HW2
Self-Supervised Learning

Yi-Lun Wu 吳易倫
yilun.ee08@nycu.edu.tw

Submission Deadline:
2022/4/27 23:59
Submit to E3

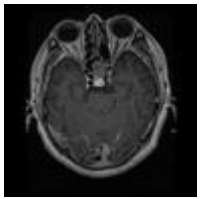
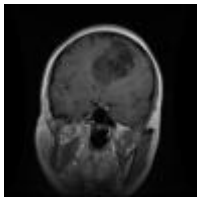
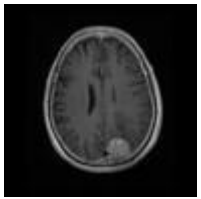
Hard deadline, No extensions

Goals

- Implement self-supervised representation learning methods.
- 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 by leave-one-out cross validation with KNN. The evaluation metric is the maximum **accuracy** over K_neighbor=1, 10, 50, 100.
- It is highly recommended to evaluate your method in the same manner.

Grading Scheme

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

- In general, the evaluation result on **unlabeled set** would be better than the **test set**.

Spec

- Do this homework in Python 3.6 or later version.
- For deep learning model, only fundamental frameworks are allowed to be used (TensorFlow, Keras, PyTorch, etc.).
- Machine learning packages are also allowed to be used (sklearn, xgboost, etc.).
- You can refer to the methods on the internet, but you should write your own code.
- Pretrained weights and external data are **forbidden**, so you should train your model from scratch (random initialization).

Submission Files - 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 location in readme.
- Do not include model weights.

Submission Files - 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 it is recommended to convert your embedding to numpy array and save it by function `numpy.save` .
- The dtype of embedding array must be `numpy.float32`.
- The embedding size must be `512`. The shape of embedding should be (7294, 512). The file size is about 15,724,672 Bytes.
- The order of embedding should follow the filenames (0000.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 - 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)
```

Reminder

- Contact TA Yi-Lun if you have any question about HW2.
  - TA: Yi-Lun Wu 吳易倫
  - yilun.ee08@nycu.edu.tw
  - Send email to other TAs will no response.
- Recommend to use self-supervised learning method to do HW2.
- Share idea, not copy and paste.
- Upload source code, readme and embedding file to E3 before deadline:
  Ex.
    - 0850726.zip (include source code and readme file)
    - 0850726.npy