

# 폴더구조 아웃라인

---

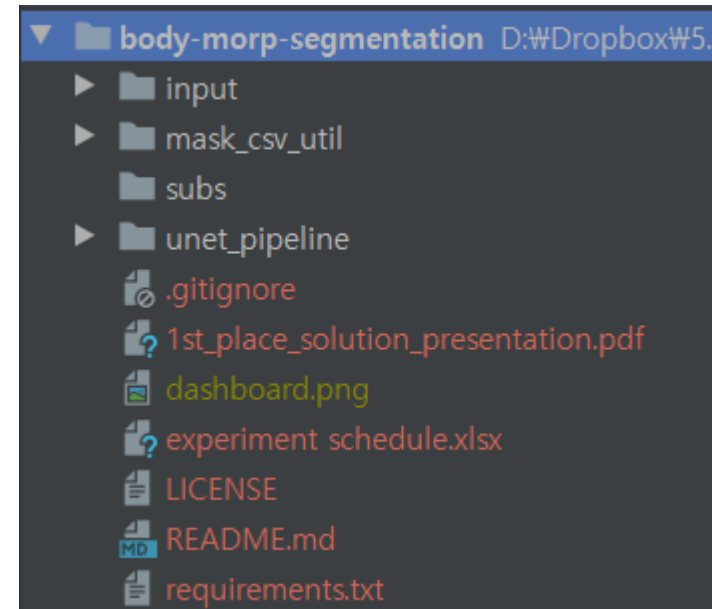
**input:** 학습/테스트용 데이터

**mask\_csv\_util:** 초기 raw 데이터 변환용 코드

**subs :** 제출용 csv와 prediction 저장 폴더

**unet\_pipeline:** 실험 전체 스크립트

나머지 파일들은 참고자료



# input: 학습/테스트용 데이터

---

mask: 학습용 마스크

test: 테스트 데이터

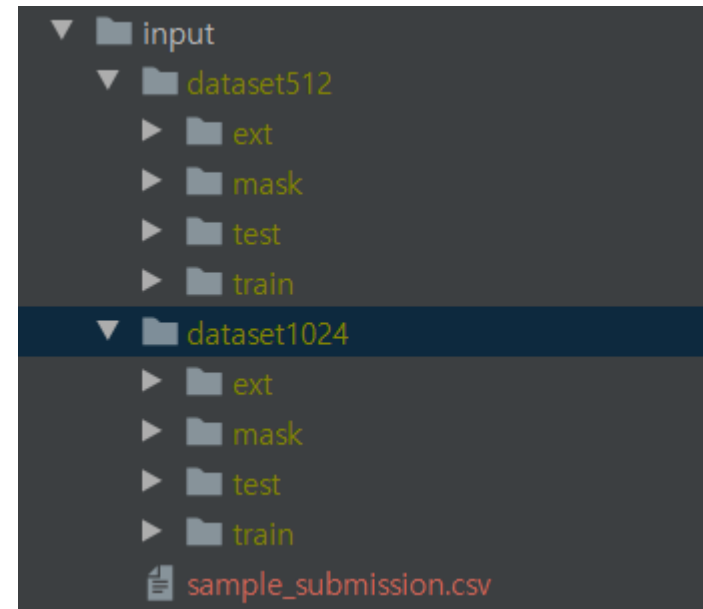
train: 학습용 데이터

ext: 외부 데이터 용 폴더

모든 데이터는 fold 인덱스를 미리 부여했음

**/UNET\_pipeline/folds/ \*.csv** 파일을 통해 관리됨.

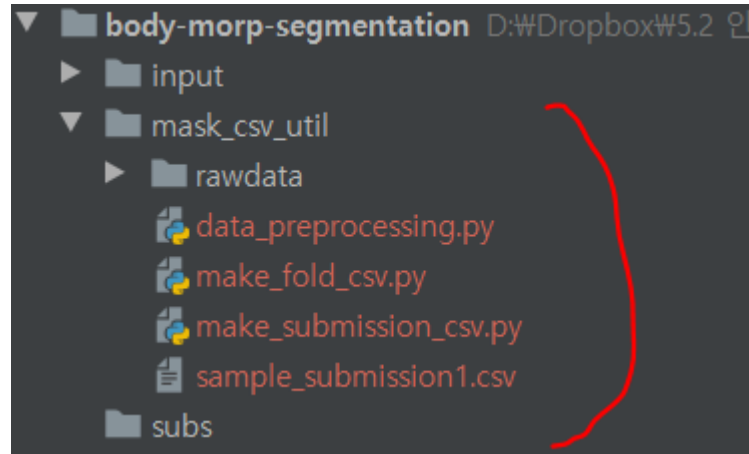
csv 부여하는 코드는 **mask\_csv\_util** 에 있음



# mask\_csv\_util: raw 데이터 변환용

---

mask\_csv\_util: 다운로드 받은 dicom 파일과 mask들을 처리하기 위한 코드가 있음



**subs** : 제출용 csv와 prediction 저장 폴더

# unet\_pipeline (1)

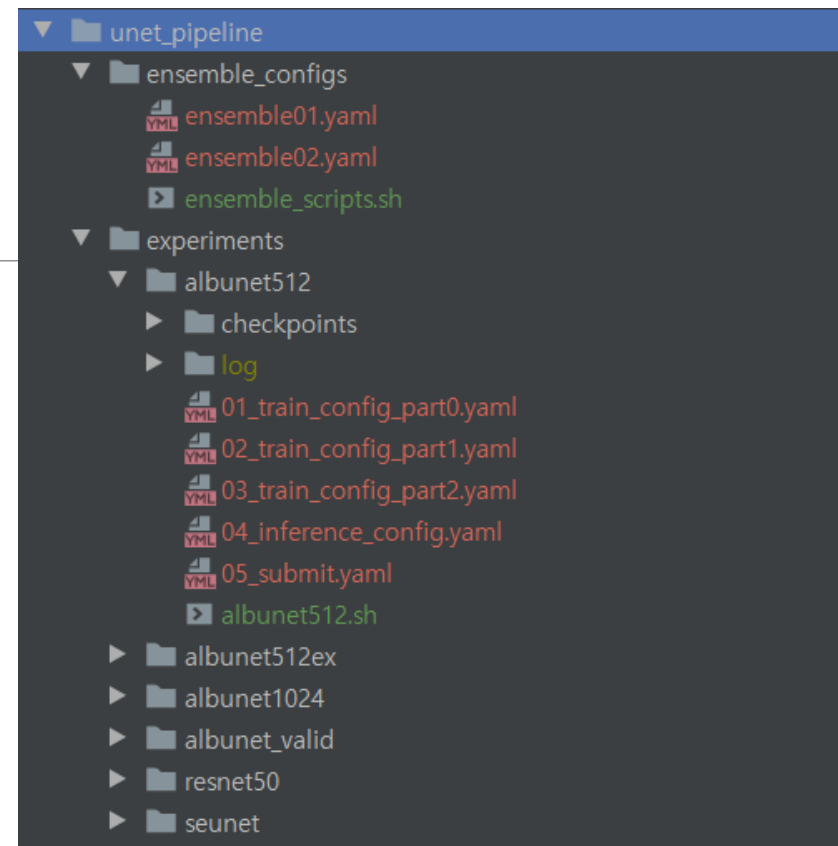
./ensemble\_configs:

앙상블을 위한 yaml 설정파일과 예제 스크립트

./experiments:

모델/세팅 별 학습 스크립트 모음집

./\*/checkpoints: fold 별 best 모델 저장



yaml 파일을 수정해서 실험 설정한 후에 testname.sh에 있는 스크립트 곁어서 실행하면 됨

# unet\_pipeline (2)

./folds: 학습실험 데이터 fold 관리용 csv

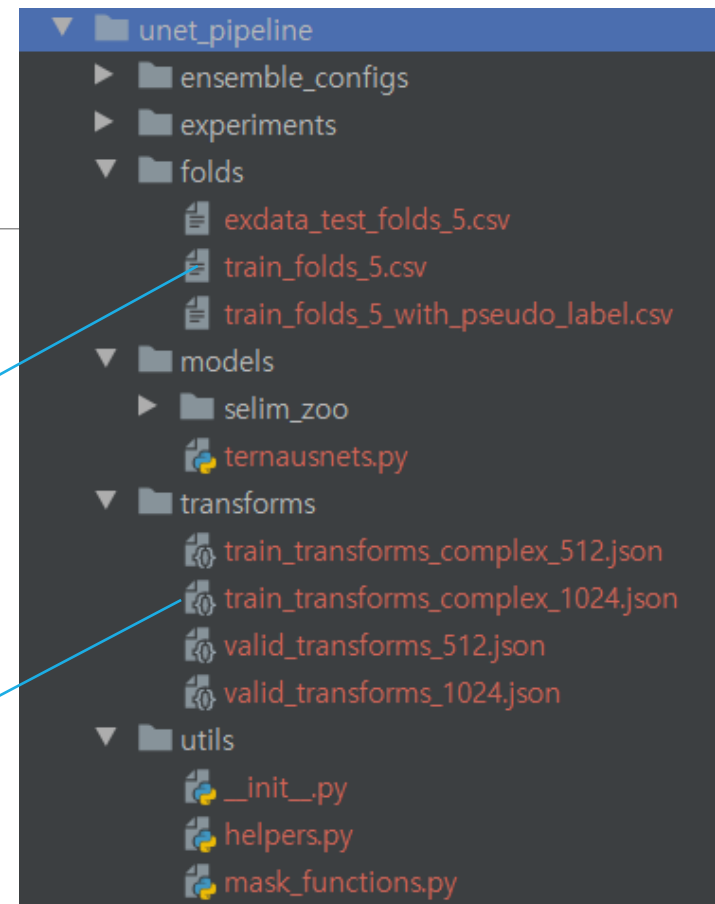
./models: 학습 가능한 모델 (일단 teranusnets만 준비됨)

./transforms: Albumentation 설정 파일

./utils: 각종 util (수정 필요없음)

```
1  fname, fold, exist_labels
2  case001.png, 0, 1
3  case002.png, 1, 1
4  case003.png, 2, 1
5  case004.png, 3, 1
6  case005.png, 4, 1
7  case006.png, 0, 1
8  case007.png, 1, 1
9  case008.png, 2, 1
10 case009.png, 3, 1
11 case010.png, 4, 1
12 case011.png, 0, 1
```

```
1  {
2      "__version__": "0.3.1",
3      "transform": {
4          "__class_fullname__": "albumentations.core.composition.Compose",
5          "p": 1.0,
6          "transforms": [
7              {
8                  "__class_fullname__": "albumentations.augmentations.transforms.HorizontalFlip",
9                  "always_apply": false,
10                 "p": 0.5
11             },
12             {
13                 "__class_fullname__": "albumentations.core.composition.OneOf",
14                 "p": 0.3,
15                 "transforms": [
16                     {
17                         "__class_fullname__": "albumentations.augmentations.transforms.RandomContrast",
18                         "always_apply": false,
19                         "p": 0.5,
20                         "limit": [
21                             -0.2,
22                             0.2
23                         ]
24                     },
25                     {
26                         "__class_fullname__": "albumentations.augmentations.transforms.RandomGamma",
27                         "always_apply": false,
```



# unet\_pipeline (3)

## TripletSubmit.py :

제출용 파일을 만들고, 마스크 후처리를 하는 곳

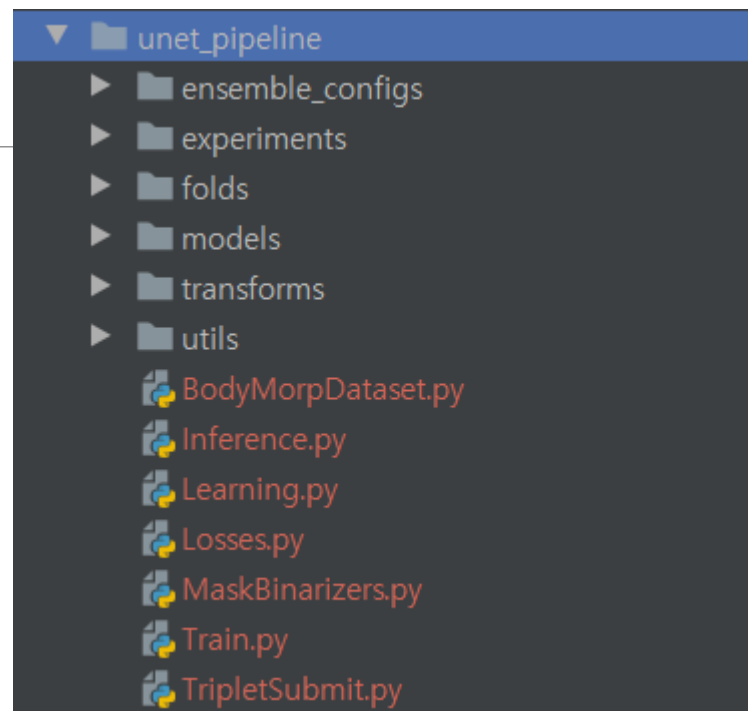
앞으로 이곳 코드를 많이 수정해야 함.

BodyMorpDataset : 데이터셋 클래스

Inference: Prediction을 위한 py

Learning: 학습용 세부 py

Training: ..... 이하 생략



# 학습 과정

총 5단계로 진행된다. exp. 폴더의 .sh 그대로 나란히 실행

	모든 과정에는 early stopping 적용
01	resnet pretrained를 이용, lr를 적당히 큰 값으로 빠르게 학습한다. (여기서 좋은 모델을 몇개 추리기)
02	01에서 얻은 best model을 이용, 적당한 lr로 uptraining.
03	02에서 얻은 best model을 이용, 적당한 lr로 uptraining. (추가설정 있음)
04	03에서 얻은 fold별 모델 5개로 추론
05	04에서 얻은 추론을 하나로 합침. submission용 csv 파일 생성

01\_train\_config\_part0.yaml  
02\_train\_config\_part1.yaml  
03\_train\_config\_part2.yaml  
04\_inference\_config.yaml  
05\_submit.yaml  
albunet512.sh

```
1 #main
2 PIPELINE_NAME: albunet512
3 DATA_DIRECTORY: '../input/dataset512'
4 LOGGER_DIR: 'log'
5
6 SEED: 42
7 DEVICE: cuda
8 DEVICE_LIST: [0]
9 WORKERS: 12
10 USE_EXT_DATA: False
11 # EXT_DATA_FOLD: folds/exdata_test_folds_5.csv
12
13
14 MODEL:
15   PRETRAINED:
16     PIPELINE_PATH: experiments/albunet512
17     CHECKPOINTS_FOLDER: checkpoints
18     PIPELINE_NAME: albunet512
19
20   FREEZE: False
21   PY: models.ternausnets
22   CLASS: AlbuNet
23
24   ARGS:
25     pretrained: True
26     num_classes: 4
27
28 FOLD:
29   NUMBER: 5
30   USEFOLDS: [0,1,2,3,4]
31   FILE: folds/train_folds_5.csv
32
33 # preprocessing - no use
34 USE_SAMPLER: False # True
35 NON_EMPTY_MASK_PROBA: 0.8
36
37 TRAIN_TRANSFORMS:
38   transforms/train_transforms_complex_512.json
39
40 VALID_TRANSFORMS:
41   transforms/valid_transforms_512.json
```

# 학습 전체 스케줄

experiment schedule 엑셀 파일 참조

No.	Experiment name	Public 리더보드	Submit info	Train info	Ave. Fold	Fold 0	Fold 1	Fold 2	Fold 3	Fold 4
1	albunet512			원사이즈 데이터						
2	albunet1024			1024로 늘린 것이 의미가 있을까?						
3	albunet512ex			외부데이터 (슈도레이블 이용)						
4	albunet512album01			앨버멘테이션 ver1						
5	albunet512album02			앨버멘테이션 ver2						
6	...									
7										
8										
9	SeUnet512									
10	SeUnet1024									
11	SeUnet512ex									
12										
13	Resnet50									
...	...									

No.	Ensemble	Public LB	Submit info
1	1+2+3+4+5		
2	1+2+3		
3	1+3+5		
4	...		
5	...		
6	...		



# 향후 진행 방향

---

주어진 테스트 데이터가 final 점수를 계산하기 때문에, 눈으로 얼핏봐도 점수가 나옴.

1. external 데이터 혹은 테스트 데이터로 슈도 레이블 만들어서 추가 학습 진행.
2. 향후 계속 추가