DL Lab 1: Dataset and DataLoader

Lab Objective:

In this assignment, you should use PyTorch, Dataset and DataLoader in Torch.nn to filter the root of files we assigned.

Dataset Introduction:

CSI_data.json contains three classes: “train”, “val” and “test”. Each class has multiple roots of different files, which formats are “CLASS_NAME/npy/THE_GENDER_AND_COUNT/POSITION/TIME/FILE_NAME”. You need to filter out what we ask you.

Example:

Give me the roots contain “TIME” of “240509”.

Part of answer shows below.

```
▾ root
    0  "val_set/npy/Female1/rand_posi/240509_111208/1715224340287274092"
    1  "val_set/npy/Female1/rand_posi/240509_111208/1715224340334126391"
    2  "val_set/npy/Female1/rand_posi/240509_111208/1715224340381526482"
    3  "val_set/npy/Female1/rand_posi/240509_111208/1715224340428644611"
    4  "val_set/npy/Female1/rand_posi/240509_111208/1715224340477015328"
    5  "val_set/npy/Female1/rand_posi/240509_111208/1715224340526830655"
    6  "val_set/npy/Female1/rand_posi/240509_111208/1715224340583395028"
    7  "val_set/npy/Female1/rand_posi/240509_111208/1715224340630836127"
    8  "val_set/npy/Female1/rand_posi/240509_111208/1715224340678857347"
    9  "val_set/npy/Female1/rand_posi/240509_111208/1715224340727267584"
    10 "val_set/npy/Female1/rand_posi/240509_111208/1715224340775495942"
    11 "val_set/npy/Female1/rand_posi/240509_111208/1715224340824162260"
    12 "val_set/npy/Female1/rand_posi/240509_111208/1715224340873160070"
    13 "val_set/npy/Female1/rand_posi/240509_111208/1715224340925519461"
    14 "val_set/npy/Female1/rand_posi/240509_111208/1715224340973282398"
    15 "val_set/npy/Female1/rand_posi/240509_111208/1715224341020797112"
    16 "val_set/npy/Female1/rand_posi/240509_111208/1715224341068565755"
```

Homework Requirements:

1. “CLASS_NAME” contain “Env3”.
2. “THE_GENDER_AND_COUNT” contain 2 females with no limit on number of males.
3. “THE_GENDER_AND_COUNT” contain 1 female without any male.
4. “TIME” contain from 5/6 18:13:07 to 5/7 23:24:34 (same as 240506_181307 to 240507_232434)
5. “CLASS_NAME” contain “Env3”, “THE_GENDER_AND_COUNT” contain just 1 male, “POSITION” contain “5_posi” and “TIME” from 5/8 09:00 to 5/8 11:00.

Save the answer of each requirement into a individual JSON file, named A1_studentID_studentName_{requirement}.json
So, your submissions will have 5 JSON files, show as below:
A1_studentID_studentName_1.json, A1_studentID_studentName_2.json, A1_studentID_studentName_3.json, A1_studentID_studentName_4.json A1_studentID_studentName_5.json.

Rules:
[1] You should implement homework by yourself. This assignment should be done individually. Please do not plagiarize the assignment. Once the T.A. finds plagiarism, these students will receive a score of zero on the assignment.
[2] Only PyTorch are allowed in this lab.
[3] If the assignment format and files are not in accordance with the regulations, the assignment score × 0.9.
[4] If the assignment is missing or incomplete training for any item, the assignment score will be deducted proportionally to the incompleteness.
[5] If you submit your assignment late, your score will be multiplied by 0.9 for each day of delay.

Submission:
[1] Please submit your code and answer JSON file, the filename should be A1_studentID_studentName.ipynb and JSON file mentioned before. And compress them into a ZIP file, the filename should be A1_studentID_studentName.zip.
[2] Implement a Dataset and DataLoader.
[3] A Report in PDF format and filename should be A1_studentID_studentName.pdf. The report needs to explain how you design the Dataset and Dataloader, and why the answer should be this. The Report has most 2 pages.

Descriptions of code:
Your code must be submitted following the format TAs give you. Any additional sections can also be included.

Assignment Evaluation:
Code & model performances (40%)
Report (60%)

Tips:
Clearly check the “Root name of file”.

Deadline: 9/29 5:20 AM

----------TA Contact information----------

Sheng Yang mail: sk774325.ai12@nycu.edu.tw

Kuan-Lin Chen mail: andy90123.ai12@nycu.edu.tw