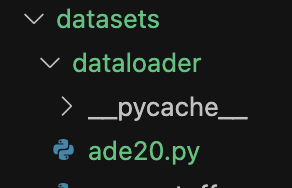
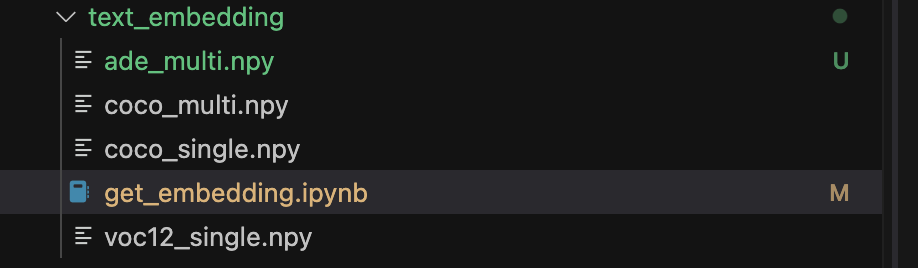
**Read Me**

Mainly based ZegCLIP open source：

（1）ADE20K dataloader：configs/\_base\_/datasets/dataloader/ade20.py

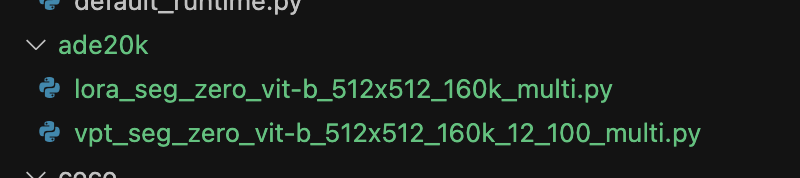


Get classes names embeddings of ADE20K from CLIP encoder: configs/\_base\_/datasets/text\_embedding/get\_embedding.ipynb



The results is saved at configs/\_base\_/datasets/text\_embedding/ade\_multi.npy

(2) Running config on ADE20K:



1. Baseline (ZegCLIP method): configs/ade20k/vpt\_seg\_zero\_vit-b\_512x512\_160k\_12\_100\_multi.py
2. Improved (this project): configs/ade20k/lora\_seg\_zero\_vit-b\_512x512\_160k\_multi.py

(3) Method update:

A screenshot of a computer

Description automatically generated

1. Main Class: models/backbone/img\_encoder.py

A computer screen with text

Description automatically generated

1. The reliable Classes of LoRACLIPVisionTransformer are written in:

models/backbone/loralib.py

models/backbone/utils.py

(4) Follow the README.md to create conda environment, and then run:

Training:

1. bash dist\_train.sh configs/ade20k/vpt\_seg\_zero\_vit-b\_512x512\_160k\_12\_100\_multi.py ./checkpoint/baseline
2. bash dist\_train.sh configs/ade20k/lora\_seg\_zero\_vit-b\_512x512\_160k\_multi.py ./checkpoint/improved

Testing:

1. python test.py configs/ade20k/vpt\_seg\_zero\_vit-b\_512x512\_160k\_12\_100\_multi.py ./checkpoint/baseline/latest.pth –eval=mIoU
2. python test.py configs/ade20k/ lora\_seg\_zero\_vit-b\_512x512\_160k\_multi.py ./checkpoint/baseline/latest.pth –eval=mIoU