble01	0.96277	6+7	MIN_CONTOUR_AREA 1300
4	ensemble6-13	0.96192	6+7+9+11+12+13	
5	ensemble671213	0.96175	6+7+12+13	변형한 DB는 사용 안함