HW2
Self-Supervised Learning

Yi-Lun Wu 吳易倫
yilun.ee08@nycu.edu.tw

Submission Deadline:
2022/11/08 08:59 a.m.
Submit to E3

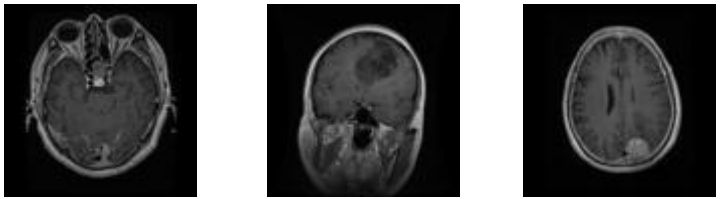
Hard deadline, No extensions

Goals

- Implement self-supervised representation learning method.
- Evaluate and compare the performance of methods.

Dataset

- Brain Magnetic Resonance Imaging (MRI) data, which contains 7294 images of resolution 96x96 in 4 categories.
- You will get:
  1. **unlabeled set**: A full dataset containing 7294 images without labels.
  2. **test set**: A small subset containing 500 images with classification label.



Evaluation Protocol

- You need to learn the image embedding of ***unlabeled set*** from given data.
- TA will evaluate your hand-in embedding of unlabeled set by ***leave-one-out cross validation with KNN***.
- The evaluation metric is the maximum ***accuracy*** over K_neighbor=1, 10, 50, 100.

Grading Scheme

Accuracy (unlabeled set)	**Points**
> 98%	100
> 97%	90
> 95%	80
> 93%	60

- Hint: The evaluation result of **unlabeled set** would be better than the **test set** in most cases.

Spec

- Do this homework in Python 3.6 or later version.
- You can google, but you need to write your own code.
- **Pretrained weights** and **external data** are **forbidden**, so you should train your model from scratch.

Python Package White List

Deeplearning Packages (equipped with autograd engine)			
TensorFlow (include Keras)	PyTorch	MXNet	JAX
Deeplearning Related Packages			
torchvision			
Machine Learning Packages			
sklearn	xgboost		
Others			
pandas	click	tqdm	tensorboardX
tensorboard	python-opencv		

Submission Files (1) - Source Code

- Zip your source code into **StudentID.zip** and submit to E3, where StudentID must be replaced with your student ID.
- You must add a readme file (.pdf, .txt, .md are ok) to clearly state how to create python environment and how to run your code to get the similar result.
- Do not include dataset in **StudentID.zip**, please add the description about dataset path in readme.
- Do not include model weights.

Submission Files (2) - Embedding

- Save your embedding to file **StudentID.npy** and submit to E3, where StudentID must be replaced with your student ID.
- The format of **StudentID.npy** is .npy, so you should convert your embedding to numpy array and save it by function `numpy.save`.
- The dtype of embedding must be `numpy.float32`.
- The embedding size must be 512. The shape of embedding should be (7294, 512). The file size is about 149,38,240 Bytes.
- The order of embedding should follow the filenames (0001.jpg ~ 7294.jpg).
- Tip: You can pad your embedding by zero if the dimension is less than 512, which does not affect the evaluation result of KNN.

Submission Files (2) - Embedding

- The example file 0850726.npy is provided for reference.
- Check the imformation of your **StudentID.npy** before upload it to new e3.

  For example,

```
In [1]: import numpy as np

In [2]: embedding = np.load('0850726.npy')

In [3]: print(embedding.dtype)
float32

In [4]: print(embedding.shape)
(7294, 512)
```